The Missing Margin:

Internal Mobility and the Measurement of Monopsony Power

Greg C. Wright

UC Merced

March 12, 2026

Abstract

Standard monopsony estimates rely on external separations, but firms also suppress internal mobility—promotions and raises—when workers cannot credibly threaten to leave. Using LinkedIn career histories matched to H-1B visa records, I show that visa constraints at PERM filing reduce internal mobility by 58–65 percent alongside the well-documented external decline. A two-channel decomposition reveals that 29–37 percent of the total mobility decline flows through the internal channel, implying that external-only estimates understate welfare costs by 33–39 percent. The co-decline is general across occupation, education, and wage level.

Keywords: Immigration; Labor Mobility; H-1B Visas; Internal Labor Markets; Monopsony

JEL: J61; J62; F22; J42

1 Introduction

A growing literature measures monopsony power from the rate at which workers separate from firms (Manning, 2003, 2021; Bassier et al., 2022). In practice, these estimates rely almost exclusively on *external* separations—workers leaving for other employers. But workers also transition *within* firms: they receive promotions, raises, and role changes. If monopsony suppresses both margins simultaneously, then external-only estimates understate the total welfare cost of monopsony power.

This paper shows that firm monopsony power does suppress both margins, and that the understatement from ignoring internal mobility is quantitatively important. The mechanism is straightforward: when workers cannot credibly threaten to leave, firms have less incentive to offer career advancement. The retention motive that drives internal offers weakens, and internal mobility declines alongside external mobility. In effect, internal mobility can be thought of as a firm amenity, but unlike other amenities it is both *worker-specific* and *continuously adjustable*, in contrast to, say, pension plans or health benefits. Moreover, the effects of internal mobility are *large in magnitude*: in these data, internal moves are associated with mean wage gains of 17–25 log points (Table 1), comparable to external moves.

I develop a three-option discrete choice framework in which the value of H-1B workers' external options declines exogenously and firms set internal offers endogenously. I test the framework's predictions using LinkedIn career histories matched to administrative visa records for H-1B workers. The H-1B visa system creates sharp, legally-enforced constraints on external mobility that vary over time as workers progress through the green card queue. When visa constraints tighten at the Program Electronic Review Management (PERM) filing date, external mobility drops directly (workers cannot easily change employers) and internal mobility drops endogenously (firms reduce promotions and raises because the retention motive weakens). By tracking both internal and external transitions before and after these constraints change, I can quantify how much of the total lock-in cost flows through each channel and how much standard external-only estimates miss.

The main empirical finding confirms that monopsony operates on both margins. Within-person estimates show that external mobility drops sharply at PERM for both Indian and Rest of World (ROW) H-1B workers relative to natives. Internal mobility *co-declines*: two-period within-person tests show internal hazard rates fall by 58–65 percent, while natives show no comparable change. This co-decline is consistent with the model's prediction that firms reduce internal offers when workers' outside options are suppressed. Post-AC21, when the worker's green card application becomes portable and external options reappear, both margins adjust—dynamics consistent with the model's predictions about how mobility responds to changes in constraint intensity.

A two-channel lock-in decomposition quantifies the cost of ignoring the internal margin. In the within-person balanced panel, 29–37 percent of the total mobility decline at PERM flows through the internal channel, implying that standard external-only estimates understate the welfare cost of lock-in by 33–39 percent.[1] Because the PERM effect measures the *incremental* lock-in from the green card process (H-1B workers already face moderate switching frictions before PERM), the total monopsony cost of the visa program is likely larger than the PERM-specific estimates suggest. More broadly, standard monopsony markdowns are estimated from separation elasticities, which capture only the external channel. But visa constraints also suppress internal advancement—a second welfare cost that separation-based methods miss entirely. External-only estimates therefore understate the total welfare cost of monopsony even if the markdown itself is correctly measured. The steady-state correction may be modest, but for discrete lock-in events—visa constraints, non-competes, occupational licensing with employer-specific investments—the omission is first-order.

The two-margin pattern holds broadly. Heterogeneity analysis across occupation (CS/Math vs Non-CS), education (advanced degree vs not), wage tercile, and institution type (Indian university vs not) shows that the internal decline persists in all subgroups within-person. The

---

[1] The balanced panel comprises 8–9 percent of identified H-1B workers—those observed in both the pre-PERM and lock-in periods. Population-weighted bounds, which account for the 91 percent of workers not observed in both periods, place φ_I between 18 and 37 percent under plausible assumptions about excluded workers' internal mobility response. I quantify these bounds in Section 7.

mechanism is general: it operates through the legal visa constraint, not through subgroup-specific labor market conditions.

This paper contributes to several literatures (Section 2). It extends the monopsony framework (Manning, 2003; Card et al., 2018; Bassier et al., 2022) by showing that standard estimates based on external separations alone miss a second margin—internal mobility—through which monopsony power operates. It provides the first evidence on how visa constraints affect within-firm career dynamics (Bound et al., 2017; Peri et al., 2015; Wang, 2021; Kroft et al., 2026), and quantifies the bias in lock-in cost estimates that ignore the internal channel.

The paper proceeds as follows. Section 2 reviews the related literature. Section 3 describes the H-1B visa system and the key milestones that generate variation in constraint intensity. Section 4 presents a three-option discrete choice framework that generates the co-decline prediction, derives the welfare decomposition, and establishes that ignoring the internal margin understates lock-in costs. Section 5 describes the data construction and H-1B identification algorithm. Section 6 presents the empirical strategy. Section 7 reports the main mobility results, quantifies the two-channel decomposition, and considers alternative explanations. Section 8 implements the welfare decomposition nonparametrically. Section 9 tests whether the monopsony pattern holds across worker subgroups. Section 10 concludes.

2   Related Literature

The modern monopsony literature, following Manning (2003), estimates the extent of firm wage-setting power from the labor supply elasticity facing the firm. While some approaches use wage variation in matched employer-employee data (Card et al., 2018; Lamadon et al., 2022) or labor market concentration (Azar et al., 2020), a leading approach measures the elasticity from *external* separation rates (Manning, 2021; Bassier et al., 2022). Naidu et al. (2016) provide direct evidence of monopsony power over migrant workers, exploiting a UAE

reform that relaxed employer-transition restrictions and finding that firms paid incumbents roughly 51 percent of marginal product pre-reform. This paper shows that separation-based estimates miss a second channel through which monopsony power operates: firms also suppress *internal* career advancement when workers cannot credibly threaten to leave. Accounting for this internal margin increases the estimated welfare cost of lock-in by 33–39 percent.

The idea that firms adjust on non-wage margins is not new. Sorkin (2018) ranks firms using revealed preference from worker flows, showing that non-wage amenities are an important component of job quality, and Madrian (1994) documents "job lock" from employer-provided health insurance. But the margin identified here—internal career advancement—differs from these in two respects: it is worker-specific (the firm targets individual workers with tailored offers) and continuously adjustable (the firm can respond in real time to changes in a worker's outside options). I show that this margin responds endogenously to the same constraints that suppress external mobility.

A growing literature studies how visa constraints affect labor market outcomes. Bound et al. (2017) study the economic impact of the H-1B program, Peri et al. (2015) find that H-1B workers complement native STEM workers, and Doran et al. (2022) use visa lotteries to estimate firm-level effects. On mobility specifically, Depew et al. (2017) find that visa constraints significantly reduce inter-firm job switching, Hunt and Xie (2019) find that green card applications reduce job mobility by approximately 20 percent, and Wang (2021) shows that this job lock reduces earnings by 4.7 percent, with mobility spiking upon permanent residency receipt. Kroft et al. (2026) study Canada's closed visa system using administrative employer-employee data and find that gaining permanent residency increases job switching by 21.7 percentage points and earnings by 5.7 percent, with workers sorting into higher-paying firms. This literature has focused on wages, employment, and external mobility; I contribute by providing the first evidence on how visa constraints affect *within-firm* career dynamics, which the monopsony framework predicts should adjust endogenously to the constraint.

The empirical setting connects to a broader literature on mobility constraints. Starr et

al. (2021) document the prevalence of non-compete agreements, Marx et al. (2009) provide causal evidence that non-compete enforcement reduces inventor mobility, and Johnson and Kleiner (2012) show that occupational licensing reduces interstate migration. Non-competes and H-1B visa constraints generate employer-specific lock-in; licensing operates differently (constraining geographic or occupational rather than employer-specific mobility), but all three suppress the external option. The H-1B setting has a key advantage for identification: visa constraints are legally imposed and time-varying, providing sharp within-worker variation in constraint intensity. The contribution here is showing that mobility constraints suppress *both* external and internal margins—a pattern that standard analyses of non-competes and licensing, which focus on external mobility, would also miss.

Finally, this paper reconceptualizes the internal labor market as an endogenous object that responds to the worker's outside options. Baker et al. (1994) document the structure of internal labor markets, and Topel (1991) and Lazear (2009) show that firm-specific human capital raises wages with tenure—a voluntary lock-in that benefits both parties. The lock-in studied here is distinct: it arises from a legal constraint rather than from mutual investment, and it *reduces* internal advancement rather than facilitating it. When visa constraints suppress external mobility, firms reduce internal offers because the retention motive weakens—the opposite of the firm-specific capital prediction.

3 Institutional Background

3.1 The H-1B Visa Program

The H-1B visa authorizes U.S. employment in "specialty occupations," subject to a statutory cap of 65,000 annually plus 20,000 for holders of U.S. advanced degrees. Initial status is granted for up to three years and may be renewed once, for a maximum of six years.

A defining feature of the H-1B is employer-specific sponsorship. The visa is tied to a particular employer, and changing jobs requires a new employer to file a transfer petition. If

H-1B to Green Card Timeline

| H-1B | PERM
Filed | | I-140
Approved | | I-485
Filed | AC21
Portability | | Green
Card |

LOCKED TO EMPLOYER
~12–18 months

CONSTRAINED
ROW: ~2–3 yr | India: 7–10+ yr

PORTABLE

Figure 1: Timeline from H-1B entry to green card receipt, showing phases of employer lock-in. See text for details.

the transfer is denied, the worker may lose lawful status. This creates moderate friction for external mobility even before the green card process begins.

3.2 The Green Card Queue

Many H-1B workers seek permanent residence (a "green card") through employer sponsorship, a process that generates employer lock-in of varying intensity (Figure 1). The process begins with a PERM (Program Electronic Review Management) labor certification, typically filed 1–3 years after initial H-1B employment, which establishes that no qualified U.S. worker is available for the position. Following PERM approval, the employer files an I-140 immigrant petition, which makes the worker's *priority date*—established at the PERM filing date— portable across employers. From PERM filing through I-140 approval, the worker is effectively locked to the sponsoring employer: changing firms means abandoning the pending application and restarting the entire process from scratch.

After I-140 approval, the priority date becomes portable—the worker retains their place in the queue even if they change employers—but the new employer must file a new PERM and I-140, a process that takes 1–2 years and that many employers are unwilling to undertake. The worker cannot file for adjustment of status (I-485) until their priority date becomes "current" according to the monthly Visa Bulletin. Due to per-country limits on employment-based green cards, the wait for a current priority date varies dramatically: workers from most

6

countries ("Rest of World") typically wait 18–24 months from their priority date (i.e., from PERM filing), while Indian nationals face waits of 7–10+ years. Chinese nationals face intermediate backlogs of 3–5 years.

Full portability arrives under Section 106(c) of the American Competitiveness in the 21st Century Act (AC21): workers with an approved I-140 and a pending I-485 application can change employers after 180 days, provided the new job is in a "same or similar" occupational category. At this point the worker can switch employers without restarting the green card process. Green card receipt restores full labor market mobility with no employer tie.

3.3   Identification: Constraint Variation

The green card process generates two sources of variation in constraint intensity that drive the empirical design. First, PERM filing creates a discrete jump in switching costs: before PERM, H-1B workers can transfer to a new employer with moderate friction; after PERM, leaving means forfeiting the green card application. Second, the per-country visa backlog creates differential constraint *duration*: ROW workers typically reach AC21 portability within 2–3 years of PERM, while Indian workers remain constrained for 7–10+ years. This generates a natural dose-response comparison—same constraint mechanism, different duration—that the event study exploits.

4   Conceptual Framework

This section develops a three-option discrete choice framework that formalizes internal mobility as an endogenous margin of firm adjustment. The framework serves three purposes: (i) it generates the testable prediction that external and internal mobility co-decline under monopsony—a prediction that could be falsified if internal mobility rose or was unaffected; (ii) it derives the welfare decomposition (Equation 6) that maps observed hazard rates and wage gains into welfare flows; and (iii) it identifies the conditions under which the co-decline

holds (the markup condition in Equation 3). The framework is not structurally estimated: the paper's main empirical results—the lock-in decomposition φ_I and the welfare-weighted φ_I^W—are nonparametric and do not depend on the logit functional form or on calibrated parameters. The reduced-form approach is deliberate: it ensures that the results are robust to misspecification of the choice model.

The firm's internal offer—promotions, raises, role changes—serves as a retention instrument: a worker-specific, continuously adjustable component of the employment package that the firm optimizes alongside the base wage. Visa constraints affect worker welfare through two channels: a direct external channel (workers cannot easily change firms) and an endogenous internal channel (firms reduce internal offers for workers who cannot credibly threaten to leave).

A worker employed at firm j at wage w_j faces three options each period: stay at w_j, accept an internal move (promotion, raise) at $w_j + g_I$, or move externally at wage $\bar{w}_X$ less switching cost τ_X^v. Utility from each option includes a logit preference shock. The key assumption is $\tau_X^H > \tau_X^N$: visa constraints make external moves costlier for H-1B workers but do not directly affect internal moves. Full derivations are in Appendix A.

The firm sets the internal offer g_I to maximize expected profit, trading off the cost of the raise against the retention benefit (fewer workers leave). The first-order condition yields:

Proposition 1 (Visa Constraints Reduce Internal Offers). *The firm's optimal internal offer g_I^* is decreasing in external switching costs:*

$$\frac{\partial g_I^*}{\partial \tau_X} < 0 \tag{1}$$

When a worker cannot easily leave (high τ_X), the firm has less incentive to offer raises or promotions. The firm's internal offer is a retention tool, and it is less valuable when the worker is already locked in.

The model treats the firm as a unitary actor choosing g_I optimally. In practice, internal

offers emerge from decentralized decisions by managers, promotion committees, and HR departments, many of whom may not directly observe the worker's visa status. The mechanism therefore requires that the firm's aggregate compensation and promotion policies respond to retention incentives—not that individual managers consciously condition on visa status. This is plausible in the large firms that employ most H-1B workers (tech, consulting, finance), where centralized compensation bands, calibrated performance reviews, and retention budgets translate firm-level incentives into individual-level outcomes.

4.1 Both Margins Constrained

This delivers the key empirical prediction:

$$\frac{\partial h_X}{\partial \tau_X} < 0 \quad \text{(direct: visa blocks external moves)} \tag{2}$$

$$\frac{\partial h_I}{\partial \tau_X} = \underbrace{\frac{\partial h_I}{\partial g_I}}_{>0} \cdot \underbrace{\frac{\partial g_I^*}{\partial \tau_X}}_{<0} + \underbrace{\left.\frac{\partial h_I}{\partial \tau_X}\right|_{g_I}}_{>0} \tag{3}$$

The first term is the endogenous firm response: the firm pulls back on internal offers because the retention motive weakens. The second term is mechanical substitution: holding g_I fixed, making external moves worse redirects some probability mass toward internal moves (a logit IIA property). Under monopsony ($MPL - w_j$ large relative to taste dispersion σ), the firm earns a substantial surplus from the worker and was investing heavily in retention through g_I. When τ_X rises and the external threat diminishes, the firm reduces g_I^* sharply. When the markup is sufficiently large and taste dispersion sufficiently low, the endogenous response dominates mechanical substitution, and internal mobility declines alongside external mobility:

$$\frac{\partial h_I}{\partial \tau_X} \leq 0 \quad \text{(both margins co-decline under monopsony)} \tag{4}$$

The co-decline of both margins is the model's key prediction: when visa constraints tighten at PERM, external mobility drops directly (workers cannot easily leave) and internal

mobility drops endogenously (firms reduce offers because the retention motive weakens). A researcher who measures the lock-in cost from external separations alone captures the first channel but misses the second.[2]

It is worth being precise about what the co-decline rules out. In the model, $\partial h_I / \partial \tau_X$ is the sum of two opposing forces: an endogenous firm response (negative) and mechanical substitution (positive). Three outcomes were possible *ex ante*: (i) internal mobility *rises* at PERM (mechanical substitution dominates—workers redirected from blocked external moves toward internal ones, with the firm's offers unchanged), implying small markups and weak monopsony; (ii) internal mobility is *flat* (the two forces offset), implying moderate markups; or (iii) internal mobility *co-declines* with external (endogenous firm response dominates), implying markups large enough to generate substantial retention investment. Outcome (i) would have been inconsistent with substantial monopsony power in this setting; outcome (ii) would have provided ambiguous evidence. The data deliver outcome (iii), and deliver it consistently across all subgroups and all three promotion definitions (Section 9.2), which rules out both (i) and (ii).

The co-decline requires the firm's markup $(MPL - w_j)$ to be sufficiently large. From the firm's first-order condition (Appendix A), the retention motive term—$(\partial h_X / \partial g_I) \cdot (MPL - w_j)$—scales with the markup. When the markup is small (workers paid close to marginal product), the retention motive is weak, the endogenous response to τ_X is correspondingly small, and mechanical substitution may dominate: internal mobility could *rise* when external mobility is blocked. The co-decline is therefore a test of whether monopsony power is present: observing it implies that the markup is large enough for the endogenous response to dominate. That the co-decline holds across all subgroups in the heterogeneity analysis (Section 9.1)—including low-wage workers, where markups might be expected to be smaller—suggests that the markup condition is satisfied broadly in this setting, though this does not guarantee it would hold in

---

[2]The logit structure imposes IIA, which governs the mechanical substitution term. Relaxing IIA (e.g., nested logit with "move" and "stay" nests) would change the magnitude of mechanical substitution but not the sign or existence of the endogenous firm response, which follows from the firm's FOC regardless of the worker's choice structure.

settings with smaller markups or weaker constraints.[3]

4.2 Welfare

The worker's expected flow utility is the logit inclusive value (McFadden, 1981), which equals the log wage plus an option value reflecting the worker's ability to switch (see Appendix B for the full derivation). An expected flow decomposition yields the welfare flow as wage drift plus the probability-weighted gains from switching:

$$\frac{d\mathcal{W}}{dt} = \frac{d\ln w}{dt} + h_I(t) \cdot \tilde{g}_I(t) + h_X(t) \cdot \tilde{g}_X(t) \tag{5}$$

where $\tilde{g}_d(t)$ is the log wage gain conditional on a type-d transition. Conditioning on survival in the current match:

$$\frac{dW^{EV}}{dt} = \frac{d\ln w}{dt} + S(t)\left[h_I(t) \cdot \tilde{g}_I(t) + h_X(t) \cdot \tilde{g}_X(t)\right] \tag{6}$$

Welfare thus has three components: wage drift, external switching value ($h_X \cdot \tilde{g}_X$), and internal switching value ($h_I \cdot \tilde{g}_I$). Visa constraints reduce welfare through both channels: the direct reduction in h_X and the endogenous reduction in h_I via lower g_I^*.

The welfare gap between H-1B and native workers exceeds the base wage gap because visa constraints reduce the option value of switching on both margins (Appendix C). At PERM, τ_X jumps and welfare drops through two channels: h_X drops directly, and g_I^* drops endogenously.

---

[3]A calibration exercise confirms that the logit model can match observed magnitudes. Setting the external wage gain to 27 percent (observed) and searching over taste dispersion σ and markup $(MPL - w)/w$, the model generates 57.6 percent internal decline alongside 27.4 percent external decline and $\varphi_I = 32.1$ percent—matching the data—at $\sigma = 0.10$ and markup = 100 percent. The co-decline requires markups of at least 75 percent and low taste dispersion ($\sigma \leq 0.12$). These markups are high relative to unconditional estimates in the monopsony literature (15–40 percent; Manning, 2021), but the calibration is conditional on the visa constraint binding—PERM-filing workers are precisely the population where employer-specific lock-in is strongest, so the relevant markup is the *constrained* markdown, not the population average. At the best-fitting parameters, the firm's optimal internal raise falls from 25.5 percent to 14.9 percent at PERM—a 42 percent reduction in internal offers when the retention motive weakens. The AC21 asymmetry also emerges: at 25 percent τ_X relaxation, external mobility recovers 25 percent of the PERM decline while internal recovers only 17 percent, because the firm's internal offer is convex in the worker's outside option.

At AC21, portability partially restores the external option, but the recovery differs by group: ROW workers recover within 1–2 years, while Indian workers remain suppressed due to the green card backlog.

4.3 Total Mobility and the Lock-in Cost

The standard approach to measuring monopsony power uses *external* separations, ignoring within-firm transitions (Manning, 2003). I now show that this omission is quantitatively important when lock-in suppresses *both* margins simultaneously.

Proposition 2 (Lock-in Cost Decomposition)**.** *Under the model (Proposition 1), visa lock-in reduces both separation rates: $\Delta s_X < 0$ (direct constraint) and $\Delta s_I < 0$ (endogenous firm response). Define the total mobility decline $\Delta s = \Delta s_X + \Delta s_I$ and the internal share of the decline:*

$$\varphi_I \equiv \frac{\Delta s_I}{\Delta s_X + \Delta s_I} \tag{7}$$

Then:

(a) *The mobility understatement factor is exactly $1/(1 - \varphi_I)$: ignoring the internal channel understates the total mobility decline by this factor. The welfare understatement depends additionally on wage gains: define $\varphi_I^W \equiv \Delta(h_I \cdot \tilde{g}_I)/[\Delta(h_X \cdot \tilde{g}_X) + \Delta(h_I \cdot \tilde{g}_I)]$, where $\tilde{g}_d$ is the mean log wage gain on channel-d transitions. In periods where $\tilde{g}_I < \tilde{g}_X$ (the typical case), $\varphi_I^W \leq \varphi_I$, and the welfare understatement factor $1/(1 - \varphi_I^W)$ is a lower bound on the mobility understatement. Empirically, $\varphi_I^W < \varphi_I$ for both H-1B groups in the decomposition.*

(b) *φ_I is directly observable from hazard rates without structural estimation: $\varphi_I = \Delta h_I/(\Delta h_X + \Delta h_I)$, where Δh_d is the within-person change in channel-d hazard rates between pre-PERM and lock-in periods.*

(c) *The decomposition is nonparametric—it does not depend on the logit functional form or*

on calibrated switching costs. It requires only that both margins respond to the same underlying constraint, which the data confirm: both margins change discretely at the same visa milestone, both partially recover at AC21, and neither changes for natives (Section 7).

The intuition is straightforward. When a worker's visa status changes at PERM, two things happen: external mobility drops directly (the worker cannot easily change employers), and internal mobility drops endogenously (the firm reduces offers because the retention motive weakens). A researcher who measures the lock-in cost from external mobility alone captures the first channel but misses the second. The internal share φ_I quantifies exactly how much is missed.

As I show in Section 8, φ_I ranges from 0.29 to 0.37 for H-1B workers, and the welfare-weighted internal share φ_I^W is 0.25–0.28, implying a welfare understatement of 33–39 percent.

Identifying assumptions.   For φ_I to have a causal interpretation, two assumptions are required beyond standard event-study identification. First, *common shock*: both Δh_X and Δh_I must respond to the same underlying constraint (PERM lock-in), not to coincidental shocks that independently affect each margin. The model establishes this theoretically (Proposition 1), and the empirical evidence supports it: both margins change discretely at the same visa milestone, both recover at AC21, and neither changes for natives. Second, *no mechanical substitution bias*: the within-person estimator must not conflate the causal internal decline with a mechanical artifact of how transitions are classified. Because external and internal transitions are mutually exclusive within a spell (each spell ends with exactly one type of transition or is censored), there is no accounting identity linking Δh_X and Δh_I—a worker who does not leave externally is not mechanically recorded as an internal transition. The two margins are mutually exclusive event types within the same underlying person-time, so the co-decline reflects two distinct behavioral responses to the same constraint, not an accounting artifact.

Each model object maps directly to an observable data counterpart (Appendix D): hazard rates h_X and h_I are computed from LinkedIn career histories, wage gains $\tilde{g}_X$ and $\tilde{g}_I$ from salary data, and the survival probability $S(t)$ from Kaplan–Meier estimation. The model predicts that external switching value dominates internal ($h_X \gg h_I$ and $\tilde{g}_X > \tilde{g}_I$), which the data confirm: the internal share of switching welfare is only 9–17 percent.

4.4 Heterogeneous Constraint Duration

The per-country green card cap creates differential constraint duration:

$$\tau_X^v(t) = \begin{cases} \tau_X^{pre} & t < 0 \text{ (pre-PERM)} \\ \tau_X^{high} & 0 \le t < T^{*v} \text{ (PERM to green card)} \\ \tau_X^{post} \approx \tau_X^N & t \ge T^{*v} \text{ (post-green card)} \end{cases} \tag{8}$$

where $T^{*,ROW} \approx$ 2–3 years and $T^{*,India} \approx$ 10–15+ years. The cumulative welfare loss $\Delta W^v = \int_0^{T^{*v}} [\mathcal{W}^N(t) - \mathcal{W}^v(t)]\, dt$ is approximately proportional to T^{*v}, implying that the welfare cost for Indian workers is 4–7 times larger than for ROW workers—driven entirely by the per-country cap, not by differences in worker quality or preferences.

5 Data

5.1 LinkedIn Career Histories

The primary data source consists of employment histories from LinkedIn profiles, covering approximately 5 million U.S. workers observed over 2005–2023, yielding 9.1 million person-year records. Berry et al. (2024) validate LinkedIn profile data against ACS and EEO-1 benchmarks, finding that LinkedIn provides reasonably reliable measures of workforce composition and employment outcomes at individual companies. Each profile contains a sequence of job episodes with employer name, job title, start date, end date, and location. I restrict to

workers with at least three observed jobs, which provides cleaner mobility estimates and sufficient variation for within-person analysis.

I define a *transition* as a change between consecutive job episodes. Transitions are classified as:

- **External (h_X):** Employer changes between episodes—the model's external mobility margin.

- **Internal (h_I):** Consecutive episodes at the same employer—the model's internal mobility margin, reflecting the firm's endogenous offer g_I.[4]

First jobs are excluded from transition analysis, as they cannot represent external transitions (there is no prior employer to leave).

Measurement of internal transitions. LinkedIn-reported title changes are an imperfect proxy for the model's internal mobility margin h_I. Several limitations deserve note. First, profile updates are self-reported and may lag actual transitions, introducing timing noise. Second, title changes can reflect corporate restructurings or cosmetic profile updates rather than genuine promotions with wage gains (g_I in the model). Third, many promotions and raises occur without title changes, so measured h_I captures only the visible subset of internal mobility. Fourth, H-1B workers may update profiles at different frequencies than natives (e.g., for visa documentation), potentially creating differential measurement error. These concerns attenuate measured internal mobility effects toward zero, making it harder to detect the model's predictions in either direction. The implications of measurement error differ for the two claims in this paper:

---

[4]Internal transitions are validated as genuine career advancement: 96 percent involve a title change, and 67 percent are accompanied by positive wage growth. Median wage growth on internal moves (+11.8 percent) exceeds median wage growth on external moves (+8.9 percent), confirming that measured internal moves represent promotions and raises rather than noise in company classification. (The median ranking reverses for means—Table 1 reports mean $\tilde{g}_X > \tilde{g}_I$—because external moves have a larger right tail.) These patterns are similar across India (median internal wage growth +9.4 percent), ROW (+8.8 percent), and natives (+16.1 percent).

1. **Sign of the internal channel:** If internal mobility were unaffected by visa constraints, the measured Δh_I should be zero (or positive, if mechanical substitution dominated). Attenuation bias works *for* this test: misclassification error in h_I (under-detection of true internal transitions) biases the measured change toward zero, making it easier to miss a positive effect if one exists. Finding a statistically significant *negative* effect (with bootstrap 95 percent CIs entirely below zero) is therefore conservative—measurement error cannot generate the observed negative sign.

2. **Magnitude claim (large internal decline):** The two-period within-person tests show that internal mobility declines by 58–65 percent for H-1B workers, accounting for 29–37 percent of the total mobility decline (φ_I). Here, attenuation bias works *against* the paper: the true internal decline is at least as large as the measured decline, and the true φ_I is at least as large as the estimated φ_I. Measurement error in h_I therefore makes the estimated welfare understatement a lower bound on the true understatement from ignoring the internal channel.

In both cases, the direction of measurement error is conservative for the paper's conclusions.

Profile updating and the AC21 test. A distinct concern is that the measured internal decline reflects changes in LinkedIn *reporting behavior* rather than genuine mobility suppression. If locked-in workers update their profiles less frequently—because they are not job-searching and have weaker incentives to maintain a current profile—then the observed decline in title changes could be a reporting artifact rather than a real reduction in internal offers. The AC21 portability threshold provides a sharp test against this interpretation. At AC21, workers gain statutory permission to change employers, which restores the firm's competitive incentive to offer internal advancement—but AC21 does not change workers' incentives to update LinkedIn profiles. If the internal decline were driven by reduced profile maintenance, there is no reason updating behavior would recover at a legal portability threshold that affects employer incentives but not reporting incentives. Yet both external and internal mobility

recover at AC21 (Figure 3), consistent with the monopsony mechanism and inconsistent with a reporting artifact. A direct test reinforces this: total profile activity—measured as all job-episode entries per person-year, regardless of whether the associated spell is completed or censored—does not decline at PERM—if anything, it rises slightly for H-1B workers relative to natives—while internal mobility measured from completed spells declines sharply. If the internal decline were a reporting artifact, total profile activity should decline in parallel; instead, the divergence between rising total activity and falling internal mobility confirms that the measured internal decline reflects genuine suppression of within-firm career advancement, not reduced profile maintenance. The native comparison further reinforces this conclusion: if profile updating varied systematically with career stage, natives should show a parallel decline at their career-stage-matched pseudo-PERM—but native internal mobility is flat throughout.

Sample selection and external validity. LinkedIn users are not a random sample of the workforce: they skew toward white-collar, college-educated, urban professionals. Berry et al. (2024) find that LinkedIn overrepresents workers at large firms and in professional/technical occupations relative to ACS and EEO-1 benchmarks. For this paper, the selection concern is mitigated by three features of the research design. First, the comparison is *within* LinkedIn—India vs. ROW vs. Native, all observed on the same platform—so platform selection affects levels but not the event-study differences, provided selection is similar across visa groups. Second, the H-1B population is itself concentrated in professional/technical occupations where LinkedIn coverage is strongest; the relevant external validity question is whether the monopsony mechanism operates for H-1B workers *not* in these occupations (e.g., specialty occupation workers in healthcare or fashion modeling), which is a small share of the H-1B population. Third, matching on metro × detailed occupation ensures that the native comparison group is drawn from the same professional labor markets where LinkedIn coverage is high, so differential coverage across occupations does not confound the comparison.

5.2 H-1B Identification and PERM Linkage

I identify H-1B workers through a multi-stage probabilistic linkage to administrative records (full algorithm details in Appendix H):

1. **PERM Linkage:** Match LinkedIn profiles to Department of Labor PERM labor certification filings using exact matches on canonicalized employer name, 2-digit SOC occupation group, and a timing window of -90 days to $+36$ months relative to the worker's first U.S. job start date. The PERM filing date anchors the event study: it marks the moment τ_X jumps from τ_X^{pre} to τ_X^{high} in the model.

2. **LCA Enrichment:** Supplement with Labor Condition Application (LCA) records filed for H-1B positions, matched on employer name, 2-digit SOC, and a ±180-day window around the first U.S. job start date.

3. **Origin Classification:** Classify workers by country of origin using pre-U.S. work history:

   - **India:** Pre-U.S. employment in India. These workers face $T^{*,India} \approx 10$–$15+$ years.

   - **ROW (Rest of World):** Pre-U.S. employment in other countries (Canada, UK, South Korea, Mexico, etc.), excluding China. Chinese workers are excluded because they face intermediate backlog durations ($T^* \approx 3$–5 years)—longer than ROW but shorter than India—making them neither a clean "short constraint" nor "long constraint" group. These ROW workers face $T^{*,ROW} \approx 2$–3 years.

4. **Predictive Model:** Train a logistic regression on a "gold set" of high-confidence labeled workers: gold positives are India/China-origin workers matched to PERM or LCA records in specialty occupations; gold negatives are ROW-origin workers with no administrative match. The classifier uses 13 features—including PERM/LCA match indicators, country-of-origin history, job timing (H-1B cap season, October start, summer OPT timing), employer H-1B sponsorship intensity, STEM occupation, and

metropolitan area—to assign H-1B probability scores. Workers scoring ≥ 0.75 are retained. The enriched sample includes education, occupation (SOC), and PERM wage data for 96.2 percent of identified H-1B workers.

5.3   Native Comparison Sample

The comparison group consists of U.S. workers without indicators of H-1B status, matched to H-1B workers within metro area $\times$ detailed occupation (ONET 6-digit) cells. This ensures natives work in the same local labor markets and narrow occupations as H-1B workers, controlling for occupation-specific career dynamics and local labor market conditions.

For within-person analysis, natives require a pseudo-PERM date to align event time. I assign each native a pseudo-PERM date equal to their first job date plus the cell-average years-to-PERM for H-1B workers in the same metro $\times$ occupation cell. This career-stage alignment anchors each native's event time to their own career trajectory, matching the career phase at which H-1B workers typically file PERM (median 5.8 years; India 6.7, ROW 5.2). Calendar-time pseudo-PD creates an artificial 10 percentage point native decline over the event window; the career-stage method reduces this to approximately 3 percentage points, reflecting genuine career lifecycle effects rather than compositional artifacts. Note that native and H-1B career profiles differ substantially in the cross section: native external mobility rises steeply with job sequence number (from 0.28 at job 2 to 0.61 at job 6), while H-1B external mobility is flat at approximately 0.21 across all job sequences. This level difference reflects the career-stage mismatch between groups—H-1B workers at PERM are typically on their 5th job, while natives in the same cell are on their 2nd–3rd—but is absorbed by person fixed effects in the within-person design. Restricting to natives with 5+ observed jobs (who better match H-1B career stage) yields a flatter native baseline ($\sim$0.30 across the event window), confirming that the within-person native trend is robust to career-stage alignment.

5.4 Hazard Rates as Model Objects

The model's transition hazards h_X and h_I map directly to empirical hazard rates computed at the cell level:

$$h_d^{g,q} = \frac{\sum_i \mathbf{1}[\text{transition type } d \text{ for worker } i]}{\sum_i T_i}, \quad d \in \{X, I\} \tag{9}$$

where g indexes group (India, ROW, Native), q indexes quarters since PERM, and T_i is person-years contributed by worker i to the cell. This duration-weighted cell-level computation avoids the bias that arises from averaging spell-level rates, which inflates estimates by a factor of approximately four due to overweighting short spells.

These are independent measures—they need not sum to any particular value, and both could increase, both decrease, or move in opposite directions. The model's key prediction (Section 4) is that under monopsony, internal mobility fails to rise when external constraints tighten: $\partial h_X / \partial \tau_X < 0$ and $\partial h_I / \partial \tau_X \leq 0$.

5.5 India vs ROW: Same Constraint, Different Duration

India and ROW H-1B workers face the same institutional constraint—employer-specific sponsorship during the green card process—but differ in constraint duration T^*. This generates a natural dose-response comparison. The per-country cap creates $T^{*,India} \approx 10\text{--}15+$ years versus $T^{*,ROW} \approx 2\text{--}3$ years, so the cumulative welfare loss $\Delta W^v \propto T^{*v}$ is approximately 4–7 times larger for Indian workers. Post-AC21, ROW workers recover as their constraint expires, while Indian workers remain suppressed—generating the divergence predicted by Equation (8).

5.6 Summary Statistics

Table 1 presents descriptive statistics for the three groups at pre-PERM periods ($h < 0$), when visa constraints are weakest.

Table 1: Summary Statistics: Pre-PERM Means by Group (Cross-Sectional)

	India	ROW	Native
Sample sizes (within-person analysis):			
Workers	75,349	89,726	192,088
Hazard rates (transitions per person-year):			
External (h_X)	0.249	0.268	0.266
Internal (h_I)	0.058	0.054	0.038
Stay probability (h_S)	0.692	0.679	0.696
Wage gains conditional on transition (log points):			
External ($\tilde{g}_X$)	0.276	0.265	0.231
Internal ($\tilde{g}_I$)	0.250	0.171	0.234
H-1B worker characteristics:			
Median PERM offered wage	$93,000	$93,000	—
Advanced degree (%)	41		—
CS/Math occupation (%)	26		—

Notes: Pre-PERM means computed over half-years $h \in \{-4, -3, -2, -1\}$ relative to PERM filing. Hazard rates are duration-weighted cell-level rates (transitions/person-years). Wage gains are mean log wage changes conditional on the transition type. Worker counts for within-person analysis require three or more observed jobs. H-1B characteristics from enriched panel (96.2 percent match rate); advanced degree and CS/Math shares are pooled across India and ROW. Native workers matched within metro area $\times$ detailed occupation (ONET 6-digit) cells.

Three features are noteworthy. First, pre-PERM external hazard rates are similar across groups (0.25–0.27), with natives slightly higher—consistent with modest employer-specificity of the H-1B visa even before PERM ($\tau_X^{pre} > \tau_X^N$). Second, H-1B workers have substantially higher internal hazard rates than natives (0.054–0.058 vs 0.038). This is consistent with the model's prediction that firms offer more internal mobility to workers with constrained external options, though it could also reflect selection—H-1B workers may sort into firms or industries with higher baseline internal mobility. Third, external wage gains are larger for H-1B workers than natives (0.27 vs 0.23 log points), suggesting positive selection among H-1B workers who do change employers—consistent with higher switching costs filtering out low-gain moves. Mean external wage gains exceed mean internal gains ($\tilde{g}_X > \tilde{g}_I$), but this ranking reverses at the median (see footnote 4): external moves have a larger right tail, pulling the mean above the median. Internal moves represent steady career advancement with less dispersion.

Figure 2 provides a career-level view of total mobility. Native workers reweighted to the H-1B metro $\times$ occupation distribution exhibit a smooth decline in total hazard rates from 0.37 to 0.34 over eight career years—the natural settling pattern as workers find better matches. H-1B workers, by contrast, are persistently flat at approximately 0.27 transitions per person-year, yielding a gap of roughly 0.07 at every career stage. India and ROW are nearly indistinguishable in this cross-sectional comparison, consistent with both groups facing the same institutional constraint. The career trajectory is a cross-sectional comparison across career years and cannot identify the within-person effect of PERM filing; the event study design in the next section provides this identification.

6 Empirical Strategy

The empirical design tests the model's predictions by tracking how h_X and h_I evolve around the PERM filing and AC21 milestones—the moments when τ_X jumps and partially relaxes. I
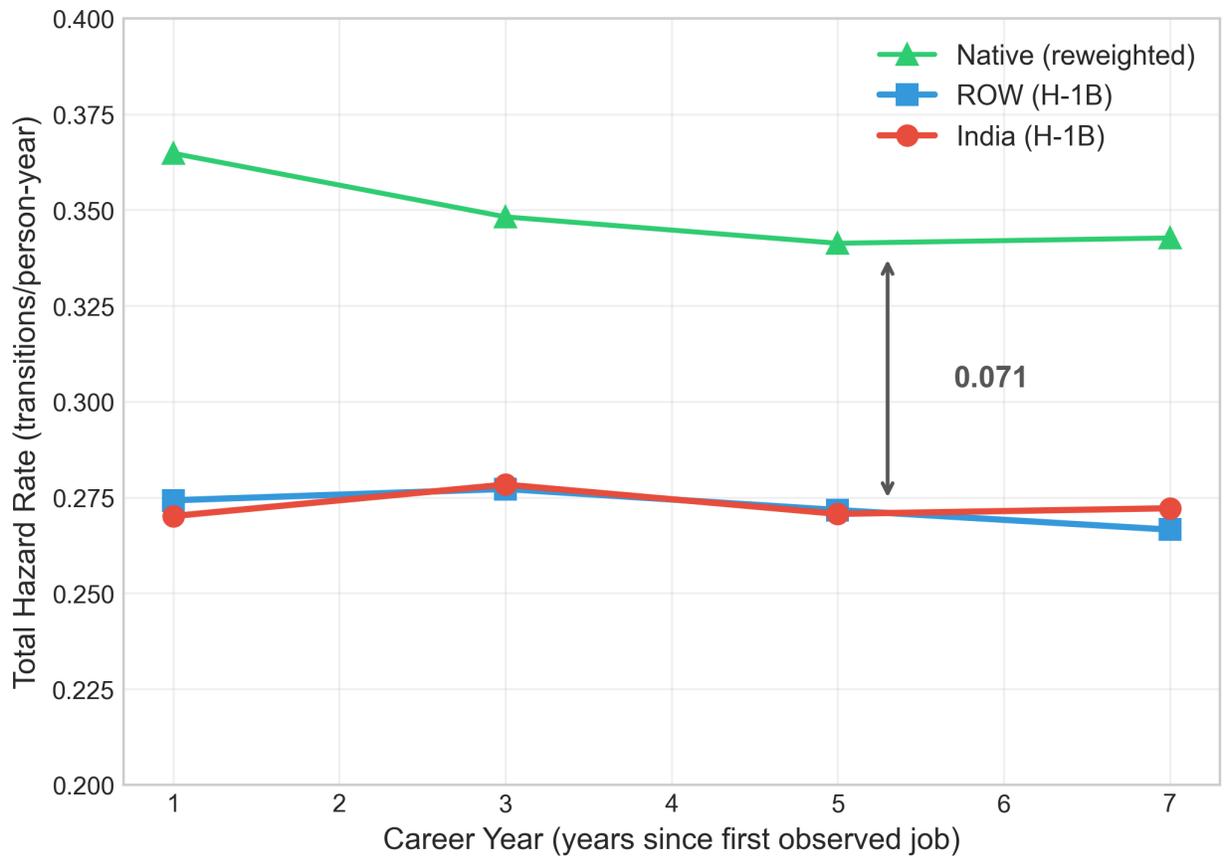
Figure 2: Total mobility over the career: H-1B vs. native workers. Total hazard rates (external + internal transitions per person-year) by career year (years since first observed job), in two-year bins with year-FE adjustment. Native workers are reweighted to the H-1B metro × detailed occupation cell distribution to ensure compositional comparability. H-1B workers show persistently lower total mobility at every career stage (gap ∼0.07), with India and ROW nearly indistinguishable in this cross-sectional comparison.

employ three complementary approaches: cross-sectional event studies, within-person identification, and a two-channel decomposition that quantifies the internal margin's contribution to total lock-in.

6.1 Event Study: Cell-Level Hazards by Group and Quarter

The primary specification estimates group-specific hazard rate profiles as a function of quarters since PERM (q):

$$h_d^{g,q} = \alpha_d^g + \sum_{q \neq -1} \beta_{d,q}^g \cdot \mathbf{1}[q] + \gamma_d^g X_t + \varepsilon_{d,t}^g \tag{10}$$

where $d \in \{X, I\}$ indexes the transition margin, $g \in \{\text{India}, \text{ROW}, \text{Native}\}$ indexes group, and X_t includes calendar year fixed effects. Year fixed effects are implemented via residualization: I demean hazard rates by calendar year and re-aggregate, which removes secular trends in mobility that affect all groups. The event study is normalized to $q = -1$ (one quarter before PERM).

The key comparisons are:

- **PERM effect ($q = 0$ vs $q = -1$):** The jump in τ_X at filing should reduce h_X and, under monopsony, h_I.

- **AC21 effect ($q = 2$ vs $q = 1$):** Partial portability should increase h_X as τ_X partially relaxes.

- **India–ROW divergence (post-AC21):** ROW workers approach $t = T^{*,ROW}$ and recover; Indian workers remain constrained.

6.2 Within-Person Identification

The cross-sectional event study may confound changes in τ_X with compositional shifts: workers observed at later quarters since PERM may differ systematically from those observed

earlier. To address this, I use two within-person estimators, both applied to workers with three or more observed jobs.

Primary estimator: two-period test. The cleanest test compares the same workers' hazard rates before and during lock-in. For each worker i observed in both the pre-PERM ($q \in [-9, -1]$) and lock-in ($q \in [0, 3]$) periods, I compute the within-person change $\Delta h_{d,i} = h_{d,i}^{\text{lock-in}} - h_{d,i}^{\text{pre}}$ and test whether the group mean $\overline{\Delta h_{d,i}}$ differs from zero.[5] This requires no functional form assumptions beyond comparing the same worker in two time periods, and it is robust to any fixed person-level heterogeneity.

Event-study visualization: Mundlak-style adjustment. To visualize the within-person hazard profile across all quarters since PERM, I use a correlated random effects adjustment:

$$h_{d,it}^{WP} = h_d^{g,q} - \left(\bar{h}_{d,i} - \bar{h}_d^g \right) \tag{11}$$

where $\bar{h}_{d,i}$ is person i's duration-weighted mean hazard rate and $\bar{h}_d^g$ is the group mean. This is algebraically equivalent to demeaning each worker's hazard rate by their own mean under duration-weighted aggregation, removing fixed person-level heterogeneity in mobility. The Mundlak adjustment generates the event-study figures (Figure 3), while the two-period test provides the primary statistical inference (Table 2).

The within-person estimates serve as the primary identification, not merely a robustness check. The model's predictions concern how *the same worker's* mobility changes as τ_X varies—exactly what within-person identification recovers.

Selection into PERM. Workers observed at PERM filing are selected: they remained with their sponsoring employer long enough to initiate the green card process, which selects on low baseline external mobility. This selection is a feature, not a threat, for the within-person

---

[5]The pre-PERM window spans approximately 4.5 years before filing ($q = -9$ to $q = -1$ in half-year bins). Results are not sensitive to the window length: shortening to $q \in [-6, -1]$ or lengthening to $q \in [-12, -1]$ yields two-period changes within 0.02 of the baseline estimates.

design: person fixed effects absorb the level of each worker's mobility, and the identifying variation comes from how *the same worker's* mobility changes at PERM versus before.

Endogeneity of PERM timing. A potential concern is that PERM filing is endogenous: workers or employers may time the filing in response to anticipated mobility changes, which would confound the event study. Three pieces of evidence mitigate this concern. First, pre-trend slope tests on within-person hazard rates over $q \in [-9, -1]$ show that H-1B mobility is trending *upward* before PERM (India external: slope $= +0.009$ per quarter, $p = 0.001$; ROW external: $+0.010$, $p = 0.005$), while native mobility trends slightly downward (-0.001, $p < 0.001$). If anything, the pre-existing upward H-1B trend makes the subsequent PERM dip more conservative: the discrete decline at PERM reverses a rising trajectory, ruling out the concern that a pre-existing downward trend generates a spurious break. One might worry that the upward pre-trend reflects a transitory component that naturally reverts, generating a spurious dip at PERM. However, the magnitude of the PERM decline far exceeds what the pre-trend slope could produce through reversion: extrapolating the pre-PERM slope forward, the linear trend predicts an *increase* of 0.03–0.04 over the lock-in window, while the observed change is a decline of 0.11–0.12—an excess decline of 0.15–0.16 beyond any trend-predicted change. The same pattern holds for internal mobility: the pre-PERM internal trend predicts an increase of 0.02, while internal hazard rates fall by 0.06—an excess decline of 0.07–0.08. Natives show no excess decline on either margin (<0.001). Moreover, the AC21 recovery—a sharp increase at a statutory milestone unrelated to any trending process—is inconsistent with mean reversion. Second, PERM timing is primarily driven by institutional factors outside the worker's control—employer immigration counsel capacity, Department of Labor processing backlogs, and the prevailing wage determination process—rather than worker-level mobility decisions. The median H-1B worker files PERM 5.2 years (ROW) to 6.7 years (India) after initial employment, reflecting employer administrative timelines rather than strategic timing by workers. Third, native workers matched within metro $\times$ occupation

cells show no change in mobility at their career-stage-aligned pseudo-PERM date, confirming that the H-1B mobility break at PERM is specific to the visa constraint rather than reflecting a general career-stage phenomenon.

6.3 Identifying the Internal Channel

The model predicts that under monopsony, both margins co-decline at PERM:

Prediction 1 (Both Margins Co-Decline). *Under monopsony ($MPL - w_j$ large), the endogenous offer reduction dominates mechanical substitution, so $\partial h_I / \partial \tau_X < 0$. Both external and internal hazards decline at PERM. The internal share of the total mobility decline, φ_I, measures how much external-only analysis misses.*

I test this using the within-person estimates, which control for compositional changes that could obscure the true within-person effect on internal mobility. If internal mobility declines within-person at PERM—while natives show no change—this confirms that monopsony operates on both margins and that external-only estimates understate the lock-in cost.

7 Results: Mobility

This section documents how visa constraints affect both external and internal mobility, quantifying the two-channel response to lock-in. The cross-sectional event study (Appendix Figure 4) shows clear external mobility patterns—a PERM dip and sustained post-PERM decline for both H-1B groups, with natives flat—but also an apparent internal spike at PERM that is entirely compositional and disappears within-person. I therefore focus on the within-person estimates throughout.

7.1 External Mobility: PERM Dip and AC21 Recovery

The model predicts that visa constraints directly reduce external mobility: $\partial h_X / \partial \tau_X < 0$. External hazard rates should drop at PERM (when τ_X jumps) and partially recover at AC21

(when τ_X partially relaxes). Indian workers should show a larger and more persistent dip than ROW workers due to longer constraint duration $T^{*,India} \gg T^{*,ROW}$.

Year-FE-adjusted external hazard rates confirm the external channel. Pre-PERM, India and ROW workers have similar external mobility ($\sim$0.26–0.28 transitions per person-year), while natives are higher and flat throughout ($\sim$0.30–0.31). At PERM ($q = 0$), both H-1B groups dip sharply; post-PERM, external mobility continues to decline, reaching $\sim$0.21 for India and $\sim$0.23 for ROW by $q = 12$. In the cross-section, the AC21 recovery is muted— consistent with the within-person finding that the cross-sectional AC21 bounce is largely compositional (workers who move at AC21 exit the sample, leaving a selected post-AC21 pool). The India–ROW divergence is exactly the pattern predicted by Equation (8): same constraint mechanism, different duration. Indian workers, facing 7–10 years of per-country backlog, plateau well below natives; ROW workers, whose constraints expire after $T^{*,ROW} \approx 2$–3 years, gradually converge. The within-person external decline at PERM is large: $\Delta h_X = -0.193$ for India and -0.196 for ROW (two-period test, user-level equal-weighted means; these differ from the cell-level duration-weighted rates in Table 2 because of weighting, but the implied φ_I is preserved), representing a 24 percent decline from pre-PERM levels. This magnitude establishes that the visa constraint generates substantial variation in external mobility—the "first stage" for identifying the internal channel response.

7.2   Internal Mobility: The Second Channel

The model predicts that monopsony suppresses internal mobility alongside external: firms reduce internal offers when workers lack credible exit options ($\partial g_I^* / \partial \tau_X < 0$, Proposition 1). This is the channel that external-only estimates miss.

Cross-sectionally, internal hazard rates show an apparent spike at PERM ($q = 0$) for H-1B workers, but this is compositional: within-person estimates show the PERM spike disappears entirely. More importantly, two-period within-person tests—comparing the same workers before PERM ($q \in [-9, -1]$) and during lock-in ($q \in [0, 3]$)—reveal a substantial *decline*

28

in internal mobility: $\Delta h_I = -0.079$ for India (95 percent CI $[-0.119, -0.040]$, $t = -3.86$, one-sided $p < 0.001$) and $\Delta h_I = -0.116$ for ROW (95 percent CI $[-0.143, -0.089]$, $t = -7.95$, one-sided $p < 0.001$). Native workers show no such decline ($\Delta h_I = +0.040$, $p = 0.048$, which does not survive Bonferroni correction for six tests—three groups $\times$ two margins—at the $\alpha/6 = 0.0083$ threshold). The borderline native result likely reflects residual career-stage trends: despite career-stage pseudo-PD alignment, the native pseudo-PERM occurs at a slightly earlier career phase (median job 2–3 vs job 5 for H-1B workers), where internal mobility naturally rises with tenure. I report the two-period test rather than quarter-by-quarter event study coefficients because internal transitions are substantially less frequent than external transitions ($h_I \approx 0.06$ vs $h_X \approx 0.25$), making quarterly estimates noisy. Pooling quarters into pre-PERM and lock-in periods maximizes statistical power while preserving the key comparison.

The internal decline confirms that monopsony operates on two margins, not one. This is the result that standard external-only estimates miss: the same visa constraint that reduces employer-to-employer transitions also reduces within-firm career advancement, because firms face less competitive pressure to offer advancement when workers are locked in.

The magnitude is substantial. In the cell-level within-person estimates (Table 2), internal hazard rates fall from 0.074 to 0.031 for India and from 0.080 to 0.028 for ROW—declines of 58 percent and 65 percent, respectively, computed as $(h_I^{\text{pre}} - h_I^{\text{lock-in}})/h_I^{\text{pre}}$.[6] At the lock-in rate of $h_I \approx 0.03$, only about 6 percent of workers receive an internal promotion or raise in a given year, compared to 7–8 percent at the pre-PERM baseline. Over a typical 2–3 year lock-in period for ROW workers, the probability of receiving at least one internal move falls from roughly 16 percent to 6 percent; for Indian workers facing 7–10 years of constraint, the cumulative effect is substantially larger. The contrast with natives—whose internal mobility

---

[6]The bootstrap Δh_I values reported above (-0.079 and -0.116) are computed at the user level: each worker's mean hazard rate in the pre-PERM period minus their mean in the lock-in period, then averaged across workers with equal weight per person. The cell-level hazard rates in Table 2 are duration-weighted within group $\times$ half-year cells, which gives more weight to workers with longer spells. The two approaches weight workers differently but yield the same implied φ_I because the ratio $\Delta h_I/(\Delta h_X + \Delta h_I)$ is preserved.

is flat throughout—confirms that this decline is driven by the visa constraint, not by career lifecycle or firm-level trends.

7.3 Both Margins Co-Decline: Quantifying the Two-Channel Response

Figure 3 visualizes the two-channel response. Panel A shows external mobility: at PERM, H-1B hazard rates drop sharply and partially recover at AC21, while natives are flat. Panel B shows internal mobility at its own scale: H-1B internal hazard rates fall from $\sim$0.07–0.09 pre-PERM to $\sim$0.03 at lock-in, while natives remain flat at $\sim$0.06. Both margins decline simultaneously, and natives show no break on either margin. The AC21 recovery is asymmetric across channels: external mobility recovers strongly for both India (+0.100, 95 percent CI $[0.052, 0.152]$) and ROW (+0.129, $[0.086, 0.179]$), while internal mobility recovers modestly for India (+0.045, $[0.006, 0.087]$, $p = 0.012$) and not at all for ROW (+0.002, $[-0.028, 0.035]$, $p = 0.45$). The incomplete internal recovery likely reflects the partial nature of AC21 portability: workers can only transfer to a "same or similar" occupation, and the new employer must be willing to take over a pending green card case, so τ_X drops but does not return to τ_X^N.[7] The asymmetry does not invalidate the co-decline at PERM—the paper's central finding—but it does suggest that the relationship between external constraints and internal offers is not fully symmetric on the way down and the way up. The India–ROW asymmetry in internal recovery is not fully explained by the model (which predicts stronger ROW recovery given shorter constraint duration) and remains an open question for future work.

Table 2 quantifies the decomposition (note that within-person hazard rate levels differ from the cross-sectional Table 1 due to the person-FE adjustment and career-stage alignment;

---

[7]Two additional factors may contribute. First, internal advancement may exhibit hysteresis: workers passed over for promotions during lock-in lose momentum in the pipeline, and restoring their trajectory takes time. Second, the internal decline may partly reflect reduced firm investment in developing workers whose long-term retention remains uncertain.
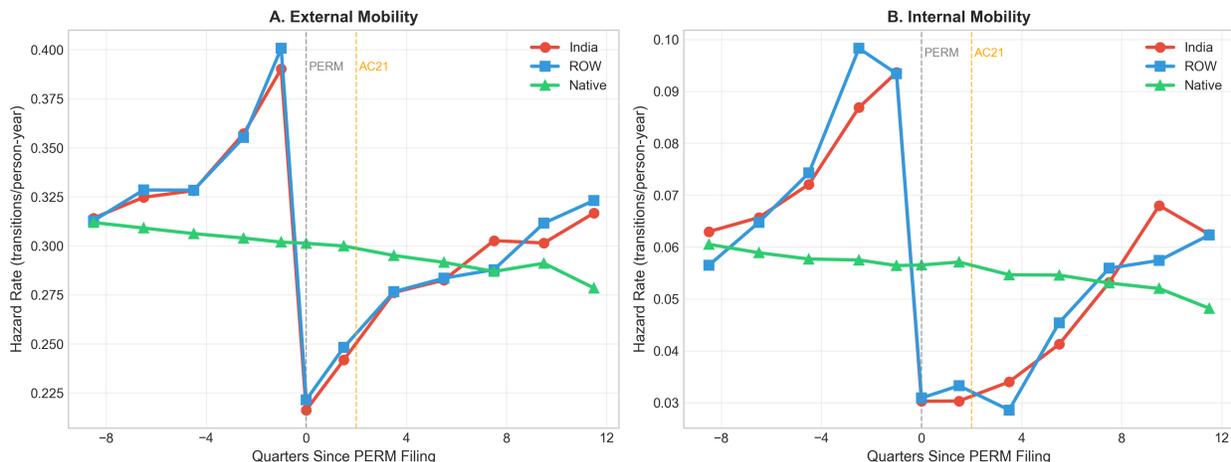
Figure 3: External and internal mobility around PERM filing (within-person balanced panel, half-year bins). Panel A: external hazard rates (h_X). Panel B: internal hazard rates (h_I), shown at their own scale to make the co-decline visible. For H-1B workers (India, ROW), *both* margins decline sharply at PERM and partially recover at AC21; for natives, mobility is essentially flat on both margins (no lock-in).

see Table 2 notes). Within-person estimates show that Indian workers' total separation rate drops from 0.41 to 0.28 at PERM—a decline of 0.14—of which the external channel accounts for 0.09 and the internal channel for 0.04 ($\varphi_I = 0.29$, 95 percent CI [0.15, 0.47]). ROW workers show a similar pattern ($\varphi_I = 0.37$, 95 percent CI [0.28, 0.50]). Natives are essentially flat. Bootstrap 95 percent CIs for the individual channels confirm that both external and internal declines are statistically significant: for India, $\Delta h_X = -0.193$ [$-0.280, -0.109$] and $\Delta h_I = -0.079$ [$-0.119, -0.040$]; for ROW, $\Delta h_X = -0.196$ [$-0.268, -0.123$] and $\Delta h_I = -0.116$ [$-0.143, -0.089$]. The total decline is $\Delta h_T = -0.272$ for India and -0.312 for ROW.[8] A researcher who measures lock-in costs from external separations alone misses 29–37 percent of the total mobility decline. Accounting for lower wage gains on internal moves ($\tilde{g}_I < \tilde{g}_X$), the welfare-weighted understatement is 33–39 percent (Table 2, φ_I^W column).

Balanced panel representativeness. The balanced panel requires workers with job spells in both the pre-PERM ($q \in [-9, -1]$) and lock-in ($q \in [0, 3]$) windows. This selects on two conditions: (a) remaining observable (staying in the U.S. and on LinkedIn) through lock-in,

---

[8]Bootstrap statistics are computed on within-person user-level mean hazard rate differences between the pre-PERM ($q \in [-9, -1]$) and lock-in ($q \in [0, 3]$) periods, using $B = 1,000$ resamples of balanced-panel users. The user-level scale differs from the cell-level hazard rates in Table 2, but the implied φ_I is similar.

Table 2: Lock-in Cost Decomposition: Pre-PERM vs. Lock-in Period (Within-Person)

Group	Pre-PERM ($q \in [-9, -1]$)			Lock-in ($q \in [0, 3]$)			Internal Share	
	h_X	h_I	$h_X + h_I$	h_X	h_I	$h_X + h_I$	φ_I	φ_I^W
India	0.338	0.074	0.412	0.244	0.031	0.275	29.1%	28.0%
							[15.2, 46.7]	
ROW	0.339	0.076	0.415	0.247	0.032	0.279	37.2%	24.8%
							[27.7, 50.3]	
Native	0.307	0.058	0.366	0.299	0.057	0.356	—	—

Notes: Within-person estimates (person FE + group-specific year FE). H-1B groups use the balanced panel (workers observed both pre-PERM and during lock-in); native baseline uses the full sample with career-stage pseudo-PD. $\varphi_I = \Delta h_I / (\Delta h_X + \Delta h_I)$ is the mobility internal share; bootstrap 95 percent CIs in brackets ($B = 1,000$). $\varphi_I^W = \Delta(h_I \cdot \tilde{g}_I) / [\Delta(h_X \cdot \tilde{g}_X) + \Delta(h_I \cdot \tilde{g}_I)]$ is the welfare-weighted internal share, accounting for lower wage gains on internal moves ($\tilde{g}_I < \tilde{g}_X$). The welfare understatement factor from ignoring the internal channel is $1/(1 - \varphi_I^W) - 1$: for India, $1/(1 - 0.280) - 1 = 0.389$ (39 percent); for ROW, $1/(1 - 0.248) - 1 = 0.330$ (33 percent). Native shares are not reported because the total decline is negligible ($\Delta h_T = -0.010$; $\Delta h_X = -0.008$, $\Delta h_I = -0.002$), serving as a placebo: matched natives show no break on either margin at the career-stage-aligned pseudo-PERM date. Native internal hazard rates here (0.058) differ from Table 1 (0.038) for two reasons: (i) Table 1 pools all pre-PERM half-years cross-sectionally, while this table uses within-person estimates that adjust for person composition; and (ii) career-stage pseudo-PD alignment shifts the native event window to later career stages, where internal mobility is naturally higher (Figure 2, Panel B). Native external hazard rates are less affected (0.266 vs 0.307) because external mobility declines with tenure while internal mobility rises. The balanced panel comprises 8.2 percent of Indian and 9.3 percent of ROW workers from the full sample; these workers are similar to the full sample on observables (advanced degree share within 2 percentage points, median number of jobs identical, CS/Math occupation share within 5 percentage points, median PERM offered wage similar).

and (b) having job episodes that span a 12+ quarter window around PERM. Condition (a) excludes workers who left the PERM employer quickly or left the U.S., which likely biases *against* finding a large internal decline: excluded workers with the weakest employer attachment are plausibly the most affected by lock-in on both margins. Condition (b) selects workers with longer careers in the LinkedIn data, not workers with more transitions—a worker who stays at the PERM employer continuously from $q = -5$ to $q = 3$ qualifies. The 8–9 percent figure primarily reflects the panel structure of LinkedIn data (most workers are not observed across enough years to span the event window) rather than a selection on mobility. A logit of balanced-panel membership on observables confirms that selection is driven primarily by data coverage rather than worker characteristics: the strongest predictors are PERM filing year (+0.4–0.5 percentage points per year, reflecting better LinkedIn coverage in later cohorts) and number of observed jobs (+0.2 percentage points per additional job, reflecting longer career histories). Substantive worker characteristics have small marginal effects: advanced degree (-0.4 percentage points), CS/Math occupation (-1.0 percentage points for India, insignificant for ROW), and log PERM wage (-1.0 to -1.3 percentage points). Balanced and non-balanced workers are similar on observables (advanced degree shares within 1 percentage point, median number of jobs within 1), and the full-sample within-person estimates (which do not require balance) show qualitatively identical patterns, suggesting that the balanced sample provides a conservative estimate of φ_I. A bounding exercise confirms this: if excluded workers (91 percent of the sample) experienced zero internal decline, the population-weighted φ_I would fall to 1.7–3.1 percent—but this zero-effect assumption is extreme. A more plausible bound, where excluded workers experience half the balanced panel's internal decline, yields $\varphi_I = 17.6$–23.8 percent, still economically substantial. Even under the adversarial assumption that excluded workers experienced *positive* internal mobility growth of 5 percentage points, φ_I remains negative (the external decline dominates), confirming that the qualitative result—both margins co-decline—is not an artifact of balanced panel selection.

This bias in measured mobility has a direct implication for estimates of the welfare cost of monopsony. In the standard framework (Manning, 2003, 2021), the firm's wage markdown depends on the labor supply elasticity it faces, which is estimated from separation rates. Internal moves are not separations—they are a retention instrument—so the separation elasticity and the implied markdown may be correctly measured on their own terms. But the markdown captures only the external channel of monopsony power. When visa constraints suppress internal advancement alongside external mobility, workers bear an additional welfare cost that separation-based methods miss entirely.

It is worth distinguishing two quantities. The *lock-in cost decomposition* asks how much of the total mobility decline at PERM flows through the internal channel—and the answer is 29–37 percent, a first-order omission. The *steady-state markdown* based on external separations alone may require only modest correction, because internal transitions are infrequent relative to external ($h_I \approx 0.06$ vs $h_X \approx 0.25$). The internal margin matters most not for correcting steady-state markdowns, but for measuring the total welfare cost of discrete lock-in events like visa constraints, non-competes, and other institutional frictions that suppress both margins simultaneously.

Table 3 presents the results in a formal difference-in-differences specification. Panel A pools all three groups and estimates the differential H-1B mobility change relative to natives: Lock-in × India and Lock-in × ROW capture the excess within-person decline for H-1B workers beyond any native career-stage trend. Both DID estimates are large and highly significant on all three margins: the total mobility DID is -0.41 for India and -0.44 for ROW ($p < 0.001$). The native baseline change is economically small and statistically insignificant on all margins. Panel B confirms that the individual within-person changes for India and ROW are each significant at the 1 percent level, while natives show no change. The DID-implied φ_I (0.30 for India, 0.35 for ROW) is consistent with the decomposition estimates in Table 2.

Table 3: Difference-in-Differences: Mobility Changes at PERM

	(1) External	(2) Internal	(3) Total
Panel A: Difference-in-differences (relative to natives)			
Lock-in × India	−0.284***	−0.123***	−0.407***
	(0.071)	(0.031)	(0.073)
Lock-in × ROW	−0.286***	−0.155***	−0.441***
	(0.067)	(0.028)	(0.068)
Lock-in (Native baseline)	0.044	0.023	0.067
	(0.057)	(0.023)	(0.059)
Panel B: Within-person changes by group			
Δh (India)	−0.240***	−0.100***	−0.340***
	(0.042)	(0.021)	(0.043)
Δh (ROW)	−0.242***	−0.132***	−0.374***
	(0.035)	(0.015)	(0.035)
Δh (Native)	0.044	0.023	0.067
	(0.057)	(0.023)	(0.059)
Person FE	✓	✓	✓
Year FE (residualized)	✓	✓	✓
Persons (balanced panel)	5,093	5,093	5,093
India	1,316	1,316	1,316
ROW	2,170	2,170	2,170
Native	1,607	1,607	1,607
φ_I (India)	—	—	0.303
φ_I (ROW)	—	—	0.352

Notes: Each column reports estimates from a first-differenced specification equivalent to person FE with $T = 2$. The dependent variable is the within-person change in hazard rate (transitions per person-year) between pre-PERM ($q \in [-9, -1]$) and lock-in ($q \in [0, 3]$) periods. Panel A reports pooled DID estimates: Lock-in × India and Lock-in × ROW are the differential H-1B changes relative to natives; Lock-in (Native baseline) captures the native change at career-stage-matched pseudo-PERM. Panel B reports within-person changes separately by group. Year FE are residualized at the spell level using group-specific year effects before collapsing to person × period. Heteroskedasticity-robust standard errors in parentheses. φ_I is the internal share of the total DID mobility decline: Panel A column (3) internal coefficient divided by total coefficient. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

DID robustness. Three additional tests confirm the DID results are not driven by the native comparison or by pre-trends. First, a *placebo test* assigns random pseudo-PERM dates to natives (500 iterations, each drawn from the career-stage distribution in each native's metro $\times$ occupation cell). The distribution of placebo two-period changes is centered near zero (mean $+0.022$ for external, $+0.019$ for internal), with 95 percent of draws in $[-0.043, +0.093]$ for external mobility—far from the H-1B effects of -0.19 to -0.20. Second, the H-1B within-person declines are highly significant *without any native comparison*: bootstrap 95 percent CIs (1,000 resamples of balanced-panel users) for the within-person change are entirely below zero for both groups and both margins (India external: $[-0.283, -0.110]$, $t = -4.49$; ROW internal: $[-0.144, -0.088]$, $t = -7.95$), with zero sign-flip probability across all four group $\times$ margin combinations. The DID against natives adds a valuable counterfactual but is not necessary for the main result. Third, as discussed in the endogeneity section above, detrending the pre-PERM slope shows that the PERM decline is 3–4 times larger than what mean reversion of the upward pre-trend could explain.

7.4 Alternative Explanations

The co-decline of external and internal mobility at PERM—with natives unaffected—is the central empirical result. Several alternative mechanisms could potentially generate a decline in one or both margins. I consider each and explain why the monopsony model provides the most parsimonious account of the full pattern of results.

Information revelation. PERM filing reveals the employer's willingness to sponsor a green card, potentially changing the worker's information set. If this revelation reduces search effort (the worker decides to stay), external mobility would decline. However, this mechanism does not explain the internal decline: learning that the employer values the worker should, if anything, increase the worker's bargaining position for internal promotions. Moreover, native workers—who also learn about employer investment through other retention signals—show

no discrete mobility break at career-stage-matched pseudo-PERM dates.

Firm-side information updating. A related but distinct channel operates on the firm side: at PERM, the firm learns that the worker has chosen to pursue a green card through this employer, revealing a preference to remain in the job. The firm could then Bayesian-update downward on the worker's reservation utility—inferring that the worker values the match highly—and reduce both external counteroffers and internal promotion offers accordingly. This mechanism generates a co-decline without requiring the monopsony channel. However, the timing pattern is inconsistent with pure information updating: the firm's posterior should stabilize shortly after PERM (the informative event), yet the mobility suppression persists for the entire lock-in period and reverses discretely at AC21—a statutory threshold that changes the worker's *legal* options but conveys no new information about preferences. The AC21 recovery implies that what constrains mobility is the binding legal restriction, not the firm's beliefs about the worker's type.

Career lifecycle. Mobility naturally declines over a career as workers settle into better matches. If PERM timing correlates with career stage, this could generate a declining trend around PERM even absent visa constraints. Two features of the research design address this concern. First, person fixed effects absorb the level of each worker's mobility, so the identifying variation comes from within-person changes relative to the worker's own baseline. Second, the career-stage pseudo-PD for natives explicitly aligns event time with career progression: each native's pseudo-PERM occurs at the career stage when H-1B workers in the same metro $\times$ occupation cell typically file PERM. Natives show no mobility break at this date, confirming that the H-1B pattern is visa-specific.

Worker effort reallocation. Workers locked to a firm might reallocate effort from job search toward on-the-job productivity, potentially improving internal promotion prospects. This predicts *higher* internal mobility (or at least higher productivity) during lock-in—the

opposite of what the data show. The sharp decline in internal mobility at PERM is inconsistent with effort reallocation driving the observed patterns.

Reduced search intensity. Workers who know they cannot change employers may reduce search effort, lowering external mobility through a demand-side (fewer applications) rather than supply-side (higher switching costs) channel. Even if this reduction were discrete at PERM (triggered by learning that the employer has committed to sponsorship), it does not explain the internal decline: reduced search effort should have no effect on within-firm promotions and role changes, which are initiated by the employer, not by worker search activity. This is the strongest argument against the search intensity channel—it cannot generate the co-decline of both margins that the data show. The AC21 recovery in external mobility provides additional evidence: the sharp increase at the statutory portability threshold is more consistent with a binding constraint that relaxes than with a resumption of search effort at a milestone that does not directly change workers' search incentives.

Why the endogenous firm response is preferred. The model's mechanism—firms reduce internal offers when the retention motive weakens—uniquely predicts the full pattern: (i) external decline at PERM (direct constraint), (ii) internal decline at PERM (endogenous offer reduction), (iii) discrete timing at visa milestones (legal constraint, not gradual adjustment), and (iv) natives unaffected (no visa constraint, no mobility break). No single alternative mechanism generates all four features. The co-decline of both margins is particularly informative: without endogenous firm behavior, models in which the firm's offers are fixed would predict that internal mobility should *rise* when external mobility is blocked (mechanical substitution), not decline.

7.5   Summary of Quantitative Results

The main quantitative findings are as follows. Within-person external mobility declines by $\Delta h_X = -0.193$ for India and -0.196 for ROW at PERM (bootstrap 95 percent CIs entirely

below zero; $t = -4.49$ and -5.49). Internal mobility co-declines: $\Delta h_I = -0.079$ for India ($t = -3.86$) and -0.116 for ROW ($t = -7.95$), representing 58–65 percent declines from pre-PERM levels. Natives show no change on either margin. The internal share of the total mobility decline is $\varphi_I = 29.1$ percent for India (95 percent CI [15.2, 46.7]) and 37.2 percent for ROW ([27.7, 50.3]). The welfare-weighted internal share is $\varphi_I^W = 25$–28 percent, implying that external-only welfare estimates understate the total lock-in cost by 33–39 percent. The DID against natives confirms these magnitudes: Lock-in × India $= -0.41$ and Lock-in × ROW $= -0.44$ for total mobility ($p < 0.001$). The co-decline holds across all subgroups tested (occupation, education, wage tercile, Indian university) and across three independent promotion definitions (Table 4).

8 Welfare Decomposition

The mobility results establish that visa constraints suppress both external and internal margins, with the internal channel accounting for 29–37 percent of the total mobility decline (Table 2). But how much do these constraints cost workers in welfare terms? The welfare decomposition derived in Section 4.2 provides a framework for answering this question.

Before proceeding, it is worth being explicit about what each result requires. The paper's results form a hierarchy of increasing assumption-dependence:

1. **The co-decline** (Section 7): both external and internal mobility fall at PERM for H-1B workers while natives are unaffected. This is a purely empirical fact that requires no model.

2. **The mobility decomposition** $\varphi_I = 29$–37 percent (Table 2): the internal share of the total mobility decline. This uses only within-person hazard rates and requires no structural assumptions, no functional form, and no wage data—it is a nonparametric accounting identity.

3. **The welfare-weighted decomposition** $\varphi_I^W = 25$–28 percent: translating mobility shares into welfare shares requires the three-option framework (Proposition 2), which assumes workers optimize over stay/internal/external options and that the value of each option maps to observed wage gains. This additionally requires salary data (predicted rather than administrative; see Section 5 for validation).

The central claim—that standard external-only estimates miss a quantitatively important margin—rests on levels 1 and 2, which are assumption-free. The welfare translation (level 3) quantifies the cost in welfare-weighted terms but is not necessary for the main conclusion. This distinction matters because the salary data used in level 3 are model-predicted rather than administrative (the prediction model achieves a log correlation of 0.27 with PERM offered wages; see Section 8.1 for details and validation). The mobility decomposition φ_I is entirely immune to salary measurement error, and the welfare-weighted φ_I^W provides a lower bound on the mobility-based understatement ($\varphi_I^W \leq \varphi_I$), so even substantial noise in predicted wages cannot reverse the qualitative conclusion.

A note on identification strategy: the paper's central understatement result—the mobility-based φ_I of 29–37 percent—is estimated from the within-person balanced panel (Table 2), which provides the cleanest causal identification. The welfare decomposition in this section uses the full cross-sectional sample, which provides substantially more precise estimates of hazard rates, wage gains, and their products at each event-time horizon (the balanced panel contains only 8–9 percent of workers and yields noisy half-year estimates of conditional wage gains). The cross-sectional welfare flows should therefore be interpreted as descriptive—they show how switching welfare evolves around visa milestones—rather than as within-person causal estimates. The causal content comes from the within-person φ_I, which establishes how much of the mobility decline is missed by external-only measures; the cross-sectional welfare flows then translate this into approximate dollar terms.

Cross-sectional wage growth patterns are consistent with the monopsony mechanism. At PERM ($q = 0$), wage growth spikes for H-1B workers—ROW to 28 percent and India to

19 percent, both above the native baseline of approximately 16 percent—consistent with retention raises offered before the constraint binds. Post-PERM, wage growth drops sharply: India falls to 9.4 percent and ROW to 10.4 percent at $q = 2$, well below the native baseline. ROW recovers to native levels by $q = 10$–12 as the constraint expires, while India remains 1–3 percentage points below. These patterns mirror the mobility dynamics: the same constraint that suppresses mobility also suppresses wage growth, with recovery tracking constraint duration. I do not report within-person wage growth estimates because the cross-sectional analysis exhibits non-flat pre-trends that complicate causal interpretation, but the cross-sectional patterns are consistent with the monopsony mechanism and motivate the welfare decomposition that follows.

8.1   Implementation

I compute the empirical welfare decomposition by group × half-year bin (using half-year bins rather than quarters for smoothing, as quarterly estimates are noisy):

$$\left.\frac{dW^g}{dt}\right|_{\text{half-year } h} = \underbrace{\overline{\Delta \ln w}^{g,h}}_{\text{wage drift}} + S^g(h) \cdot \left[\underbrace{h_I^{g,h} \cdot \tilde{g}_I^{g,h}}_{\text{internal switching}} + \underbrace{h_X^{g,h} \cdot \tilde{g}_X^{g,h}}_{\text{external switching}} \right] \tag{12}$$

where $\tilde{g}_d^{g,h}$ is the mean log wage gain conditional on a type-d transition in group g during half-year h, and $S^g(h)$ is the Kaplan–Meier survival probability computed from the group-specific hazard rates. The survival probability declines substantially over the event window: $S(0) \approx 0.43$ at PERM for both H-1B groups, falling to $S(6) \approx 0.19$ by three years post-PERM, which attenuates post-PERM welfare flows considerably. Salary data are generated by a predictive model that combines multiple sources—H-1B filing wages, job posting data, CPS occupational wage distributions, and platform-reported salaries—using an ensemble of gradient-boosted trees with company, title, location, seniority, and tenure embeddings as features. The model produces salary estimates for all observations, which enter through the wage drift term $\overline{\Delta \ln w}^{g,h}$ and the conditional wage gains $\tilde{g}_d$. Because these are model-predicted

rather than individually reported wages, the welfare decomposition should be interpreted as using the best available salary estimates rather than administrative wage records.[9]

8.2 Results

External switching value—$h_X \cdot \tilde{g}_X$—dominates the welfare decomposition, accounting for 83–91 percent of total switching welfare across groups and time periods. The internal share of switching welfare ranges from 9 to 17 percent, smaller than the mobility share ($\varphi_I = 29$–37 percent) because $\tilde{g}_X > \tilde{g}_I$. The welfare-weighted internal share φ_I^W is 25–28 percent, implying that external-only welfare estimates understate the total lock-in cost by 33–39 percent.

The India–ROW divergence is sharp. Defining Δ as the change in switching welfare relative to natives between pre-PERM and post-AC21 periods, ROW is essentially flat ($\Delta = -0.001$) while India drops substantially ($\Delta = -0.010$). The cumulative welfare cost for Indian workers is approximately 4–7 times larger, driven entirely by the per-country green card cap (Section 4.4).

9 Robustness and Extensions

9.1 Heterogeneity: Mechanism Tests

The two-margin monopsony pattern—external and internal mobility co-declining at PERM— could be driven by a particular subpopulation rather than the general mechanism. If only

---

[9]To validate the salary model, I compare LinkedIn-predicted salaries to PERM offered wages— administrative wages certified by the Department of Labor—for H-1B workers at their PERM job. The log correlation is 0.27 ($n = 216{,}477$), with 68 percent of predictions within 50 percent of the PERM wage. Prediction quality is similar across visa groups (India median absolute error 37.9 percent; ROW 41.7 percent), so any measurement error is approximately symmetric across groups. More importantly, the key results rely on within-person *changes* in salary across jobs, not salary levels. Prediction error in model-generated salaries—which shrinks estimates toward the conditional mean, a form of nonclassical measurement error— attenuates wage growth estimates symmetrically across groups and time periods, so the India–ROW–Native *differences* in wage growth are not systematically biased by the prediction model. Importantly, the paper's main understatement result—the mobility-based φ_I of 29–37 percent (Table 2)—uses only hazard rates and does not depend on salary predictions at all. The welfare-weighted φ_I^W of 25–28 percent uses salary data but provides a *lower bound* on the mobility-based understatement, so even substantial salary prediction noise would not reverse the qualitative conclusion.

low-wage workers or only CS workers show the pattern, the mechanism may be more specific than the model implies.

This section tests whether the co-decline persists across four dimensions of heterogeneity: occupation, education, wage level, and institution type. All estimates use within-person identification (person FE + year FE), which isolates the causal effect of the constraint from compositional differences across subgroups. I focus primarily on external mobility heterogeneity, where sample sizes support subgroup comparisons; internal mobility patterns by subgroup are noisier but show parallel declines across all groups (consistent with the main results in Table 2).

9.1.1 Occupation: CS/Math vs Non-CS

If monopsony power operates through external switching costs that are occupation-specific, the pattern might differ between high-demand occupations (CS/Math, SOC 15) where outside options are abundant and other occupations where they are scarcer.

Within-person external mobility by occupation group reveals a striking result: CS/Math and Non-CS workers track identically within-person. The PERM dip, the post-AC21 partial recovery, and the India–ROW divergence are all present in both occupation groups with nearly identical magnitudes. This rules out explanations that depend on occupation-specific outside options or occupation-specific employer market power. The monopsony mechanism operates through the visa constraint itself—τ_X—not through occupation-specific labor market conditions.

For internal mobility, CS/Math workers show slightly higher internal hazard rates than Non-CS workers for India (0.06 vs 0.04), but the within-person trajectory shapes are parallel.

9.1.2 Education: Advanced Degree vs Not

Workers with advanced degrees (masters or PhD) may have better outside options and stronger bargaining positions. If the endogenous offer reduction depends on the firm's baseline

43

markdown, more educated workers with smaller markups might show weaker monopsony effects.

As with occupation, the education dimension generates no within-person heterogeneity in external mobility. Advanced-degree and non-advanced-degree workers show overlapping profiles, with parallel PERM dips and post-AC21 trajectories. This is consistent with the model: the constraint τ_X^H is determined by visa status, not by education, so both groups face the same external switching cost increase at PERM.

9.1.3 Wage Tercile: PERM Offered Wage

Wage level provides the most natural test of the monopsony mechanism. The model predicts that firms with large markups ($MPL - w_j$ large) invest heavily in retention through g_I, and pull back sharply when τ_X rises. If monopsony power varies by wage level, we should see heterogeneous responses.

Wage terciles generate the clearest within-person heterogeneity. Workers are divided by PERM offered wage into Low (<\$82k), Mid (\$82–112k), and High (>\$112k) terciles:

- **Level differences.** Low-Wage workers have the highest baseline external mobility ($\sim$0.27 transitions/person-year), consistent with weaker attachment and more frequent job changes. High-Wage workers have the lowest ($\sim$0.20), reflecting stronger attachment and higher switching costs.

- **Lock-in depth.** High-Wage India workers drop to the lowest absolute floor ($\sim$0.14 at $q = 1$), the largest relative decline. This is consistent with the model: high-wage workers have the most to lose from an employer change during constraint (forfeiting a high salary and restarting the green card process), so τ_X binds most tightly for them.

- **Similar absolute lock-in.** Despite the level differences, the absolute magnitude of the PERM dip is similar across wage groups ($\sim$0.05–0.06 transitions per person-year), suggesting that the constraint shock $\Delta\tau_X$ at PERM is similar regardless of wage level.

9.1.4 Indian University (India Only)

For Indian H-1B workers, institution prestige may affect outside options. Workers who attended top-tier Indian institutions (IITs, IIMs, etc.) may have stronger professional networks and more outside options, potentially moderating the constraint's effect. Workers who attended Indian universities and then completed U.S. education may have different labor market attachment than those who did not.

As with occupation and education, the Indian university dimension generates minimal within-person heterogeneity in external mobility. Workers who attended Indian universities and those who did not show parallel trajectories with similar PERM dips and post-AC21 paths. This suggests that the visa constraint is the dominant force: regardless of institutional prestige, the legal barrier to employer change (τ_X^H) binds similarly for all Indian H-1B workers.

9.1.5 Summary: Monopsony is General

The heterogeneity analysis delivers a clear message: the two-margin monopsony pattern is general, not driven by a particular subpopulation.

- **Occupation:** No within-person heterogeneity. CS/Math and Non-CS workers track identically.

- **Education:** No within-person heterogeneity. Advanced degree and non-advanced-degree overlap.

- **Wage:** Level differences in baseline mobility, but the co-decline holds across all terciles. Lock-in is deepest for High-Wage workers in absolute terms.

- **Indian University:** No within-person heterogeneity. Institutional background does not moderate the constraint.

These comparisons are based on visual inspection of overlapping within-person hazard rate profiles rather than formal interaction tests (triple-difference specifications of the form

H-1B × post-PERM × subgroup). The balanced panel—already only 8–9 percent of the full sample—split further into subgroups produces cells too small for precise interaction estimates (e.g., India × balanced × CS/Math contains fewer than 2,000 workers). Minimum detectable effect (MDE) calculations confirm that these interaction tests are underpowered to detect moderate effects. At 80 percent power ($\alpha = 0.05$, two-sided), the MDE for the occupation × lock-in interaction on external mobility is 0.056 transitions per person-year—roughly twice the observed interaction of 0.013—and the MDE for the education interaction is similarly large (0.042–0.057). Only the balanced-vs-full-sample contrast is adequately powered. The near-identical trajectory shapes across subgroups within each dimension—particularly for occupation and education, where the profiles are virtually superimposed—are therefore consistent with either no heterogeneity or moderate heterogeneity below the detection threshold. Importantly, any detectable heterogeneity would need to be large enough to reverse the qualitative conclusion that the co-decline holds within a subgroup, which seems implausible given that all subgroup trajectories show the same shape and sign.

This generality is itself evidence for the model's mechanism. The monopsony channel operates through the legal constraint ($\tau_X^H > \tau_X^N$) and the firm's endogenous response ($\partial g_I^* / \partial \tau_X < 0$). Because the legal constraint is identical across worker subgroups—all H-1B workers in the green card queue face the same employer-specific restriction—the mechanism should produce similar within-person patterns regardless of occupation, education, or institution. That is precisely what the data show.

The wage tercile heterogeneity in *levels* (not within-person trajectory shapes) is consistent with the model's additional prediction that workers with higher w_j face larger opportunity costs of changing employers during constraint, leading to deeper lock-in even when the constraint τ_X is identical.

9.2 Robustness of the Internal Decline: Three Promotion Definitions

LinkedIn title changes are an imperfect proxy for internal career advancement. To confirm that the internal decline reflects genuine promotion suppression rather than measurement artifacts, I test three independent definitions of promotion and show that all three produce the same pattern.

- **Title-based:** A seniority hierarchy extracted from LinkedIn job titles (intern $\to$ junior $\to$ associate $\to$ mid $\to$ senior $\to$ manager $\to$ director $\to$ VP $\to$ C-suite). An internal transition to a higher level counts as a promotion. This captures 39 percent of internal transitions.

- **Wage-based:** An internal transition accompanied by a salary increase exceeding 10 percent. This definition is completely independent of job titles and captures 55 percent of internal transitions. If the title-based decline were driven by cosmetic title changes or profile noise, the wage-based measure—which relies on salary data rather than title keywords—should show a weaker or absent pattern.

- **Strict:** The intersection of title upgrade and positive wage growth, the most conservative definition. This captures 32 percent of internal transitions—cases where the worker both moved up the seniority hierarchy *and* received a pay raise, leaving little doubt that genuine career advancement occurred.

All three definitions produce quantitatively similar results in the two-period within-person test (Table 4). Title-based promotions decline by 59 percent for both India and ROW. Wage-based promotions decline by 66–69 percent—the *largest* decline, which is the opposite of what measurement noise in titles would predict. The strict definition shows 63 percent declines. Natives show no comparable change on any definition (-2 to -10 percent, not statistically significant after Bonferroni correction). The consistency across three independent definitions—especially the wage-based measure, which uses entirely different data from the

title classification—confirms that the internal decline at PERM reflects genuine suppression of career advancement, not artifacts of how LinkedIn title changes are measured.

Table 4: Promotion Decline at PERM: Three Independent Definitions

Definition	India	ROW	Native
Title seniority upgrade	−58.8%	−59.1%	−1.6%
Wage increase >10%	−66.2%	−68.6%	−10.3%
Title upgrade + wage increase	−63.7%	−62.5%	−1.7%
Total internal (all types)	−64.7%	−69.5%	−10.6%

Notes: Two-period within-person test: percent change in hazard rate from pre-PERM ($q \in [-9, -1]$) to lock-in ($q \in [0,3]$) for balanced-panel workers observed in both periods ($n_{\text{India}} = 1{,}316$; $n_{\text{ROW}} = 2{,}170$; $n_{\text{Native}} = 2{,}191$). Title-based: internal transition to a higher seniority rank in a 10-level hierarchy. Wage-based: internal transition with >10 percent salary increase (independent of title). Strict: title upgrade *and* positive wage growth.

10   Conclusion

This paper shows that monopsony operates on two margins of labor mobility, not one. When visa constraints tighten at PERM filing, external mobility drops directly and internal mobility drops endogenously—firms reduce promotions, raises, and role changes because locked-in workers pose no credible exit threat. Within-person estimates confirm that both margins decline sharply for H-1B workers while natives are unaffected, and a two-channel decomposition reveals that 29–37 percent of the total mobility decline flows through the internal channel. Standard monopsony estimates based on external separations alone miss this second margin, understating the welfare cost of lock-in by 33–39 percent.

Three features of the results strengthen the case for two-margin monopsony. First, the pattern is general: heterogeneity analysis confirms the internal decline across occupation, education, wage level, and institution type. Second, the promotion decline is robust to measurement: three independent definitions of promotion—title seniority upgrade, wage increase exceeding 10 percent (independent of titles), and the intersection of both—all show

59–69 percent declines for H-1B workers with natives essentially unaffected (Table 4). Third, the internal decline is large in absolute terms: internal hazard rates fall by 58–65 percent relative to pre-PERM baseline, effectively shutting down within-firm career advancement during the constraint period.

For policy, these findings imply that visa portability reforms would operate through two channels, not one. Policies that reduce external switching costs (such as expanding AC21 or reducing green card backlogs) would restore external mobility directly and should increase internal offers endogenously, as firms once again compete to retain workers who can plausibly leave. One caveat: the AC21 recovery is asymmetric across margins—external mobility recovers strongly while internal mobility recovers only modestly (India) or not at all (ROW)—suggesting that the internal channel may exhibit hysteresis or adjustment lags, so reforms may restore internal mobility more slowly than external. The gains would be largest for Indian workers, whose cumulative welfare loss is 4–7 times that of ROW workers due entirely to the per-country cap.

When does the internal margin matter beyond immigration? The model's mechanism requires three conditions: (i) switching costs must be *employer-specific*, so that a particular firm gains monopsony power over its own workers; (ii) the firm must be able to adjust *individualized* internal offers in response to the worker's outside option; and (iii) internal mobility must be frequent enough that its suppression is quantitatively important. Non-compete agreements satisfy all three conditions: they are employer-specific, firms can respond with targeted retention offers, and internal promotions are a meaningful share of career advancement for affected workers. Employer-provided health insurance (before the ACA) is a similar case. Occupational licensing creates geographic or occupational barriers rather than employer-specific lock-in per se, but when combined with employer-financed training investments, it can effectively tie workers to the sponsoring employer. By contrast, geographic monopsony in thin labor markets operates through a different channel—workers lack outside options due to few nearby employers, not due to employer-specific switching costs—and the

firm's incentive to adjust internal offers may be weaker when the constraint is market-wide rather than firm-specific. The decomposition framework (φ_I) is directly portable: applying it to non-compete or job-lock contexts would establish whether the magnitude of the internal channel in these settings is comparable to the 29–37 percent found here. The H-1B context is extreme (workers face loss of legal work authorization), and the internal share likely depends on both the severity of the external constraint and the size of the surplus at stake in the match.

A Model Derivations

This appendix presents the full three-option discrete choice framework summarized in Section 4.

Worker's Problem

A worker i employed at firm j at wage w_j faces three options each period:

- **Stay (S)**: Remain in current role at wage w_j.

- **Internal move (I)**: Accept a within-firm transition (promotion, raise) at wage $w_j + g_I$, where g_I is the internal wage gain offered by the firm.

- **External move (X)**: Move to a new firm k at wage $\bar{w}_X$, incurring switching cost τ_X^v that depends on visa status $v \in \{H, N\}$.

Utility from each option includes a deterministic component and an idiosyncratic Type I Extreme Value (Gumbel) preference shock:

$$U_{id} = V_d + \sigma \epsilon_{id}, \quad d \in \{S, I, X\} \tag{E.1}$$

where $\sigma > 0$ governs the variance of idiosyncratic preferences, and:

$$V_S = \ln w_j \tag{E.2}$$

$$V_I = \ln(w_j + g_I) - \tau_I \tag{E.3}$$

$$V_X = \ln \bar{w}_X - \tau_X^v \tag{E.4}$$

Here $\tau_I \geq 0$ is a hassle cost of internal transitions (common to all workers), and τ_X^v captures switching costs that are higher for H-1B workers: $\tau_X^H > \tau_X^N \geq 0$.

Choice Probabilities

Standard logit gives the transition hazards:

$$h_d^v = \frac{\exp(V_d/\sigma)}{\Phi^v}, \quad d \in \{S, I, X\} \tag{E.5}$$

where the inclusive value is $\Phi^v = \exp(V_S/\sigma) + \exp(V_I^v/\sigma) + \exp(V_X^v/\sigma)$. Note that V_I^v depends on visa status through the firm's endogenous offer g_I^v.

Firm's Internal Offer Problem

The firm employs the worker at current wage w_j with marginal product MPL. Expected profit:

$$\Pi(g_I) = h_S(g_I) \cdot (MPL - w_j) + h_I(g_I) \cdot (MPL - w_j - g_I)$$
$$= (1 - h_X) \cdot (MPL - w_j) - h_I \cdot g_I \tag{E.6}$$

The first-order condition for the optimal internal offer g_I^*:

$$\frac{\partial \Pi}{\partial g_I} = \underbrace{-\frac{\partial h_X}{\partial g_I}}_{>0} \cdot (MPL - w_j) - h_I - g_I \cdot \underbrace{\frac{\partial h_I}{\partial g_I}}_{>0} = 0 \tag{E.7}$$

How Visa Constraints Reduce Internal Offers

From the logit structure:
$$\frac{\partial h_X}{\partial g_I} = -\frac{h_X \cdot h_I}{\sigma(w_j + g_I)} < 0 \tag{E.8}$$

Increasing the internal offer draws workers away from external moves (the *retention motive*).

$$\frac{\partial h_I}{\partial g_I} = \frac{h_I(1 - h_I)}{\sigma(w_j + g_I)} > 0 \tag{E.9}$$

When τ_X increases: (i) h_X decreases directly; (ii) the retention motive $(\partial h_X/\partial g_I) \cdot (MPL-$

w_j) decreases in magnitude; (iii) the firm's optimal response is to lower g_I^*. By the implicit function theorem applied to the FOC, $\partial g_I^* / \partial \tau_X < 0$ (Proposition 1 in the main text).

B   Welfare Derivation

The worker's expected flow utility from the three-option choice is the logit inclusive value (McFadden, 1981):

$$\mathcal{W}(t) = \sigma \ln \Phi^v(t) = \sigma \ln \left[\exp \frac{V_S(t)}{\sigma} + \exp \frac{V_I^v(t)}{\sigma} + \exp \frac{V_X^v(t)}{\sigma} \right] \qquad \text{(A.1)}$$

Normalizing by $\exp(V_S/\sigma)$:

$$\mathcal{W}(t) = V_S(t) + \sigma \ln \left[1 + \exp \frac{\Delta V_I}{\sigma} + \exp \frac{\Delta V_X}{\sigma} \right]$$
$$= \ln w_j(t) + \sigma \ln \left[1 + \exp \frac{\Delta V_I}{\sigma} + \exp \frac{\Delta V_X}{\sigma} \right] \qquad \text{(A.2)}$$

where $\Delta V_I = V_I - V_S$ and $\Delta V_X = V_X - V_S$ are the utility gains from internal and external moves relative to staying. Welfare equals the log wage plus an *option value* term reflecting the worker's ability to switch.

Taking the time derivative and using the logit choice probabilities:

$$\frac{d\mathcal{W}}{dt} = \frac{d \ln w_j}{dt} + h_I(t) \cdot \frac{d\Delta V_I}{dt} + h_X(t) \cdot \frac{d\Delta V_X}{dt} \qquad \text{(A.3)}$$

Equation (5) in the main text is not a direct derivative of A.2; rather, it is an expected flow decomposition in a continuous-time hazard framework. In each period, the worker either stays (gaining $d \ln w_j/dt$), makes an internal move (gaining $\tilde{g}_I = \ln(w_j + g_I) - \ln w_j$), or makes an external move (gaining $\tilde{g}_X = \ln \bar{w}_X - \ln w_j$). The expected welfare flow is therefore

the weighted sum of these outcomes:

$$\frac{d\mathcal{W}}{dt} = \frac{d\ln w_j}{dt} + h_I(t) \cdot \tilde{g}_I(t) + h_X(t) \cdot \tilde{g}_X(t) \tag{A.4}$$

where the weights are the hazard rates (transition probabilities per unit time). This is consistent with A.3 when utility gaps ΔV_d are approximately equal to the log wage gains $\tilde{g}_d$ at the margin of transition (i.e., when switching costs are small relative to wage gains, or when the relevant comparison is between workers facing different switching costs at the same wage gains). Conditioning on still being in the current match introduces the survival function $S(t - t_0) = \exp\left(-\int_{t_0}^{t}[h_I(s) + h_X(s)]\,ds\right)$, yielding Equation (6).

C Monopsony Markdown

Comparing an H-1B worker (type H) to an identical native worker (type N) at the same firm:

$$\mathcal{W}^H - \mathcal{W}^N = \underbrace{(\ln w_j^H - \ln w_j^N)}_{\text{base wage markdown}} + \underbrace{\sigma \ln \frac{\Phi^H}{\Phi^N}}_{\text{option value loss}} \tag{B.1}$$

Since $\Delta V_X^H < \Delta V_X^N$ (visa constraint) and $\Delta V_I^H < \Delta V_I^N$ (endogenous firm response), the option value term is unambiguously negative. The total welfare gap exceeds the base wage gap.

At PERM ($t = 0$), τ_X jumps and welfare drops through two channels: h_X drops directly, and g_I^* drops endogenously, reducing h_I. At AC21, portability partially restores the external option, but the recovery differs by group: ROW workers ($T^* \approx 2$–3 years) recover within 1–2 years, while Indian workers ($T^* \approx 10$–15+ years) remain suppressed due to the green card backlog.

The total welfare gap between constrained and unconstrained workers includes both

channels:

$$\underbrace{\mathcal{W}^N - \mathcal{W}^H}_{\text{total welfare gap}} = \underbrace{(\ln w^N - \ln w^H)}_{\text{base wage gap}} + \underbrace{\sigma \ln \frac{\Phi^N}{\Phi^H}}_{\text{option value gap (both margins)}} \tag{B.2}$$

D   Model-to-Data Mapping

The key empirical objects—hazard rates, wage gains, and survival probabilities—are measured directly from the data without structural estimation. The model provides the theoretical framework that motivates these objects and generates the predictions tested in the main text, but the lock-in decomposition (φ_I) and welfare decomposition are nonparametric.

Model Object	Data Counterpart	Source
$h_X(t)$	External hazard rate (transitions/person-year)	LinkedIn panel
$h_I(t)$	Internal hazard rate (same-company transitions/person-year)	LinkedIn panel
$\tilde{g}_X(t)$	Mean log wage gain on external moves	LinkedIn salaries
$\tilde{g}_I(t)$	Mean log wage gain on internal moves	LinkedIn salaries
$S(t)$	Kaplan–Meier survival probability	Computed from hazards

Table 5: Mapping model objects to empirical counterparts. All objects are measured directly from the data without logit inversion or calibrated switching costs.

The empirical welfare decomposition (Equation (6) in the main text) decomposes welfare flows into three observable components: wage drift ($\overline{\Delta \ln w}$), external switching value ($h_X \cdot \tilde{g}_X$), and internal switching value ($h_I \cdot \tilde{g}_I$). External switching dominates: the internal share of switching welfare is 9–17 percent across groups and time periods, reflecting $h_X \gg h_I$ and $\tilde{g}_X > \tilde{g}_I$.

E   Cross-Sectional Event Study

Table 6 reports the corresponding event-study coefficients from Equation (10), with bootstrap standard errors (200 resamples of users within each group). The PERM effect ($h = 0$) is negative and significant for external mobility (India -0.008, ROW -0.015; natives flat) and positive for internal mobility (India $+0.007$, ROW $+0.007$)—the compositional PERM spike
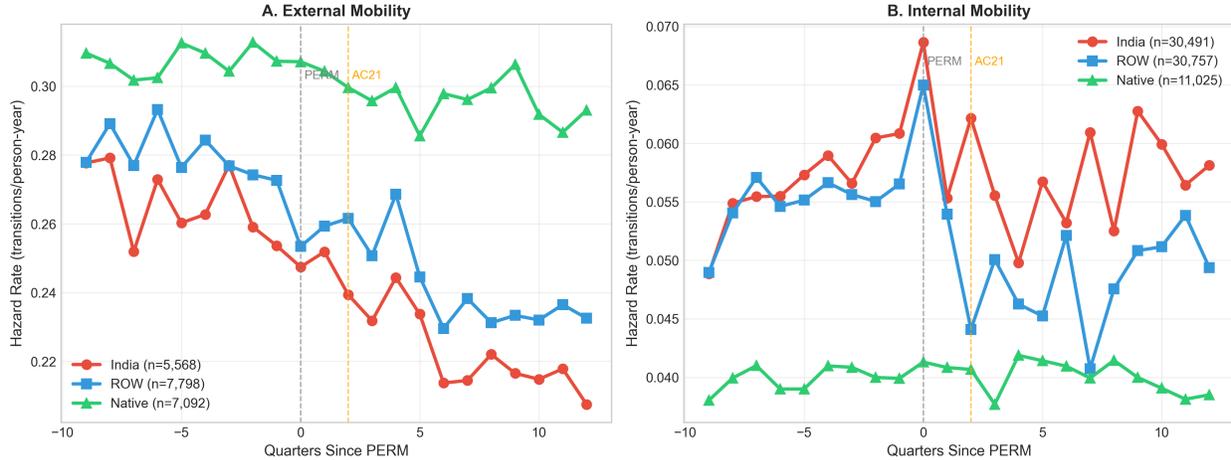
Figure 4: Cross-sectional event study: external and internal hazard rates by quarter since PERM (year-FE adjusted, metro × detailed occupation matching, career-stage pseudo-PD for natives). Panel A: external mobility drops at PERM and declines post-PERM for H-1B workers; natives are flat. The muted cross-sectional AC21 recovery reflects compositional selection: workers who move at AC21 exit the sample. Panel B: internal mobility shows an apparent compositional spike at PERM for H-1B workers. This spike disappears entirely within-person (Figure 3 in the main text), confirming it reflects compositional shifts in who is observed at each event-time quarter rather than a genuine within-person increase.

that disappears within-person (Section 7). Post-PERM internal coefficients are negative for both H-1B groups and positive for natives, consistent with the lock-in pattern.

F Sample Construction

Table 7 documents the sample construction from raw LinkedIn profiles to the final analysis samples.

G Welfare Decomposition Details

This appendix provides the full welfare decomposition results summarized in Section 8.

External switching dominance. The product $h_X \cdot \tilde{g}_X$ accounts for 83–91 percent of total switching welfare across groups and time periods. This follows from two empirical regularities: external transitions are approximately five times more frequent than internal ($h_X \gg h_I$), and external wage gains exceed internal ($\tilde{g}_X > \tilde{g}_I$).

Table 6: Event-Study Coefficients: External and Internal Mobility by Half-Year Since PERM

Half-year	External Mobility			Internal Mobility		
	India	ROW	Native	India	ROW	Native
$[-9, -8]$	+0.014***	+0.005***	−0.002	−0.006***	−0.003**	+0.001
	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)
$[-7, -6]$	+0.009***	+0.005***	−0.006***	−0.003*	+0.001	+0.001
	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)
$[-5, -4]$	+0.007***	+0.002	+0.001	−0.001	+0.000	+0.002**
	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)
$[-3, -2]$	+0.009***	+0.000	−0.001	−0.001	−0.001	+0.000
	(0.002)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)
-1	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
0	−0.008***	−0.015***	−0.001	+0.007***	+0.007***	+0.001
	(0.001)	(0.001)	(0.003)	(0.001)	(0.001)	(0.002)
$[1, 2]$	+0.008***	+0.001	−0.007***	−0.005***	−0.004***	+0.001
	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)
$[3, 4]$	+0.029***	+0.033***	−0.011***	−0.010***	−0.010***	+0.004***
	(0.006)	(0.005)	(0.003)	(0.004)	(0.003)	(0.002)
$[5, 6]$	+0.015***	+0.010**	−0.016***	−0.007**	−0.010***	+0.006***
	(0.005)	(0.004)	(0.003)	(0.004)	(0.003)	(0.002)
$[7, 8]$	+0.009*	+0.009*	−0.010***	−0.006*	−0.014***	+0.003
	(0.005)	(0.005)	(0.003)	(0.003)	(0.003)	(0.002)
$[9, 10]$	+0.008	+0.008	−0.008**	−0.001	−0.008**	+0.005**
	(0.005)	(0.005)	(0.004)	(0.004)	(0.004)	(0.002)
$[11, 12]$	+0.006	+0.012**	−0.018***	−0.006	−0.008**	+0.005**
	(0.006)	(0.006)	(0.004)	(0.004)	(0.004)	(0.002)
$h = -1$ level	0.216	0.228	0.309	0.064	0.058	0.052

Notes: Cross-sectional event-study coefficients $\beta_{d,h}^{g}$ from Equation (10), normalized to $h = -1$. Bootstrap standard errors in parentheses (200 resamples of users within each group). Year fixed effects are group-specific (residualized within each group separately). Half-year bins pool adjacent quarters except at $h = -1$ (reference) and $h = 0$ (PERM) to preserve the discontinuity. Stars: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table 7: Sample Flow

Step	India	ROW	Native
Starting sample:			
LinkedIn profiles (stratified sample)	$\sim$5,000,000		
U.S.-based profiles	$\sim$2,600,000		
H-1B identification (direct match):			
Matched to PERM/LCA records	$\sim$165,000		—
Classified as India-born	47,401	—	—
Classified as ROW	—	43,645	—
H-1B identification (probabilistic):			
+ Name/nationality algorithm	(expands H-1B pool)		
Analysis sample restrictions:			
Three or more observed jobs	75,349	89,726	192,088
Completed spells (event = 1, job_seq > 1)	(spell-level restrictions)		
Within-person balanced panel:			
Observed in both pre-PERM and lock-in	6,179	8,344	—
(percent of analysis sample)	(8.2%)	(9.3%)	—

Notes: The starting sample is a stratified random sample of LinkedIn profiles that oversamples workers in occupations and metros with high H-1B concentration. H-1B identification uses a multi-stage probabilistic linkage to PERM labor certifications and LCA filings (Section 5; Appendix H). The analysis sample counts exceed the direct-match counts because the probabilistic identification stage adds workers identified through name/nationality algorithms who were not directly linked to PERM/LCA records. The "three or more observed jobs" restriction provides sufficient within-person variation for the hazard rate analysis. Native workers are matched to H-1B workers within metro area $\times$ detailed occupation (ONET 6-digit) cells. The balanced panel requires job spells in both the pre-PERM ($q \in [-9, -1]$) and lock-in ($q \in [0, 3]$) windows. Native balanced panel statistics are not reported because native pseudo-PD balanced selection is artificial (Section 7).

Internal share dynamics. The internal share of switching welfare—$h_I \cdot \tilde{g}_I / (h_I \cdot \tilde{g}_I + h_X \cdot \tilde{g}_X)$— ranges from 9 to 17 percent. Two patterns stand out. First, India's internal share rises from 13 percent to 18 percent post-AC21, as ROW workers recover externally while Indian workers remain constrained—the internal channel's relative contribution grows, but not enough to compensate ($\Delta = -0.010$ for India vs. $\Delta = -0.001$ for ROW). Second, for Indian workers post-AC21, where internal share reaches 18 percent, the understatement from ignoring the internal channel is most severe.

India–ROW divergence. Both groups experience the same constraint shock at PERM. Post-AC21, ROW approaches $T^{*,ROW}$ and switching welfare returns to near pre-PERM levels, while India remains far from $T^{*,India}$ and switching welfare drops substantially. The cumulative welfare cost—the integral of this gap over the constraint period—is approximately $T^{*,India}/T^{*,ROW} \approx$ 4–7 times larger for Indian workers, driven entirely by the per-country green card cap.

H H-1B Identification Algorithm

This appendix describes the full multi-stage algorithm used to identify H-1B workers in the LinkedIn data and link them to administrative records.

Step 1: Company Name Canonicalization

LinkedIn employer names are standardized to match PERM and LCA records: transliterate to ASCII, convert to lowercase, remove corporate suffixes (`inc`, `llc`, `corp`, etc.), collapse whitespace, and trim. This produces a canonical employer string for exact-match joins.

Step 2: PERM Linkage

LinkedIn profiles are matched to Department of Labor PERM labor certification filings using three criteria:

- **Employer:** Exact match on canonical company name (LinkedIn employer vs. PERM employer of record).

- **Timing:** PERM decision date falls within a window of 90 days before to 36 months after the worker's first U.S. job start date on LinkedIn. The asymmetric window reflects that PERM filings typically occur after employment begins, with processing times of 6–16 months.

- **Occupation:** The worker's first U.S. job SOC code must match the PERM filing SOC code at the 2-digit major group level (e.g., both in SOC 15: Computer and Mathematical). Among candidates matching on employer and timing, exact 6-digit SOC matches are preferred, followed by 2-digit matches; candidates with no SOC overlap are dropped.

When multiple PERM records match a single LinkedIn profile, the algorithm selects the closest match by SOC proximity, then by date distance. The PERM decision date anchors the event study at $q = 0$.

Step 3: LCA Enrichment

Workers not matched to PERM are supplemented with Labor Condition Application (LCA) records filed for H-1B positions. The matching criteria are:

- **Employer:** Exact match on canonical company name.

- **Timing:** LCA start date falls within ± 180 days of the worker's first U.S. job start date.

- **Occupation:** 2-digit SOC match between the LCA and the worker's first U.S. job.

LCA matches provide an indicator of H-1B status but do not supply a PERM filing date. Workers with LCA matches but no PERM match are used in the gold set (below) and predictive model but are excluded from the event-study analysis, which requires a PERM date.

Step 4: Origin Classification

Workers are classified by country of origin using pre-U.S. work history on LinkedIn:

- **India:** At least one job located in India before the first U.S. job.

- **China:** At least one job located in China before the first U.S. job.

- **ROW (Rest of World):** At least one job located outside the U.S., India, and China before the first U.S. job.

Chinese workers are excluded from the main analysis because they face intermediate green card backlog durations ($T^* \approx 3$–5 years), making them neither a clean short-constraint nor long-constraint group (see Section 5).

Step 5: Valid H-1B Role Filter

Only workers whose first U.S. job falls in specialty occupation categories are retained as potential H-1B workers. The valid 2-digit SOC codes are:

SOC	Occupation Group
11	Management
13	Business and Financial Operations
15	Computer and Mathematical
17	Architecture and Engineering
19	Life, Physical, and Social Science
25	Educational Instruction and Library
29	Healthcare Practitioners and Technical

Workers outside these SOC groups receive an H-1B probability of zero regardless of other indicators.

Step 6: Gold Set Construction

The predictive model is trained on a "gold set" of workers with high-confidence H-1B status labels:

- **Gold positive:** Workers who (i) have a PERM match or LCA match, (ii) have pre-U.S. work history in India or China, and (iii) are in a valid H-1B role. The India/China restriction ensures that gold positives are overwhelmingly true H-1B workers, since these countries dominate H-1B usage and their nationals rarely qualify for other work visas. Chinese workers are included in the gold set for classifier training (where their high H-1B base rate improves identification) but excluded from the analysis sample due to intermediate backlog durations. Excluding Chinese workers from the gold set and retraining the classifier yields nearly identical predicted probabilities (correlation > 0.99), confirming that the classifier is not sensitive to this choice.

- **Gold negative:** Workers who (i) have no PERM match, (ii) have no LCA match, and (iii) have pre-U.S. work history only in ROW countries (excluding India and China). These are foreign-born professionals with no administrative link to the H-1B program.

The gold set label is binary: 1 for gold positive, 0 for gold negative. If the gold set exceeds 20,000 observations, a random subsample of 20,000 is drawn for computational efficiency.

Step 7: Predictive Model

A logistic regression is trained on the gold set using the following 13 features:

Feature	Definition
has_pd	Has a PERM decision date match (= 1 if PERM-linked)
has_lca	Has any LCA match (= 1 if LCA-linked)
india_hist	Pre-U.S. work history in India
china_hist	Pre-U.S. work history in China
row_hist	Pre-U.S. work history in other countries
cap_season	First U.S. job started in Aug–Nov (H-1B cap season)
summer_start	First U.S. job started in May–Jul (OPT timing)
oct1_start	First U.S. job started in October (H-1B fiscal year)
early_move	First U.S. job lasted < 365 days
top_sponsor	Employer in top decile of H-1B sponsorship intensity
is_stem	First job SOC in Computer/Math (15), Engineering (17), or Science (19)
is_top_metro	First job in a top H-1B metro area (San Jose, San Francisco, Seattle, New York, Boston, Austin, Dallas)
is_high_skill	First job SOC in {15, 17, 19, 25, 29}

The model produces a continuous H-1B probability score for each worker. The analysis sample uses a threshold of ≥ 0.75, retaining workers for whom the model is confident in H-1B status.

Step 8: Student Boost and Enrichment

A post-hoc adjustment boosts likely OPT-to-H-1B converts (model probability 0.50–0.80 with LCA match, May–July start, high-skill occupation, non-India/China origin) to probability 0.85, targeting foreign graduates who transition from OPT to H-1B but are underweighted in the gold positive set. The final sample is enriched with PERM offered wage, detailed SOC codes, and education level from administrative records (96.2 percent match rate).

References

Azar, José A., Ioana Marinescu, Marshall I. Steinbaum, and Bledi Taska, "Concentration in U.S. Labor Markets: Evidence from Online Vacancy Data," *The Quarterly Journal of Economics*, 2020, *135* (3), 1101–1165.

Baker, George, Michael Gibbs, and Bengt Holmstrom, "The Internal Economics of the Firm: Evidence from Personnel Data," *The Quarterly Journal of Economics*, 1994, *109* (4), 881–919.

Bassier, Ihsaan, Arindrajit Dube, and Suresh Naidu, "Monopsony in Movers: The Elasticity of Labor Supply to Firm Wage Policies," *Journal of Human Resources*, 2022, *57* (S), S50–S86.

Berry, Alexander, Elizabeth M. Maloney, and David Neumark, "The Missing Link? Using LinkedIn Data to Measure Race, Ethnic, and Gender Differences in Employment Outcomes at Individual Companies," *NBER Working Paper*, 2024, (32294).

Bound, John, Gaurav Khanna, and Nicolas Morales, "Understanding the Economic Impact of the H-1B Program on the U.S.," in "High-Skilled Migration to the United States and its Economic Consequences," University of Chicago Press, 2017. NBER Working Paper 23153.

Card, David, Jörg Heining, and Patrick Kline, "Firms and Labor Market Inequality: Evidence and Some Theory," *Journal of Labor Economics*, 2018, *36* (S1), S13–S70.

Depew, Briggs, Peter Norlander, and Todd A. Sørensen, "Inter-Firm Mobility and Return Migration Patterns of Skilled Guest Workers," *Journal of Population Economics*, 2017, *30* (2), 681–721.

Doran, Kirk, Alexander Gelber, and Adam Isen, "The Effects of High-Skilled Immigration on Firms: Evidence from H-1B Visa Lotteries," *The Quarterly Journal of Economics*, 2022, *137* (2), 127–198.

Hunt, Jennifer and Bin Xie, "How Restricted is the Job Mobility of Skilled Temporary Work Visa Holders?," *Journal of Policy Analysis and Management*, 2019, *38* (1), 41–64.

Johnson, Janna E. and Morris M. Kleiner, "Is Occupational Licensing a Barrier to Interstate Migration?," *Journal of Human Resources*, 2012, *47* (4), 883–912.

Kroft, Kory, Isaac Norwich, Matthew J. Notowidigdo, and Stephen Tino, "The Labor Market Return to Permanent Residency," *NBER Working Paper*, 2026, (34630).

Lamadon, Thibaut, Magne Mogstad, and Bradley Setzler, "Imperfect Competition, Compensating Differentials, and Rent Sharing in the U.S. Labor Market," *American Economic Review*, 2022, *112* (1), 169–212.

Lazear, Edward P., "Firm-Specific Human Capital: A Skill-Weights Approach," *Journal of Political Economy*, 2009, *117* (5), 914–940.

Madrian, Brigitte C., "Employment-Based Health Insurance and Job Mobility: Is There Evidence of Job-Lock?," *The Quarterly Journal of Economics*, 1994, *109* (1), 27–54.

Manning, Alan, *Monopsony in Motion: Imperfect Competition in Labor Markets*, Princeton, NJ: Princeton University Press, 2003.

— , "Monopsony in Labor Markets: A Review," *ILR Review*, 2021, *74* (1), 3–26.

Marx, Matt, Deborah Strumsky, and Lee Fleming, "Mobility, Skills, and the Michigan Non-Compete Experiment," *Management Science*, 2009, *55* (6), 875–889.

McFadden, Daniel, "Econometric Models of Probabilistic Choice," in Charles F. Manski and Daniel McFadden, eds., *Structural Analysis of Discrete Data with Econometric Applications*, Cambridge, MA: MIT Press, 1981, pp. 198–272.

Naidu, Suresh, Yaw Nyarko, and Shing-Yi Wang, "Monopsony Power in Migrant Labor Markets: Evidence from the United Arab Emirates," *Journal of Political Economy*, 2016, *124* (6), 1735–1792.

Peri, Giovanni, Kevin Shih, and Chad Sparber, "STEM Workers, H-1B Visas, and Productivity in US Cities," *Journal of Labor Economics*, 2015, *33* (S1), S225–S255.

Sorkin, Isaac, "Ranking Firms Using Revealed Preference," *The Quarterly Journal of Economics*, 2018, *133* (3), 1331–1393.

Starr, Evan, J. J. Prescott, and Norman D. Bishara, "Noncompetes in the U.S. Labor Force," *Journal of Law and Economics*, 2021, *64* (1), 53–84.

Topel, Robert H., "Specific Capital, Mobility, and Wages: Wages Rise with Job Seniority," *Journal of Political Economy*, 1991, *99* (1), 145–176.

Wang, Xuening, "US Permanent Residency, Job Mobility, and Earnings," *Journal of Labor Economics*, 2021, *39* (3), 639–671.