Discounted Degrees: Employer Frictions and the Misallocation of High-Skilled Immigrants

Willy Chen*

November 2026
Latest Version†

Abstract

Immigrants in the US are more often over-educated for their jobs than natives, but the reason is not well understood. I study one explanation: a cost that employers bear when they hire an immigrant. When firms cannot pay an immigrant less than a native in the same job, the cost falls on the job instead of the wage. In a signaling model with this constraint, immigrants keep their degrees and take jobs that require less education, and the gap shrinks with the degree. I test these predictions with a LinkedIn panel that follows college graduates across countries. I find that in their first US job, immigrant master's holders are 3.1 percentage points more likely than comparable US-educated workers to be over-educated if they hold a US degree and 9.1 points more likely if they hold only foreign degrees. The gap is smaller for doctorate holders and largest, 20.9 points, for foreign-trained physicians, lawyers, and other professionals who must relicense. Among workers who stay, foreign-trained doctorate and professional-degree holders start in occupations that pay less than those of comparable US-educated workers and close that deficit within ten years, whereas master's holders keep a small premium. The 2008 extension of STEM Optional Practical Training lengthened work authorization without changing training. It raised the share of immigrant STEM master's graduates who started their careers in the US by 3.7 percentage points, and those who stayed more often worked in their field and held occupations paying about 8 log points more, although over-education did not fall. The results suggest that at least part of the hiring cost is created by visa policy and can be removed by it.

Keywords: Occupational Mismatch, Over-Education, Signaling, High-skilled Immigration
JEL: J15, J24, J31, J61, D82

*Michigan State University: willyc@msu.edu

†Acknowledgements: I thank Todd Elder, Alfonso Flores-Lagunes, Raoul Minetti, Stephen Woodbury, and Hanzhe Zhang for their continuing inspiration, advice, and support of this project. All errors are my own.

1 Introduction

College-educated immigrants in the US often work in jobs that require less schooling than they have (Bohn, 2010; Chiswick and Miller, 2009, 2013; Rendon and Cuecuecha, 2010). In 2019, 21 percent of them, about 2 million workers, were unemployed or in jobs requiring no more than a high school diploma, against 16 percent of US-born graduates (Batalova and Fix, 2021). The gap is not confined to immigrants educated abroad: the rate was 24 percent for those educated abroad and 17 percent for those educated in the US. In fact, in the American Community Survey 2010–2019, 42 percent of non-citizens who likely earned their terminal post-secondary degree in the US work in occupations where citizens in the same occupation have a lower most-common degree, as shown in Figure 16.

The literature has measured immigrant over-education and its wage penalty, but the explanations it proposes have largely gone untested. I formalize and test one explanation: a cost that employers bear when they hire an immigrant. I add a generic friction to the canonical education-signaling model (Cho and Kreps, 1987; Mailath, 1987; Spence, 1978), in which firms infer a worker's productivity from their education. Firms observe which workers are immigrants, but prevailing-wage rules and posted pay scales prevent them from paying an immigrant less than a native in the same job. Lang and Manove (2011) show that firms pass group-specific friction to workers through lower wages; barred from doing so to immigrants, firms offer them worse jobs instead. In equilibrium, immigrants keep their degrees and take jobs that natives with less education hold, appearing in the data as immigrant over-education. When the equilibrium wage schedule is convex, the same cost displaces a more educated immigrant by fewer degree levels, so the over-education gap should shrink with the degree. Visa sponsorship is the most visible source of the friction, but the friction in the model can also stand for discrimination against foreign credentials, whether taste-based or statistical, and for language barriers. The size and persistence of these sources likely differ by how a worker entered the US labor market.

I use a panel of LinkedIn profiles that follows college graduates across countries, including those who leave the US. Given that nativity is unobservable on LinkedIn, I classify workers into three groups: Came-to-Study (CtS) workers, who earned their terminal degree at a US institution after a first degree abroad; Came-to-Work (CtW) workers, who hold only foreign degrees and then work in the US; and fully US-educated (FUS) workers, who stand in for natives. I estimate the gap in the likelihood of over-education between immigrants and FUS workers with the same degree, institution tier, and field, and I interpret a positive gap as consistent with a positive friction. Because individual pay also reflects bargaining and preferences, I measure occupational outcomes with occupation-level median wages and revealed-preference rankings of occupations and firms. The estimates describe immigrants who take US jobs, not those who start their careers abroad.

I report four findings. First, I find that the over-education gap at entry shrinks with the degree, as the model predicts. In their first US job, CtS master's holders are 3.1 percentage points more likely than comparable FUS workers to be over-educated, and CtW master's holders are 9.1 points more likely. Pooled across employers, the CtS doctorate gap is -0.6 points, but the pool mixes in universities, which are exempt from the H-1B cap and account for 36 to 40 percent of doctorate first jobs. At employers subject to the cap, CtS doctorate holders are 1.4 points more likely to be over-educated, about half the master's gap of 2.9 points at the same employers. CtW doctorate holders show the same pattern: at cap-subject employers they are 1.7 points more likely to be over-educated, against 8.7 points for CtW master's holders. At these employers, the CtW doctorate gap is close to the CtS gap, whereas a foreign degree predicts a larger one. This may reflect who obtains a visa. Foreign-trained professionals such as physicians and lawyers have the largest gap, 20.9 points, consistent with a second barrier after arrival: they must relicense before they can work in the occupations for which they were trained.

Second, I find that the occupational-pay gap depends on the degree, and that for master's holders over-education does not appear to mean a lower-paying occupation. The occupations of CtS master's holders pay 11.9 log points more at the median than those of FUS master's holders, and those of CtW master's holders pay 5.8 log points more. A Gelbach (2016) decomposition suggests that this premium reflects where these workers sort. For CtS master's holders, job characteristics account for 5.5 log points, and the occupation's rank and demand for H-1B labor for 5.3. The mismatch measures work in the direction the model predicts, but they contribute only -1.2 log points, and little of the gap remains for either pathway once every covariate is held fixed. One reason higher pay can coexist with over-education is that occupational pay need not rise with required schooling above the bachelor's level: master's-level jobs such as teaching often pay less than jobs that require a bachelor's degree (Newburger, 2022). The occupations of CtW doctorate and professional-degree holders instead pay 9.2 and 8.2 log points less, the sign the model predicts, and job characteristics account for 7.4 and 12.1 log points of these deficits. Because the panel does not record individual pay, these are differences in which occupations workers hold, not in pay within an occupation.

Third, I explore whether these deficits fade with US experience, as they would if part of the friction is transient, such as statistical discrimination or the cost of sponsorship. Among workers who stay in the US for ten years, I find that the gaps shrink for foreign-trained doctorate and professional-degree holders. Relative to comparable FUS workers, the occupations of CtW doctorate holders pay 3.9 log points less in year one and 3.4 log points more by year ten. The gap for CtW professional-degree holders moves from -9.7 to 0.7 log points. These stayers start above the -17.3 of all entrants, and their raw gap rises by only 4.7 points, so this path overstates how much the occupations of the group improve. I find little change for master's holders, who start with a

small premium: it moves from 3.3 to 4.0 log points for CtS workers and from 2.6 to 2.2 log points for CtW workers. These paths describe stayers, and they do not separate employer learning from job search, US-specific skills, or the end of sponsorship.

Fourth, I find that US placement responds to work-authorization policy, as the model predicts when the friction falls. In April 2008, the US extended Optional Practical Training (OPT) for F-1 graduates in STEM fields from 12 to 29 months, which delays the point at which an employer must sponsor an H-1B visa. The rule changed how long graduates could work, not what they had learned, so a pure productivity gap should not have moved placement. Relative to non-STEM graduates, the share of CtS STEM master's graduates of the 2008–2011 cohorts who took their first job in the US rose by 3.7 percentage points. Five years after graduation, those who stayed were 7.6 points less likely to work outside their field and held occupations that paid 8 log points more, but they were no less likely to be over-educated: the estimate is 0.5 points. This suggests that the gain lies within an education level, which a measure in degree levels does not register. Retention at year five changed by only -0.1 percentage points, so longer stays do not appear to drive these gains, although the estimates describe those who stayed, not a fixed group of workers. Across the 2008–2021 cohorts, most gains are larger for graduates whose first employer was subject to the H-1B cap, as a lottery channel predicts, though few master's graduates start at exempt employers. Doctorate estimates are similar, but their pre-period coefficients are less stable, so I treat them as suggestive.

The paper makes three contributions, reviewed in Section 1.1. First, it formalizes the screening explanation of immigrant over-education (Chiswick and Miller, 2009). In Lang and Manove (2011), a group-specific friction lowers the group's wage. I show that when the wage cannot fall, the friction lowers the job instead, and the shape of the wage schedule predicts how the gap varies with the degree. Second, I separate, at labor-market entry, the part of the gap tied to a foreign credential from the part that all immigrants face. Foreign schooling is valued below domestic schooling (Friedberg, 2000), but immigrants with foreign degrees also need sponsorship, so the two are hard to tell apart. Because the panel records where each degree was earned, I can compare immigrants with foreign and with US degrees, who both need sponsorship. Because it follows fixed cohorts for ten years, I can also follow each group's gap with US experience; cross-sectional estimates confound such assimilation with changes in cohort quality (Borjas, 1985). Third, the literature on high-skilled visa policy measures effects on firms and on who stays; I measure whether graduates' US jobs match their education.

This matters for welfare. If immigrants are placed below their degrees because they are less productive, the displacement is efficient; if they are placed there because of a cost created by visa rules, it is a misallocation of talent in the sense of Hsieh et al. (2019). A rule that changed only the length of work authorization moved both entry and placement, suggesting that at least part of the friction is of the second kind.

After a discussion of the related literature, the paper proceeds as follows: Section 2 presents the model extension, Section 3 the construction, validation, and measurement of the data, and Section 4 the empirical strategy. Sections 5 to 7 report results at entry, over the first decade, and under the STEM-OPT extension. Section 8 concludes, and Appendix J discusses welfare.

1.1 Related Literature

The paper relates to three strands of literature: the mechanisms of over-education, the recognition of foreign credentials, and high-skilled immigration policy in the US.

1.1.1 Immigrant Over-Education: Mechanisms and Consequences

Over-education is common and it is costly. Across 296 estimates from 39 countries, McGuinness et al. (2018) report an average over-education incidence of 21.5 to 25.9 percent and, across 74 estimates, an average wage penalty of 13.6 percent. They distinguish vertical mismatch, where attained schooling exceeds or falls short of what the job requires, from horizontal mismatch between a worker's field of study and occupation; I measure both. A recent survey names imperfect information about worker skill as one of the two primary sources of mismatch, alongside search frictions (Coraggio et al., 2026); the friction studied in this paper is of the first kind.

Over-educated workers earn less than adequately matched workers with the same schooling (Duncan and Hoffman, 1981; Verdugo and Verdugo, 1989). The evidence is mixed on whether the penalty persists. Over-educated workers in the PSID move up more often, consistent with a stepping stone (Sicherman, 1991), whereas among Flemish school-leavers accepting such a job lowers the transition rate into an adequate one by 51 to 98 percent (Baert et al., 2013). For immigrants, a stepping stone predicts that the gap closes with US experience; a trap predicts that it does not.

Following work on the general labor force (Hartog, 2000; Kiker et al., 1997; Verdugo and Verdugo, 1989), studies of immigrants document larger and more persistent penalties (Chiswick and Miller, 2009, 2013; Li and Lu, 2023; Lu and Hou, 2020; Mattoo et al., 2008; Warman and Worswick, 2015), in the US and in Europe (Aleksynska and Tritah, 2013; Nieto et al., 2013; Visintin et al., 2015; Weber et al., 2024). Cassidy and Gaulke (2024) report a 56 percent increase in the US immigrant over-education penalty between 1993 and 2019, and Lu and Li (2021) attribute a substantial share of nativity sorting gaps to educational mismatch. Using the 2000 Census, Chiswick and Miller (2008) find that immigrants and natives in occupations matched to their schooling earn the same return to it. Immigrants' lower overall return comes mainly from under-education, which they link to favorable selection. They link over-education to the imperfect transferability of foreign skills but leave open why so many immigrants work in occupations whose typical schooling is well below their own. Chiswick and Miller (2009) find that immigrants are more often over-educated

the more work experience they acquired before migrating, which they attribute to US employers' uncertainty about foreign experience, and less often over-educated the longer they have lived in the US. They add a screening hypothesis to the search, human capital, and assignment accounts reviewed by Hartog (2000), but they only state it verbally. They interpret their evidence as favorable selection that is realized once immigrants overcome the imperfect transferability of their skills.

It is also unsettled whether over-education reflects unobserved skill gaps. Among UK graduates it partly does: graduates who are over-educated and dissatisfied with the match have a much lower skill endowment than matched graduates and a pay penalty of 22 to 26 percent, against 5 to 11 percent for those whose skills resemble matched graduates' (Chevalier, 2003; Chevalier and Lindley, 2009). In the US PIAAC, however, first-generation immigrants are more often over-educated than natives, but over-educated workers score above average in literacy and numeracy, and numeracy does not predict the odds of mismatch once other characteristics are held fixed (Pivovarova and Powers, 2022). In short, the work on immigrants establishes the size and persistence of the penalty but not why immigrants end up in jobs below their education.

The closest precedent for a group-specific friction in a signaling model is Lang and Manove (2011): employers observe Black workers' productivity less precisely, so Black workers acquire more education than white workers of similar ability. In their model, competitive firms pay each worker their expected productivity given education and the noisy signal, so the friction operates entirely through the wage, and there are no occupations. For immigrants, prevailing-wage rules keep a firm from discounting the wage within an occupation, so the friction moves the worker into a job below their education. I test this prediction with the gap by terminal degree in Section 5, which is larger for master's than for doctorate holders.

1.1.2 Foreign Credentials, Assimilation, and Employer Learning

Where a credential was earned shapes what it is worth. Foreign schooling and experience are valued below their domestic equivalents in Israel (Friedberg, 2000); in the US, immigrants with US schooling earn more and earn higher returns on their foreign schooling (Bratsberg and Ragan, 2002); and in a Toronto resume audit, foreign experience or Indian, Pakistani, Chinese, or Greek names lowered callbacks, and listing language fluency, multinational-firm experience, or education from highly selective schools did not close the gap (Oreopoulos, 2011). Immigrants who entered on student or temporary work visas outperform natives, largely through education and field (Hunt, 2011). My comparison of CtS and CtW workers follows the same logic, but I ask how each group's jobs compare with its own education over the career.

Language and cultural adaptation are the other leading account: immigrant earnings rise with English proficiency, especially in occupations that require it (Chiswick and Miller, 2010), and longer US residence reduces over-education (Chiswick and Miller, 2009). Cross-sectional ev-

idence indexed by years since arrival, however, confounds assimilation with changes in cohort quality (Borjas, 1985), and selective emigration overstates assimilation among those who remain (Lubotsky, 2007). My estimates of assimilation condition on cohort and are restricted to workers who stay in the US for ten years, so they describe stayers.

Studies of the task content of immigrants' jobs reach a similar conclusion. Male immigrants to Canada move from source-country occupations that require high cognitive and low manual skills into first Canadian jobs that require the opposite; the gap is small for immigrants fluent in English or French and narrows with time in Canada for the rest (Imai et al., 2019). In the US, immigrants' jobs require fewer analytical and interactive tasks and more manual tasks than natives' jobs, and the gap widened across arrival cohorts as occupational assimilation slowed, least so for immigrants with graduate degrees (Cassidy, 2019). Both studies point to language proficiency as a main source of the gap, a transferability deficit that should be small for fluent, highly educated immigrants and should narrow with time in the host country. The workers in this paper are highly educated, and many hold US degrees, so a language deficit is unlikely to explain much of their gap.

Employer learning offers a distinct account: wages come to depend less on schooling and more on productivity that employers learn over time (Altonji and Pierret, 2001), and how much there is to learn depends on the credential: for US college graduates, ability appears to be observed nearly perfectly at entry, while for high school graduates it is revealed gradually (Arcidiacono et al., 2010). Applied to immigrants, the gap should close fastest for the group whose credentials employers evaluate least accurately at entry. Comparing immigrants with foreign and with US degrees isolates that group: both need sponsorship, but only the first hold credentials that US employers cannot readily evaluate. In Section 6, I find that the occupational-pay deficits of foreign-trained doctorate and professional-degree holders close within ten years, and that the deficit of US-trained doctorate holders closes about four-fifths as far. Because the outcome there is the median wage of the occupation held, this closing coincides with occupational mobility, which job search, US-specific skills, and the end of sponsorship at permanent residence can also produce.

1.1.3 High-Skilled Immigration Policy

Research on US high-skilled visa policy has focused on firms and natives. H-1B admissions raised immigrant science and engineering employment and patenting, with limited effects on natives (Kerr and Lincoln, 2010), and an additional H-1B visa crowds out about 1.5 other workers at the firm (Doran et al., 2022). The 2008 STEM-OPT extension raised the likelihood that international STEM graduates initially stay in the US (Beine et al., 2023; Demirci, 2019) and foreign students' choice of STEM majors (Amuedo-Dorantes et al., 2019; Khoo, 2025); Guo et al. (2024) and Kwon et al. (2026) study its effects on natives. This work does not measure how well the jobs these graduates take match their education. In Section 7, I estimate the extension's effect on that outcome.

2 Conceptual Framework

The mismatch literature has proposed employer screening as a cause of immigrant over-education (Chiswick and Miller, 2009; Leuven and Oosterbeek, 2011) but has not modeled it. Signaling models explain over-education in the general population: pooling equilibria can leave workers over-educated when ability is imperfectly correlated with schooling costs (Ordine and Rose, 2009), and job assignment can itself signal ability to outside employers (Waldman, 1984). Neither model has a friction that applies to one group of workers and not another. I start from the canonical signaling model with a unique separating equilibrium under regularity conditions (Cho and Kreps, 1987; Mailath, 1987; Spence, 1978), and add immigrants as a small share of players who arrive after the natives' equilibrium has formed.

2.1 Separating Equilibrium of Natives

As in Spence (1978) and Mailath (1987), two identical firms compete for a unit measure of workers.

- Each worker has a private innate productivity type $\theta \in \Theta \subset \mathbb{R}_+$.
- Each worker chooses an observable education $e \in \mathcal{E} \subset \mathbb{R}_+$.
- The cost of education attainment is represented by the function $C(\theta, e)$ such that $\frac{\partial}{\partial \theta} C < 0$, $\frac{\partial}{\partial e} C > 0$, $\frac{\partial^2}{\partial e^2} C > 0$, and $\frac{\partial^2}{\partial \theta \partial e} C < 0$.
- The production of a worker in a firm is represented by the function $F(\theta, e)$ such that $\frac{\partial}{\partial \theta} F, \frac{\partial}{\partial e} F, \frac{\partial^2}{\partial \theta \partial e} F > 0, \forall \theta \in \Theta$.[1]
- Firms simultaneously offer workers a wage W after observing e.
- The worker's utility function is $U(\theta, e, W) = W - C(\theta, e)$.
- The firm's profit function is $\pi(\theta, e, W) = F(\theta, e) - W$.

Suppose that firms cannot perfectly observe each worker's type. Observing a worker's education e, the firm forms a belief $\widehat{\Theta}(e) \subseteq \Theta$ about the worker's type and offers them $W = E\left[F(\theta, e) \mid \theta \in \widehat{\Theta}(e)\right]$. Let $\sigma : \Theta \to \mathcal{E}$ denote a worker's strategy for education attainment.

Definition 1 (Incentive Compatibility)**.** *The strategy σ is incentive compatible if*

$$\sigma(\theta) \in \underset{e}{argmax}\ U\left(\theta, e, W(e)\right),\ \forall \theta \in \Theta \tag{1}$$

[1] The original model in Spence (1978) assumes that signals are costly and unproductive, i.e., $\frac{\partial}{\partial e} F(\theta, e) = 0$. Relaxing this assumption does not affect the existence of separating equilibria (Mas-Colell et al., 1995).

The game has many equilibria, pooling and separating (Spence, 1978); with reasonable beliefs, only the most efficient separating equilibrium survives (Cho and Kreps, 1987). In that case, σ is one-to-one, so firms perfectly infer each worker's type from their education and pay their productivity: $W(e) = F(\sigma^{-1}(e), e)$. Let $\bar{e} \equiv \sigma(\bar{\theta})$ denote the education the top type chooses in this equilibrium, so that a type-$\bar{\theta}$ native holds the contract $W(\bar{e})$.

Assumption 1 (Reasonable Beliefs of Education Signal)**.** *In equilibrium, a type-θ_w worker can only be inferred as θ_w if they choose education $\sigma(\theta_w) = e_w > \max\{\sigma(\theta) \mid \theta < \theta_w\}$, where σ is the incentive compatible separating equilibrium strategy that survives the Cho-Kreps intuitive criterion.*

Mailath (1987) proves that a continuous separating equilibrium exists in this environment (Appendix B; Appendix B.1 derives the schedule graphically). Because foreign education and the need for sponsorship are observable, I take the natives' separating equilibrium as given and let immigrants choose among its contracts.

2.2 Extending the Model with Immigrant Workers

Suppose that immigrants draw types from the same distribution Θ as natives. A firm that hires an immigrant bears a hiring friction $d > 0$, so an immigrant of type θ with education e is worth $F(\theta, e) - d$ to the firm, against $F(\theta, e)$ for an otherwise identical native; hiring them into the job that pays $W(e)$ is profitable only if $F(\theta, e) - d \geq W(e)$. For intuition, let F be linear in e for each type, as drawn in Figure 1, and let $\sup \Theta = \bar{\theta}$ and $\inf \Theta = \underline{\theta}$. I characterize the contract of the top type $\bar{\theta}$ and then compare it with that of a lower type.

The friction d represents any friction that lowers the net value of a foreign worker to the firm relative to an otherwise identical native. The most visible component is visa sponsorship. Work authorization for high-skilled immigrants is rationed through capped programs and time-limited permits, and sponsorship fees, legal costs, and lottery risk are closer to a fixed cost per hire than to a share of output. As such, d enters in levels and shifts the immigrant worker's net production function down in parallel to that of an otherwise identical native. Doran et al. (2022) compare firms that won and lost H-1B lotteries and find that an additional visa crowds out other hires, which suggests that sponsored workers are costly and imperfectly substitutable. Uncertain work authorization also lowers the firm's return to training the worker. A second component is discrimination, taste-based (Becker, 1957) or statistical, when a foreign credential is a noisier signal of productivity (Altonji and Pierret, 2001; Friedberg, 2000).[2] A third is a genuine productivity difference, for instance

[2] A level shift fits this component less well: the cost of a noisier signal plausibly rises with the worker's productivity, and a friction that grows with the wage would work against the prediction below that the gap shrinks with the degree.

from language barriers (Imai et al., 2019). Placement alone cannot separate a productivity difference from a cost; the STEM-OPT extension in Section 7, which changed the cost without changing training, separates them in part. The components also differ in persistence. Statistical discrimination should fade as employers learn a worker's productivity, whereas taste-based discrimination is persistent and sponsorship costs last until the worker no longer needs sponsorship.

Assumption 2 (Stable Occupation-Level Posted Wages)**.** *The contract $W(e)$ is tied to the occupation rather than to the individual worker: a firm offering a position with equilibrium contract $W(e)$ pays $W(e)$ to whomever it hires and cannot adjust that wage downward for an immigrant.*

For H-1B workers, the assumption is a legal requirement. An employer sponsoring an H-1B worker must attest in a Labor Condition Application that it will pay at least the higher of the occupation's prevailing wage in the area and the wage it pays its own workers in the same position.[3] The certified offers are consistent with the rule binding: only 2 percent of sponsored offers in the occupations CtS master's holders enter fall below the occupation's tenth wage percentile, against 10 percent for the occupation's own workforce (Section 5.4).[4] Workers on Optional Practical Training are not yet covered, but an employer that intends to keep them must eventually file an H-1B petition, and it anticipates the prevailing wage when it hires. For an OPT worker, the friction is the expected cost of that petition, including exogenous risks of lottery loss; the STEM-OPT extension studied in Section 7 lowers it by delaying the filing and increasing the number of available draws.[5] When the friction cannot be passed on through the wage, the menu of job offers absorbs it.

Under Assumption 2, immigrants choose among the contracts on the natives' equilibrium wage schedule. By Assumption 1, a type-$\bar{\theta}$ immigrant must still choose $\bar{e}$ to be identified as $\bar{\theta}$. Since d enters as a level shift, it leaves the marginal return to education unchanged, so the optimal education of an immigrant of each type θ is the same as that of a native of the same type. Their net value to the firm is then $F(\bar{\theta}, \bar{e}) - d = W(\bar{e}) - d < W(\bar{e})$, so no firm hires them into the job $W(\bar{e})$. Under Bertrand competition, firms bid for them until the job offered is worth their net value. The equilibrium contract is the under-placement contract

$$W(\tilde{e}) = W(\bar{e}) - d, \tag{2}$$

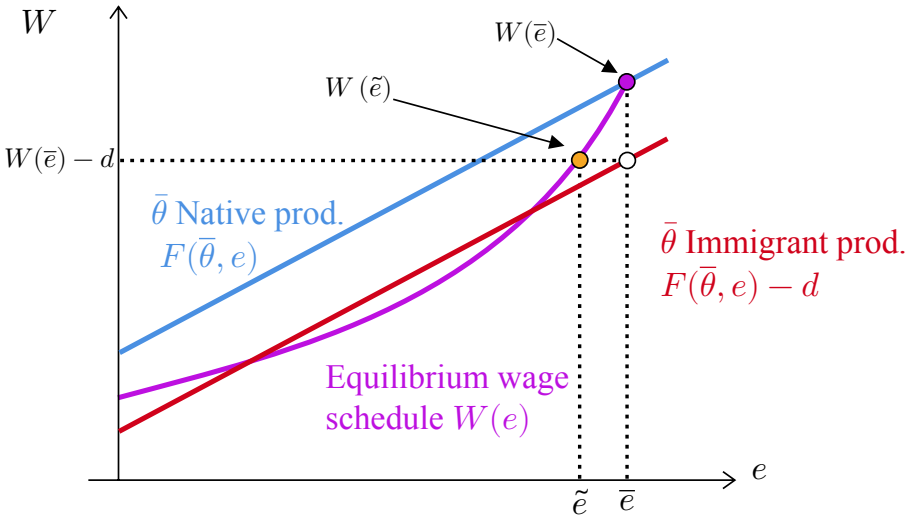
[3] The attestation is tied to the position described in the petition, so an employer cannot certify a worker for one job and assign them to another. A material change in duties, worksite, or pay requires an amended petition, and the Fraud Detection and National Security Directorate of US Citizenship and Immigration Services verifies, through unannounced site visits, that the worker's actual duties, location, hours, and salary match the petition.

[4] Consistent with pay set by the job, immigrants and natives in occupations matched to their schooling earn the same return to it (Chiswick and Miller, 2008).

[5] Internal pay structures, under which a posted position carries one wage whoever fills it, point the same way. Since 2016, employers of STEM-OPT workers must also attest that the worker's pay is commensurate with that of similarly situated US workers, report material changes to the worker's training plan, and accept DHS site visits that check the job against that plan.

a job on the natives' schedule that is typically held by natives with education $\tilde{e} < \bar{e}$ (Figure 1).

Figure 1: Under-Placement of a Type-$\bar{\theta}$ Immigrant



Notes: The blue line is the type-$\bar{\theta}$ native's productivity schedule $F(\bar{\theta}, e)$ and the red line the type-$\bar{\theta}$ immigrant's net productivity $F(\bar{\theta}, e) - d$. The purple curve is the natives' equilibrium wage schedule $W(e)$; each point on it is a job typically held by a worker with education e, and a type-$\bar{\theta}$ native holds $W(\bar{e})$. A type-$\bar{\theta}$ immigrant still chooses $\bar{e}$ to separate but is worth $W(\bar{e}) - d$ to the firm. Because the firm cannot pay them less in the job $W(\bar{e})$ (Assumption 2), it offers the job whose wage equals their net value, $W(\tilde{e}) = W(\bar{e}) - d$, which is typical for education $\tilde{e} < \bar{e}$.

The immigrant holds $\bar{e}$ only to be identified as $\bar{\theta}$; the job itself is one that natives with education $\tilde{e}$ hold. The friction is output the firm loses: its net output from the immigrant is $F(\bar{\theta}, \bar{e}) - d = W(\tilde{e})$, against $F(\bar{\theta}, \bar{e}) = W(\bar{e})$ from an otherwise identical native, which is why the immigrant is offered $W(\tilde{e})$. Appendix J interprets the lost output d in a welfare context.

2.2.1 From Theoretical Friction to Observable Choices

The under-placement contract shows up not in the worker's education but in the distance between their education $\bar{e}$ and the typical education of their job $\tilde{e}$. An immigrant who holds $\bar{e}$ and works at $W(\tilde{e})$ is over-educated relative to a native with the same credential $\bar{e}$.

The contract also predicts how this gap varies with education. Consider a lower type who chooses $\bar{e}' < \bar{e}$ and, as an immigrant, is placed in the job $\tilde{e}'$ with $W(\bar{e}') - W(\tilde{e}') = d$. Because d is the same amount at every education level, when the equilibrium wage schedule is convex, as in Figure 1, the same shortfall spans a shorter stretch of the schedule at higher education, so $\bar{e} - \tilde{e} < \bar{e}' - \tilde{e}'$: the more educated the immigrant, the smaller the distance between their education and that of their job. The over-education gap between immigrants and natives should therefore be positive and shrink with the degree. Convexity is not automatic. Let $\theta(e) \equiv \sigma^{-1}(e)$. Along the separating schedule, $W'(e) = C_e$ and $\theta'(e) = (C_e - F_e)/F_\theta$, so $W''(e) = C_{ee} + C_{\theta e}\,\theta'(e)$, and the schedule is convex wherever $C_{ee}F_\theta \geq -C_{\theta e}\,(C_e - F_e)$: the curvature of the cost of education must outweigh sorting. Appendix B states the conditions of Mailath (1987) under which the separating

schedule exists.

The friction need not be the same for every group of immigrants: a friction in evaluating foreign credentials raises d for workers with foreign degrees. It is even higher for foreign-trained professional-degree holders, who must be relicensed in the US before they can practice.

2.3 Comparative Statics: Lowering the Friction

A lower d for one group of immigrants, holding their types fixed, has two predictions in the model. The first concerns the contract available to a type-$\bar{\theta}$ immigrant who takes a US job. Because the under-placement contract satisfies $W(\tilde{e}) = W(\bar{e}) - d$, lowering d raises $\tilde{e}$ toward $\bar{e}$: the job moves closer to the credential. The improvement can appear in over-education, in whether the worker works in their field, or in the pay and rank of the occupation. Over-education is measured in degree levels, so a small fall in d can improve the job without moving the worker across a degree boundary.

The model does not say which jobs open up. Outside the model, when d falls because work authorization lasts longer, the jobs that open up are likely to be those that require on-the-job training. Following Becker (1962), firms and workers share the cost of firm-specific training so that neither party wants to end the match, and whether they share it depends on how easily they can agree on the worker's value once the training is done (Hashimoto, 1981); where matches last longer, earnings rise more steeply with tenure, consistent with more such training (Hashimoto and Raisian, 1985). An OPT period that expires or an H-1B lottery that is lost ends the match for reasons neither party controls, so jobs that require substantial training are those a firm is least willing to offer an immigrant. Longer work authorization lengthens the expected match and raises the firm's return to training, which moves the immigrant toward these jobs.

The second prediction concerns who takes a US job at all. The model as stated assumes that every immigrant does. Suppose instead that each graduate i also has a best foreign option worth v_i, and takes the US contract only if it is worth at least v_i. Let

$$V(d) = W(\tilde{e}) = W(\bar{e}) - d \tag{3}$$

denote the value of that contract. Because $V(d)$ decreases in d, the share of graduates who take a US job, $\Pr(v_i \leq V(d))$, rises when d falls. The graduates it draws in are those whose foreign option exceeded $V(d)$ before the change and falls short of it after: the graduates whose friction-distorted US offers were closest to their foreign options. If d varies across workers within the group, the marginal graduates need not hold the same jobs as those who would have stayed regardless, so the average job among stayers can move by more or less than the job of any worker who would have stayed. I separate the two in Section 4.3.

3 Data and Measurement

The core source is a panel of worker records assembled by People Data Labs (PDL) from public LinkedIn profiles, which PDL reports reconciling against the other sources in its data union (People Data Labs, 2022b) . The panel consists of each worker's employment history, education, employers, job titles, and locations. I map job titles to six-digit Standard Occupational Classification (SOC) codes with a custom classifier trained on the LinkUp job-title taxonomy (LinkUp, 2025). Occupational median wages come from the Occupational Employment and Wage Statistics (OEWS) (Bureau of Labor Statistics, 2026), job characteristics from O*NET (National Center for O*NET Development, 2025), and H-1B sponsorship measures from Labor Condition Applications (U.S. Department of Labor, 2026). The American Community Survey (ACS) supplies an external benchmark for the coverage of the LinkedIn panel (Ruggles et al., 2025). Integrated Postsecondary Education Data System (IPEDS) admissions and completions records and QS rankings supply the institutional information used to construct education-selectivity measures (Quacquarelli Symonds, 2019; U.S. Department of Education, National Center for Education Statistics, 2019).

The underlying panel contains approximately 44 million individuals who either report US work experience or a post-secondary degree from a US institution and have at least one job that can be assigned a valid SOC. I retain workers with a post-secondary degree and sufficient educational history to classify the location of their first and last post-secondary degrees. The analysis distinguishes three mutually exclusive education pathways. Fully US-educated (FUS) workers completed all observed post-secondary education at US institutions. Came-to-Study (CtS) workers completed their first post-secondary degree outside the US and their terminal degree at a US institution. Came-to-Work (CtW) workers completed all observed post-secondary education outside the US before working in the US. The labels describe education and career histories, not birthplace or citizenship; FUS workers are the analogue to the model's natives.

I group terminal post-secondary degrees into four levels, aligned with the Census classification: bachelor's, master's, doctorate, and professional. Professional degrees, such as the MD, JD, DDS, and PharmD, are degrees in fields that require a US or state license to practice. The doctorate group contains the PhD and other research doctorates, such as the EdD and the DMA, that do not lead to a license. The two groups involve similar schooling; they differ in whether a third party must license the holder before practice, which in the model raises the friction for foreign-trained professionals.

I assign each post-secondary institution to one of nine selectivity tiers, with unmatched institutions in a separate category. US institutions are tiered by an IPEDS index of acceptance rates and test scores, with tier 9 containing the top 1 percent. A foreign institution receives the tier of the US institution with the closest average QS World University Ranking. Each worker carries the tier of their terminal-degree institution, which enters the regressions interacted with the degree. The full

procedure is documented in Appendix H.

The entry analysis uses each worker's first US job. For CtS and FUS workers, it is the first job-year after the terminal degree, kept when it starts within five years after their terminal degree and is located in the US. For CtW workers, it is the first job-year located in the US after the degree, kept when it starts within five years of the degree; non-US experience measures the time between the degree and that job. The five-year window is the same for all three pathways and is observed in full for every graduation cohort from 2003 to 2021. Appendix I extends the CtW window to fifteen years. The longitudinal analyses keep later job-years and record a departure from the US when the profile reports later employment abroad. Of 9.5 million first jobs of CtS and FUS graduate-degree holders, 4.2 million, or 44.5 percent, enter the entry analysis; of 25.9 million first jobs in the CtW–FUS pipeline, which also includes bachelor's degrees, 12.3 million, or 47.6 percent, do. Most of the loss reflects the scope of the analysis rather than missing data. A first job that is abroad or has no recorded location removes 17.9 and 11.2 percent of the initial counts, and a degree completed outside 2003–2021 removes another 22.8 and 22.6 percent. For CtW workers, the initial count already includes only the 5.3 million who hold a US job within fifteen years of the degree; 72.4 percent of them start it within five years, and the location restriction removes none of them. Within this scope, 82.0 percent of CtS–FUS and 82.6 percent of CtW–FUS workers have complete data, and the rate ranges only from 82.0 to 83.8 percent across the four groups. About three-quarters of the remaining loss comes from the employer measures, which require an employer in the connected set of worker flows between firms. Table 2 reports the counts at each stage. The estimates describe college-educated LinkedIn users, not the US labor force.

Taking the data to the model requires the equilibrium wage schedule, and individual wages do not reveal it: realized pay reflects bargaining over rents (Card et al., 2014; Mortensen and Pissarides, 1994), compensating differentials for non-pecuniary attributes (Rosen, 1986), and unobserved preferences (Lachowska et al., 2026), all of which generate dispersion at a given productivity-education pair.[6] Each point on the schedule is an occupation, so I use occupational median wages from the OEWS as the empirical schedule, which removes the variation that comes from individual bargaining and preferences. Revealed-preference rankings of firms and occupations recovered from worker mobility (Sorkin, 2018) capture the non-wage value of a job. Occupational task content enters through the first three principal components of the O*NET skill and ability measures, described in Section 3.2.

[6] Recent structural and empirical work suggests that this confound is large. Taber and Vejlin (2020) estimate a model that nests Roy selection, search frictions, and compensating differentials, and find that non-pecuniary compensating differentials account for a substantial share of the variance of wages alongside human capital. Consistent evidence comes from the search literature: Sullivan and To (2014) and Hall and Mueller (2018) recover dispersion in the non-wage value of jobs that is comparable to, or larger than, the dispersion in offered wages, while the discrete-choice experiment of Mas and Pallais (2017) shows workers forgoing sizable fractions of their wage for amenities such as scheduling flexibility.

3.1 The LinkedIn Panel

I fine-tune a Llama-3.1 language model with 8 billion parameters (*Llama 3.1 (8B)* 2024) to classify job titles, using a proprietary LinkUp dataset with thousands of real job-posting titles for each of the 1,016 detailed O*NET-SOC codes. The classifier assigns a valid eight-digit code to 74,280,970 of 76,084,134 unique titles, or 97.6 percent, with an average confidence of 0.83; I collapse codes to six digits to match the ACS. The 97.6 percent is a coverage rate, not an accuracy rate. Against blind coding of 1,005 US first jobs by a separate language model, the classifier agrees on 56.1 percent of six-digit and 78.8 percent of two-digit occupations, counting a match on the blind coder's alternative code, and its agreement is similar across pathways. These rates are at or above the agreement between two expert coders and that of other automated occupation classifiers in the literature on comparable data (Appendix F, Table F1).

The classifier reads job titles in any language, so the panel keeps the careers of workers who leave the US, and I observe exits directly (Figures 22 and 23). A worker who leaves but does not update their profile appears to stay, with their occupation frozen at the time of exit; this is likely more common in recent graduation cohorts.

3.1.1 Non-random Selection into LinkedIn

LinkedIn users are more educated and more often in white-collar jobs than the US workforce. Master's holders are 31.9 percent of FUS workers in the panel against 26.0 percent in the ACS, and 51.0 percent of FUS workers graduated between 2015 and 2021 (Table 3). Against IPEDS completions, coverage rises with institutional selectivity and is higher in business and technical fields (Appendix D). The estimates are therefore most informative about graduates of selective institutions in these fields.

3.2 O*NET Job Characteristics

A mismatch that begins with a hiring friction can persist through occupation-specific skills (Gathmann and Schönberg, 2010; Lise and Postel-Vinay, 2020), and the distance between a worker's skills and their occupation's requirements lowers wages and slows wage growth (Guvenen et al., 2020). I therefore measure the skill content of occupations from the O*NET skills and abilities. The literature usually sorts these into predetermined categories, such as cognitive, manual, or routine, and takes the first principal component within each (Autor et al., 2003; Cassidy, 2019; Guvenen et al., 2020; Imai et al., 2019; Lise and Postel-Vinay, 2020; Warman and Worswick, 2015). I instead run one principal component analysis on the level scores of all 87 skills and abilities; the first three components account for about 73 percent of the variance.

To label each component, I give a large language model, Microsoft Copilot, the O*NET descriptions of the 20 skills and abilities with the highest loadings on that component (Microsoft, 2025); the labels are descriptive and do not enter the estimation. Table A3 lists the highest-loading skills and abilities for the three components: (1) Cognitive and Interpersonal Competence (CIC); (2) Attention and Coordination (AC); and (3) Practical Job Capabilities (PJC).

3.3 Over-education and Horizontal Mismatch

Measurement of over-education generally follows one of three approaches: (a) job analysis, which relies on objective occupational requirements (e.g., O*NET); (b) worker self-assessment, based on perceived educational needs; and (c) realized matches, which compare individual schooling to the distribution of education within an occupation (Allen and Velden, 2001; Chevalier, 2003; Chiswick and Miller, 2009; Hartog, 2000; McGuinness et al., 2018). Within the realized-match framework, over-education is often defined as schooling exceeding the occupational mean by one standard deviation (Verdugo and Verdugo, 1989) or as holding a degree above the modal level (Kiker et al., 1997). Fredriksson et al. (2018) refine the realized-match benchmark by comparing a worker's skills with those of tenured workers in the same occupation and firm, and find that poor match quality at entry lowers entry wages and predicts later job mobility. I use the job-analysis approach for over-education and a realized-match approach for horizontal mismatch.

Horizontal mismatch has received less attention (Somers et al., 2019); among US graduates, working in a job unrelated to one's major lowers earnings, by different amounts in different fields (Robst, 2007). I classify a job as horizontally matched when it is among the occupations that FUS graduates of the same degree and major most commonly hold. For each degree-by-major group, I order occupations by their share of FUS jobs and take the smallest set that covers at least half of them, including ties at the margin. A share-based cutoff covers the same fraction of FUS workers in every group, so the FUS mismatch rate does not depend on how widely a major's graduates spread. The shares count every job over the career, so an occupation reached through promotion counts as a match.

Figure 17 shows what the definition picks out. Within a degree-by-major group, the occupations outside the matched set pay 27 and 24 log points less at the median than those inside it for FUS bachelor's and master's graduates, and 15 log points less for doctorate and professional-degree graduates, though imprecisely estimated. They also score 0.34 to 0.45 standard deviations lower on cognitive and interpersonal competence, but they do not differ systematically in attention and coordination or in practical capabilities. Mismatched jobs are not uniformly worse: management, which accounts for 14 to 21 percent of mismatched jobs, pays about 30 log points more than the matched set, and excluding it leaves a wage gap of 13 to 20 log points. The two definitions are:

- **Over-education**: Worker i is over-educated in job j if their terminal degree is above the modal entry education requirement of occupation j in O*NET. All occupations are kept, and a worker whose degree is below the requirement is coded as not over-educated.

- **Horizontal mismatch**: Worker i is horizontally mismatched in job j if j is outside the smallest set of occupations that together hold half of the jobs of FUS workers with i's degree and major.

Because most occupations require at most a bachelor's degree for entry, over-education is common among FUS workers with graduate degrees: at the first job, 79.6 percent of FUS master's holders, 54.8 percent of FUS doctorate holders, and 22.4 percent of FUS professional-degree holders are over-educated. Horizontal mismatch is less common: 37.4 percent of FUS master's holders are horizontally mismatched. Because these levels largely reflect how O*NET sets entry requirements, I focus on the gap between immigrants and FUS workers with the same degree. A postdoctoral appointment requires a doctorate, so a doctorate holder in a postdoctoral first job is never over-educated.

The two measures are defined independently for each education-occupation pair. In Section 5, I show that they are correlated but that working outside one's field explains little of the immigrant over-education gap.

3.4 Occupation and Firm Sorting

I construct three measures: a firm-level utility ranking, an occupation-level ranking estimated within degree-major strata, and a descriptive measure of how general a major or occupation is.

Firm PageRank utility I order employers by revealed preference following Sorkin (2018). A worker employed at firm k who receives an offer from firm j accepts if $V_j + \varepsilon_{ijk} > V_k + \varepsilon_{ikk}$, where V_j is a valuation of firm j common to every worker who might compare the two and ε_{ijk} is idiosyncratic taste. Under this choice rule, a firm that draws a disproportionate share of leavers from firms that are themselves valuable must be highly valued; the resulting valuation $\tilde{V}$ is the leading eigenvector of the firm-to-firm transition-share matrix. Because $\tilde{V}$ is built from realized inflows, a firm that hires more workers receives more mass for that reason alone, so $\tilde{V}_j$ measures a firm's draw on movers from valuable firms, which combines value with scale. A size-netted index, $\log V_j^e = \log \tilde{V}_j - \log g_j$ with g_j the firm's size, serves as a robustness check. Appendix C states the construction, each assumption, and what its failure would do to the ranking. I estimate on five-year rolling windows and interpolate the rankings for 2020–2022. All regressions use the rankings in logs (Appendix E).

Within-education SOC PageRank utility The occupation ranking applies the same logic to moves between occupations, separately for workers with the same highest degree and field, because which occupations compete for a worker depends on the credential. Flows within a stratum can be thin, so I add a uniform baseline with a damping factor of 0.85 (Brin and Page, 1998), which keeps every occupation comparable (Appendix C).

Major/SOC generality From the network of moves between majors and occupations, I compute a generality statistic, DUH, for every major and occupation within a degree group (Chen, 2024). A major with high DUH_{major} sends its graduates into many occupations; an occupation with high DUH_{SOC} draws workers from many majors. Robustness checks replace DUH with the Gini-Simpson index (Gini, 1912; Simpson, 1949).

4 Empirical Strategy

This section describes the estimating equations, sample selection, and identifying assumptions used in the three results sections that follow: mismatch and the friction at labor market entry (Section 5), the dynamics of the friction (Section 6), and the STEM-OPT extension (Section 7).

4.1 Identifying Occupational Mismatch at Labor Market Entry

The model's prediction is about mismatch, not wages. At the under-placement contract, point (b) of Figure 4, the worker holds $\bar{e}$ but works in an occupation whose typical requirement is $\tilde{e} < \bar{e}$. Immigrants should therefore be more likely than comparable natives with the same degree to be over-educated, by less at higher degrees. Table 1 lists the predicted sign by degree and pathway, with $\bar{e}$ a doctorate and $\bar{e}'$ a master's degree.

For each mismatch outcome $m \in$ {over-education, horizontal mismatch} and each immigrant pathway, I estimate a linear probability model for worker i in first job j, adding controls in blocks:

$$
\begin{aligned}
Y_{ij}^{m} = \alpha^{m} + \beta^{m} P_i + \sum_{e \neq \text{MA}} \left[\eta_e^{m} \mathbb{1}\{E_i = e\} + \beta_e^{m} P_i \mathbb{1}\{E_i = e\} \right] + \mu_{\text{grad}(i)} + \varepsilon_i^{m} \qquad (4) \\
+ \sum_{e,\tau} \kappa_{e\tau}^{m} \mathbb{1}\{E_i = e\} \times \mathbb{1}\{T_i = \tau\} \qquad Col.(2) \\
+ \lambda_1^{m} \text{STEM}_i + \lambda_2^{m} DUH_{f(i)} + \delta_{e,f(i)}^{m} \qquad Col.(3) \\
+ \phi^{m} F_j + \rho^{m} CE_j + \psi^{m} P_i \times CE_j \qquad Col.(4) \\
+ \xi^{m} M_{ij}^{-m}, \qquad Col.(5)
\end{aligned}
$$

where P_i denotes the CtS (or CtW) pathway and FUS workers are the omitted comparison pathway. E_i is the terminal degree, the counterpart of e in the model, with the master's degree as the reference, and $\mu_{\text{grad}(i)}$ are graduation-cohort fixed effects. In the CtW comparison they are labor-market-entry-cohort fixed effects: the year of the first US job for CtW workers and the graduation year for FUS workers. The CtW specification also controls for non-US experience, the years between the degree and the first US job, and its square, and for whether the first US employer is the same as their last employer abroad. T_i is the tier of the terminal institution; tier effects differ by degree but not by pathway, because in the model a credential carries the same signal wherever it was earned. $f(i)$ is the coarse field of worker i's terminal major: STEM_i indicates a major on the DHS STEM list, and $DUH_{f(i)}$ is the generality of the major described in Appendix E. $\delta_{e,f(i)}^{m}$ are degree-by-major-field fixed effects, so that Column 3 compares an immigrant with FUS workers who hold the same degree, from an institution of the same tier, in the same field of study. The object of interest is the gap for degree e, $\beta^{m} + \beta_e^{m}$; for master's holders it is β^{m}. I take Column 3 as the main estimate,

because every control in it is fixed before the worker enters the labor market.

F_j is a vector containing the log firm PageRank utility, the firm's 3-, 6-, 12-, and 24-month churn rates, and an indicator for whether the firm filed any Labor Condition Application between 2011 and 2025. CE_j indicates that the employer is a higher-education institution, which is exempt from the annual H-1B cap. M_{ij}^{-m} is the complementary mismatch measure: horizontal mismatch when Y_{ij}^m is over-education, and over-education when it is horizontal mismatch; Column 5 is estimated only for these two outcomes. CtW specifications also include non-US experience and its square, an indicator for a same-employer transfer, and labor-market-entry-cohort fixed effects.

Columns 4 and 5 condition on outcomes of the same job match that produces the mismatch. Column 4 tests one alternative explanation: immigrants accept mismatched jobs in exchange for better or visa-sponsoring employers, a compensating differential instead of a friction. If that were the case, the gap would shrink once the employer is held fixed. Because the friction may itself steer immigrants toward sponsoring employers, those who reach high-utility firms are positively selected, which biases the Column 4 gap toward zero. Column 5 asks whether the two types of mismatch move together.

Employers exempt from the H-1B cap need separate treatment. Universities, their affiliated nonprofits, and nonprofit and government research organizations can file petitions at any time of year, and their hires face no lottery.[7] Universities account for 36 to 40 percent of the first jobs of doctorate holders in each pathway, against one to seven percent for master's holders, so pooling employers mixes, for doctorate holders, jobs where the lottery applies with jobs where it does not. Column 4 includes $P_i \times CE_j$, so its pathway coefficients measure the gap at cap-subject employers. I also re-estimate Column 3 separately at cap-subject and cap-exempt first employers, which lets every coefficient differ by employer type.

The first job is the first non-intern, non-volunteer job-year that starts within five years after the terminal degree; for CtW workers, it is the first such job-year in the US. Assistantships and "PhD student" titles do not count, and a worker who only continues a pre-graduation job has no first job. I keep US first jobs of the 2003–2021 graduation cohorts, because both mismatch measures are scored against US benchmarks. All columns use one sample, the workers observed on every covariate in the richest column.

I then compare the rank and pay of the occupations that immigrants and FUS workers hold. Here the question is pay given the type of job, so the occupation's characteristics enter as controls:

[7] 8 U.S.C. §1184(g)(5)(A)–(B). CE_j covers higher-education institutions only.

$$
\begin{aligned}
Y_{ij} = \alpha + \beta P_i + \sum_{e \neq \text{MA}} [\eta_e \mathbb{1}\{E_i = e\} + \beta_e P_i \mathbb{1}\{E_i = e\}] + \mu_{\text{grad}(i)} + \varepsilon_i \qquad (5)\\
+ \zeta_1 OE_{ij} + \zeta_2 HM_{ij} + \zeta_3 OE_{ij} \times HM_{ij} \qquad Col.(2)\\
+ \sum_{e,\tau} \kappa_{e\tau} \mathbb{1}\{E_i = e\} \times \mathbb{1}\{T_i = \tau\} \qquad Col.(3)\\
+ \lambda Z_j + \omega_{o(j)} \qquad Col.(4)\\
+ \phi \ln(\text{Firm Sorkin}_j) \qquad Col.(5)\\
+ \rho CE_j + \psi P_i CE_j, \qquad \text{Last col.}
\end{aligned}
$$

with $Y_{ij} = \ln(\text{SOC PageRank}_j)$ and $Y_{ij} = \ln(\text{OEWS median wage}_j)$. Z_j contains the three standardized O*NET principal components and the standardized importance of certification, and $\omega_{o(j)}$ are fixed effects for the occupation's on-the-job training category. The median-wage regressions add $\ln(\text{SOC PageRank}_j)$ in a further column. These specifications include degree-by-tier effects but not degree-by-field fixed effects. For CtW workers every column also includes labor-market-entry-cohort fixed effects, non-US experience, and a same-employer transfer indicator, so the first column is not a raw gap.

4.1.1 Decomposing the Occupational Wage Gap

To ask which characteristics account for the gap in occupational pay, I use the decomposition of Gelbach (2016), which assigns the change in the gap to blocks of covariates without depending on the order in which they are added. The covariates form seven blocks. *Degree major* comprises indicators for the major groups that define horizontal mismatch. *Worker characteristics* comprise STEM, graduation in 2008 or later and its interaction with STEM, major generality, and school tier, together with years since degree and its square for CtS workers, and non-US labor-market experience and its square and a same-employer transfer indicator for CtW workers. *English-language origin* comprises an indicator that English is an official language of the country of the worker's first post-secondary degree, with the US coded as English-speaking. *Job characteristics* comprise the three skill components of Section 3.2, the importance of certification, and SOC generality. *Firm sorting* comprises the employer's PageRank utility, its churn rates at 3, 6, 12, and 24 months, an indicator that it ever sponsored an H-1B worker, and an indicator for cap-exempt educational employers. *Job sorting* comprises the occupation's PageRank utility and indicators for deciles of its H-1B demand. *Occupational mismatch* comprises over-education and horizontal mismatch, each also interacted with the log education signal.

4.2 Examining the Dynamics of the Hiring Friction

The part of d that reflects employers' uncertainty about foreign credentials should fade as they learn a worker's productivity, whereas sponsorship costs and taste-based discrimination should not. The path of the adjusted gap over a career is informative of what d contains. I measure the gap with an Oaxaca–Blinder decomposition of the occupation's median wage, first at the first US job and then year by year over the first decade. At the first US job, the decomposition also links the two parts of the analysis: set against the decomposition of Section 4.1.1, which holds the occupation's characteristics fixed, it shows how much of the entry gap lies in which occupation a worker reaches.

For each pathway $p \in \{\text{CtS}, \text{CtW}\}$ and terminal degree e, I decompose the immigrant–FUS gap in the residualized log median OEWS wage of the occupation held at the first US job, defined as in Section 4.1:

$$\bar{Y}_{pe} - \bar{Y}_{FUS,e} = \underbrace{\left(\bar{X}_{pe} - \bar{X}_{FUS,e}\right)\hat{\beta}_{FUS,e}}_{\text{Explained}} + \underbrace{\bar{X}_{pe}\left(\hat{\beta}_{pe} - \hat{\beta}_{FUS,e}\right)}_{\text{Unexplained}}. \tag{6}$$

The decomposition is estimated separately for each graduate degree level. The covariate vector X_{ipe} contains degree-by-terminal-institution-tier indicators, field indicators, STEM status, an indicator for graduation after 2008 and its interaction with STEM, terminal-major generality, experience, and experience squared. For CtW workers it additionally contains non-US experience and its square and an indicator for a same-employer transfer.

$$\begin{aligned} X_{ipe} = \sum_{\tau} \kappa_{e\tau} \mathbb{1}\{T_i = \tau\} + \gamma_1 \text{Field}_i + \gamma_2 \text{STEM}_i + \gamma_3 \text{Post2008}_i \\ + \gamma_4(\text{STEM}_i \times \text{Post2008}_i) + \gamma_5 \text{DUHMajor}_i + \gamma_6 \text{Exp}_i + \gamma_7 \text{Exp}_i^2 \\ + \mathbb{1}\{p = \text{CtW}\}\left[\gamma_8 \text{NonUSExp}_i + \gamma_9 \text{NonUSExp}_i^2 + \gamma_{10} \text{Transfer}_i\right]. \end{aligned} \tag{7}$$

For CtS, Y residualizes log OEWS median wages by terminal-degree completion cohort; for CtW, it residualizes by labor-market-entry cohort. Unlike the decomposition in Section 4.1.1, this one excludes occupation, firm, mismatch, and task characteristics, which can themselves result from the friction. I call the unexplained component the adjusted placement gap, U_{pe}. Because $W(\tilde{e}) = W(\bar{e}) - d$, a positive d predicts a negative U_{pe}.

To observe dynamics, I estimate the unexplained component U_{pet}^{stay} for each year since entry into the US labor market, among stayers, workers who remain in the US labor market for ten years. For

each pathway p and degree e, I re-estimate Equation (6) on the job observations held at t:

$$\bar{Y}_{pet} - \bar{Y}_{FUS,et} = \underbrace{\left(\bar{X}_{pet} - \bar{X}_{FUS,et}\right)\hat{\beta}_{FUS,et}}_{\text{Explained}} + \underbrace{\bar{X}_{pet}\left(\hat{\beta}_{pet} - \hat{\beta}_{FUS,et}\right)}_{U^{\text{stay}}_{pet}}, \tag{8}$$

with the same covariates as Equation (6), except that an indicator for holding the first job replaces experience, on the US jobs held at t by workers who graduated in 2003 or later, so that every year is priced with realized OEWS wages. For CtS and FUS workers, t counts years since the terminal degree; for CtW workers, it counts years since their first year in the US labor market. I report the paths from year one, because at $t = 0$ many workers have not yet started their first job.

Many immigrants leave the US: a large share of CtS graduates at every degree leave at graduation, and among those who stay, annual exit hazards are low and fall with tenure (Figure 22); CtW workers exit at higher rates (Figure 23). Fixing the population of stayers and the cohort means that changes in U^{stay}_{pet} follow a fixed population of workers, not changes in cohort quality (Borjas, 1985). A worker still enters the estimate at t only if a US job is observed that year, so employment gaps and temporary spells abroad change who is observed. The paths describe stayers; if lower-earning immigrants are the ones who leave (Lubotsky, 2007), stayers are positively selected.

4.3 Difference-in-Differences: STEM-OPT Extension

Identifying the causal effect of the friction needs a change in the cost of hiring a foreign worker that leaves the worker's productivity unchanged. The April 2008 interim rule extended Optional Practical Training (OPT) for F-1 graduates of STEM programs by 17 months, from 12 to 29, and left non-STEM graduates' OPT unchanged. It lengthened the period in which a firm can hire a STEM graduate without sponsoring an H-1B. Because H-1B petitions are filed once a year against the 85,000-cap, the extension gives a STEM graduate at least one additional filing season before work authorization lapses. That lowers the risk of involuntary separation, which the employer prices into hiring and training decisions (Hashimoto, 1981).[8]

I compare STEM with non-STEM graduates before and after 2008. The two predictions of Section 2.3 map onto two margins: the share of graduates whose first job is in the US should rise, and the jobs of those who stay should move closer to their credentials. Because the extension also changes who stays, the second margin mixes a change in jobs with a change in the composition of stayers (Section 4.3.2). The sample is CtS graduates whose terminal degree is a master's or a doctorate; few professional degrees are STEM-designated. For worker i with terminal degree e, I

[8]The same rule also extended status for all F-1 graduates with a pending H-1B petition (the "cap-gap" provision); because it applies to STEM and non-STEM graduates alike, it is differenced out below. A 2016 rule further lengthened the STEM extension to 24 months for graduates who applied after May 2016, including much of the 2015 cohort, so these graduates receive a larger dose of the same treatment.

estimate separately by degree:

$$\begin{aligned}\text{US_Job}_i = \alpha_e + \beta_e \,\text{STEM}_i + \gamma_e \,\text{Post}_{c(i)} + \delta_e \left(\text{STEM}_i \times \text{Post}_{c(i)}\right) \\ + \sum_{\tau} \kappa_{e\tau} \mathbb{1}\{T_i = \tau\} + \lambda_e \,\text{DUHMajor}_i + \varepsilon_i,\end{aligned} \tag{9}$$

$$\begin{aligned}Y_{it}^k = \alpha_{et}^k + \beta_{et}^k \,\text{STEM}_i + \gamma_{et}^k \,\text{Post}_{c(i)} + \delta_{et}^k \left(\text{STEM}_i \times \text{Post}_{c(i)}\right) \\ + \sum_{\tau} \kappa_{et\tau}^k \mathbb{1}\{T_i = \tau\} + \lambda_{et}^k \,\text{DUHMajor}_i + u_{it}^k, \qquad \text{for } S_{it} = 1.\end{aligned} \tag{10}$$

Here STEM_i indicates that the major of i's terminal degree is on the 2008 STEM Designated Degree Program List,[9] $c(i)$ is the year i completed the terminal degree, $\text{Post}_c = \mathbb{1}\{c \geq 2008\}$, and t is years since the terminal degree. In Equation (9), US_Job_i indicates that the first post-degree job is in the US, and δ_e is the difference-in-differences in that share. In the model, a lower friction raises this share because more graduates' US offers now exceed their foreign options.

On the intensive margin, I estimate Equation (10) separately at each t on graduates whose first job is in the US and who are still US-resident at t ($S_{it} = 1$). The outcomes Y_{it}^k are the residualized log OEWS median wage of the occupation held, over-education, horizontal mismatch, log SOC PageRank score, and log firm PageRank score. I report Equation (10) only at values of t at which every STEM×Post group retains at least 100 workers and at least 20 percent of its $t = 0$ count.

Two further outcomes describe what happens after entry, among graduates whose first job is in the US. Retention, Stay_{it}, indicates that i has not left the US by year t. Employer attachment, NoSwitch_{it}, indicates that i has not yet moved from the first employer to another employer by year t while in the US.

For $m \in \{\text{Stay}, \text{NoSwitch}\}$ I estimate, separately by degree and at each t,

$$\begin{aligned}m_{it} = \alpha_{et}^m + \beta_{et}^m \,\text{STEM}_i + \gamma_{et}^m \,\text{Post}_{c(i)} + \delta_{et}^m \left(\text{STEM}_i \times \text{Post}_{c(i)}\right) \\ + \sum_{\tau} \kappa_{et\tau}^m \mathbb{1}\{T_i = \tau\} + \lambda_{et}^m \,\text{DUHMajor}_i + v_{it}^m,\end{aligned} \tag{11}$$

on graduates whose status at t is known: those who have left by t and those still observed at t.

A graduate whose profile ends before t without an exit is dropped from Equation (11) at t. For employer attachment this makes little difference, but for retention the dropped graduates are disproportionately stayers in recent cohorts, whose profiles are observed less often in the last years

[9] I assign STEM status from the terminal degree's own major. A graduate whose earlier degree is in a STEM field but whose master's or doctorate is not, such as an engineer who completes a master's in management, is coded non-STEM, because under the 2008 rule the extension applied to the degree on which OPT was granted. From 2016, a graduate can also qualify on a STEM degree earned at a US institution in the previous ten years. In this case, some post-2016 graduates coded non-STEM may become eligible, which biases the estimates for those cohorts toward zero.

of the data. I report retention with the Kaplan–Meier estimator: within each STEM×Post group, the share still in the US at t is the product of one-year survival rates $\prod_{j \le t} \left(1 - \text{exits}_j / \text{at risk}_j\right)$, where a graduate whose profile ends counts as at risk until the last year observed. The retention difference-in-differences is the same two-by-two contrast of these shares. A positive $\delta_{et}^{\text{NoSwitch}}$ means that the extension reduced employer switching.

The comparison cohorts graduated between 2003 and 2006. I drop the 2007 cohort because eligibility for the extension turned on the date a worker's initial OPT period ended, not on graduation year, so 2007 graduates still on OPT in April 2008 were partly treated. I report three post-period windows: all cohorts graduating in 2008 or later, a local window of 2008–2011 graduates, and a window of 2012–2015 graduates that places the post-period after the financial crisis.

Standard errors. STEM status is assigned by major: every graduate of a given major is either eligible or not, according to the 2008 list. I therefore cluster standard errors by terminal-degree major, of which there are about 550 in the master's samples and 415 in the doctorate samples, in the event study, the stayer outcomes, and employer switching (Abadie et al., 2023). For retention, I report Greenwood standard errors for the Kaplan–Meier shares, which treat graduates as independent.

4.3.1 Identifying Assumptions

Interpreting δ_e as the causal effect of a lower hiring friction requires three assumptions.

Assumption 3 (Parallel trends)**.** *Absent the extension, the share of STEM graduates taking a first job in the US would have evolved in parallel with the share of non-STEM graduates, conditional on tier and major generality.*

Assumption 3 is not testable in the post-period, but it has a testable implication in the pre-period, which I examine with the event study in Equation (12) below.

Assumption 4 (No differential concurrent shocks)**.** *Nothing else that changed around 2008 moved the US prospects of STEM graduates relative to non-STEM graduates.*

Two events could violate Assumption 4: the 2008–2009 recession and the H-1B cap itself. If STEM employment was more resilient to the recession, δ_e would overstate the policy effect; the 2012–2015 window, which leaves out the recession cohorts entirely, is designed to address this. The recession can also bias the intensive-margin estimates. Graduating into a downturn raises the probability of over-education (Summerfield and Theodossiou, 2017), pushes new hires into less complex jobs (Baley et al., 2022), and lowers earnings for a decade or more (Kahn, 2010; Oreopoulos et al., 2012; Schwandt and Wachter, 2019). If this entry penalty was smaller for STEM

graduates, the 2008–2011 cohorts would show a STEM advantage in the jobs they take with no change in d, so δ^k_{et} would overstate the policy effect; the 2012–2015 window addresses this channel as well. The H-1B cap was not reached in the first week of filing for fiscal years 2010 through 2013, so for the cohorts filing in those years an additional filing season was worth less than when the cap binds. This works against finding an effect in the local 2008–2011 window.

Assumption 5 (Stable treatment assignment)**.** *The composition of STEM and non-STEM graduates does not respond to the policy.*

For the 2008–2011 cohorts this is plausible, since most of them had chosen a program before the April 2008 rule. It is less plausible for the 2012–2015 cohorts, who enrolled after the extension was known: if foreign students who intended to stay in the US shifted into STEM programs, the share of STEM graduates starting in the US would rise even with no change in hiring friction. This response is documented: Amuedo-Dorantes et al. (2019) find that foreign students became 18 percent more likely to major in STEM after 2008, in part through non-STEM bachelor's holders entering STEM master's programs, and Khoo (2025) finds higher STEM enrollment and completion among international students at Ohio public universities. The 2012–2015 window therefore addresses Assumption 4 at the cost of weakening Assumption 5, and the two windows are best interpreted together rather than as one robustness check.

Event study. To examine the timing of the extensive-margin effect and pre-trends, I replace the single post indicator with a full set of graduation-cohort fixed effects and interact STEM with each cohort:

$$\begin{aligned}\text{US_Job}_i =& \mu_{c(i)} + \beta_e \,\text{STEM}_i + \sum_{c \neq 2006} \phi_{ec} \,\text{STEM}_i \times \mathbb{1}\{c(i) = c\} \\ &+ \sum_{\tau} \kappa_{e\tau} \mathbb{1}\{T_i = \tau\} + \lambda_e \,\text{DUHMajor}_i + \varepsilon_i,\end{aligned} \tag{12}$$

with the 2006 cohort as the reference and 2007 excluded. The coefficients ϕ_{ec} for $c \leq 2005$ test Assumption 3: under parallel trends they should be close to zero, with a break at 2008, and I test them jointly with a Wald test. The coefficients for $c \geq 2008$ trace how the effect evolves across cohorts, including any change for the later cohorts, who received the 24-month extension.

4.3.2 Selection on the Intensive Margin

Equation (10) conditions on $S_{it} = 1$, a US first job and US residence at t, both of which can respond to the treatment: the extension raised the share of STEM graduates who start in the US (Figure 12) and, by less, how long master's graduates stay (Section 7.2).

Assumption 6 (Monotone selection)**.** *The extension never causes a worker who would otherwise have stayed to leave:* $S_{it}(1) \geq S_{it}(0)$.

Under Assumption 6, treated stayers are either always-stayers, with $S_{it}(0) = S_{it}(1) = 1$, or marginal stayers, with $S_{it}(0) = 0$ and $S_{it}(1) = 1$; let $\omega_t = \Pr(S_{it}(0) = 1 \,|\, S_{it}(1) = 1,\, D_i = 1)$ denote the always-stayer share.

Assumption 7 (Parallel trends among would-be stayers)**.** *Absent the extension, the mean outcome among STEM graduates who would have stayed would have changed across the 2008 cutoff by the same amount as the mean outcome among non-STEM graduates who would have stayed:* $E[Y_{it}^k(0) \mid S_{it}(0) = 1, \text{STEM}, \text{Post}] - E[Y_{it}^k(0) \mid S_{it}(0) = 1, \text{STEM}, \text{Pre}]$ *equals the corresponding change for non-STEM graduates.*

Assumption 7 is stated on stayers because the estimator uses only stayers, but it restricts two things at once: how untreated outcomes evolve, and how the composition of would-be stayers evolves. Appendix G.3 shows what this requires and how it can be checked.

Under Assumption 7, the difference-in-differences estimator identifies

$$\delta_{et}^k = E\big[Y_{it}^k(1) \,\big|\, S_{it}(1) = 1, D_i = 1\big] - E\big[Y_{it}^k(0) \,\big|\, S_{it}(0) = 1, D_i = 1\big]\,. \tag{13}$$

Writing the first term as an ω_t-weighted average over always-stayers and marginal stayers, and noting that under Assumption 6, $S_{it}(0) = 1$ selects exactly the always-stayers, gives

$$\begin{aligned}\delta_{et}^k = {} & \underbrace{E\big[Y_{it}^k(1) - Y_{it}^k(0) \,\big|\, \text{always-stayer},\, D_i = 1\big]}_{\text{effect on always-stayers}} \\ & + \underbrace{(1-\omega_t)\Big(E\big[Y_{it}^k(1) \,\big|\, \text{marginal},\, D_i = 1\big] - E\big[Y_{it}^k(1) \,\big|\, \text{always-stayer},\, D_i = 1\big]\Big)}_{\text{composition}}\,. \end{aligned} \tag{14}$$

Appendix G derives Equation (14) step by step. The composition term is zero only if marginal stayers hold jobs of the same quality as always-stayers, which the model implies when d is the same for every worker in the group. If d varies across workers, the sign of the term depends on how the friction and the foreign option are jointly distributed. If the foreign option is independent of d and has a log-concave distribution, marginal stayers faced on average a larger friction than always-stayers and hold lower-quality jobs; the composition term is then negative, and a positive $\widehat{\delta}_{et}^k$ understates the effect on always-stayers. Without such a restriction the sign is not determined. I therefore interpret Equation (10) as the effect of the extension on the job quality of the population that remains in the US, not as the effect on a fixed set of workers.

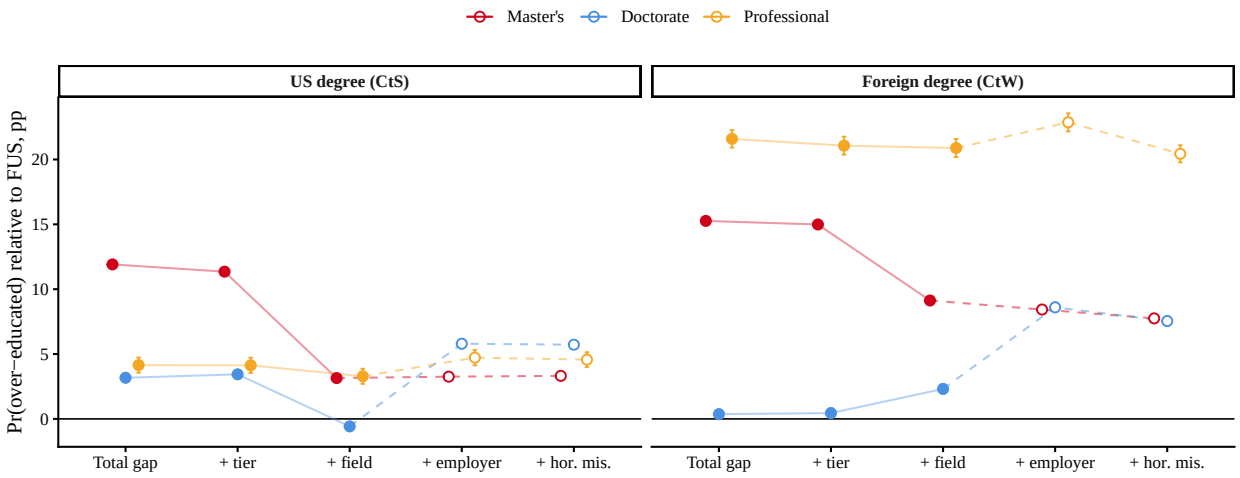
5 Occupational Mismatch at Labor Market Entry

This section reports what happens at the first US job: where immigrants sit relative to the contracts in Figure 4, and how the rank and pay of the occupations they hold differ from those of FUS workers.

5.1 Mismatched Contracts

Figures 2 and 3 summarize the over-education and horizontal-mismatch results, and Tables 6 through 9 report the estimates. Each point in the figures is the gap between an immigrant pathway and FUS workers with the same terminal degree, $\beta^m + \beta^m_e$, as blocks of controls are added.

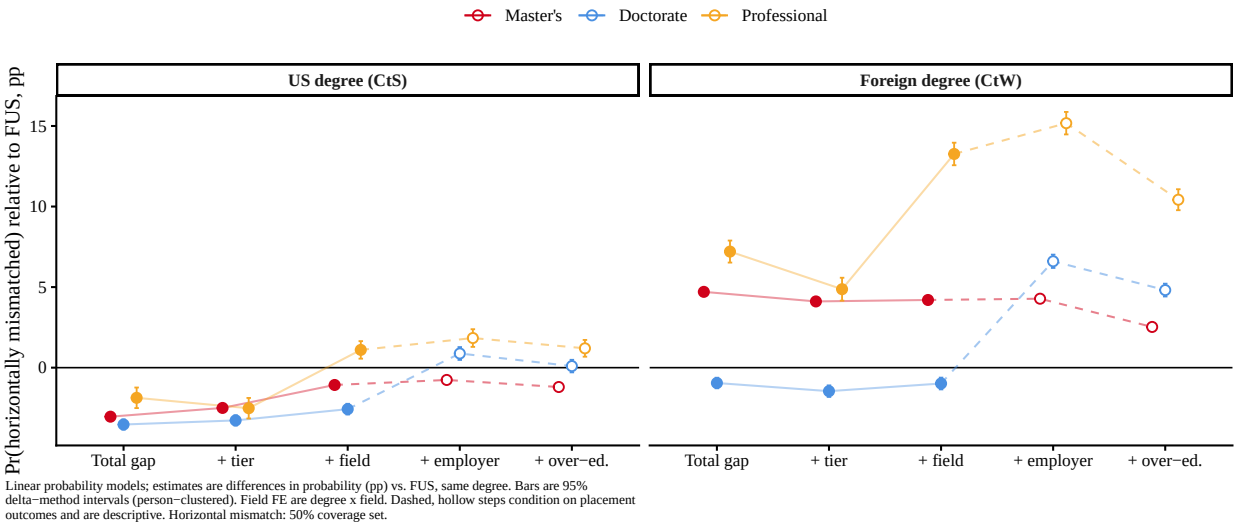
Figure 2: Over-Education Gap at the First US Job, Relative to FUS Workers with the Same Degree



Notes: Estimates of $\beta^{OE} + \beta^{OE}_e$ from Equation (4), in percentage points, with 95% confidence intervals. Moving right, the specification adds degree-by-tier effects, degree-by-field fixed effects with STEM and major generality, the employer block, and horizontal mismatch. Dashed, hollow steps condition on placement outcomes and are descriptive; from "+ employer" on, they are gaps at cap-subject employers. Doctorate pools the PhD and non-licensure doctorates; professional denotes licensure doctorates.

CtS Workers Without controls, CtS master's holders are 11.9 percentage points more likely than FUS master's holders to be over-educated at their first US job (Table 6). Institution tier barely changes the gap, but holding the field fixed reduces it to 3.1 points in Column 3, which suggests that most of the raw gap reflects field-sorting. For doctorate holders, the Column 3 gap is −0.6 points, small but statistically significant; for professional-degree holders it is 3.3 points. CtS master's and doctorate holders are 1.1 and 2.6 points less likely than FUS workers to work outside their field, and CtS professional-degree holders 1.1 points more likely (Table 7). Optional Practical Training and the H-1B both require a job related to the degree, which may explain part of this advantage.

Figure 3: Horizontal-Mismatch Gap at the First US Job, Relative to FUS Workers with the Same Degree



Notes: Same as in Figure 2, for horizontal mismatch. The last step adds over-education.

CtW Workers CtW master's holders are 9.1 percentage points more likely than FUS master's holders to be over-educated in Column 3, about three times the CtS gap (Table 8). CtW doctorate holders are 2.3 points more likely. CtW professional-degree holders are 20.9 points more likely, the largest gap of any group. Horizontal mismatch follows the same order (Table 9): 4.2 points for master's holders, −1.0 for doctorate holders, and 13.3 for professional-degree holders.

Employer Sorting Column 4 of Tables 6 and 8 adds firm utility, churn, H-1B filing, and cap-exempt status. The master's over-education gaps barely move, to 3.3 points for CtS and 8.4 for CtW, which provides little evidence that immigrants trade job match for better or sponsoring employers. If the friction itself steers immigrants toward such employers, those who reach high-utility firms are positively selected, which pushes the Column 4 gap toward zero; the stable gap is therefore, if anything, conservative.

Interpreting the Estimates through the Framework The positive master's gaps in both pathways are consistent with the under-placement contract, point (b) of Figure 4. Pooled across employers, the doctorate gaps are −0.6 points for CtS and 2.3 for CtW, but the pool includes universities, where the H-1B lottery does not apply. At cap-subject first employers (Table 10), CtS doctorate holders are 1.4 percentage points more likely than FUS doctorate holders to be over-educated, about half the CtS master's gap of 2.9 points at the same employers. A gap that is positive and smaller at the higher degree is consistent with the convexity of the wage schedule. CtW doctorate holders show the same decline: at cap-subject employers their gap is 1.7 points, against 8.7 points for CtW

master's holders. Their gap is, however, close to the CtS doctorate gap, whereas Table 1 predicts a larger gap for a foreign degree. CtW professional-degree holders have the largest gap of any group, 20.9 points, whereas the CtS professional gap, 3.3 points, is about the same as the CtS master's gap, not smaller as Table 1 predicts.

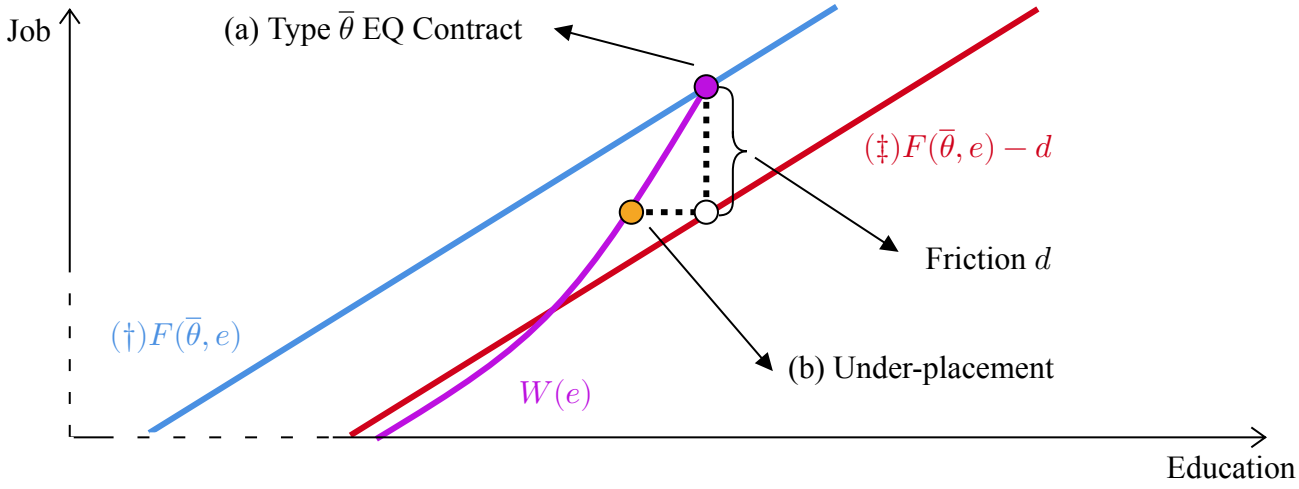
For master's and professional-degree holders, the gap is larger with a foreign degree than with a US degree, as a friction in evaluating foreign credentials would predict; at cap-subject employers, the doctorate gaps differ little. And conditioning on horizontal mismatch in Column 5, which enters positively at 8.4 and 16.0 points, leaves the master's over-education gaps almost unchanged, at 3.3 points for CtS and 7.7 for CtW, which suggests that working outside one's field explains little of this gap.

Selection at Two Gates I hypothesize that the CtW professional gap of 20.9 points and the low occupational pay of CtW doctorate holders in Section 6 reflect who reaches the sample, whereas the model assumes that immigrants draw types from the same distribution as natives. The first gate is entry into the US labor market. A CtS graduate can start work on Optional Practical Training without an employer petition or a lottery. A CtW worker arrives on an employment-based visa (an H-1B petition, a cap-exempt employer, an intra-company transfer, or a visa for extraordinary ability) or on a family-based or other visa that does not depend on an employer's assessment. The mix of channels likely differs by degree.

For doctorate holders trained abroad, employment-based channels are likely to dominate. This gate may explain why their over-education gap is close to the CtS gap at cap-subject employers, although the CtW master's gap at those employers is three times the CtS gap. In Appendix I, the pooled CtW gap rises by 6.3 points when the window adds workers who arrive six to fifteen years after the degree, which suggests that early arrivals are more selected. The gate also appears to show in occupational pay: at the first US job, the adjusted gap is -9.3 log points for CtW doctorate holders and -2.3 for CtS doctorate holders (Section 6). This is consistent with the gate steering CtW doctorate holders toward academic and research positions, which require a doctorate but pay little. Universities account for a similar share of doctorate first jobs in every pathway, 36 to 40 percent, but post-doctoral scholar appointments, at universities and elsewhere, account for 37 percent of CtW doctorate first jobs, against 27 percent for CtS and 21 percent for FUS doctorate holders. CtW professional-degree holders face a second gate after they arrive. A doctorate earned abroad can be used in a US job immediately; a foreign medical, law, or dental degree cannot until its holder passes US licensing examinations and, for a physician, completes a US residency. A holder who does not relicense works in an occupation that does not require the license and is over-educated by construction. Family-based visas, which do not screen on whether the credential can be used, are also more likely to admit such workers than an employer recruiting for a licensed position. Part of

the CtW professional gap may therefore reflect licensing rules rather than employers' assessments (Mattoo et al., 2008). Because relicensing is costly and those who complete it are a selected group (Kugler and Sauer, 2005), this part need not fade with US tenure. The panel records neither visa type nor licensure, so neither gate is observed directly.

Figure 4: Separating Equilibrium with Immigrants



Notes: $(\dagger)$ marks the type-$\bar{\theta}$ native worker's net productivity schedule $F(\bar{\theta}, e)$. $(\ddagger)$ marks the type-$\bar{\theta}$ immigrant worker's net productivity schedule $F(\bar{\theta}, e) - d$. (a) marks the equilibrium contract for type-$\bar{\theta}$ native workers $W(\bar{e})$. (b) marks the under-placement contract for type-$\bar{\theta}$ immigrant workers, who hold $\bar{e}$ but work a job typical for $\tilde{e} < \bar{e}$ at $W(\tilde{e}) = W(\bar{e}) - d$; the vertical distance is the friction d.

5.2 Occupational Sorting

Tables 11 through 14 report estimates of Equation (5) for the rank (log PageRank score) and the pay (log OEWS median wage) of the first US occupation. In the full specification, the occupations of CtS master's holders rank 21.9 log points higher and pay 1.5 log points more than those of FUS master's holders; for CtW master's holders the figures are 11.5 and 1.6. Immigrant master's holders thus hold higher-ranked occupations even with mismatch, tier, tasks, and employer held fixed, which suggests that over-education does not move them down the occupational ranking. The clearest deficit at any degree is that of CtW professional-degree holders, whose occupations rank 49.7 log points below those of FUS professional-degree holders. Task content matters for pay: for CtW master's holders, the gap in occupational pay falls from 7.9 to 2.9 log points when Column 4 adds the occupation's training category and task content.

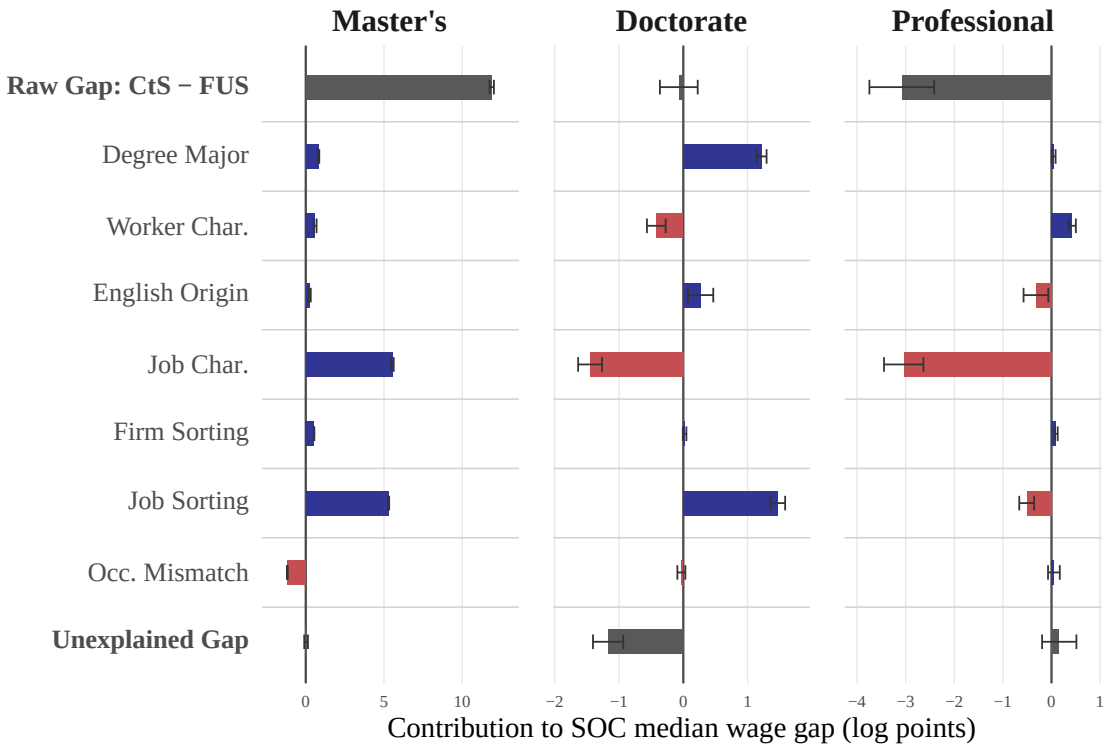
5.3 What Accounts for the Occupational Wage Gap

Figures 5 and 6 decompose the gap in the occupation's log OEWS median wage into the contributions of the seven blocks defined in Section 4.1.1; the last bar in each panel is the gap left once

every block is held fixed. The blocks include characteristics of the occupation and employer, so the decomposition describes match characteristics that contribute to the gaps, not why workers entered those occupations.

The occupations of CtS master's holders pay 11.9 log points more than those of FUS master's holders. Job characteristics account for 5.5 log points, job sorting (the occupation's rank and its demand for H-1B labor) for 5.3, and mismatch for −1.2, so little of the gap remains unaccounted for. For CtS doctorate holders, the raw gap is −0.1 log points, with job sorting contributing 1.5 and job characteristics −1.5. For CtS professional-degree holders, job characteristics account for the entire gap of −3.1 points. Tables A5 and A6 report the contribution of each covariate.

Figure 5: Decomposing First Job-Year Sorting Gap Drivers For CtS Workers



Notes: Gelbach (2016) decomposition of the gap between CtS and FUS workers with the same terminal degree in the log OEWS median wage of the occupation held at the first US job. Bars are block contributions in log points with 95 percent intervals; the last bar is the gap with every block held fixed. Doctorate pools the PhD and non-licensure doctorates.

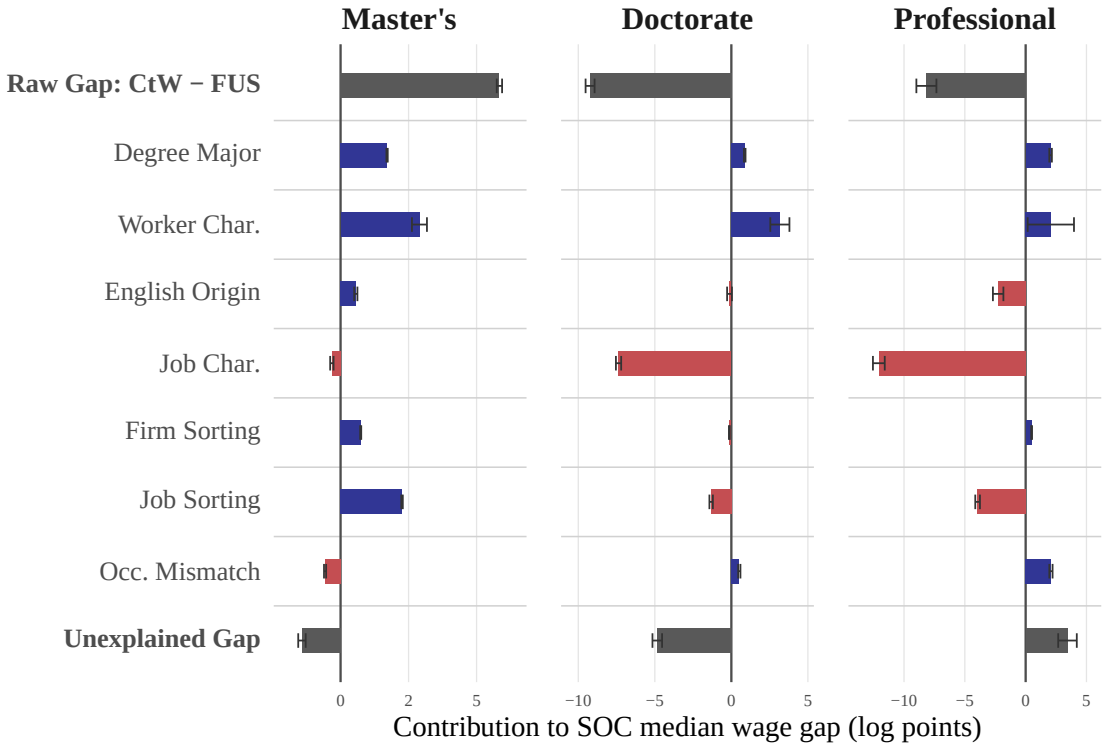
The CtW gaps are 5.8 log points for master's holders, −9.2 for doctorate holders, and −8.2 for professional-degree holders.

For master's holders, the covariates more than account for the gap, leaving −1.4 log points. For doctorate and professional-degree holders, job characteristics account for −7.4 and −12.1 log points, and −4.8 and +3.4 points remain. Like CtS master's holders, CtW master's holders are over-educated and in higher-paying occupations. This is possible because occupational pay need not rise with required schooling above the bachelor's level: master's-level occupations such as

teaching, counseling, and social work often pay less than occupations that require a bachelor's degree (Newburger, 2022), a feature that the model's single wage schedule abstracts from.

Once job characteristics, the occupation's rank, and the employer are held fixed, little of the gap remains for CtS workers at any degree or for CtW master's holders. This suggests that the master's premium reflects where these workers sort. The outcome is the occupation's median wage, so the decomposition cannot detect a discount in pay within an occupation, which Section 5.4 examines. Foreign credentials are valued below their US equivalents (Friedberg, 2000; Oreopoulos, 2011), and if that discount operates here, it appears to work through which occupations CtW workers reach: for CtW doctorate holders, job characteristics alone account for -7.4 log points. The adjusted placement gap in Section 6, which does not hold these characteristics fixed, measures this margin. The remaining CtW doctorate gap of -4.8 points can also reflect unobserved differences in skill.

Figure 6: Decomposing First Job-Year Sorting Gap Drivers For CtW Workers



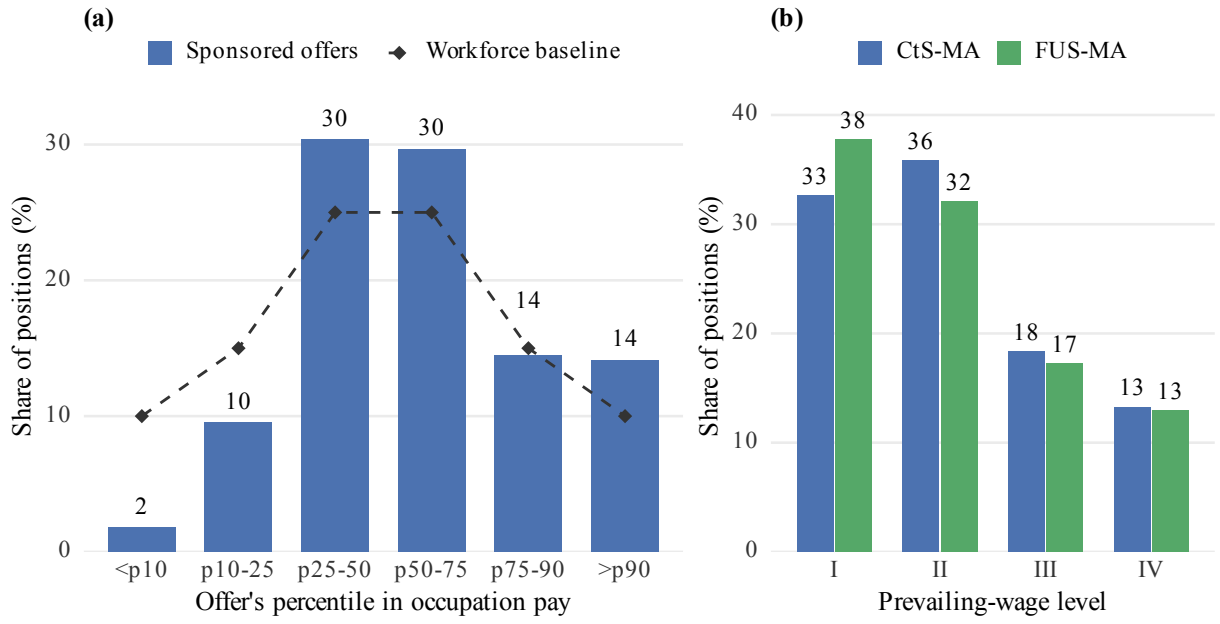
Notes: As in Figure 5, for CtW workers relative to FUS workers with the same terminal degree. Worker characteristics additionally include non-US experience and a same-employer transfer indicator.

5.4 Is the Friction a Discounted Wage? Evidence from the H-1B Wage Floor

This subsection examines Assumption 2. If employers could pay sponsored workers less for the same job, CtS master's holders might hold high-wage occupations only because they are cheaper,

and sponsored offers in those occupations would concentrate at the bottom of each occupation's wage distribution. Labor Condition Application (LCA) disclosures record the wage offered for each certified H-1B position. Panel (a) of Figure 7 places every certified offer in the national OEWS wage distribution of its occupation, weighting occupations by how often CtS master's holders enter them, and compares the result with the distribution of the occupation's own workforce. Only two percent of offers fall below the tenth percentile, and 11 percent below the twenty-fifth; the rest lie above, mostly between the twenty-fifth and seventy-fifth percentiles. This provides little evidence that sponsored workers are hired as cheap substitutes at the bottom of the occupation.

Figure 7: **H-1B sponsored offers are not concentrated at the occupation wage floor.**



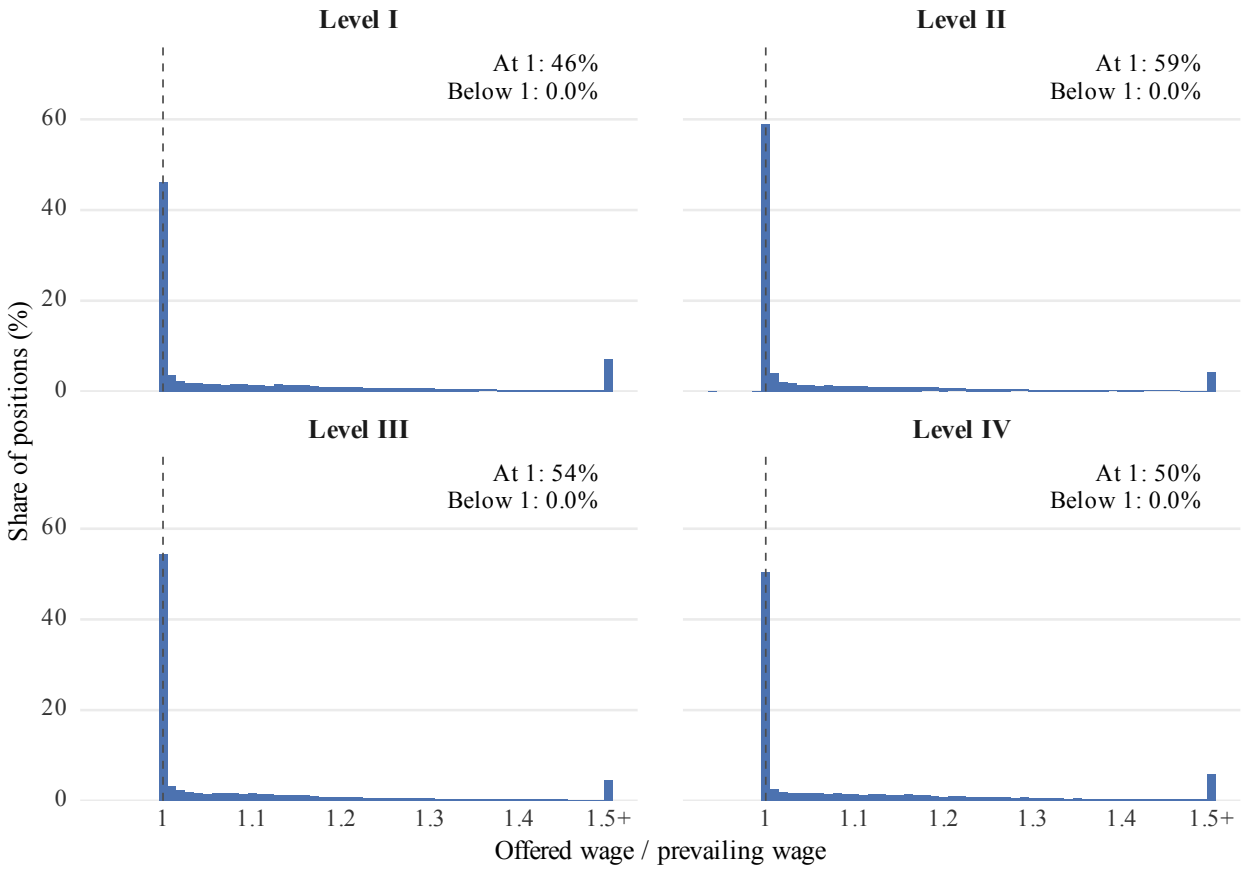
Notes: Certified full-time H-1B Labor Condition Application filings, weighted by the number of positions. Panel (a): fiscal years 2011–2025. Share of positions whose offered wage falls in each band of the occupation's national OEWS wage distribution, weighting occupations by how often CtS master's holders enter them. The dashed line is the share expected if offers matched the occupation's own workforce: 10, 15, 25, 25, 15, and 10 percent. Panel (b): fiscal years 2015–2025, because the wage level is not recorded earlier. Share of positions at each prevailing-wage level in the occupations that CtS and FUS master's holders enter, weighting each group's occupations by where its entrants go. Entry occupations are those of the first US job within five years of the degree.

An employer that wanted to pay less could also certify the position at a low wage level. Each LCA position is certified at one of four statutory levels, and Level I, near the seventeenth OEWS percentile, is the lowest. Panel (b) shows that the occupations CtS master's holders enter are, if anything, less often filed at Level I than those FUS master's holders enter: weighting each group's occupations by where its entrants go, 33 percent of H-1B positions in the occupations CtS master's holders enter are at Level I, against 38 percent in the occupations FUS master's holders enter, and about two-thirds of the former are at Level II or higher. The benchmark is the occupation's whole workforce, not comparable native new hires, so the two panels are consistent with equal pay within

occupations but do not establish it.

Whether the attestation itself sets pay is a separate question. Figure 8 compares each certified offer with the prevailing wage the employer attested to. If the floor binds, offers bunch at it. Across all certified H-1B positions, 46 to 59 percent at each wage level are offered within half a percent of the prevailing wage; counting each filing once instead of each position, the shares are 31 to 46 percent. This suggests that for about half of sponsored positions, the prevailing wage, not the employer, sets the offer, as Assumption 2 requires. Because the prevailing wage rises with the level, from between the tenth and twenty-fifth OEWS percentiles for about half of Level I positions to above the seventy-fifth for most Level IV positions, bunching at a level's floor is not bunching at the bottom of the occupation. The figure covers every certified position, because LCA records cannot be linked to workers in the panel.

Figure 8: Certified H-1B Offers Relative to the Prevailing Wage, by Wage Level



Notes: Distribution of the ratio of the offered wage to the prevailing wage, both annualized, for certified full-time H-1B positions in fiscal years 2015–2025, by prevailing-wage level. Each filing counts by the number of positions it requests; occupations are not reweighted. Bins have width 0.01; the bar at 1 covers ratios from 0.995 to 1.005, and ratios above 1.5 are pooled. Ratios below 0.5 or above 5, 0.1 percent of positions, are dropped. Fiscal years 2011–2014 are excluded because the wage level is not recorded.

6 The Dynamics of the Hiring Friction

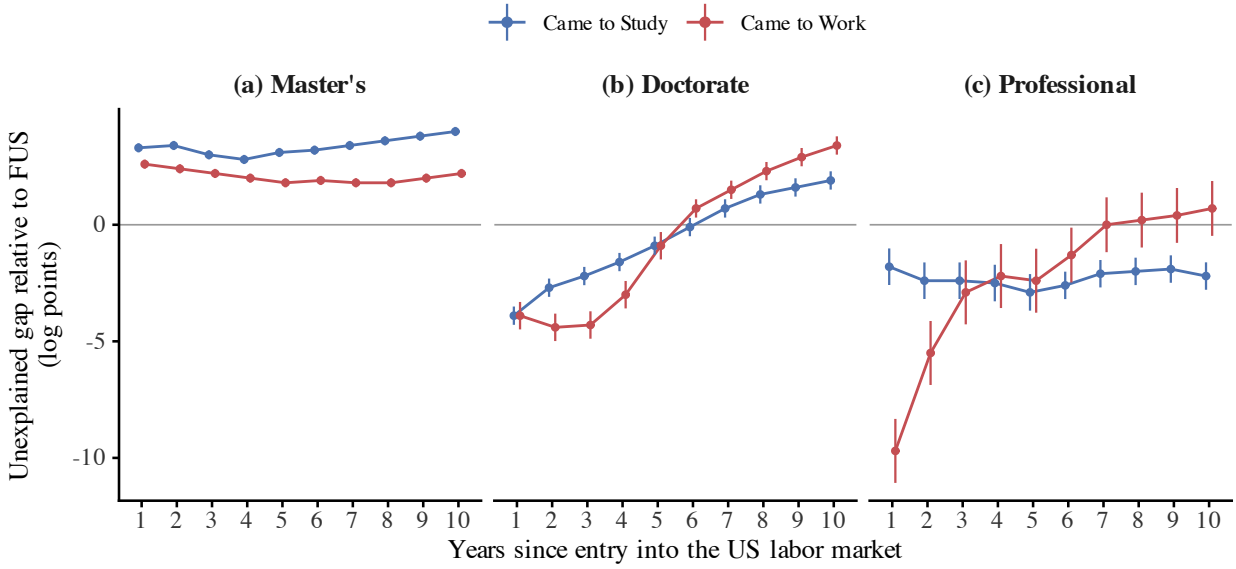
This section follows each group's occupational median wage gap relative to FUS workers over the first decade. As stated in Section 4.2, the object of interest here is the unexplained component in the Oaxaca-Blinder decomposition. If the unexplained component at entry reflects transient components of d, it should move toward zero as employers learn or as workers assimilate. The path need not be symmetric. Because the outcome is the occupation's median wage, the gap moves only when workers change occupation. A worker whose occupation pays less than the FUS benchmark predicts has an incentive to move to a better-paying occupation, so a negative unexplained component, the sign the under-placement contract generates, should shrink. A worker whose occupation pays more has no such incentive, so a positive unexplained component can persist or grow. Both immigrant pathways need sponsorship, but only CtW workers hold credentials that US employers cannot readily evaluate, so employer learning predicts a larger fall in CtW deficits than in CtS deficits.

From Sorting to Placement: The Gap at Entry The Gelbach decomposition in Section 5.3 holds job characteristics, the occupation's rank, and the employer fixed, so it shows how much of the occupational wage gap comes from where immigrants sort. The friction in the model acts on that sorting. Measuring it therefore requires the gap before these characteristics are held fixed. I take that step at the first US job with the Oaxaca–Blinder decomposition of Equation (6), which holds fixed only characteristics determined before labor-market entry, and then follow the same object over time.

Tables 15 and 16 report the result. At the first US job, the adjusted placement gap, i.e. the unexplained component, U_{pe} is 3.2, -2.3, and -3.1 log points for CtS master's, doctorate, and professional-degree holders, and -1.3, -9.3, and -17.3 for their CtW counterparts. Comparing the two decompositions locates the gap. For CtW doctorate holders, -9.3 log points remain unexplained by pre-market characteristics, against -4.8 once the occupation's characteristics and the employer are also held fixed in Section 5.3. This suggests that about half of their gap is tied to the measured characteristics of the occupations and employers they reach, the margin on which the friction acts in the model. For CtW master's holders, the two decompositions agree, at -1.3 and -1.4 log points, so there is little gap to locate. Five of the six gaps have the negative sign the under-placement contract predicts, and each CtW gap is below the CtS gap at the same degree, by 4.5, 7.0, and 14.2 log points. The exception is the CtS master's gap, a premium that rises from 2.3 log points at the tenth percentile of the occupation's wage distribution to 3.5 at the ninetieth (Table 17). These entry gaps are the starting point for the paths below, which re-estimate the same decomposition in each year among workers who stay.

Figure 9 shows that, among ten-year stayers, negative unexplained components shrink and positive unexplained component persist. The gap of CtW doctorate holders is −3.9 log points in year one and −4.4 in year two; it crosses zero in year six, and is a 3.4-point premium by year ten. The gap of CtW professional-degree holders moves from −9.7 to 0.7, reaching zero in year seven. The CtS doctorate gap moves from the same −3.9 to 1.9, about four-fifths as far as the CtW doctorate gap. Employer learning predicts a larger fall for CtW workers, but this difference is small, which provides little evidence for that account. For CtS professional-degree holders, their small deficit does not shrink: it is 1.8 log points in year one and 2.2 in year ten. Master's holders have no deficit to close. The CtS premium rises from 3.3 to 4.0, and the CtW premium stays between 1.8 and 2.6. From year one on, the CtS deficits never exceed 3.9 log points, which suggests that the part of the friction that persists for workers with US degrees is small. Each path is priced against its own FUS sample, so I compare them only in direction and approximate size.

Figure 9: Unexplained Occupational-Wage Gap of Ten-Year Stayers, CtS and CtW

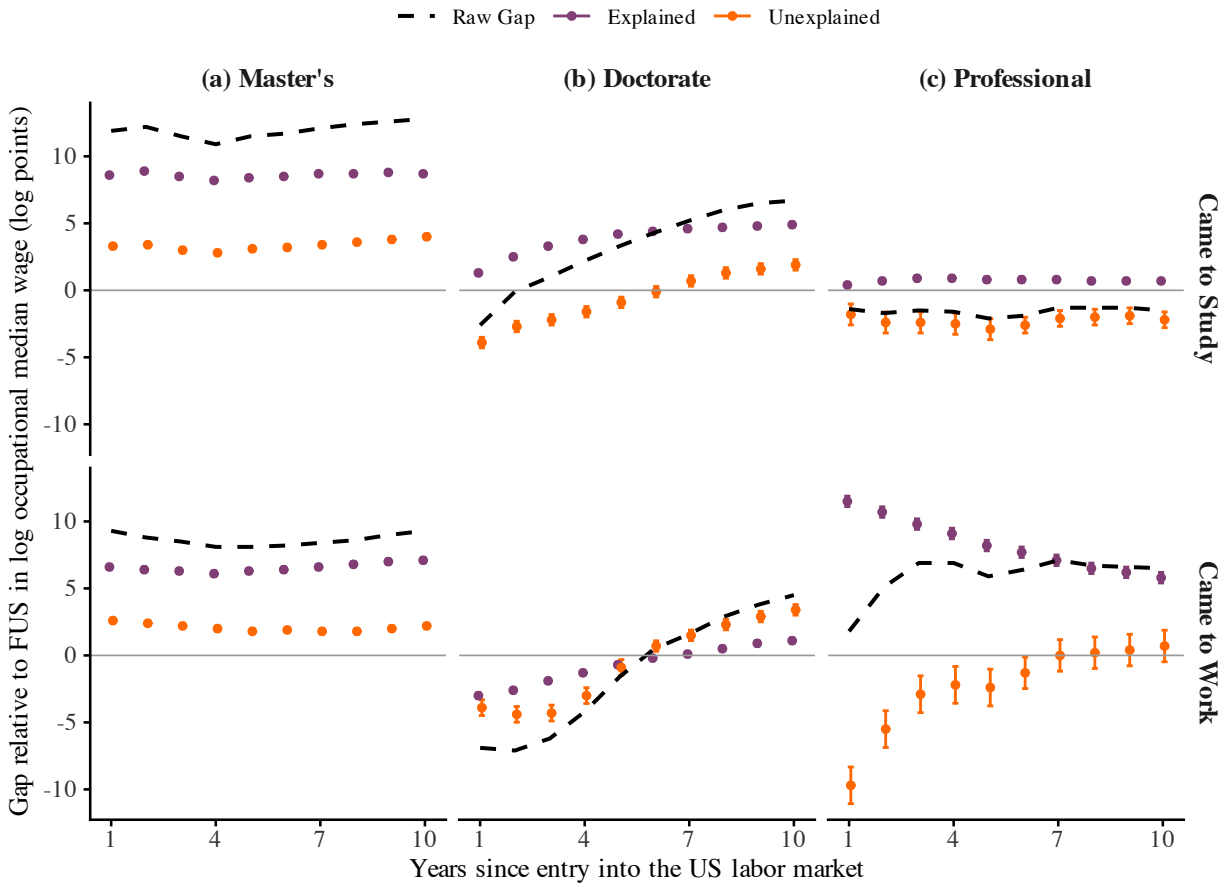


Notes: Unexplained components of the decompositions in Figure 10, with 95 percent confidence intervals. Each pathway's component is priced at the coefficients of its own FUS sample, and the CtW decompositions include additional covariates, so the vertical distance between the two paths is not itself an estimate. Years count from the terminal degree for CtS workers and from the first US labor-market year for CtW workers.

Figure 10 shows the full decomposition for each group. Among CtS stayers (top row), the gaps do not shrink in a common way. For master's holders, the raw advantage in occupational median wages is 11.9 log points a year after graduation and 12.8 at year ten; the explained component is about 9 points throughout, and the unexplained premium rises slightly, from 3.3 to 4.0. For doctorate holders, the unexplained gap moves from a 3.9-point deficit at year one to a 1.9-point premium at year ten, crossing zero between years six and seven. Professional-degree holders keep a small deficit, 2.2 log points at year ten.

Among CtW stayers (bottom row), master's holders have a small unexplained premium: 2.6 log points at year one, 1.8 to 2.4 from year two through year nine, and 2.2 by year ten, with an explained component of 6.1 to 7.1 points throughout. The doctorate deficit, 3.9 log points at year one and 4.4 at year two, closes steadily and becomes a 3.4-point premium by year ten, while the explained component rises from about -3 to about $+1$. For professional-degree holders, the unexplained gap moves from a 9.7-point deficit to a 0.7-point premium, but the explained component falls from $+11.5$ to $+5.8$, so the raw gap rises only from 1.8 to 6.5 log points. Their convergence in Figure 9 therefore overstates how much their occupational position improves relative to FUS workers.

Figure 10: Occupational-Wage Gap of Ten-Year Stayers, by Pathway and Degree



Notes: Oaxaca–Blinder decomposition, FUS coefficients as reference, of the gap relative to FUS workers in the log OEWS median wage of the occupation held in each year, among workers who remain in the US through year ten; graduation cohorts from 2003 on that reach year ten. Each pathway, degree, and year is estimated separately (Tables A7 to A12). Years count from the terminal degree for CtS workers and from the first US labor-market year for CtW workers; year zero is reported in the tables but not shown (Section 4.2). The CtW decompositions also hold fixed non-US labor-market experience, its square, and a same-employer transfer indicator. Bars are 95 percent confidence intervals. Doctorate pools the PhD and non-licensure doctorates; Professional denotes licensure doctorates.

Selection into Staying The starting points of the paths show how selected the stayers are. Among CtS workers, the unexplained gap of ten-year stayers in year one is close to that of all entrants at

their first US job: 3.3, −3.9, and −1.8 log points for master's, doctorate, and professional-degree holders, against the entry gaps of 3.2, −2.3, and −3.1 (Tables A7 to A9). Among CtW workers it is higher: 2.6, −3.9, and −9.7 log points, against −1.3, −9.3, and −17.3 (Tables A10 to A12). CtW workers who remain for ten years thus began in better-paying occupations, relative to comparable FUS workers, than CtW entrants as a whole, by 3.9, 5.4, and 7.6 log points, as expected if lower-earning immigrants are the ones who leave (Lubotsky, 2007). Their paths understate the gaps of all CtW entrants, most of all for professional-degree holders. The comparison is approximate, because the two decompositions use different FUS benchmarks and different job samples.

Interpreting the Estimates through the Framework For CtW doctorate holders convergence is what a component of d that fades with employer learning would produce: employers resolve their uncertainty about a foreign credential as they observe the worker. The CtS doctorate gap, however, closes about four-fifths as far, which suggests that much of this convergence is not specific to a foreign credential. For CtW professional-degree holders, convergence could also reflect relicensing or moves between occupations that do not require a US license, and the fall in their explained component, from +11.5 to +5.8 log points, indicates that their occupations change along the way.[10] I find no occupational-pay deficit for master's stayers to close: their gaps are premiums in both pathways, which is consistent with the asymmetry above. The evidence on the master's friction therefore rests on the over-education gap at entry and on the policy response in Section 7.

Convergence need not reflect employer learning. Sponsorship ends at permanent residence, which the two pathways reach at different points in their careers, and US-specific skills accumulate with tenure. The outcome also moves only when a worker changes occupation, and a worker whose occupation pays less than comparable FUS workers' has a reason to change it whatever the cause of the gap. These estimates do not separate employer learning from job search or assimilation.

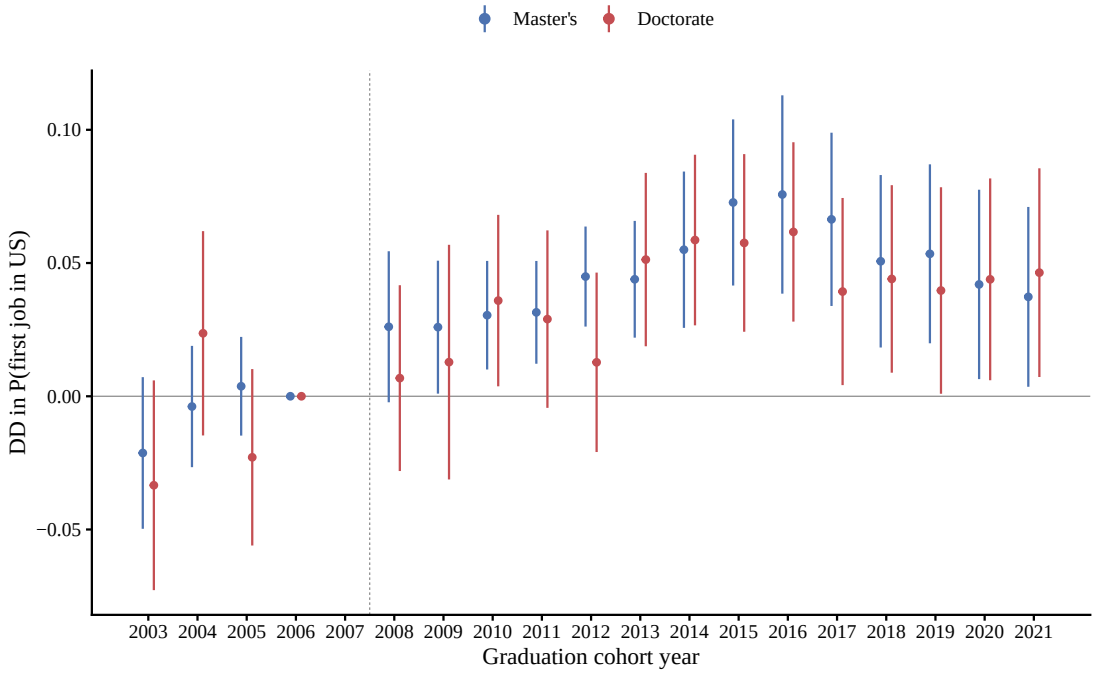
[10] Relicensing is unlikely to be the main channel. I define licensed occupations as those that professional degrees train for and that require a US license to practice: physicians and surgeons, dentists, pharmacists, optometrists, podiatrists, chiropractors, physical therapists, audiologists, veterinarians, lawyers, and judges, identified by their SOC codes. Occupations that require a less costly license, such as nursing, are not included. Among ten-year stayers, the share of CtW professional-degree holders working in a licensed occupation is 52 percent in year one, 56 in year four, and 49 in year ten, against 60, 62, and 58 percent for FUS professional-degree holders.

7 The Hiring Friction at the Margins

This section asks whether lowering the cost of hiring an international graduate changes where graduates start their careers and the jobs they hold. I first examine when the placement effect appears, then the jobs of graduates who stay, and finally whether the effects are concentrated at employers subject to the H-1B lottery.

7.1 Extensive Margin: Initial US Placements

Figure 11: Cohort Event Study: STEM–Non-STEM Difference in the Probability of a First Job in the US



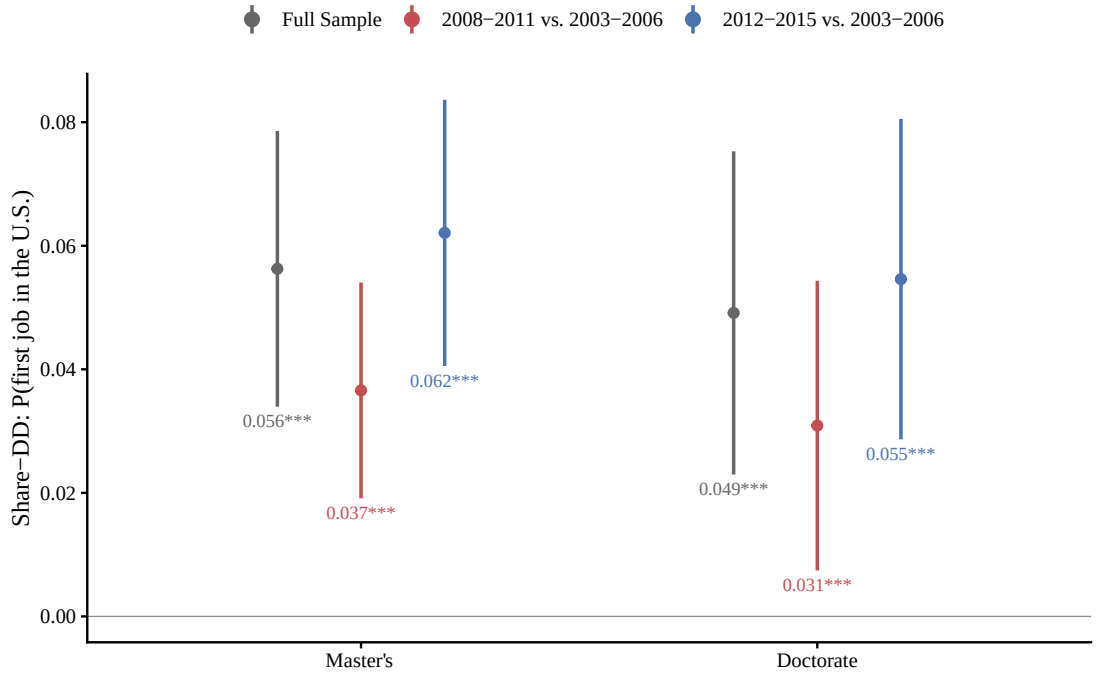
Note: Coefficients ϕ_{ec} from Equation (12), CtS master's and doctorate graduates, estimated separately by degree and offset horizontally. Doctorate pools the PhD and non-licensure doctorates. The outcome is an indicator for a first job located in the US, so a positive coefficient is a relative increase in US placement. STEM indicates that the graduate's own terminal-degree major is on the 2008 STEM Designated Degree Program List. The 2006 cohort is the reference and 2007 is omitted. Bars are 95 percent confidence intervals clustered by major.

Figure 11 plots the event-study coefficients ϕ_{ec} from Equation (12); a positive coefficient means that STEM graduates of that cohort were more likely than non-STEM graduates to start in the US, relative to 2006. For master's graduates, the 2003–2005 coefficients are -2.1, -0.4, and $+0.4$ percentage points. They rise toward the reference year, which would overstate the post-2008 estimates if it reflected a trend, but the coefficients are jointly statistically insignificant. For doctorate graduates, the coefficients are -3.3, $+2.4$, and -2.3 points. They alternate in sign, and

the joint test is noisy and insignificant at the 5% level. The doctorate sample is about a fifth the size of the master's sample, and I treat the doctorate estimates as suggestive.

The master's effect appears with the first treated cohort and grows. For 2008–2011 graduates, the increase in US placement ranges from 2.6 to 3.2 percentage points; the 2008 and 2009 graduates, who had enrolled before the rule was announced, are both at 2.6 points, and every coefficient but the 2008 one is statistically significant. The increase is 4.4 to 5.5 percentage points for 2012–2014 graduates, rises to 6.6 to 7.6 for 2015–2017 graduates, and is 5.1 and 5.4 for 2018 and 2019 and 4.2 and 3.7 for 2020 and 2021, all statistically significant. Two features of the policy environment match this path. The H-1B cap was not reached in the first week of filing for fiscal years 2010 through 2013, so an additional filing season was worth less to graduates filing in those years. The 2016 rule also lengthened the extension from 17 to 24 months for graduates who applied after May 2016, including much of the 2015 cohort.

Figure 12: Increase in CtS US Placement Rate after STEM OPT Extension



Notes:
Difference-in-differences estimates of δ_e from Equation (9): the change in the probability of a first job in the US for STEM relative to non-STEM CtS graduates. Full Sample compares the 2008–2021 cohorts with the 2003–2006 cohorts, and the other two series compare the window named in the legend with the same cohorts; the 2007 cohort is excluded. All estimates control for school tier and major generality. Labels give the point estimates. Bars are 95 percent confidence intervals clustered by major. Doctorate pools the PhD and non-licensure doctorates.

Both rules lower the same friction, the risk that a hire is lost before winning the lottery. The coefficients suggest that the 2008 rule alone raised master's placement by about 3 percentage points, and the cohorts that received the longer extension raise the pooled estimate below. These later cohorts also enrolled after the extension was known, and foreign students shifted into STEM majors after 2008 (Amuedo-Dorantes et al., 2019), so the post-2015 coefficients combine a larger

dose of the treatment with any selection into STEM (Assumption 5). For doctorate graduates, the coefficients are 0.7, 1.3, 2.9, and 1.3 percentage points for 2008, 2009, 2011, and 2012 and not statistically significant, whereas the coefficient of 3.6 for 2010 and those of 3.9 to 6.2 for every cohort from 2013 on are all statistically significant.

Figure 12 shows that relative to non-STEM graduates, STEM graduates became more likely to take their first job in the US after the extension. Pooling the 2008–2021 cohorts, the difference-in-differences in US placement is 5.6 percentage points for master's graduates, about 6.5 percent of the pre-period rate of 86.3 percent at which STEM master's graduates started in the US, and 4.9 percentage points for doctorate graduates. The 2008–2011 estimates are 3.7 and 3.1 percentage points, and for 2012–2015 graduates they are 6.2 and 5.5; all six estimates are statistically significant with standard errors clustered by major. The master's estimate of 3.7 exceeds the cohort coefficients in Figure 11 partly because its comparison group is the 2003–2006 average, not the 2006 cohort. The 2012–2015 window also includes 2015 graduates who received the longer extension. Because the effect persists after the recession cohorts, the downturn alone appears unlikely to explain it. Under the 2008 rule, the relative US placement of master's graduates rose by about 3 to 3.7 percentage points, and it rose further once the 2016 rule applied; the doctorate increase is of similar size and is statistically significant for 2010 and from the 2013 cohort on.

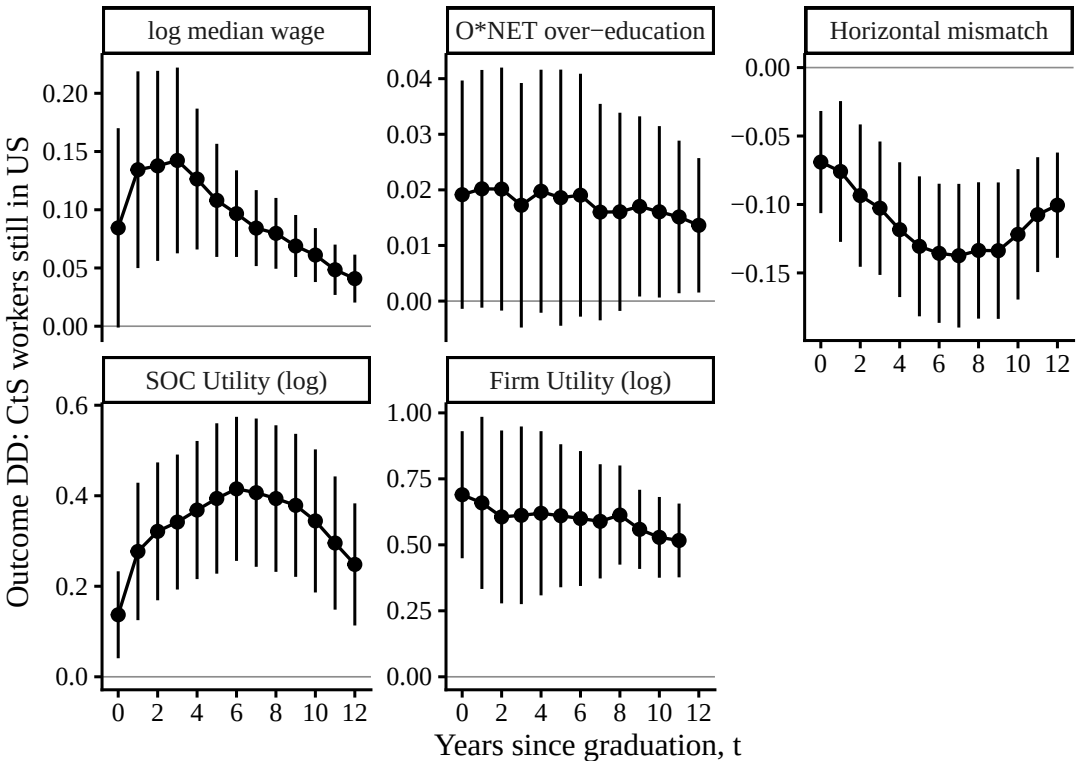
7.2 Job Quality among Stayers

The job-quality estimates combine a change in the jobs available to a given worker with a change in who stays (Equation (14)); I return to the second at the end of this subsection. Table 19 reports the difference-in-differences of Equation (10) five years after graduation, and Figures 13 and 14 trace them year by year. I begin with the 2008–2011 window, which excludes the cohorts that received the longer extension or enrolled after it was known. Among master's graduates of those cohorts still working in the US five years after graduation, the share of STEM graduates working outside their field fell by 7.6 percentage points relative to non-STEM graduates. Relative to non-STEM graduates, their occupations rose by 19 log points in revealed-preference rank and 8 log points in median pay, and their employers by 21 log points. Over-education does not change: the estimate is 0.5 percentage points. Over the career, horizontal mismatch is lower from $t = 0$ on, by 4.9 to 8.3 percentage points from $t = 2$, and the occupational median wage is unchanged at $t = 0$ and 7 to 11 log points higher at $t = 1$ to $t = 4$.

Pooling the 2008–2021 cohorts raises the estimates. At $t = 5$, horizontal mismatch is 13.1 percentage points lower, occupational PageRank score 39 and firm PageRank score 61 log points higher, and the occupational median wage 11 log points higher; at $t = 11$ the mismatch and PageRank estimates are still -10.7 percentage points and 30 log points. Over-education again does not

fall: it is 1.7 to 2.0 points higher in the first five years, not statistically significant, and 1.4 to 1.7 points higher, and statistically significant, at $t = 9$ to $t = 12$. The post-recession 2012–2015 window shows the same pattern: horizontal mismatch falls by 6.9 to 13.9 percentage points, PageRank score rises by 26 to 37 log points from $t = 1$ on, firm PageRank score is 64 log points higher at $t = 5$, and the occupational median wage 10 log points higher, while over-education is 1.1 to 1.9 points higher and statistically significant only at $t = 11$ and $t = 12$. Lower horizontal mismatch with no fall in over-education indicates that master's graduates moved within an education level, into their own field and into higher-ranked, better-paid occupations, a change that over-education, measured in degree levels, does not register (Section 2.3). The positive estimates at $t = 9$ to $t = 12$ are not what the model predicts, and may reflect the changing composition of stayers.

Figure 13: Trends of Occupational Outcome of Stayers, Master's Obtained 2008–2021

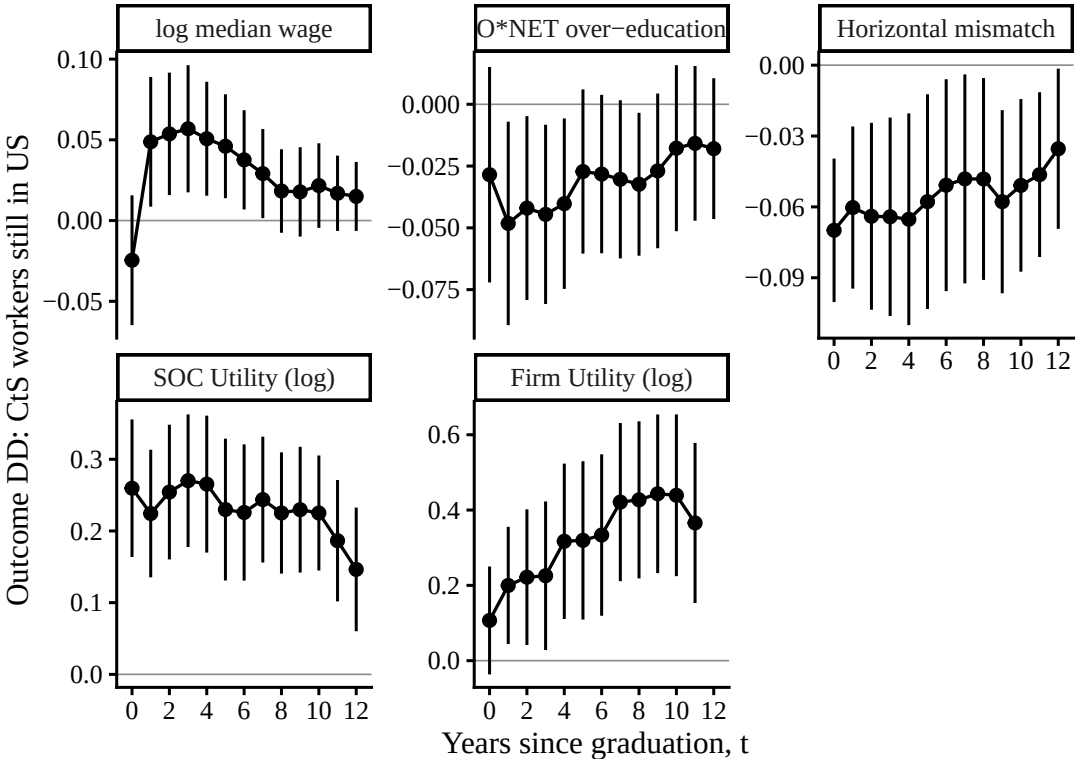


Notes: CtS master's graduates. Estimates of δ^k_{et} from Equation (10) by years since graduation, for graduates whose first job is in the United States and who are still US-resident at t. Post-period cohorts graduated in 2008–2021 and comparison cohorts in 2003–2006. Log median wage is the residualized log OEWS median wage of the occupation held; SOC utility and firm utility are log occupational and firm PageRank; over-education and horizontal mismatch are in probability units. Horizons are shown when every STEM×Post group retains at least 100 workers and at least 20 percent of its $t = 0$ size. Bars are 95 percent confidence intervals clustered by major. Because the sample conditions on staying, the estimates combine changes in jobs with changes in who stays (Equation (14)).

For doctorate graduates the gains are similar, and over-education also falls. Pooling the 2008–2021 cohorts, horizontal mismatch is 3.5 to 7.0 percentage points lower at every horizon, occupational PageRank score is 15 to 27 log points higher at every horizon, and the occupational median wage is 4 to 6 log points higher from $t = 1$ to $t = 6$, though not statistically significant from

$t = 8$. Firm PageRank rises later, to 43 and 44 log points at $t = 8$ and $t = 10$. Over-education is 4.0 to 4.8 percentage points lower from $t = 1$ to $t = 4$, not statistically significant at $t = 5$, and 3.2 points lower at $t = 8$. In the 2008–2011 window the doctorate estimates are smaller, and the mismatch estimates are not statistically significant after $t = 5$; firm PageRank score, for example, is 31 and 32 log points higher at $t = 9$ and $t = 10$, against 54 and 51 in the 2012–2015 window. The evidence for improved placement is therefore similar for the two degrees in the full post-period and less precise for doctorate graduates within the local windows.

Figure 14: Trends of Occupational Outcome of Stayers, Doctorate Obtained 2008–2021



Notes: CtS doctorate graduates. Estimates of δ^k_{et} from Equation (10) by years since graduation, for graduates whose first job is in the United States and who are still US-resident at t. Post-period cohorts graduated in 2008–2021 and comparison cohorts in 2003–2006. Log median wage is the residualized log OEWS median wage of the occupation held; SOC utility and firm utility are log occupational and firm PageRank; over-education and horizontal mismatch are in probability units. Horizons are shown when every STEM×Post group retains at least 100 workers and at least 20 percent of its $t = 0$ size. Bars are 95 percent confidence intervals clustered by major. Because the sample conditions on staying, the estimates combine changes in jobs with changes in who stays (Equation (14)). Doctorate pools the PhD and non-licensure doctorates.

The extension could raise these estimates by changing who stays: more STEM graduates started in the US, and more may have stayed. Both changes are small relative to the population of stayers. Relative to non-STEM graduates, the share of STEM master's graduates still in the US rose by 1.0 percentage point by $t = 5$ and 1.9 by $t = 10$; for 2008–2011 graduates the estimates are -0.1 and 0.6. For doctorate graduates they are 0.3 and 0.9. Table G2 reports the estimates at other horizons and for the 2012–2015 window, together with a linear probability version of Equation (11) with standard errors clustered by major, which gives similar magnitudes. Counting both the additional

entrants and the additional retention, marginal stayers are about one in thirteen of STEM master's stayers at year five, and about one in nine at the upper limits of the 95 percent confidence intervals. Even at those limits, for composition alone to produce the 13.1-point fall in horizontal mismatch, they would need a mismatch rate more than 100 percentage points below that of always-stayers. The argument does not bound composition for over-education or for the continuous outcomes, so the stayer estimates do not identify a job-quality gain for a fixed group of workers.

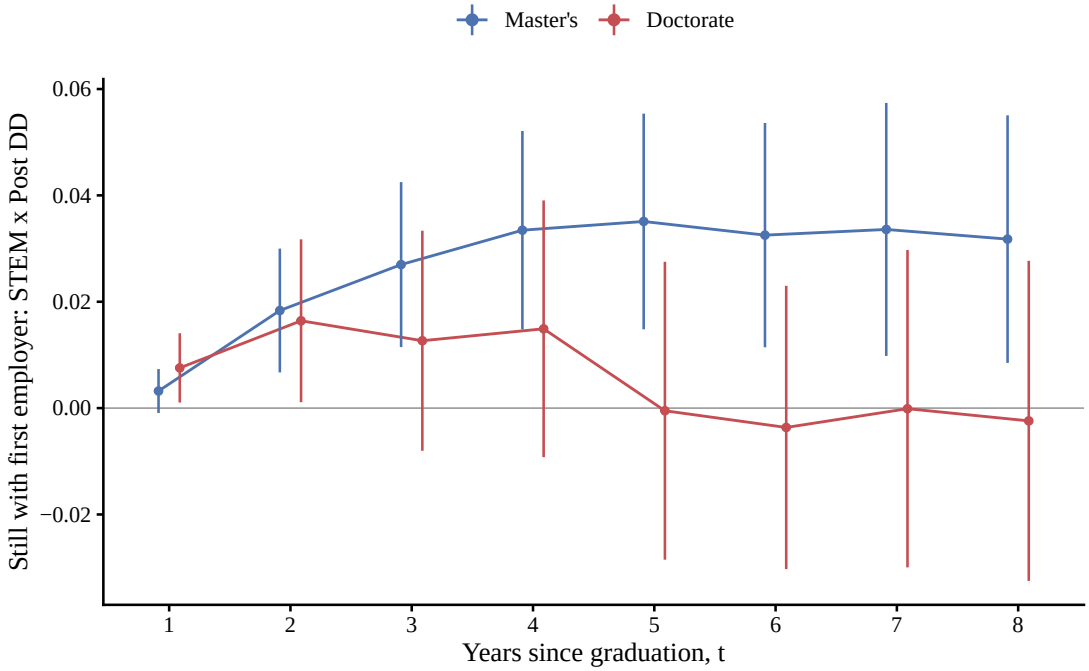
Retention measures whether graduates leave the US; I next ask whether they leave their first employer. Figure 15 plots $\delta_{et}^{\text{NoSwitch}}$ from Equation (11). Graduates of later cohorts change employers sooner in every group: five years after graduation, 58 percent of post-2008 non-STEM master's graduates are still with their first employer, against 71 percent of pre-2008 graduates. The decline is smaller for STEM master's graduates, 9.4 percentage points against 12.7. As a result, STEM master's graduates of post-2008 cohorts are relatively more likely to still be with their first employer: by 0.3 percentage points at $t = 1$, 1.8 at $t = 2$, 2.7 to 3.5 from $t = 3$ to $t = 5$, and 3.2 to 3.4 through $t = 8$, with standard errors of 0.8 to 1.2 from $t = 3$ on. For doctorate graduates the estimates are 0.8 and 1.6 percentage points at $t = 1$ and $t = 2$, and within 1.5 points of zero from $t = 3$ on, with standard errors of 1.1 to 1.5.

Three mechanisms could reduce switching, and they differ in timing. First, before 2008 a STEM graduate whose first H-1B petition failed lost work authorization twelve months after graduation and had to leave the employer; removing these involuntary separations predicts an effect at $t = 1$ and $t = 2$. Second, a graduate waiting on the lottery or on a filed petition has reason to stay with the employer until status is secured, and the extension keeps graduates in this position for up to three filing seasons; this predicts an effect that builds over the first three years. Third, a longer expected match raises the employer's return to training the worker, and when the two share the cost of firm-specific training, neither wants to end the match (Becker, 1962; Hashimoto, 1981); this predicts an effect that builds with tenure and persists after status is secured. The master's path, starting small at $t = 1$, rising through $t = 5$, and holding at about 3.2 points at $t = 8$, fits the second and third mechanisms, and its persistence fits the third best, although sponsorship for permanent residence can also keep a worker with the employer. The doctorate path, small, early, and gone by $t = 5$, is closer to the first, though its estimates are imprecise. Timing alone does not separate the three, but under each, the extension acted on the first employment relationship more than on the decision to remain in the US.

7.3 Cap-Exempt First Employers

A graduate whose first employer is exempt from the H-1B cap faces no lottery at that employer, so if the extension works by adding lottery draws, the direct effect should be smaller for these

Figure 15: Employer Switching after a US First Job: STEM–Non-STEM Difference-in-Differences



Note: Estimates of $\delta_{et}^{\text{NoSwitch}}$ from Equation (11), CtS master's and doctorate graduates whose first job is in the US, estimated separately by degree and offset horizontally. The outcome indicates that the graduate has not yet moved from the first employer to another employer by year t, so a positive estimate means less switching. Bars are 95 percent confidence intervals clustered by major. Doctorate pools the PhD and non-licensure doctorates.

graduates. It need not be zero, because cap-subject employers remain their outside option, both when they take the job and afterward. I use the indicator CE_j in two ways: as an outcome, and to split the stayer estimates.

Cap-exempt employers account for 3.6 percent of the US first jobs of pre-period STEM master's graduates and 34.5 percent of those of doctorate graduates. Panel A of Table 18 shows that the extension did not change these shares: the difference-in-differences in the cap-exempt share is -0.2 percentage points for master's graduates and 0.6 for doctorate graduates. Because the shares are computed among graduates whose first job is in the US, and the extension changed who takes one, these are not effects on the probability of starting at a cap-exempt employer among all graduates.

Panel B splits the stayer estimates by the type of first employer. Most of the gains are larger for graduates who started at cap-subject employers, as the lottery channel predicts. Five years after graduation, horizontal mismatch is 13.2 percentage points lower for master's graduates whose first employer was cap-subject and 6.9 points lower for those whose first employer was cap-exempt; for doctorate graduates the estimates are -8.7 and -1.7. Firm PageRank score is 63 and 47 log points higher for master's and doctorate graduates who started at cap-subject employers and unchanged for those who started at cap-exempt ones, and master's occupational PageRank score rises by 41 log points for those who started at cap-subject employers against 8 for those who started at cap-exempt ones. The gains at cap-exempt employers are not zero. The occupational median wage rises for both types of first employer, by 11 and 7 log points for master's graduates and by 3.1 and 3.5 for doctorate graduates, and the doctorate PageRank gain is similar for both, 22 and 17. The comparison has four limits: the master's difference in horizontal mismatch, 6.3 points, is about one standard error when the two samples are treated as independent; the master's cap-exempt sample is small, about 400 workers in its smallest group; firm PageRank score compares poorly across the split when most cap-exempt employers are universities; and the split conditions on where graduates start: Panel A shows no shift in the shares, but it cannot show whether different graduates start at each type of employer.

I find that the extension changed where graduates start and the jobs they hold more than how long they stay. It raised US placement of STEM master's graduates, and those who stayed work more often in their field and hold higher-ranked occupations at higher-ranked employers, while retention rose little and over-education did not fall. The match gains are mostly larger where the lottery binds, which is consistent with the friction the extension relaxed, though the master's cap-exempt sample is small.

8 Conclusion

In the model I develop, a cost of hiring immigrants downshifts the menu of offers when firms cannot pay an immigrant less than a native in the same job: immigrant workers still reveal their type via education, but they take jobs that require less education, and the gap shrinks with the degree. I find evidence consistent with this prediction for most groups.

At their first US job, CtS and CtW master's holders are 3.1 and 9.1 percentage points more likely than comparable FUS workers to be over-educated. CtS and CtW doctorate holders show a smaller gap, 1.4 and 1.7 points at cap-subject employers, against 2.9 and 8.7 points for master's holders at those employers, and CtW professional-degree holders, who must relicense, have the largest gap at 20.9 points. For master's holders, over-education does not appear to mean a lower-paying occupation: the occupations of CtS and CtW master's holders pay 11.9 and 5.8 log points more than those of FUS master's holders. The occupations of CtW doctorate and professional-degree holders instead pay 9.2 and 8.2 log points less than those of FUS workers with the same degree.

Over the next decade following entry to the US labor market, I find that the occupational-pay deficits of foreign-trained workers (CtW) appear to fade for doctorate and professional-degree holders. Among ten-year stayers, their gaps relative to FUS workers move from -3.9 and -9.7 log points in year one to 3.4 and 0.7 by year ten, although the professional-degree path overstates the improvement in their occupations. The CtS doctorate deficit closes about four-fifths as far, and master's holders in both pathways keep a premium of under 5 log points that changes little. These paths describe stayers, and they do not separate employer learning from other sources of occupational mobility.

The 2008 STEM-OPT extension results suggest that work authorization affects both whether graduates start in the US and the jobs they then hold. I find that the extension raised the share of CtS STEM master's graduates of the 2008–2011 cohorts who started in the US by 3.7 percentage points relative to non-STEM graduates. Five years later, those who stayed were 7.6 points less likely to work outside their field and held occupations that paid 8 log points more, while over-education did not change. Retention for these cohorts changed by only -0.1 percentage points at year five, and most gains are larger for graduates whose first employer was subject to the H-1B lottery.

What d contains matters for welfare. If immigrants are placed below their degrees because they are less productive, the placement is efficient; if a cost created by visa rules places them there, it is misallocation. The table below sets the results against two channels. An information friction, in which employers evaluate foreign credentials less accurately, predicts a gap that is larger for foreign degrees and fades with US experience. I find that over-education gaps at entry are larger for foreign master's and professional degrees, and that occupational-pay deficits appear to fade for

foreign-trained doctorate and professional-degree holders. The CtS doctorate deficit also closes, so the second result does not isolate this channel. A hiring cost such as sponsorship predicts a gap that persists until sponsorship ends and that a longer work permit reduces; the STEM-OPT extension appears to have reduced it. The evidence is consistent with both channels and suggests that at least part of d is a cost that policy can remove, though this rests on the identifying assumptions of the STEM-OPT design.

What the Evidence Says about the Friction d

Channel	Predicts	Finding	Evidence
Information (fades with experience)	Gap larger for foreign degrees	Yes for master's and professional degrees: CtW $>$ CtS at entry. Little difference for doctorate degrees at cap-subject employers (1.7 vs. 1.4)	Descriptive
	Occupational-pay deficit shrinks with US experience, more for foreign degrees	Yes for CtW doctorate and professional degrees: deficits close within ten years. The CtS doctorate deficit also closes, about four-fifths as far; the small CtS professional deficit does not. Master's gaps are premiums in both pathways	Descriptive
Hiring cost (persists until sponsorship ends)	A longer work permit raises US placement	Yes: $+3.7$ pp for 2008–2011 master's graduates	Difference-in-differences
	…and improves stayers' jobs	Yes, except over-education: horizontal mismatch falls, occupational PageRank score, median wage, and firm PageRank rise, by more at cap-subject employers for most outcomes; the fall in mismatch is too large to come from extra staying	Difference-in-differences

Notes: Summary of Sections 5.1, 6 and 7.

Three limits bound these conclusions. The outcomes describe occupations and employers, not individual pay, so they establish differences in placement, not in wages within a job. The panel observes neither visa type nor licensure, so the selection and relicensing accounts of the CtW doctorate and professional gaps remain untested. And the STEM-OPT design measures the part of d that the extension removed for CtS graduates, not the level of d for all immigrants.

The next step is to measure the welfare cost set out in Appendix J, which requires the share of workers displaced and a credible counterfactual without the friction. CtW workers offer a further test. The extension did not cover them, but they draw from the same H-1B lottery, in which a 29-month OPT gives a STEM CtS graduate up to three draws instead of one. If employers treat the two groups as substitutes, cheaper CtS hiring should lower the placement of STEM CtW arrivals after 2008, which a triple difference can check.

References

Abadie, Alberto et al. (2023). "When Should You Adjust Standard Errors for Clustering?" In: *Quarterly Journal of Economics* 138.1, pp. 1–35.

Alba-Ramirez, Alfonso (1993). "Mismatch in the Spanish labor market: overeducation?" In: *Journal of Human Resources*, pp. 259–278.

Aleksynska, Mariya and Ahmed Tritah (2013). "Occupation–education mismatch of immigrant workers in Europe: Context and policies". In: *Economics of Education Review* 36, pp. 229–244.

Allen, Jim and Rolf van der Velden (2001). "Educational Mismatches versus Skill Mismatches: Effects on Wages, Job Satisfaction, and On-the-Job Search". In: *Oxford Economic Papers* 53.3, pp. 434–452. DOI: 10.1093/oep/53.3.434.

Altonji, Joseph G. and Charles R. Pierret (2001). "Employer Learning and Statistical Discrimination". In: *The Quarterly Journal of Economics* 116.1, pp. 313–350. DOI: 10.1162/003355301556329.

Amuedo-Dorantes, Catalina, Delia Furtado, and Huanan Xu (2019). "OPT Policy Changes and Foreign Born STEM Talent in the U.S." In: *Labour Economics* 61, p. 101752.

Arcidiacono, Peter, Patrick Bayer, and Aurel Hizmo (2010). "Beyond Signaling and Human Capital: Education and the Revelation of Ability". In: *American Economic Journal: Applied Economics* 2.4, pp. 76–104.

Autor, David H, Frank Levy, and Richard J Murnane (2003). "Skill demand, inequality, and computerization: Connecting the dots". In: *Technology, growth, and the labor market*. Springer, pp. 107–129.

Baert, Stijn, Bart Cockx, and Dieter Verhaest (2013). "Overeducation at the Start of the Career: Stepping Stone or Trap?" In: *Labour Economics* 25, pp. 123–140. DOI: 10.1016/j.labeco.2013.04.013.

Baley, Isaac, Ana Figueiredo, and Robert Ulbricht (2022). "Mismatch Cycles". In: *Journal of Political Economy* 130.11, pp. 2943–2984.

Batalova, Jeanne and Michael Fix (2021). "Leaving Money on the Table: The Persistence of Brain Waste among College-Educated Immigrants." In: *Migration Policy Institute*.

Becker, Gary S. (1957). *The Economics of Discrimination*. Chicago: University of Chicago Press.

— (1962). "Investment in Human Capital: A Theoretical Analysis". In: *Journal of Political Economy* 70.5, Part 2, pp. 9–49.

Beenstock, Michael, Barry R Chiswick, and Ari Paltiel (2010). "Testing the immigrant assimilation hypothesis with longitudinal data". In: *Review of Economics of the Household* 8.1, pp. 7–27.

Beine, Michel, Giovanni Peri, and Morgan Raux (2023). "International College Students' Impact on the US Skilled Labor Supply". In: *Journal of Public Economics* 223, p. 104917. DOI: 10.1016/j.jpubeco.2023.104917.

Berry, Alexander, Elizabeth M. Maloney, and David Neumark (Mar. 2024). "The Missing Link? Using LinkedIn Data to Measure Race, Ethnic, and Gender Differences in Employment Outcomes at Individual Companies". In: *NBER Working Paper* 32294. DOI: 10.3386/w32294. URL: https://www.nber.org/papers/w32294.

Bohn, Sarah E (2010). "The quantity and quality of new immigrants to the US". In: *Review of Economics of the Household* 8.1, pp. 29–51.

Borjas, George J (1985). "Assimilation, changes in cohort quality, and the earnings of immigrants". In: *Journal of labor Economics* 3.4, pp. 463–489.

Bratsberg, Bernt and James F. Ragan Jr. (2002). "The Impact of Host-Country Schooling on Earnings: A Study of Male Immigrants in the United States". In: *Journal of Human Resources* 37.1, pp. 63–105.

Brin, Sergey and Lawrence Page (1998). "The anatomy of a large-scale hypertextual web search engine". In: *Computer networks and ISDN systems* 30.1-7, pp. 107–117.

Bureau of Labor Statistics (2026). *Occupational Employment and Wage Statistics*. Tech. rep. Accessed: May 15, 2026. U.S. Department of Labor. URL: https://www.bls.gov/oes/.

Card, David, Francesco Devicienti, and Agata Maida (2014). "Rent-sharing, Holdup, and Wages: Evidence from Matched Panel Data". In: *Review of Economic Studies* 81.1, pp. 84–111. DOI: 10.1093/restud/rdt030.

Cassidy, Hugh (2019). "Occupational attainment of natives and immigrants: A cross-cohort analysis". In: *Journal of Human Capital* 13.3, pp. 375–409.

Cassidy, Hugh and Amanda Gaulke (2024). "The increasing penalty to occupation-education mismatch". In: *Economic Inquiry* 62.2, pp. 607–632.

Chen, Willy (2024). "Descriptive Units of Heterogeneity: An Axiomatic Approach to Measuring Heterogeneity". In: *Available at SSRN 4952556*.

Chevalier, Arnaud (2003). "Measuring Over-Education". In: *Economica* 70.279, pp. 509–531. DOI: 10.1111/1468-0335.t01-1-00296.

Chevalier, Arnaud and Joanne Lindley (2009). "Overeducation and the Skills of UK Graduates". In: *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 172.2, pp. 307–337.

Chiswick, Barry R and Paul W Miller (2008). "Why is the payoff to schooling smaller for immigrants?" In: *Labour Economics* 15.6, pp. 1317–1340.

— (2009). "The international transferability of immigrants' human capital". In: *Economics of Education Review* 28.2, pp. 162–169.

Chiswick, Barry R and Paul W Miller (2010). "Occupational language requirements and the value of English in the US labor market". In: *Journal of Population Economics* 23.1, pp. 353–372.

— (2013). "The impact of surplus skills on earnings: Extending the over-education model to language proficiency". In: *Economics of Education Review* 36, pp. 263–275.

Cho, In-Koo and David M Kreps (1987). "Signaling games and stable equilibria". In: *The Quarterly Journal of Economics* 102.2, pp. 179–221.

Collins, William J and Ariell Zimran (2023). "Working their way up? US immigrants' changing labor market assimilation in the age of mass migration". In: *American Economic Journal: Applied Economics* 15.3, pp. 238–269.

Coraggio, Luca et al. (2026). "Mismatch in the 21st Century: An Overview". In: *Labour Economics* 102, p. 102955. DOI: 10.1016/j.labeco.2026.102955.

Dahl, Christian Møller, Torben Johansen, and Christian Vedel (2024). *Breaking the HISCO Barrier: Automatic Occupational Standardization with OccCANINE*. arXiv preprint 2402.13604. URL: https://arxiv.org/abs/2402.13604.

Demirci, Murat (2019). "Transition of International Science, Technology, Engineering, and Mathematics Students to the U.S. Labor Market: The Role of Visa Policy". In: *Economic Inquiry* 57.3, pp. 1367–1391. DOI: 10.1111/ecin.12795.

Doran, Kirk, Alexander Gelber, and Adam Isen (2022). "The Effects of High-Skilled Immigration Policy on Firms: Evidence from Visa Lotteries". In: *Journal of Political Economy* 130.10, pp. 2501–2533. DOI: 10.1086/720467.

Duncan, Greg J. and Saul D. Hoffman (1981). "The Incidence and Wage Effects of Overeducation". In: *Economics of Education Review* 1.1, pp. 75–86. DOI: 10.1016/0272-7757(81)90028-5.

Fredriksson, Peter, Lena Hensvik, and Oskar Nordström Skans (2018). "Mismatch of Talent: Evidence on Match Quality, Entry Wages, and Job Mobility". In: *American Economic Review* 108.11, pp. 3303–3338.

Friedberg, Rachel M (2000). "You can't take it with you? Immigrant assimilation and the portability of human capital". In: *Journal of labor economics* 18.2, pp. 221–251.

Garibaldi, Pietro, Pedro Gomes, and Thepthida Sopraseuth (2025a). "Economic Consequences of Vertical Mismatch". In: *Quantitative Economics* 16.2, pp. 535–564.

— (2025b). "Output Costs of Education and Skill Mismatch in OECD Countries". In: *Economics Letters* 250, p. 112278.

Gathmann, Christina and Uta Schönberg (2010). "How general is human capital? A task-based approach". In: *Journal of Labor Economics* 28.1, pp. 1–49.

Gelbach, Jonah B. (2016). "When Do Covariates Matter? And Which Ones, and How Much?" In: *Journal of Labor Economics* 34.2, pp. 509–543. DOI: 10.1086/683668.

Gini, Corrado (1912). "Variabilità e mutabilità: contributo allo studio delle distribuzioni e delle relazioni statistiche". Italian. In: *Studi Economico-Giuridici della R. Università de Cagliari* 3, pp. 3–159.

Guo, Xue, Jing Gong, and Min-Seok Pang (2024). "Creation or Destruction? STEM OPT Extension and Employment of Information Technology Professionals". In: *MIS Quarterly* 48.2, pp. 715–730.

Guvenen, Fatih et al. (2020). "Multidimensional Skill Mismatch". In: *American Economic Journal: Macroeconomics* 12.1, pp. 210–244.

Hall, Robert E. and Andreas I. Mueller (2018). "Wage Dispersion and Search Behavior: The Importance of Nonwage Job Values". In: *Journal of Political Economy* 126.4, pp. 1594–1637. DOI: 10.1086/697739.

Hartog, Joop (2000). "Over-education and earnings: where are we, where should we go?" In: *Economics of education review* 19.2, pp. 131–147.

Hashimoto, Masanori (1981). "Firm-Specific Human Capital as a Shared Investment". In: *American Economic Review* 71.3, pp. 475–482.

Hashimoto, Masanori and John Raisian (1985). "Employment Tenure and Earnings Profiles in Japan and the United States". In: *American Economic Review* 75.4, pp. 721–735.

Hsieh, Chang-Tai et al. (2019). "The Allocation of Talent and U.S. Economic Growth". In: *Econometrica* 87.5, pp. 1439–1474.

Hunt, Jennifer (2011). "Which Immigrants Are Most Innovative and Entrepreneurial? Distinctions by Entry Visa". In: *Journal of Labor Economics* 29.3, pp. 417–457.

Imai, Susumu, Derek Stacey, and Casey Warman (2019). "From engineer to taxi driver? Language proficiency and the occupational skills of immigrants". In: *Canadian Journal of Economics/Revue canadienne d'économique* 52.3, pp. 914–953.

Kahn, Lisa B. (2010). "The Long-Term Labor Market Consequences of Graduating from College in a Bad Economy". In: *Labour Economics* 17.2, pp. 303–316.

Kambourov, Gueorgui and Iourii Manovskii (2008). "Rising Occupational and Industry Mobility in the United States: 1968–97". In: *International Economic Review* 49.1, pp. 41–79.

Kerr, William R. and William F. Lincoln (2010). "The Supply Side of Innovation: H-1B Visa Reforms and U.S. Ethnic Invention". In: *Journal of Labor Economics* 28.3, pp. 473–508. DOI: 10.1086/651934.

Khanna, Gaurav and Nicolas Morales (Dec. 2024). *The IT Boom and Other Unintended Consequences of Chasing the American Dream*. Working Paper 460. Center for Global Development. URL: https://www.cgdev.org/publication/it-boom-and-other-unintended-consequences-chasing-american-dream.

Khoo, Pauline (2025). "If You Extend It, They Will Come: The Effects of the STEM OPT Extension". In: *Applied Economics*. Online first. DOI: 10.1080/00036846.2025.2551288.

Kiker, Billy Frazier, Maria C Santos, and M Mendes De Oliveira (1997). "Overeducation and undereducation: Evidence for Portugal". In: *Economics of Education Review* 16.2, pp. 111–125.

Kugler, Adriana D. and Robert M. Sauer (2005). "Doctors without Borders? Relicensing Requirements and Negative Selection in the Market for Physicians". In: *Journal of Labor Economics* 23.3, pp. 437–465.

Kwon, Seoyoung, Jongkwan Lee, and Joan Monràs (2026). *The Effect of Retaining High-Skilled International Graduates: Evidence from the STEM OPT Extension*. CEPR Discussion Paper 21224. Centre for Economic Policy Research.

Lachowska, Marta et al. (2026). "Work hours mismatch". In: *Econometrica* 94.3, pp. 991–1025.

Lang, Kevin and Michael Manove (2011). "Education and Labor Market Discrimination". In: *American Economic Review* 101.4, pp. 1467–1496.

Lee, David S. (2009). "Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects". In: *Review of Economic Studies* 76.3, pp. 1071–1102.

Leuven, Edwin and Hessel Oosterbeek (2011). "Overeducation and mismatch in the labor market". In: *Handbook of the Economics of Education* 4, pp. 283–326.

Li, Xiaoguang and Yao Lu (2023). "Education–occupation mismatch and nativity inequality among highly educated US workers". In: *Demography* 60.1, pp. 201–226.

LinkUp (2025). *Job Records*. Dewey Data. DOI: 10.82551/V091-HN53. URL: https://doi.org/10.82551/V091-HN53.

Lise, Jeremy and Fabien Postel-Vinay (Aug. 2020). "Multidimensional Skills, Sorting, and Human Capital Accumulation". In: *American Economic Review* 110.8, pp. 2328–76. DOI: 10.1257/aer.20162002. URL: https://www.aeaweb.org/articles?id=10.1257/aer.20162002.

Llama 3.1 (8B) (2024). Model Card. Version Release Date: July 23, 2024. Licensed under the Llama 3.1 Community License Agreement. URL: https://huggingface.co/meta-llama/Llama-3.1-8B.

Lu, Yao and Feng Hou (2020). "Immigration system, labor market structures, and overeducation of high-skilled immigrants in the United States and Canada". In: *International migration review* 54.4, pp. 1072–1103.

Lu, Yao and Xiaoguang Li (2021). "Vertical education-occupation mismatch and wage inequality by race/ethnicity and nativity among highly educated US workers". In: *Social Forces* 100.2, pp. 706–737.

Lubotsky, Darren (2007). "Chutes or ladders? A longitudinal analysis of immigrant earnings". In: *Journal of Political Economy* 115.5, pp. 820–867.

Mailath, George J (1987). "Incentive compatibility in signaling games with a continuum of types". In: *Econometrica: Journal of the Econometric Society*, pp. 1349–1365.

Mas, Alexandre and Amanda Pallais (2017). "Valuing Alternative Work Arrangements". In: *American Economic Review* 107.12, pp. 3722–3759. DOI: 10.1257/aer.20161500.

Mas-Colell, Andreu, Michael D Whinston, and Jerry R Green (1995). *Microeconomic theory*. Vol. 1. Oxford University Press New York.

Massing, Natascha et al. (2019). "How Standardized is Occupational Coding? A Comparison of Results from Different Coding Agencies in Germany". In: *Journal of Official Statistics* 35.1.

Mattoo, Aaditya, Ileana Cristina Neagu, and Çağlar Özden (2008). "Brain waste? Educated immigrants in the US labor market". In: *Journal of development economics* 87.2, pp. 255–269.

McGuinness, Seamus, Konstantinos Pouliakas, and Paul Redmond (2018). "Skills Mismatch: Concepts, Measurement and Policy Approaches". In: *Journal of Economic Surveys* 32.4, pp. 985–1015. DOI: 10.1111/joes.12254.

Mellow, Wesley and Hal Sider (1983). "Accuracy of Response in Labor Market Surveys: Evidence and Implications". In: *Journal of Labor Economics* 1.4, pp. 331–344.

Microsoft (2025). *Microsoft Copilot*. https://copilot.microsoft.com. Used to construct names for the three principal components of skills and abilities from O*NET.

Mortensen, Dale T and Christopher A Pissarides (1994). "Job creation and job destruction in the theory of unemployment". In: *The review of economic studies* 61.3, pp. 397–415.

National Center for O*NET Development (2025). *O*NET Database (Version 29.2)*. O*NET 29.2 released February 2025. Sponsored by the U.S. Department of Labor, Employment & Training Administration. URL: https://www.onetcenter.org/db_releases.html.

Newburger, Jennifer Cheeseman (2022). *Average Teachers' Earnings Declining, Lower Than Similarly Educated Workers*. U.S. Census Bureau. URL: https://www.census.gov/library/stories/2022/07/teachers-among-most-educated-yet-pay-lags.html.

Nieto, Sandra, Alessia Matano, and Raul Ramos (2013). *Skill Mismatches in the EU: Immigrants vs. Natives*. Working Paper 18. Research Institute of Applied Economics (IREA), University of Barcelona.

Ordine, Patrizia and Giuseppe Rose (2009). "Overeducation and Instructional Quality: A Theoretical Model and Some Facts". In: *Journal of Human Capital* 3.1, pp. 73–105.

Oreopoulos, Philip (2011). "Why Do Skilled Immigrants Struggle in the Labor Market? A Field Experiment with Thirteen Thousand Resumes". In: *American Economic Journal: Economic Policy* 3.4, pp. 148–171. DOI: 10.1257/pol.3.4.148.

Oreopoulos, Philip, Till von Wachter, and Andrew Heisz (2012). "The Short- and Long-Term Career Effects of Graduating in a Recession". In: *American Economic Journal: Applied Economics* 4.1, pp. 1–29.

People Data Labs (2022a). *Company Insights*. Dewey Data. DOI: 10.82551/ZSJV-Z052. URL: https://doi.org/10.82551/ZSJV-Z052.

— (2022b). *Person*. Dewey Data. DOI: 10.82551/ZSJV-Z052. URL: https://doi.org/10.82551/27H2-5E56.

Pivovarova, Margarita and Jeanne M. Powers (2022). "Do Immigrants Experience Labor Market Mismatch? New Evidence from the US PIAAC". In: *Large-scale Assessments in Education* 10, p. 9.

Quacquarelli Symonds (2019). *QS World University Rankings (2008–2019)*. Tech. rep. Historical tracking panel covering cycles 2008 through 2019. URL: https://www.topuniversities.com/world-university-rankings.

Rémen, Thomas et al. (2018). "Development of a Coding and Crosswalk Tool for Occupations and Industries". In: *Annals of Work Exposures and Health* 62.7, pp. 796–807.

Rendon, Silvio and Alfredo Cuecuecha (2010). "International Job Search: Mexicans in and out of the US". In: *Review of Economics of the Household* 8.1, pp. 53–82.

Robst, John (2007). "Education and job match: The relatedness of college major and work". In: *Economics of Education review* 26.4, pp. 397–407.

Rosen, Sherwin (1986). "The Theory of Equalizing Differences". In: *Handbook of Labor Economics*. Ed. by Orley C. Ashenfelter and Richard Layard. Vol. 1. Amsterdam: North-Holland. Chap. 12, pp. 641–692.

Ruggles, Steven et al. (2025). *IPUMS USA: Version 16.0*. Minneapolis, MN: IPUMS. DOI: 10.18128/D010.V16.0. URL: https://doi.org/10.18128/D010.V16.0.

Russ, Daniel E., Kwan-Yuet Ho, Joanne S. Colt, et al. (2016). "Computer-Based Coding of Free-Text Job Descriptions to Efficiently Identify Occupations in Epidemiological Studies". In: *Occupational and Environmental Medicine* 73.6, pp. 417–424.

Schmitz, Matthew and Linda Forst (2016). "Industry and Occupation in the Electronic Health Record: An Investigation of the National Institute for Occupational Safety and Health Industry and Occupation Computerized Coding System". In: *JMIR Medical Informatics* 4.1, e5.

Schwandt, Hannes and Till von Wachter (2019). "Unlucky Cohorts: Estimating the Long-Term Effects of Entering the Labor Market in a Recession in Large Cross-Sectional Data Sets". In: *Journal of Labor Economics* 37.S1, S161–S198.

Sicherman, Nachum (1991). "" Overeducation" in the labor market". In: *Journal of labor Economics* 9.2, pp. 101–122.

Simpson, Edward H. (1949). "Measurement of Diversity". In: *Nature* 163, p. 688. DOI: 10.1038/163688a0.

Somers, Melline A. et al. (2019). "Horizontal Mismatch between Employment and Field of Education: Evidence from a Systematic Literature Review". In: *Journal of Economic Surveys* 33.2, pp. 567–603.

Sorkin, Isaac (2018). "Ranking firms using revealed preference". In: *The quarterly journal of economics* 133.3, pp. 1331–1393.

Spence, Michael (1978). "Job market signaling". In: *Uncertainty in economics*. Elsevier, pp. 281–306.

Sullivan, Paul and Ted To (2014). "Search and Nonwage Job Characteristics". In: *Journal of Human Resources* 49.2, pp. 472–507. DOI: 10.3368/jhr.49.2.472.

Summerfield, Fraser and Ioannis Theodossiou (2017). "The Effects of Macroeconomic Conditions at Graduation on Overeducation". In: *Economic Inquiry* 55.3, pp. 1370–1387.

Taber, Christopher and Rune Vejlin (2020). "Estimation of a Roy/Search/Compensating Differential Model of the Labor Market". In: *Econometrica* 88.3, pp. 1031–1069. DOI: 10.3982/ECTA14441.

Team, MLX Explore (2026). *mlx-lm*. Version 0.31.0. Python package for running, fine-tuning, and quantizing large language models on Apple Silicon using the MLX framework. URL: https://pypi.org/project/mlx-lm/.

U.S. Department of Education, National Center for Education Statistics (2019). *Integrated Postsecondary Education Data System (IPEDS): Completions Component*. Tech. rep. URL: https://nces.ed.gov/ipeds/datacenter/DataFiles.aspx.

U.S. Department of Labor (2026). *Labor Condition Application (LCA) Disclosure Data 2011-2025*. Employment and Training Administration (ETA), Office of Foreign Labor Certification. Accessed: 2025-12-15. URL: https://www.dol.gov/agencies/eta/foreign-labor/performance.

Van Dijcke, David, Florian Gunsilius, and Austin Wright (May 2024). *Return to Office and the Tenure Distribution*. Working Paper. University of Michigan.

Verdugo, Richard R and Naomi Turner Verdugo (1989). "The impact of surplus schooling on earnings: Some additional findings". In: *Journal of human resources*, pp. 629–643.

Visintin, Stefano, Kea Tijdens, and Maarten van Klaveren (2015). "Skill Mismatch among Migrant Workers: Evidence from a Large Multi-Country Dataset". In: *IZA Journal of Migration* 4, pp. 1–34.

Waldman, Michael (1984). "Job Assignments, Signalling, and Efficiency". In: *RAND Journal of Economics* 15.2, pp. 255–267.

Warman, Casey and Christopher Worswick (2015). "Technological change, occupational tasks and declining immigrant outcomes: Implications for earnings and income inequality in Canada". In: *Canadian Journal of Economics/Revue canadienne d'économique* 48.2, pp. 736–772.

Weber, Rosa, Mathieu Ferry, and Mathieu Ichou (2024). “Which degree for which occupation? Vertical and horizontal mismatch among immigrants, their children, and grandchildren in France”. In: *Demography* 61.6, pp. 1923–1948.

9 Tables

Table 1: Predicted Equilibrium Contract and Over-Education Gap by Degree and Degree Origin

	Worker Type (Terminal Degree Origin)		
Pathway (Education location)	Native	Immigrant (US)	Immigrant (Foreign)
Master's ($\bar{e}'$) [Rate of Over-education relative to Natives]	$W(\bar{e}')$	$W(\tilde{e}')$ [++]	$W(\tilde{e}')$ [+ + +]
Doctorate ($\bar{e}$) [Rate of Over-education relative to Natives]	$W(\bar{e})$	$W(\tilde{e})$ [+]	$W(\tilde{e})$ [++]
Professional ($\bar{e}$) [Rate of Over-education relative to Natives]	$W(\bar{e})$	$W(\tilde{e})$ [+]	$W(\tilde{e})$, larger d [+ + +]

Notes: Every immigrant holds the under-placement contract $W(\tilde{e}) = W(\bar{e}) - d$ at their own degree. Signs give the predicted over-education gap relative to natives with the same degree; more signs denote a larger gap. The gap is smaller at higher degrees because $W(e)$ is convex, and larger with a foreign degree if the friction includes difficulty evaluating foreign credentials. Doctorate comprises the PhD and other research doctorates that do not lead to a professional license. Professional comprises degrees that require a US or state license to practice, such as the MD and the JD; a license earned abroad does not transfer, which adds to d for foreign-educated holders.

Table 2: Sample Coverage of the First-Job Analysis

Pipeline	Initial first jobs	In scope	Analytic sample	Retained of initial	Retained in scope
CtS–FUS, graduate degrees	9,501,780	5,157,285	4,229,371	44.5%	82.0%
CtW–FUS, bachelor's and above	25,945,203	14,933,262	12,339,407	47.6%	82.6%

Notes: FUS = only US post-secondary education; CtS = foreign post-secondary education, then US terminal degree; CtW = only non-US post-secondary education, then US work.One observation per person. In scope: the first post-degree job starts within five years of the terminal degree and is located in the United States, and the degree was completed in 2003–2021. For CtW workers, the first job is the first US job-year after the degree, and the initial count includes only CtW workers with a US job within fifteen years of the degree. The analytic sample is the common-support sample used in every column of Tables 6 through 9. Common support requires the employer block (firm utility, firm churn, H-1B sponsorship, and cap-exempt status) and both mismatch measures; the CtW pipeline also requires market-entry cohort and non-US experience. Most of the loss from the initial count is out of scope: the first job is abroad or has no recorded location, or the degree falls outside the cohort window. Within scope, about three-quarters of the loss comes from the employer measures, which require an employer in the connected set of worker flows between firms; the retained share ranges from 82.0 to 83.8 percent across CtS, CtW, and the two FUS groups. Estimates apply to the covered population of credentialed LinkedIn users and are not population-representative.

Table 3: Degree and Graduation Year Distribution

Pathway	LinkedIn									ACS
	Fully US-educated			Came-to-Study			Came-to-Work			
Observations	17.0M			1.5M			9.4M			3.8 M
	Row %	N	Col.%	Row %	N	Col.%	Row %	N	Col.%	
Terminal Degree										
Bachelor's	66.2%	10.4M	60.8%	1.5%	0.2M	15.7%	32.3%	5.1M	53.7%	64.2%
Master's	52.5%	5.4M	31.9%	9.8%	1.0M	66.0%	37.7%	3.9M	41.4%	26.0%
Doctorate	46.4%	0.5M	3.2%	19.8%	0.2M	15.1%	33.8%	0.4M	4.2%	4.0%
Professional	85.5%	0.7M	4.2%	6.0%	0.05M	3.3%	8.5%	0.1M	0.8%	5.8%
Graduation/Entering Cohorts										
'03–'04	62.5%	0.8M	4.6%	4.7%	0.1M	3.9%	32.7%	0.4M	4.4%	–
'05–'09	61.9%	2.8M	16.4%	4.9%	0.2M	14.6%	33.2%	1.5M	16.0%	–
'10–'14	61.7%	4.8M	28.0%	5.2%	0.4M	26.2%	33.1%	2.6M	27.2%	–
'15–'19	60.0%	6.0M	35.5%	5.9%	0.6M	39.1%	34.0%	3.4M	36.4%	–
'20–'21	60.0%	2.6M	15.5%	5.6%	0.2M	16.1%	34.4%	1.5M	16.1%	–

Notes: FUS = only US post-secondary education; CtS = foreign post-secondary education, then US terminal degree; CtW = only non-US post-secondary education, then US work.Entries are row percentages, person counts, and column percentages. Counts use the terminal (highest-ranked) degree per person. Doctorate comprises the PhD and doctorates that do not lead to a professional license (EdD, DBA, DMin, DMA, DSc, DrPH); Professional comprises licensure doctorates (MD, JD, DDS, PharmD, and similar). The ACS column reports doctorate degrees and professional degrees beyond a bachelor's degree, which follow the same division. The LinkedIn universe is restricted to terminal-degree graduation cohorts 2003–2021. The ACS column is an independent extract and is never pooled with the LinkedIn columns; it is retained only for the terminal-degree comparison because the supplied ACS extract has no comparable graduation-cohort measure. Percentages may not sum to 100 because of rounding.

Table 4: Post-Graduation Observation Years by Cohort and Status

	LinkedIn								
	Fully US-educated			Came-to-Study			Came-to-Work		
	17.0M			1.5M			9.4M		
Graduation/Entering Cohorts	(Observed Years) Total	After Grad/Entry	in US After Grad/Entry						
'03–'04	26.2	23.5	11.3	26.6	23.5	7.7	25.0	23.5	2.5
'05–'09	23.1	19.8	9.9	23.6	19.8	6.7	21.8	19.8	2.0
'10–'14	19.2	14.8	7.8	19.6	14.8	5.1	17.7	14.8	1.2
'15–'19	15.5	10.0	5.3	15.7	10.0	3.5	13.6	9.9	0.5
'20–'21	12.4	6.5	3.3	13.2	6.5	2.1	10.9	6.5	0.2

Notes: FUS = only US post-secondary education; CtS = foreign post-secondary education, then US terminal degree; CtW = only non-US post-secondary education, then US work.Entries are mean distinct calendar years in which a person appears in the panel: total observed years, observed years at or after terminal-degree graduation, and the latter restricted to `live_in_us_raw = 1`. The universe and the inclusive 2003–2021 graduation-cohort restriction are identical to the first table. The ACS has no comparable longitudinal observation and is therefore omitted.

Table 5: Post-Secondary Institution Selectivity Bins

Tier	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
Selectivity Ranking	Unranked	Open or not reported	Bottom 25%	75%–50%	50%–33%	33%–20%	20%–10%	10%–5%	5%–1%	Top 1%

Notes: A worker's tier is that of the institution that granted their terminal degree. Tier 0 (unranked): the terminal institution has no tier. Tier 1: the institution is matched but reports no admissions selectivity or admits all applicants.

Table 6: Probability of First Job Over-Education by Postgraduate Degree Holders, CtS vs. FUS

	(1)	(2)	(3)	(4)	(5)
			Pr(Over-Education)		
CtS	0.119***	0.114***	0.031***	0.033***	0.033***
	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)
Doctorate	-0.249***	–	–	–	–
	(0.001)				
Professional	-0.572***	–	–	–	–
	(0.001)				
CtS × Doctorate	-0.087***	-0.079***	-0.037***	0.025***	0.024***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
CtS × Professional	-0.078***	-0.072***	0.001	0.015***	0.012***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Cap-Exempt				-0.226***	-0.221***
				(0.001)	(0.001)
CtS × Cap-Exempt				-0.171***	-0.164***
				(0.002)	(0.002)
Hor. Mis.					0.084***
					(0.000)
Baseline Pr. (FUS, Master's)	0.796	0.796	0.796	0.796	0.796
N	4,229,371	4,229,371	4,229,367	4,229,367	4,229,367
Cohort FE	✓	✓	✓	✓	✓
Degree × School Tier		✓	✓	✓	✓
Degree × field FE, STEM, generality			✓	✓	✓
Employer block				✓	✓

Notes: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. FUS = only US post-secondary education; CtS = foreign post-secondary education, then US terminal degree; CtW = only non-US post-secondary education, then US work.Column 3 is the main specification. Doctorate pools the PhD and non-licensure doctorates; Professional denotes licensure doctorates (MD, JD, DDS, PharmD, and similar). The employer block contains log firm utility, 3-, 6-, 12-, and 24-month churn, an indicator for any H-1B filing, cap-exempt status, and its interaction with the pathway, so pathway coefficients in Columns 4 and 5 are gaps at cap-subject employers. From Column 2 on, degree main effects are absorbed by the degree-by-tier and degree-by-field effects and marked –. Baseline: share of FUS master's holders with the outcome.

Table 7: Probability of First Job Horizontal Mismatch by Postgraduate Degree Holders, CtS vs. FUS

	(1)	(2)	(3)	(4)	(5)
	Pr(Horizontal Mismatch)				
CtS	-0.030***	-0.025***	-0.011***	-0.008***	-0.012***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Doctorate	-0.054***	–	–	–	–
	(0.001)				
Professional	-0.003**	–	–	–	–
	(0.001)				
CtS × Doctorate	-0.005**	-0.008***	-0.015***	0.016***	0.013***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
CtS × Professional	0.012***	-0.000	0.022***	0.026***	0.024***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Cap-Exempt				-0.052***	-0.021***
				(0.001)	(0.001)
CtS × Cap-Exempt				-0.092***	-0.069***
				(0.002)	(0.002)
Over Edu.					0.135***
					(0.001)
Baseline Pr. (FUS, Master's)	0.374	0.374	0.374	0.374	0.374
N	4,229,371	4,229,371	4,229,367	4,229,367	4,229,367
Cohort FE	✓	✓	✓	✓	✓
Degree × School Tier		✓	✓	✓	✓
Degree × field FE, STEM, generality			✓	✓	✓
Employer block				✓	✓

Notes: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. FUS = only US post-secondary education; CtS = foreign post-secondary education, then US terminal degree; CtW = only non-US post-secondary education, then US work.Column 3 is the main specification. Doctorate pools the PhD and non-licensure doctorates; Professional denotes licensure doctorates (MD, JD, DDS, PharmD, and similar). The employer block contains log firm utility, 3-, 6-, 12-, and 24-month churn, an indicator for any H-1B filing, cap-exempt status, and its interaction with the pathway, so pathway coefficients in Columns 4 and 5 are gaps at cap-subject employers. From Column 2 on, degree main effects are absorbed by the degree-by-tier and degree-by-field effects and marked –. Baseline: share of FUS master's holders with the outcome.

Table 8: Probability of First Job Over-Education by Postgraduate Degree Holders, CtW vs. FUS

	(1)	(2)	(3)	(4)	(5)
	Pr(Over-Education)				
CtW	0.153***	0.150***	0.091***	0.084***	0.077***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Bachelor's	-0.439***	–	–	–	–
	(0.000)				
Doctorate	-0.249***	–	–	–	–
	(0.001)				
Professional	-0.572***	–	–	–	–
	(0.001)				
CtW × Bachelor's	-0.234***	-0.229***	-0.127***	-0.124***	-0.119***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
CtW × Doctorate	-0.149***	-0.145***	-0.068***	0.002	-0.002
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
CtW × Professional	0.063***	0.061***	0.118***	0.144***	0.127***
	(0.003)	(0.004)	(0.004)	(0.004)	(0.003)
Cap-Exempt				-0.161***	-0.160***
				(0.001)	(0.001)
CtW × Cap-Exempt				-0.190***	-0.156***
				(0.002)	(0.002)
Hor. Mis.					0.160***
					(0.000)
Baseline Pr. (FUS, Master's)	0.796	0.796	0.796	0.796	0.796
N	12,339,407	12,339,407	12,339,402	12,339,402	12,339,402
Cohort FE	✓	✓	✓	✓	✓
Degree × School Tier		✓	✓	✓	✓
Degree × field FE, STEM, generality			✓	✓	✓
Employer block				✓	✓

Notes: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. FUS = only US post-secondary education; CtS = foreign post-secondary education, then US terminal degree; CtW = only non-US post-secondary education, then US work.Column 3 is the main specification. Doctorate pools the PhD and non-licensure doctorates; Professional denotes licensure doctorates (MD, JD, DDS, PharmD, and similar). The employer block contains log firm utility, 3-, 6-, 12-, and 24-month churn, an indicator for any H-1B filing, cap-exempt status, and its interaction with the pathway, so pathway coefficients in Columns 4 and 5 are gaps at cap-subject employers. From Column 2 on, degree main effects are absorbed by the degree-by-tier and degree-by-field effects and marked –. Baseline: share of FUS master's holders with the outcome. Quadratic non-U.S. experience and within-employer U.S.-bound transfer are controlled for in all columns.

Table 9: Probability of First Job Horizontal Mismatch by Postgraduate Degree Holders, CtW vs. FUS

	(1)	(2)	(3)	(4)	(5)
	Pr(Horizontal Mismatch)				
CtW	0.047***	0.041***	0.042***	0.043***	0.025***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Bachelor's	0.046***	–	–	–	–
	(0.000)				
Doctorate	-0.054***	–	–	–	–
	(0.001)				
Professional	-0.002**	–	–	–	–
	(0.001)				
CtW × Bachelor's	-0.044***	-0.056***	-0.032***	-0.034***	-0.008***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
CtW × Doctorate	-0.057***	-0.056***	-0.052***	0.023***	0.023***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
CtW × Professional	0.025***	0.008*	0.091***	0.109***	0.079***
	(0.003)	(0.004)	(0.004)	(0.004)	(0.003)
Cap-Exempt				-0.007***	0.026***
				(0.001)	(0.001)
CtW × Cap-Exempt				-0.215***	-0.175***
				(0.002)	(0.002)
Over Edu.					0.208***
					(0.000)
Baseline Pr. (FUS, Master's)	0.374	0.374	0.374	0.374	0.374
N	12,339,407	12,339,407	12,339,402	12,339,402	12,339,402
Cohort FE	✓	✓	✓	✓	✓
Degree × School Tier		✓	✓	✓	✓
Degree × field FE, STEM, generality			✓	✓	✓
Employer block				✓	✓

Notes: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. FUS = only US post-secondary education; CtS = foreign post-secondary education, then US terminal degree; CtW = only non-US post-secondary education, then US work.Column 3 is the main specification. Doctorate pools the PhD and non-licensure doctorates; Professional denotes licensure doctorates (MD, JD, DDS, PharmD, and similar). The employer block contains log firm utility, 3-, 6-, 12-, and 24-month churn, an indicator for any H-1B filing, cap-exempt status, and its interaction with the pathway, so pathway coefficients in Columns 4 and 5 are gaps at cap-subject employers. From Column 2 on, degree main effects are absorbed by the degree-by-tier and degree-by-field effects and marked –. Baseline: share of FUS master's holders with the outcome. Quadratic non-U.S. experience and within-employer U.S.-bound transfer are controlled for in all columns.

Table 10: Over-Education Gap Relative to FUS Workers with the Same Degree, by Type of First Employer

		Cap-exempt	Cap-subject	Difference
Master's	CtS	5.1 [4.4, 5.9]	2.9 [2.8, 3.0]	2.3 [1.5, 3.0]
	CtW	0.4 [−0.7, 1.5]	8.7 [8.6, 8.8]	−8.3 [−9.4, −7.2]
Doctorate	CtS	−2.7 [−3.2, −2.3]	1.4 [1.0, 1.7]	−4.1 [−4.7, −3.5]
	CtW	−3.1 [−3.8, −2.4]	1.7 [1.3, 2.1]	−4.9 [−5.7, −4.0]
Professional	CtS	6.9 [4.7, 9.1]	3.1 [2.5, 3.7]	3.8 [1.5, 6.1]
	CtW	7.4 [5.3, 9.6]	22.2 [21.4, 22.9]	−14.7 [−17.0, −12.5]

Notes: FUS = only US post-secondary education; CtS = foreign post-secondary education, then US terminal degree; CtW = only non-US post-secondary education, then US work.Column 3 of Tables 6 and 8, estimated separately on workers whose first employer is cap-exempt and cap-subject. Cap-exempt employers account for 7.1, 4.1, and 1.3 percent of the first jobs of FUS, CtS, and CtW master's holders and 40.3, 36.9, and 36.4 percent of those of doctorate holders. Estimates are in percentage points, with 95 percent confidence intervals from person-clustered standard errors in brackets. The difference assumes the two samples share no clusters. Cap-exempt employers are higher-education institutions. The first employer is chosen in the same search as the job, so the split locates the gap; it is not an effect of cap exemption.

Table 11: First Job Occupational Utility Sorting of CtS Workers (vs. FUS)

	(1)	(2)	(3)	(4)	(5)	(6)
	ln(SOC PageRank Score)					
CtS	0.403***	0.320***	0.292***	0.230***	0.223***	0.219***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Doctorate	0.673***	0.618***	0.598***	0.571***	0.557***	0.555***
	(0.003)	(0.002)	(0.004)	(0.004)	(0.004)	(0.004)
Professional	1.167***	1.290***	1.335***	1.119***	1.126***	1.127***
	(0.003)	(0.002)	(0.006)	(0.005)	(0.005)	(0.005)
CtS × Doctorate	-0.219***	-0.211***	-0.191***	-0.163***	-0.167***	-0.195***
	(0.005)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
CtS × Professional	-0.437***	-0.402***	-0.364***	-0.291***	-0.284***	-0.286***
	(0.012)	(0.009)	(0.009)	(0.007)	(0.007)	(0.007)
Cap-Exempt						0.014***
						(0.002)
CtS × Cap-Exempt						0.087***
						(0.005)
Over Edu.		0.182***	0.175***	0.070***	0.064***	0.068***
		(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Hor. Mis.		-2.031***	-2.030***	-2.024***	-2.029***	-2.028***
		(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
OE × HM		0.150***	0.157***	0.481***	0.486***	0.486***
		(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
N	4,266,561	4,266,561	4,266,561	4,266,561	4,266,561	4,266,561
Cohort FE	✓	✓	✓	✓	✓	✓
Occ. Mis.		✓	✓	✓	✓	✓
Degree × School Tier			✓	✓	✓	✓
O*NET Job Char.				✓	✓	✓
Firm Utility.					✓	✓

Notes: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ FUS = only US post-secondary education; CtS = foreign post-secondary education, then US terminal degree; CtW = only non-US post-secondary education, then US work.Doctorate pools the PhD and non-licensure doctorates; Professional denotes licensure doctorates (MD, JD, DDS, PharmD, and similar). Unlike Column 3 of Tables 6–9, these specifications control for degree-by-tier effects but not for field of study, so the pathway gaps are not directly comparable with those tables. Occupational mismatch controls are indicator variables for over-education, horizontal mismatch, and both over-education and horizontal mismatch.Firm utility controls include firm pagerank utility a la Sorkin (2018) and firm 3-, 6-, 12-, and 24-month churn rates. Cap-exempt refers to H-1B cap-exempt status of the employer, proxied by institution of higher education in this table.O*NET job characteristics include the on-the-job training fixed effects, standardized importance of certification, and the three principal components labeled Cognitive and Interpersonal Capabilities, Attention and Coordination, and Practical Job Capabilities.

Table 12: First Job Occupational Utility Sorting of CtW Workers (vs. FUS)

	(1)	(2)	(3)	(4)	(5)	(6)
	ln(SOC PageRank Score)					
CtW	0.112***	0.267***	0.250***	0.144***	0.125***	0.115***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Bachelor's	-0.246***	-0.368***	-0.444***	-0.061***	-0.056***	-0.066***
	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)
Doctorate	0.674***	0.465***	0.427***	0.544***	0.531***	0.575***
	(0.003)	(0.002)	(0.004)	(0.005)	(0.005)	(0.005)
Professional	1.168***	0.897***	0.931***	1.086***	1.095***	1.090***
	(0.003)	(0.002)	(0.005)	(0.005)	(0.005)	(0.005)
CtW × Bachelor's	0.078***	-0.109***	-0.107***	-0.052***	-0.052***	-0.047***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
CtW × Doctorate	-0.074***	-0.243***	-0.228***	-0.152***	-0.146***	-0.172***
	(0.005)	(0.003)	(0.004)	(0.004)	(0.004)	(0.004)
CtW × Professional	-0.870***	-0.791***	-0.724***	-0.623***	-0.619***	-0.612***
	(0.011)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Cap-Exempt						-0.147***
						(0.002)
CtW × Cap-Exempt						0.086***
						(0.005)
Over Edu.		-0.456***	-0.458***	-0.102***	-0.106***	-0.113***
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Hor. Mis.		-1.761***	-1.761***	-1.710***	-1.713***	-1.713***
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
OE × HM		-0.023***	-0.019***	0.294***	0.297***	0.296***
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Non-US Exp.	-0.045***	-0.004**	-0.003**	-0.001	0.002	0.002
	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Non-US Exp.^{2}	0.012***	0.006***	0.006***	0.004***	0.003***	0.003***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Same Employer (Transfer)	0.263***	0.187***	0.186***	0.104***	0.079***	0.077***
	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
N	12,462,649	12,462,649	12,462,649	12,462,649	12,462,649	12,462,649
Market-Entry Cohort FE	✓	✓	✓	✓	✓	✓
Occ. Mis.		✓	✓	✓	✓	✓
Degree × School Tier			✓	✓	✓	✓
O*NET Job Char.				✓	✓	✓
Firm Utility.					✓	✓

Notes: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ FUS = only US post-secondary education; CtS = foreign post-secondary education, then US terminal degree; CtW = only non-US post-secondary education, then US work.The market-entry cohort is the year of the first US job for CtW workers and the graduation year for FUS workers. Doctorate pools the PhD and non-licensure doctorates; Professional denotes licensure doctorates (MD, JD, DDS, PharmD, and similar). Unlike Column 3 of Tables 6–9, these specifications control for degree-by-tier effects but not for field of study, so the pathway gaps are not directly comparable with those tables. Occupational mismatch controls are indicator variables for over-education, horizontal mismatch, and both over-education and horizontal mismatch.Firm utility controls include firm pagerank utility a la Sorkin (2018) and firm 3-, 6-, 12-, and 24-month churn rates. Cap-exempt refers to H-1B cap-exempt status of the employer, proxied by institution of higher education in this table.O*NET job characteristics include the on-the-job training fixed effects, standardized importance of certification, and the three principal components labeled Cognitive and Interpersonal Capabilities, Attention and Coordination, and Practical Job Capabilities. Quadratic non-U.S. experience and within-employer U.S.-bound transfer are controlled for in all columns.

Table 13: First U.S. Job Wage Sorting of CtS Workers (vs. FUS)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	ln(OEWS Median Wage)						
CtS	0.114***	0.110***	0.097***	0.030***	0.028***	0.013***	0.015***
	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)
Doctorate	-0.026***	-0.022***	-0.050***	-0.032***	-0.037***	-0.077***	-0.069***
	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
Professional	0.337***	0.322***	0.306***	0.251***	0.254***	0.175***	0.172***
	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
CtS × Doctorate	-0.119***	-0.115***	-0.103***	-0.062***	-0.063***	-0.053***	-0.038***
	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
CtS × Professional	-0.147***	-0.136***	-0.125***	-0.046***	-0.044***	-0.025***	-0.023***
	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)
Cap-Exempt							-0.039***
							(0.001)
CtS × Cap-Exempt							-0.046***
							(0.001)
Over Edu.		0.074***	0.069***	0.099***	0.098***	0.092***	0.087***
		(0.001)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)
Hor. Mis.		0.149***	0.148***	0.072***	0.071***	0.209***	0.209***
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
OE × HM		-0.371***	-0.366***	-0.111***	-0.110***	-0.144***	-0.145***
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
ln(SOC PageRank)						0.069***	0.069***
						(0.000)	(0.000)
N	3,977,618	3,977,618	3,977,618	3,977,618	3,977,618	3,977,618	3,977,618
Cohort FE	✓	✓	✓	✓	✓	✓	✓
Occ. Mis.		✓	✓	✓	✓	✓	✓
Degree × School Tier			✓	✓	✓	✓	✓
O*NET Job Char.				✓	✓	✓	✓
Firm Utility.					✓	✓	✓

Notes: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ FUS = only US post-secondary education; CtS = foreign post-secondary education, then US terminal degree; CtW = only non-US post-secondary education, then US work.Doctorate pools the PhD and non-licensure doctorates; Professional denotes licensure doctorates (MD, JD, DDS, PharmD, and similar). Unlike Column 3 of Tables 6–9, these specifications control for degree-by-tier effects but not for field of study, so the pathway gaps are not directly comparable with those tables. Firm utility controls include firm pagerank utility a la Sorkin (2018) and firm 3-, 6-, 12-, and 24-month churn rates. Cap-exempt refers to H-1B cap-exempt status of the employer, proxied by institution of higher education in this table.O*NET job characteristics include the on-the-job training fixed effects, standardized importance of certification, and the three principal components labeled Cognitive and Interpersonal Capabilities, Attention and Coordination, and Practical Job Capabilities.

Table 14: First U.S. Job Wage Sorting of CtW Workers (vs. FUS)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	ln(OEWS Median Wage)						
CtW	0.015***	0.086***	0.079***	0.029***	0.024***	0.018***	0.016***
	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)
Bachelor's	-0.183***	-0.347***	-0.352***	-0.016***	-0.015***	-0.013***	-0.019***
	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Doctorate	-0.026***	-0.129***	-0.173***	-0.071***	-0.075***	-0.105***	-0.085***
	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
Professional	0.337***	0.113***	0.095***	0.208***	0.211***	0.149***	0.147***
	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
CtW × Bachelor's	0.057***	-0.045***	-0.047***	-0.020***	-0.020***	-0.017***	-0.016***
	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)
CtW × Doctorate	-0.140***	-0.206***	-0.204***	-0.106***	-0.104***	-0.097***	-0.056***
	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
CtW × Professional	-0.148***	-0.102***	-0.129***	-0.108***	-0.107***	-0.072***	-0.058***
	(0.004)	(0.004)	(0.004)	(0.003)	(0.003)	(0.002)	(0.002)
Cap-Exempt							-0.069***
							(0.000)
CtW × Cap-Exempt							-0.108***
							(0.001)
Over Edu.		-0.350***	-0.350***	-0.017***	-0.018***	-0.013***	-0.019***
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Hor. Mis.		-0.040***	-0.040***	-0.018***	-0.019***	0.073***	0.072***
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
OE × HM		-0.157***	-0.155***	0.009***	0.010***	-0.006***	-0.007***
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
ln(SOC PageRank)						0.054***	0.054***
						(0.000)	(0.000)
Non-US Exp.	-0.002**	0.005***	0.005***	0.007***	0.008***	0.008***	0.008***
	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)
Non-US Exp.^{2}	0.004***	0.002***	0.002***	0.000*	0.000	-0.000	-0.000**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Same Employer (Transfer)	0.110***	0.095***	0.093***	0.042***	0.035***	0.031***	0.027***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
N	11,643,822	11,643,822	11,643,822	11,643,822	11,643,822	11,643,822	11,643,822
Market-Entry Cohort FE	✓	✓	✓	✓	✓	✓	✓
Occ. Mis.		✓	✓	✓	✓	✓	✓
Degree × School Tier			✓	✓	✓	✓	✓
O*NET Job Char.				✓	✓	✓	✓
Firm Utility.					✓	✓	✓

Notes: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ FUS = only US post-secondary education; CtS = foreign post-secondary education, then US terminal degree; CtW = only non-US post-secondary education, then US work.The market-entry cohort is the year of the first US job for CtW workers and the graduation year for FUS workers. Doctorate pools the PhD and non-licensure doctorates; Professional denotes licensure doctorates (MD, JD, DDS, PharmD, and similar). Unlike Column 3 of Tables 6–9, these specifications control for degree-by-tier effects but not for field of study, so the pathway gaps are not directly comparable with those tables. Firm utility controls include firm pagerank utility a la Sorkin (2018) and firm 3-, 6-, 12-, and 24-month churn rates. Cap-exempt refers to H-1B cap-exempt status of the employer, proxied by institution of higher education in this table.O*NET job characteristics include the on-the-job training fixed effects, standardized importance of certification, and the three principal components labeled Cognitive and Interpersonal Capabilities, Attention and Coordination, and Practical Job Capabilities. Quadratic non-U.S. experience and within-employer U.S.-bound transfer are controlled for in all columns.

Table 15: Decomposing CtS Worker SOC Median Wage Sorting, First U.S. Job

	(1) Master's	(2) Doctorate	(3) Professional
CtS	0.053***	-0.081***	0.233***
	(0.001)	(0.001)	(0.003)
FUS	-0.065***	-0.087***	0.261***
	(0.000)	(0.001)	(0.001)
Difference	0.118***	0.005***	-0.028***
	(0.001)	(0.001)	(0.003)
Explained	0.086***	0.028***	0.003**
	(0.000)	(0.001)	(0.001)
Unexplained	0.032***	-0.023***	-0.031***
	(0.001)	(0.001)	(0.003)
Explained, by covariate			
STEM	0.006***	-0.022***	0.000***
	(0.000)	(0.001)	(0.000)
Grad $\geq$ 2008	0.002***	0.001***	-0.000***
	(0.000)	(0.000)	(0.000)
STEM $\times$ Grad $\geq$ 2008	0.016***	0.016***	-0.002***
	(0.000)	(0.001)	(0.000)
Major Generality	-0.000***	-0.007***	0.003***
	(0.000)	(0.000)	(0.000)
Years since Degree	-0.005***	-0.002***	-0.008***
	(0.000)	(0.000)	(0.001)
Years since Degree^2	0.001***	-0.001***	0.006***
	(0.000)	(0.000)	(0.000)
Unexplained, by covariate			
STEM	-0.010***	0.002	0.001
	(0.002)	(0.008)	(0.004)
Grad $\geq$ 2008	0.003	0.012	0.016*
	(0.002)	(0.007)	(0.008)
STEM $\times$ Grad $\geq$ 2008	0.025***	-0.001	-0.004
	(0.002)	(0.007)	(0.004)
Major Generality	0.041***	-0.025**	-0.078***
	(0.003)	(0.008)	(0.015)
Years since Degree	-0.010***	-0.003	-0.015*
	(0.001)	(0.004)	(0.007)
Years since Degree^2	0.002**	0.007***	0.002
	(0.001)	(0.002)	(0.004)
N	3,590,577	440,037	505,626

Notes: Standard errors in parentheses. All columns use only the first job-year observation of a worker after their terminal degree. First jobs in the United States, 2003–2021 graduation cohorts. The outcome is the log OEWS median wage of the occupation, net of graduation-cohort effects estimated on the full panel, as in Figures 5 and 6 and the ten-year decompositions. Covariates are STEM, graduation in 2008 or later and its interaction with STEM, major generality, years since degree and its square, school tier, and major field; school-tier and major-field contributions are included in the totals but not shown. Doctorate pools the PhD and non-licensure doctorates; Professional denotes licensure doctorates (MD, JD, DDS, PharmD, and similar). FUS = only US post-secondary education; CtS = foreign post-secondary education, then US terminal degree; CtW = only non-US post-secondary education, then US work.* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 16: Decomposing CtW Worker SOC Median Wage Sorting, First U.S. Job

	(1) Master's	(2) Doctorate	(3) Professional
CtW	0.087***	-0.085***	0.287***
	(0.000)	(0.001)	(0.004)
FUS	0.026***	0.005***	0.353***
	(0.000)	(0.001)	(0.001)
Difference	0.061***	-0.090***	-0.066***
	(0.000)	(0.001)	(0.004)
Explained	0.074***	0.003**	0.107***
	(0.000)	(0.001)	(0.002)
Unexplained	-0.013***	-0.093***	-0.173***
	(0.001)	(0.002)	(0.004)
Explained, by covariate			
STEM	0.003***	-0.015***	0.001***
	(0.000)	(0.001)	(0.000)
Grad $\geq$ 2008	-0.002***	-0.002***	-0.000***
	(0.000)	(0.000)	(0.000)
STEM $\times$ Grad $\geq$ 2008	0.006***	0.010***	-0.007***
	(0.000)	(0.001)	(0.000)
Major Generality	-0.000**	-0.006***	0.012***
	(0.000)	(0.000)	(0.000)
Years since Degree	0.017***	0.009***	-0.069***
	(0.000)	(0.001)	(0.002)
Years since Degree^2	-0.003***	0.004***	0.059***
	(0.000)	(0.001)	(0.002)
Same Employer (Transfer)	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)
Unexplained, by covariate			
STEM	-0.004***	0.024***	0.018***
	(0.001)	(0.006)	(0.005)
Grad $\geq$ 2008	-0.016***	0.014*	0.071***
	(0.001)	(0.006)	(0.012)
STEM $\times$ Grad $\geq$ 2008	0.013***	-0.021***	-0.012**
	(0.001)	(0.005)	(0.005)
Major Generality	0.015***	-0.058***	-0.030
	(0.002)	(0.009)	(0.019)
Years since Degree	-0.016***	0.034***	-0.052**
	(0.002)	(0.006)	(0.017)
Years since Degree^2	0.011***	0.007	0.021
	(0.001)	(0.004)	(0.012)
Same Employer (Transfer)	0.004***	0.004***	-0.001
	(0.000)	(0.000)	(0.000)
N	4,099,978	452,277	502,536

Notes: Standard errors in parentheses. All columns use only the first US job-year observation of a worker after their terminal degree. First jobs in the United States, 2003–2021 graduation cohorts. The outcome is the log OEWS median wage of the occupation, net of market-entry-cohort effects estimated on the full panel, as in Figures 5 and 6 and the ten-year decompositions. Covariates are STEM, graduation in 2008 or later and its interaction with STEM, major generality, years since degree and its square, non-US experience and its square, a same-employer transfer indicator, school tier, and major field; school-tier and major-field contributions are included in the totals but not shown. Doctorate pools the PhD and non-licensure doctorates; Professional denotes licensure doctorates (MD, JD, DDS, PharmD, and similar). FUS = only US post-secondary education; CtS = foreign post-secondary education, then US terminal degree; CtW = only non-US post-secondary education, then US work.* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 17: Decomposing CtS Master's Degree Holder SOC Wage Distribution, First U.S. Job

	(1)	(2)	(3)	(4)	(5)
OEWS Quantile	10	25	50	75	90
CtS	0.081***	0.093***	0.099***	0.099***	0.100***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
FUS	-0.012***	-0.013***	-0.014***	-0.014***	-0.014***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Difference	0.092***	0.106***	0.113***	0.113***	0.113***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Explained	0.069***	0.079***	0.083***	0.081***	0.078***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Unexplained	0.023***	0.027***	0.030***	0.032***	0.035***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Explained, by covariate					
STEM	0.012***	0.010***	0.007***	0.003***	0.002***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Grad $\geq$ 2008	-0.000***	-0.000***	-0.000***	-0.000***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
STEM $\times$ Grad $\geq$ 2008	0.008***	0.012***	0.015***	0.016***	0.016***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Major Generality	0.003***	0.001***	-0.000***	-0.001***	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Years since Degree	-0.003***	-0.004***	-0.005***	-0.006***	-0.004***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Years since Degree^2	-0.000***	-0.000	0.000***	0.001***	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Unexplained, by covariate					
STEM	-0.008***	-0.009***	-0.010***	-0.009***	-0.010***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Grad $\geq$ 2008	-0.006**	-0.001	0.002	0.005*	0.007**
	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)
STEM $\times$ Grad $\geq$ 2008	0.025***	0.025***	0.025***	0.024***	0.028***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Major Generality	0.041***	0.041***	0.034***	0.019***	0.015***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Years since Degree	-0.004**	-0.006***	-0.009***	-0.012***	-0.006***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Years since Degree^2	-0.000	0.000	0.002*	0.004***	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
N	3,590,577	3,590,577	3,590,577	3,558,157	3,125,708

Notes: Standard errors in parentheses. All columns use only the first job-year observation of a worker after their terminal degree. First jobs in the United States, 2003–2021 graduation cohorts; each OEWS percentile is net of graduation-cohort effects. School-tier and major-field contributions are included in the explained and unexplained totals but not shown. FUS = only US post-secondary education; CtS = foreign post-secondary education, then US terminal degree; CtW = only non-US post-secondary education, then US work.Covariates are school tier, major field, major generality, STEM, graduation in 2008 or later and its interaction with STEM, and years since degree and its square. Outcome is net of terminal degree graduation cohort fixed-effects.* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 18: STEM-OPT and Cap-Exempt First Employers

Panel A. Type of first US employer

	Master's		Doctorate	
	Pre-period share	DD	Pre-period share	DD
Cap-exempt	0.036	−0.002	0.345	0.006
		(0.004)		(0.018)
Cap-subject	0.849	0.011	0.569	−0.000
		(0.010)		(0.017)
Unidentified	0.115	−0.009	0.087	−0.006
		(0.007)		(0.007)

Panel B. Stayer outcomes at $t = 5$, by type of first employer

	Master's		Doctorate	
	Cap-subject	Cap-exempt	Cap-subject	Cap-exempt
Log occupational median wage	0.114	0.069	0.031	0.035
	(0.025)	(0.029)	(0.019)	(0.014)
Over-education	0.024	−0.034	−0.021	−0.046
	(0.013)	(0.028)	(0.017)	(0.017)
Horizontal mismatch	−0.132	−0.069	−0.087	−0.017
	(0.026)	(0.048)	(0.030)	(0.019)
Log occupational PageRank	0.409	0.075	0.224	0.169
	(0.084)	(0.114)	(0.062)	(0.047)
Log firm PageRank	0.633	0.108	0.473	−0.026
	(0.136)	(0.184)	(0.157)	(0.081)
Smallest STEM × Post group	9,858	410	1,181	1,173

Notes: FUS = only US post-secondary education; CtS = foreign post-secondary education, then US terminal degree; CtW = only non-US post-secondary education, then US work.CtS graduates whose first job is in the United States, 2003–2006 and 2008–2021 graduation cohorts. Doctorate pools the PhD and non-licensure doctorates. Panel A: share of US first jobs at each type of employer among pre-2008 STEM graduates, and the STEM × Post difference-in-differences in that share. The three difference-in-differences sum to zero. Panel B: Equation (10) at $t = 5$, estimated separately by the type of first employer; graduates whose first employer is unidentified are excluded. Standard errors, clustered by major, in parentheses. The smallest group is the smallest STEM × Post group across the five outcomes.

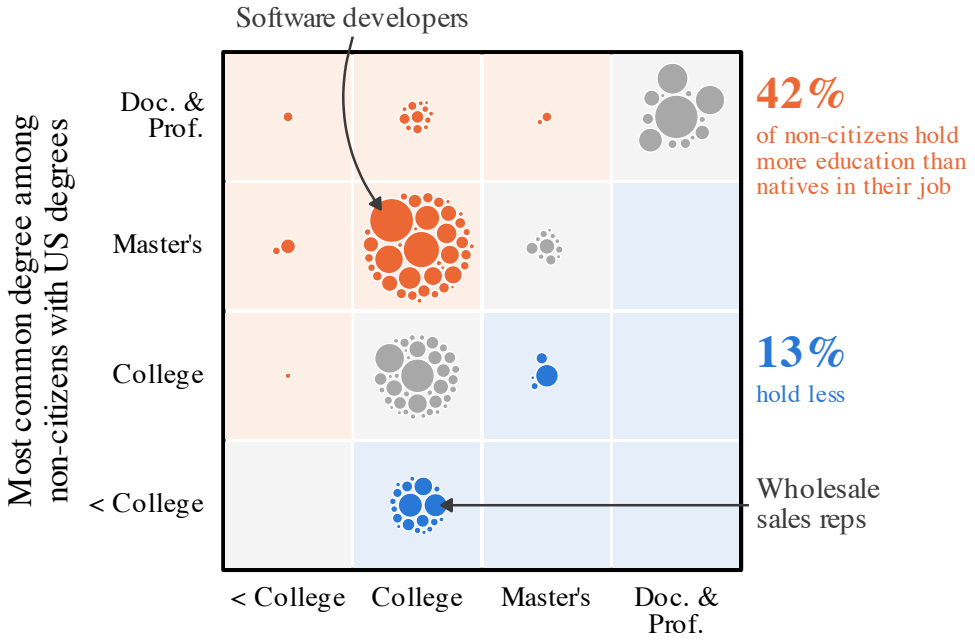
Table 19: STEM-OPT and the Jobs of Master's Stayers Five Years after Graduation

	Graduates of 2008–2011	Graduates of 2008–2021
Panel A. Stayer outcomes at $t = 5$		
Over-educated (pp)	0.5	1.9
	[−0.6, 1.6]	[−0.4, 4.2]
Horizontally mismatched (pp)	−7.6	−13.1
	[−10.6, −4.7]	[−18.2, −7.9]
Log occupational PageRank	0.19	0.39
	[0.10, 0.28]	[0.23, 0.56]
Log firm PageRank	0.21	0.61
	[0.10, 0.33]	[0.34, 0.88]
Log occupational median wage	0.08	0.11
	[0.05, 0.11]	[0.06, 0.16]
Panel B. Entry and retention		
US placement at graduation (pp)	3.7	5.6
	[1.9, 5.4]	[3.4, 7.9]
Still in the US at $t = 5$ (pp)	−0.1	1.0
	[−0.7, 0.5]	[0.5, 1.5]

Notes: FUS = only US post-secondary education; CtS = foreign post-secondary education, then US terminal degree; CtW = only non-US post-secondary education, then US work.Difference-in-differences, STEM versus non-STEM CtS master's graduates, post- versus pre-2008 cohorts (2003–2006; 2007 excluded). Panel A: Equation (10) at $t = 5$ among graduates working in the United States. US placement: Equation (9). Still in the US: Kaplan–Meier retention among US starters (Table G2). 95 percent confidence intervals in brackets, clustered by terminal-degree major except for retention, which uses Greenwood standard errors.

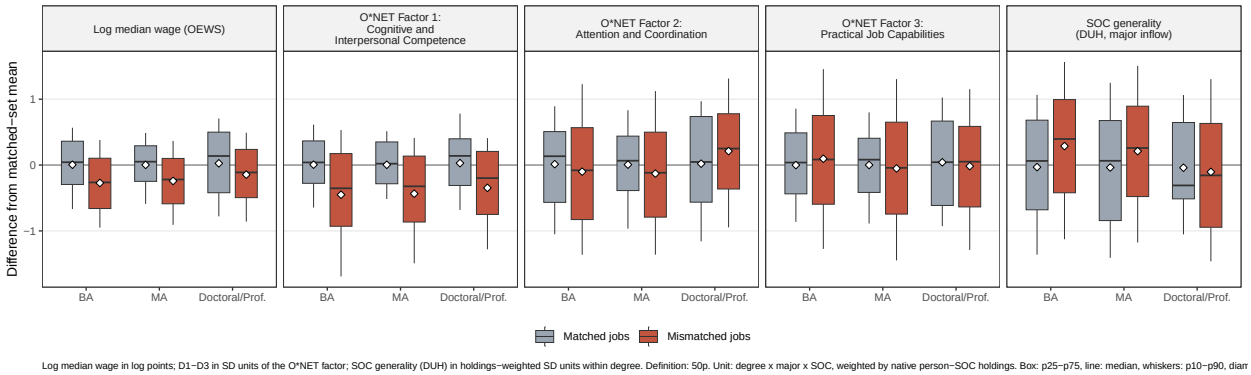
10 Figures

Figure 16: Occupational Mismatch of Non-Citizens with US Degrees



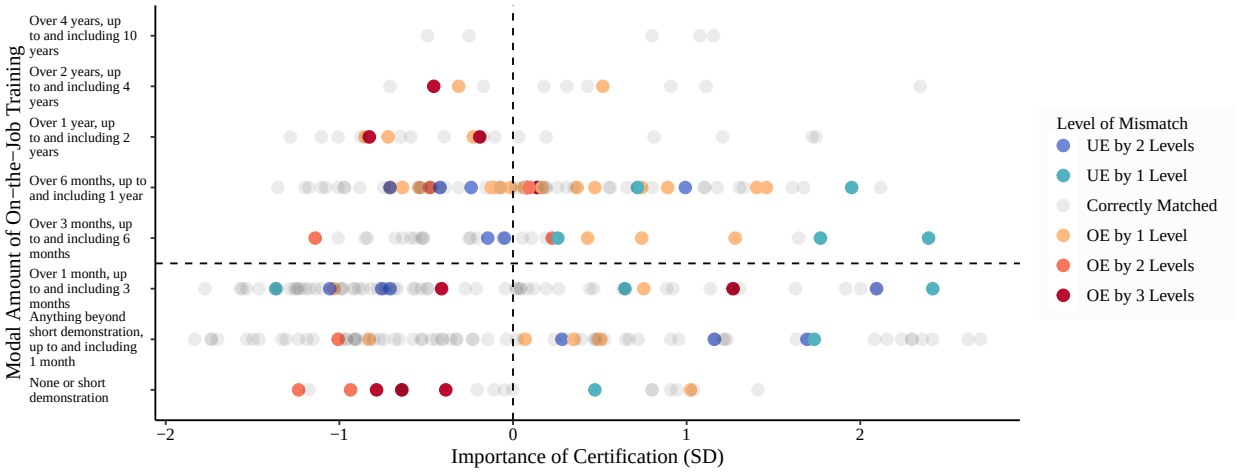
Notes: ACS 2010–2019. Wage and salary workers at work 50–52 weeks, ages 18–65, not enrolled in school. Natives are citizens by birth; non-citizens with US degrees are non-citizens who arrived before the typical completion age of their highest degree. Each circle is one of the 142 occupations with at least 30 respondents in each group and a most common degree of college or above in either group, sized by non-citizen employment. Orange circles: the non-citizens' most common degree is higher; blue: lower; gray: equal. Shares are weighted by non-citizen employment.

Figure 17: What Horizontally Mismatched Jobs Look Like



Notes: Occupations held by FUS workers, by degree-by-major group. Matched jobs are the occupations in the smallest set that holds half of the group's FUS jobs (Section 3.3); mismatched jobs are all others. Each outcome is the occupation's value minus the mean of the group's matched set, weighted by FUS holdings. Log median wage is the national OEWS median; the three O*NET factors (Section 3.2) and SOC generality are in standard-deviation units. Doctoral/Prof. pools doctorate and professional degrees. Boxes span the 25th to 75th percentiles, lines mark medians, whiskers the 10th and 90th percentiles, and diamonds means.

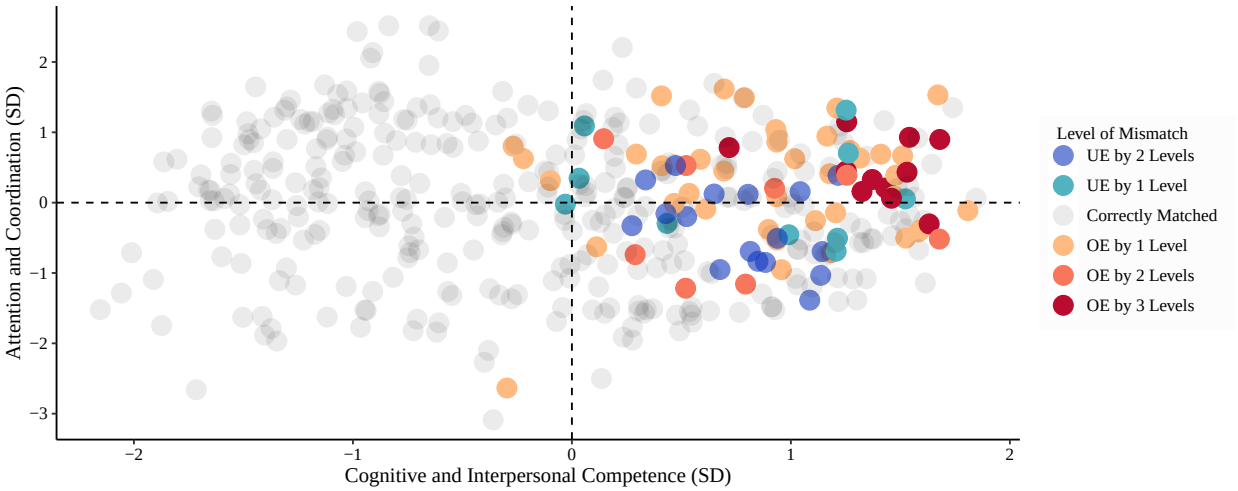
Figure 18: Educational Mismatch of Non-Citizens, Highest Degree Obtained in US



Notes: Level of mismatch is the difference between the modal degree of a non−citizen worker, whose highest degree is obtained inside the United States, and the modal degree of a citizen worker. The degree levels, from highest to lowest, are: Doctorate, Professional, Master's, Bachelor's, Associates, High School, and less than High School. Job characteristics are from the O*NET dataset. Mismatch realizations are from the American Community Survey 2010−2019.

Notes: ACS 2010–2019 and O*NET. Each circle is an occupation. Mismatch is the difference between the modal degree of non-citizen workers whose highest degree was obtained in the United States and the modal degree of citizen workers in the occupation. OE and UE denote over- and under-education by that many levels; the levels, from highest to lowest, are doctorate, professional, master's, bachelor's, associate's, high school, and less than high school. The horizontal axis is the standardized O*NET importance of certification, and the vertical axis is the occupation's modal amount of on-the-job training. The dashed horizontal line separates occupations with more than three months of training from the rest, and the dashed vertical line marks zero.

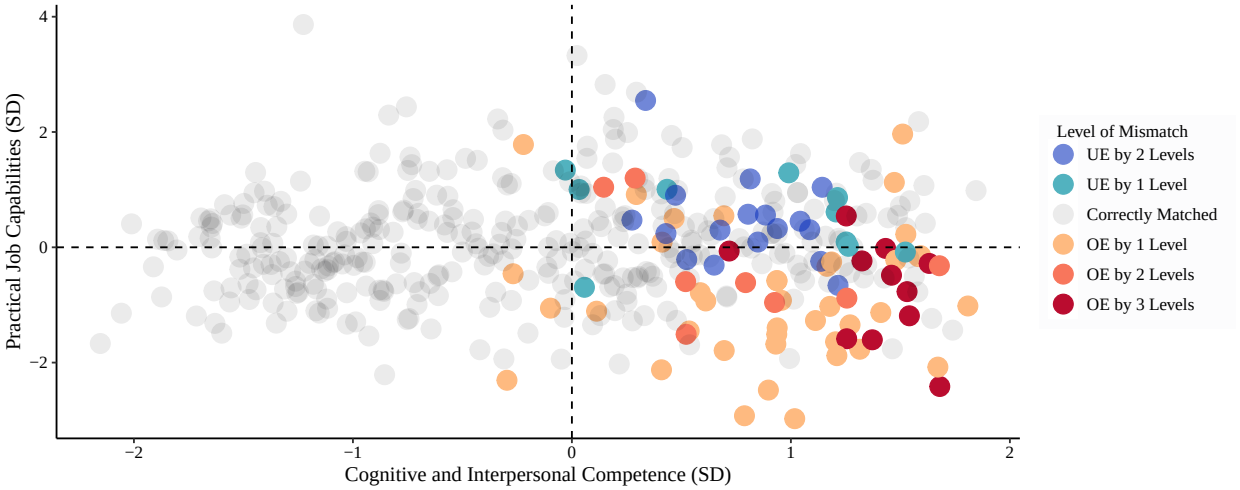
Figure 19: Occupations with Non-Citizen Mismatch Require Above-Average Cognitive and Interpersonal Competence



Notes: Level of mismatch is the difference between the modal degree of a non−citizen worker, whose highest degree is obtained inside the United States, and the modal degree of a citizen worker. The degree levels, from highest to lowest, are: Doctorate, Professional, Master's, Bachelor's, Associates, High School, and less than High School.

Notes: ACS 2010–2019 and O*NET. Each circle is an occupation. Mismatch is the difference between the modal degree of non-citizen workers whose highest degree was obtained in the United States and the modal degree of citizen workers in the occupation. OE and UE denote over- and under-education by that many levels; the levels, from highest to lowest, are doctorate, professional, master's, bachelor's, associate's, high school, and less than high school. Axes are the standardized principal components of Section 3.2, and dashed lines mark zero.

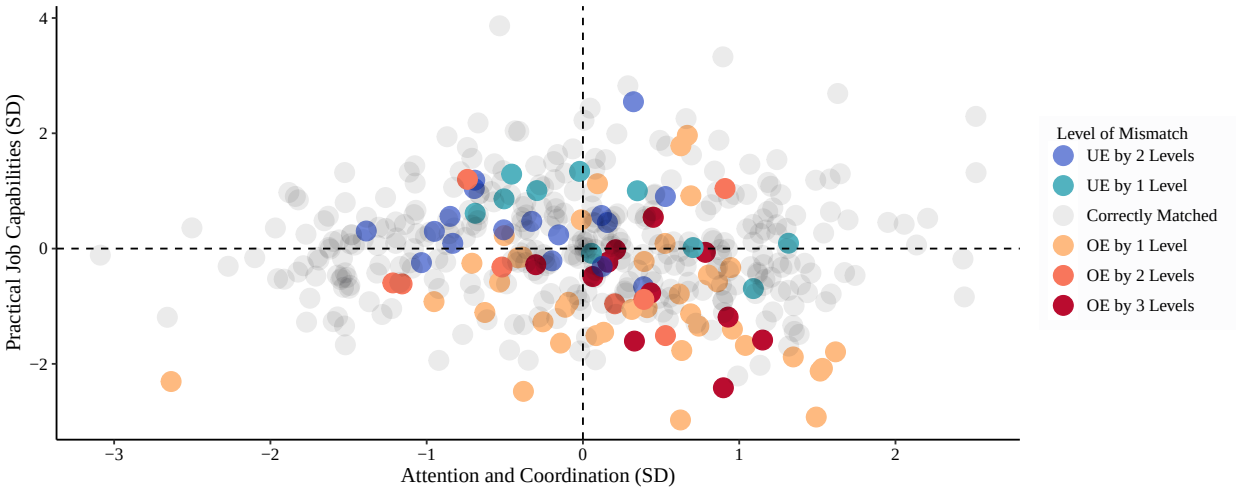
Figure 20: Negative Relationship between Over-Education and Job Requirements of Practical Capabilities



Notes: Level of mismatch is the difference between the modal degree of a non−citizen worker, whose highest degree is obtained inside the United States, and the modal degree of a citizen worker. The degree levels, from highest to lowest, are: Doctorate, Professional, Master's, Bachelor's, Associates, High School, and less than High School.

Notes: ACS 2010–2019 and O*NET. Each circle is an occupation. Mismatch is the difference between the modal degree of non-citizen workers whose highest degree was obtained in the United States and the modal degree of citizen workers in the occupation. OE and UE denote over- and under-education by that many levels; the levels, from highest to lowest, are doctorate, professional, master's, bachelor's, associate's, high school, and less than high school. Axes are the standardized principal components of Section 3.2, and dashed lines mark zero.

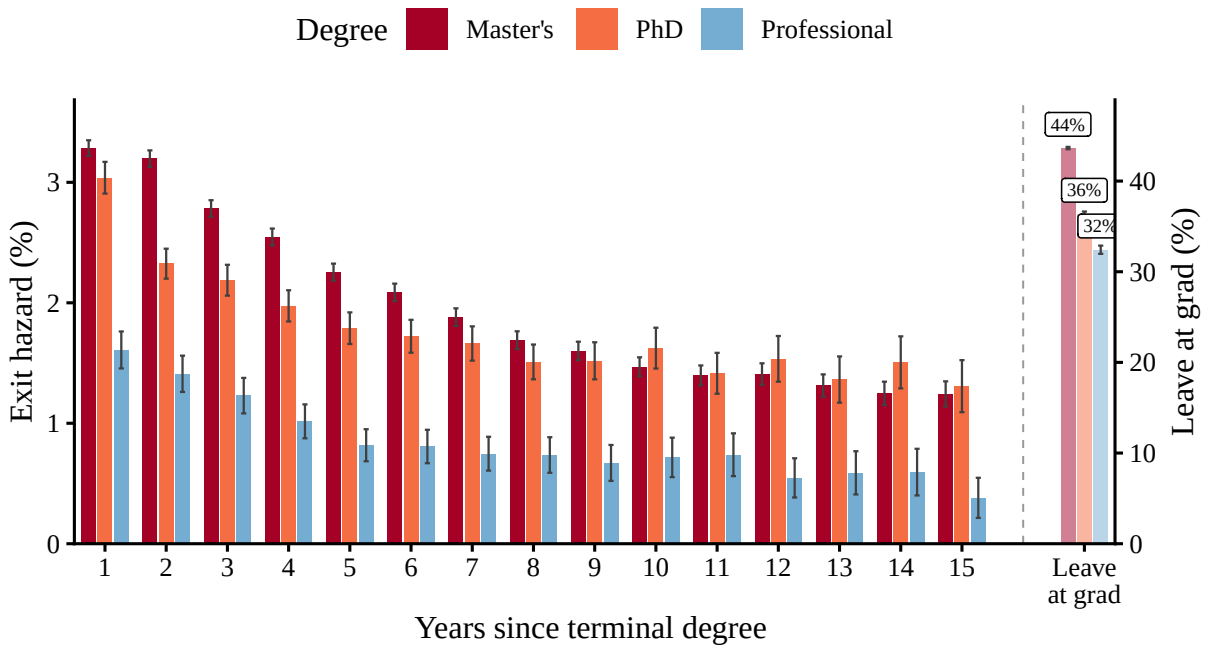
Figure 21: No Clear Relationship between Educational Mismatch and Job Requirements of Attention and Coordination



Notes: Level of mismatch is the difference between the modal degree of a non−citizen worker, whose highest degree is obtained inside the United States, and the modal degree of a citizen worker. The degree levels, from highest to lowest, are: Doctorate, Professional, Master's, Bachelor's, Associates, High School, and less than High School.

Notes: ACS 2010–2019 and O*NET. Each circle is an occupation. Mismatch is the difference between the modal degree of non-citizen workers whose highest degree was obtained in the United States and the modal degree of citizen workers in the occupation. OE and UE denote over- and under-education by that many levels; the levels, from highest to lowest, are doctorate, professional, master's, bachelor's, associate's, high school, and less than high school. Axes are the standardized principal components of Section 3.2, and dashed lines mark zero.

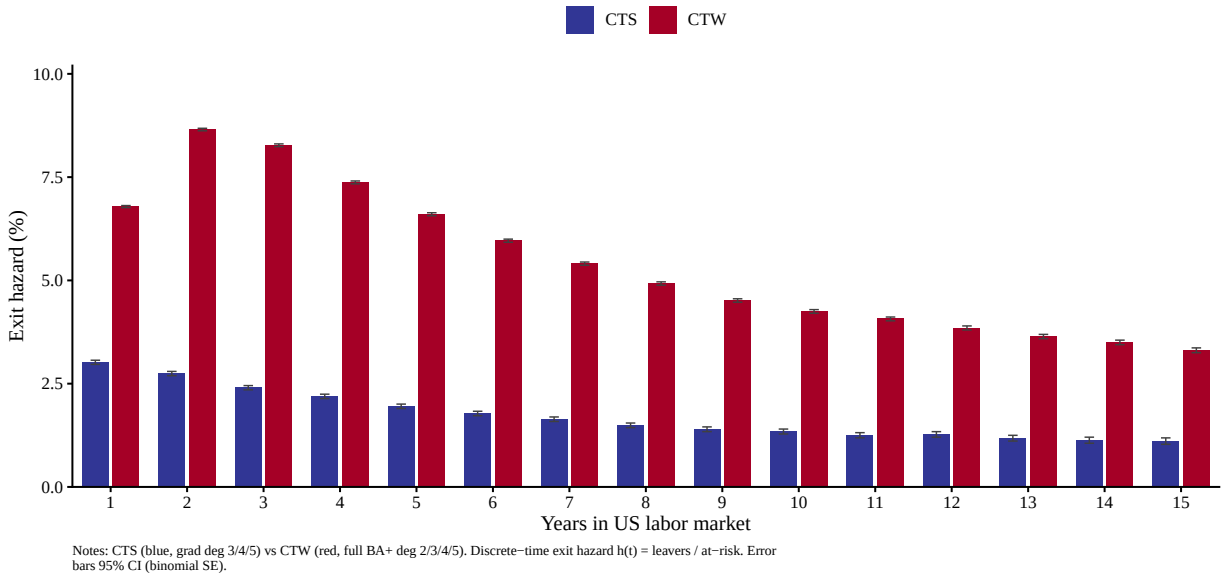
Figure 22: Low Exits After Initial US Placement, Came-to-Study, 1970-2025



Notes: Came to Study. Bars grouped by degree within each year. Years 1+: discrete−time exit hazard h(ce) = leavers / at−risk; at−risk conditions on being in US and observed (right−censored); left axis. Rightmost bars: leave−at−graduation, share not in US at graduation; right axis. Error bars 95% CI (binomial SE).

Notes: CtS graduates, pooled across STEM and non-STEM majors; t counts years since the terminal degree. Bars for years 1 to 15, on the left axis, are discrete-time exit hazards: workers who leave the United States in year t divided by workers in the United States and observed in year t. Hazard bars with fewer than 100 workers at risk are omitted. Error bars are 95 percent confidence intervals from binomial standard errors. The bars on the right, on the right axis, are the share of graduates not in the US at graduation, among all graduates observed at $t = 0$. Doctorate pools the PhD and non-licensure doctorates.

Figure 23: Exit Hazard Rates by Years in US Labor Market, 1970-2025



Notes: CtS workers with a graduate degree, for whom t counts years since the terminal degree, and CtW workers with a bachelor's degree or above, for whom t counts years since the first US job. Bars for years 1 to 15, on the left axis, are discrete-time exit hazards: workers who leave the United States in year t divided by workers in the United States and observed in year t. Hazard bars with fewer than 100 workers at risk are omitted. Error bars are 95 percent confidence intervals from binomial standard errors. The bar on the right, on the right axis, is the share of CtS graduates not in the US at graduation. CtW workers enter the sample at their first US job, so they have no such share.

A Appendix Tables

Table A1: Occupational Quantile Example

Occupation	Doc./Pro.	MA	BA	Assoc.	HS	<HS	Occ. Quantile
Surgeon	99.5	0.5	0	0	0	0	100
Post-Secondary Teacher	37.8	33.2	19.4	2.4	6.8	0.4	97.1
Sales Managers	1.0	10.8	50.2	7.5	29.5	1.0	61.0
Respiratory Therapists	0.9	3.1	25.0	56.6	14.1	0.3	59.5
Fence Erectors	0	0.7	3.7	4.4	69.4	21.8	3.0
Earth Drillers, not Oil	0	0.2	4.0	6.0	76.3	13.5	2.4

Table A2: Age-Education Schedule

Highest Degree	Age at Completion
Kindergarten or below	6
First Grade	7
Second Grade	8
Third Grade	9
Fourth Grade	10
Fifth Grade	11
Sixth Grade	12
Seventh Grade	13
Eighth Grade	14
Ninth Grade	15
Tenth Grade	16
Eleventh Grade	17
Twelfth Grade or High School Diploma	18
GED or Some College	19
Associate's	20
Bachelor's Degree	22
Master's Degree	24
Professional Degree Beyond Bachelor's	28
Doctoral Degree	30

Table A3: List of Skills and Abilities used to Label Each Principal Component

Cognitive and Interpersonal Competence	**Attention and Coordination**	**Practical Job Capabilities**
Written Expression	*Visualization*	*Programming*
Speaking	*Operations Monitoring*	*Service Orientation*
Active Listening	*Quality Control Analysis*	*Social Perceptiveness*
Reading Comprehension	*Perceptual Speed*	*Time Sharing*
Writing	*Troubleshooting*	*Explosive Strength*
Active Learning	*Visual Color Discrimination*	*Stamina*
Written Comprehension	*Operation and Control*	*Coordination*
Oral Expression	*Equipment Selection*	*Gross Body Coordination*
Critical Thinking	*Depth Perception*	*Equipment Selection*
Judgment and Decision Making	*Hearing Sensitivity*	*Gross Body Equilibrium*
Oral Comprehension	*Auditory Attention*	*Negotiation*
Deductive Reasoning	*Far Vision*	*Repairing*
Speech Clarity	*Repairing*	*Static Strength*
Systems Evaluation	*Equipment Maintenance*	*Near Vision*
Complex Problem Solving	*Reaction Time*	*Technology Design*
Systems Analysis	*Selective Attention*	*Equipment Maintenance*
Fluency of Ideas	*Response Orientation*	*Installation*
Speech Recognition	*Flexibility of Closure*	*Trunk Strength*
Monitoring	*Technology Design*	*Persuasion*
Inductive Reasoning	*Rate Control*	*Speech Recognition*

Table A4: Example of Predicted SOC Codes using Preliminarily Fine-Tuned Llama-3.1

LinkedIn Title	**O*NET SOC**	**Corresponding O*NET Title**
director-field support	27-2012.05	Media Technical Directors/Managers
senior page designer and copy editor	27-3041.00	Editors
gfliwn askdm	00-0000.00	NA
taylor mattis fellow	51-9195.05	Potters, Manufacturing
solar analyst intern	17-2199.11	Solar Energy Systems Engineers
international visiting attorney	23-1011.00	Lawyers
director of lower school admission	11-9032.00	Education Administrators, Kindergarten through Secondary
asset & reliability engineer	17-2199.00	NA
sales / acct manager	41-1011.00	First-Line Supervisors of Retail Sales Workers
attest intern	13-2099.00	NA
immersion associate	25-1194.00	Career/Technical Education Teachers, Postsecondary
medicaid waiver case manager	11-9151.00	Social and Community Service Managers
網頁工程師	15-1299.08	Computer Systems Engineers/Architects
博士生	25-9044.00	Teaching Assistants, Postsecondary

Table A5: Gelbach Decomposition of the First-Job Occupational-Wage Gap, CtS vs. FUS

	Master's	Doctorate	Professional
Raw gap (no controls)	0.1189*** (0.0007)	-0.0007 (0.0015)	-0.0308*** (0.0034)
Major Generality	-0.0030*** (0.0000)	-0.0018*** (0.0001)	0.0038*** (0.0002)
STEM	-0.0040*** (0.0003)	-0.0158*** (0.0005)	-0.0002*** (0.0001)
Grad $\geq$ 2008	0.0024*** (0.0001)	0.0007*** (0.0001)	0.0004*** (0.0001)
STEM $\times$ Grad $\geq$ 2008	0.0084*** (0.0003)	0.0134*** (0.0005)	-0.0005*** (0.0001)
English-Speaking Origin	0.0024*** (0.0004)	0.0027** (0.0010)	-0.0032* (0.0013)
Years since Degree	-0.0033*** (0.0001)	-0.0012*** (0.0001)	-0.0020*** (0.0002)
Years since Degree^2	0.0010*** (0.0001)	0.0005*** (0.0001)	0.0016*** (0.0001)
Skill Factor 1	0.0079*** (0.0003)	-0.0129*** (0.0005)	-0.0064*** (0.0012)
Skill Factor 2	0.0233*** (0.0002)	0.0229*** (0.0006)	-0.0101*** (0.0013)
Skill Factor 3	0.0317*** (0.0002)	0.0089*** (0.0004)	0.0045*** (0.0003)
Cert. Importance	-0.0069*** (0.0001)	-0.0156*** (0.0002)	-0.0092*** (0.0009)
SOC Generality	-0.0006*** (0.0000)	-0.0178*** (0.0003)	-0.0092*** (0.0005)
H-1B Demand Decile 1 (of 10)	0.0004*** (0.0000)	-0.0004*** (0.0000)	-0.0001 (0.0001)
H-1B Demand Decile 2 (of 10)	0.0001*** (0.0000)	-0.0005*** (0.0001)	-0.0001 (0.0001)
H-1B Demand Decile 3 (of 10)	0.0006*** (0.0000)	-0.0025*** (0.0002)	-0.0004*** (0.0001)
H-1B Demand Decile 4 (of 10)	0.0019*** (0.0000)	-0.0016*** (0.0002)	-0.0006*** (0.0001)
H-1B Demand Decile 5 (of 10)	0.0037*** (0.0000)	0.0011*** (0.0001)	-0.0001 (0.0001)
H-1B Demand Decile 6 (of 10)	0.0006*** (0.0000)	-0.0029*** (0.0001)	0.0004*** (0.0001)
H-1B Demand Decile 7 (of 10)	0.0011*** (0.0000)	-0.0001 (0.0001)	0.0004*** (0.0001)
H-1B Demand Decile 8 (of 10)	0.0015*** (0.0000)	-0.0025*** (0.0001)	-0.0015*** (0.0004)
H-1B Demand Decile 9 (of 10)	-0.0018*** (0.0000)	-0.0026*** (0.0001)	-0.0000 (0.0003)
H-1B Demand Decile 10 (of 10)	0.0204*** (0.0001)	0.0119*** (0.0002)	-0.0005* (0.0002)
Firm Util.	0.0030*** (0.0000)	-0.0011*** (0.0001)	0.0002* (0.0001)
Churn 3mo	0.0000 (0.0000)	0.0000** (0.0000)	0.0000 (0.0000)
Churn 6mo	0.0004*** (0.0000)	0.0000 (0.0000)	0.0000 (0.0000)
Churn 12mo	-0.0001*** (0.0000)	0.0003*** (0.0000)	0.0000 (0.0000)
Churn 24mo	0.0003*** (0.0000)	-0.0002*** (0.0000)	-0.0000 (0.0000)
Employer Ever Sponsors H-1B	0.0000 (0.0000)	0.0010*** (0.0001)	-0.0001* (0.0000)
SOC Util.	0.0244*** (0.0002)	0.0148*** (0.0004)	-0.0026*** (0.0005)
OE$\times$ ln(Edu. Signal)	0.0000 (0.0000)	0.0042*** (0.0002)	-0.0001 (0.0002)
HM$\times$ ln(Edu. Signal)	-0.0001*** (0.0000)	0.0001 (0.0001)	0.0011*** (0.0003)
Over Edu.	-0.0094*** (0.0001)	0.0006*** (0.0001)	0.0007** (0.0003)
Hor. Mismatch	-0.0023*** (0.0001)	-0.0052*** (0.0002)	-0.0012** (0.0004)
Cap-Exempt Employer (PDL Educational)	0.0018*** (0.0000)	0.0002*** (0.0000)	0.0009*** (0.0001)
Total Explained ($b_0 - b_{\text{full}}$)	0.1188*** (0.0007)	0.0109*** (0.0016)	-0.0324*** (0.0033)
Final gap (full controls)	0.0002 (0.0006)	-0.0117*** (0.0012)	0.0016 (0.0018)
N (common sample)	3,128,221	398,464	415,174

Notes: Gelbach (2016) decomposition of the gap in the log OEWS median wage of the occupation held at the first US job between CtS and FUS workers with the same terminal degree. Rows are the contributions of each covariate, in log units, with standard errors in parentheses; the raw gap minus the total explained equals the final gap. School-tier and major-field contributions are included in the total but not shown. Figure 5 groups the contributions into the blocks defined in Section 4.1.1. Doctorate pools the PhD and non-licensure doctorates; Professional denotes licensure doctorates. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A6: Gelbach Decomposition of the First-Job Occupational-Wage Gap, CtW vs. FUS

	Master's	Doctorate	Professional
Raw gap (no controls)	0.0584*** (0.0005)	-0.0923*** (0.0015)	-0.0816*** (0.0042)
Major Generality	-0.0003*** (0.0000)	-0.0009*** (0.0001)	0.0132*** (0.0003)
STEM	-0.0020*** (0.0001)	-0.0107*** (0.0004)	-0.0010*** (0.0002)
Grad $\geq$ 2008	-0.0019*** (0.0000)	-0.0009*** (0.0001)	0.0015*** (0.0001)
STEM $\times$ Grad $\geq$ 2008	0.0033*** (0.0001)	0.0079*** (0.0003)	-0.0022*** (0.0002)
Same Employer (Transfer)	0.0009*** (0.0001)	0.0012*** (0.0001)	-0.0006* (0.0003)
English-Speaking Origin	0.0056*** (0.0003)	-0.0012 (0.0008)	-0.0226*** (0.0022)
Non-US Labor Mkt. Exp.	0.0275*** (0.0011)	0.0416*** (0.0026)	-0.0093 (0.0079)
Non-US Labor Mkt. Exp.^{2}	-0.0024** (0.0008)	-0.0068*** (0.0017)	0.0137* (0.0056)
Skill Factor 1	-0.0106*** (0.0002)	-0.0323*** (0.0005)	-0.0628*** (0.0014)
Skill Factor 2	-0.0053*** (0.0002)	-0.0021*** (0.0005)	0.0248*** (0.0013)
Skill Factor 3	0.0163*** (0.0001)	0.0063*** (0.0003)	0.0260*** (0.0005)
Cert. Importance	-0.0066*** (0.0001)	-0.0229*** (0.0003)	-0.0503*** (0.0012)
SOC Generality	0.0030*** (0.0000)	-0.0227*** (0.0003)	-0.0584*** (0.0009)
H-1B Demand Decile 1 (of 10)	0.0001*** (0.0000)	-0.0003*** (0.0000)	-0.0010*** (0.0001)
H-1B Demand Decile 2 (of 10)	-0.0001*** (0.0000)	-0.0012*** (0.0001)	-0.0017*** (0.0002)
H-1B Demand Decile 3 (of 10)	0.0002*** (0.0000)	-0.0045*** (0.0002)	-0.0034*** (0.0002)
H-1B Demand Decile 4 (of 10)	0.0010*** (0.0000)	-0.0060*** (0.0002)	-0.0045*** (0.0003)
H-1B Demand Decile 5 (of 10)	0.0015*** (0.0000)	0.0010*** (0.0000)	-0.0014*** (0.0002)
H-1B Demand Decile 6 (of 10)	0.0003*** (0.0000)	-0.0045*** (0.0001)	0.0005*** (0.0001)
H-1B Demand Decile 7 (of 10)	0.0007*** (0.0000)	-0.0011*** (0.0001)	0.0015*** (0.0001)
H-1B Demand Decile 8 (of 10)	0.0007*** (0.0000)	-0.0016*** (0.0001)	0.0031*** (0.0004)
H-1B Demand Decile 9 (of 10)	-0.0014*** (0.0000)	-0.0034*** (0.0001)	0.0073*** (0.0004)
H-1B Demand Decile 10 (of 10)	0.0129*** (0.0001)	0.0060*** (0.0002)	-0.0063*** (0.0002)
Firm Util.	0.0034*** (0.0001)	-0.0017*** (0.0001)	0.0029*** (0.0001)
Churn 3mo	0.0000 (0.0000)	-0.0000** (0.0000)	-0.0001*** (0.0000)
Churn 6mo	0.0001*** (0.0000)	-0.0000 (0.0000)	0.0001*** (0.0000)
Churn 12mo	0.0001*** (0.0000)	0.0006*** (0.0001)	-0.0000 (0.0000)
Churn 24mo	0.0002*** (0.0000)	-0.0007*** (0.0000)	-0.0000 (0.0000)
Employer Ever Sponsors H-1B	0.0002*** (0.0000)	0.0002*** (0.0000)	-0.0001 (0.0001)
SOC Util.	0.0067*** (0.0001)	0.0024*** (0.0004)	-0.0336*** (0.0006)
OE$\times$ ln(Edu. Signal)	-0.0015*** (0.0001)	0.0031*** (0.0002)	-0.0000 (0.0001)
HM$\times$ ln(Edu. Signal)	0.0006*** (0.0001)	0.0003* (0.0001)	-0.0006*** (0.0002)
Over Edu.	-0.0112*** (0.0001)	0.0001 (0.0000)	0.0127*** (0.0005)
Hor. Mismatch	0.0064*** (0.0001)	0.0017*** (0.0003)	0.0087*** (0.0004)
Cap-Exempt Employer (PDL Educational)	0.0034*** (0.0000)	0.0003*** (0.0000)	0.0021*** (0.0001)
Total Explained ($b_0 - b_{\text{full}}$)	0.0727*** (0.0007)	-0.0439*** (0.0019)	-0.1160*** (0.0051)
Final gap (full controls)	-0.0142*** (0.0007)	-0.0484*** (0.0016)	0.0344*** (0.0039)
N (common sample)	3,602,198	410,236	413,798

Notes: Gelbach (2016) decomposition of the gap in the log OEWS median wage of the occupation held at the first US job between CtW and FUS workers with the same terminal degree. Rows are the contributions of each covariate, in log units, with standard errors in parentheses; the raw gap minus the total explained equals the final gap. School-tier and major-field contributions are included in the total but not shown. Figure 6 groups the contributions into the blocks defined in Section 4.1.1. Doctorate pools the PhD and non-licensure doctorates; Professional denotes licensure doctorates. Worker characteristics additionally include non-US experience and a same-employer transfer indicator. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A7: Decomposing the Occupational-Wage Gap by Year since Degree, CtS Master's Holders, Ten-Year Stayers

	Years since terminal degree										
	0	1	2	3	4	5	6	7	8	9	10
CtS	-0.144***	-0.001	0.029***	0.038***	0.050***	0.070***	0.081***	0.092***	0.101***	0.107***	0.111***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
FUS	-0.189***	-0.121***	-0.094***	-0.077***	-0.060***	-0.045***	-0.036***	-0.029***	-0.023***	-0.018***	-0.016***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Difference	0.046***	0.119***	0.122***	0.115***	0.109***	0.115***	0.117***	0.121***	0.124***	0.126***	0.128***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Explained	0.064***	0.086***	0.089***	0.085***	0.082***	0.084***	0.085***	0.087***	0.087***	0.088***	0.087***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Unexplained	-0.018***	0.033***	0.034***	0.030***	0.028***	0.031***	0.032***	0.034***	0.036***	0.038***	0.040***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Explained, by covariate											
First Job (this spell)	-0.002***	-0.002***	-0.000***	0.000***	0.000***	0.000**					
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)					
STEM	0.013***	0.009***	0.009***	0.009***	0.008***	0.009***	0.010***	0.011***	0.011***	0.012***	0.011***
	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Grad $\geq$ 2008	0.001***	0.002***	0.002***	0.001***	0.001***	0.001***	0.001***	0.001***	0.001***	0.000***	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
STEM $\times$ Grad $\geq$ 2008	0.004***	0.011***	0.013***	0.011***	0.011***	0.010***	0.009***	0.007***	0.006***	0.005***	0.004***
	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Major Generality	-0.002***	-0.001***	-0.001***	-0.001***	-0.001***	-0.000***	-0.000	0.000***	0.000***	0.001***	0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Unexplained, by covariate											
First Job (this spell)	-0.114***	0.004***	-0.000	-0.000***	-0.000**	-0.000***					
	(0.004)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)					
STEM	-0.010**	-0.011***	-0.013***	-0.008***	-0.006*	-0.004	0.000	0.001	0.003	0.004	0.005*
	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Grad $\geq$ 2008	-0.001	0.007*	0.004	0.001	0.001	0.001	0.000	-0.001	-0.003	-0.003	-0.002
	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
STEM $\times$ Grad $\geq$ 2008	0.020***	0.026***	0.027***	0.021***	0.016***	0.014***	0.012***	0.011***	0.009***	0.009***	0.008***
	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Major Generality	0.020***	0.029***	0.021***	0.016***	0.018***	0.018***	0.022***	0.027***	0.023***	0.022***	0.021***
	(0.005)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
N	1,899,315	2,050,563	2,139,411	2,163,412	2,182,373	2,245,379	2,282,945	2,318,256	2,352,860	2,390,462	2,490,133

Notes: Oaxaca–Blinder decomposition, FUS coefficients as reference, of the log OEWS median wage of the occupation held in each year since the terminal degree, net of graduation-cohort effects, among workers who remain in the US through year ten; graduation cohorts from 2003 on that reach year ten. Each column is estimated separately. Year zero is reported for completeness; the text and Figures 10 and 9 start at year one (Section 4.2). Covariates are STEM, graduation in 2008 or later and its interaction with STEM, major generality, an indicator for the first job spell, school tier, and major field; tier and field contributions are included in the totals but not shown. Standard errors in parentheses. Doctorate pools the PhD and non-licensure doctorates; Professional denotes licensure doctorates. $^{*}\ p < 0.05$, $^{**}\ p < 0.01$, $^{***}\ p < 0.001$

Table A8: Decomposing the Occupational-Wage Gap by Year since Degree, CtS Doctorate Holders, Ten-Year Stayers

	Years since terminal degree										
	0	1	2	3	4	5	6	7	8	9	10
CtS	-0.390***	-0.169***	-0.100***	-0.055***	-0.015***	0.018***	0.039***	0.057***	0.069***	0.076***	0.078***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
FUS	-0.261***	-0.143***	-0.098***	-0.065***	-0.037***	-0.015***	-0.004***	0.005***	0.009***	0.012***	0.010***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Difference	-0.129***	-0.026***	-0.001	0.010***	0.022***	0.033***	0.043***	0.052***	0.060***	0.065***	0.067***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Explained	-0.045***	0.013***	0.025***	0.033***	0.038***	0.042***	0.044***	0.046***	0.047***	0.048***	0.049***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Unexplained	-0.083***	-0.039***	-0.027***	-0.022***	-0.016***	-0.009***	-0.001	0.007***	0.013***	0.016***	0.019***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Explained, by covariate											
First Job (this spell)	-0.001***	-0.000	-0.000***	-0.000***	-0.000	-0.000					
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)					
STEM	-0.015***	-0.014***	-0.012***	-0.006***	-0.001	0.004***	0.008***	0.012***	0.014***	0.018***	0.019***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Grad $\geq$ 2008	0.000***	0.001***	0.001***	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
STEM $\times$ Grad $\geq$ 2008	-0.009***	0.003*	0.006***	0.006***	0.007***	0.007***	0.007***	0.006***	0.004***	0.002**	0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Major Generality	-0.011***	-0.008***	-0.007***	-0.006***	-0.006***	-0.005***	-0.004***	-0.004***	-0.004***	-0.004***	-0.003***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Unexplained, by covariate											
First Job (this spell)	-0.119***	0.001	0.000	0.000	0.000	-0.000					
	(0.015)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)					
STEM	-0.001	-0.004	-0.006	-0.013	-0.007	-0.010	-0.011	-0.008	-0.007	-0.008	-0.009
	(0.011)	(0.010)	(0.009)	(0.009)	(0.008)	(0.008)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Grad $\geq$ 2008	-0.035***	-0.001	0.002	-0.004	-0.004	-0.005	-0.008	-0.008	-0.010	-0.013*	-0.015**
	(0.010)	(0.009)	(0.008)	(0.008)	(0.007)	(0.007)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
STEM $\times$ Grad $\geq$ 2008	-0.006	-0.008	-0.002	0.002	-0.002	0.002	0.008	0.006	0.008	0.010*	0.012*
	(0.009)	(0.008)	(0.007)	(0.007)	(0.006)	(0.006)	(0.006)	(0.005)	(0.005)	(0.005)	(0.005)
Major Generality	-0.024	-0.008	0.002	0.012	0.001	0.009	0.009	0.011	0.017	0.017	0.016
	(0.013)	(0.012)	(0.011)	(0.011)	(0.011)	(0.010)	(0.010)	(0.009)	(0.009)	(0.009)	(0.009)
N	234,328	248,052	257,085	260,850	263,284	269,288	273,186	277,223	281,464	285,515	297,095

Notes: Oaxaca–Blinder decomposition, FUS coefficients as reference, of the log OEWS median wage of the occupation held in each year since the terminal degree, net of graduation-cohort effects, among workers who remain in the US through year ten; graduation cohorts from 2003 on that reach year ten. Each column is estimated separately. Year zero is reported for completeness; the text and Figures 10 and 9 start at year one (Section 4.2). Covariates are STEM, graduation in 2008 or later and its interaction with STEM, major generality, an indicator for the first job spell, school tier, and major field; tier and field contributions are included in the totals but not shown. Standard errors in parentheses. Doctorate pools the PhD and non-licensure doctorates; Professional denotes licensure doctorates. $^{*}\ p < 0.05$, $^{**}\ p < 0.01$, $^{***}\ p < 0.001$

Table A9: Decomposing the Occupational-Wage Gap by Year since Degree, CtS Professional Degree Holders, Ten-Year Stayers

	Years since terminal degree										
	0	1	2	3	4	5	6	7	8	9	10
CtS	0.138***	0.222***	0.243***	0.252***	0.254***	0.245***	0.246***	0.248***	0.248***	0.248***	0.246***
	(0.005)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
FUS	0.141***	0.236***	0.260***	0.267***	0.270***	0.266***	0.265***	0.261***	0.261***	0.261***	0.261***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Difference	-0.003	-0.014**	-0.017***	-0.015***	-0.016***	-0.021***	-0.019***	-0.013***	-0.013***	-0.013***	-0.015***
	(0.005)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Explained	0.002	0.004**	0.007***	0.009***	0.009***	0.008***	0.008***	0.008***	0.007***	0.007***	0.007***
	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Unexplained	-0.005	-0.018***	-0.024***	-0.024***	-0.025***	-0.029***	-0.026***	-0.021***	-0.020***	-0.019***	-0.022***
	(0.005)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Explained, by covariate											
First Job (this spell)	0.001	-0.000**	-0.000***	0.000***	0.000	-0.000					
	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)					
STEM	0.000*	0.000	0.000	-0.000	-0.000**	-0.000***	-0.000***	-0.000***	-0.001***	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Grad $\geq$ 2008	-0.001	-0.001***	-0.001***	-0.000***	-0.000	0.000**	0.000**	0.000***	0.000***	0.000***	0.000**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
STEM $\times$ Grad $\geq$ 2008	-0.001	-0.000	-0.000	-0.000	-0.001**	-0.000**	-0.000**	-0.000*	-0.000***	-0.000	0.000
	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Major Generality	0.003***	0.002***	0.003***	0.003***	0.003***	0.002***	0.002***	0.002***	0.002***	0.002***	0.002***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Unexplained, by covariate											
First Job (this spell)	-0.082***	-0.000	-0.002*	-0.001	-0.001	-0.001					
	(0.017)	(0.002)	(0.001)	(0.001)	(0.001)	(0.000)					
STEM	-0.001	-0.001	0.004	0.006	0.005	0.003	0.000	-0.001	-0.002	-0.001	0.000
	(0.006)	(0.005)	(0.005)	(0.004)	(0.004)	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Grad $\geq$ 2008	0.026*	0.022*	0.013	0.022**	0.019*	0.013	0.011	0.013*	0.016*	0.014*	0.014*
	(0.012)	(0.010)	(0.009)	(0.008)	(0.008)	(0.007)	(0.007)	(0.007)	(0.006)	(0.006)	(0.006)
STEM $\times$ Grad $\geq$ 2008	-0.010	-0.007	-0.006	-0.005	-0.003	0.000	0.002	0.004	0.004	0.003	0.002
	(0.006)	(0.004)	(0.004)	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Major Generality	-0.039	-0.033	-0.022	-0.037	-0.017	-0.031	-0.013	0.009	-0.008	-0.015	-0.022
	(0.025)	(0.022)	(0.021)	(0.020)	(0.020)	(0.019)	(0.018)	(0.018)	(0.017)	(0.017)	(0.017)
N	234,526	273,801	294,486	307,050	314,564	321,862	327,617	332,837	337,497	342,626	355,495

Notes: Oaxaca–Blinder decomposition, FUS coefficients as reference, of the log OEWS median wage of the occupation held in each year since the terminal degree, net of graduation-cohort effects, among workers who remain in the US through year ten; graduation cohorts from 2003 on that reach year ten. Each column is estimated separately. Year zero is reported for completeness; the text and Figures 10 and 9 start at year one (Section 4.2). Covariates are STEM, graduation in 2008 or later and its interaction with STEM, major generality, an indicator for the first job spell, school tier, and major field; tier and field contributions are included in the totals but not shown. Standard errors in parentheses. Doctorate pools the PhD and non-licensure doctorates; Professional denotes licensure doctorates. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A10: Decomposing the Occupational-Wage Gap by Year since Degree, CtW Master's Holders, Ten-Year Stayers

	Years since first US labor-market year										
	0	1	2	3	4	5	6	7	8	9	10
CtW	0.035***	0.064***	0.087***	0.099***	0.112***	0.127***	0.137***	0.147***	0.155***	0.162***	0.168***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
FUS	-0.096***	-0.028***	-0.001***	0.015***	0.032***	0.046***	0.055***	0.062***	0.068***	0.073***	0.075***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Difference	0.132***	0.093***	0.088***	0.085***	0.081***	0.081***	0.082***	0.084***	0.086***	0.090***	0.093***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Explained	0.042***	0.066***	0.064***	0.063***	0.061***	0.063***	0.064***	0.066***	0.068***	0.070***	0.071***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Unexplained	0.090***	0.026***	0.024***	0.022***	0.020***	0.018***	0.019***	0.018***	0.018***	0.020***	0.022***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Explained, by covariate											
First Job (this spell)	-0.020***	0.003***	0.001***	0.001***	0.000***	0.000***					
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)					
STEM	0.007***	0.005***	0.005***	0.005***	0.005***	0.006***	0.006***	0.007***	0.007***	0.007***	0.007***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Grad $\geq$ 2008	-0.005***	-0.007***	-0.007***	-0.007***	-0.007***	-0.006***	-0.006***	-0.006***	-0.005***	-0.003***	-0.002***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
STEM $\times$ Grad $\geq$ 2008	0.001***	0.004***	0.004***	0.004***	0.004***	0.003***	0.003***	0.002***	0.002***	0.002***	0.002***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Major Generality	-0.001***	-0.000***	-0.001***	-0.000***	-0.000***	-0.000***	-0.000	0.000***	0.000***	0.000***	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Non-US Labor Mkt. Exp.	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Non-US Labor Mkt. Exp.^{2}	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Same Employer (Transfer)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Unexplained, by covariate											
First Job (this spell)	-0.077***	0.000	0.000	0.000	0.000	0.000					
	(0.002)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)					
STEM	-0.014***	-0.008***	-0.008***	-0.004*	0.000	0.001	0.003	0.003	0.004*	0.005**	0.005**
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Grad $\geq$ 2008	0.017***	0.004*	-0.000	-0.004*	-0.009***	-0.012***	-0.018***	-0.022***	-0.025***	-0.024***	-0.020***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
STEM $\times$ Grad $\geq$ 2008	0.025***	0.018***	0.016***	0.012***	0.007***	0.007***	0.006***	0.007***	0.007***	0.007***	0.008***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Major Generality	-0.018***	-0.002	-0.007	-0.009	-0.006	-0.006	0.004	0.009*	0.010*	0.014**	0.016***
	(0.005)	(0.004)	(0.004)	(0.005)	(0.005)	(0.005)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Non-US Labor Mkt. Exp.	-0.022***	-0.015***	-0.012***	-0.010***	-0.010***	-0.009***	-0.009***	-0.007**	-0.007***	-0.007***	-0.006**
	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Non-US Labor Mkt. Exp.^{2}	0.025***	0.024***	0.021***	0.018***	0.017***	0.016***	0.014***	0.011***	0.010***	0.009***	0.008***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Same Employer (Transfer)	0.002***	0.002***	0.002***	0.002***	0.002***	0.002***	0.002***	0.002***	0.002***	0.002***	0.002***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
N	1,973,663	2,094,238	2,166,358	2,183,189	2,197,628	2,252,522	2,286,921	2,321,492	2,357,050	2,397,236	2,496,189

Notes: Oaxaca–Blinder decomposition, FUS coefficients as reference, of the log OEWS median wage of the occupation held in each year since the worker's first year in the US labor market (for FUS workers, since the terminal degree), net of labor-market-entry-cohort effects, among workers who remain in the US through year ten; graduation cohorts from 2003 on that reach year ten. CtW workers enter if their first US labor-market year is within five years of the terminal degree. Each column is estimated separately. Year zero is reported for completeness; the text and Figures 10 and 9 start at year one (Section 4.2). Covariates are STEM, graduation in 2008 or later and its interaction with STEM, major generality, an indicator for the first job spell, school tier, and major field, and for CtW workers non-US labor-market experience, its square, and a same-employer transfer indicator; tier and field contributions are included in the totals but not shown. Standard errors in parentheses. Doctorate pools the PhD and non-licensure doctorates; Professional denotes licensure doctorates. $^{*}\ p < 0.05$, $^{**}\ p < 0.01$, $^{***}\ p < 0.001$

Table A11: Decomposing the Occupational-Wage Gap by Year since Degree, CtW Doctorate Holders, Ten-Year Stayers

	Years since first US labor-market year										
	0	1	2	3	4	5	6	7	8	9	10
CtW	-0.162***	-0.120***	-0.077***	-0.036***	0.012***	0.060***	0.093***	0.112***	0.129***	0.140***	0.147***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
FUS	-0.168***	-0.051***	-0.006***	0.026***	0.055***	0.076***	0.088***	0.096***	0.101***	0.103***	0.102***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Difference	0.006*	-0.069***	-0.071***	-0.062***	-0.042***	-0.016***	0.005*	0.016***	0.029***	0.038***	0.045***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Explained	0.020***	-0.030***	-0.026***	-0.019***	-0.013***	-0.007***	-0.002	0.001	0.005***	0.009***	0.011***
	(0.003)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Unexplained	-0.014***	-0.039***	-0.044***	-0.043***	-0.030***	-0.009***	0.007**	0.015***	0.023***	0.029***	0.034***
	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Explained, by covariate											
First Job (this spell)	0.057***	0.001	-0.001***	-0.001***	-0.000	-0.000					
	(0.002)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)					
STEM	-0.011***	-0.010***	-0.009***	-0.005***	-0.001	0.003***	0.006***	0.009***	0.011***	0.014***	0.014***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Grad $\geq$ 2008	-0.007***	-0.009***	-0.009***	-0.007***	-0.007***	-0.006***	-0.006***	-0.006***	-0.005***	-0.004***	-0.003***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
STEM $\times$ Grad $\geq$ 2008	-0.003***	0.001*	0.002***	0.002***	0.003***	0.003***	0.003***	0.002***	0.002***	0.001*	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Major Generality	-0.009***	-0.006***	-0.006***	-0.005***	-0.005***	-0.004***	-0.003***	-0.003***	-0.003***	-0.003***	-0.003***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Non-US Labor Mkt. Exp.	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Non-US Labor Mkt. Exp.^{2}	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Same Employer (Transfer)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Unexplained, by covariate											
First Job (this spell)	0.116***	0.000	-0.000	-0.000	-0.000	-0.000					
	(0.006)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)					
STEM	-0.008	0.001	-0.001	-0.013	-0.012	-0.013	-0.008	-0.004	-0.014*	-0.018**	-0.016*
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.007)	(0.007)	(0.007)	(0.007)	(0.006)	(0.006)
Grad $\geq$ 2008	0.036***	0.012	0.011	0.008	0.002	-0.004	-0.009	-0.014*	-0.026***	-0.027***	-0.027***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.006)	(0.006)	(0.006)	(0.006)	(0.005)	(0.005)
STEM $\times$ Grad $\geq$ 2008	0.020***	-0.003	-0.011	-0.008	-0.002	0.002	0.002	0.003	0.011*	0.012**	0.014**
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.004)
Major Generality	-0.097***	-0.051**	-0.023	-0.011	-0.003	0.007	0.016	0.001	-0.005	0.009	0.012
	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)	(0.015)	(0.014)	(0.014)	(0.013)	(0.013)	(0.013)
Non-US Labor Mkt. Exp.	0.032***	-0.019**	-0.007	0.002	0.017**	0.025***	0.026***	0.028***	0.023***	0.023***	0.020***
	(0.007)	(0.006)	(0.006)	(0.006)	(0.006)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Non-US Labor Mkt. Exp.^{2}	0.024***	0.051***	0.042***	0.032***	0.020***	0.009*	0.001	-0.003	-0.003	-0.005	-0.004
	(0.005)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)
Same Employer (Transfer)	0.002***	0.002***	0.001***	0.001***	0.001***	0.001***	0.001***	0.001***	0.001***	0.001***	0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
N	223,441	232,098	238,172	240,181	241,429	246,146	249,560	252,994	256,763	260,479	270,473

Notes: Oaxaca–Blinder decomposition, FUS coefficients as reference, of the log OEWS median wage of the occupation held in each year since the worker's first year in the US labor market (for FUS workers, since the terminal degree), net of labor-market-entry-cohort effects, among workers who remain in the US through year ten; graduation cohorts from 2003 on that reach year ten. CtW workers enter if their first US labor-market year is within five years of the terminal degree. Each column is estimated separately. Year zero is reported for completeness; the text and Figures 10 and 9 start at year one (Section 4.2). Covariates are STEM, graduation in 2008 or later and its interaction with STEM, major generality, an indicator for the first job spell, school tier, and major field, and for CtW workers non-US labor-market experience, its square, and a same-employer transfer indicator; tier and field contributions are included in the totals but not shown. Standard errors in parentheses. Doctorate pools the PhD and non-licensure doctorates; Professional denotes licensure doctorates. $^{*}\ p < 0.05$, $^{**}\ p < 0.01$, $^{***}\ p < 0.001$

Table A12: Decomposing the Occupational-Wage Gap by Year since Degree, CtW Professional Degree Holders, Ten-Year Stayers

	Years since first US labor-market year										
	0	1	2	3	4	5	6	7	8	9	10
CtW	0.306***	0.345***	0.402***	0.427***	0.430***	0.415***	0.419***	0.422***	0.417***	0.417***	0.416***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
FUS	0.234***	0.328***	0.351***	0.358***	0.361***	0.356***	0.355***	0.351***	0.351***	0.351***	0.351***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Difference	0.072***	0.018*	0.051***	0.069***	0.069***	0.059***	0.064***	0.071***	0.067***	0.066***	0.065***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Explained	0.089***	0.115***	0.107***	0.098***	0.091***	0.082***	0.077***	0.071***	0.065***	0.062***	0.058***
	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Unexplained	-0.018*	-0.097***	-0.055***	-0.029***	-0.022**	-0.024***	-0.013*	0.000	0.002	0.004	0.007
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Explained, by covariate											
First Job (this spell)	-0.053***	0.004***	0.002***	-0.001***	-0.000	0.000***					
	(0.002)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)					
STEM	0.002***	0.000**	0.000	-0.000	-0.000**	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Grad $\geq$ 2008	0.008***	0.002***	0.002***	-0.000*	-0.001***	-0.002***	-0.003***	-0.003***	-0.003***	-0.002***	-0.002***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
STEM $\times$ Grad $\geq$ 2008	-0.009***	-0.005***	-0.004***	-0.003***	-0.002***	-0.001***	-0.001***	-0.001***	-0.001***	-0.000	0.000
	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Major Generality	0.016***	0.011***	0.012***	0.011***	0.009***	0.006***	0.006***	0.006***	0.006***	0.006***	0.006***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Non-US Labor Mkt. Exp.	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Non-US Labor Mkt. Exp.^{2}	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Same Employer (Transfer)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Unexplained, by covariate											
First Job (this spell)	-0.166***	0.000	0.000	-0.000	-0.000	0.000					
	(0.015)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)					
STEM	0.031***	0.029***	0.027***	0.022***	0.019**	0.021***	0.016**	0.016**	0.019***	0.021***	0.013**
	(0.007)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Grad $\geq$ 2008	0.134***	0.092***	0.096***	0.069***	0.050***	0.014	-0.010	-0.020*	-0.014	-0.008	-0.009
	(0.013)	(0.013)	(0.012)	(0.012)	(0.012)	(0.011)	(0.010)	(0.010)	(0.010)	(0.010)	(0.009)
STEM $\times$ Grad $\geq$ 2008	-0.017**	-0.013*	-0.011*	-0.006	-0.005	-0.006	-0.004	-0.006	-0.005	-0.005	0.000
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Major Generality	-0.073*	-0.021	0.061	0.100**	0.113***	0.088**	0.109***	0.108***	0.114***	0.126***	0.116***
	(0.036)	(0.035)	(0.035)	(0.034)	(0.033)	(0.032)	(0.031)	(0.030)	(0.029)	(0.029)	(0.028)
Non-US Labor Mkt. Exp.	0.016	-0.035	-0.007	-0.010	-0.005	-0.018	-0.022	-0.012	-0.010	-0.003	-0.005
	(0.022)	(0.021)	(0.020)	(0.019)	(0.019)	(0.019)	(0.018)	(0.018)	(0.017)	(0.017)	(0.016)
Non-US Labor Mkt. Exp.^{2}	0.015	0.028	0.001	0.002	-0.000	0.012	0.014	0.008	0.003	-0.007	-0.001
	(0.015)	(0.015)	(0.015)	(0.015)	(0.014)	(0.014)	(0.014)	(0.013)	(0.013)	(0.012)	(0.012)
Same Employer (Transfer)	0.000	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
N	231,835	268,829	288,204	299,641	306,284	312,942	318,414	323,386	327,914	332,820	345,047

Notes: Oaxaca–Blinder decomposition, FUS coefficients as reference, of the log OEWS median wage of the occupation held in each year since the worker's first year in the US labor market (for FUS workers, since the terminal degree), net of labor-market-entry-cohort effects, among workers who remain in the US through year ten; graduation cohorts from 2003 on that reach year ten. CtW workers enter if their first US labor-market year is within five years of the terminal degree. Each column is estimated separately. Year zero is reported for completeness; the text and Figures 10 and 9 start at year one (Section 4.2). Covariates are STEM, graduation in 2008 or later and its interaction with STEM, major generality, an indicator for the first job spell, school tier, and major field, and for CtW workers non-US labor-market experience, its square, and a same-employer transfer indicator; tier and field contributions are included in the totals but not shown. Standard errors in parentheses. Doctorate pools the PhD and non-licensure doctorates; Professional denotes licensure doctorates. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

B Existence of a Natives' Separating Equilibrium (Mailath, 1987)

The general model with minimal assumptions on functional form follows the model outlined in (Mailath, 1987). The original paper made no assumptions on whether a high θ denotes higher productivity. For the purpose of this paper, I make this assumption and specify the conditions from the paper as such.

Let $U\left(\theta, \hat{\theta}, e\right)$ be the utility function of a type θ worker with education level e when the firms believe that they are type $\hat{\theta}$. Conditions 1 through 5 below are regularity conditions:

Condition 1 (Smoothness)**.** $U\left(\theta, \hat{\theta}, e\right)$ *is twice differentiable on* $\Theta^2 \times \mathcal{E}$.

Condition 2 (Belief Monotonicity)**.** $\dfrac{\partial}{\partial \hat{\theta}} U\left(\theta, \hat{\theta}, e\right) > 0.$

In other words, fixing the education level and actual productivity type, being recognized as having higher productivity is strictly better. The condition incentivizes high productivity workers to signal their productivity type.

Condition 3 (Type monotonicity)**.** $\dfrac{\partial^2}{\partial \theta \partial e} U\left(\theta, \hat{\theta}, e\right) > 0$

Given that $U\left(\theta, \hat{\theta}, e\right) = w\left(\hat{\theta}\right) - C\left(\theta, e\right)$, this condition ensures that the marginal cost of education decreases in productivity.

Condition 4 ("Strict" Quasi-Concavity[11])**.** *The equation* $\frac{\partial}{\partial e} U\left(\theta, \hat{\theta}, e\right) = 0$ *has a unique solution in* e*, denoted* $\phi(\theta)$ *such that* $\phi(\theta) \in \underset{e}{argmax}\ U\left(\theta, \hat{\theta}, e\right)$*, and* $\dfrac{\partial^2}{\partial e^2} U\left(\theta, \hat{\theta}, e\right)\Big|_{e=\phi(\theta)} < 0$

This condition ensures that there is a unique optimal education level for each productivity type, fixing $\hat{\theta}$.

Condition 5 (Boundedness)**.** $\exists k > 0$ *such that*

$$\forall (\theta, e) \in \Theta \times \mathcal{E}, \frac{\partial^2}{\partial e^2} U(\theta, \hat{\theta}, e) \geq 0 \ \Rightarrow\ \left|\frac{\partial}{\partial e} U(\theta, \hat{\theta}, e)\right| > k.$$

Conditions 4 and 5 together implies that, for arbitrary $\sigma(\theta)$, the set of possible signals that are reasonable, given the firms' belief update, is bounded (Mailath, 1987). In other words, condition 4 ensures the existence of an optimal signaling education level while condition 5 ensures that signals must not be free.

These next two conditions help characterize the separating equilibrium in the game:

Condition 6 (Initial Value [IV])**.** $\sigma\left(\tilde{\theta}\right) = \phi\left(\tilde{\theta}\right)$*, where*

$$\tilde{\theta} = \begin{cases} \sup \Theta & \text{if } \frac{\partial}{\partial \hat{\theta}} U(\theta, \hat{\theta}, e) < 0 \\ \inf \Theta & \text{if } \frac{\partial}{\partial \hat{\theta}} U(\theta, \hat{\theta}, e) > 0 \end{cases} \tag{B.1}$$

[11] Mailath (1987) notes that this condition is stronger than the typical strict quasi-concavity but weaker than concavity.

In other words, IV is satisfied if the marginal utility with respect to the worker's true type being negative (positive) implies that the lowest (highest) productivity type chooses the the optimal education level in equilibrium. This condition helps with the solvability of Equation (B.3) below.

Condition 7 (Single Crossing [SC])**.**

$$\frac{d}{d\theta}\frac{\frac{\partial}{\partial e}U(\theta,\hat{\theta},e)}{\frac{\partial}{\partial\hat{\theta}}U(\theta,\hat{\theta},e)} \gtrless 0 \tag{B.2}$$

SC ensures that the indifference curves of any two informed workers of different types crosses at most once.

Mailath (1987) theorem 1: If conditions 1 through 5 are satisfied and $\sigma(\theta)$ is one-to-one and satisfies incentive compatibility, then $\sigma(\theta)$ has at most one discontinuity on Θ and where it is continuous, it is differentiable and satisfies:

$$\frac{d\sigma}{d\theta} = -\frac{\frac{\partial}{\partial\hat{\theta}}U(\theta,\hat{\theta},e)}{\frac{\partial}{\partial e}U(\theta,\hat{\theta},e)}. \tag{B.3}$$

Mailath (1987) theorem 2: If conditions 1 through 5 are satisfied, and $\sigma(\theta)$ is one-to-one and satisfies incentive compatibility, then

SC $\Rightarrow$ σ is strictly monotonic on Θ and σ satisfies Eq. (3)

IV $\Rightarrow$ σ is strictly monotonic on Θ, satisfies Eq. (3), and $sign\left(\frac{d\sigma}{d\theta}\right) = sign\left[\frac{\partial^2}{\partial\theta\partial e}U\left(\theta,\hat{\theta},e\right)\right]$

Mailath (1987) corollary: Suppose conditions 1 through 5 are satisfied and IV holds, then the equilibrium σ is the optimal σ, which uniquely solves Equation (B.3).

Definition 2 (Strict Incentive Compatibility [SIC])**.** *The strategy $e(\theta)$ satisfies strict incentive compatibility if*

$$\{\sigma(\theta)\} = \underset{e}{argmax}\ U(\theta,\hat{\Theta}(e),e),\ \forall\theta\in\Theta \tag{B.4}$$

Mailath (1987) theorem 3: Suppose that conditions 1 through 5 are satisfied. A strictly increasing function σ, continuous at $\underline{\theta}$, satisfies strict incentive compatibility if and only if

(i) σ solves Equation (B.3).

(ii) When $\frac{\partial}{\partial\hat{\theta}}U(\theta,\hat{\theta},e) > (<)0$, $\frac{\frac{\partial}{\partial e}U(\theta,\hat{\theta},e)}{\frac{\partial}{\partial\hat{\theta}}U(\theta,\hat{\theta},e)}$ is a strictly increasing (decreasing) function of θ for $\left(\hat{\theta},e\right)$ in the graph of σ.

B.1 Graphical derivation of the continuous type model

Step 1: Suppose $\Theta = \{\bar{\theta}, \underline{\theta}\}$.

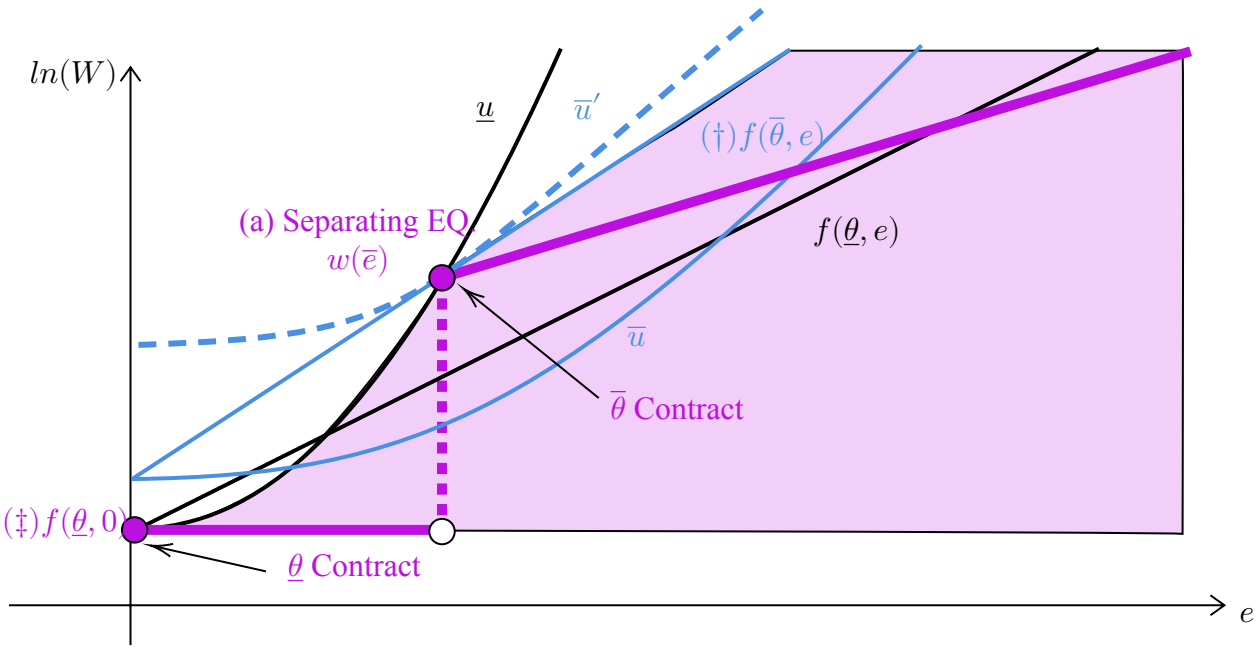
In Figure B1 below, the curves $\underline{u}$ and $\bar{u}$ indicates the indifference curve of a worker of type $\underline{\theta}$ and $\bar{\theta}$ with no education, respectively. The linear production functions of workers of respective types are indicated on the graph. Where $\underline{u}$ intersects $f(\bar{\theta}, e)$ creates a separating contract where the type $\underline{\theta}$ worker is indifferent between getting the education level $\bar{e}$ required for that contract and accepting a contract that requires no education.

At that same point, the $\bar{\theta}$ worker will obtain the education level $\bar{e}$ for the contract, as indicated by the $\bar{\theta}$ worker's higher level indifference curve $\bar{u}'$. A full equilibrium wage schedule is thus represented by the purple discontinuous function on the graph:

$$w(e) = \begin{cases} f(\underline{\theta}, 0) & e < \bar{e} \\ f(\bar{\theta}, \bar{e}) & e = \bar{e} \\ f(\theta, e) & e > \bar{e},\ \theta \in \Theta \end{cases}$$

Note that the equilibrium wage schedule can be any wage schedule that is in the shaded area as long as it passes through $f(\underline{\theta}, 0)$ and $f(\bar{\theta}, \bar{e})$

Figure B1: Separating Equilibrium with $\Theta = \{\bar{\theta}, \underline{\theta}\}$



Notes: (†) marks the type-$\bar{\theta}$ worker's productivity schedule $f(\bar{\theta}, e)$. (‡) marks the type-$\underline{\theta}$ worker's productivity schedule $f(\underline{\theta}, e)$. (a) marks the separating equilibrium contract for type-$\bar{\theta}$ workers, $w(\bar{e})$. $\bar{u}'$ represents type-$\bar{\theta}$ workers' indifference curve at the separating contract. $\bar{u}$ represents type-$\bar{\theta}$ workers' indifference curve at the $\underline{\theta}$ contract.

Step 2: Add an intermediate productivity: $\Theta = \{\bar{\theta}, \theta_M, \underline{\theta}\}$ where $\bar{\theta} > \theta_M > \underline{\theta}$.

The result here is achieved in the same way as step 1. In Figure B2 below, the curves $\underline{u}$, $\underline{u}_M$, and $\bar{u}$ indicates the indifference curve of a worker of type $\underline{\theta}$, θ_M, and $\bar{\theta}$ with no education, respectively. The linear production functions of workers of respective types are indicated on the graph. Where $\underline{u}$ intersects $f(\theta_M, e)$ creates a separating contract $w(e_M)$where the type $\underline{\theta}$ worker is indifferent between getting the education level e_M required for that contract and accepting a contract that requires no education.

At the contract $w(e_M)$, the θ_M worker will obtain the education level e_M for the contract, as indicated by the θ_M worker's higher level indifference curve $\bar{u}_M$.

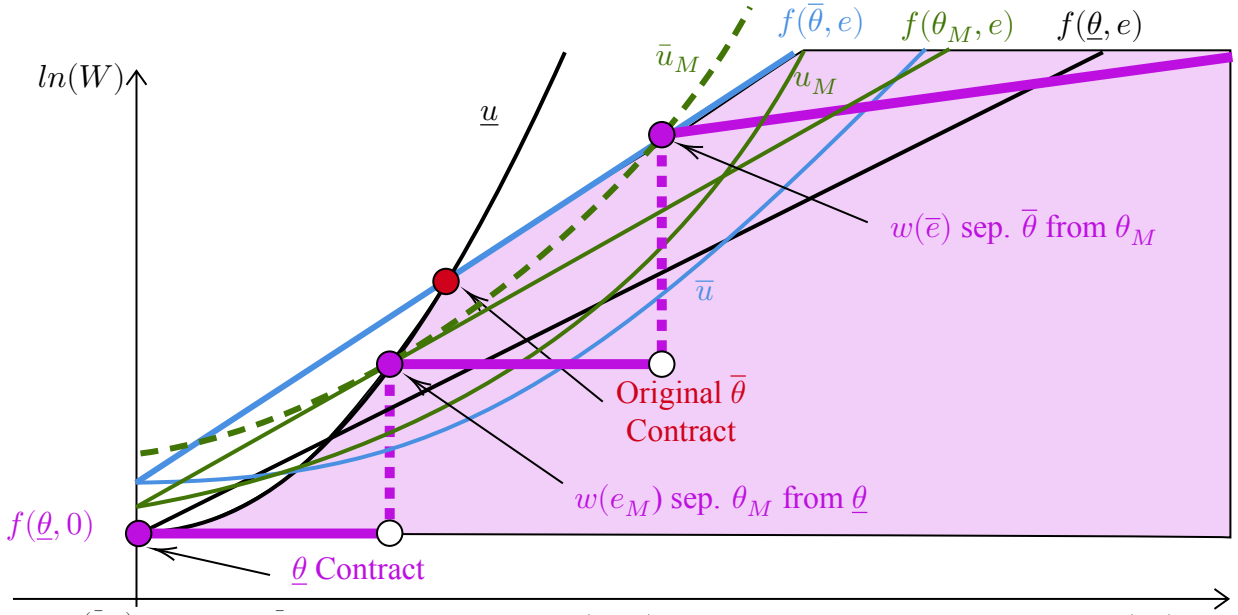
Similarly, the indifference curve $\bar{u}_M$ intersects $f\left(\bar{\theta}, e\right)$ at the contract $w(\bar{e})$. Since both $w(\bar{e})$ and $w(e_M)$ are on $\bar{u}_M$, the θ_M worker is indifferent between the two education levels while the $\bar{\theta}$ worker will choose $\bar{e}$ as that contract achieves a higher utility for them (indifference curve not shown).

A full equilibrium wage schedule is thus represented by the purple discontinuous function on the graph:

$$w(e) = \begin{cases} f\left(\underline{\theta}, 0\right) & e < e_M \\ f\left(\theta_M, e_M\right) & e_M \leq e \leq \bar{e} \\ f\left(\bar{\theta}, \bar{e}\right) & e = \bar{e} \\ f\left(\theta, e\right) & e > \bar{e},\ \theta \in \Theta \end{cases}$$

Note that the equilibrium wage schedule can be any wage schedule that is in the shaded area as long as it passes through $f\left(\underline{\theta}, 0\right)$, $f\left(\theta_M, e_M\right)$ and $f\left(\bar{\theta}, \bar{e}\right)$.

Figure B2: Separating Equilibrium with $\Theta = \{\bar{\theta}, \theta_M, \underline{\theta}\}$ where $\bar{\theta} > \theta_M > \underline{\theta}$



Notes: $f(\bar{\theta}, e)$ marks the type-$\bar{\theta}$ worker's productivity schedule. $f(\theta_M, e)$ marks the type-θ_M worker's productivity schedule. $f(\underline{\theta}, e)$ marks the type-$\underline{\theta}$ worker's productivity schedule. $w(\bar{e})$ marks the separating equilibrium contract for type-$\bar{\theta}$ workers, $w(\bar{e})$; $w(e_M)$ marks the separating equilibrium contract for type-θ_M workers, $w(e_M)$. $\bar{u}_M$ represents type-θ_M workers' indifference curve at the separating contracts where they are indifferent between obtaining e_M and $\bar{e}$. $\bar{u}$ represents type-$\bar{\theta}$ workers' indifference curve at the $\underline{\theta}$ contract.

Step 3: A Continuum of Types: $\Theta = \left[\underline{\theta}, \bar{\theta}\right]$.

Notice that because each point on the separating wage schedule is determined by the ordered productivity functions, the straight line between $w(e_M)$ and $w(\bar{e})$ necessarily lies below the $w(\bar{e})$ in step 1, as indicated on the Figure B2. This implies that, as more and more types are added to Θ, the equilibrium wage schedule will conform to a convex function as shown in Figure B3 below.

The equilibrium wage schedule is thus:

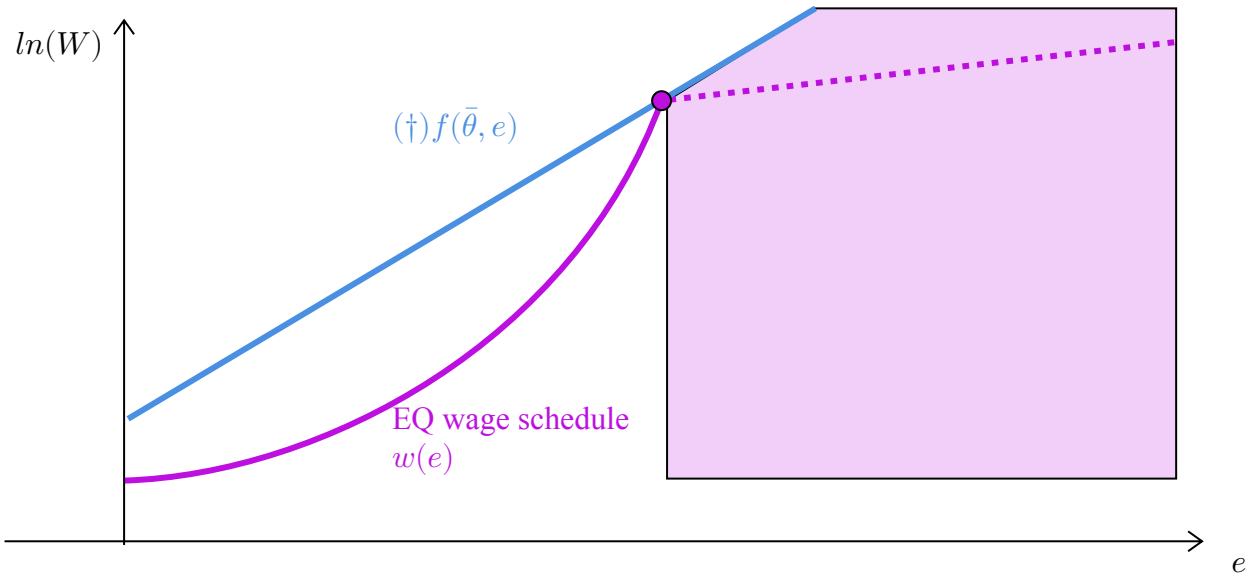
$$w(e) = \begin{cases} f(\theta, \sigma(\theta)) & , e \in \left[\sigma\left(\underline{\theta}\right), \sigma\left(\bar{\theta}\right)\right] \\ f(\theta, e) & , otherwise. \end{cases}$$

Contracts where $e > \sigma\left(\bar{\theta}\right)$—represented by the dotted wage schedule and shaded region—exhibit over-signaling, as the marginal wage return to education strictly exceeds its marginal cost. The shaded region spans the complete set of equilibrium wages these contracts can sustain, while the dotted schedule serves as an illustrative benchmark.

Assumption 8 (Occupation-Level Posted Wages)**.** *The contract $w(e)$ attaches to the occupation rather than to the worker: a firm filling a position whose equilibrium contract is $w(e)$ pays $w(e)$ to whomever it hires, and cannot adjust that wage downward for an immigrant.*

Assumption 2 is what makes the immigrant extension bite, and it is worth stating why. Firms directly observe immigrant status, including foreign education and sponsorship needs. Absent a

Figure B3: Separating Equilibrium with $\Theta = [\underline{\theta}, \bar{\theta}]$



Notes: $(\dagger)$ marks the type-$\bar{\theta}$ worker's productivity schedule $f(\bar{\theta}, e)$. The intersection of $(\dagger)$ and the equilibrium wage schedule is the separating equilibrium contract for type-$\bar{\theta}$ workers.

restriction on wage setting, Bertrand competition among firms that observe it would simply offer an immigrant the contract $f(\theta, e) - d$ at every e: a smooth downward shift of the schedule that leaves no education level unavailable and generates no discrete mismatch. Under Assumption 2 the firm cannot make that offer. Employing a type-$\bar{\theta}$ immigrant in an occupation whose contract is $W(e)$ yields $f(\bar{\theta}, e) - d - w(e)$, and where this is negative the firm excludes the immigrant from the occupation rather than underpaying her. The H-1B program makes the assumption concrete rather than merely convenient: a sponsoring employer must attest to paying at least the Department of Labor prevailing wage for the occupation, so nativity-specific downward wage adjustment within an occupation is legally foreclosed and exclusion into a lower-education occupation becomes the binding margin. Posted occupational wage scales and equal-pay norms deliver the same restriction more broadly.

C Theoretical Foundation of the PageRank Score

Firm PageRank utility follows Sorkin (2018), who utilizes systematic mobility between firms to create an order over employers under the revealed-preference framework.

Denote the set of workers $\mathcal{I}$ and the set of firms $\mathcal{J} = \{1, \dots, J\}$. Consider a worker $i \in \mathcal{I}$, currently employed at firm k, who receives an offer from firm j; the worker accepts if $V_j + \varepsilon_{ijk} > V_k + \varepsilon_{ikk}$, where V_j and V_k are common utility attached to firms j and k—the same for every worker who might compare them—and ε_{ijk} is worker i's idiosyncratic taste for firm j given current employer k, unrelated to V_j on average across workers. Let M_{jk} be the number of workers observed transitioning from firm k to firm j over the sample period, and let $S_k = \sum_{j \in \mathcal{J}} M_{jk}$ be firm k's total observed outflow. The observed share of k's leavers who land at j is then $A_{jk} = M_{jk}/S_k$, and under the choice rule above this share is informative about V_j relative to all other firms that k's workers are observed moving to: firms that draw a disproportionate share of leavers, from firms that are themselves valuable, must be highly valued. This is a self-referential definition—a firm's value is revealed by the value of the firms whose workers it attracts—and it is exactly encoded in the fixed point $\tilde{V}_j = \sum_k A_{jk}\tilde{V}_k$, where $\tilde{V} = (\tilde{V}_1, \dots, \tilde{V}_J)$ is the leading eigenvector of the $J \times J$ matrix A. The size-adjusted index used in the analysis, $V_j^e \approx \tilde{V}_j + \log(g_j)$, augments the pure choice-revealed value $\tilde{V}_j$ with the log of firm j's size g_j, on the view that part of a firm's realized flow reflects scale of opportunity rather than worker preference.

This logic requires three assumptions, each with a direct economic interpretation:

- **A common value index**: Any worker moving from k to j is assumed to agree, up to the idiosyncratic term ε_{ijk}, that j is worth V_j—i.e., firms are vertically ranked (some firms are simply better places to work), rather than workers sorting on match-specific comparative advantage. If firm value were instead worker-specific, the share A_{jk} would reveal relative match quality for the mover, not a firm attribute, and no single ranking $\tilde{V}$ would be recoverable.

- **Selection on value, not on outside constraints**: A transition captured in M_{jk} is assumed to reflect the worker choosing the best available offer, not being forced to move (e.g., by a layoff) or by a search friction unrelated to V_j. To the extent involuntary separations are common and correlated with firm quality, A_{jk} conflates "workers prefer j" with "workers were pushed out of k," attenuating the revealed-preference interpretation of $\tilde{V}$.

- **A connected and stable labor market**: Every firm's value V_j must be comparable to every other's through some chain of observed moves captured in M (workers routed through the non-employment node count as one such link), and the ranking $\tilde{V}$ is assumed stable over the window used to estimate it. To account for moving market conditions, I estimate the ranking using 5-year rolling windows and interpolate the rankings for the period 2020-2022.

C.1 Within-education occupation ranking mechanics.

Within a given degree-major pair containing n distinct occupations, the estimated ranking adds a small baseline term representing an occupation's share of workers independent of where they came

from,

$$r = d\,A\,r + \frac{1-d}{n}\mathbf{1}, \tag{C.1}$$

where $A = [A_{ba}]$ is the stratum's occupation-to-occupation share matrix, $\mathbf{1}$ is the n-vector of ones, and the damping factor d ($= 0.85$ per Brin and Page (1998)) is the weight placed on the observed-transition term relative to the uniform baseline term. This ensures every occupation retains some minimal comparability even absent a direct or well-populated chain of observed moves connecting it to the rest of the stratum's network.

Feasible-set truncation and the estimation sample Revealed-preference rankings presume that observed moves reveal preference over a commonly valued bundle. For immigrant workers that presumption is not innocuous: their feasible set is truncated, so a move into a lower-quantile occupation may reflect the constraint rather than the ranking, compressing the ranking across the interval of interest and attenuating d toward zero. I therefore re-estimate the occupation ranking on native workers alone, whom the model places on the undistorted schedule and whose mobility reveals unconstrained preference. The two rankings agree closely—person-year-weighted Spearman of 0.979 across the 37 postgraduate strata, with no loss of network connectivity — so the full-sample ranking is retained. Firm rankings use the full sample: exclusion operates at the occupation contract rather than the firm, and immigrant moves supply the panel's only coverage of foreign employers.

D LinkedIn Coverage with IPEDS Benchmark

This appendix evaluates the representativeness of the People Data Labs (PDL) LinkedIn education records by benchmarking it against the Integrated Postsecondary Education Data System (IPEDS) Completions census, maintained by the National Center for Education Statistics. Because IPEDS tracks the universe of degrees conferred by Title IV-eligible institutions within the United States, this exercise establishes the extent to which the LinkedIn panel reflect the true population of U.S. graduates. International institutions fall outside the scope of this benchmark as they lack a corresponding IPEDS institutional identifier (UNITID).

The rest of this appendix proceeds as follows: Appendix D.1 outlines the alignment and filtering protocols used to construct the comparative sample. Appendix D.2 assesses absolute coverage across institutional tiers. Appendix D.3 examines the fidelity of the within-school field mix. Appendix D.4 documents systematic field-level representation biases; and finally, I present data-quality diagnostics and robustness checks.

D.1 Data Alignment and Sample Construction

I compare the population census (IPEDS) and the sample panel (LinkedIn/PDL) at the cell level, defined as the institution$\times$degree level$\times$graduation year$\times$field of study level.

- **Population Data:** The IPEDS Completions baseline is restricted to bachelor's degrees (award level 5) and master's degrees (award level 7), capturing first majors conferred between 2008 and 2019.
- **Sample Data:** The LinkedIn/PDL panel is structured at the individual degree level. Each record is resolved to an IPEDS UNITID, and free-text major fields are mapped onto standardized two-digit Classification of Instructional Programs (CIP2) codes via a curated crosswalk. Fields of study are evaluated as CIP2 share vectors within each institution-year cell.

To ensure statistical reliability, two primary filters are applied to the data:

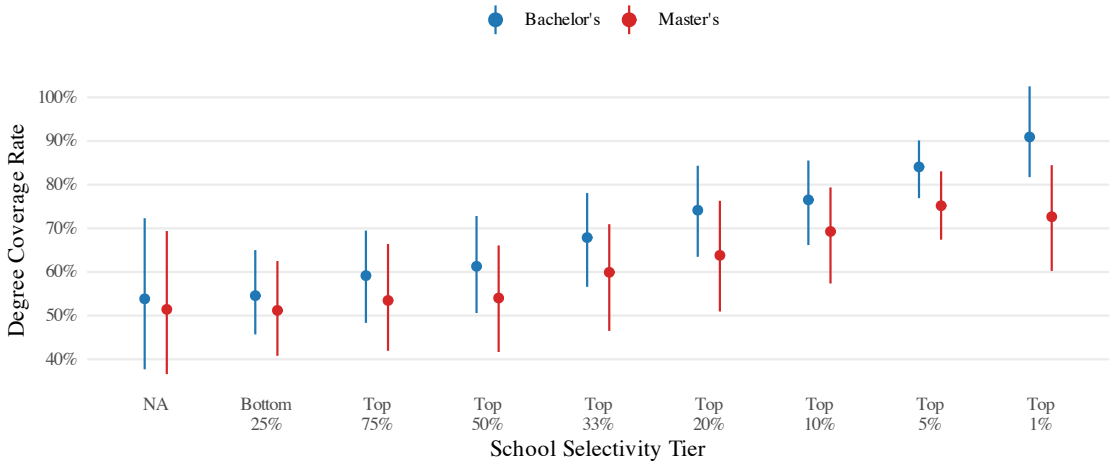
1. **Overlap requirement:** An institution-year cell is retained only if it contains at least 50 graduates in the IPEDS census and 20 matched profiles in the LinkedIn panel. Out of 54,277 potential school-year cells, 29,536 satisfy this criterion. While this percentage can seem low, the majority of drops can be attributed to misclassified schools in the PDL data.
2. **Crosswalk Collision Filter:** Institutional cells subject to crosswalk attribution errors—where a single institutional identifier in LinkedIn erroneously aggregates a multi-campus system or an online-only university chain—are identified by an impossible coverage rate exceeding 100%. These collided cells are excluded from the primary analytic samples to preserve the integrity of campus-level tiering.

D.2 Coverage and Institutional Selectivity

Coverage is defined as the ratio of captured LinkedIn profiles to the total IPEDS graduate count. Aggregated across all fields and graduation years, the panel captures **56.3%** of U.S. bachelor's graduates and **50.6%** of master's graduates within the collision-free, volume-weighted sample.

However, coverage is not uniform across higher education institutions. Figure D1 demonstrates that the median school-level coverage rate rises monotonically with institutional admissions selectivity. Median coverage ranges from approximately 57% for institutions in the lowest selectivity tier (open or unreported admissions) to 83% for institutions in the most selective tier, indicating that graduates of elite institutions are significantly more likely to maintain professional online profiles.

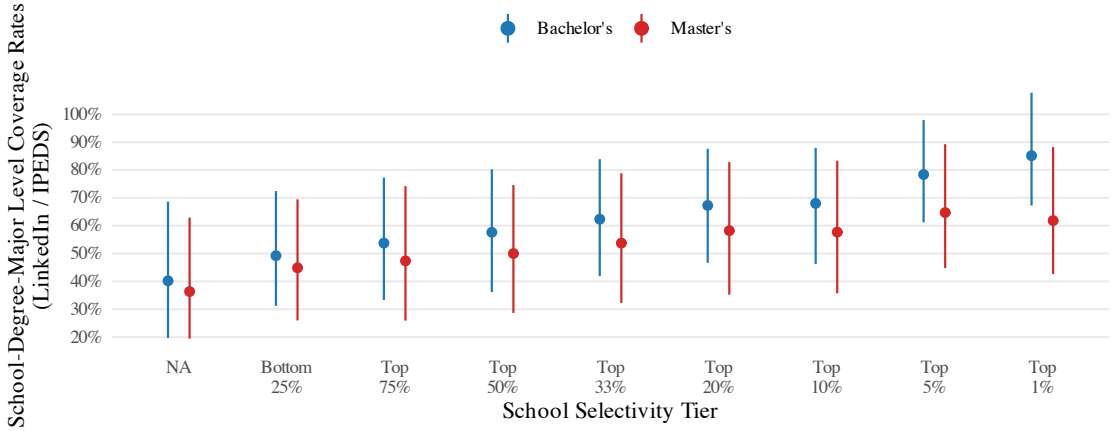
Figure D1: **Institutional Coverage by Admissions Selectivity Tier**



Median (point) with IQR (range). Tier 1 = open / not reported.

This institutional gradient persists at the finer granular level at which the dataset is utilized (institution × degree × major), as illustrated in Figure D2. Volume-weighted cell-level coverage climbs from 26% in the lowest selectivity tier to 88% in the most selective tier for bachelor's degrees, and from 28% to 77% for master's degrees.

Figure D2: **Granular Cell-Level Coverage by Selectivity Tier**

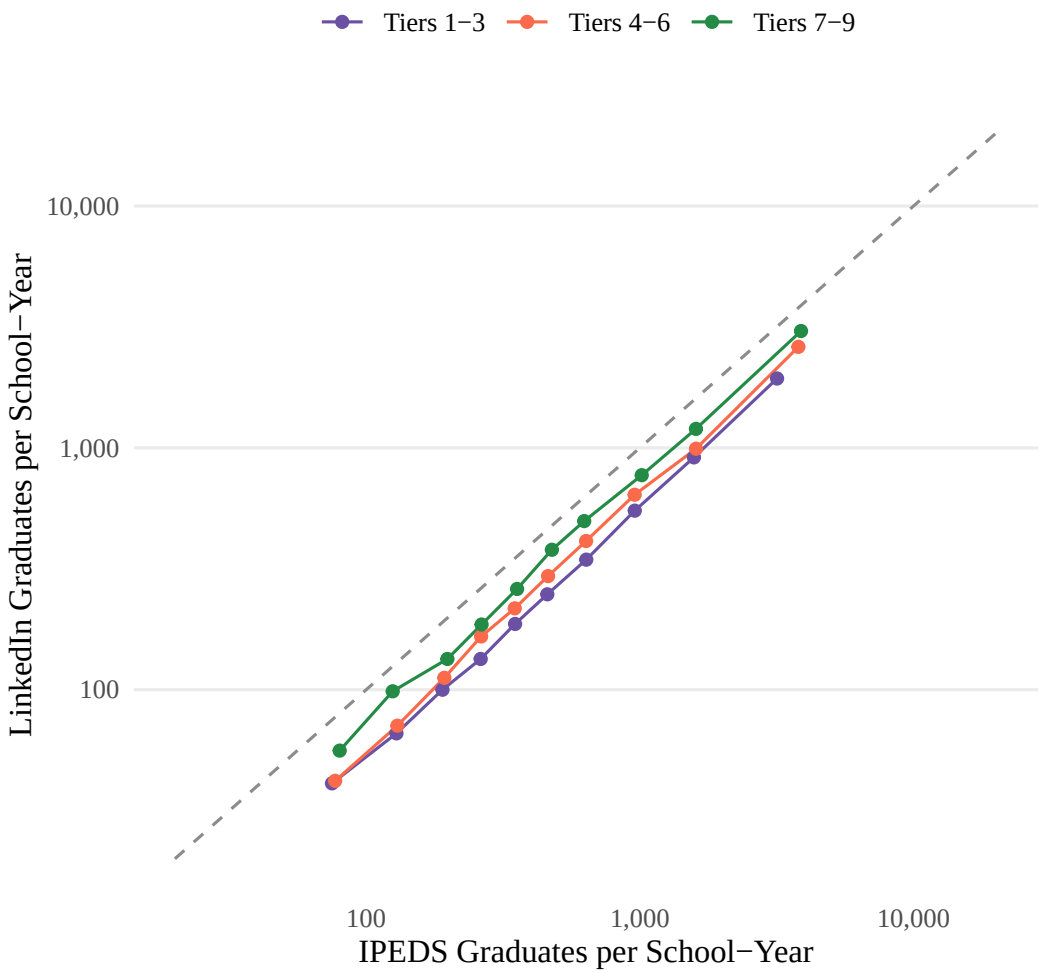


Median (point) + IQR (range) of cell coverage rates among school x major cells with at least one LinkedIn grad (IPEDS >= 25, CIP 15/46 excluded). Zero-coverage cells are excluded and are common at low tiers (~half of bachelor's cells at tier 1 (NA)), so these are coverage rates CONDITIONAL on any LinkedIn presence.

Crucially, within any given selectivity tier, coverage exhibits size-neutrality. Figure D3 plots log LinkedIn graduate counts against log IPEDS graduate counts. Within each selectivity band,

the relationship runs strictly parallel to the 45-degree identity line. A formal elasticity estimate indicates that LinkedIn volume scales one-for-one with cohort size (pooled elasticity $\approx$ 1.02, $R^2 \approx 0.90$). Consequently, the proportional under-reporting shortfall is invariant to the size of an academic program, and the vertical dispersion between bands reflects the pure institutional selectivity gradient.

Figure D3: Size-Neutrality of Coverage Within Selectivity Bands



Notes: Binned medians of graduate counts within IPEDS−size deciles by tier band

D.3 Within-School Composition: The Field Mix

While absolute coverage captures level differences, establishing the representativeness of the panel also requires verifying that the internal *distribution* of academic fields within schools is accurately reproduced. Isummarize the fidelity of this within-school major composition using the Dissimilar-

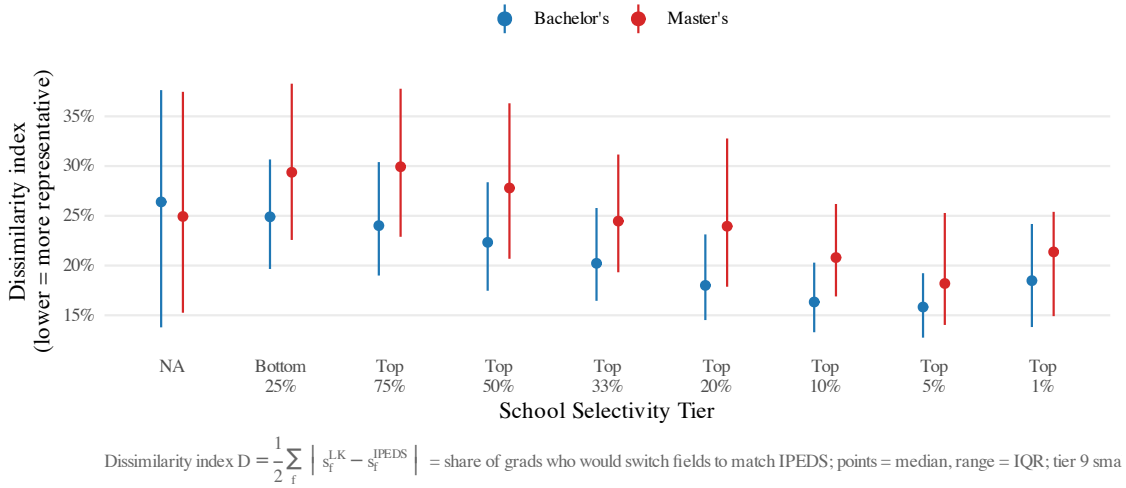
ity Index (D), which is mathematically equivalent to the total variation distance:

$$D = \frac{1}{2} \sum_{f} \left| s_f^{LK} - s_f^{IPEDS} \right| \tag{D.1}$$

where s_f^{LK} and s_f^{IPEDS} represent the shares of field f within an institution-year in LinkedIn and IPEDS, respectively. The index represents the proportion of graduates in the LinkedIn sample who would need to reallocate to alternative fields for the sample distribution to perfectly replicate the population baseline. As such, a higher dissimilarity index value corresponds to an institution where certain fields are reported disproportionally higher than others.

As shown in Figure D4, internal compositional divergence falls as institutional selectivity increases. The median dissimilarity index decreases from approximately 26% in the lowest tier to 16% in the highest tier. Master's programs exhibit slightly higher compositional divergence than bachelor's programs (overall medians of 27% versus 22%), which is consistent with smaller cell counts and greater heterogeneity in specialized professional program titles.

Figure D4: Within-School Field Mix Dissimilarity Across Selectivity Tiers



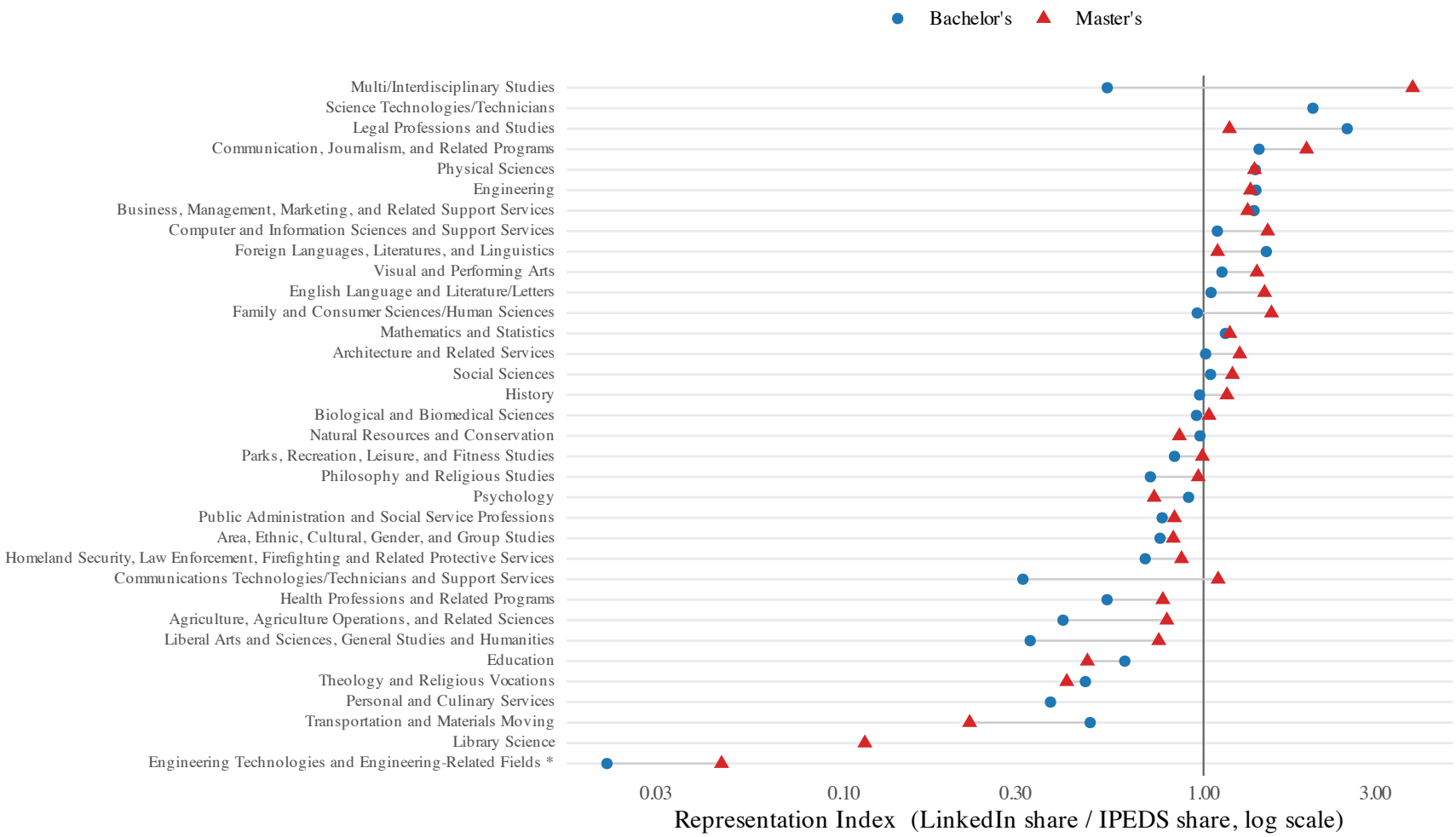
D.4 Field Bias and Classification Artifacts

A main concern with using LinkedIn data is that compositional errors are systematic and driven by field-level preferences for using LinkedIn. Figure D5 tracks the representation index, defined as the ratio of a field's share in the LinkedIn panel to its share in the IPEDS census. Values above unity signify that a field is over-represented in the sample relative to the population field mix.

The panel consistently over-represents professional, commercial, and quantitative fields, such as Business, Engineering, Communications, and Legal Studies. Conversely, it under-represents care-, teaching-, and public-oriented disciplines, including Health Professions, Education, and Liberal Arts/General Studies. This pattern holds true across both bachelor's and master's degree levels. Importantly, the nature of mapping free-text majors to CIP2 codes result in mis-categorization

in certain cases such as inputting ”Engineering Technologies and Engineering-Related Fields” as ”Engineering” and hence getting mapped to Engineering (CIP2-14) instead of the proper CIP2-15). However, the large distortions in representation seen in Figure D5 does not translate to the overall sample as major-relevant variables are calculated using the set of canonical majors in Appendix K.

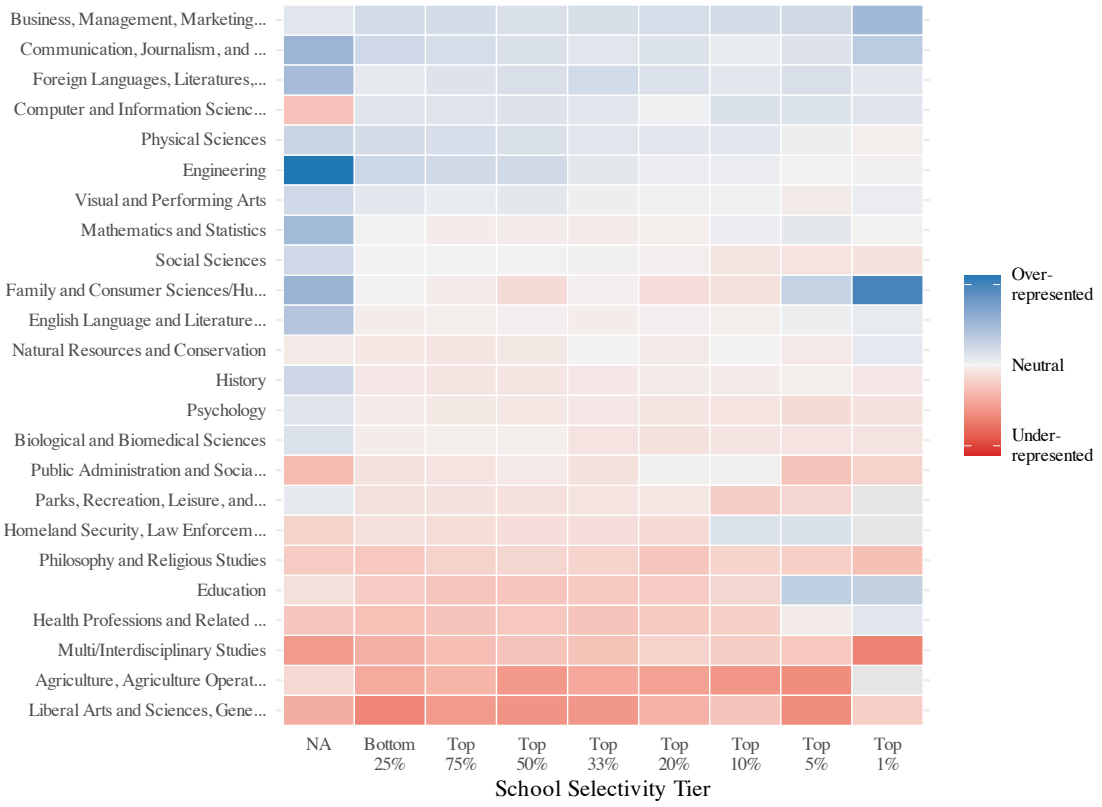
Figure D5: Field-Level Representation Index Relative to IPEDS Census Baseline



*Pooled over years, collision-free, IPEDS >= 5,000 grads. >1 over-, <1 under-represented. Circles = bachelor's, triangles = master's. * = major->CIP self-report seam (CIP 15 eng-tech, 46 construction): extreme values are matching artifacts, not sampling, showing free-text major->CIP matching is hard; Engineering (CIP 14) is correspondingly inflated.*

This field bias is uniform across institutional tiers rather than driven by specific subsets of schools. As shown in Figure D6, the institutional selectivity gradient and the field-level composition bias operate as roughly additive components: Health and Education remain under-represented across virtually all tiers, while Business and Communications are consistently over-represented.

Figure D6: Joint Distribution of Representation Across Fields and Tiers



Representation index = within-tier LinkedIn field share / IPEDS field share. Blue over-, red under-represented. Eng-tech (CIP 15) excluded; Engineering (CIP 14) remains inflated as eng-tech grads self-label as 'Engineering'.

The cell-grain coverage heatmap in Figure D7 synthesizes these dual margins. While coverage generally increases from left to right with selectivity across all fields, care and teaching fields remain compressed across nearly all tiers. Notably, several cells in the highest selectivity tiers exhibit coverage rates exceeding 1.0, most visibly in Business (CIP 52), which reaches 1.27 and 2.71 in the top two tiers.

This phenomenon does not reflect genuine over-capture of business degrees. Rather, elite institutions confer very few undergraduate business degrees, resulting in an exceptionally small IPEDS denominator. Concurrently, LinkedIn users who graduate from elite institutions with degrees in Economics or other quantitative social sciences and subsequently enter corporate or financial sectors frequently self-label their major as "Business." For example, Harvard reportedly conferred 0 business BAs from 2009 to 2018, but there are 3,641 LinkedIn users reporting obtaining a business bachelor's in the same period. This inflation in CIP 52 is mirrored by a corresponding deflation in Social Sciences (CIP 45), which remains below unity at those same elite tiers.

Figure D7: **Volume-Weighted Cell Coverage Heatmap**

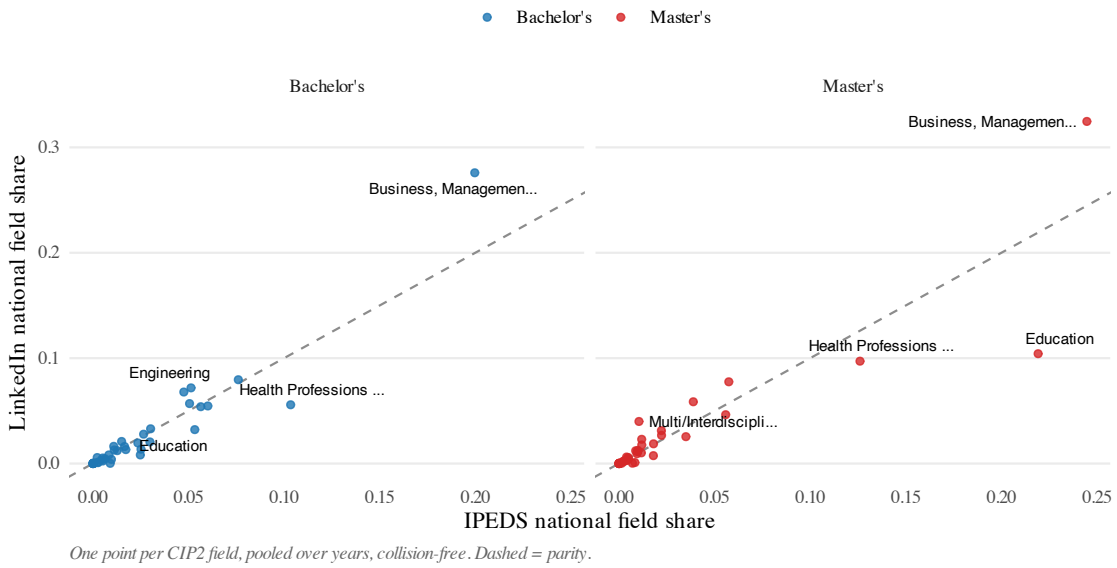


Figures reported are volume-weighted coverage rates; Grey = fewer than 50 IPEDS grads (too thin). Engineeting tech (CIP 15) and construction (CIP 46) are excluded due to string matching misrepresentation.

D.5 Aggregate Robustness and Data Quality Diagnostics

National Portfolio Tracking: As a broad aggregate validation, Figure D8 plots the national CIP2 field shares in LinkedIn against those in IPEDS, pooling all available years. The overall composition tracks the population distribution closely, yielding Pearson correlation coefficients of 0.948 for bachelor's degrees and 0.897 for master's degrees.

Figure D8: **National Field Shares: LinkedIn versus IPEDS Baseline**



One point per CIP2 field, pooled over years, collision-free. Dashed = parity.

Institutional Crosswalk Collisions: As noted in Appendix D.1, coverage rates above 100% at the institutional level indicate mapping errors where LinkedIn's institutional identifiers merge distinct campuses or online chains into a single corporate hub. These account for 14.1% of the raw LinkedIn degree records and are concentrated heavily among lower-tier institutions (representing 57% of the total volume in the lowest selectivity tier). Rather than aggregating these records and losing institutional granularity, these cells are dropped from the primary sample.

Classification Seams: Certain severe field deviations stem from mapping artifacts rather than true sampling bias. For instance, "Engineering Technologies" (CIP 15) approaches zero in LinkedIn because graduates likely systematically label their majors as "Engineering" (CIP 14), causing a corresponding artificial inflation in the latter. Similarly, the text string "construction" initially inflated "Construction Trades" (CIP 46) before being algorithmically corrected to Construction Management under Business (CIP 52).

Sensitivity Checks: Three additional tests verify that the core results are robust:

1. *Match Quality:* Every in-scope degree is resolved via direct, un-ambiguous institutional identifiers rather than fuzzy name matching, ensuring that coverage patterns are not artifacts of alignment errors.

2. *Date Imputation:* Restricting the sample to the 2.8% of records with precise, month-resolved graduation dates leaves the national field-share correlations virtually unchanged (the bachelor's correlation shifts minimally from 0.948 to 0.967).

3. *Profile-Completion Lag:* Over time, total LinkedIn degree counts remain below IPEDS and exhibit an inverted-U trajectory that peaks around 2015-2016. This pattern is the signature of a multi-year lag in profile creation and completion by recent graduates, rather than a shifting structural sampling bias. Manual spot-checks against public institutional records confirm the validity of the underlying counts.

D.6 Methodological Implications

Because absolute profile coverage scales with institutional selectivity and the sample composition tilts toward corporate and quantitative fields, the raw LinkedIn panel over-represents graduates from elite schools and professional programs relative to the true U.S. graduate population. For research designs that are sensitive to these margins, these systematic discrepancies can be corrected. Researchers can construct post-stratification weights that align LinkedIn cells (at the institution $\times$ CIP2 $\times$ year level) to the true IPEDS population margins, effectively eliminating first-order selection bias.

Table D1: **Cell-Level Coverage and Capture Rates by Selectivity Tier (Bachelor's Degrees)**

Selectivity Tier	Number of Cells	Volume-Weighted Coverage	% IPEDS Captured
1 (Open / Unreported)	39,843	0.265	26.5
2 (Bottom 25%)	59,259	0.479	47.9
3 (Top 75%)	74,124	0.557	55.7
4 (Top 50%)	50,202	0.577	57.7
5 (Top 33%)	40,316	0.622	62.2
6 (Top 20%)	27,166	0.635	63.5
7 (Top 10%)	14,705	0.658	65.8
8 (Top 5%)	12,665	0.798	79.8
9 (Top 1%)	3,087	0.882	88.2
All Tiers	**321,367**	**0.563**	**56.3**

Notes: Volume-weighted coverage measures the percentage of the total IPEDS graduation cohort successfully matched to corresponding profiles in the professional panel.

E Data Construction

E.1 Why a Longitudinal Panel

Employer-learning is a within-worker process, and the workhorse sources for immigrant labor-market research—repeated cross-sections such as the decennial Census and the ACS—cannot follow the same individual across waves. Estimated experience-earnings and assimilation profiles therefore identify the true within-person path only if cohort composition is stable across waves, and that assumption fails when out-migration is selective. Lubotsky (2007) matches survey respondents to longitudinal Social Security earnings records and shows that lower-earning immigrants disproportionately leave the United States, so the surviving cross-section mechanically appears to assimilate; the longitudinal profile is markedly flatter than its cross-sectional counterpart, which overstates immigrant earnings growth by roughly half. Panel data does not fully resolve this, since longitudinal samples are themselves subject to non-random attrition (Beenstock et al., 2010), and linked historical records reveal analogous selection among earlier immigrant waves (Collins and Zimran, 2023). I therefore construct a novel longitudinal dataset.

E.2 Classifying Job Titles

To operationalize the LinkedIn panel, it is necessary to map self-reported, unstructured job titles into the standardized O*NET Standard Occupational Classification (SOC) codes. Because user-generated profiles frequently contain idiosyncratic, highly specific, or non-standard occupational phrasing, traditional fuzzy-matching algorithms or dictionary-based approaches yield low match rates.

During training, the fine-tuned model is prompted to evaluate the self-reported job title and output only the corresponding 8-digit SOC code. To prevent forced matches on irrelevant data, the training data explicitly instructs the model to return a null identifier (00-0000.00) for gibberish or un-mappable titles (e.g., "gfliwn askdm"). An analysis of the preliminary predictions demonstrates the model's robust semantic understanding. It successfully maps highly specific titles (e.g., "asset & reliability engineer" to Industrial Engineers) and academic appointments (e.g., "international visiting attorney" to Lawyers). Crucially for a study examining immigrant labor market trajectories, the model inherently processes and classifies non-English job titles without requiring ex-ante translation; for example, successfully mapping "網頁工程師" (Website Engineer) directly to Computer Systems Engineers/Architects and "博士生" (PhD Student) to Post-secondary Teaching Assistants. More examples are reported in Table A4.

E.3 Online-Profile Data in Labor Research

The tradeoff between platform selection and panel coverage is well understood in the emerging literature that repurposes online-profile and resume data for labor-market research. Berry et al. (2024) show that LinkedIn-derived samples misrepresent the demographic composition of employment in known and measurable ways, while demonstrating that the data can nonetheless be benchmarked against administrative population totals, and resume data at sufficient scale now supports credible causal designs (Van Dijcke et al., 2024). The analysis here follows that logic, trading the

return-migration bias that contaminates survey data for a platform selection bias whose structure is observable against the IPEDS benchmark and can be addressed with post-stratification weights.

E.4 The Conventional Mismatch Taxonomy

The main taxonomy of occupational mismatch a la Hartog (2000) can be visualized in the following way. Let the xy-plane represent the educational modality of an occupation and relabel it the fa-plane. The f-axis represents the field of specialization and the a-axis represents education attainment. For worker $i \in \boldsymbol{\mathcal{I}}$ with education $e_i = (f_i, a_i)$. The typical education requirement for job j is defined by the modal education e_M^j of workers in job j, i.e., $e_M^j = \underset{e}{argmax} \sum_{i \in \{i \in \boldsymbol{\mathcal{I}} | i \text{ works at job } j\}} \mathbb{1}\{e = e_i\}$. The conventional way to define vertical and horizontal mismatch is: (1) Worker i is vertically mismatched in job j if $a_i \neq \bar{a}^j$; and (2) Worker i is horizontally mismatched in job j if $f_i \neq \bar{f}^j$.

However, this taxonomy introduces two imprecision: (1) It implicitly assumes weak correlation between these two mismatch types, which can cause a bias in estimation using equilibrium observational data; (2) it assumes that education attainment is fixed upon job placement. To account for (1), it is prudent to define the two mismatch types separately and evaluate the correlation between them. To account for (2), I use the O*NET occupation-level modal required education for entry to define over-education.

A reasonable way to define horizontal mismatch, adopted by Li and Lu (2023), is to look at mismatches in the underlying tasks requirement of a previous job. However, it is difficult to cleanly separate transient jobs from skill-forming jobs in observational data. Moreover, Lise and Postel-Vinay (2020) show that workers skill-sorting towards occupations that are desirable is a non-negligible mechanism in determining their career paths. This contaminates the identification of horizontal mismatch via task comparisons in observational data.

E.5 Horizontal-Mismatch Cutoff

The matched set for each degree-by-major group is defined by the share of FUS jobs it covers rather than by a fixed number of occupations. An earlier version of the measure counted a job as matched if it was among the 20 occupations most commonly held by FUS graduates with the same degree and major. A fixed count covers a different share of workers in every group: graduates of a major that leads to few occupations are almost always matched, and graduates of a major that spreads across many are often mismatched, so the FUS mismatch rate under a count rule mainly reflects how dispersed a major's graduates are. Dispersion also falls with the degree. The median matched set under the share rule contains 34 occupations for bachelor's groups, 22 for master's groups, and 11 for doctoral groups, so a 20-occupation cutoff was strict for bachelor's and lenient for doctoral graduates, and a lower doctoral mismatch rate under it would have been mechanical. The share rule fixes the matched set at half of each group's FUS jobs, with occupations tied at the margin included, and counts jobs over the whole career, so that an occupation reached through promotion counts as a match (Section 3.3). Figure 17 shows which jobs it classifies as mismatched.

E.6 Certification and On-the-Job Training

I hypothesize that the presence of institutional entry barriers, specifically professional certifications and on-the-job training (OJT) requirements, alters the signaling value of formal education. If a professional certification is a prerequisite for employment, the certification itself serves as a more informative signal of worker ability than the underlying degree. Conversely, in occupations where OJT is essential, educational attainment can be more informative to the employer. This assumes a structural link between formal education and a worker's latent productivity type (θ), where higher attainment signals a lower cost of training or a higher rate of skill acquisition during the OJT phase. While Figure 18 explores the relationship between occupational mismatch and the requirements for certification and OJT, the empirical patterns yield few deterministic implications.

E.7 Principal Components of O*NET Skills and Abilities

O*NET provides both level and importance scores for 87 distinct SAs associated with each occupation; for example, the occupation "Chief Executives" requires an average level score of 4.88 (on a 0–7 scale) for oral comprehension.

To interpret these factors, I extract the detailed O*NET descriptions for the 20 SAs with the highest loadings on each component. I then utilize a Large Language Model (LLM), Microsoft Copilot, to synthesize these descriptions into encompassing labels, weighting the semantic content by the magnitude of the respective loadings (Microsoft, 2025).

These components account for 73% of the variation in SAs with the next principal component accounting for roughly 3%.

By comparing the modal degree of citizens vs. non-citizens in an occupation, we can explore the relationship between these characteristics and the probability of mismatches. Figure 20 illustrates the relationship between these skill dimensions and occupational mismatch. Both over-education (OE) and under-education (UE) occur almost exclusively in occupations requiring above-average CIC. Furthermore, the direction of mismatch is highly correlated with PJC requirements: occupations with high PJC demands are significantly more likely to be characterized by under-education. In contrast, Figure 19 shows that the second principal component, AC, exhibits no discernible relationship with occupational mismatch.

E.8 Why Under-Education Is Not Classified

The analysis keeps every occupation, and a worker whose degree falls below the occupation's entry requirement is coded as not over-educated (Section 3.3). Under-education is not treated as a separate outcome because the literature suggests that it arises from a different process than over-education. In stable under-education matches, workers substitute unobserved existing human capital, such as innate ability, extensive tenure, or OJT, for formal academic credentials (). This asymmetry aligns with the classical human capital theory: over-educated workers accept lower-level positions to acquire general experience, whereas under-educated workers accumulate occupation-specific human capital that tightly binds them to their current roles (). Mobility patterns corroborate this distinction; Alba-Ramirez (1993) demonstrates that under-educated workers are far more likely to change firms within the same occupation rather than switch occupational categories en-

tirely. Consequently, under-education reflects a robust match where a worker's specific skill bundle satisfies occupational demands, largely eliminating the incentive for occupational switching.

Empirical training patterns provide further evidence of this structural divergence. As Hartog (2000) notes, under-educated workers are significantly more likely to receive job-specific training, whereas over-educated workers are less likely to require or receive it. Because this training constitutes a direct investment in occupation-specific human capital, the duration of an under-educated match positively correlates with specific training accumulation. This reinforces the premise that under-educated workers sort into higher-level roles requiring targeted learning, and that the resulting skill accumulation further entrenches them within their occupational track. Collectively, this evidence implies that under-educated workers operate under a distinct search and matching process governed by occupational lock and the relative permanence of specific human capital investments.

E.9 Ranking Details

Recovering $\tilde{V}$ requires that observed moves reflect choice rather than displacement, and that the mobility network is connected and stable over the window used to estimate it.

All regressions use the rankings as $\ln(\text{PageRank score})$. Log-transformation removes the mechanical dependence of raw scores on stratum size, converting a multiplicative scaling into an additive shift absorbed by stratum fixed effects; it compresses the heavy right skew that concentrated worker flows induce, preventing hub occupations from dominating; and, unlike ordinal ranks, it preserves the ratio comparisons that a recursively multiplicative ranking makes meaningful.[12]

Within a given degree-major pair, a worker observed leaving occupation a and entering occupation b is treated as having revealed that b was the most attractive reachable alternative at the time.

The generality statistic DUH is not a revealed-preference object and does not feed into the PageRank estimation above. This statistic is between 0 and 1, and it reflects the distribution of realized SOC/degree-major within a degree-major/SOC.

[12] Ordinal ranks discard magnitude and impose equal spacing between adjacent positions, neither of which the network construction supports. A level specification, by contrast, has no interpretable units. In the log-median-wage regressions the log specification yields a standard semi-elasticity: the percentage change in SOC median wage associated with a one percent increase in an occupation's or firm's utility.

F Validation of Occupation Coding

Occupation is the unit on which every outcome in this paper is measured: over-education compares a worker's degree with the O*NET entry requirement of their occupation, and the wage, ranking, and task measures are occupation-level. This appendix documents how accurately the classifier described in Section 3.1 assigns occupations, and compares that accuracy with the benchmarks the measurement literature reports for human coders and for automated coding tools.

F.1 What the Match Rate Measures

Of 76.1 million unique job titles, the classifier returns a code for 97.6 percent and abstains, with the null code, for the rest. The match rate is therefore a coverage rate, not an accuracy rate. Fewer than 0.05 percent of titles receive a code outside the 867 SOC-2018 detailed occupations; these are set to missing before any analysis. Weighted by person-years rather than by unique titles, 96.4 percent of employed job-years carry a valid code. Coverage is the margin on which the classifier differs most from conventional approaches: a token-weighted (TF-IDF) dictionary match of the same titles left roughly 80 percent of them unmatched.

F.2 Benchmarks from the Literature

Occupation coding is an inexact procedure even when done by trained people, and agreement falls sharply as the classification becomes more detailed. Table F1 collects the benchmarks most comparable to this setting. In the January 1977 CPS, workers' and employers' reports of the same job agree on the three-digit occupation for 57.6 percent of workers and on the one-digit occupation for 81 percent (Mellow and Sider, 1983); Kambourov and Manovskii (2008) show that errors of this size are large enough to distort measured occupational mobility. When two expert coders independently code the same 1,000 job histories into SOC-2010, they agree on 48.1 percent of detailed (six-digit) codes and on 62.9 percent at three digits (Rémen et al., 2018). Among automated tools, SOCcer, which uses the job title, task description, and industry, agrees with expert coders on 44.5 percent of six-digit codes and 76.3 percent of two-digit codes (Russ et al., 2016); the NIOSH coding system reaches a kappa of 0.84 at two digits and 0.56–0.70 at four digits against manual coding (Schmitz and Forst, 2016); and German coding agencies coding the same responses into ISCO reach a kappa of about 0.5 at four digits (Massing et al., 2019). Much higher figures are reported only where the input text is short and highly standardized, such as historical census occupation strings coded into HISCO (Dahl et al., 2024).

F.3 Validation Design

I validate the classifier in two samples.

Held-out training-source titles. The classifier is fine-tuned on LinkUp job-posting titles, each labeled by LinkUp with an O*NET-SOC code. I draw a seeded random sample of 1,000 titles from the held-out portion of these data, which the model never saw during training, and code each one with the production prompt and settings. Agreement with LinkUp's label measures how closely

Table F1: Occupation-Coding Agreement in the Literature and in This Paper

Source	Classification	Comparison	Detailed	Broad
Mellow and Sider (1983)	1970 Census	Worker vs. employer report, CPS	57.6 (3-digit)	81 (1-digit)
Rémen et al. (2018)	SOC-2010	Two expert coders, same jobs	48.1 (6-digit)	62.9 (3-digit)
Russ et al. (2016) (SOCcer)	SOC-2010	Algorithm vs. expert codes; title, task, industry	44.5 (6-digit)	76.3 (2-digit)
Schmitz and Forst (2016) (NIOCCS)	SOC	Algorithm vs. manual codes, kappa	0.56–0.70 (4-digit)	0.84 (2-digit)
Massing et al. (2019)	ISCO	Coding agencies, same responses, kappa	≈0.5 (4-digit)	—
Dahl et al. (2024) (OccCANINE)	HISCO	Algorithm vs. source codes, historical records	96	—
This paper	SOC-2018	Classifier vs. independent blind coding, US first jobs	56.1 (6-digit)	78.8 (2-digit)
			68.5 (4-digit)	70.5 (3-digit)

Notes: Percent agreement unless marked as kappa. "Detailed" and "Broad" report the most and least detailed levels each source gives, with the digit level in parentheses. Mellow and Sider's figures are as reported in Kambourov and Manovskii (2008). The figures in the bottom row are weighted by sampling stratum and count a match on the independent coder's alternative code (Table F2).

the classifier reproduces its training source, not its accuracy: LinkUp's labels are themselves noisy, and for non-English titles they are frequently wrong.

Independent blind coding of first jobs. To measure accuracy on the titles that enter the analysis, I draw 1,005 US first jobs, one per worker, from the first-job samples of the 2003–2021 graduation cohorts. The sample is stratified by pathway (FUS, CtS, CtW) and by whether the title is written in plain Latin script. Each job is coded independently by a separate general-purpose large language model (Anthropic's Claude), which sees the title, employer, industry, seniority, country, and year, but not the classifier's code or the worker's pathway, and uses no external lookup. For titles that support more than one defensible code, the independent coder records an alternative code and an ambiguity flag. The same coder also codes the 1,000 held-out LinkUp titles from the title alone, without seeing either LinkUp's label or the classifier's prediction. [PENDING: A human reviewer adjudicated, blind to source, a random subsample of 50 disagreements and 25 agreements; the classifier's code was judged correct or equally defensible in [x] percent of disagreements and [y] percent of agreements, implying a six-digit accuracy of [z] percent (95 percent CI [..]).]

F.4 Results

Table F2 reports agreement at each level of the SOC hierarchy. Three findings stand out.

First, the classifier performs at the level of human coding. Against independent coding of US first jobs, it agrees on 56.1 percent of detailed (six-digit) occupations, 68.5 percent at four digits, and 78.8 percent at the two-digit major group. These rates are at or above the agreement between two expert coders on SOC-2010 and those of SOCcer, which also uses task descriptions. Most disagreement is within a four-digit group: agreement rises by 12 points from six to four digits, and by 3 points from six to five digits. On held-out LinkUp titles, the classifier agrees with the independent coder (49.6 percent at six digits) as often as LinkUp's own labels do (49.0 percent),

so departures from the training labels are not, on net, departures from the correct code.

Second, the classifier's confidence score separates correct from incorrect codes. Among first jobs coded with confidence of at least 0.9, agreement with the independent coder is 76.1 percent; below 0.8, it is 31.0 percent. On held-out LinkUp titles, agreement with LinkUp's label is 85.7 percent at confidence of at least 0.95 and 9.2 percent between 0.5 and 0.8.

Third, coding error is similar across pathways on the margin that matters for the outcomes. Six-digit agreement is 56.8 percent for FUS workers and 51.4 and 52.0 percent for CtS and CtW workers; the difference is about 1 to 1.5 standard errors with roughly 335 jobs per group, and the minimum difference detectable with 80 percent power is about 11 points. Over-education depends on the occupation's education requirement, not on the six-digit code itself: when the two codings disagree, the occupation's modal degree differs for 18.5, 19.7, and 19.5 percent of FUS, CtS, and CtW first jobs. Non-English titles, on which the classifier performs worst (29.7 percent agreement with LinkUp's labels, against 52.2 percent for English titles), are 0.1 percent of FUS and CtS US first jobs and 0.8 percent of CtW US first jobs.

Table F2: Agreement of the Occupation Classifier by SOC Level

	2-digit (major)	3-digit (minor)	4-digit	5-digit (broad)	6-digit (detailed)
A. US first jobs: classifier vs. independent blind coding (n = 1,000)					
Pooled	78.8	70.5	68.5	59.4	56.1
FUS	79.2	71.6	69.6	60.4	56.8
CtS	80.7	68.7	64.6	55.4	51.4
CtW	74.8	63.2	61.7	52.8	52.0
B. Held-out LinkUp titles					
Classifier vs. LinkUp label (n = 1,000)	72.5	64.0	62.5	58.1	50.8
Classifier vs. independent coding (n = 943)	73.6	66.3	63.7	57.2	49.6
LinkUp label vs. independent coding (n = 943)	70.4	63.1	61.1	54.9	49.0

Notes: Percent agreement after mapping SOC-2010 codes to SOC-2018. Comparisons with independent coding count a match on the independent coder's alternative code. Panel A is weighted by sampling stratum (pathway by title script), and the pooled row by population. Panel A excludes the five jobs the independent coder judged uncodable; panel B excludes 44 titles whose labels the independent coder had seen before coding and the titles it judged uncodable. The four-digit level is the first four digits (e.g., 11-10), the level to which the ACS and CPS mask some detailed codes.

F.5 Limitations

Three limitations qualify these results. First, the independent coder is itself a language model; agreement with it is not accuracy, and the two models may share blind spots on genuinely ambiguous titles. The human adjudication above bounds this concern. Second, about half of first-job titles are, in the independent coder's judgment, ambiguous at six digits (for example, "associate," "analyst," "project manager"); on titles it judges unambiguous, agreement is 74.1 percent. Much of the residual disagreement reflects the granularity of the taxonomy rather than classifier error, which is one reason the ACS and CPS mask some detailed codes. Generic titles held by highly educated workers are also the case in which coding error mechanically produces over-education. The leading examples are "associate" held by JD holders, which the title alone codes as customer service representatives, and "resident" held by MD holders, coded as residential advisors. For a targeted set of 2.7 million title–employer combinations (19.1 million job-years) the classifier is therefore

re-run with the employer, industry, and seniority added to the prompt. The targets are generic titles, codes with confidence below 0.5, and licensed doctorate holders in occupations requiring less than a master's degree. The re-coded occupation is kept only if it does not lower the occupation's modal education, which corrects about 96 percent of the misclassified "associate" and "resident" records. Adding context to every title does not help: on the validation sample it improves generic titles but degrades specific ones, and leaves overall agreement unchanged. Third, the title normalization applied before classification keeps only the text before the first separator ("/", "|", " – "). This removes occupational content from 7.5 percent of US first jobs (7.6, 8.4, and 6.5 percent for FUS, CtS, and CtW), on which agreement is 33 percent, against 58 percent elsewhere. The affected share is similar across pathways.

G Selection on the Intensive Margin of the STEM-OPT Design

Table G1 collects the notation of Section 4.3, which this appendix uses throughout. Symbols used only in this appendix are defined where they first appear.

Table G1: Notation for the STEM-OPT Design

Symbol	Meaning
i	Worker
e	Terminal degree (master's or doctorate, where doctorate pools the PhD and non-licensure doctorates), the empirical counterpart of the education choice e in the model
$c(i)$	Year worker i completed the terminal degree (graduation cohort)
Post_c	$\mathbb{1}\{c \geq 2008\}$
STEM_i	Terminal-degree major is STEM
D_i	Treatment: $\text{STEM}_i \times \text{Post}_{c(i)}$
T_i, τ	Terminal-institution tier; τ indexes tiers
DUHMajor_i	Major generality
t	Years since the terminal degree
k	Job-quality outcome
Abroad_i	First post-degree job is outside the United States
S_{it}	Holds a US first job and is still US-resident at t; $S_{it}(1)$ and $S_{it}(0)$ with and without the extension
Y_{it}^k	Outcome k at t; $Y_{it}^k(1)$ and $Y_{it}^k(0)$ with and without the extension
ω_t	Share of treated stayers at t who would have stayed without the extension
δ_e	Extensive-margin difference-in-differences
δ_{et}^k	Intensive-margin difference-in-differences among stayers at t
ϕ_{ec}	Event-study coefficient for graduation cohort c

This appendix derives Equation (14) and states what Assumption 7 requires. The difference-in-differences in Equation (10) is estimated on workers who hold a US first job and are still US-resident t years after graduation. The extension changes who is in that sample, so the four group means average over different populations, and the estimate cannot be interpreted as an average effect on a fixed set of workers without further steps. Two distinct selection problems arise. The first is that who stays may change across the 2008 cutoff for reasons unrelated to the policy; this enters the identifying assumption. The second is that the policy itself changes who stays; this enters the interpretation of the estimand. I treat them in that order.

Throughout, I drop the outcome, degree, and time indices k, e, and t, and I ignore covariates. The derivation is therefore exact for the difference-in-differences of group means. With tier and major generality as controls, the same argument applies among workers with the same characteristics, and the regression estimate combines these comparisons with weights determined by OLS, so the decomposition holds for it approximately.

G.1 Setup

Index the four groups by (s, a), where $s = 1$ for STEM and $a = 1$ for graduation in or after 2008, and let $D_i = \mathbb{1}\{s = 1,\, a = 1\}$. Each worker has two potential outcomes, $Y_i(1)$ and $Y_i(0)$, with

and without the extension, and two potential stayer indicators, $S_i(1)$ and $S_i(0)$, equal to one if the worker would hold a US first job and still be US-resident at t under each regime. The data reveal $(Y_i, S_i) = (Y_i(1), S_i(1))$ in group $(1,1)$ and $(Y_i, S_i) = (Y_i(0), S_i(0))$ in the other three groups.

Each outcome is a benchmark attached to an occupation or an employer: the OEWS median wage, SOC PageRank, firm value, over-education, and horizontal mismatch. $Y_i(0)$ and $Y_i(1)$ are therefore defined for jobs held abroad as well as in the United States. The estimator never uses these outcomes: it averages only over stayers, and Assumption 7 is a statement about stayer means. Their role is interpretive. They make the mean outcome over all graduates in a group, stayers and leavers together, a well-defined object, which lets me write each stayer mean as a group mean plus a selection term in Appendix G.3, and they make that selection term observable in the untreated groups. For an outcome defined only for US jobs, the split would not exist, and Assumption 7 would have to be read directly as a restriction on stayer means, with the same content.

G.2 What the estimator recovers

Let $m_{sa} = E[Y_i \mid S_i = 1,\ (s,a)]$ denote the observed mean among stayers in group (s,a). The estimator is

$$\delta = (m_{11} - m_{10}) - (m_{01} - m_{00}). \tag{G.1}$$

Because no worker outside group $(1,1)$ is treated,

$$m_{sa} = E[Y_i(0) \mid S_i(0) = 1,\ (s,a)] \qquad \text{for } (s,a) \neq (1,1), \tag{G.2}$$

$$m_{11} = E[Y_i(1) \mid S_i(1) = 1,\ (1,1)]. \tag{G.3}$$

Define the unobserved counterfactual

$$m_{11}^0 \equiv E[Y_i(0) \mid S_i(0) = 1,\ (1,1)], \tag{G.4}$$

the mean outcome that STEM graduates in the post-period would have had without the extension, among those who would have stayed without it. Adding and subtracting m_{11}^0 gives the identity

$$\delta = \underbrace{(m_{11} - m_{11}^0)}_{\text{(A)}} + \underbrace{\left[(m_{11}^0 - m_{10}) - (m_{01} - m_{00})\right]}_{\text{(B)}}. \tag{G.5}$$

I define the counterfactual by conditioning on $S_i(0)$ rather than $S_i(1)$ for two reasons. First, by Equation (G.2), every term in (B) is then a mean of untreated outcomes among workers who would stay without the extension, so (B) is a difference-in-differences of like objects. Second, conditioning on $S_i(1) = 1$ would require the untreated outcome of workers who stay only because of the extension.

Assumption 7 states that (B) is zero. (A) is the object of interest and is taken up in Appendix G.4.

G.3 What (B) $= 0$ requires

The standard parallel-trends assumption restricts means of untreated outcomes over whole groups. Every term in (B) is instead a mean over a selected subset of each group: the workers who would stay without the extension. This subsection separates the two.

Step 1: a stayer mean is a group mean plus a selection term. Fix a group (s, a) and suppress it in the notation. Write $\rho = \Pr(S_i(0) = 1)$ for the untreated stay rate and

$$\nu^S = E[Y_i(0) \mid S_i(0) = 1], \qquad \nu^L = E[Y_i(0) \mid S_i(0) = 0], \qquad \nu = E[Y_i(0)] \tag{G.6}$$

for the mean untreated outcome among would-be stayers, among would-be leavers, and over the whole group. By the law of total expectation,

$$\nu = \rho\, \nu^S + (1 - \rho)\, \nu^L. \tag{G.7}$$

Subtracting this from ν^S,

$$\nu^S - \nu = \nu^S - \rho\, \nu^S - (1 - \rho)\, \nu^L = (1 - \rho)\, (\nu^S - \nu^L). \tag{G.8}$$

Define the selection term

$$\sigma \equiv \nu^S - \nu = (1 - \rho)\, \Delta, \qquad \Delta \equiv \nu^S - \nu^L. \tag{G.9}$$

The stayer mean is then

$$E[Y_i(0) \mid S_i(0) = 1] = \nu + \sigma. \tag{G.10}$$

σ is how far would-be stayers sit above the average graduate in their group. It is the product of two things: Δ, how different stayers are from leavers, and $1 - \rho$, how many leavers there are. If nobody leaves ($\rho = 1$) or stayers and leavers are alike ($\Delta = 0$), the stayer mean equals the group mean.[13]

Step 2: (B) splits into outcome trends and selection trends. Applying Equation (G.10) to each of the four terms in (B), with ν_{sa} and σ_{sa} indexing groups,

$$\text{(B)} = \underbrace{[(\nu_{11} - \nu_{10}) - (\nu_{01} - \nu_{00})]}_{\text{(B1): parallel trends in outcomes}} + \underbrace{[(\sigma_{11} - \sigma_{10}) - (\sigma_{01} - \sigma_{00})]}_{\text{(B2): parallel trends in selection}}. \tag{G.11}$$

(B1) is the standard parallel-trends condition, stated over whole groups. (B2) is new. Assumption 7 is the statement that (B1) + (B2) $= 0$. The natural sufficient condition is that each is zero separately.

[13] Equivalently, $\sigma = \text{Cov}(Y_i(0), S_i(0))/\rho$, using $E[Y \mid S = 1] = E[YS]/\Pr(S = 1)$ and $E[YS] = E[Y]\Pr(S = 1) + \text{Cov}(Y, S)$, together with $\text{Cov}(Y, S) = \rho(1 - \rho)\Delta$ for binary S. The Δ form is easier to interpret.

Offsetting violations would also satisfy the assumption, but I do not rely on them.

Step 3: necessary conditions for (B2) $= 0$. (B2) is zero when the change in σ across the 2008 cutoff is the same in STEM as in non-STEM:

$$(1-\rho_{11})\Delta_{11} - (1-\rho_{10})\Delta_{10} = (1-\rho_{01})\Delta_{01} - (1-\rho_{00})\Delta_{00}. \tag{G.12}$$

Three observations clarify what this does and does not require.

First, it does not require that stayers be the same kind of worker before and after 2008. σ may change over time within each field, provided it changes by the same amount in both. Only the difference across fields has to cancel, exactly as for outcomes in (B1).

Second, σ has two channels: how many workers would stay, ρ, and how selected the stayers are, Δ. Within field s, the change across the cutoff separates exactly into the two:

$$\sigma_{s1} - \sigma_{s0} = -\Delta_{s0}\,(\rho_{s1} - \rho_{s0}) + (1-\rho_{s1})\,(\Delta_{s1} - \Delta_{s0}). \tag{G.13}$$

(B2) can fail through either channel. It fails through Δ if, for example, the 2008–2009 recession pushed less productive non-STEM graduates out of the United States while leaving the composition of STEM stayers unchanged. Non-STEM stayer quality would then rise through selection alone, making (B2) negative even when (B1) holds.

Third, parallel trends in stay rates is not enough. Suppose untreated stay rates change by the same amount r in both fields, $\rho_{11} - \rho_{10} = \rho_{01} - \rho_{00} = r$, and the stayer–leaver gap is stable over time within each field, $\Delta_{s1} = \Delta_{s0} = \Delta_s$. By Equation (G.13), (B2) $= -(\Delta_1 - \Delta_0)\,r$, which is zero only if staying is equally selective in the two fields, $\Delta_1 = \Delta_0$. If staying in the United States is more selective among STEM graduates than among non-STEM graduates, a common shift in stay rates alone violates (B2).

Step 4: sufficient conditions. Each of the following makes (B2) zero:

(i) Staying is unrelated to untreated outcomes in every group: $\Delta_{sa} = 0$. This is the case in which conditioning on stayers is innocuous; I do not expect it to hold.

(ii) Selection into staying is stable over time within each field: $\rho_{s1} = \rho_{s0}$ and $\Delta_{s1} = \Delta_{s0}$ for $s \in \{0, 1\}$.

(iii) Staying is equally selective in both fields and stable over time ($\Delta_{sa} = \Delta$), and untreated stay rates follow parallel trends.

Condition (iii) is the closest to what the design already assumes elsewhere, and it shows what Assumption 7 adds beyond that assumption.

Relation to the extensive-margin analysis. The extensive-margin difference-in-differences is identified under parallel trends in untreated stay rates, $(\rho_{11} - \rho_{10}) = (\rho_{01} - \rho_{00})$. Two implications follow.

First, the extensive-margin estimate does not show that (B) is nonzero. That estimate measures workers with $S_i(1) = 1$ and $S_i(0) = 0$ in group $(1, 1)$, stayers induced by the extension. Only

$S_i(0)$ appears in (B). Treatment-induced staying enters (A), not (B). Reading the extensive-margin estimate as causal already assumes that untreated stay rates trend in parallel, so the same data cannot also be read as evidence of differential untreated selection.

Second, the extensive-margin assumption does not imply (B2). By Step 3, parallel stay rates deliver (B2) $= 0$ only when staying is equally selective in the two fields and stable over time. Assumption 7 restricts the composition of stayers, which the extensive-margin assumption leaves unrestricted.

Step 5: what can be checked. Because outcomes are defined for jobs held abroad, ρ_{sa}, Δ_{sa}, and therefore σ_{sa} are observed in the three untreated groups, and across graduation cohorts within the pre-period. Stability of Δ within the pre-period, and similar Δ across fields, are informative about condition (iii). An event study of stayer means over pre-period cohorts tests (B1) and (B2) jointly. The comparison with US natives, who face the same field-specific shocks to staying but no OPT, is informative about (B2) to the extent that such shocks affect CtS and native selection similarly. None of these checks the counterfactual for group $(1, 1)$ directly, which is unobserved by construction. Two caveats apply to the observed Δ: leavers' outcomes rely on profiles that may be updated incompletely after departure, and workers without an observed job at t contribute to ρ but not to Δ.

G.4 From δ to an effect on a fixed group

Under Assumption 7, Equation (G.5) reduces to

$$\delta = E[Y_i(1) \mid S_i(1) = 1,\ D_i = 1] - E[Y_i(0) \mid S_i(0) = 1,\ D_i = 1]. \tag{G.14}$$

The two terms average over different workers: those who stay with the extension and those who would stay without it. To relate them, classify treated workers by $(S_i(0), S_i(1))$:

Type	$S_i(0)$	$S_i(1)$	In the estimation sample?
Always-stayer	1	1	Yes
Marginal stayer	0	1	Yes, because of the extension
Pushed out	1	0	No
Never-stayer	0	0	No

Assumption 6 rules out the pushed-out type. The workers with $S_i(1) = 1$ are then the always-stayers and the marginal stayers, and the workers with $S_i(0) = 1$ are exactly the always-stayers. Let

$$\omega \equiv \Pr(S_i(0) = 1 \mid S_i(1) = 1,\ D_i = 1) \tag{G.15}$$

denote the share of treated stayers who would have stayed without the extension, and write:

$$\bar{y}_A^1 = E[Y_i(1) \mid \text{always-stayer},\ D_i = 1], \tag{G.16}$$

$$\bar{y}_A^0 = E[Y_i(0) \mid \text{always-stayer},\ D_i = 1], \tag{G.17}$$

$$\bar{y}_M^1 = E[Y_i(1) \mid \text{marginal},\ D_i = 1]. \tag{G.18}$$

By the law of total expectation over the two types in the sample,

$$E[Y_i(1) \mid S_i(1) = 1,\ D_i = 1] = \omega\, \bar{y}_A^1 + (1-\omega)\, \bar{y}_M^1. \tag{G.19}$$

Under Assumption 6, the second term of Equation (G.14) is $\bar{y}_A^0$. Substituting and adding and subtracting $\bar{y}_A^1$,

$$\delta = \omega\, \bar{y}_A^1 + (1-\omega)\, \bar{y}_M^1 - \bar{y}_A^0 = (\bar{y}_A^1 - \bar{y}_A^0) + (1-\omega)\,(\bar{y}_M^1 - \bar{y}_A^1)\,, \tag{G.20}$$

which is Equation (14):

$$\delta = \underbrace{E[Y_i(1) - Y_i(0) \mid \text{always-stayer},\ D_i = 1]}_{\text{effect on always-stayers}} \tag{G.21}$$

$$+ \underbrace{(1-\omega)\Big(E[Y_i(1) \mid \text{marginal},\ D_i = 1] - E[Y_i(1) \mid \text{always-stayer},\ D_i = 1]\Big)}_{\text{composition}}. \tag{G.22}$$

The first term is an average effect on a fixed group. The second compares the two types at the outcomes they hold under the extension, the US jobs they hold, so it never uses the US outcome that marginal stayers would have had without the extension. It is zero only if marginal stayers hold jobs of the same quality as always-stayers. Without Assumption 6, the workers with $S_i(0) = 1$ would include the pushed-out type, whose treated outcome is never observed, and Equation (G.22) would acquire a third term of unknown sign.

A numerical illustration. Suppose 80 percent of STEM post-period stayers would have stayed anyway ($\omega = 0.8$), always-stayers score $\bar{y}_A^0 = 9$ without the extension and $\bar{y}_A^1 = 10$ with it, and marginal stayers score $\bar{y}_M^1 = 6$. The effect on always-stayers is 1, but $\delta = 0.8(10) + 0.2(6) - 9 = 0.2$, and Equation (G.22) gives $1 + 0.2(6 - 10) = 0.2$. When marginal stayers hold lower-quality jobs than always-stayers, a positive δ understates the effect on always-stayers.

G.5 Identifying ω and bounding the effect on always-stayers

Because $\{S_i(0) = 1\} \subseteq \{S_i(1) = 1\}$ under Assumption 6,

$$\omega = \frac{\Pr(S_i(0) = 1 \mid D_i = 1)}{\Pr(S_i(1) = 1 \mid D_i = 1)}. \tag{G.23}$$

The denominator is the observed stayer share in group $(1, 1)$. The numerator is identified by a difference-in-differences on stayer shares under parallel trends in untreated stay rates, the same assumption that identifies the extensive-margin estimate. Given ω, trimming the upper and lower $1-\omega$ tails of the treated stayers' outcome distribution bounds $\bar{y}_A^1$ in the spirit of Lee (2009). Since $\bar{y}_A^0 = m_{11}^0$ is identified under Assumption 7, this bounds the effect on always-stayers. I leave these bounds to future drafts.

G.6 Retention estimates

Table G2 reports the retention difference-in-differences that enter the bound on composition in Section 7.2, for each degree and post-period window. Panel A is the Kaplan–Meier estimate described in Section 4.3; its Greenwood standard errors treat graduates as independent. Panel B is the linear probability version of Equation (11), which drops graduates whose profile ends before t without an exit, with standard errors clustered by major. Where the two can be compared, the clustered standard errors are up to about one and a half times as large.

Table G2: STEM-OPT Extension and Retention in the United States

	Master's			Doctorate		
Post cohorts	2008–2021	2008–2011	2012–2015	2008–2021	2008–2011	2012–2015
Panel A. Kaplan–Meier, Greenwood standard errors						
$t = 1$	0.15	0.07	0.14	0.01	-0.05	-0.03
	(0.05)	(0.06)	(0.05)	(0.09)	(0.12)	(0.10)
$t = 2$	1.09	0.29	1.33	0.30	-0.28	0.30
	(0.11)	(0.15)	(0.14)	(0.21)	(0.28)	(0.25)
$t = 3$	1.23	0.06	1.45	0.09	-1.06	0.12
	(0.16)	(0.21)	(0.19)	(0.33)	(0.43)	(0.39)
$t = 5$	0.99	-0.09	1.28	0.32	-0.67	0.43
	(0.24)	(0.30)	(0.28)	(0.47)	(0.61)	(0.54)
$t = 8$	2.08	0.95	2.28	0.65	-0.61	0.55
	(0.33)	(0.40)	(0.37)	(0.65)	(0.80)	(0.72)
$t = 10$	1.86	0.61	1.96	0.94	-0.54	0.75
	(0.37)	(0.44)	(0.41)	(0.72)	(0.88)	(0.79)
Panel B. Linear probability, standard errors clustered by major						
$t = 1$	0.17	0.09	0.16	0.01	-0.05	-0.03
	(0.04)	(0.06)	(0.05)	(0.08)	(0.11)	(0.10)
$t = 2$	1.16	0.35	1.42	0.31	-0.29	0.31
	(0.18)	(0.19)	(0.23)	(0.24)	(0.29)	(0.28)
$t = 3$	1.36	0.17	1.64	0.12	-1.09	0.14
	(0.27)	(0.27)	(0.34)	(0.34)	(0.36)	(0.40)
$t = 5$	1.44	0.13	1.62	0.25	-0.67	0.48
	(0.37)	(0.37)	(0.40)	(0.51)	(0.59)	(0.56)
$t = 8$	2.43	1.30	2.79	0.30	-0.52	0.60
	(0.39)	(0.46)	(0.42)	(0.61)	(0.67)	(0.70)
$t = 10$	1.75	1.01	2.38	-0.17	-0.53	0.22
	(0.46)	(0.49)	(0.53)	(0.76)	(0.82)	(0.89)
Graduates	479,668	115,319	163,398	123,678	32,460	44,113
Majors	552	522	526	409	362	390

Notes: Difference-in-differences, STEM versus non-STEM CtS graduates whose first job is in the United States, post- versus pre-2008 graduation cohorts (2003–2006; 2007 excluded), in the share that has not left the United States by t, in percentage points. Doctorate pools the PhD and non-licensure doctorates. Panel A: Kaplan–Meier estimates; a graduate whose profile ends counts as at risk until the last year observed. Panel B: linear probability version of Equation (11), which drops graduates whose profile ends before t without an exit. Graduates: graduates in the four groups at $t = 0$. Majors: terminal-degree majors at $t = 1$. At $t = 10$, the 2012–2015 window contains only the 2012–2014 cohorts.

H School Selectivity Tier Construction

The analysis assigns each degree-granting institution to one of nine selectivity tiers (1 = least selective, 9 = most selective) drawn from two sources joined by institution type. U.S. institutions are tiered using the admissions-based index constructed from undergraduate admission selectivity data maintained by the Integrated Postsecondary Education Data System (IPEDS). Non-U.S. institutions are tiered by a nearest-U.S.-comparator procedure that maps onto the domestic selectivity tiers using the QS World University Rankings (). In the panel, the two sources are merged via a conditional prioritization on institutional identifier, with a source indicator column recording which authority assigned each observation's tier.

U.S. Tier: Admissions Index The domestic tier is drawn from a composite admissions index constructed from institutional undergraduate admission characteristics—acceptance rate and SAT/ACT score distributions—for all U.S. degree-granting institutions. Schools are sorted on the composite and divided into nine tiers, with tier 9 corresponding to the most selective institutions. The index is keyed by a standard institutional identifier and joined to the education records through a school-name crosswalk maintained separately.

The QS rankings are deliberately used only to assign tiers for non-U.S. schools. Empirically, all 204 ranked U.S. institutions in QS ranking already carry an admissions-based tier, so global rankings would add no domestic coverage. More importantly, global rankings exhibit a systematic research-prestige bias that demotes highly selective liberal arts colleges relative to their true admissions standing (e.g., Swarthmore College, Williams College, and Lehigh University each fall two to four tiers lower on global rankings than on the admissions index).

Ranking Crosswalk and Bridge Sample Extending the tier to international schools first requires identifying which global ranking entries correspond to which domestic institutions, so that the U.S. bridge sample can be constructed. This crosswalk is built using fuzzy name matching against the full institutional characteristics directory (≈6,500 institutions). A match is accepted automatically only under high confidence: exact name agreement, a high similarity ratio with a margin over the runner-up, or agreement between the fuzzy match and an independent institutional data chain.

Ambiguous cases are exported for manual review; 41 institutions required hand curation (verified by state and city), including several flagship universities whose common names the matcher could not surface unambiguously (e.g., University of Virginia, University of Washington—Seattle, and Ohio State University). University system aggregate entries with no single identifier are excluded. Of 228 U.S. entries in the ranking file, 227 are resolved to a distinct identifier, yielding 204 bridge institutions that carry both a global average rank and a domestic admissions tier.

International Tier: Nearest-U.S.-Comparator For each non-U.S. institution in the crosswalk, the selectivity tier is assigned by identifying the U.S. bridge school whose global average rank is closest in absolute distance and inheriting that school's admissions tier:

$$\hat{\tau}_i = \tau_{j^*(i)}, \quad j^*(i) = \arg\min_{j \in \mathcal{B}} \left| R_i - R_j \right|, \tag{H.1}$$

where R_i is the global average rank of international school i, $\mathcal{B}$ is the set of 204 U.S. bridge schools, and τ_j is the domestic admissions tier of bridge school j. The matching is implemented using a nearest-neighbor matching strategy on the sorted average rank.

The procedure yields tiers for 3,964 non-U.S. institutions across 116 countries. Sanity checks on named institutions are consistent with expectations: Oxford, Cambridge, ETH Zürich, and the National University of Singapore map to tier 9; the University of Toronto, Tsinghua University, University of Tokyo, and the University of Melbourne to tier 8; IIT Delhi to tier 7.

Panel Integration and Country Assignment The panel joins domestic and international tiers by institutional identifier via a prioritization of the domestic tier over the international tier. Because the international file contains only non-U.S. institutions, no school can be claimed by both sources.

Whether a worker's degree is domestic or foreign—a key stratum in the analysis—is determined by a school-level country field. The resolver applies a four-step precedence rule:

1. If the institution has a domestic institutional identifier, classify as U.S.
2. If it appears in the global ranking crosswalk with a non-U.S. country, classify as non-U.S.
3. Otherwise, take a modal vote over all raw school location records for that school identifier (≈590 million observations), accepting the modal value if it clears a minimum purity threshold.
4. If the modal country lies outside North America, classify as non-U.S.

Schools that remain unresolved after all four steps are excluded from stratum-specific analyses. Validation against the two authoritative sources yields a 100 percent agreement rate for the domestic subsample and 95.7 percent for the global ranking non-U.S. subsample.

Limitations Several caveats apply to the international tier. First, global rankings score only approximately 1,500 institutions worldwide; unranked foreign schools receive no tier assignment. Second, the nearest-U.S.-comparator procedure assumes that global rank is comparable across countries as a signal of institutional reputation—an assumption that holds approximately for research-intensive universities but may be less accurate in national systems where ranking coverage is sparse or systematically biased. Third, the name match is the sole authority for non-U.S. schools, so match errors propagate directly into tier assignments. Fourth, the modal-country vote is approximately 96 percent reliable against the authoritative sources; the residual uncertainty is concentrated among small or obscure institutions with few records.

H.1 From Admissions Scores to the Continuous Education Signal.

This subsection details the reputation score underlying the tiered signal used in Section 3.1. To capture school reputation, I assign each U.S. degree-granting institution a composite score based on the following specification:

$$\text{score} = 0.65 \times \text{Avg. Std. Midpoint SAT/ACT Scores} + 0.35 \times \text{Std. Rejection Rate}. \quad \text{(H.2)}$$

This score serves as a composite measure of institutional selectivity, proxying the cost of attaining an undergraduate degree at a given institution. Due to program and degree availability, these scores are interpreted relative to their respective degree levels. Nevertheless, estimating signal strength using just undergraduate admission statistics directly requires a strong *Independence of Irrelevant Alternatives* (IIA) assumption across degrees. Because the scores are derived from undergraduate admission data, institutions providing other degree levels are considered subsets of the universe used to establish the score's induced order. Under IIA, this ordering can be applied within each degree level to compute a standardized reputation score for worker i's education attainment a_i, denoted as $\tilde{s}_{i,a} = \frac{score_i - \bar{score}_a}{sd(score_a)}$. The latent education signal is then constructed as follows in Equation (H.3):

$$\hat{s}_i = \hat{\beta}_0 + \sum_{a=\text{Master}}^{\text{PhD}} \hat{\delta}_a \cdot \mathbb{1}\{Deg_i = a\} + \sum_{a=\text{Bachelor}}^{\text{PhD}} \hat{\beta}_a \cdot \mathbb{1}\{Deg_i = a\} \cdot \tilde{s}_{i,a}. \tag{H.3}$$

Recognizing that the IIA assumption may be difficult to satisfy when comparing institutions in specialized fields, I implement a tiered classification of institutions instead of using the raw scores. This discretization ensures that the induced order is robust to minor fluctuations in selectivity and mitigates the potential for IIA violations when the choice set includes institutions with highly similar signaling values. While the constructed $\hat{s}$ remains a coarse measure, it more flexibly accounts for the heterogeneous cost structure $c(e)$ by allowing different degree levels to yield equivalent education signals via the coefficients β_a, thereby avoiding the biases inherent in a strictly monotonic ordering of degree-institution pairs. "'

I Specification Choices and Comparisons

Controls in the entry regressions. The controls in Columns 2 and 3 of Equation (4) are fixed before the worker enters the labor market. I do not condition on characteristics of the occupation the worker takes, because both mismatch measures are defined by that occupation. Major generality stays because it describes the major, not the job. I do not separately classify under-education: under-educated workers typically substitute experience and on-the-job training for the credential and move between firms within an occupation (); Appendix E.8 reviews this evidence.

Origin-region fixed effects. The CtW specifications do not include fixed effects for the world region of the foreign school. Every FUS worker was educated in North America, so such fixed effects would identify the CtW gap mainly from CtW workers educated there, nearly all of them Canadian, and from those whose school country is unknown, not from the CtW population as a whole.

Cap-exempt employers. The indicator CE_j covers higher-education institutions only, so some cap-exempt employers, such as nonprofit research organizations, are counted as cap-subject. A postdoctoral appointment requires a doctorate, so its holder is never over-educated, which matters because universities account for 36 to 40 percent of doctorate first jobs. Column 4 of Equation (4) includes $P_i \times CE_j$, but ψ^m is common to all degrees, so it imposes the same cap-exempt shift on master's and doctorate holders. Re-estimating Column 3 separately by employer type (Table 10) lets every coefficient differ.

The CtW entry window. The CtW first US job is kept when it starts within five years of the degree, the window used for CtS and FUS workers. A longer window would add CtW workers who arrive later, but the panel ends in 2026, so a fifteen-year window is observed in full only for the 2003–2011 graduation cohorts, and the CtW sample would then differ by cohort for a mechanical reason: among CtW workers of the 2021 cohort with a first US job within fifteen years, 36 percent held an earlier post-degree job, against 56 to 58 percent in the 2003–2011 cohorts. With a fifteen-year window, the Column 3 over-education gaps of CtW master's, doctorate, and professional-degree holders are 10.3, 8.6, and 24.5 percentage points, against 9.1, 2.3, and 20.9 in the main sample, and the horizontal-mismatch gaps are 4.7, 3.8, and 15.9, against 4.2, −1.0, and 13.3. The ranking across degrees is the same, but the doctorate gap is then close to the master's gap: the CtW workers whom the longer window adds, those who reach the US six to fifteen years after the degree, raise the doctorate over-education gap by 6.3 points and the master's gap by 1.2.

Regressions and the decomposition for CtW master's holders. The regressions and the decomposition in Section 5.3 differ in size but not in sign. In the first column of Table 14, the occupations of CtW master's holders pay 1.5 log points more than those of FUS master's holders, and the raw gap in the decomposition is 5.8 log points. Both net out market-entry-cohort effects, the regressions through fixed effects and the decomposition by residualizing the occupation's log median wage on every job-year of the panel. The first column, however, also holds non-US labor-market experience and transfer status fixed, whereas the decomposition counts both as worker characteristics that

explain the gap. Together they account for 2.6 of the 4.3-point difference, and the rest appears to reflect the different samples and specifications.

The first year in the dynamic decompositions. The decompositions in Equation (8) include an indicator for holding the first job. At $t = 0$ it accounts for -11.4, -11.9, and -8.2 log points of the unexplained gap for CtS master's, doctorate, and professional-degree holders and -7.7, $+11.6$, and -16.6 for their CtW counterparts; from $t = 1$ on, its contribution is within one log point of zero in every group. The gap at $t = 0$ therefore reflects when workers start, not which occupations they hold. The tables report $t = 0$ for completeness.

Entry gaps and stayer paths. The comparison of entry gaps with the first-year gaps of stayers in Section 6 is approximate for two reasons. First, each decomposition prices covariates at the coefficients of its own FUS sample: FUS ten-year stayers in year t for the paths, and all FUS entrants at their first US job for the entry gaps. FUS stayers are themselves selected, so a difference between the two unexplained gaps combines selection among immigrants with a change in the benchmark. Second, the entry sample takes each worker's first US job within five years of the degree, not the job held in a fixed year. Neither decomposition holds the job or employer fixed, so neither reproduces the final gaps in Figures 5 and 6.

Comparing the CtW and CtS paths. Each pathway's unexplained gap is priced at the coefficients of its own FUS sample, and the CtW decompositions hold fixed additional covariates, so the distance between the paths is not a single estimate and has no standard error. The clocks also differ: t counts years since the degree for CtS workers and years since the first US labor-market year for CtW workers, who may arrive with experience abroad. The comparison aligns time in the US labor market, the relevant clock for employer learning, but not career stage. A worker whose occupation pays more than comparable FUS workers' has no reason to change it, which is consistent with the persistent CtS master's premium. Because that premium is not a friction, I do not interpret the distance between the paths as what a foreign credential adds.

The H-1B wage benchmark. The benchmark in panel (a) of Figure 7 is the occupation's incumbent wage distribution, not that of comparable native new hires, and entrants of any origin are paid below the occupation median early in their careers. A direct test of within-occupation pay parity would compare sponsored offers with the realized wages of native entrants in the same occupation with the same experience; the panel imputes occupational medians, not individual wages, and cannot support it. The attestation also does not police the employer's choice of wage level, which is the margin panel (b) addresses.

J Welfare Implications and Limitations

The friction d distorts the allocation of $\bar{\theta}$ immigrants away from the contract they would obtain absent the friction, making explicit the welfare loss to the market. The logic is that of Hsieh et al. (2019), who attribute between 20 and 40 percent of US growth in aggregate market output per person from 1960 to 2010 to the decline in frictions that women and Black men faced in the labor market and in accumulating human capital. Without d, a type-$\bar{\theta}$ immigrant faces the natives' equilibrium schedule and, like a native of the same type, chooses $\bar{e}$ and is employed at $W(\bar{e})$. I use this contract as the benchmark rather than the full-information choice $e^*(\bar{\theta}) \in \underset{e}{argmax}\ F(\bar{\theta}, e) - C(\bar{\theta}, e)$: natives already over-invest in education to separate, and that distortion is not caused by d. Introducing d removes the jobs on the equilibrium wage schedule between $W(\tilde{e})$ and $W(\bar{e})$ from the choice set of a type-$\bar{\theta}$ immigrant, who keeps the credential $\bar{e}$ and is placed at $\tilde{e}$ (Figure 1). Because their education is unchanged, the distortion falls entirely on the job. The per-worker welfare loss, the induced reduction in output, is

$$L(\bar{\theta}) = F\left(\bar{\theta}, \bar{e}\right) - \left[F\left(\bar{\theta}, \bar{e}\right) - d\right] = d = W(\bar{e}) - W(\tilde{e}). \tag{J.1}$$

The firm loses output d on the immigrant, and the immigrant bears it as a wage sacrifice $W(\bar{e}) - W(\tilde{e})$ in an occupation for which less education is typical. Empirically, the loss is concentrated where the over-education gap is largest: among master's holders in both pathways and foreign-trained professional-degree holders (Section 5). Because the wage schedule is convex, the same d moves a doctorate holder a shorter distance down the schedule: the output lost per worker is the same at every degree, but the mismatch it produces is smaller at the top.

The magnitude, and even the interpretation, of $L(\bar{\theta})$ depends on what d represents. If d is a real resource cost of sponsorship, such as legal and processing costs, then not hiring the immigrant at $W(\bar{e})$ frees up resources the firm would have spent, and the resulting allocation is efficient given that cost; L then measures what the sponsorship system costs, which policy can reduce. If d is a transfer, such as filing fees paid to the government, it acts as a tax and L is deadweight loss. If d is taste-based discrimination, the firm saves nothing by displacing the worker and the match surplus is simply destroyed: pure allocative deadweight with no offsetting benefit. If d instead reflects statistical discrimination, the distortion is transient: as employer-learning erodes d, the worker's allocation converges toward $\bar{e}$, so the present value of L is smaller and front-loaded. The dynamics of the adjusted gap documented in Section 6 therefore bear on the size of the welfare loss, not merely on its source. A persistent gap implies a durable allocative cost, whereas an attenuating one implies a loss that employer learning partly undoes.

The STEM-OPT extension offers the closest evidence on how the loss responds when d falls. The extension lowered the cost of hiring a STEM graduate without changing who the graduate is, and Section 7 estimates two responses that map onto Equation (J.1). The extensive-margin estimate δ_e measures how many graduates, who would otherwise have taken a first job abroad instead, take one in the US. The intensive-margin estimates δ^k_{et} measure how the jobs held by those who remain change, which bears on the first term of $L(\bar{\theta})$. Because Equation (14) interprets δ^k_{et} as an effect on the population that remains rather than on a fixed set of workers, these estimates provide evidence on the direction in which the allocation moves when d falls, not a measure of

$L(\bar{\theta})$ for any one worker. For CtS STEM master's graduates of the 2008–2011 cohorts, δ_e is 3.7 percentage points, and five years after graduation those who stayed were 7.6 percentage points less likely to work outside their field and held occupations paying 8 log points more at the median, while over-education changed by 0.5 points. The allocation therefore appears to have moved toward better-paid jobs that require the same education, not up the schedule from $W(\tilde{e})$ toward $W(\bar{e})$, and because Section 7.2 does not bound the change in who stays for the wage outcome, the 8 log points are not an estimate of the fall in $L(\bar{\theta})$.

The aggregate welfare cost scales the per-worker loss by the mass of displaced high-type immigrants,

$$\mathcal{L} = \int_{\Theta} s(\theta)\, L(\theta)\, dG(\theta), \tag{J.2}$$

where $s(\theta)$ is the share of type θ displaced off the natives' equilibrium schedule and G is the type distribution. Evaluating $\mathcal{L}$ requires the displaced share $s(\theta)$ and a $d = 0$ counterfactual, which in turn rest on an estimate of d, which the adjusted gaps in Section 6 do not provide on their own, and the stayer restriction discussed in Section 4.2. Equation (J.2) is a partial-equilibrium object: it counts the allocative loss borne by displaced immigrants, holding fixed the native equilibrium schedule, firm entry, and output prices. A full accounting would net it against the general-equilibrium consequences of high-skilled inflows, which Khanna and Morales (2024) find are substantial and mixed in sign: the same inflows that raise firm profits and lower output prices depress the wages and employment of competing natives. Removing d would move high-type immigrants up into occupations where they compete more directly with natives, so an efficiency gain in the sense of Equation (J.1) need not be a welfare gain once those margins are counted. I do not estimate these effects on natives. The STEM-OPT design offers one direct check for a single policy, by comparing the job outcomes of STEM and non-STEM natives across the 2008 rule, which I leave for future work.

Calibrated models of vertical mismatch put its output cost at 0.26 percent of US output (Garibaldi et al., 2025a) and at 3 to 4 percent of output on average across 17 OECD economies (Garibaldi et al., 2025b); both are calibrated to the workforce as a whole. I therefore treat Equation (J.2) as the framework that the estimated friction populates, and reserve its full quantification, together with the associated immigration-policy counterfactuals, for future work.

Limitations. Several features of the data limit what the estimates can show. First, every outcome is an occupation- or employer-level benchmark (the OEWS median wage, SOC PageRank, firm value, and the two mismatch measures) rather than the worker's realized wage, so the estimates describe where immigrants are placed, not what they are paid within a placement. Second, the dynamic estimates in Section 6 and the intensive-margin estimates in Section 7 condition on staying in the United States, so both describe the population that stays rather than all entrants. Third, departures are observed only if workers update their profiles after leaving, so exits are measured with error. Fourth, the STEM indicator marks majors on the 2008 list, so majors added later are coded non-STEM even for cohorts that became eligible, which biases δ_e toward zero.

K Miscellaneous Tables

Link to online appendix of canonical major groups/