Climb or Exit: Employer Learning and Migrant Careers

Preliminary and incomplete. Please do not cite without permission.

Agustín Faure*

March 2026

Abstract

Immigration has become essential for sustaining aging workforces, yet many host countries struggle to retain the workers they admit. I study how employer learning frictions shape the careers and retention of migrant workers. Using German social security data, I document three facts: migrants exit the labor market at elevated rates that converge to locals' after roughly eleven years, they fall progressively behind in wages and occupational complexity within the same establishments, and career stagnation differentially predicts exit for migrants. I develop a dynamic career model with Bayesian learning about worker ability, assignment along a job ladder, and an endogenous exit margin calibrated to these patterns. The estimated model implies that employer signals about migrants are roughly eight times noisier than those about locals, delaying belief convergence and stalling promotion. Counterfactual decompositions suggest that the information friction operates primarily through selection at the exit margin rather than through a direct glass ceiling on promotion: noisy learning leads bad matches to linger and good matches to leave prematurely. In the model, equalizing signal precision makes exit efficient and reverses the observed wage penalty among stayers from -4.5 to $+7.0$ log points.

*ZEW Mannheim; agustin.faure@zew.de. I gratefully acknowledge financial support from the Joachim Herz Foundation. I declare no conflicts of interest.

1 Introduction

Immigration has become a central instrument for sustaining the workforce in aging economies. Across the OECD, the working-age population is projected to shrink through 2060, and GDP per capita growth would slow by roughly 40% without policy intervention (OECD, 2025). With fertility persistently below replacement (Bloom et al., 2024) and welfare systems under pressure (European Central Bank., 2022), industrialized countries increasingly rely on migration to fill labor shortages (Marois et al., 2020; Basso et al., 2025). Germany, which admits hundreds of thousands of workers annually across legal channels, exemplifies this approach.

Yet in many host countries, migrants do not stay. Within five years of arrival, 67% of migrants to Germany and 75% to the Netherlands leave (OECD, 2024). Those who remain face a different problem. Across nine countries, three-quarters of the immigrant–native earnings gap reflects unequal access to higher-paying jobs rather than unequal pay within the same position (Hermansen et al., 2025). In the German administrative data studied in this paper, the pattern is visible inside firms: migrants who arrive with a small wage premium see it erode and reverse as they stagnate at lower rungs of the occupational ladder while locals advance. Among migrants, those who stagnate are more likely to leave, and both patterns converge over the same experience horizon.

Why do migrants leave? This paper proposes that a central part of the answer lies in how employers learn about their workers. When employer signals about migrant ability are noisier—for example, because performance in complex tasks is harder to disentangle from language barriers—beliefs converge more slowly, promotion is delayed, and the outside option becomes attractive. But the mechanism is richer than a simple glass ceiling. If learning is slow, the friction distorts who leaves as much as who climbs: high-ability migrants whose talent goes unrecognized are pushed toward exit, while low-ability migrants whose poor fit has not yet been revealed remain too long. The information friction does not merely stall

careers—it degrades the labor market's ability to sort workers at the exit margin.

To formalize and quantify this mechanism, I develop a dynamic career model with Bayesian learning about worker ability, assignment along a job ladder, and an endogenous exit margin, and calibrate it to German administrative data. The model combines learning from noisy performance signals whose precision varies by worker group and task complexity, an internal career ladder where promotion follows a threshold rule on employer beliefs, search frictions that generate monopsony wage-setting, and an outside option through which migrants compare the value of their German career to alternatives abroad. The framework builds on Pastorino (2024) and the Gibbons and Waldman (1999, 2006) tradition of careers in organizations, extending it to differential information frictions across demographic groups and to endogenous migration decisions.

The analysis uses German social-security records (2% Sample of Integrated Labour Market Biographies, SIAB) linked to establishment data (Establishment History Panel, BHP) for the period 2012–2022. Germany's occupational classification (KldB 2010) records the complexity level of every job, providing a built-in career ladder that I aggregate into three rungs—low, medium, and high complexity. Combined with among the highest migrant exit rates in the OECD, this makes Germany a natural setting for studying how career dynamics and retention interact.[1]

I document three stylized facts. First, migrants exit the German labor market at elevated rates, but the excess relative to locals converges to zero after approximately eleven years of experience. Second, within the same establishments, the initial migrant wage premium erodes and reverses; the probability of holding a medium- or high-complexity position diverges along the same trajectory. Third, within-plant wage stagnation differentially predicts exit for migrants, and this link is concentrated in the first decade—the same horizon over which excess exit operates.

I calibrate the model by simulated method of moments, matching the three empirical

[1]High exit rates are not unique to Germany: across OECD countries, the majority of immigrants who leave do so within the first five years of arrival (OECD, 2024).

patterns jointly. The estimates imply that employer signals about migrants are roughly eight times noisier in variance terms.[2] After fourteen years of experience, employers have resolved 80% of their initial uncertainty about a local's ability but only 34% about a migrant's. Migrants also accumulate general human capital roughly 28% more slowly, contributing a growing wage shortfall alongside the learning penalty. The outside option interacts with noisy learning to weaken exit selection: at entry, exit is nearly ability-blind, and even late in the career a significant fraction of exiters are workers the occupation should have retained. The calibrated model reproduces the qualitative patterns in the data—converging exit rates, diverging career trajectories, and the stagnation–exit link—though some moments are matched more closely than others, leaving room for refinement through additional moment conditions or greater model flexibility. The counterfactual magnitudes reported below should therefore be read as illustrative of the mechanism rather than as precise point estimates.

Counterfactual decompositions deliver the paper's main finding. The direct allocation effect—the glass ceiling that delays promotions—accounts for less than one percent of the observed wage penalty. The dominant channel is selection at the exit margin. Under noisy learning, the labor market sorts poorly: low-ability migrants linger because their beliefs stay imprecise and wages hover near the pooling level, while high-ability migrants receive unlucky signal draws, go unrecognized, and leave. Equalizing signal precision leads to *more* exit—the voluntary exit rate rises from 13% to 21%—but exit becomes efficient: it is the low-ability workers who now leave, while the high-ability ones stay and advance. As a result, the observed wage penalty reverses from roughly -4.5% to $+7\%$, revealing positive ability selection among migrants that the friction masks. The cost of the information friction is not slow promotions—it is wrong exits.

This paper relates to four strands of literature. First, a large body of work studies immigrant wage assimilation and return migration. Chiswick (2001) and Borjas (1985, 1995)

[2]The noise multiplier is $R = 2.78$, yielding a variance ratio of $R^2 \approx 7.7$. This translates into Kalman gains that are six times lower at the entry rung and eleven times lower at the top rung; see Section 5 for details.

established the empirical regularity of immigrant earnings convergence and showed that cross-sectional estimates can be biased by changes in cohort quality. For Germany, Lehmer and Ludsteck (2015) decompose wage assimilation and find that firm-specific human capital and search gains each account for roughly 40% of the catch-up, with general human capital explaining only a fifth; alternative explanations emphasize monopsonistic wage-setting (Hirsch and Jahn, 2015) or show that unconditional income gaps can widen rather than narrow at the population level (Berbée and Stuhler, 2025). A parallel literature models return migration as a decision shaped by location preferences, planned horizons, and investment incentives (Dustmann, 1997; Dustmann and Görlach, 2016). It is well established that high-ability migrants can be among those who leave: under the purchasing-power motive, high earners return earlier (Dustmann, 2003), and under the Roy model, return migration intensifies the selection pattern of the initial flow (Borjas and Bratsberg, 1996). Adda et al. (2022) develop a structural lifecycle model that combines ability heterogeneity with persistent location preferences, showing that behavioural selection operates alongside ability selection. Their analysis, however, does not benchmark migrant survival against natives, and the friction lies entirely on the worker's side, not on the employer's. I contribute to this literature by placing the friction on the firm side and connecting employer learning to both career progression and the exit decision.

Second, recent work has begun to study the role of firms in immigrant assimilation. Arellano-Bover and San (2023) show that sorting into higher-paying firms explains a substantial share of the immigrant–native wage gap among former Soviet immigrants in Israel; Gorshkov (2024) finds that job-ladder climbing accounts for roughly a quarter of early wage assimilation in Denmark; and Hermansen et al. (2025) document across nine countries that unequal access to high-paying jobs, rather than unequal pay within the same job, drives three-quarters of the immigrant–native earnings gap. These studies highlight that firm-level processes matter, but focus on sorting *across* firms. Javdani and McGee (2018) study within-firm promotion gaps for immigrants in Canada, but do not connect career progression

to the exit decision. This paper shifts the focus to what happens *within* the employment relationship: how firms learn about migrant workers and how that learning shapes both career progression and the decision to stay.

Third, the information friction at the center of this paper has its roots in the statistical-discrimination literature. Phelps (1972) and Arrow (1973) showed that rational use of group-level information can generate differential treatment in the absence of taste-based prejudice. Farber and Gibbons (1996) derived the empirical implication that if employers learn over time, hard-to-observe correlates of productivity should gain explanatory power for wages as experience accumulates; Altonji and Pierret (2001) confirmed this prediction. Lange (2007) estimated the speed of employer learning and found it to be fast, but estimated a single speed for all workers. Subsequent work refined the empirical framework: Kahn and Lange (2014) jointly estimated learning and productivity dynamics, and Aryal et al. (2022) developed an instrumental-variables approach that separates signaling from learning. Lepage (2024) showed that when employers have less experience evaluating a demographic group, noisier signals generate persistent statistical discrimination without taste-based bias—the mechanism formalized in this paper. I introduce differential learning across demographic groups and connect it to an endogenous exit margin, showing that the quantitative consequences flow primarily through sorting at the exit margin rather than through promotion delays. Coate and Loury (1993) showed that statistical discrimination can become self-confirming when it distorts investment incentives, a channel related to the investment–signal-quality complementarity discussed in Section 7. Lehrer et al. (2025) provide complementary evidence that employer exposure to immigrants reduces differential treatment; Benson and Lepage (2024) show that managers' hiring experiences with demographic groups shape subsequent decisions; and Dodini and Willen (2025) find that immigrant wage discrimination fades with sustained employer–immigrant interaction. These findings are consistent with the view that the friction is informational.

Fourth, the model draws on the personnel-economics literature on careers inside firms.

Baker et al. (1994a,b) documented the basic stylized facts of internal labor markets using personnel records from a single firm; Benson et al. (2019) showed that firms systematically promote based on current-job performance rather than predicted managerial ability; Waldman (1984) showed that job assignments reveal information about ability to outside firms; and Gibbons and Waldman (1999, 2006) developed the theoretical framework for careers in organizations with learning and human-capital acquisition (see also DeVaro and Waldman, 2012). The most direct structural predecessor is Pastorino (2024), who estimates a career-ladder model with Bayesian learning and human-capital acquisition using personnel records from a single firm. I extend her framework in two directions: differential information frictions across worker groups and an endogenous exit margin that connects career dynamics to migration decisions. The exit mechanism builds on Jovanovic (1979): bad matches dissolve as information accumulates, but here the dissolution speed differs across groups and the outside option is emigration. Recent work expands the frontier in complementary directions (Chen et al., 2025; Bircan et al., 2024; Hoffman and Stanton, 2025).

These literatures have developed largely in isolation. Migration economics models the worker's decision but treats the firm as a black box; personnel economics models how firms learn and assign workers but considers a single group in a closed labor market. The experimental evidence of Kaas and Manger (2012), who find that hiring discrimination against migrants in Germany disappears entirely when detailed reference letters are provided, supports the view that the differential treatment is an information problem rather than a preference. This paper bridges the gap by embedding a Bayesian-learning career model in a setting with endogenous migration.

The contribution is threefold. First, I document three linked empirical regularities—excess exit convergence, within-plant career divergence, and the stagnation–exit link—that jointly characterize the migrant career gap. To the best of my knowledge, the convergence of excess exit rates and the link between within-establishment career stagnation and exit have not been characterized. Second, I show that a single friction—differential signal precision—can

generate all three patterns in a calibrated structural model, without requiring taste-based discrimination, skill differences at entry, or differential location preferences. Third, and most important substantively, the counterfactual decomposition suggests that the dominant channel is not the direct allocation effect but distorted sorting at the exit margin: noisy learning pushes out high-ability migrants whose talent goes unrecognized, while retaining poor matches. Equalizing signal precision does not reduce exit—it makes exit efficient, so that low-ability workers leave and high-ability ones stay and advance. The policy implication shifts from targeting promotion barriers to improving signal quality—through credential recognition, structured performance evaluation, and language investment (Dustmann, 1994).

The remainder of the paper is organized as follows. Section 2.2 describes the institutional setting, data construction, and the three stylized facts. Section 3 develops the model. Section 4 specifies functional forms, describes the identification strategy, and presents the calibration procedure. Section 5 reports parameter estimates and model fit. Section 6 presents counterfactual decompositions. Section 7 concludes.

2 Setting, Data, and Stylized Facts

2.1 Institutional Setting

Germany is a natural setting for studying how information frictions shape migrant careers. The country combines three features that this paper exploits: an employer-reported occupational classification that records task complexity for every job (Section 2.1.3), a credential system centered on the nationally standardized apprenticeship that creates a sharp contrast with foreign qualifications (Section 2.1.2), and among the highest migrant exit rates in the OECD (Section 2.1.4). The observation window 2012–2022 is dictated by the availability of consistent occupational coding in the administrative data (Section 2.2) and spans a period of rising migrant inflows and large humanitarian arrivals in the mid-2010s.

2.1.1 Migrant Composition and Entry Regimes

During the observation window, migrants entered Germany through three legal regimes that structure the composition of the migrant workforce: intra-EU free movement, purpose-tied residence permits for third-country nationals, and humanitarian protection.

EU nationals can reside and work in Germany without a work permit and with equal treatment in access to jobs and pay. They constitute approximately 42% of migrants in the sample (Section 2.2), with Central and Eastern European nationals forming the largest EU subgroup — reflecting the completion of free-movement access for Bulgarian and Romanian citizens in 2014 and Croatian citizens in 2015. Southern EU migrants (Greek, Italian, Spanish, Portuguese) form a second substantial group, with origins in the guest-worker recruitment of the 1960s–70s and sustained inflows through family reunification.

Non-EU migrants enter through purpose-tied permits (for employment, education, or family reunification) or through the asylum system. Turkish nationals, the largest single non-EU group in the sample, reflect the same guest-worker legacy as Southern EU migrants but face stricter legal constraints on mobility and residence. The remaining non-EU migrants are heterogeneous: they include labor migrants screened through employment-visa channels, family reunification entrants who were not selected on labor-market criteria, and refugees who entered the labor market only after delays imposed by the asylum procedure. Appendix A.2 reports the full origin-group classification and Table 8 reports the distribution.

2.1.2 What Employers Know and How They Learn

The career dynamics studied in this paper depend on how quickly employers learn about a worker's productive ability. Two features of the German institutional environment suggest that this learning process may differ systematically between locals and migrants: the legibility of credentials at the point of hire, and the role of language in on-the-job evaluation as tasks grow more complex.

Credential legibility. German employers rely heavily on formal credentials to assess workers at entry. For locals, the most common credential is the occupation-specific certificate produced by the dual apprenticeship system (*Ausbildung*): a nationally standardized qualification that certifies two to three years of structured firm-based training. Beyond its role in human-capital formation, the *Ausbildung* also serves as an extended screening period — training firms observe apprentices performing occupation-specific tasks for years before deciding whether to retain them. The resulting credential is highly legible: it tells a prospective employer, with reasonable precision, what tasks the worker can perform. Foreign qualifications carry no such standardization. Curricula differ, assessment standards are unfamiliar, and professional references are harder to verify, so the employer must rely more heavily on on-the-job observation to learn what a migrant worker can do.

Since 2012, the Recognition Act (*Anerkennungsgesetz*) provides a standardized procedure for assessing foreign professional qualifications. Brücker et al. (2021) show that obtaining full recognition raises immigrant wages by roughly 20% and closes the native–immigrant earnings gap, confirming that credential non-recognition is a quantitatively important friction. However, uptake remains concentrated in regulated professions (health care alone accounts for nearly 70% of applications), and fewer than half of applicants receive immediate full equivalence. For the majority of migrants in unregulated occupations, formal recognition is neither required nor sought. Because the observation window starts in 2012, the sample mixes workers who entered before and after the Recognition Act.[3]

None of this implies that German employers cannot identify talented migrants. In the data, migrants earn *more* than locals at entry within the same establishment and initial occupation (Section 2.3.2), suggesting that firms screen migrant applicants more aggressively and hire positively selected individuals. The credential gap is therefore not about whether firms can spot good migrants but about how much residual uncertainty remains after that initial screening, and how quickly it is resolved through subsequent on-the-job observation.

[3]Entry-cohort fixed effects in the empirical specification (Section 2.3.2) absorb level shifts associated with this reform.

Language, workplace culture, and task complexity. If credentials shape what the employer knows at the point of hire, language and familiarity with workplace norms may shape how quickly the employer learns thereafter. At lower-complexity positions (*Helfer*), tasks tend to be manual and directly observable: output is countable, errors are visible, and performance assessment requires little verbal exchange or cultural context. As tasks grow more complex — coordinating teams, writing reports, navigating regulatory environments, managing client relationships — evaluation becomes harder to separate from communication style, professional conventions, and institutional know-how. For a worker operating in a second language and an unfamiliar professional culture, the employer faces an inference problem: does a less polished report reflect lower ability, or a language barrier and unfamiliarity with local conventions orthogonal to skill? Dustmann (1994) documents that writing fluency has a larger effect on migrant earnings in Germany than speaking fluency, consistent with the idea that the informational cost of these barriers is concentrated in positions where written output matters for evaluation. More broadly, Peri and Sparber (2009) show that immigrants systematically specialize in manual rather than communication-intensive occupations, consistent with a comparative disadvantage in roles where language and cultural fluency are central to both production and evaluation.

If the difficulty of evaluating a migrant's performance grows with the complexity of the assigned task, then the informational cost of promotion is higher for migrants than for locals. Section 3 formalizes this channel; the descriptive evidence in Section 2.3.2 documents its empirical counterpart.

Finally, evidence from a German field experiment supports the view that the differential treatment of migrants in hiring reflects an information problem rather than a preference. Kaas and Manger (2012) find that callback discrimination against Turkish-named applicants disappears entirely when the application includes a detailed reference letter — that is, when better information about productivity is directly available. Employers appear to use observable characteristics as proxies for unobserved ability precisely when more informative

signals are lacking.

2.1.3 Occupational Hierarchy and Career Progression

The model in Section 3 represents careers as movement along an internal ladder ordered by task complexity. This abstraction maps onto a concrete feature of how German firms organize work. The German occupational classification (KldB 2010), which is reported by employers to the social-security system and available in the administrative data used in this paper (Section 2.2), assigns every job a *requirement level* (*Anforderungsniveau*) that distinguishes four tiers of task complexity: *Helfer* (unskilled/helper), *Fachkraft* (skilled), *Spezialist* (specialist), and *Experte* (expert). In addition, the classification identifies supervisory and managerial roles through a separate subgroup code, capturing leadership responsibilities that cut across complexity levels (Schuster and Strahl, 2019). These levels are not merely statistical categories. In firms covered by collective bargaining agreements — roughly half of all employees in Germany (Jäger et al., 2022) — requirement levels map directly to pay grades (*Entgeltgruppen*), so that moving from *Fachkraft* to *Spezialist* typically entails a discrete wage jump defined by the collective agreement. Even in uncovered firms, the KldB hierarchy reflects genuine differences in task content, autonomy, and responsibility (Paulus and Matthes, 2013): a *Fachkraft* performs routine professional tasks within established procedures, while a *Spezialist* is expected to develop solutions to non-routine problems, and an *Experte* designs processes or manages projects with substantial discretion.

The descriptive evidence in Section 2.3.2 confirms that within-establishment transitions along these levels are sufficiently common to make the career-ladder framework empirically relevant: the growing migrant–local gap in the probability of holding a mid or high position is a central feature of the data. Section 2.2 describes the construction of the ladder measure and Section 3 formalizes the assignment process.

2.1.4 Outflows and Retention

The model in Section 3 features an endogenous exit margin: migrants compare their German career prospects to an outside option and leave when the outside option dominates. Two things matter for interpreting this margin: what "exit" means in the administrative data, and what institutional forces shape the exit decision.

In the social-security records used in this paper, an exit is defined as permanent disappearance from the German employment registry (Section 2.3.1). This can reflect several distinct outcomes: return migration to the country of origin, onward migration to a third country, transition to self-employment or civil-servant status (both outside the social-security system), or withdrawal from the labor force. The data do not distinguish between these destinations. What they do establish is that the exit is permanent within the observation window — individuals who leave never reappear. For migrants, the most common interpretation is emigration, but the outside option in the model should be understood broadly as any alternative to continued formal employment in Germany.

Cross-country evidence supports the interpretation that administrative exits largely reflect actual emigration. Among European OECD countries, Germany has the second-highest five-year exit rate at 67%, behind only the Netherlands at 75% (OECD, 2024).[4]

Two institutional features shape exit asymmetrically for migrants and locals. First, EU citizens face no legal barrier to leaving and re-entering Germany, which lowers the cost of exit and makes the decision more reversible in principle — though in the data, exits are empirically permanent. For non-EU migrants, the decision is shaped by residence permit conditions. Many permits are tied to employment: losing a job can trigger a limited period to find new employment before the permit expires, which means that an exogenous separation (a plant closure, a layoff) can force exit even when the worker would prefer to stay. This creates a channel through which the exogenous separation rate in the model may have differential

[4]Forced removals and assisted voluntary return account for only a small fraction of these outflows. The vast majority of exit is spontaneous. EU migrants are especially mobile: on average across European OECD countries, only 48% of EU immigrants remain after five years of residence.

consequences for migrants, beyond the voluntary exit margin.[5]

Second, the outside option itself differs between migrant groups. A migrant from an EU country with similar wage levels (Austria, the Netherlands) faces a relatively attractive outside option even if their German career is progressing well. A migrant from a lower-income country of origin may find exit unattractive despite career stagnation, because the wage gap between Germany and the alternative is large. The model captures this variation in reduced form through the outside-option parameters (Section 4), but it does not model the destination choice or origin-country heterogeneity.[6]

2.2 Data and Sample Construction

I use German administrative microdata from the Institute for Employment Research (IAB) Research Data Centre (FDZ). The core individual-level dataset is the Sample of Integrated Labour Market Biographies (SIAB),[7] a 2% random sample of the Integrated Employment Biographies (IEB). The SIAB contains employer-reported employment histories with start and end dates, daily wages subject to social security reporting, and a set of worker characteristics.[8] In addition, the SIAB records non-employment episodes administered by the Federal Employment Agency, including unemployment benefit receipt under SGB III and SGB II,[9] registered job search, and participation in labour-market programmes (Schmucker, 2025). I link jobs to establishments using the Establishment History Panel (BHP). The BHP aggregates the same employment registers to establishment-by-year cross-sections at the

[5]The model separates exogenous separations $\delta(t)$, which are common to both groups, from endogenous exit, which is migrant-specific. In practice, some fraction of what appears as exogenous separation for non-EU migrants may trigger involuntary exit from Germany. This ambiguity is a limitation of the administrative data.

[6]Retention rates vary substantially by region of origin: across European OECD countries, immigrants from Latin America and North Africa have the highest five-year retention rates (around 78%), while those from North America have the lowest (43%) (OECD, 2024).

[7]I use the SIAB 1975–2023, version 7523 v1.

[8]The SIAB also includes marginal employment spells from 1999 onward.

[9]In the German Social Code, SGB III (*Arbeitsförderung*) governs insurance-based unemployment benefits and active labour market programmes administered by the Federal Employment Agency. In the SIAB this appears in the *Leistungsempfängerhistorik* (LeH). By contrast, SGB II (*Grundsicherung für Arbeitsuchende*) provides means-tested basic income support (ALG II in my 2012–2022 window), recorded in the *Leistungshistorik Grundsicherung* (LHG). See Schmucker (2025) for definitions and coverage of LeH and LHG.

June 30 reference date and provides a stable establishment identifier that enables longitudinal linkage and matching to the SIAB records (Ganzer et al., 2024).

The SIAB records individual employment histories as day-exact episodes reported by employers. From these episodes, I construct job spells—continuous employment relationships between a worker and an establishment. If an episode ends but the worker reappears at the same establishment within three months without registered unemployment, I treat the interruption as part of the same spell.[10] For each spell end, I classify the worker's next observed state into five mutually exclusive transition types: (1) *censored* if the spell is ongoing at the panel end, (2) *to non-employment* if the spell is followed by registered unemployment or a gap longer than three months, (3) *job-to-job* if the worker starts a new full-time regular job at a different establishment without an intervening non-employment spell, (4) *exit* if no further social-security records are observed for the individual,[11] and (5) *to part-time* if the next observed job without intervening non-employment is part-time.[12]

I construct a three-rung career ladder from the KldB 2010 occupational classification. Each KldB 2010 code has five digits. The first four identify the occupation (occupational area, subarea, group, and subgroup). The fifth digit records the *requirement level* (*Anforderungsniveau*), which distinguishes four tiers of task complexity: Helfer (1, unskilled), Fachkraft (2, skilled), Spezialist (3, specialist), and Experte (4, expert) (Paulus and Matthes, 2013). In addition, a value of 9 in the fourth digit identifies supervisory roles (when the requirement level is Spezialist or below) and managerial roles (when the requirement level is Experte). Managing directors and executives have a dedicated code (71104) (Schuster and Strahl, 2019).

[10]This measure follows standard recommendations (vom Berge and Schmucker, 2025) and has also been used as such in the literature (e.g. Hirsch and Jahn, 2015).

[11]This definition captures emigration but also transitions to self-employment, civil service (*Beamte*), or other states outside the social-security system. I cannot distinguish among these destinations in the data. For migrants, emigration is likely the dominant component; for locals, exits are rarer and more heterogeneous. Section 2.3.1 documents the resulting exit patterns.

[12]Classifying transitions to part-time employment is important because, while my model does not target job-intensity choices and thus focuses solely on full-time regular employment, pooling part-time shifts by for example treating them as censored or job-to-job would bias transition probabilities.

I aggregate these categories into three rungs ordered by task complexity. The **low** rung ($j = 0$) comprises Helfer and Fachkraft positions. The **mid** rung ($j = 1$) comprises Spezialist positions and supervisory roles. The **high** rung ($j = 2$) comprises Experte positions, managerial roles, and executives. This grouping preserves the substantive distinction between routine tasks, specialist knowledge, and expert or managerial responsibility while keeping the model tractable (Section 4).

Daily wages are right-censored at the social-security contribution ceiling. I replace censored observations with the FDZ imputed daily wage, which follows the imputation strategy in Drechsler and Ludsteck (2025)[13] and is calibrated for full-time employees and ages 14–64. I deflate all wages to 2020 euros using the CPI from the Statistisches Bundesamt (2025).

A key advantage of the longitudinal IEB is that it allows me to construct biographical variables from complete employment histories. Using each individual's first appearance in the data (observed from 1975 onward) and the subsequent sequence of employment and non-employment spells, I construct job tenure as time since the start of the current active spell, plant tenure as total time worked at the establishment since the first observed spell at the establishment, and experience in Germany as total time employment since first observed IEB appearance (without vocational training periods), following FDZ-IAB guidance (vom Berge and Schmucker, 2025). I classify migration status using an individual's first observed nationality in the IEB and keep this status time-invariant, consistent with the convention in SIAB/IEB-based work (Hirsch and Jahn, 2015). In addition to a binary migrant–local indicator (non-German vs. German at first observation), I also group migrants into nine origin categories tailored to the German context (Appendix A.2). Finally, the FDZ reports schooling in three categories—no vocational training, vocational training, and university education—which I use as a coarse measure of formal skill (Fitzenberger et al., 2006; Schmucker, 2025).

[13]The approach developed by Drechsler and Ludsteck (2025) builds on the daily wages including one-off payments, `tentgelt_bonus`, uses leave-one-out means to mitigate uncongeniality, and relies on updated models for the upper tail.

2.2.1 Sample construction

I restrict the analysis window to 2012–2022. The year 2011 contains a reporting break that affects two key inputs. First, working-time status (full-time vs. part-time) was subject to a reporting change that temporarily increased missing values and distorted within-establishment switches (Ludsteck and Thomsen, 2016). Second, occupational coding transitioned from KldB88 to KldB2010, resulting in mixed coding and many missing entries in 2011.[14] Because my analysis relies on consistent full-time status and KldB2010-based rungs, I begin in 2012. I end the window in 2022 and use the 2023 data to classify post-spell transitions for spells ending in 2022.

To prepare the underlying employment histories, I first construct a spell-level file of non-overlapping full-time regular job spells and merge establishment-year characteristics from the BHP using the establishment identifier.[15] If spells overlap, I keep the spell with the longest tenure and break ties by the higher wage. I restrict the spell file to workers aged 18–54, and I keep workers who first appear in the IEB between ages 18 and 30. Using the BHP, I exclude spells in establishments experiencing major workforce shocks—either a workforce reduction of more than 25% relative to the previous year or a closure—following Hirsch and Jahn (2015).

From the spell file, I construct the main analysis dataset as an unbalanced annual person-year panel. For each worker-year, I select the full-time regular job spell that covers June 30, aligning the worker panel to the establishment-year covariates from the BHP, which are also measured on June 30. All biographical variables (tenure, experience) are timestamped at June 30. The person-year panel is the unit of observation for all regression analyses. The underlying spell file is retained for workers in the panel, including within-year job changes that do not cover June 30, and is used to date exits from full-time regular employment

[14] Only about 60% of KldB88 four-digit codes map uniquely to KldB2010 five-digit codes (Schmucker, 2025; Paulus and Matthes, 2013)

[15] I follow standard SIAB preparation choices; see Stüber et al. (2023) for guidance on handling overlapping spells and related cleaning steps.

(Section 2.3.1).

2.2.2 Sample Composition

My final analysis sample contains 360,021 full-time jobs from 233,829 unique individuals. Of these individuals, 40,762 are classified as migrants (17.4%). The migrant share among full-time workers rises from about 10% in 2012 to about 18% in 2022. Among migrants, 42% hold EU citizenship. Table 8 reports the distribution of countries of origin.

Table 1 describes the composition of the sample. Panel A reports individual characteristics. Migrants are, on average, about three years younger than locals (37.39 vs. 40.62). Migrants also enter the German labor market later on average: they start their careers in Germany at age 23, compared to age 21 for locals. Educational attainment differs sharply across groups. A substantially larger share of migrants has no professional qualification (21.1%) compared to locals (4.9%). This gap mainly reflects differences in vocational training participation, where migrants are about 12 percentage points below locals. Differences in tertiary education are smaller: 32.4% of locals hold a university degree (including universities of applied sciences) compared to 28.2% of migrants.

Panel B reports daily wages and job composition. Migrants earn lower daily wages on average, and 14.7% of local wages are right-censored at the contribution ceiling compared to 8.0% among migrants. Migrants are more concentrated in manufacturing, hospitality, transport, and cleaning; locals in business, law, and finance. The distribution across rungs differs markedly: 72.6% of migrant worker-years fall on the low rung compared to 60.5% for locals, and migrants are underrepresented at both the mid rung (10.7% vs. 18.2%) and the high rung (16.7% vs. 21.4%). The leadership indicator shows a similar gap (4.0% vs. 8.0%).

Panel C reports labor-market histories. Migrants have accumulated less labor-market experience and shorter tenure with their current employer than locals. On average, migrants have about four fewer years of experience and about two fewer years of employer tenure across jobs in the sample.

Table 1: Sample composition

	All	Locals	Migrants
Panel A: Demographics			
Age	40.18 (9.33)	40.62 (9.31)	37.39 (8.92)
Female (%)	33.4	34.3	27.3
Education: vocational (%)	61.0	62.7	50.7
Education: university (%)	31.8	32.4	28.1
Panel B: Job characteristics			
Wage (top-coded)	125.96 (50.90)	128.01 (50.95)	113.08 (48.69)
Censored wage (%)	13.8	14.7	8.0
Wage (imputed)	135.90 (70.39)	138.64 (71.29)	118.77 (61.74)
Occupation segment (%)			
Business, Law, and Finance	26.2	28.3	13.0
Production technology	14.6	14.5	15.4
Transport & Logistics	9.9	8.9	16.5
Manufacturing	9.6	8.7	15.4
Health care	8.4	8.4	8.2
Commerce and Trade	8.1	8.4	6.0
Construction	5.9	5.8	6.7
Social and Cultural Services	5.7	6.0	3.9
Information Technologies	4.7	4.7	4.3
Gastronomy and Tourism	2.1	1.7	4.3
Natural Sciences	1.9	1.9	1.7
Agriculture & horticulture	1.2	1.2	1.6
Safety and Security	1.0	1.0	1.0
Cleaning Services	0.7	0.5	2.0
Rung (%)			
Low (Helfer + Fachkraft)	62.2	60.5	72.6
Mid (Spezialist + Supervisor)	17.1	18.2	10.7
High (Experte + Manager/Exec.)	20.7	21.4	16.7
Leadership (%)	7.5	8.0	4.0
Panel C: Employment histories			
Age at entry	21.20 (3.38)	20.88 (3.18)	23.23 (3.80)
Experience in Germany	14.56 (8.53)	15.14 (8.41)	10.94 (8.37)
Tenure at current employer	6.62 (6.95)	6.88 (7.01)	5.03 (6.33)
Sample size			
Worker-years	1,284,491 (100.0%)	1,107,725 (86.2%)	176,766 (13.8%)
Full-time jobs	360,021 (100.0%)	296,955 (82.5%)	63,066 (17.5%)
Individuals	233,829 (100.0%)	193,067 (82.6%)	40,762 (17.4%)

Notes: The unit of observation is a full-time job-year (worker-year). All moments are unweighted. Individuals may appear in multiple years and are not observed in years without full-time employment.

2.3 Stylized Facts

I now turn to three empirical patterns that motivate and discipline the structural model developed in Section 3. The first documents migrant excess exit and its convergence with

experience. The second shows that within the same establishments, migrants fall behind in wages and occupational complexity. The third links career stagnation to the exit decision.

2.3.1 Migrants Exit at Higher Rates — But Only Early On

I study permanent exits from the German social-security administrative records among workers observed in full-time regular employment. I define an exit as the end date of an individual's last observed spell in the administrative data, provided that this end date occurs at least twelve months before the end of the observation window. I restrict the analysis window to 2012–2022, and use 2023 only to implement this twelve-month non-return rule. Temporary interruptions followed by re-entry are therefore not coded as exits.

Time is measured as cumulative labor-market experience in Germany since the first observed employment spell. The experience clock advances only during observed employment. Individuals enter the risk set when they first hold a full-time regular job, which implies delayed entry in experience time.

Figure 1: KM Survival Function

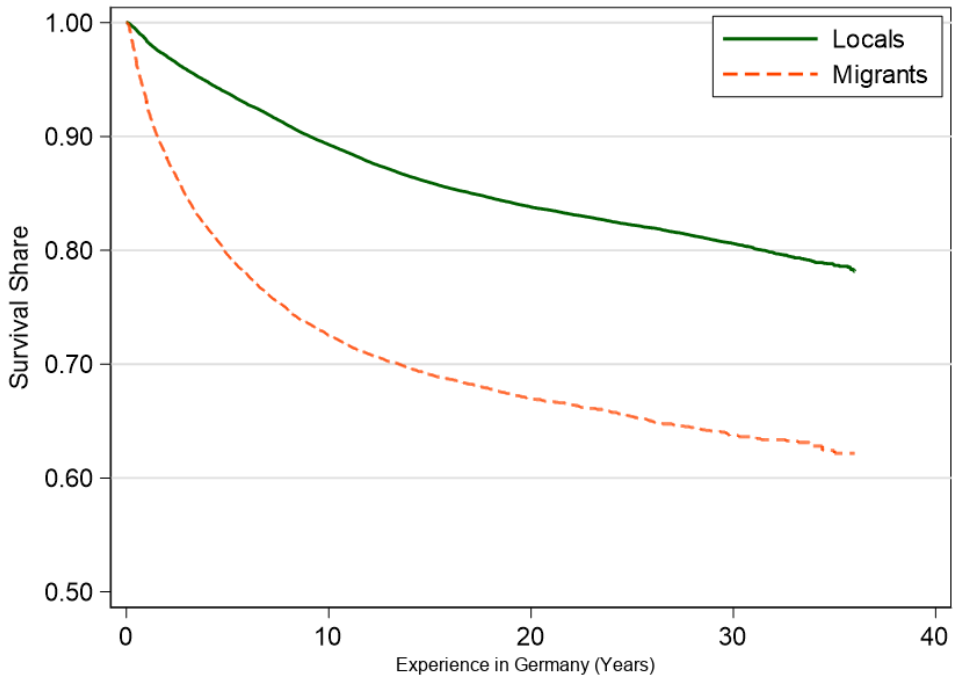


Figure 1 plots Kaplan–Meier survival curves using experience as the analysis time (Table 9 in Appendix B.1 reports the underlying risk set; Figure 9 in Appendix B.2 plots the corresponding hazard functions). Over the 2012–2022 window, migrants are substantially more likely to exit the German social-security data. By ten years of experience, 27.5% of migrants have left compared to 10.7% of locals. These exits represent a permanent loss from the administrative records — individuals who leave never return within the observation window.

To quantify the experience profile of the exit differential and assess its robustness to composition, I estimate a linear probability model of permanent exit on migrant status interacted with experience dummies:

$$\text{Exit}_{it} = \sum_{k} \alpha_k \cdot \mathbf{1}[\text{exp}_{it} = k] \times \text{Migrant}_i + \beta_1 \text{age}_{it} + \beta_2 \text{age}_{it}^2 + \delta_t + \eta_c + \mathbf{x}_i' \boldsymbol{\beta} + \varepsilon_{it} \quad (1)$$

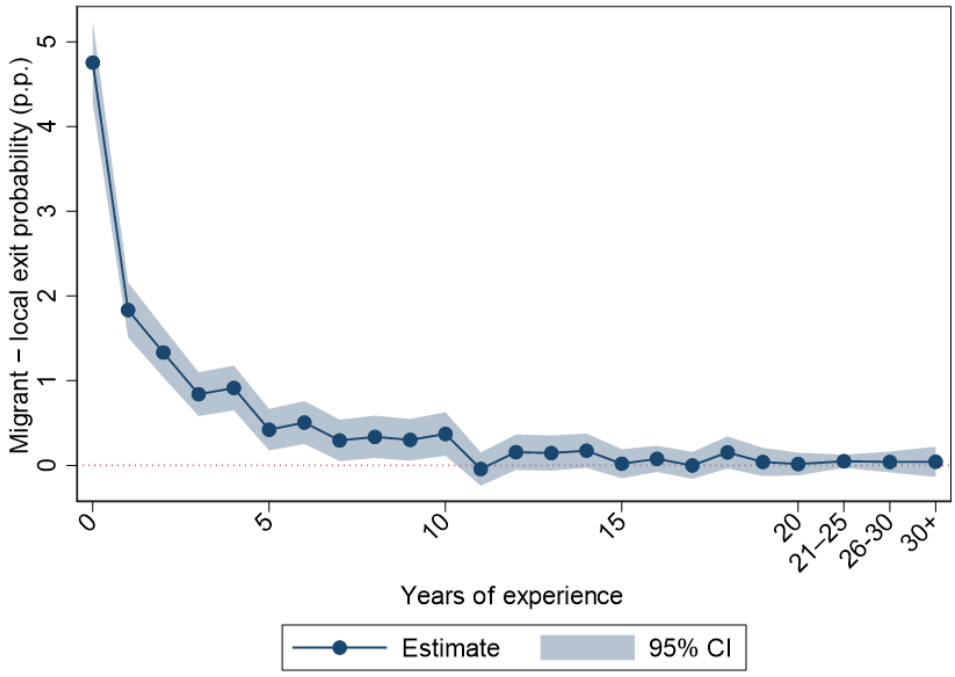
where δ_t and η_c are year and entry-cohort fixed effects, age enters as a quadratic, and $\mathbf{x}_i$ includes education and gender. The coefficients α_k trace the migrant excess-exit probability at each level of experience, net of compositional differences in age, education, gender, and cohort.

Figure 2 plots the $\hat{\alpha}_k$ sequence. The migrant excess-exit probability is large and statistically significant in the first years of experience, then declines steadily. From eleven years of experience onward, the differential collapses to zero. It does not merely attenuate, it vanishes entirely and remains flat thereafter. The pattern is robust to the inclusion of establishment fixed effects, though the level of the gap is attenuated (from roughly 5 to 3 percentage points at onset), indicating that part of the differential reflects sorting into higher-turnover firms while the time profile is a within-firm phenomenon.[16]

To assess the economic magnitude and the prevalence of the exit gap across occupations,

[16]I cannot distinguish permanent emigration from permanent withdrawal into self-employment, informal work, or non-participation. What I can establish is that migrants leave the registry of regular full-time employment at higher rates, despite arguably lower outside options: restricted access to civil-service jobs, weaker self-employment networks, and — for non-EU nationals — visa conditions that make labor-market attachment valuable.

Figure 2: Excess Exit Probability



I complement the LPM with a complementary-log-log (cloglog) discrete-time hazard model. The cloglog specification is better suited than the LPM for this purpose: it maps directly to a continuous-time proportional hazard, which allows me to compute model-implied survival probabilities and cumulative retention measures from the estimated hazard. The specification follows the same structure as equation (1), interacting migrant status with baseline occupational segment.

Table 2 reports two objects derived from the model. The first three columns show $\widehat{S}(10)$, the model-implied probability of remaining in the data at ten years of experience, separately for locals, migrants, and their gap. The first row shows the adjusted survival function for all sectors. Locals' estimated adjusted survival at ten years, 91.9%, barely differs from the raw Kaplan–Meier estimate of 89.3% (Table 9). For migrants, the adjusted figure is 82.5% — ten percentage points above the raw estimate of 72.5% — indicating that observable composition accounts for a substantial share of the level difference. The adjusted gap nevertheless remains −9.34 percentage points.

Table 2: Retention gaps by baseline occupational segment

	$\widehat{S}(10)$			RMST$(0,10)$				
Occ. segment	Locals	Migrants	Gap (pp)	Locals	Migrants	Gap (yr)	N	Migr.%
All segments	91.9	82.5	−9.34***	9.55	8.87	0.68***	131,144	23.5
	(0.37)	(0.80)	(0.95)	(0.03)	(0.06)	(0.07)		
Business, Law & Finance	90.8	78.2	−12.55***	9.48	8.58	0.91***	31,245	12.8
	(0.63)	(1.99)	(1.99)	(0.04)	(0.15)	(0.15)		
Production Technology	95.6	84.2	−11.39***	9.75	8.98	0.78***	15,376	23.5
	(1.02)	(2.00)	(2.29)	(0.06)	(0.15)	(0.16)		
Health Care	92.0	90.1	−1.88	9.55	9.36	0.19	14,259	20.7
	(1.08)	(1.54)	(1.98)	(0.07)	(0.11)	(0.13)		
Transport & Logistics	94.6	86.1	−8.56***	9.69	9.08	0.61***	12,808	40.2
	(1.04)	(1.46)	(1.80)	(0.07)	(0.10)	(0.12)		
Manufacturing	94.2	86.9	−7.30***	9.67	9.14	0.53***	11,934	33.0
	(1.16)	(1.67)	(1.82)	(0.07)	(0.12)	(0.13)		
Social & Cultural Serv.	75.7	74.0	−1.78	8.60	8.29	0.32	10,023	15.1
	(1.51)	(3.20)	(3.91)	(0.12)	(0.25)	(0.30)		
Commerce & Trade	94.9	86.5	−8.47***	9.71	9.11	0.60***	9,843	17.3
	(0.96)	(2.48)	(2.81)	(0.06)	(0.18)	(0.20)		
Construction	92.9	80.7	−12.23***	9.60	8.74	0.86***	6,893	30.4
	(1.81)	(2.41)	(2.87)	(0.11)	(0.17)	(0.20)		
Information Technologies	93.9	82.4	−11.52***	9.66	8.86	0.80***	6,074	23.4
	(1.33)	(3.13)	(3.52)	(0.08)	(0.23)	(0.25)		
Gastronomy & Tourism	92.2	77.8	−14.35***	9.55	8.53	1.02***	5,280	40.9
	(1.89)	(2.45)	(3.19)	(0.12)	(0.18)	(0.23)		
Natural Sciences	93.6	80.7	−12.96***	9.65	8.75	0.90***	2,499	19.8
	(2.24)	(4.65)	(4.92)	(0.14)	(0.34)	(0.35)		
Agriculture & Horticulture	91.2	68.9	−22.33***	9.51	7.93	1.57***	2,159	32.8
	(3.39)	(3.48)	(4.70)	(0.22)	(0.27)	(0.33)		
Safety & Security	87.9	89.8	1.87	9.32	9.34	−0.02	1,425	24.1
	(4.30)	(5.97)	(7.64)	(0.28)	(0.42)	(0.52)		
Cleaning Services	94.1	84.8	−9.35*	9.67	9.00	0.66**	1,326	55.8
	(4.28)	(3.24)	(5.04)	(0.27)	(0.23)	(0.34)		

Notes: $\widehat{S}(10)$ is the model-implied probability (%) of remaining in the data at ten years of experience. RMST$(0,10)$ is the restricted mean survival time (years) over the first ten years. Both are computed from predicted annual exit probabilities under counterfactual migrant status from a complementary-log-log model, averaged within segment × experience-year cells and aggregated over $t = 1, \ldots, 10$. Bootstrap standard errors (in parentheses) are clustered at the individual level. N is the number of individuals. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ (two-sided).

To interpret the economic magnitude of this gap, I turn to the restricted mean survival time (RMST), reported in the next three columns. The RMST gives the expected number of years a worker remains in full-time employment over the first decade. The migrant–local RMST gap measures cumulative experience lost to excess migrant exit: a gap of 0.68 years implies that for every 100 new migrant workers entering full-time employment, approximately 68 person-years of labor-market experience are lost relative to comparable locals over the first ten years.

This number, however, hides heterogeneity across sectors. The remaining rows in Table 2

display the results dividing individuals by their first observed occupation category (in the window of observation, 2012–2022). The excess exit gap is significant in eleven of fourteen occupational segments, spanning white-collar professions (Business, Law, and Finance: 12.6pp; Natural Sciences: 13.0pp), technical occupations (Production Technology: 11.4pp; IT: 11.5pp), manual work (Construction: 12.2pp; Manufacturing: 7.3pp), and services (Gastronomy: 14.4pp). Importantly, the average gap is being driven down by some sectors that do not display a gap: Only Health Care, Social and Cultural Services, and Safety and Security show statistically insignificant gaps — segments where, respectively, institutional credentialing may partly substitute for on-the-job evaluation, baseline attrition is high for both groups, or sample sizes are small.

The convergence is the first key fact of this paper. It implies that whatever drives excess migrant exit is a phenomenon that resolves as experience in Germany accumulates. The next two subsections document what happens to migrants' careers within German firms over this same horizon.

2.3.2 Within-Establishment Career Divergence

What happens to migrants' careers inside German establishments? To answer this, I estimate experience profiles for migrants and locals within the same establishments:

$$\begin{aligned} y_{it} = \quad & \gamma_{g(i)} + \textstyle\sum_{k\neq 0} \alpha_k \cdot \mathbf{1}[\exp_{it} = k] \times \text{Migrant}_i + \sum_{k\neq 0} \delta_k \cdot \mathbf{1}[\exp_{it} = k] + \\ & \psi_{j(i,t)} + \beta_1 \text{age}_{it} + \beta_2 \text{age}_{it}^2 + d_t + \eta_c + \mathbf{x}_i' \boldsymbol{\beta} + \varepsilon_{it} \end{aligned} \tag{2}$$

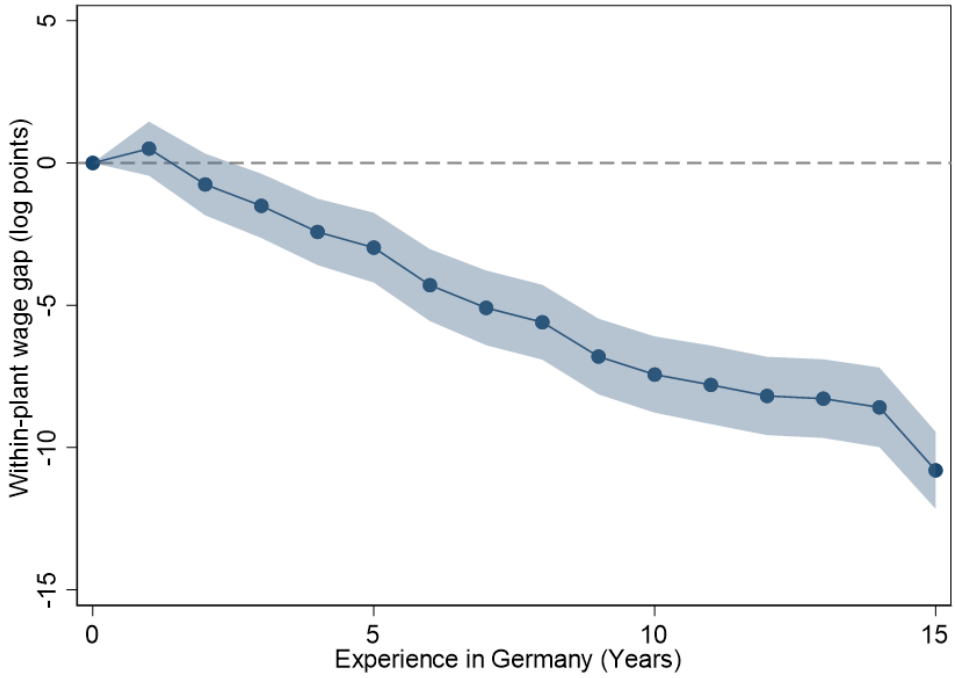
where $g(i) \in \{\text{Migrant}, \text{Local}\}$, $\psi_{j(i,t)}$ is an establishment fixed effect, age enters as a quadratic, d_t and η_c are year and entry-cohort fixed effects[17], and $\mathbf{x}_i$ includes education and gender. The coefficients of interest are the within-establishment experience profile gaps α_k. I deliberately exclude occupation, task complexity, tenure, and leadership status from the controls. These are outcomes of the employer's job-assignment process — the channel I want to study, not

[17]Since experience measures accumulated employment, the identity Entry Year = Year − Experience does not hold. Therefore, I control for entry-cohort fixed effects.

absorb. I apply this specification to several outcome variables: log daily wages, an indicator for holding a mid- or high-rung position, and an indicator for a leadership role (supervisor or management).

Wages. Figure 3 plots the $\hat{\alpha}_k$ sequence for log daily wages. At entry, migrants earn a small and statistically insignificant premium over locals within the same establishment (baseline gap $= 0.003$). This initial near-parity erodes within 3 years of experience and reverses: by fifteen years, the within-plant gap has shifted by approximately -11 log points.

Figure 3: Within-Establishment Wage Gap



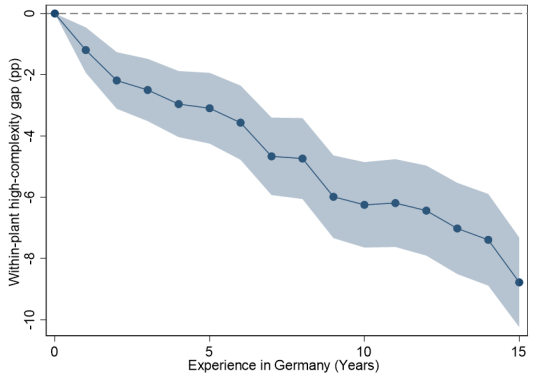
This pattern is robust along several dimensions.[18] Table 3, Panel A shows several additional specifications alongside the baseline. Column (2) adds fixed effects for the initial three-digit occupation and tenure controls. This is noteworthy: the baseline migrant gap turns positive and significant (0.016 log points), indicating that conditional on starting occupation, migrants

[18]The shape of the divergence is also stable across entry cohorts: estimating equation (2) on rolling three-year entry-cohort windows (2012–2014, 2013–2015, and so on) yields consistent profiles, with only precision varying as sample sizes change.

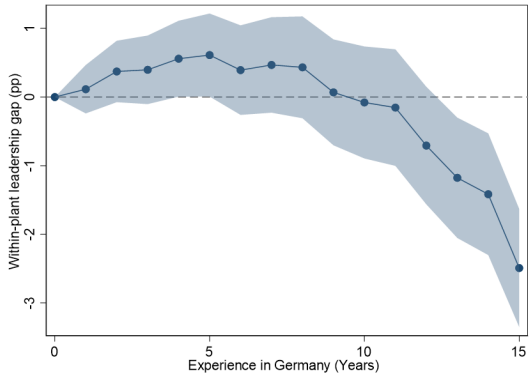
are initially positively selected within establishments. Yet the divergence persists — by fifteen years the total gap (baseline + α_{15}) reaches approximately -7 log points, compared to -11 in the baseline specification. Column (3) replaces education and gender with individual fixed effects, isolating within-person, within-establishment wage dynamics; the divergence is confirmed from approximately five years onward, consistent with the penalty reflecting a dynamic process rather than fixed differences between workers. Column (4) restricts to a balanced panel of workers with at least ten years of experience, ruling out compositional shifts from selective exit; the divergence, if anything, is larger.

Occupational complexity. If the within-plant wage penalty reflects differential access to higher-skilled roles — rather than unequal pay for identical work — it should be visible in occupational assignment. I replace log wages in equation (2) with a binary indicator for holding a mid- or high-rung position. Figure 4a reports the within-plant gap in the probability of reaching a mid or high position. The pattern closely mirrors the wage gap: migrants become progressively less likely to hold mid or high positions within the same establishment, and the gap grows at a comparable rate.

(a) Job Complexity Gap



(b) Leadership Gap



A separate analysis of leadership positions (identified by a 9 in the fourth digit of the KldB 2010 code, indicating supervisor or management roles) reveals a different time profile (Figure 4b): the gap is near zero and statistically insignificant for the first eight to nine

years of experience, then turns negative and widens sharply, reaching approximately −2.5 percentage points by fifteen years.

Panel B of Table 3 reports the corresponding specifications for occupational complexity. The pattern tells a consistent story. In the baseline (column 1), migrants start in significantly lower-complexity positions within the same establishment (baseline gap = −1.9 percentage points). Once initial occupational segment is conditioned on (column 2), this deficit vanishes (baseline gap = 0.6 pp, insignificant) — just as with wages, migrants who enter the same occupation as locals are, if anything, positively selected. Yet the within-plant complexity divergence persists across all specifications, confirming that the wage penalty operates substantially through differential job assignment.

The second key fact is that migrants face a persistent and growing within-plant career penalty. The parallel trajectories of the wage and complexity gaps suggest that job assignment is a substantial channel through which the within-plant wage penalty operates: migrants are less likely to advance to mid and high positions, and the complexity gap grows at a rate consistent with the wage divergence.

2.3.3 Career Progression and the Exit Decision

The previous subsections document two patterns: migrants exit at excess rates that converge after roughly eleven years of experience, and migrants fall progressively behind locals in within-plant career outcomes. A natural question is whether these patterns are related — whether migrants who experience slower within-plant career progression are more likely to leave. I measure recent within-plant career progress as the log-wage change over the past year at the same establishment.[19] I then estimate:

$$\text{Exit}_{it} = \alpha \cdot \Delta \ln w_{it}^{\text{wp}} + \sum_{k} \phi_k \cdot \mathbf{1}[\text{exp}_{it} = k] + \delta_t + \mathbf{x}_{it}'\boldsymbol{\beta} + \varepsilon_{it} \tag{3}$$

[19]This requires that the worker remained at the same establishment for at least one year.

Table 3: Within-plant gaps across specifications

	(1) Baseline Plant FE	(2) Init. occ. + tenure	(3) Person + Plant FE	(4) Balanced (10yr exp)
Panel A: Log-wage gap				
α_5	-0.0297***	-0.0238***	-0.0129**	-0.0567
	(0.0063)	(0.0061)	(0.0054)	(0.0419)
α_{10}	-0.0743***	-0.0613***	-0.0331***	-0.1022**
	(0.0068)	(0.0066)	(0.0066)	(0.0419)
α_{15}	-0.1080***	-0.0889***	-0.0423***	-0.1397***
	(0.0069)	(0.0066)	(0.0074)	(0.0420)
Baseline gap	0.0030	0.0164***	0.0000	0.0352
	(0.0059)	(0.0057)	(0.0000)	(0.0419)
Panel B: Position Complexity Gap				
α_5	-0.0309***	-0.0309***	-0.0228***	-0.0649**
	(0.0059)	(0.0051)	(0.0039)	(0.0293)
α_{10}	-0.0625***	-0.0578***	-0.0337***	-0.0971***
	(0.0071)	(0.0062)	(0.0053)	(0.0300)
α_{15}	-0.0878***	-0.0671***	-0.0260***	-0.1191***
	(0.0075)	(0.0063)	(0.0064)	(0.0308)
Baseline gap	-0.0189***	0.0057	0.0000	0.0139
	(0.0056)	(0.0046)	(0.0000)	(0.0300)
Establishment FE	Yes	Yes	Yes	Yes
Individual FE	No	No	Yes	No
Init. occupation FE	No	Yes	No	No
Tenure controls	No	Yes	No	No
Education, gender	Yes	Yes		Yes
Age, year, cohort FE	Yes	Yes	Yes	Yes
Sample	Full	Full	Full	exp $\geq$ 10
N	1,232,284	1,232,284	1,202,994	963,494

Notes: Each cell reports $\hat{\alpha}_k$, the within-plant migrant–local wage gap at experience levels 5, 10, and 15 relative to the gap at experience 0. Negative values indicate migrants falling behind locals within the same establishment. The outcome is log daily wage. Column (1) is the baseline specification with establishment and year fixed effects, entry-cohort fixed effects, and controls for age (quadratic), education, and gender. Column (2) adds fixed effects for the first observed 3-digit KldB 2010 occupation and tenure-category indicators. Column (3) replaces education, gender, and the migrant main effect with individual fixed effects (absorbed); N is smaller because individuals with only one observation are dropped when person fixed effects are included. Column (4) restricts to workers who accumulate at least 10 years of experience. Standard errors clustered at the establishment level in parentheses.

separately for migrants and locals, and in a pooled specification that interacts $\Delta \ln w_{it}^{\text{wp}}$ with migrant status. The controls include experience fixed effects, year fixed effects, education, gender, and age (quadratic). The coefficient α captures the conditional association between recent within-plant wage growth and permanent exit: among workers at the same career stage, in the same calendar year, and with the same observable characteristics, are those with slower recent wage growth within their establishment more likely to permanently leave the data?

This is a descriptive association, not a causal estimate, and two sources of confounding are worth stating explicitly. First, reverse causality: a migrant who has already decided to leave may disengage — reduce effort, decline opportunities — causing wages to flatten because exit is imminent, not the other way around. Second, common shocks: a bad match or a struggling establishment may simultaneously produce low wage growth and exit, with both driven by the same underlying factor rather than one causing the other. The descriptive evidence cannot fully separate these channels from a direct link running from career dynamics to exit decisions. What it can do, as I discuss below, is establish features of the association that constrain the set of plausible explanations.

Results. Table 4 reports the estimates. For migrants, a one-unit increase in recent within-plant log-wage growth is associated with a reduction in exit probability of 1.4 percentage points. For locals, the corresponding estimate is 0.5 percentage points. In the pooled specification, the interaction between within-plant wage growth and migrant status is negative and statistically significant ($\hat{\beta}^{\text{interaction}} = -0.83$ pp), confirming that the association is differentially strong for migrants. The result is robust to the inclusion of establishment fixed effects, which absorb time-invariant firm characteristics and address the concern that common plant-level shocks drive both wage dynamics and exit.[20]

The most informative feature of the results is how the association varies with experience.

[20]With establishment fixed effects, the identifying variation is across workers within the same firm: among migrants at the same establishment, in the same year, at the same experience level, do those with worse recent wage growth exit at higher rates? The coefficient remains negative and significant, though attenuated.

Table 4: Career progression and permanent exit

	Overall Exit				By experience		
	(1) Mig only	(2) Loc only	(3) Pooled	(4) + FE	(5) Exp 0–5	(6) Exp 6–10	(7) Exp 11+
$\Delta \ln w_{it}^{wp}$	-0.0136***	-0.0048***	-0.0048***	-0.0038***	-0.0299***	-0.0088	-0.0030
	(0.0032)	(0.0009)	(0.0009)	(0.0009)	(0.0067)	(0.0069)	(0.0038)
$\Delta \ln w_{it}^{wp} \times$ Migrant			-0.0083**	-0.0080**			
			(0.0033)	(0.0031)			
Experience FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Education, gender	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age (quadratic)	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Entry cohort FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Exp × Migrant	No	No	Yes	Yes	No	No	No
Establishment FE	No	No	No	Yes	No	No	No
N	114,957	813,070	928,027	900,898	30,012	23,272	61,673

Notes: Outcome is a binary indicator for permanent exit from social-security records. $\Delta \ln w_{it}^{wp}$ is one-year within-plant log wage growth (same establishment in t and t-1). Columns (1)–(4) are overall specifications: migrants only, locals only, pooled with interaction, pooled with interaction + establishment FE. Columns (5)–(7) are migrant-only estimates by experience group (0–5, 6–10, 11+ years). All columns include experience FE, year FE, entry-cohort FE, age controls (quadratic), education, and gender. Standard errors clustered at the individual level in parentheses. *, **, *** indicate $p < 0.10$, $p < 0.05$, and $p < 0.01$.

Estimating equation (3) separately for migrants at different career stages, the coefficient is −3.0 percentage points and highly significant at 0–5 years of experience, but already statistically insignificant at 6–10 years (−0.9 pp) and beyond eleven years (−0.3 pp). Using cumulative two-year within-plant wage growth as the regressor sharpens the time variation: the 6–10 year coefficient becomes marginally significant, consistent with the association extending through the first decade but concentrated in its early years (Table 10 in Appendix B.4).

Interpretation. The concentration of the stagnation–exit link in the early career — and its attenuation toward zero over the same horizon at which excess exit itself vanishes — is the key result, and it helps adjudicate between competing explanations. A disengagement story, in which migrants who plan to leave reduce effort and thereby experience slower wage growth, predicts a stagnation-exit link at all career stages: disengagement before departure is not specific to early-career workers. The concentration of the association in the first years of experience is inconsistent with this explanation operating uniformly. A planned-return

story, in which migrants with predetermined short horizons both underinvest in their German careers and exit on schedule, is harder to distinguish. If all planned returners have horizons below ten years, the association would mechanically vanish at longer horizons as this group exits the sample. This is observationally equivalent to a model in which migrants respond to accumulated career evidence and exit when progress falls short of expectations: both mechanisms predict the same experience-dependent pattern. What the descriptive evidence establishes is the following. Within-plant career dynamics and exit decisions are linked, and this link is substantially stronger for migrants than for locals. The association is concentrated in the early career, aligning with the horizon over which excess migrant exit operates. After eleven years of experience, within-plant wage growth no longer predicts exit for migrants — not because career disadvantages disappear (the within-plant divergence in Section 2.3.2 continues), but because the population of migrants responsive to career dynamics has been depleted through earlier exit. What the descriptive evidence cannot do is separate career-driven exit from predetermined return plans: both generate the same experience-dependent pattern. The structural model in Section 3 provides one framework for quantifying the career-driven channel.

3 Model

Section 2.3 documents three empirical regularities: migrants exit at elevated rates that converge with experience, they fall progressively behind locals in wages and occupational complexity within the same establishments, and career stagnation in the early years predicts exit. This section builds a model in which differential signal precision (the idea that employers learn about migrant ability more slowly) is the central driver of these patterns, working alongside human-capital differences. The mechanism is simple: Firms observe noisy signals of worker ability. When those signals are noisier for migrants, beliefs converge later, promotions are delayed, and wages stagnate relative to locals. As careers stagnate, the option of leaving

Germany becomes attractive.

The model combines four ingredients: (i) Bayesian learning from noisy performance signals, with precision that varies by worker group and task complexity; (ii) an internal career ladder where promotions follow a threshold rule in beliefs; (iii) search frictions that generate monopsony power and a wage markdown; and (iv) an endogenous exit margin through which migrants compare their German career prospects to an outside option. The framework builds on the internal-ladder model of Pastorino (2024) and the learning-and-assignment literature (Gibbons and Waldman, 1999, 2006), extending them to study how differential information frictions shape career divergence and migration decisions.

3.1 Environment

Firms and Workers. Firms are risk–neutral, produce a homogeneous good, and face perfect competition in the output market. The product price is normalised to one. Each firm f operates an internal career ladder with rungs (or "jobs") indexed $j = 0, \ldots, J^f$ and ordered by task complexity and span of control. The only variable input at assignment is the worker's effective labour, a composite of accumulated human capital and latent ability.

Throughout this section I fix a worker i and suppress the index except where needed. Each worker has labor-market experience t and a permanent latent ability θ, drawn once at entry from $N(m_{0g}, \sigma_\theta^2)$, where the prior mean m_{0g} is group-specific. Workers are either locals ($g = L$) or migrants ($g = M$).

Production. A worker's log output in firm f and rung j at period t is

$$y_{fjt} = \phi(j, f, g, t) + c_j \times (\theta + \sigma_{gj}\varepsilon_t) + \eta_{fjt}, \tag{4}$$

where $\phi(j, f, g, t)$ captures all observable, deterministic components of productivity (task-specific baseline productivity at the rung–firm pair (j, f), and human capital accumulated by group g over experience t) while c_j is the productivity leverage of ability at rung j,

$\varepsilon_t \sim \mathcal{N}(0,1)$ is appraisal noise, and $\eta_{fjt} \sim$ EV1 is a transitory productivity shock.[21] The structural assumption is additive separability: latent ability θ, scaled by c_j, enters production separately from observables. Demeaning and scaling therefore yields the belief-relevant signal (Gibbons and Waldman, 2006)

$$z_t = \theta + \sigma_{gj}\varepsilon_t, \tag{5}$$

where σ_{gj} is the signal noise at rung j for group g. Ability leverage is increasing in the ladder ($c_{j'} > c_j$ for $j' > j$), so higher rungs amplify talent. Signal noise is allowed to differ across groups and rungs. Section 4.1 specifies parametric forms for σ_{gj} and all other components of (4).

Information and Bayesian Updating. Let $m_{t+1} \equiv \mathbb{E}[\theta|z_t]$ and $v_{t+1} \equiv \text{Var}[\theta|z_t]$ denote the posterior mean and variance after the period-t appraisal but *before* job assignment in period $t+1$. Conditional on worker migration status g and assigned rung j, the posterior expected ability and variance are

$$m_{t+1} = m_t + K_{gj}(v_t)\big(z_t - m_t\big) \tag{6}$$

and

$$v_{t+1} = v_t(1 - K_{gj}(v_t)). \tag{7}$$

$K_{gj}(v_t)$ is the Kalman filter and is given by $K_{gj}(v_t) = \dfrac{v_t}{v_t + \sigma_{gj}^2}$.

Search frictions and separations. Each period, the worker receives a competing offer with probability $\lambda_g \in (0,1)$, where $\lambda_M \leq \lambda_L$.[22] When a competitor arrives, Bertrand-style wage renegotiation follows; otherwise, the incumbent's posted wage stands. Employment relationships also dissolve exogenously with per-period probability $\delta(t) \in [0,1)$, declining in

[21] The EV1 assumption on η delivers the log-sum object used in the wage formula; see Pastorino (2024).

[22] Migrants are allowed to face thinner markets, consistent with German evidence on monopsonistic frictions (Hirsch and Jahn, 2015).

experience. Section 4.1 specifies a parametric form for δ.

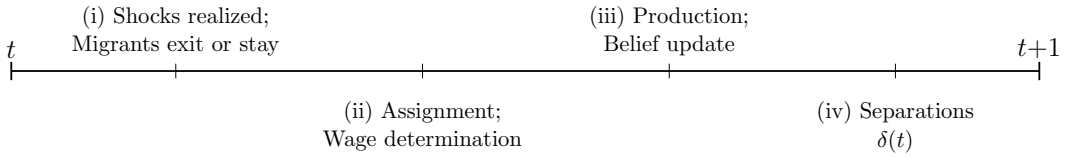
Outside option for migrants. A migrant who exits Germany moves to an absorbing outside state with value

$$V_t^{\text{exit}} \;=\; f(m_t) \;+\; \xi_t, \tag{8}$$

where $f(m_t)$ is a reduced-form flow that depends on the posterior mean ability, capturing Roy-style selection on perceived type (Roy, 1951; Borjas, 1987): migrants whose ability is believed to be high command better opportunities abroad, so the outside value rises with m_t. The taste shock ξ_t is iid Gumbel, independent of German-side shocks, and captures non-pecuniary exit motives (family ties, cultural preferences) orthogonal to career dynamics. Section 4.1 specifies f as a polynomial of degree two in m_t.

3.2 Timing

Figure 5: Within–period timing.



Within each period t, events unfold as follows. (i) Shocks are realized and the public state $s_t = (j_t, m_t, v_t, g)$ is observed. Migrants choose *Exit* or *Stay*. (ii) The incumbent assigns the worker to a rung and posts a wage. With probability λ_g a competitor arrives and triggers renegotiation; otherwise, the posted wage stands. (iii) Production occurs, the appraisal signal z_t is realized, and beliefs update to (m_{t+1}, v_{t+1}). (iv) An exogenous separation occurs with probability $\delta(t)$, and the period finishes.

3.3 Equilibrium

The public state at the start of period t is $s_t = (j_t, m_t, v_t, g)$. Each firm observes s_t and its realized productivity shocks η_t, then chooses a rung assignment and posts a wage. The worker observes all offers (from the incumbent and, with probability λ_g, from a poacher) and accepts the best one. Migrants additionally compare their best German offer to the outside option V_t^{exit}.

I focus on a symmetric, stationary, Markov-perfect equilibrium (MPE) in which strategies depend only on (s_t, η_t). Firms maximize the expected net present value of profits and workers maximize the expected net present value of wages. Beliefs evolve by the Kalman update. As a tie-breaking refinement, whenever a poacher arrives, the hiring wage makes the best non-employing firm indifferent between hiring and not—a "dynamic second price" in the sense of Pastorino (2024). When no poacher arrives, the incumbent posts a certainty-equivalent wage.[23]

The environment inherits the core structure of Pastorino (2024), so Pastorino's efficient-assignment result carries over: equilibrium job assignment maximizes dynamic social surplus (Result 1). My extensions (group-specific search frictions, differential signal noise, and an endogenous exit margin) generate three additional results: promotions follow a cutoff rule in beliefs whose threshold depends on signal precision (Result 2), wages combine a monopsony markdown with competitive renegotiation, and migrant exit is governed by a stay-exit comparison shaped by career prospects. The next two subsections develop these results.

3.4 Job Assignment

A central property of the equilibrium is that job assignment is efficient: in every period, the employing firm and the assigned rung are those that maximize the joint surplus of the match.

[23]The formal strategy spaces, Bellman equations, and the robust-indifference condition are stated in Appendix C.1.

Conditional on the same public beliefs, a migrant and a local are assigned to the same rung and face the same promotion standard.

Result 1 (Efficient assignment)**.** *In any symmetric, stationary, robust MPE, the equilibrium employer and rung assignment at t maximizes the dynamic social surplus—the sum of worker and firm values—over the set of feasible employers. Conditional on a poacher's arrival, both the employer identity and the assigned rung are efficient; without arrival, the incumbent chooses the surplus-maximizing rung among its own rungs.*

The intuition follows Pastorino (2024, Proposition 1): public productivity shocks and dynamic second-price competition ensure that the employing firm and rung generate the highest joint value. The extensions I introduce—group-specific arrival rates, differential signal noise, and the migrant exit margin—affect belief dynamics and outside options but do not alter the within-period assignment mechanism. The formal argument is in Appendix C.1.

Because assignment is efficient, the within-firm promotion decision reduces to comparing surplus across adjacent rungs.

Result 2 (Promotion cutoff)**.** *For any firm f, group g, and adjacent rungs $j \to j{+}1$, there exists a unique cutoff $m^{\star}_{g,j}(v)$ such that*

$$\textit{promote } j \to j{+}1 \iff m \geq m^{\star}_{g,j}(v). \tag{9}$$

The cutoff $m^{\star}_{g,j}(v)$ is weakly increasing in posterior uncertainty v. See Appendix C.2 for the proof.

The cutoff exists because higher rungs load more heavily on ability ($c_{j+1} > c_j$), so the surplus gain from promotion is strictly increasing in the posterior mean and crosses zero exactly once. Where the cutoff sits depends on two conditions on the noise technology, each with a distinct implication for careers. First, signals are noisier at higher rungs ($\sigma_{gj'} > \sigma_{gj}$ for $j' > j$): more complex tasks are harder to evaluate. This creates an informational cost

of promotion—moving a worker up means observing noisier signals—and makes the cutoff decline with experience. Early in a career, posterior uncertainty v is large and the firm prefers to gather more signals before committing to a promotion. As v falls, the option value of waiting diminishes and the cutoff relaxes, generating naturally "slow-release" careers in which promotions accelerate over time.

Second, noisier migrant signals affect promotion timing through three competing forces. First, noisier signals slow belief convergence: the Kalman gain $K_{gj} = v/(v + \sigma^2_{gj})$ is smaller when σ^2_{gj} is larger, so posterior beliefs about migrants evolve more slowly toward true ability and migrants take longer to reach any given promotion threshold. Second, what matters for the threshold itself is the *relative* information loss from promotion—how much precision the firm sacrifices by moving the worker to a higher, noisier rung. A larger relative loss raises the option value of staying and pushes the cutoff up. Third, when signals are noisier *uniformly* across all rungs, the relative loss actually shrinks, because the gap between rungs narrows in proportional terms. This reduces the option value of waiting and pushes the cutoff down. The net effect on migrant–local promotion gaps is therefore theoretically ambiguous: slower convergence delays promotion, but a uniformly noisier environment can lower the threshold that must be reached. Section 4 estimates the model and evaluates which force dominates. In either case, the resulting delay in promotion produces the widening complexity gap documented in Section 2.3.2, and it emerges entirely from the information structure.

3.5 Wages and Migrant Exit

Results 1–2 characterize job assignment taking wages as given. I now close the model by deriving the wage equation and the migrant exit decision. Because the stay value that governs exit is built from the wage stream, the two results are tightly linked: anything that compresses wages also pushes migrants toward exit.

Result 3 (Expected wage)**.** *The expected log wage combines competitive renegotiation (prob-*

ability λ_g) with a certainty-equivalent posting (probability $1-\lambda_g$) and simplifies to

$$\mathbb{E}[w_t] \;=\; \left[\lambda_g + (1-\lambda_g)\,M\right] c_j\, m_t \;+\; h_{gt} \;+\; \varepsilon_t, \tag{10}$$

where h_{gt} denotes markdown-inclusive human capital: the true return from the production technology is also scaled by $[\lambda_g + (1-\lambda_g)\,M]$, but since this factor is common across groups it is absorbed into the estimated human-capital parameters (Appendix C.3); ε_t absorbs transitory wage variation; and

$$M \;\equiv\; \frac{\lambda_g\,\gamma}{1-\gamma(1-\lambda_g)}, \qquad \gamma \equiv \beta\,(1-\delta(t)), \tag{11}$$

is the monopsony markdown.

The derivation, detailed in Appendix C.3, proceeds in two steps. First, consider the two wage regimes. With probability λ_g a competitor arrives and triggers Bertrand renegotiation: competition forces the losing firm to bid up to its own valuation, so the renegotiation wage is pinned by the runner-up's flow output at the assigned rung (equation (38) in the Appendix). When no competitor arrives (probability $1-\lambda_g$), the incumbent posts a certainty-equivalent (CE) wage that makes the worker just indifferent between accepting and searching. The renegotiation wage is competitive and the CE wage reflects monopsony power.

Second, the CE wage depends on the worker's outside option, which has a recursive structure: each period the worker may draw a competing offer (probability λ_g) or wait another period and try again. Under stationarity, this geometric recursion collapses to the scalar markdown M in (11).[24] As $\lambda_g \to 1$, $[\lambda_g + (1-\lambda_g)M] \to 1$ and wages approach the competitive bound—the worker captures the full ability rent. As $\lambda_g \to 0$, $M \to 0$ and the incumbent extracts all surplus.

The wage-setting mechanism combines the monopsony-through-search insight of Burdett and Mortensen (1998) with the Bertrand-competition structure of Postel-Vinay and Robin (2002), in which the renegotiation wage is pinned by the runner-up's valuation and workers

[24] See equations (39)–(44) in the Appendix.

hold no independent bargaining power. What distinguishes the current framework from these predecessors is the interaction with Bayesian learning and an internal career ladder (Pastorino, 2024): wage dynamics arise not only from job-to-job transitions but also from within-firm belief updating and promotion.

Equation (10) identifies three channels through which migrant–local wage differences arise. First, a *search-friction* channel: if migrants face thinner markets ($\lambda_M < \lambda_L$), their markdown is larger and wages are lower holding the state fixed.[25] Second, a *human-capital* channel: group-specific accumulation rates enter the additive term h_{gt} directly. Third, a *dynamic* channel operating through $c_j\, m$: noisier signals slow belief convergence for migrants at every rung (the Kalman gain K_{gj} is smaller when σ^2_{gj} is larger) so the posterior mean m_t evolves more slowly, and delayed promotions keep migrants at lower rungs where the ability return c_j is small.

Result 4 (Migrant exit)**.** *At each period, a migrant compares the lifetime value of remaining in Germany, $W_t^{stay}(s_t)$, to an outside option $V_t^{\text{exit}}(m_t)$ that depends on perceived ability. Both values are convex in the posterior mean m_t. If the exit value is sufficiently convex relative to the stay value, the gap between them is U-shaped in m_t and exit is elevated at both tails of the belief distribution, generating Roy-style two-sided selection in outmigration. See Appendix C.4 for details.*

The exit decision depends on the interplay of two forces. First, the *option value of learning*: early in a career, posterior uncertainty is large and further signals may substantially revise beliefs upward, so migrants stay even when current wages are low. As uncertainty dissipates, the option value shrinks and stagnant careers trigger reevaluation. Second, the *outside option's dependence on beliefs*: because $f(m_t)$ rises with perceived ability, migrants whose signals have been unfavorable (low m_t) find exit unattractive, while those whose true ability exceeds what the German market has recognized may find exit attractive—the outside

[25]The calibration finds this channel to be quantitatively small: arrival rates are near-equal across groups. See Section 4.2 for the parameter choices and supporting evidence.

option rewards their type even if German wages do not yet.

Migrants stay only in the interior of the belief distribution, generating two-sided selection into staying. At the lower tail, workers who are a poor match for the German labor market learn this relatively quickly—even with noisy signals, sufficiently low realizations push beliefs down fast—and exit because their German career prospects are poor. At the upper tail, workers whose ability exceeds what the German market has recognized find that the outside option already rewards their type, so they exit despite having good signals. The two margins operate sequentially in the cross section. Early on, bad matches filter out quickly, generating high early excess exit. Next, a sustained period of elevated exit follows as workers who are good matches but whose careers progress too slowly find that the outside option dominates. Finally, the exit rate converges: the surviving migrants are those in the middle of the belief distribution, for whom the stay value comfortably exceeds the outside option. Within individuals, the model also predicts a period of waiting—while uncertainty is high and the option value of learning keeps migrants in—followed by a rise in exit as beliefs converge and stagnant careers are revealed.[26]

4 Parameterization and Calibration

To take the model to data, each primitive—the noise technology σ_{gj}, the human-capital accumulation rule, the production leverage c_j, and the exit value—must be given a concrete functional form. The guiding principle is parsimony: each free parameter should be disciplined by a distinct block of data moments, yet the specification must be flexible enough to reproduce the joint dynamics of wages, promotions, and exit documented in Section 2.3.2. This section specifies those functional forms, describes which parameters are normalized or set externally and which are calibrated, and lays out the simulated method of moments procedure.

[26]Preliminary evidence is consistent with this within-individual pattern: with entry as the base year, the cumulative migrant–local exit gap starts at zero, increases significantly over the first years of experience, and then plateaus—consistent with workers waiting to learn about their match quality before exiting. A formal test of this prediction is left for future work.

Throughout, $g \in \{L, M\}$ denotes the worker's group (local, migrant), $j = 0, 1, 2$ indexes the career ladder, m_t and v_t are the posterior mean and variance of ability, and t counts years of labor-market experience.

Career ladder. The three-rung career ladder defined in Section 2.2 maps directly onto the model. The low rung ($j = 0$) groups Helfer and Fachkraft positions, the mid rung ($j = 1$) groups Spezialist and supervisory roles, and the high rung ($j = 2$) groups Experte, managerial, and executive roles. Three rungs require only two promotion thresholds and, with the structured noise specification below, four noise parameters instead of six. All workers enter at $j = 0$.[27]

I focus estimation on the first fifteen years of labor-market experience ($t = 0, \ldots, 14$). This is the window over which Section 2.3.2 documents the sharpest divergence in wages and occupational complexity between migrants and locals, and where migrant excess exit is concentrated. Leadership gaps, by contrast, emerge later in the career, when exit is no longer a first-order concern; extending the horizon to capture those dynamics is left for future work.

4.1 Functional Forms

Signal noise technology. The noise technology is the central mechanism of the model: it governs how fast employers learn about each group and, through the Kalman gain, how quickly beliefs converge to true ability. An unrestricted specification would require six free parameters (one signal variance per group-rung pair). I impose a linear structure that reduces this to four while separating two economically distinct channels through which migrants can face noisier signals.

Assign each rung a task-complexity index $\zeta_j \in \{0, 0.5, 1\}$, increasing from low to high.

[27] Transitions between Helfer and Fachkraft are frequent and carry small wage premia; they are treated as within-rung variation absorbed by the measurement error.

The signal variance for locals at rung j is

$$\sigma_{Lj}^2 \;=\; \sigma_{L0}^2 \left(1 + \kappa_L \, \zeta_j\right), \tag{12}$$

where σ_{L0}^2 is the baseline variance at the entry rung and $\kappa_L \geq 0$ governs how fast noise grows with task complexity. For migrants,

$$\sigma_{Mj}^2 \;=\; \sigma_{L0}^2 \, R^2 \left(1 + \kappa_M \, \zeta_j\right), \tag{13}$$

with $R > 1$ and $\kappa_M \geq \kappa_L$. The two migrant-specific parameters capture distinct sources of the learning penalty. The level shifter R scales migrant noise at every rung. It reflects factors such as language barriers or unfamiliarity with credentials that make signals uniformly less informative, regardless of task complexity. The excess gradient $\kappa_M - \kappa_L$ captures how much faster noise grows for migrants as tasks become more complex (for instance, because cultural or institutional knowledge matters more at mid and high rungs). Section 4.3 shows that these two channels are separately identified: within-rung wage gaps pin down R, while cross-group differences in promotion timing pin down the gradient.

Human capital. While the noise technology drives differential *learning*, human capital drives differential *wage growth*. I let human capital accumulate deterministically in total labor-market experience,

$$h_{gt} \;=\; \alpha_g \, t, \tag{14}$$

where $\alpha_L = \alpha_0$ and $\alpha_M = \alpha_0 \, \rho_\alpha$ with $\rho_\alpha \in (0, 1]$. The ratio ρ_α captures any migrant penalty in human-capital accumulation—for instance, from imperfect transferability of pre-migration experience or slower on-the-job learning in a foreign institutional environment.

This specification is deliberately simpler than the stochastic, ability-dependent human capital in Pastorino (2024) and the task-specific human capital in Gibbons and Waldman (2006). In those frameworks, human capital depends on ability and the assigned job, so it

enters the surplus comparison and shapes assignment decisions. Here, h_{gt} depends only on group and experience—not on the rung or on ability—so it shifts the value of every assignment equally and drops out of the promotion comparison. The benefit is that assignment depends on beliefs alone, which keeps the state space tractable and isolates the learning channel as the sole driver of differential promotions. The cost is that the model cannot generate an independent human-capital motive for promotion.[28]

Ability leverage and promotion thresholds. The ability leverage and promotion thresholds determine when evolving beliefs translate into upward mobility. The productivity leverage of ability at rung j is c_j with $c_0 = 1$ (normalization), $c_1 \geq 1$, and $c_2 \geq c_1$. Because c_j is increasing, higher rungs amplify the return to talent: a worker whose ability is believed to be high produces relatively more at a mid or high position than at the low rung. This comparative advantage is the force that makes promotion valuable. The surplus from assigning a worker with posterior mean m to rung j at experience t is

$$\mathcal{S}_j(m, v, t) \;=\; c_j\, m \;+\; \gamma(t)\, \mathbb{E}\big[\mathcal{S}_{j'}(m', v')\big], \tag{15}$$

where $\gamma(t) = \beta\,(1 - \delta(t))$ is the effective discount factor and the expectation integrates over next-period beliefs via the Kalman update. Promotion from rung j to $j + 1$ requires that the surplus gain exceed a threshold,

$$\mathcal{S}_{j+1}(m, v, t) - \mathcal{S}_j(m, v, t) \;\geq\; \kappa_j, \tag{16}$$

with κ_0 governing the low $\to$ mid transition and κ_1 governing mid $\to$ high. The thresholds are a reduced-form representation of organizational frictions to promotion (administrative costs, training investments, or institutional rigidity) beyond the pure surplus comparison.

The interaction with the noise technology is what generates the migrant promotion penalty.

[28] Allowing rung-specific or ability-dependent accumulation rates would add a promotion motive beyond learning but would also expand the state space considerably. I leave this extension for future work.

The surplus gain from promotion grows as the posterior mean m rises, but noisier signals slow posterior convergence. A migrant whose signals are scaled by $R > 1$ accumulates favorable beliefs more slowly than an otherwise identical local, even though both have the same true ability. The migrant therefore crosses the promotion threshold later—producing the widening complexity gap documented in Section 2.3.2—without requiring any migrant-specific promotion friction.

Outside option and exit. The exit margin is what distinguishes migrant careers from local careers in this model. Locals face only exogenous separations; migrants additionally have access to an outside option that locals do not, and the value of that option depends on how the German labor market perceives their ability.[29] This creates a selection channel: migrants who exit are disproportionately those for whom the German career path is insufficiently attractive relative to their outside opportunities. Crucially, this can generate exit from *both* ends of the belief distribution. Workers with low posterior means exit because their German career prospects are poor. Workers with high posterior means may also exit if their beliefs are evolving too slowly—the German labor market has not yet recognized their ability, but the outside option already rewards their revealed type.

A migrant who exits at experience t with posterior mean m_t receives the present value

$$V_t^{\text{exit}}(m_t) \;=\; \left(\mu + \mu_1\, m_t + \mu_2\, m_t^2\right) \;\times\; \sum_{s=0}^{T-t} \beta^s, \tag{17}$$

where μ is a baseline flow, μ_1 and μ_2 govern the Roy-style return to perceived ability, and the summation converts flows into a net present value over the remaining horizon. The quadratic in m_t is what allows the two-sided selection just described: when $\mu_1 > 0$, the outside value rises with beliefs, so that high-ability migrants whose German careers stagnate find exit attractive; when $\mu_2 \neq 0$, the curvature can match the non-linear relationship between beliefs

[29] In practice, exiting the German labor market encompasses return migration, onward migration to a third country, or withdrawal into non-covered employment. I do not model the destination choice; the outside option represents the best alternative to continued formal employment in Germany.

and exit observed in the data.

Exit decisions are smoothed by an iid Gumbel taste shock with scale ξ, following the standard logit framework (Train, 2009, ch. 3). The worker's value at the start of t is

$$W_t \;=\; \xi \,\ln\!\Big(\exp\big(W_t^{\text{stay}}/\xi\big) + \exp\big(V_t^{\text{exit}}/\xi\big)\Big), \tag{18}$$

and the exit probability is $\Pr(\text{exit}) = \big[1 + \exp\big((W_t^{\text{stay}} - V_t^{\text{exit}})/\xi\big)\big]^{-1}$. The scale parameter ξ captures dispersion in non-pecuniary motives for exit (family ties, cultural preferences, risk attitudes) that are orthogonal to career dynamics. Larger ξ smooths exit probabilities, preventing the model from generating knife-edge exit decisions at a single belief cutoff.

Exogenous separations. Independently of the exit decision, matches dissolve exogenously at rate

$$\delta(t) \;=\; \delta_0 \,\exp(-\delta_1\, t), \tag{19}$$

which captures the empirical regularity that separation rates decline with experience. This rate is common to locals and migrants and reflects permanent exits from the administrative data. The differential migrant exit documented in Section 2.3.2 is endogenous and governed by (17)–(18).

Wages. In the symmetric equilibrium, wages combine a monopsony markdown with competitive renegotiation. Let $M \equiv \frac{\lambda\,\gamma}{1-\gamma(1-\lambda)}$ be the markdown from Section 3.5. When a competitor arrives (probability λ), wages equal the competitive bound; otherwise, the incumbent posts the certainty-equivalent markdown. The expected log wage for a worker at rung j with posterior mean m at experience t is

$$\mathbb{E}[w_t] \;=\; \big[\lambda + (1-\lambda)\,M\big]\, c_j\, m \;+\; h_{gt} \;+\; \varepsilon_t, \qquad \varepsilon_t \sim \mathcal{N}(0,\, \sigma_\varepsilon^2), \tag{20}$$

where σ_ε is a measurement-error standard deviation that absorbs transitory wage variation not captured by the model.[30][31]

4.2 Parameters

The model has 23 parameters in total. Three are normalizations that fix scales without loss of generality: the local prior mean $m_{0,L} = 0$ anchors the belief scale, the prior variance $v_0 = 1$ sets the units of uncertainty, and the ability leverage at the entry rung $c_0 = 1$ means that c_1 and c_2 are measured relative to the baseline return to talent. Table 11 in the Appendix lists each parameter, its value or role, and its source.

Five parameters are set before the calibration. The annual discount factor is $\beta = 0.96$. The competitor arrival rate $\lambda = 0.157$ is calibrated from observed job-to-job transition rates and imposed as common across groups.[32] The exogenous separation parameters $\delta_0 = 0.0065$ and $\delta_1 = 0.070$ are estimated from local workers' permanent exit rates. I first estimate a linear probability model of permanent exit on experience dummies with the full set of controls (plant, year, and entry-year fixed effects, demographics, initial occupation), then fit $\delta(t) = \delta_0\, e^{-\delta_1 t}$ to the 15 composition-constant conditional exit rates by nonlinear least squares ($R^2 = 0.96$).[33] The wage measurement-error standard deviation is set to $\sigma_\varepsilon = 0.10$.

The migrant prior mean $m_{0,M}$ is not a free parameter either. At entry, all workers occupy

[30] As shown in Appendix C.3, human capital is also marked down by $[\lambda + (1-\lambda)\,M]$; since this factor is common across groups, it is absorbed into the estimated α_g parameters, so h_{gt} in (20) denotes the markdown-inclusive return.

[31] This simplification follows from the symmetry of firms' technology. The rung-specific productivity shock η_{fjt} and the log-assignment-probability terms from the wage derivation (Appendix C.3) average out across the firm population, leaving the deterministic flow as the wage base. The markdown term $[\lambda + (1-\lambda)M]$ collapses the two wage regimes into a single factor.

[32] The estimate divides the observed annual job-to-job transition rate for locals (7.9%) by an assumed switching probability of one half, consistent with symmetric Bertrand competition between incumbent and poacher. The common-λ assumption is supported by the modest gap between local and migrant J2J rates (7.9% vs. 10.1%); the higher migrant rate is consistent with endogenous differences in acceptance thresholds rather than differential arrival rates (see Appendix D).

[33] The calibration results reported in this chapter were obtained using preliminary separation estimates ($\delta_0 = 0.026$, $\delta_1 = 0.10$) derived from a broader definition of local separations that included transitions to non-employment and part-time work. The final estimates ($\delta_0 = 0.0065$, $\delta_1 = 0.070$) restrict attention to permanent exits, which is the conceptually correct measure. All other targeted moments are virtually unchanged between the preliminary and final data extracts (differences at the fourth decimal place). Re-estimation with the corrected separation rates is in progress.

rung $j = 0$ with $c_0 = 1$, so the initial wage gap is a closed-form function of $m_{0,M}$ and the pre-set parameters λ, β, and δ_0. I pin $m_{0,M}$ at each parameter evaluation so that the model reproduces the observed entry wage gap of 1.6 log points exactly. This is a useful discipline: the calibration starts from the correct wage level at entry, so the optimizer must match the *dynamics* of wages, promotions, and exit rather than adjusting levels to compensate for a misaligned starting point.

This leaves fourteen free parameters for calibration. Four govern the signal-noise technology: the baseline variance σ^2_{L0}, the migrant noise amplifier R, and the complexity gradients κ_L and $\kappa_M - \kappa_L$. Two govern human capital: the local accumulation rate α_0 and the migrant ratio ρ_α. Two set the ability leverage at the upper rungs, c_1 and c_2, governing how much more talent matters at mid and high positions. Two are promotion thresholds, κ_0 and κ_1. And four govern exit: the baseline, slope, and curvature of the outside-option value (μ, μ_1, μ_2) and the Gumbel scale ξ. The next subsection shows how distinct blocks of data moments discipline each of these parameters.

4.3 Targeted Moments and Identification

All moment targets are constructed from within-plant career gaps, fixing the initial occupational allocation.[34] The initial-occupation control ensures that the gaps I target reflect differential career dynamics within the same starting conditions, not differences in occupational sorting at entry. It also reveals a telling baseline pattern. At experience zero, migrants and locals who enter the same plant in the same occupation hold indistinguishable positions on the complexity ladder (Table 13, Panel D: +0.3 pp, insignificant), yet migrants earn 1.6 log points more (Panel A: significant at 1%). This positive initial wage gap—higher perceived ability conditional on the same occupational assignment—is what pins $m_{0,M} > 0$ in the calibration. The moment targets then discipline how quickly this initial advantage erodes as

[34]Specifically, the specification includes establishment fixed effects and fixed effects for the worker's first observed three-digit KldB 2010 occupation. This corresponds closely to column (2) of Table 3, with the difference that I omit tenure controls to avoid conditioning on a partially endogenous variable; the results are nearly identical with or without tenure.

differential learning, promotion, and exit dynamics unfold.

I organize 32 moments into nine blocks, chosen so that each subset of free parameters is disciplined by a distinct block of data variation (see Appendix D.3 and Table 13 for block definitions and target values). With 14 free parameters and 32 conditions, the model has 18 overidentifying restrictions. The identification argument has three pillars.

The local experience-wage profile pins down human-capital accumulation (α_0), and the cross-group wage gap at multiple horizons identifies the migrant HC ratio (ρ_α) and, jointly with promotion timing, the ability leverage parameters (c_1, c_2). The migrant–local gaps in occupational complexity shares—which widen from near zero at entry—discipline the promotion thresholds (κ_0, κ_1) and the noise gradient (κ_L, $\kappa_M - \kappa_L$): slower learning and a steeper noise gradient both delay the point at which migrants cross the promotion cutoffs. Exit rates and their covariance with within-plant wage stagnation identify the outside-option parameters (μ, μ_1, μ_2, ξ): the level and shape of the exit value, and the dispersion of non-pecuniary exit motives.

The key identification insight is that conditioning on the *rung* separates the learning channel from the assignment channel. Two workers at the same rung differ in wages only through their posterior means m_t and their human-capital flows, not through occupational sorting. The within-rung wage gap therefore provides sharp identification of the migrant noise amplifier R—the core parameter of the model—independent of the promotion thresholds. Without this conditioning, R and κ_j would be jointly identified only through the total wage gap, which confounds slower learning with differential assignment.

4.4 Calibration Procedure

I estimate the 14 free parameters by simulated method of moments (SMM), minimizing the weighted distance between model-implied and data moments,

$$Q(\theta) \;=\; \sum_{k=1}^{32} w_k \left(\hat{m}_k(\theta) - d_k\right)^2, \tag{21}$$

where $\hat{m}_k(\theta)$ is the model moment at parameter vector θ, d_k is the data target, and w_k is the inverse-variance weight $1/\text{SE}_k^2$, scaled by block-specific factors. Exit moments are upweighted most heavily, because selective exit of low-m migrants narrows the wage gap among survivors: without upweighting, the optimizer can reduce the wage-gap loss by suppressing exit.[35]

Each evaluation of $Q(\theta)$ has two stages. First, the dynamic programming problem is solved by backward induction over beliefs, variance, rungs, and experience, separately for each group. Second, a panel of 50,000 workers per group is simulated forward for 15 experience periods, drawing ability, signals, exit shocks, and separations from the model's distributions, and the 32 moments are computed from the simulated panel. The 14-dimensional parameter space is searched globally using differential evolution, followed by local gradient-based refinement. Details are in Appendix D.

5 Results

The calibrated model points to a large learning penalty as the primary driver of the migrant career gap. Employer signals about migrant ability are roughly eight times noisier than those about locals, slowing the convergence of beliefs and delaying promotion. This learning disadvantage is reinforced by two secondary forces: slower human-capital accumulation and a belief-sensitive outside option that selects the most promising migrants out of the German labor market. I first discuss the parameter estimates and their economic interpretation, then evaluate how well the model reproduces the targeted moments.

5.1 Parameter Estimates

Table 5 summarizes the key parameter estimates; full details including signal variances, implied Kalman gains by rung, and pre-set parameters are in Table 12 (Appendix D).

[35] I have verified that the qualitative pattern of calibrated parameters is robust to moderate changes in the relative block weights (halving or doubling the exit upweighting factor). Appendix D reports the full weighting scheme and optimization details.

Table 5: Key parameter estimates

Parameter	Estimate
Learning speed	
R (noise multiplier)	2.78
κ_L (local noise gradient)	8.87
κ_M (migrant noise gradient)	13.24
Human capital & ability	
α_L (local HC rate)	0.040
ρ_α (migrant/local ratio)	0.72
c_1 (ability leverage, mid)	1.18
c_2 (ability leverage, high)	1.28
Promotion friction	
κ_0 (low $\rightarrow$ mid)	3.85
κ_1 (mid $\rightarrow$ high)	1.73
Exit selection	
μ (baseline)	0.08
μ_1 (belief slope)	0.55
μ_2 (belief curvature)	0.10
ξ (taste shock scale)	1.03

Notes: Point estimates from simulated method of moments with differential evolution (14 free parameters; see Appendix D for full parameter values and computational details). Standard errors are not reported. The simulation-based objective precludes reliable numerical Hessian inversion. Bootstrap inference, which requires re-running the full global optimization on resampled data, is left to future work.

Learning speed. The most striking result is the magnitude of the learning differential. The migrant noise multiplier is $R = 2.78$, so baseline signal variance for migrants is roughly eight times that of locals ($R^2 \approx 7.7$). Translated into Kalman gains, employers update beliefs about a local worker at the low rung at rate $K_{L,0} = 0.222$ (about one-fifth of the prior uncertainty resolved each period). The corresponding migrant gain is $K_{M,0} = 0.036$, roughly six times slower. In career terms, this gap is stark: after fourteen years of experience, employers have resolved 80% of their initial uncertainty about a local's ability but only 34% about a migrant's. A migrant at the end of the observation window carries the same residual uncertainty as a local after fewer than two years. The penalty worsens up the ladder, reflecting the two channels identified in Section 4.1: the level shifter R generates the low-rung disadvantage through uniformly noisier signals, while the excess gradient $\kappa_M - \kappa_L$ amplifies it at higher complexity, where cultural and institutional knowledge matters more. At the

mid rung, migrants learn roughly ten times slower than locals. At the high rung, the ratio approaches eleven. These estimates are large; the parsimonious noise structure may overstate the true learning penalty to the extent that other frictions also contribute to the career gap.

Human capital and ability leverage. A secondary but important force is differential human-capital accumulation. The local rate $\alpha_L = 0.040$ implies roughly 56 log points of wage growth over fourteen years from experience alone; migrants accumulate at 72% of this rate ($\rho_\alpha = 0.72$), generating a shortfall of about 16 log points by the end of the observation window—even without differential promotions. Ability leverage rises modestly across rungs ($c_1 = 1.18$, $c_2 = 1.28$), meaning that higher-rung assignments amplify the return to talent but not dramatically. The modest leverage is worth noting: the model likely understates the true return to promotion.

Promotion thresholds. Promotion requires a minimum surplus gain that is common to both groups ($\kappa_0 = 3.85$ for low $\to$ mid; $\kappa_1 = 1.73$ for mid $\to$ high). The *relative information loss* from promotion—the fraction of per-period precision sacrificed by moving to a noisier rung—is nearly group-neutral: 82% for locals versus 87% for migrants at the first transition, and 45% versus 47% at the second.[36] This resolves the theoretical ambiguity identified in Section 3: noisier signals could either raise or lower the promotion threshold, depending on whether the relative information cost of moving up is larger or smaller for migrants. The calibration shows it is essentially the same. The promotion delay is therefore driven by two reinforcing forces: slower belief convergence means migrants take longer to accumulate the evidence needed to justify promotion, and persistently high posterior variance sustains the option value of waiting, keeping the effective threshold elevated long after it has relaxed for locals.

[36]Computed as $\rho_g(j) = 1 - \sigma^2_{gj}/\sigma^2_{g,j+1}$; under the linear noise specification this reduces to $\rho_g(j) = \kappa_g(\zeta_{j+1} - \zeta_j)/(1 + \kappa_g\,\zeta_{j+1})$, which depends only on the complexity gradient κ_g and rung indices. See equation (35) in Appendix C.2 for the derivation.

Outside option. The outside-option parameters reveal a Roy-style selection channel that operates on both tails of the ability distribution. The outside option value is quadratic in perceived ability, with a large positive slope ($\mu_1 = 0.55$) and positive curvature ($\mu_2 = 0.10$). Exit is negatively selected on average: bad matches stagnate and eventually leave as even a modest outside option dominates their depressed career prospects. But the right tail of the exiter distribution is thick—a substantial fraction of exiters are workers the occupation should have retained. Under noisy learning, these high-ability migrants receive unlucky signal draws, the market undervalues them, wages stagnate, and the outside option becomes attractive. Noisy learning thus weakens selection at both margins: bad matches are not identified quickly enough to exit early, and good matches receive misleadingly negative signals that push them out. The Gumbel scale $\xi = 1.03$ implies substantial dispersion in non-pecuniary exit motives, so that exit probabilities respond smoothly to beliefs rather than switching at a single cutoff.

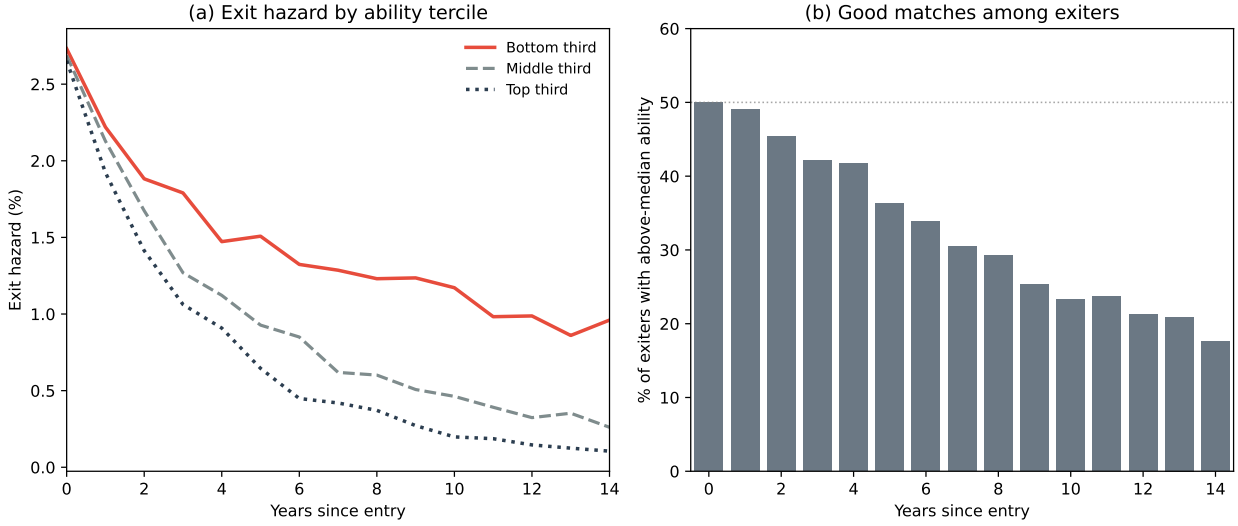
Figure 6 illustrates these patterns. Panel (a) shows that all ability terciles start at similar exit hazard rates—noisy beliefs cannot yet distinguish good from bad matches. Over time, the bottom tercile remains elevated while the middle and top terciles converge toward zero: the market gradually learns who the bad matches are, but it takes a decade. Panel (b) shows the composition of exiters: at entry, roughly half have above-median ability; by experience 14, fewer than one in five is a worker the occupation should have retained.

5.2 Model Fit

Figure 7 and Table 13 compare the model's predictions to the 32 targeted moments. The model fits the wage gap and exit dynamics closely, the local wage-experience profile less so, and struggles most with promotion timing.[37]

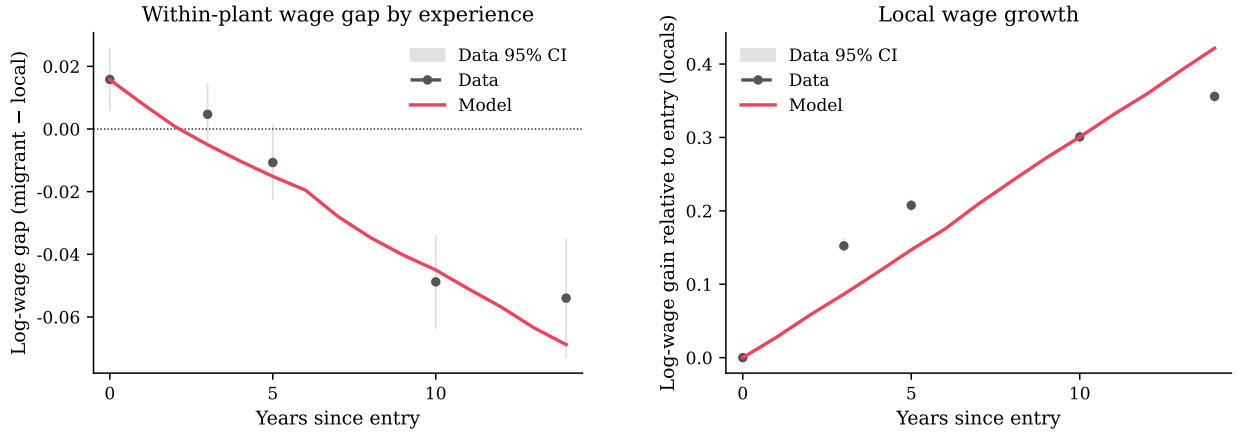
[37] Enriching the human-capital specification—for instance, with stochastic, ability-dependent accumulation as in Pastorino (2024)—would strengthen the promotion margin and is a priority for the next iteration of the model.

Figure 6: Exit selection under noisy learning



Notes: Panel (a) shows the exit hazard rate (probability of voluntary exit conditional on survival) by tercile of true ability θ, which is unobserved by employers. Panel (b) shows the fraction of exiters at each experience level whose true ability exceeds the population median ($\theta > 0$); the dashed line marks 50%. Model simulated with $N = 200{,}000$ migrant workers.

Figure 7: Wage dynamics: data vs. model



(a) Migrant–local wage gap

(b) Local wage-experience profile

Notes: Panel (a): the gap is defined as migrant minus local log wages within the same establishment, controlling for initial occupation. Panel (b): cumulative log-wage gain for locals relative to entry. Grey bars/bands show 95% confidence intervals. Model simulated with $N = 100{,}000$ workers per group.

Wages. The model tracks the migrant–local wage gap closely (Panel A of Table 13; Figure 7a): starting at +1.6 log points at entry (pinned by construction), the gap turns negative by experience 5 and widens to −4.5 log points by experience 10 (data: −4.9). The local

wage-experience profile fits less well (Panel B; Figure 7b): the linear human-capital specification generates a nearly straight line through the data's concave trajectory, underpredicting early wage growth (6.2 log points too low at experience 3) and overpredicting late growth (6.7 log points too high at experience 14). Because the wage gap is the difference of two profiles shaped by the same specification, the local-profile misfit propagates into the gap: the overshoot at experience 14 (-6.9 vs. -5.4 log points) reflects the model stretching late wage growth while compressing it early. The within-gap fit nevertheless remains good at all intermediate horizons, with deviations under one log point through experience 10.

Exit. The model reproduces the declining profile of migrant excess exit well (Panel H): 2.5 percentage points at entry (data: 3.0), falling to 0.6 at experience 10 (data: 0.4). The stagnation–exit gradient has the correct sign (Panel I)—migrants with stagnating wages are more likely to leave—though the model captures roughly half of the empirical magnitude.

Promotions. Promotion dynamics are the weakest dimension of fit, though the model captures the correct qualitative patterns on both margins. Local mid-and-high shares rise over the career, and migrants fall behind—both as in the data. Quantitatively, however, both movements are too muted. On the *level* (Panel F): local promotions reach 1.7 percentage points above entry by experience 10, versus 3.4 in the data, closing to two-thirds of the empirical magnitude by experience 14 (2.7 vs. 4.0). On the *cross-group gap* (Panel D): the migrant shortfall at experience 10 is -1.6 percentage points against -5.1 in the data. That said, the within-rung conditional wage gap—which isolates learning from assignment—is matched more closely (Panel C: -2.4 vs. -3.7 log points at experience 10), confirming that the model captures the differential speed of learning even where it underestimates the speed of promotion. The quantitative shortfall likely reflects the modest ability leverage ($c_1 = 1.18$, $c_2 = 1.28$): understating the surplus gain from assigning high-ability workers to complex tasks dampens the incentive to promote and compresses the resulting migrant–local gap. Increasing ability leverage or enriching the promotion margin (e.g., through rung-specific

human capital) are natural extensions.

Summary. The calibrated model captures the central narrative: migrants start with a small wage advantage that erodes as employers learn faster about locals, promoting them sooner and accumulating more human capital along the way. Exit is concentrated early and responds to career stagnation, consistent with the belief-driven exit mechanism described in Section 3.5. Quantitatively, the model fits the wage gap and exit dynamics well, though the linear human-capital specification distorts the local wage-experience profile. The main weakness is promotion, where the model underestimates both the speed and differential pace of assignment—a limitation of the parsimonious specification that mutes the assignment channel relative to the data.

6 Counterfactuals

The calibrated model attributes the migrant career gap to three forces: differential learning speed, slower human-capital accumulation, and selective exit. This section isolates their contributions through counterfactual experiments. A caveat is in order: as shown in Section 5, the model underestimates promotion dynamics, so the decomposition likely understates the role of the assignment channel. The selection channel emerges as the primary mechanism through which the information friction affects the career gap, but a richer promotion margin would plausibly redistribute some of the explanatory weight toward assignment.

6.1 Decomposing the Career Gap

I begin with the paper's central question: how much of the career gap is driven by the information friction? I construct counterfactual experiments along two dimensions: learning speed (differential vs. equal) and exit regime (endogenous exit vs. no endogenous exit). "Equal learning" sets $R = 1$ and $\kappa_M = \kappa_L$, giving migrants the same signal precision as

locals. "No endogenous exit" shuts down voluntary exit ($\mu \to -\infty$) while retaining exogenous separation $\delta(t)$. The human-capital accumulation gap ($\alpha_M/\alpha_L = 0.72$) is retained throughout.

Table 6: Counterfactual decomposition of the migrant career gap

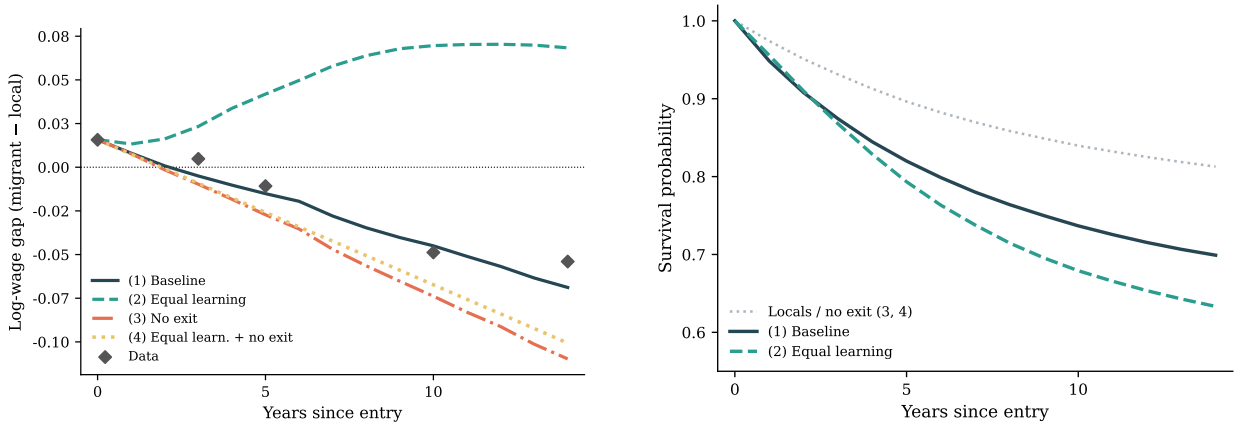
	Endogenous exit		No endogenous exit	
	(1)	(2)	(3)	(4)
A. Log-wage gap (migrant − local)				
Experience 5	-0.0151	+0.0419	-0.0272	-0.0259
Experience 10	-0.0450	+0.0696	-0.0738	-0.0673
Experience 14	-0.0688	+0.0683	-0.1097	-0.1007
Average (1–14)	-0.0306	+0.0512	-0.0509	-0.0466
B. Migrant survival rate				
Experience 5	0.820	0.793	0.896	0.896
Experience 14	0.699	0.633	0.813	0.813
C. Mid-and-high share gap (migrant − local)				
Experience 10	-0.0158	+0.0042	-0.0156	+0.0011
Equal learning	No	Yes	No	Yes

Notes: "Equal learning" sets $R = 1$ and $\kappa_M = \kappa_L$, giving migrants the same signal precision as locals. "No endogenous exit" shuts down voluntary exit ($\mu \to -\infty$); exogenous separation $\delta(t)$ remains active. HC rate differences ($\alpha_M/\alpha_L = 0.72$) are retained throughout. Model simulated with $N = 100{,}000$ workers per group.

Table 6 reports the results. Column (1) reproduces the baseline. Column (2) equalizes learning speed while keeping endogenous exit active. The migrant wage penalty does not merely close, it reverses from −4.5 log points at experience ten to +7.0. The fourteen-year average swings from −3.1 to +5.1 log points. Figure 8a traces this reversal across the full experience profile, moving from line (1) to (2).

This result, however, combines several effects. Columns (3) and (4) shut down endogenous exit, removing the selection channel, and allow a decomposition of the baseline gap. In Figure 8a, these two scenarios trace nearly overlapping paths at the bottom of the panel—the common floor that reflects the structural disadvantage in human-capital accumulation. Column (4) (equal learning, no exit) isolates the pure human-capital penalty at −6.7 log points at experience ten. Moving from Column (4) to (3) reintroduces differential learning while keeping exit shut down; the gap moves modestly, from −6.7 to −7.4 log points. The

Figure 8: Counterfactual wage gaps and survival



(a) Wage gap

(b) Migrant survival

Notes: Each line corresponds to a column in Table 6; line styles and colors distinguish scenarios. Panel (b) additionally shows locals (gray) for reference. "Equal learning" sets $R = 1$ and $\kappa_M = \kappa_L$; "no endogenous exit" sets $\mu \to -\infty$ while retaining exogenous separation $\delta(t)$. HC rate differences ($\alpha_M/\alpha_L = 0.72$) are retained throughout. Diamonds in Panel (a) show the targeted wage-gap moments from Section 4.3. Model simulated with $N = 100{,}000$ workers per group.

direct effect of the information friction on allocation accounts for only 0.7 log points. Moving from Column (3) to (1) restores endogenous exit, and the gap narrows from -7.4 to -4.5, a $+2.9$ log-point gain from selective exit even under differential learning. The full baseline gap thus decomposes as -6.7 (HC) $-$ 0.7 (learning) $+$ 2.9 (selection) $= -4.5$.

The visual distance between the baseline and the no-exit floor in Figure 8a is the selection effect: $+2.9$ log points under differential learning. But comparing columns (2) and (4) reveals how dramatically the information friction suppresses this channel. Under equal learning, restoring endogenous exit swings the gap by 13.7 log points (from -6.7 to $+7.0$), compared to only 2.9 under differential learning—a fivefold difference. The information friction thus operates primarily by weakening exit selection, not through direct effects on allocation: occupational complexity gaps remain small in magnitude across all scenarios (Panel C). Figure 8b illustrates the mechanism: under equal learning, migrant survival drops well below the baseline because faster learning identifies bad matches sooner, pulling them out of the occupation. Under differential learning, this sorting is blunted—bad matches linger and good matches leave, as I show in the next subsection.

6.2 Who Exits?

The decomposition shows that selection at the exit margin is the dominant channel through which the information friction affects the career gap. This raises the question: who are the migrants that leave?

Table 7: Who exits? Ability selection under baseline and equalized learning

	Baseline		Equal learning	
	Exiters	Stayers	Exiters	Stayers
A. Overview				
N	13,169	86,831	20,500	79,500
Share (%)	13.2	86.8	20.5	79.5
B. Ability (θ)				
Mean	-0.275	+0.061	-0.735	+0.218
p25	-0.982	-0.606	-1.408	-0.407
p50	-0.276	+0.053	-0.802	+0.189
p75	+0.429	+0.727	-0.129	+0.823
C. At moment of exit				
Employer belief (m_t)	-0.261		-0.734	
Wage growth (last)	-0.024	+0.024	-0.091	+0.034

Notes: "Exiters" are migrants who voluntarily leave the occupation; "stayers" remain until the end of the observation window or separate exogenously. "Equal learning" sets $R = 1$ and $\kappa_M = \kappa_L$, giving migrants the same signal precision as locals. θ is true (unobserved) ability, drawn $\mathcal{N}(0, 1)$ at entry; percentiles refer to the within-group distribution. m_t is the employer's posterior mean belief about θ at the time of exit. "Wage growth (last)" is the change in log wages between the penultimate and final period before exit (for exiters) or between the last two periods in the sample (for stayers). All exits occur from the lowest rung. Model simulated with $N = 100{,}000$ migrant workers per scenario.

Table 7 compares the composition of exiters and stayers under baseline and equalized learning. Under the baseline, 13% of migrants voluntarily exit. Exit is negatively selected on average: exiters have mean ability -0.28 versus $+0.06$ for stayers (Panel B). Exiters experience negative wage growth in their last period (-0.024) while stayers average $+0.024$ (Panel C), consistent with a deteriorating match value triggering the exit decision. All exits occur from the low rung.

The ability distribution of exiters, however, reveals that exit operates at both ends of the distribution. Comparing across columns shows how learning speed reshapes who leaves.

Bad matches stay too long. Low-ability migrants are the primary exiters: the 25th percentile of exiter ability is −0.98 and the median is −0.28 (Panel B, baseline). But the exit rate of 13% is low relative to what faster learning would produce. Under equalized learning, the exit rate rises to 21%, mean exiter ability falls from −0.28 to −0.73, and employer beliefs at exit drop from −0.26 to −0.73 (Panel C). Noisy signals delay the identification of low types: their employer beliefs remain imprecise, wages stay close to the pooling level, and voluntary exit is postponed. These workers accumulate experience in a career that will not reward them. Faster learning would be welfare-improving: they would exit sooner and redirect to their outside option.

Good matches leave unnecessarily. The right tail of the exiter distribution is thick under baseline learning: exiter p75 is +0.43, close to stayer p75 of +0.73 (Panel B). A substantial fraction of exiters are workers the occupation should have retained. Under noisy learning, these high-ability migrants receive unlucky signal draws, the market undervalues them, wages stagnate, and the outside option becomes attractive. Under equalized learning, the right tail thins dramatically: exiter p75 drops to −0.13, while stayer p75 rises to +0.82. These workers are correctly identified and their wages rise with ability, making exit unattractive.

Together, these two forces explain why noisy learning produces weak selection at the exit margin. Too few bad matches leave, and too many good matches leave—a pattern visible in Figure 6, where the composition of exiters shifts from near-random at entry to predominantly low-ability only after a decade. The net effect is a surviving pool that is less positively selected, which is why migrant survival is higher under baseline learning than under equalized learning (Figure 8b; Table 6, Panel B): 70% of migrants survive to experience fourteen under baseline, compared to only 63% under equal learning.

7 Conclusion

This paper studies how employer learning shapes the careers and retention of migrant workers. Using German social-security records, I document three empirical regularities: migrants exit at elevated rates that converge to locals' after roughly eleven years, they fall progressively behind in wages and occupational complexity within the same establishments, and career stagnation differentially predicts exit. I develop a dynamic career model with Bayesian learning, a three-rung occupational ladder, and an endogenous exit margin, and calibrate it to match these patterns jointly. The estimates imply that employer signals about migrants are roughly eight times noisier in variance terms, delaying belief convergence substantially: after fourteen years, employers have resolved 80% of their uncertainty about a local but only 34% about a migrant.

The paper's central finding is that this information friction operates primarily through selection at the exit margin, rather than through a direct glass ceiling on promotion. Counterfactual decompositions show that the learning effect accounts for around 15% of the observed wage penalty, while equalizing signal precision reverses the wage gap from -4.5 to $+7.0$ log points—not by reducing exit, but by making it efficient. Under faster learning, bad matches leave sooner and good matches stay, producing a fivefold increase in the selection effect relative to the baseline. The policy implication shifts from targeting promotion barriers to improving signal quality: structured performance evaluation, credential recognition, and early language investment make migrant ability legible to employers, accelerating belief convergence and tightening the link between true ability and the exit decision.

Several dimensions of the analysis call for further development. The model underpredicts promotion dynamics, suggesting that stochastic, ability-dependent human capital accumulation following Pastorino (2024) would strengthen the assignment channel. Introducing firm-level heterogeneity in signal precision would test whether some employers learn about migrants faster than others. Integrating the employer-side learning friction with the worker-

side investment channel of Adda et al. (2022)—where host-country investment may improve not only wages but also signal quality—is a natural next step. Finally, bootstrap inference and sensitivity analysis of the calibrated parameters remain to be completed.

References

Adda, J., C. Dustmann, and J.-S. Görlach (2022): "The Dynamics of Return Migration, Human Capital Accumulation, and Wage Assimilation," *The Review of Economic Studies*, 89, 2841–2871.

Altonji, J. G. and C. R. Pierret (2001): "Employer Learning and Statistical Discrimination*," *The Quarterly Journal of Economics*, 116, 313–350.

Arellano-Bover, J. and S. San (2023): "The Role of Firms and Job Mobility in the Assimilation of Immigrants: Former Soviet Union Jews in Israel," Discussion Paper 16389, IZA.

Arrow, K. J. (1973): "The Theory of Discrimination," in *Discrimination in Labor Markets*, ed. by O. Ashenfelter and A. Rees, Princeton, NJ: Princeton University Press, 3–33.

Aryal, G., M. Bhuller, and F. Lange (2022): "Signaling and Employer Learning with Instruments," *American Economic Review*, 112, 1669–1702.

Baker, G., M. Gibbs, and B. Holmstrom (1994a): "The Internal Economics of the Firm: Evidence from Personnel Data," *Quarterly Journal of Economics*, 109, 881–919.

Baker, G., M. Gibbs, and B. Holmstrom (1994b): "The Wage Policy of a Firm," *Quarterly Journal of Economics*, 109, 921–955.

Basso, G., E. Gentili, S. Lattanzio, and G. Roma (2025): "Migration Flows and Policies in Italy and in Other European Countries," Occasional Paper 923, Bank of Italy, Rome.

Benson, A. and L. P. Lepage (2024): "Learning to Discriminate on the Job," Working Paper Discussion Paper No. 25116, Rationality and Competition, CRC TRR 190.

Benson, A., D. Li, and K. Shue (2019): "Promotions and the Peter Principle," *Quarterly Journal of Economics*, 134, 2085–2134.

Berbée, P. and J. Stuhler (2025): "The Integration of Migrants in the German Labour Market: Evidence over 50 Years," *Economic Policy*, 40, 481–549.

Bircan, Ç., G. Friebel, and T. Stahl (2024): "Gender Promotion Gaps in Knowledge Work: The Role of Task Assignment in Teams," CEPR Discussion Paper No. 20338.

Blackwell, D. (1953): "Equivalent Comparisons of Experiments," *The Annals of Mathematical Statistics*, 24, 265–272.

Bloom, D. E., M. Kuhn, and K. Prettner (2024): "Fertility in High-Income Countries: Trends, Patterns, Determinants, and Consequences," *Annual Review of Economics*, 16, 159–184.

Borjas, G. (1987): "Self-Selection and the Earnings of Immigrants," *The American Economic Review*, 77, 531–553.

Borjas, G. J. (1985): "Assimilation, Changes in Cohort Quality, and the Earnings of Immigrants," *Journal of Labor Economics*, 3, 463–489.

Borjas, G. J. (1995): "Assimilation and Changes in Cohort Quality Revisited: What Happened to Immigrant Earnings in the 1980s?" *Journal of Labor Economics*, 13, 201–245.

Borjas, G. J. and B. Bratsberg (1996): "Who Leaves? The Outmigration of the Foreign-Born," *The Review of Economics and Statistics*, 78, 165–176.

Brücker, H., A. Glitz, A. Lerche, and A. Romiti (2021): "Occupational Recognition and Immigrant Labor Market Outcomes," *Journal of Labor Economics*.

Brücker, H. and E. J. Jahn (2011): "Migration and Wage-setting: Reassessing the Labor Market Effects of Migration," *The Scandinavian Journal of Economics*, 113, 286–317.

Burdett, K. and D. T. Mortensen (1998): "Wage Differentials, Employer Size, and Unemployment," *International Economic Review*, 39, 257–273.

Chen, Y., T. Jungbauer, and M. Waldman (2025): "Education Signaling and Employer Learning Heterogeneity," Working paper.

Chiswick, B. R. (2001): "The Effect of Americanization on the Earnings of Foreign-born Men," in *The New Immigrant in the American Economy*, Routledge.

Coate, S. and G. C. Loury (1993): "Will Affirmative-Action Policies Eliminate Negative Stereotypes?" *American Economic Review*, 83, 1220–1240.

DeVaro, J. and M. Waldman (2012): "The Signaling Role of Promotions: Further Theory and Empirical Evidence," *Journal of Labor Economics*, 30, 91–147.

Dodini, S. and A. Willen (2025): "The Power to Discriminate," Working Paper Discussion Paper No. 17830, IZA.

Drechsler, J. and J. Ludsteck (2025): "Imputation Strategies for Rightcensored Wages in Longitudinal Datasets," Working Paper arXiv:2502.12967, arXiv.

Dustmann, C. (1994): "Speaking Fluency, Writing Fluency and Earnings of Migrants," *Journal of Population Economics*, 7, 133–156.

Dustmann, C. (1997): "Return Migration, Uncertainty and Precautionary Savings," *Journal of Development Economics*, 52, 295–316.

Dustmann, C. (2003): "Return Migration, Wage Differentials, and the Optimal Migration Duration," *European Economic Review*, 47, 353–369.

Dustmann, C. and J.-S. Görlach (2016): "The Economics of Temporary Migrations," *Journal of Economic Literature*, 54, 98–136.

European Central Bank. (2022): *The Macroeconomic and Fiscal Impact of Population Ageing.*, LU: Publications Office.

Farber, H. S. and R. Gibbons (1996): "Learning and Wage Dynamics," *Quarterly Journal of Economics*, 111, 1007–1047.

Fitzenberger, B., A. Osikominu, and R. Völter (2006): "Imputation Rules to Improve the Education Variable in the IAB Employment Subsample," *Journal of Contextual Economics – Schmollers Jahrbuch*, 126, 405–436.

Ganzer, A., A. Schmucker, and S. Wolter (2024): "Betriebs-Historik-Panel 1975–2023," FDZ-Datenreport 09/2024, Institut für Arbeitsmarkt- und Berufsforschung (IAB), Nürnberg.

Gibbons, R. and M. Waldman (1999): "Chapter 36 Careers in Organizations: Theory and Evidence," in *Handbook of Labor Economics*, Elsevier, vol. 3, 2373–2437.

Gibbons, R. and M. Waldman (2006): "Enriching a Theory of Wage and Promotion Dynamics inside Firms," *Journal of Labor Economics*, 24, 59–107.

Gorshkov, A. (2024): "Job Ladders and Labour Market Assimilation of Immigrants," *Labour Economics*, 90, 102594.

Hermansen, A. S., A. Penner, I. Boza, M. M. Elvira, O. Godechot, M. Hällsten, L. F. Henriksen, F. Hou, Z. Lippényi, T. Petersen, M. Reichelt, H. Sabanci, M. Safi, D. Tomaskovic-Devey, and E. Vickstrom (2025): "Immigrant–Native Pay Gap Driven by Lack of Access to High-Paying Jobs," *Nature*.

Hirsch, B. and E. J. Jahn (2015): "Is There Monopsonistic Discrimination against Immigrants?" *ILR Review*, 68, 501–528.

Hoffman, M. and C. T. Stanton (2025): "People, Practices, and Productivity: A Review

of New Advances in Personnel Economics," in *Handbook of Labor Economics*, Elsevier, vol. 5, chap. 9.

Jäger, S., S. Noy, and B. Schoefer (2022): "The German Model of Industrial Relations: Balancing Flexibility and Collective Action," *Journal of Economic Perspectives*, 36, 53–80.

Javdani, M. and A. McGee (2018): "Labor Market Mobility and the Early-Career Outcomes of Immigrant Men," *IZA Journal of Development and Migration*, 8, 20.

Jovanovic, B. (1979): "Job Matching and the Theory of Turnover," *Journal of Political Economy*, 87, 972–990.

Kaas, L. and C. Manger (2012): "Ethnic Discrimination in Germany's Labour Market: A Field Experiment," *German Economic Review*, 13, 1–20.

Kahn, L. B. and F. Lange (2014): "Employer Learning, Productivity, and the Earnings Distribution: Evidence from Performance Measures," *Review of Economic Studies*, 81, 1575–1613.

Lange, F. (2007): "The Speed of Employer Learning," *Journal of Labor Economics*, 25, 1–35.

Lehmer, F. and J. Ludsteck (2015): "Wage Assimilation of Foreigners: Which Factors Close the Gap? Evidence From Germany," *Review of Income and Wealth*, 61, 677–701.

Lehrer, S. F., L. P. Lepage, and N. S. Pereira (2025): "Breaking Barriers: The Impacts of Employer Exposure to Immigrants," *Labour Economics*, 102805.

Lepage, L. P. (2024): "Experience-Based Discrimination," *American Economic Journal: Applied Economics*, 16, 288–321.

Ludsteck, J. and U. Thomsen (2016): "Imputation of the Working Time Information for the Employment Register Data," FDZ-Methodenreport 01/2016, Institut für Arbeitsmarkt- und Berufsforschung (IAB), Nürnberg.

Marois, G., A. Bélanger, and W. Lutz (2020): "Population Aging, Migration, and Productivity in Europe," *Proceedings of the National Academy of Sciences*, 117, 7690–7695.

McFadden, D. L. (1974): "Conditional Logit Analysis of Qualitative Choice Behavior," in *Frontiers in Econometrics*, ed. by P. Zarembka, New York: Academic Press, 105–142.

Milgrom, P. and C. Shannon (1994): "Monotone Comparative Statics," *Econometrica*, 62, 157–180.

OECD (2024): *Return, Reintegration and Re-migration: Understanding Return Dynamics and the Role of Family and Community*, OECD.

OECD (2025): *OECD Employment Outlook 2025: Can We Get Through the Demographic Crunch?*, Paris: OECD Publishing.

Pastorino, E. (2024): "Careers in Firms: The Role of Learning about Ability and Human Capital Acquisition," *Journal of Political Economy*, 132, 1994–2073.

Paulus, W. and B. Matthes (2013): "Klassifikation der Berufe 2010 - Struktur, Codierung und Umsteigeschlüssel," FDZ-Methodenreport 08/2013, Institut für Arbeitsmarkt- und Berufsforschung (IAB), Nürnberg.

Peri, G. and C. Sparber (2009): "Task specialization, immigration, and wages," *American Economic Journal: Applied Economics*, 1, 135–169.

Phelps, E. S. (1972): "The Statistical Theory of Racism and Sexism," *American Economic Review*, 62, 659–661.

Postel-Vinay, F. and J.-M. Robin (2002): "Equilibrium Wage Dispersion with Worker and Employer Heterogeneity," *Econometrica*, 70, 2295–2350.

Roy, A. D. (1951): "Some Thoughts on the Distribution of Earnings," *Oxford Economic Papers*, 3, 135–146.

Royston, P. and M. K. B. Parmar (2002): "Flexible Parametric Proportional-Hazards and Proportional-Odds Models for Censored Survival Data, with Application to Prognostic Modelling and Estimation of Treatment Effects," *Statistics in Medicine*, 21, 2175–2197.

Schmucker, A. (2025): "Stichprobe der Integrierten Arbeitsmarktbiografien (SIAB) 1975-2023," FDZ-Datenreport 02/2025, Institut für Arbeitsmarkt- und Berufsforschung (IAB), Nürnberg.

Schuster, E. and J. Strahl (2019): "Beschäftigte mit Leitungsfunktion," Methodenbericht, Statistik der Bundesagentur für Arbeit, Nürnberg.

Statistisches Bundesamt (2025): "Verbraucherpreisindex: Deutschland, Jahre," https://www-genesis.destatis.de/datenbank/online/statistic/61111/table/61111-0001.

Stüber, H., W. Dauth, and J. Eppelsheimer (2023): "A Guide to Preparing the Sample of Integrated Labour Market Biographies (SIAB, Version 7519 v1) for Scientific Analysis," *Journal for Labour Market Research*, 57, 7.

Topkis, D. M. (1998): *Supermodularity and Complementarity*, Princeton University Press.

Train, K. E. (2009): *Discrete Choice Methods with Simulation, Second Edition*, New York: Cambridge University Press.

vom Berge, P. and A. Schmucker (2025): "Creating Cross-Sectional Data and Biographical Variables with the Sample of Integrated Labour Market Biographies 1975-2023 - Programming Examples for Stata," Tech. rep., Research Data Centre of the Federal Employment Agency (BA) at the Institute for Employment Research (IAB).

Waldman, M. (1984): "Job Assignments, Signalling, and Efficiency," *RAND Journal of Economics*, 15, 255–267.

A Data

A.1 Occupational-domain classification

The KldB 2010 classification organizes occupations along two dimensions. The first four digits identify the substantive occupational content, grouped hierarchically into 10 occupational areas (*Berufsbereiche*), 37 subareas (*Berufshauptgruppen*), and finer occupational groups. The fifth digit records the requirement level (*Anforderungsniveau*), which distinguishes four tiers of task complexity (Section 2.1.3). I construct the 14 occupational segments used in the analysis by regrouping the KldB 2010 two-digit subareas into substantively coherent domains that reflect the German sectoral structure: Agriculture and Horticulture, Manufacturing, Production Technology, Construction, Natural Sciences, Information Technologies, Transport and Logistics, Safety and Security, Commerce and Trade, Gastronomy and Tourism, Business Law and Finance, Social and Cultural Services, Health Care, and Cleaning Services. The mapping follows standard practice in work based on the KldB 2010 (Paulus and Matthes, 2013) and ensures that each domain contains a sufficient number of observations for both migrant and local workers. Table 1 reports the distribution of worker-years across these domains.

A.2 Migrant Status

Migration status is coded from the IEB nationality field, which records the citizenship reported by the employer at the time of each employment spell. An individual is classified as a migrant if their first observed nationality is non-German. This status is held fixed over time, following standard practice in the German administrative data literature (e.g. Brücker and Jahn, 2011; Hirsch and Jahn, 2015; Lehmer and Ludsteck, 2015). Using first-observed nationality rather than current nationality avoids the endogeneity of naturalization: acquiring German citizenship is correlated with labor-market integration and individual characteristics,

so conditioning on current citizenship would introduce selection bias. Because naturalization rates in Germany are non-trivial—particularly among long-resident migrants—this concern is empirically relevant (Brücker and Jahn, 2011). A limitation of the nationality-based approach is that second-generation immigrants who retain their parents' foreign citizenship are classified as migrants even though they may not face the same informational frictions as first-generation arrivals. To the extent that second-generation workers are better integrated into the German labor market, their inclusion in the migrant group works against finding differential treatment, and the estimated migrant–local gaps can be interpreted as a lower bound on the career penalties faced by first-generation migrants.

Beyond the binary migrant–local split, migrants are partitioned into eight mutually exclusive and exhaustive origin groups tailored to the German context: (1) German-speaking (Austria, Switzerland, Liechtenstein), (2) Southern EU (Greece, Italy, Spain and Portugal), (3) Central and Eastern EU countries (including Bulgaria, Romania and Croatia), (4) Other EU-15, UK and EFTA, (5) Turkey, (6) Balkans (non-EU), (7) Other OECD (non-EU), and (8) Rest of World. The mapping is constructed explicitly from FDZ nationality keys to groups. Each group falls unambiguously under migrant vs. local and is time-invariant by construction. Table 8 reports the distribution.

Table 8: Distribution of migrants by origin

Nationality groups	N (%)
AT/CH/LI	891 (2.2%)
Southern EU	4,509 (11.1%)
EU-CEEC	10,275 (25.2%)
Other EU-15, UK & EFTA	2,167 (5.3%)
Turkey	6,557 (16.1%)
Balkans (non-EU)	3,877 (9.5%)
Other OECD (non-EU)	681 (1.7%)
Rest of World	11,804 (29.0%)
Total	40,762 (100%)

B Descriptive Statistics and Reduced Form Regressions

B.1 Kaplan–Meier Risk Table

Table 9 reports the Kaplan–Meier life table underlying Figure 1 at five-year intervals of cumulative experience in Germany, separately for locals and migrants. Each row summarizes the interval from the listed time point to the next grid point. *At risk* is the number of individuals carried over from the previous interval (at the start of this interval, before new entries). *Enter* is the number of episodes that begin during the interval (delayed entries as workers first appear in full-time employment). *Censored* and *Fail* are, respectively, the number of episodes that end in censoring or in permanent exit during the interval. The accounting identity $\text{At risk}_{\text{next}} = \text{At risk} + \text{Enter} - \text{Censored} - \text{Fail}$ links successive rows. *Surv.* is the Kaplan–Meier survival estimate at the listed time point with Greenwood standard errors: at time 0 it equals 1 by definition (no one has yet been at risk), and it declines as failures accumulate. For example, at ten years of experience, 89.3% of locals but only 72.5% of migrants remain in the social-security records.

Table 9: Kaplan–Meier risk table at selected experience years

Time	At risk	Enter	Censored	Fail	Surv. (SE)	[95% conf. int.]
Locals						
0	0	340,144	294,339	2,115	1.000 (.)	[., .]
5	43,690	336,601	336,328	2,148	0.938 (0.001)	[0.935, 0.941]
10	41,815	292,168	292,283	1,547	0.893 (0.002)	[0.890, 0.896]
15	40,153	273,737	270,101	1,026	0.860 (0.002)	[0.856, 0.863]
20	42,763	265,711	272,067	805	0.838 (0.002)	[0.834, 0.842]
25	35,602	179,643	214,591	654	0.822 (0.002)	[0.818, 0.826]
Migrants						
0	0	137,851	124,359	2,871	1.000 (.)	[., .]
5	10,621	69,839	73,063	792	0.796 (0.003)	[0.789, 0.803]
10	6,605	46,369	47,181	291	0.725 (0.004)	[0.718, 0.733]
15	5,502	36,166	36,893	162	0.691 (0.004)	[0.683, 0.699]
20	4,613	25,648	27,350	86	0.669 (0.004)	[0.661, 0.678]
25	2,825	15,459	18,220	64	0.654 (0.005)	[0.645, 0.663]

Notes: Kaplan–Meier life table reported at five-year intervals of cumulative labor-market experience in Germany. Analysis time is experience (years) since first observed full-time employment spell. The failure event is permanent exit from the social-security records (no further spells observed within at least twelve months). Each row covers the interval from the listed time to the next grid point. *At risk*: number of individuals carried over from the previous interval. *Enter*: number of episodes beginning during the interval (delayed entries). *Censored* and *Fail*: episodes ending in censoring or permanent exit during the interval. Rows satisfy $\text{At risk}_{\text{next}} = \text{At risk} + \text{Enter} - \text{Censored} - \text{Fail}$. *Surv.*: Kaplan–Meier survival estimate at the listed time point, with Greenwood standard errors in parentheses. Generated via Stata's `sts list, survival by(imm2) risktable(0 5 10 15 20 25) enter`.

B.2 Kaplan–Meier Hazard Function

Figure 9 plots the Kaplan–Meier smoothed hazard functions corresponding to the survival curves in Figure 1. The migrant exit hazard starts at approximately 5% per year and declines steeply over the first ten to fifteen years of experience, converging to the local hazard of roughly 0.4% per year. This nonparametric pattern mirrors the LPM excess-exit coefficients in the main text and confirms that the convergence is not an artifact of the linear specification.

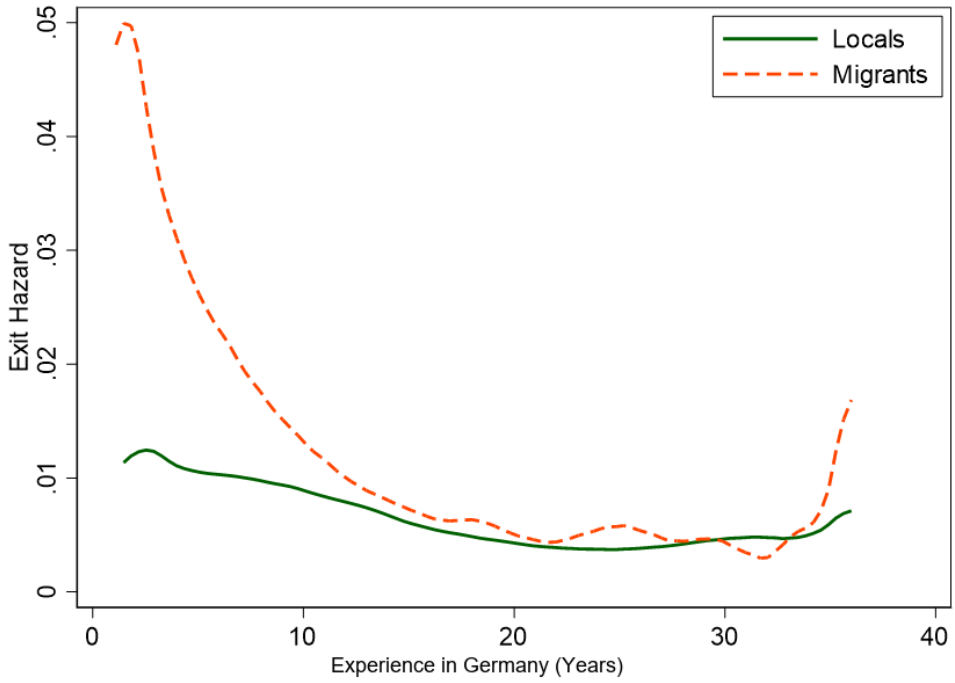


Figure 9: KM Hazard Function

Notes: Kaplan–Meier smoothed hazard estimates by group, using cumulative labor-market experience in Germany as the analysis time. The failure event is permanent exit from the social-security records (no further spells within at least twelve months).

B.3 Royston-Parmar Flexible Hazard Models

The main text documents the migrant excess exit pattern using a linear probability model (LPM) estimated on the annual person-year panel. While the LPM is transparent and straightforward to interpret, it operates on a discrete annual grid and imposes a linear

structure on the exit probability. This section verifies that the same pattern holds in a nonlinear survival framework estimated on actual continuous-time employment spells. I use Royston-Parmar flexible parametric survival models (Royston and Parmar, 2002), which approximate the baseline hazard using restricted cubic splines rather than leaving it unspecified as in a Cox model. This yields a fully parametric model that avoids the proportional-hazards restriction while remaining smooth and easy to interpret.

The model approximates the log cumulative hazard as a smooth function of log-time using natural cubic splines—piecewise cubic polynomials that are constrained to be linear beyond the boundary knots and to join smoothly at the internal knots. This spline-based representation is flexible enough to capture a wide range of hazard shapes (monotone, bathtub, hump-shaped) without imposing a specific distributional assumption. For migrants, the model includes a separate spline interaction that allows the migrant hazard ratio to vary freely over time rather than being forced to remain constant, so that it can capture patterns such as elevated early exit risk that converges toward parity at longer tenures. Formally, the specification is

$$\ln H\big(t \mid \mathrm{Mig}_i, \mathbf{X}_i(t)\big) = S_0\big(\ln t; \boldsymbol{\delta}, \kappa\big) + \boldsymbol{\beta}'\mathbf{X}_i(t) + \beta_{\mathrm{Mig}}\,\mathrm{Mig}_i + \mathrm{Mig}_i \times S_{\mathrm{Mig}}\big(\ln t; \boldsymbol{\delta}_{\mathrm{Mig}}, \kappa_{\mathrm{Mig}}\big), \quad (22)$$

where $H(t \mid \mathrm{Mig}_i, \mathbf{X}_i(t))$ denotes the cumulative hazard as a function of experience t in the German labour market. The term $S_0(\ln t; \boldsymbol{\delta}, \kappa)$ is a baseline log-cumulative hazard modelled as a natural cubic spline in $\ln t$ with parameters $\boldsymbol{\delta}$ and internal knots κ. The vector $\mathbf{X}_i(t)$ collects (potentially time-varying) covariates with proportional effects $\boldsymbol{\beta}$. Migrants differ from locals through a proportional shift β_{Mig} in the log-cumulative hazard and through a time-varying component captured by the interaction of the migrant indicator Mig_i with a second spline in $\ln t$, $S_{\mathrm{Mig}}(\ln t; \boldsymbol{\delta}_{\mathrm{Mig}}, \kappa_{\mathrm{Mig}})$. In other words, for locals ($\mathrm{Mig}_i = 0$), the log-cumulative hazard is given by the baseline spline and the proportional covariate effects. For migrants ($\mathrm{Mig}_i = 1$),

the model allows both a level shift and a flexible deviation from the baseline shape over experience. I use four internal knots for the baseline spline ($\kappa = 4$) and two internal knots for the migrant-specific spline ($\kappa_{\text{Mig}} = 2$).

I estimate three versions of equation (22) that differ in the set of covariates included in $\mathbf{X}_i(t)$. The first specification includes only entry-cohort and calendar-year dummies, so that the migrant effect is identified net of broad time and cohort differences. The second specification augments this with basic demographics (age, education, and gender), allowing for differences in observable life-cycle and skill composition. The third, richest specification additionally controls for detailed job and establishment characteristics, including rung, occupation, industry, region, and plant-level attributes, so that migrants and locals are compared within similar positions and workplaces. Figure 10 plots the implied time-varying hazard ratio for migrants relative to locals as a function of experience in Germany for each specification.

The curves show the relative exit risk of migrants compared to locals at each point in their German careers. Adding demographics and job attributes reduces the level of the migrant hazard ratio but leaves its overall shape largely unchanged: across all three specifications, migrants face substantially higher exit hazards at the beginning of their time in Germany, with the excess risk declining over experience and approaching parity with locals only at longer tenures. This confirms that the higher migrant exit rates documented in the Kaplan–Meier estimates and the LPM regressions in the main text are not solely driven by observable composition, but reflect systematic differences in retention even among workers in comparable observed states.

B.4 Two-Year Within-Plant Wage Growth and Exit

Table 10 replicates the stagnation–exit analysis of Section 2.3.3 using cumulative two-year within-plant wage growth as the regressor instead of the one-year change. The two-year measure is less noisy but restricts the sample to workers who remained at the same

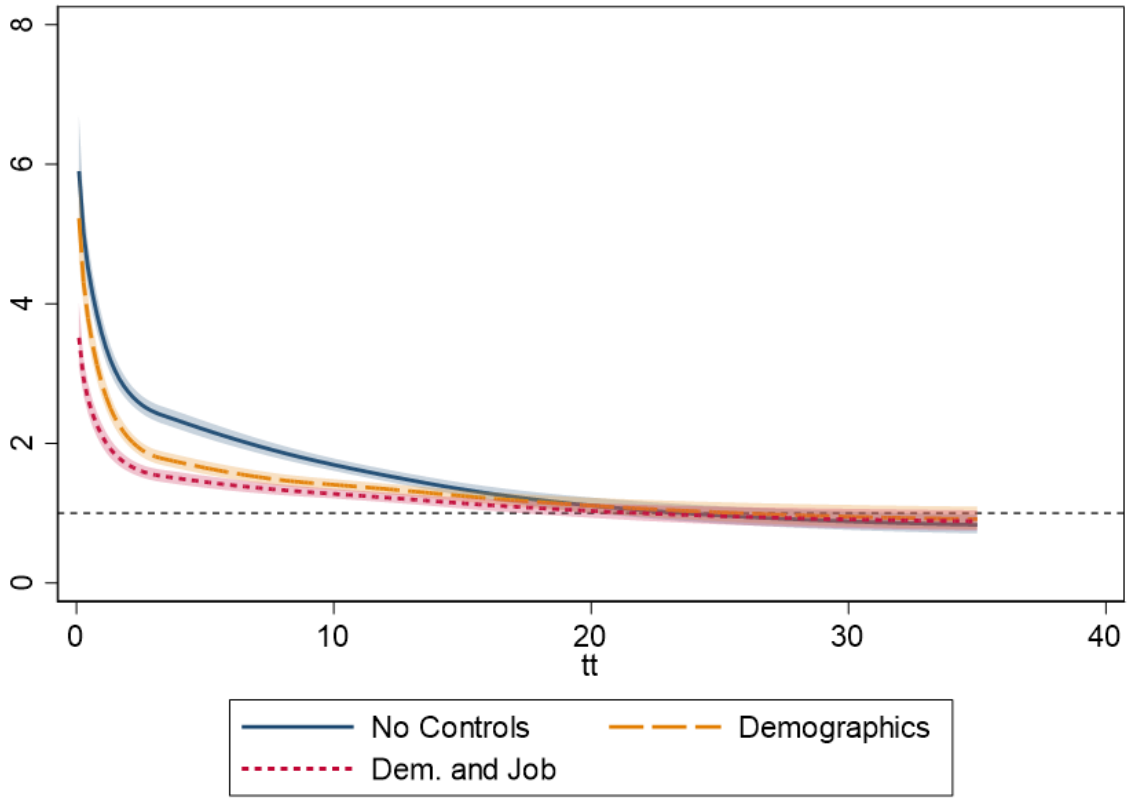


Figure 10: Time-varying hazard ratio of exit for migrants relative to locals

Notes: The figure plots the estimated hazard ratio $h_{\text{Mig}}(t)/h_{\text{Local}}(t)$ implied by equation (22) as a function of experience in Germany. The variable `tt` on the horizontal axis denotes experience in Germany, measured in years since first observed employment in the social-security records. Values above one indicate higher exit hazards for migrants. Shaded areas represent 95% confidence bands. The three curves correspond to specifications with progressively richer controls: (1) entry-cohort and calendar-year dummies only ("No Controls"), (2) adding demographics ("Demographics"), and (3) adding detailed job and establishment characteristics ("Dem. and Job").

establishment for at least two consecutive years. The pattern is qualitatively identical: the association between career stagnation and exit is concentrated among migrants and attenuates after eleven years of experience.

Table 10: Career progression and permanent exit: two-year within-plant wage growth

	Overall Exit				By experience		
	(1) Mig only	(2) Loc only	(3) Pooled	(4) + FE	(5) Exp 0–5	(6) Exp 6–10	(7) Exp 11+
$\Delta \ln w_{it}^{\mathrm{wp,2y}}$	-0.0101***	-0.0034	-0.0035	-0.0029***	-0.0214***	-0.0119*	-0.0035
	(0.0030)	(0.0000)	(0.0000)	(0.0009)	(0.0081)	(0.0066)	(0.0027)
$\Delta \ln w_{it}^{\mathrm{wp,2y}} \times$ Migrant			-0.0061	-0.0046			
			(0.0000)	(0.0028)			
Experience FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Education, gender	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age (quadratic)	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Entry cohort FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Exp × Migrant	No	No	Yes	Yes	No	No	No
Establishment FE	No	No	No	Yes	No	No	No
N	81,673	619,890	701,563	684,184	14,560	16,474	50,639

Notes: Same specification layout as Table 4, but with a two-year within-plant wage-growth regressor. $\Delta \ln w_{it}^{\mathrm{wp,2y}}$ is cumulative within-plant log wage growth over two consecutive years in the same establishment. Columns (1)–(4) are overall specifications and columns (5)–(7) are migrant-only by experience group. All columns include experience FE, year FE, entry-cohort FE, age controls (quadratic), education, and gender. Standard errors clustered at the individual level in parentheses. *, **, *** indicate $p < 0.10$, $p < 0.05$, and $p < 0.01$.

C Model

C.1 Efficient Assignment

Result 1 states that, in any symmetric, stationary, robust MPE, the equilibrium employer and rung assignment maximize the dynamic social surplus. The argument follows Pastorino (2024, Proposition 1) and proceeds in three steps: (i) the worker is indifferent across employers in equilibrium; (ii) each firm's rung choice maximizes bilateral surplus; and (iii) the equilibrium employer generates the highest total surplus.

Throughout, $\Pi^f(\cdot \mid f')$ denotes firm f's profit value when firm f' employs the worker, and $W(\cdot \mid f)$ denotes the worker's value when employed by f. Two branches arise each period: (A) no poacher arrives, so only the incumbent I makes an offer; (B) a poacher P arrives and competes for the worker.

Step 1 (worker indifference). In branch (A), I chooses a rung j and wage w_t to maximize profit subject to the participation constraint $W(s_t, \eta_t \mid I) \geq \underline{\mathrm{W}}_t$, where $\underline{\mathrm{W}}_t$ is the worker's

reservation value. At the optimum the constraint binds: otherwise I could lower the wage and still retain the worker. In branch (B), I and P engage in Bertrand competition. The hiring firm $f \in \{I, P\}$ must make the worker indifferent between f and $f' \neq f$—otherwise f could shade its wage down and still attract the worker. Combining both branches:

$$\begin{aligned} &W(s_t, \eta_t \mid f) = W(s_t, \eta_t \mid f'), \quad f' \neq f, && \text{if a poacher arrives,} \\ &W(s_t, \eta_t \mid I) = \underline{\mathrm{W}}_t && \text{otherwise.} \end{aligned} \tag{23}$$

Step 2 (within-firm surplus maximization). Define the bilateral surplus between firm f and the worker at rung j:

$$\mathcal{S}(f, j \mid s_t, \eta_t) \;\equiv\; \Pi^f(s_t, \eta_t \mid f, j) \;+\; W(s_t, \eta_t \mid f).$$

By Step 1, the worker's value is pinned—either at $\underline{\mathrm{W}}_t$ or equalized across firms—so maximizing profit over j is equivalent to maximizing $\mathcal{S}$ over j. In branch (A), this gives

$$\mathcal{S}(I, j_t^* \mid s_t, \eta_t) = \max_{j \in \mathcal{J}^I} \Big\{ y_I(s_t, j) + \eta_{I,j,t} + \gamma\, \mathbb{E}\big[\mathcal{S}(I, j_{t+1}^* \mid s_{t+1}, \eta_{t+1}) \mid s_t, j\big] \Big\}, \tag{24}$$

where $y_f(s_t, j) \equiv \phi(j, f, g, t) + c_j\, m_t$ is expected output net of the transitory shock from eq. (4), $\gamma \equiv \beta(1 - \delta(t))$, and the expectation is over posterior beliefs, future shocks, and poacher arrival.

The same recursion extends to competition. The equilibrium refinement (robustness) requires the non-employing firm to be indifferent between hiring and not at the equilibrium wage: otherwise it could raise its offer and attract the worker at weakly positive profit. Formally, if firm f does not employ in period t, then $\Pi^f(\cdot \mid f) = \Pi^f(\cdot \mid f')$; if it does employ, $\Pi^f(\cdot \mid f) \geq \Pi^f(\cdot \mid f')$. Either way, f's unconditional profit value equals its hiring value. It

follows that, for any employing firm f,

$$\mathcal{S}(f, j_t^* \mid s_t, \eta_t) = \max_{j \in \mathcal{J}^f} \Big\{ y_f(s_t, j) + \eta_{f,j,t} + \gamma \, \mathbb{E}\big[\mathcal{S}(f, j_{t+1}^* \mid s_{t+1}, \eta_{t+1}) \mid s_t, j\big] \Big\}, \tag{25}$$

analogously to (24). In both cases, the employing firm assigns the worker to the rung that maximizes bilateral surplus.

Step 3 (equilibrium employer maximizes total surplus). In branch (A), total surplus is $\mathcal{S}(I, j_t^* \mid s_t, \eta_t)$. In branch (B), total surplus is $\Pi^I + \Pi^P + W$. To determine which firm generates higher total surplus, compare $\mathcal{S}(s_t, \eta_t \mid I) = \Pi^I(\cdot \mid I) + \Pi^P(\cdot \mid I) + W(\cdot \mid I)$ with $\mathcal{S}(s_t, \eta_t \mid P) = \Pi^I(\cdot \mid P) + \Pi^P(\cdot \mid P) + W(\cdot \mid P)$. Suppose the incumbent hires in equilibrium. Worker indifference (Step 1) gives $W(\cdot \mid I) = W(\cdot \mid P)$; the robustness condition gives $\Pi^P(\cdot \mid I) = \Pi^P(\cdot \mid P)$. These common terms cancel in the difference $\mathcal{S}(\cdot \mid I) - \mathcal{S}(\cdot \mid P)$, leaving

$$\Pi^I(\cdot \mid I) \geq \Pi^I(\cdot \mid P) \tag{26}$$

which holds by I's revealed preference (it chose to retain the worker). The symmetric argument applies when P hires. Together with (24)–(25), this proves that the equilibrium employer generates the highest total surplus and assigns the worker to the surplus-maximizing rung.

C.2 Promotion Rule

Result 2 states that, for any firm f, group g, and adjacent rungs $j \to j{+}1$, promotion obeys a unique cutoff in the belief mean m, with the cutoff weakly increasing in posterior uncertainty v. I establish this in two parts: first, existence and uniqueness of the cutoff via single crossing. Second, comparative statics of the cutoff in v and in signal noise.

Let $Q_f(m, v;\, j)$ denote the value of assigning the worker to rung j at firm f when the

current belief is (m, v), and define the promotion gain

$$\Delta(m, v) \;\equiv\; Q_f(m, v;\, j{+}1) \;-\; Q_f(m, v;\, j). \tag{27}$$

A cutoff $m^{\star}_{g,j}(v)$ is any solution to

$$\Delta\big(m^{\star}_{g,j}(v), v\big) = 0, \tag{28}$$

with the policy $j \to j{+}1$ if and only if $m \geq m^{\star}_{g,j}(v)$. It therefore suffices to prove that $\Delta(\cdot, v)$ is continuous and strictly increasing in m, and that $\Delta(m, v) \to -\infty$ as $m \to -\infty$ while $\Delta(m, v) \to +\infty$ as $m \to +\infty$; by the intermediate value theorem, there is then a unique root $m^{\star}_{g,j}(v)$. I will also establish how $m^{\star}_{g,j}(v)$ moves with v and with a noise scale (e.g., a multiplicative factor on σ_{gj}) using monotone comparative statics in the sense of Milgrom and Shannon (1994).

Crucially, by Result 1 (efficient assignment), conditional on firm f being the employer, the within-firm rung choice solves the autarky problem. Fixing f, g, and t, write $\phi_j \equiv \phi(j, f, g, t)$ for the observable component of output at rung j from equation (4), and let $\gamma \equiv \beta(1 - \delta(t))$. The Bellman recursion for the action value is

$$Q_f(m, v;\, j) \;=\; \phi_j \;+\; c_j\, m \;+\; \eta_{fjt} \;+\; \gamma\, \mathbb{E}_{s_j}\!\left[\, \max_{j' \in \mathcal{J}_f} Q_f\big(m', v';\, j' | j\big) \,\right], \tag{29}$$

where belief transitions follow the Kalman updating in equations (6)–(7): the posterior mean satisfies $m' \sim \mathcal{N}\big(m,\, K_{gj}(v)\, v\big)$ and the posterior variance updates to $v\big(1 - K_{gj}(v)\big)$, with $K_{gj}(v) = v/(v + \sigma^2_{gj})$. Because the within-firm rung choice maximizes surplus (Result 1), wage terms cancel and I can analyze $\Delta(m, v)$ directly.

Let $H(m', v' \mid j) \equiv \max_{j' \in \mathcal{J}_f} Q_f(m', v';\, j' \mid j)$ denote the within-firm continuation value.

The argument uses two properties of the action values—shape regularity and strict increasing differences—that follow from the Gaussian location structure and the ability

leverage ordering $c_{j+1} > c_j$, combined with standard monotone comparative statics results (Milgrom and Shannon, 1994; Topkis, 1998).

Lemma 1 (Shape of action values)**.** *For each rung j, the fixed-point action value $Q_j(\cdot, v)$ is continuous, nondecreasing, and convex in m. Consequently, the continuation value H defined above is continuous, nondecreasing, and convex in m'.*

Proof. I construct the fixed point by value iteration. Initialize $H^{(0)} \equiv 0$. At each iteration, compute the action value at rung j as $Q_j^{(n+1)}(m, v) = \phi_j + c_j m + \eta_{fjt} + \gamma\, \mathbb{E}[H^{(n)}(m', v(1-K_{gj}))]$ with $m' \sim \mathcal{N}(m, K_{gj}v)$, then set $H^{(n+1)} = \max_{j'} Q_{j'}^{(n+1)}$. Each iteration composes three operations that preserve continuity, nondecreasingness, and convexity in m: (i) adding the affine flow $\phi_j + c_j m + \eta_{fjt}$; (ii) taking a Gaussian location average (monotonicity by first-order stochastic dominance; convexity by Jensen's inequality); (iii) taking the pointwise maximum over rungs (Topkis, 1998, Theorem 2.7.6). Since $\gamma < 1$, the Bellman map is a contraction, so the properties pass from every iterate to the fixed point. □

Lemma 2 (Increasing differences)**.** *Fix $v > 0$ and consider two adjacent rungs j and $j{+}1$. For any $m_2 > m_1$,*

$$\Delta(m_2, v) - \Delta(m_1, v) \;\geq\; \frac{c_{j+1} - c_j}{1+\gamma}\,(m_2 - m_1) \;>\; 0. \tag{30}$$

Hence $\Delta(\cdot, v)$ is strictly increasing in m with a uniform positive slope.

Proof. With $H^{(0)} \equiv 0$, the first iterate gives slope gap $\partial_m \Delta^{(1)} = c_{j+1} - c_j$. Write $\delta_n \equiv \inf_{(m,v)} \partial_m \Delta^{(n)}$. The slope of $\Delta^{(n+1)}$ decomposes as

$$\partial_m \Delta^{(n+1)} \;=\; (c_{j+1} - c_j) \;+\; \gamma\big(\partial_m \mathbb{E}_{j+1}[H^{(n)}] - \partial_m \mathbb{E}_j[H^{(n)}]\big),$$

where each $\partial_m \mathbb{E}_r[H^{(n)}]$ is the expected subgradient of $H^{(n)}$ under the Gaussian kernel at rung r. Since $\gamma < 1$, the Bellman operator is a γ-contraction; the continuation gap's slope

contribution is therefore bounded by $\gamma\,\delta_n$ in absolute value, yielding

$$\delta_{n+1} \;\geq\; (c_{j+1} - c_j) \;-\; \gamma\,\delta_n, \qquad \delta_1 = c_{j+1} - c_j. \tag{31}$$

This alternating recursion converges to $(c_{j+1} - c_j)/(1+\gamma) > 0$, yielding the uniform lower bound on $\partial_m \Delta$ in the fixed point and hence (30). □

With these two properties in hand, I prove the two parts of Result 2.

Proof of Part 1 (unique cutoff). By Lemma 1, $\Delta(\cdot, v)$ is continuous. By Lemma 2, it is strictly increasing with a uniform positive slope. Since the flow component $(c_{j+1} - c_j)m$ dominates as $m \to \pm\infty$, the intermediate value theorem gives a unique root $m_g^\star(v)$, and the optimal policy is the cutoff rule $j \to j{+}1$ iff $m \geq m_g^\star(v)$. □

For Part 2, it is helpful to display the promotion gain in its flow–continuation decomposition:

$$\begin{aligned} \Delta(m, v) \;&\equiv\; Q_f(m, v;\, j{+}1) \;-\; Q_f(m, v;\, j) \\ &=\; (\phi_{j+1} - \phi_j) \;+\; (\eta_{f,j+1,t} - \eta_{fjt}) \;+\; (c_{j+1} - c_j)\, m \;+\; \gamma\,\Gamma(m, v), \end{aligned} \tag{32}$$

where the continuation gap is

$$\begin{aligned} \Gamma(m, v) \;\equiv\; &\mathbb{E}\big[\, H\big(m',\, v(1{-}K_{g,j+1})\big) \;\big|\; m' \sim \mathcal{N}\big(m,\, K_{g,j+1}\, v\big)\big] \\ &-\; \mathbb{E}\big[\, H\big(m',\, v(1{-}K_{gj})\big) \;\big|\; m' \sim \mathcal{N}\big(m,\, K_{gj}\, v\big)\big]\,. \end{aligned} \tag{33}$$

Proof of Part 2 (comparative statics). I establish two statements.

(i) Cutoff increases with v. The continuation value $H(m', v') = \max_{j'} Q_{j'}(m', v')$ has increasing differences in $(m', -v')$: the marginal value of higher beliefs is larger when uncertainty is lower. This property is established by iterating the Bellman operator. In each iteration, the flow $\phi_j + c_j m + \eta_{fjt}$ has no cross-effects in (m, v); the Gaussian location kernel has the monotone likelihood ratio property in m; and both the innovation variance $K_{gj}v$ and the

updated posterior variance $v(1-K_{gj})$ are increasing in v. By Topkis (1998, Theorem 2.8.1), the pointwise maximum over rungs preserves increasing differences, so the Bellman map inherits the property, which passes to the fixed point. Since higher v worsens both future variances and widens the one-step spreads while H is convex (Lemma 1), both components of $\Gamma(m, v)$ in (33) shift down when v rises, while the flow part of Δ is unchanged. Hence $\Delta(\cdot, v)$ shifts down in v and $m_g^\star(v)$ is nondecreasing by the monotone selection theorem (Milgrom and Shannon, 1994, Theorem 4).

(ii) Cutoff comparative statics in the noise technology. The cutoff $m_g^\star(v)$ depends on signal noise only through the continuation gap $\Gamma(m, v)$ in (33), which compares the learning environment at rung j with that at rung $j+1$. The economically relevant quantity is the *relative information loss* from promotion: the fraction of per-period precision sacrificed by moving the worker to a noisier task. Define the fractional precision loss as

$$\rho_g(j) \;\equiv\; 1 \;-\; \frac{\sigma_{gj}^2}{\sigma_{g,j+1}^2}\,. \tag{34}$$

When $\rho_g(j) = 0$, the signal environment is equally informative at both rungs: $K_{gj} = K_{g,j+1}$, the two expectations in (33) coincide, and $\Gamma = 0$. Promotion is then governed entirely by the flow comparison $(\phi_{j+1} - \phi_j) + (c_{j+1} - c_j)\, m$. When $\rho_g(j) > 0$, the signal at rung j is Blackwell-sufficient for the signal at rung $j+1$: the noisier signal can be expressed as the cleaner signal plus independent Gaussian noise. Since H is convex in posteriors (Lemma 1), the expected continuation value is weakly higher under the more informative signal (Blackwell, 1953), giving $\Gamma \leq 0$ and raising the cutoff. A larger $\rho_g(j)$ widens the informativeness gap between rungs and strengthens the effect. □

To characterise how the noise parameters shape $\rho_g(j)$, I discuss two forces under the calibrated specification.

Force 1: the level of noise. If the group difference is purely multiplicative—$\sigma_{Mj}^2 = R^2\, \sigma_{Lj}^2$ for all j—then the factor R^2 cancels from numerator and denominator of $\rho_g(j)$, so $\rho_M(j) =$

$\rho_L(j)$: a uniform increase in noise does not change the relative information cost of promotion. Under the linear specification (12)–(13), $\sigma_{gj}^2 = \sigma_{g0}^2(1 + \kappa_g\,\zeta_j)$, so

$$\rho_g(j) \;=\; \frac{\kappa_g\,(\zeta_{j+1} - \zeta_j)}{1 + \kappa_g\,\zeta_{j+1}}\,, \tag{35}$$

which confirms the cancellation: ρ_g depends only on the complexity gradient κ_g and the rung indices. The effect of higher uniform noise on the cutoff operates not through the relative information loss but through the absolute level of precision at every rung: when all signals are noisier, the total option value of learning is smaller, and the continuation gap Γ shrinks toward zero from both sides. By the monotone selection argument in (i), this shifts $\Delta(\cdot)$ toward its flow component, lowering $m_g^\star(v)$.

Force 2: the complexity gradient. A higher κ_g directly increases $\rho_g(j)$: differentiating (35) gives $\partial\rho_g/\partial\kappa_g = (\zeta_{j+1} - \zeta_j)/(1 + \kappa_g\,\zeta_{j+1})^2 > 0$. When noise grows faster with task complexity, promotion sends the worker into a much noisier environment relative to the current rung, widening the Kalman-gain gap and strengthening both channels described above. The option value of staying increases, making Γ more negative and raising $m_g^\star(v)$.

C.3 Equilibrium Wages

This section derives the two wage regimes stated in Section 3.5—the competitive renegotiation wage and the certainty-equivalent (CE) posting—and shows how they combine into the expected wage equation (10). The derivation follows Pastorino (2024) with the modification that the arrival rate λ_g is group-specific. Throughout, $\gamma \equiv \beta(1 - \delta(t))$ denotes the effective discount factor. The key results are: (i) the renegotiation wage equals the runner-up's flow output at the assigned rung (equation (38)); (ii) the CE wage embeds the monopsony markdown $M \equiv \lambda_g\gamma/[1 - \gamma(1 - \lambda_g)]$ from equation (11), which summarizes how search frictions compress wages below marginal product (equation (44)); and (iii) under the common production technology, firm-specific levels and assignment probabilities integrate out under

the EV1 distribution, leaving the expected-wage form used in calibration.

Remark 1 (Log-sum identity)**.** *If* $\{u_j + \eta_j\}$ *are payoffs with* η_j *iid EV1, the standard identity (McFadden, 1974) gives* $\ln \sum_j \exp(u_j) = u_k - \ln \Pr(j = k)$ *for any alternative* k*, where* $\Pr(j = k) = \exp(u_k)/\sum_{j'} \exp(u_{j'})$*. I apply this identity twice below: within firms (across rungs) and across firms (across competitors).*

Derivation of the expected wage (Section 3.5). I begin with the simpler renegotiation wage, then derive the CE wage.

Part I: Renegotiation wage. Conditional on at least one outside offer, let f' denote the runner-up (best poacher). Renegotiation makes the worker indifferent between staying at the incumbent A and moving to f':

$$w^r_{At} \;+\; \gamma\, \mathbb{E}[W(s_{t+1}, \eta_{t+1} \mid j_{At})] \;=\; W(s_t, \eta_t \mid f'). \tag{36}$$

By the robustness refinement, the runner-up's value decomposes as $W(\cdot \mid f') = V_{f'}(\cdot \mid f') - \Pi_{f'}(\cdot \mid A)$, where the runner-up's profit when A hires is the discounted difference between next-period surplus and the worker's continuation wage:

$$\Pi_{f'}(s_t, \eta_t \mid A) \;=\; \gamma\, \mathbb{E}\Big[V_{f'}(s_{t+1}, \eta_{t+1} \mid j_{At})\Big] \;-\; \gamma\, \mathbb{E}\Big[W(s_{t+1}, \eta_{t+1} \mid j_{At})\Big].$$

Substituting into (36), the continuation wage $\gamma\, \mathbb{E}[W(s_{t+1}, \eta_{t+1} \mid j_{At})]$ cancels on both sides, leaving

$$w^r_{At} \;=\; V_{f'}(s_t, \eta_t \mid f') \;-\; \gamma\, \mathbb{E}\Big[V_{f'}(s_{t+1}, \eta_{t+1} \mid j_{At})\Big]. \tag{37}$$

Within-firm surplus is a log-sum over rungs with iid EV1 shocks (Remark 1). Writing $V_{f'}$ as the inclusive value at the runner-up and applying the identity at the incumbent's rung j_{At}

cancels the continuation surplus $\gamma\,\mathbb{E}[V_{f'}(s_{t+1} \mid j_{At})]$, yielding

$$w^r_{At} \;=\; y_{f'}(j_{At}, s_t) \;-\; \ln \Pr_{f'}\big(j = j_{At} \mid s_t\big) \;+\; \eta_{f't}. \tag{38}$$

Bertrand competition pins the wage at the runner-up's flow output: the losing firm bids up to its own valuation, so the renegotiation wage reflects the runner-up's productivity at the assigned rung.

Part II: Certainty-equivalent wage. Let the incumbent be firm A. When no competitor arrives, A posts a CE wage that makes the worker indifferent between accepting and rejecting in favor of the market outside option $O_g(s_t, \eta_t)$. The outside option has a standard search-theoretic structure: with probability λ_g a firm arrives and the worker obtains the best competing wage value; otherwise the worker collects a no-match continuation value $B(s_t, \eta_t)$—the present value of the worker's payoff stream during unemployment, not a per-period flow—and retains the option to search next period:

$$O_g(s_t, \eta_t) \;=\; \lambda_g\,\mathbb{E}_K\Big[\max_{f \in K} W(s_t, \eta_t \mid f)\Big] \;+\; (1-\lambda_g)\Big\{B(s_t, \eta_t) \;+\; \gamma\,\mathbb{E}_{\eta_{t+1}}\big[O_g(s_{t+1}, \eta_{t+1})\big]\Big\}.$$

Write $\bar{O}_g(s_t) \equiv \mathbb{E}_{\eta_t}[O_g(s_t, \eta_t)]$ and $\bar{B}(s_t) \equiv \mathbb{E}_{\eta_t}[B(s_t, \eta_t)]$ for the ex-ante (pre-shock) values. Taking expectations over η_t and using no-match stationarity ($s_{t+1} = s_t$ in the no-match branch) yields the stationary fixed point

$$\bar{O}_g(s_t) = \frac{\lambda_g\,\mathbb{E}_{\eta,K}[\max_f W(s_t, \eta_t \mid f)] + (1-\lambda_g)\,\bar{B}(s_t)}{1 - \gamma(1-\lambda_g)}\,. \tag{39}$$

I pin down $\bar{B}(s_t)$ by a market-indifference condition. The present value of market-sector profits must be the same whether the worker searches or A hires: $\Pi^M_t(M) = \Pi^M_t(A)$, where $\Pi^M_t(M)$ denotes market profits when the worker is in the market and $\Pi^M_t(A)$ when A employs.

In the search state, the stationary market-profit recursion is

$$\Pi_t^M(M) \;=\; \frac{\lambda_g\,\mathbb{E}\big[\Pi_{f'}(s_t,\eta_t \mid f')\big] \;-\; (1-\lambda_g)\,\bar{B}(s_t)}{1-\gamma(1-\lambda_g)}\,,$$

where the winner's hiring profit enters with weight λ_g and the no-match continuation value $\bar{B}$ is a cost to the market sector. On the A-hires path, market-sector firms earn no flow profit; their continuation is

$$\Pi_t^M(A) \;=\; \gamma\,\mathbb{E}\Big[\Pi_{t+1}^M(M)\,\Big|\,j_{At}\Big].$$

By the robustness refinement, the winner's hiring profit on both paths equals $\Pi_{f'}(s_t,\eta_t \mid f') = \gamma\,\mathbb{E}[V_{f'}(s_{t+1} \mid j_{At})] - \gamma\,\mathbb{E}[W(s_{t+1} \mid j_{At})]$. Under stationarity, the $\max_f V_f$ terms that enter $\Pi_t^M(M)$ via the winner's profit and $\Pi_t^M(A)$ via the continuation contain the same expected maximum surplus. Equating $\Pi_t^M(M) = \Pi_t^M(A)$, these terms cancel; what remains involves only $\bar{B}$ and $\mathbb{E}[W]$, giving

$$\bar{B}(s_t) \;=\; \gamma\,(1-\gamma)\;\mathbb{E}[\,W(s_{t+1},\eta_{t+1} \mid j_{At})\,]\,. \tag{40}$$

During unemployment the worker's implicit earnings are the option value of future employment, discounted by both time preference and the separation hazard $(1-\gamma)$.

The realized no-match continuation value $B(s_t,\eta_t)$ differs from $\bar{B}(s_t)$ by a term proportional to the aggregate productivity shock. Within-firm surplus is a log-sum over rungs (Remark 1), so the winner's profit in the market-profit recursion shifts additively in the aggregate EV1 shock η_t. Conditioning the market-profit recursion on η_t, subtracting the ex-ante version, and using market indifference (which sets the left-hand side to zero) yields

$$B(s_t,\eta_t) = \bar{B}(s_t) \;+\; \frac{\lambda_g}{1-\lambda_g}\,\eta_t. \tag{41}$$

Define $\tilde{\eta}_t \equiv \frac{\lambda_g}{1-\lambda_g}\,\eta_t$. The rescaling reflects arrival-rate amplification: when offers are rare (low λ_g), the no-match value is more sensitive to aggregate shocks because the worker must wait

longer to exploit a favorable draw through employment.

With $\bar{O}_g$ and B in hand, I derive the CE wage. The worker's present value at A mixes the competitive branch (probability λ_g) with the no-arrival branch:

$$W(s_t, \eta_t|A) = \lambda_g \, \mathbb{E}\left[\max_f W(s_t, \eta_t \mid f)\right] + (1-\lambda_g)\big[w^{ce}_{At} + \gamma \, \mathbb{E}[W(s_{t+1}, \eta_{t+1} \mid j_{At})]\big]. \tag{42}$$

Equating $W(s_t, \eta_t \mid A) = O_g(s_t, \eta_t)$ (worker indifference), the λ_g branches cancel across both sides. The no-arrival branch on the market side features no learning, so its continuation equals the stationary value $\bar{O}_g(s_t)$. Rearranging:

$$w^{CE}_{At}(s_t, \eta_t) \;=\; B(s_t, \eta_t) \;+\; \gamma \, \bar{O}_g(s_t) \;-\; \gamma \, \mathbb{E}\big[W\big(s_{t+1}, \eta_{t+1} \mid j_{At}\big)\big] \, . \tag{43}$$

Substituting $B(s_t, \eta_t) = \bar{B}(s_t) + \tilde{\eta}_t$ from (41) separates the shock-contingent piece:

$$w^{CE}_{At}(s_t, \eta_t) = \bar{B}(s_t) + \gamma \, \bar{O}_g(s_t) - \gamma \, \mathbb{E}\big[W\big(s_{t+1}, \eta_{t+1} \mid j_{At}\big)\big] + \tilde{\eta}_t. \tag{44}$$

The CE wage compensates the worker for the no-match continuation value $\bar{B}$, plus the discounted option value of future search $\gamma \bar{O}_g$, minus the continuation she would earn under the current assignment, plus the shock-contingent premium $\tilde{\eta}_t$.

Part III: Expected wage. I now combine the two regimes to obtain the expected wage in (10). Throughout, I use the common production technology: all firms share the same ability return c_j and the same observable component $\phi(j, g, t)$ at each rung. Firms differ only in their additive productivity levels and their iid EV1 draws.

Step 1: Renegotiation wage in expectation. I take expectations of (38) over the identity of the runner-up f'. Substituting the output decomposition $y_{f'}(j_{At}, s_t) = \phi(j, g, t) + c_j \, m_t + \psi_{f'}$,

the renegotiation wage becomes

$$w^r_{At} \;=\; \phi(j,g,t) + c_j\, m_t \;+\; \psi_{f'} \;-\; \ln \Pr_{f'}\big(j{=}j_{At} \mid s_t\big) \;+\; \eta_{f't}.$$

The first two terms are common across firms. The remaining three vary with the runner-up's identity: $\psi_{f'}$ is the runner-up's firm-specific productivity level, so $\mathbb{E}_{f'}[\psi_{f'}]$ is a constant at given s_t; by the EV1 distributional assumption, $\mathbb{E}[-\ln \Pr_{f'}(j{=}j_{At} \mid s_t)]$ is a deterministic function of s_t; and $\mathbb{E}[\eta_{f't}] = 0$. Collecting:

$$\mathbb{E}\big[w^r_t \mid j,\, s_t\big] \;=\; \phi(j,g,t) \;+\; c_j\, m_t \;+\; \varepsilon^r_t, \tag{45}$$

where ε^r_t absorbs $\mathbb{E}[\psi_{f'}]$, the mean of the assignment premium, and the mean-zero shock. Bertrand competition delivers the full expected output at the assigned rung.

Step 2: CE wage in expectation. Recall the monopsony markdown $M \equiv \frac{\lambda_g\,\gamma}{1-\gamma(1-\lambda_g)}$ from equation (11). I substitute $\bar{O}_g(s_t)$ from (39) and $\bar{B}(s_t)$ from (40) into the CE wage (44). First, combining $\bar{B}$ and $\gamma\bar{O}_g$:

$$\bar{B}(s_t) + \gamma\,\bar{O}_g(s_t) \;=\; M\;\mathbb{E}_\eta\!\left[\max_f\, W(s_t,\eta_t \mid f)\right] \;+\; \frac{\bar{B}(s_t)}{1-\gamma(1-\lambda_g)}\,.$$

Substituting $\bar{B} = \gamma(1-\gamma)\,\mathbb{E}[W(s_{t+1} \mid j_{At})]$ from (40) and collecting the continuation-value terms:

$$w^{CE}_{At} \;=\; M\left\{\mathbb{E}_\eta\!\left[\max_f\, W(s_t,\eta_t \mid f)\right] \;-\; \gamma\,\mathbb{E}[W(s_{t+1},\eta_{t+1} \mid j_{At})]\right\} \;+\; \tilde{\eta}_t, \tag{46}$$

where the coefficient on $\mathbb{E}[W(s_{t+1})]$ simplifies to $-M\gamma$ because $\frac{\gamma(1-\gamma)}{1-\gamma(1-\lambda_g)} - \gamma = -\frac{\lambda_g\gamma^2}{1-\gamma(1-\lambda_g)} = -M\gamma$. The CE wage is proportional to the markdown M times the net competitive value: the expected best-offer value less the discounted continuation the worker would earn at the incumbent.

To convert $\mathbb{E}_\eta[\max_f W]$, note that conditional on s_t each firm's surplus is a log-sum over rungs plus an iid EV1 firm-time shock (Remark 1). Under robustness, $W(\cdot \mid f)$ inherits the additive EV1 structure from the surplus because the runner-up's profit is a deterministic function of s_t and j_{At}. Applying the log-sum identity across firms and taking expectations, firm-specific levels and assignment probabilities average to state-dependent constants by the same logic as Step 1, and the $\gamma\,\mathbb{E}[W(s_{t+1})]$ continuation terms cancel. This leaves

$$\mathbb{E}\big[w_t^{CE} \mid j,\, s_t\big] \;=\; M\,\big[\phi(j,g,t) \;+\; c_j\,m_t\big] \;+\; \varepsilon_t^{CE}. \tag{47}$$

The CE wage delivers a marked-down fraction M of expected output.

Step 3: Combining. The expected wage mixes the two regimes:

$$\mathbb{E}[w_t] \;=\; \lambda_g\,\mathbb{E}[w_t^r] \;+\; (1-\lambda_g)\,\mathbb{E}[w_t^{CE}] \;=\; \big[\lambda_g + (1-\lambda_g)\,M\big]\big[\phi(j,g,t) + c_j\,m_t\big] \;+\; \varepsilon_t.$$

Decompose $\phi(j,g,t) = \alpha_j + h_{gt} + \psi_f$ (from equation (4)), where α_j is the rung baseline, h_{gt} is accumulated human capital, and ψ_f is the firm-specific level. The firm level ψ_f averages out under the common technology. Absorbing α_j into the residual:

$$\mathbb{E}[w_t] \;=\; \big[\lambda_g + (1-\lambda_g)\,M\big]\,\big[c_j\,m_t + h_{gt}\big] \;+\; \varepsilon_t. \tag{48}$$

Human capital is portable—all firms value it equally—but during unemployment the worker earns no output, so search frictions mark down the return to human capital by the same factor as the ability rent. Define $\tilde{h}_{gt} \equiv \big[\lambda_g + (1-\lambda_g)\,M\big]\,h_{gt}$. Since the markdown is common across groups and periods, it is absorbed by the estimated human-capital parameters. Writing h_{gt} for $\tilde{h}_{gt}$ hereafter yields

$$\mathbb{E}[w_t] \;=\; \big[\lambda_g + (1-\lambda_g)\,M\big]\,c_j\,m_t \;+\; h_{gt} \;+\; \varepsilon_t,$$

which is equation (10). □

C.4 Migrant Stopping Rule

Result 4 states that two-sided exit arises when the outside option is sufficiently convex in the posterior mean m_t relative to the stay value. This section develops that claim using the quadratic parameterization of the exit value adopted in Section 4.1 and explains the logit kernel that smooths the discrete stay–exit comparison into a continuous exit probability.

Two-sided selection. The stay value combines expected flow wages with the discounted continuation value of all future wages:

$$W_t^{\text{stay}}(s_t) \;=\; \underbrace{\lambda_g\,\mathbb{E}\big[w_t^r\big] + (1-\lambda_g)\,w_t^{ce}}_{\text{expected flow wage}} + \gamma\;\underbrace{\mathbb{E}\big[W(s_{t+1})\big]}_{\text{continuation value}}. \tag{49}$$

It is increasing in m_t: the flow wage is linear in m_t (from the wage equation), while the continuation value adds strict convexity. The convexity arises because the value function $W_{t+1}(m_{t+1})$ is convex in beliefs—higher m_{t+1} unlocks promotions with steeper ability returns ($c_{j+1} > c_j$), making the value function the upper envelope of functions with different slopes—and Bayesian updating preserves this convexity, since $m_{t+1} = m_t + \varepsilon$ with innovation ε independent of m_t. The exit value $V_t^{\text{exit}}(m_t)$ depends on perceived ability through $f(m_t) = \mu_1\, m_t + \mu_2\, m_t^2$ (equation 17). This is a polynomial of degree two—the simplest parameterization that satisfies the convexity condition in Result 4 when $\mu_2 > 0$. Consider the deterministic value gap

$$D(m_t) \;\equiv\; V_t^{\text{exit}}(m_t) \;-\; W_t^{\text{stay}}(s_t).$$

Since W_t^{stay} is approximately linear in m_t at any given state (the flow wage is linear, and the continuation value's curvature is bounded), the quadratic term $\mu_2\, m_t^2$ eventually dominates at both tails. Migrants stay—$D(m_t) < 0$—only in the interior of the belief distribution, generating two-sided selection into staying:

- At the *lower tail* (low m_t), the stay value is low because German career prospects are poor—low current wages, slow promotion, limited upside from further learning—so $D(m_t) > 0$ and exit dominates. These workers are negatively selected out.

- At the *upper tail* (high m_t), the quadratic causes V_t^{exit} to rise faster than W_t^{stay}. Workers whose ability exceeds what the German market has recognized find that the outside option already rewards their type, so $D(m_t) > 0$ again. These workers are positively selected out.

Between these two regions, $D(m_t) < 0$ and the stay value dominates. The U-shaped gap generates two-sided exit consistent with the hump-shaped excess exit documented in Section 2.3.2.

Logit kernel smoothing. Without a taste shock, exit would be a deterministic, knife-edge comparison: a worker exits if and only if $V_t^{\text{exit}} > W_t^{\text{stay}}$. This is unappealing both computationally—it produces discontinuous exit probabilities that make calibration unstable—and substantively, because migrants' exit decisions reflect heterogeneous non-pecuniary factors (family ties, cultural attachment, risk attitudes) that the model does not observe.

The iid Gumbel taste shock ξ_t with scale ξ (equation 18) smooths the binary decision into a continuous probability. The Gumbel distributional assumption yields the standard logit form (Train, 2009, ch. 3): the exit probability is

$$\Pr(\text{exit} \mid m_t, s_t) \;=\; \frac{1}{1 + \exp\big(\big[W_t^{\text{stay}} - V_t^{\text{exit}}\big]/\xi\big)},$$

a logistic function of the stay–exit value gap scaled by ξ. The scale parameter governs the sharpness of the exit response. When ξ is small, exit probabilities are concentrated near zero or one and the decision approximates a deterministic cutoff rule. When ξ is large, non-pecuniary motives add substantial dispersion and exit probabilities respond gradually to changes in the value gap. The calibrated ξ (Section 4.3) balances these forces to match the

observed level and experience profile of migrant exit rates.

D Calibration Details

D.1 Parameter Summary

Table 11 lists all model parameters, organized into normalizations, parameters set outside the calibration, and free parameters estimated by simulated method of moments.

Table 11: Model Parameters

Symbol	Description	Value	Source / Role
Panel A. Normalizations			
$m_{0,L}$	Local prior mean	0	
v_0	Prior variance	1	
c_0	Ability leverage, rung 0	1	
Panel B. Set externally			
β	Annual discount factor	0.96	Standard
λ	Competitor arrival rate	0.157	Wage mixture model
δ_0	Baseline separation rate	0.0065	NLS on local perm. exit
δ_1	Separation decay	0.070	NLS on local perm. exit
σ_ε	Wage measurement error	0.10	Residual wage dispersion
$m_{0,M}$	Migrant prior mean	—	Pinned to entry wage gap
Panel C. Calibrated (14 free parameters)			
σ^2_{L0}	Local baseline signal variance	Signal noise level	
R	Migrant noise amplifier (> 1)	Migrant learning penalty (level)	
κ_L	Local noise gradient	Noise increase with complexity (locals)	
$\kappa_M - \kappa_L$	Migrant excess gradient	Extra noise increase (migrants)	
α_0	HC accumulation rate (locals)	Experience-wage profile level	
ρ_α	Migrant HC ratio (α_M/α_0)	Migrant HC penalty	
c_1	Ability leverage, mid	Talent return at rung 1	
c_2	Ability leverage, high	Talent return at rung 2	
κ_0	Promotion threshold, $0 \to 1$	Low $\to$ mid friction	
κ_1	Promotion threshold, $1 \to 2$	Mid $\to$ high friction	
μ	Outside-option baseline	Exit level	
μ_1	Roy slope (linear in m)	Belief sensitivity of exit value	
μ_2	Roy slope (quadratic in m)	Convexity of exit value in beliefs	
ξ	Gumbel scale (exit smoothness)	Dispersion of non-pecuniary exit motives	

Notes: Panel A lists quantities normalized without loss of generality. Panel B lists parameters fixed prior to calibration; "—" for $m_{0,M}$ indicates that it is pinned analytically at each evaluation to match the observed entry wage gap. Panel C lists the 14 parameters estimated by simulated method of moments.

Table 12 reports the full set of calibrated parameter values, including signal variances, implied Kalman gains by rung and group, and pre-set parameters.

Table 12: Calibrated model parameters

Parameter	Description	Value	Interpretation
A. Signal structure (learning speed)			
$\sigma^2_{L,0}$	Base signal noise (locals)	3.499	
R	Migrant noise multiplier	2.777	$\sigma^2_{M,0} = 26.980$
κ_L	Local noise gradient	8.870	
κ_M	Migrant noise gradient	13.244	
B. Implied Kalman gains at entry (learning speed)			
$K^0_{L,0}$	Lower (local)	0.2223	Signal-to-noise: 3.50
$K^0_{M,0}$	Lower (migrant)	0.0357	Signal-to-noise: 26.98
$K^0_{L,1}$	Specialist (local)	0.0500	Signal-to-noise: 19.02
$K^0_{M,1}$	Specialist (migrant)	0.0048	Signal-to-noise: 205.64
$K^0_{L,2}$	Expert (local)	0.0281	Signal-to-noise: 34.54
$K^0_{M,2}$	Expert (migrant)	0.0026	Signal-to-noise: 384.30
	Kalman gain ratio (rung 0)	0.161	Migrants learn 6.2x slower
C. Human capital accumulation			
α_L	Local HC rate	0.0402	0.562 log-pts over 14 yrs
α_M/α_L	Migrant HC ratio	0.720	$\alpha_M = 0.0289$
D. Ability leverage by rung			
c_0	Lower	1.000	Ability amplification
c_1	Specialist	1.181	Ability amplification
c_2	Expert	1.281	Ability amplification
E. Promotion friction			
$\bar{\kappa}_0$	Lower $\rightarrow$ Specialist	3.848	Surplus threshold
$\bar{\kappa}_1$	Specialist $\rightarrow$ Expert	1.727	Surplus threshold
F. Outside option (endogenous exit)			
μ	Base outside option	0.0765	
μ_m	Belief slope	0.5499	Higher ability $\rightarrow$ better outside option
μ_{m^2}	Belief curvature	0.0959	
ξ	Taste shock scale	1.0341	Gumbel dispersion
G. Fixed / pre-set			
λ	Arrival rate	0.157	Common to both groups
δ_0	Base separation	0.026	$\delta(t) = \delta_0 e^{-\delta_1 t}$
δ_1	Separation decay	0.10	

Notes: Parameters calibrated by simulated method of moments (differential evolution). Kalman gains computed at prior $v_0 = 1.000$. HC accumulation is linear: $h(t) = \alpha_g \cdot t$.

D.2 Parameter Transformations

Several transformations are applied during estimation to enforce parameter constraints. The baseline signal variance σ_{L0}^2 is log-parameterized to ensure positivity. The migrant noise amplifier is parameterized as $R = 1 + \exp(r)$ to enforce $R > 1$. The migrant noise gradient enters as $\kappa_M = \kappa_L + \exp(\cdot)$ to enforce $\kappa_M \geq \kappa_L$. The ability leverage parameters are parameterized as ratios $c_1 = 1 + \exp(\cdot)$ and $c_2 = c_1 \cdot (1 + \exp(\cdot))$ to enforce the ordering $c_0 \leq c_1 \leq c_2$.

D.3 Moment Block Descriptions

The 32 targeted moments are organized into nine blocks (A–I). Each block disciplines a distinct subset of the model's free parameters.

Wage dynamics (blocks A–B). Block A targets the migrant–local wage gap at experience 0, 3, 5, 10, and 14. Starting from the positive entry gap, the migrant advantage erodes and reverses: by experience 14, the within-plant gap has shifted by roughly -5.4 log points. Block B targets the local experience-wage profile at the same horizons (relative to entry). Together, these nine moments pin down the human-capital parameters (α_0, ρ_α) and, jointly with the promotion moments below, the interaction between ability leverage (c_j) and learning speed: a high α_0 generates steep wage growth for locals, while a low ρ_α depresses migrant wage growth even before differential promotions take effect.

Promotion dynamics (blocks C–D, G–I). Blocks C and D target the migrant–local gaps in the probability of holding a mid-or-above position and a high position, respectively, at experience 0, 5, and 10. Both gaps are near zero at entry and widen substantially: the mid-and-high gap reaches -5.1 percentage points by experience 10 and the high gap reaches -3.2 percentage points. Blocks G and I target the local promotion dynamics directly—the change in local mid-and-high and high shares relative to entry at experience 3, 5, 10, and 14. These

eight moments discipline the promotion thresholds (κ_0, κ_1), the ability leverage parameters (c_1, c_2), and the noise gradient (κ_L): higher c_j raises the return to promotion and lowers the effective promotion cutoff in beliefs, while higher κ_L increases the option value of waiting and slows promotion. The cross-group gaps in blocks C and D provide additional leverage for the migrant-specific noise parameters (R, $\kappa_M - \kappa_L$), since slower learning and a steeper noise gradient both delay the point at which migrants cross the promotion cutoffs.

Within-rung wage gaps (block H). Block H targets the wage gap between migrants and locals *within the same rung* at experience 5, 10, and 14, weighted by the local rung distribution. This conditional gap isolates the learning channel from the assignment channel: two workers at the same rung differ in wages only through their posterior means m_t and their respective human-capital flows, not through occupational sorting. Because the within-rung gap is driven almost entirely by the speed at which beliefs converge, block H provides sharp identification of the migrant noise amplifier R and of ρ_α, conditional on the local experience profile pinned by block B.

Exit dynamics (blocks E–F). Block E targets the migrant excess exit rate—the difference in exit probabilities between migrants and locals—at experience 0, 3, 5, and 10. This excess is large at entry (≈ 3 percentage points) and declines toward zero as experience accumulates. Block F targets the stagnation–exit gradient: the OLS coefficient of exit on within-plant wage growth, estimated separately for early-career (experience 0–5) and mid-career (experience 6–10) migrants. A more negative coefficient indicates that migrants whose wages stagnate are more likely to exit, a feature that disciplines the Roy slopes μ_1 and μ_2. Together, blocks E and F identify the four outside-option parameters (μ, μ_1, μ_2, ξ).

Table 13 reports the empirical targets alongside the model's simulated counterparts.

Table 13: Model fit: simulated vs. empirical moments

Moment	Data	Model	Deviation
H. Exit rate excess (migrant - local)			
Experience 0	+0.0298	+0.0250	-0.0048
Experience 3	+0.0118	+0.0144	+0.0026
Experience 5	+0.0116	+0.0102	-0.0014
Experience 10	+0.0041	+0.0060	+0.0019
C. Within-rung wage gap (conditional)			
Experience 5	-0.0060	-0.0141	-0.0081
Experience 10	-0.0374	-0.0240	+0.0134
Experience 14	-0.0417	-0.0363	+0.0054
I. Stagnation-exit link			
Experience 0–5	-0.0312	-0.0145	+0.0167
Experience 6–10	-0.0124	-0.0074	+0.0050
A. Wage gap (migrant - local)			
Experience 0	+0.0157	+0.0158	+0.0001
Experience 3	+0.0048	-0.0050	-0.0098
Experience 5	-0.0108	-0.0151	-0.0043
Experience 10	-0.0488	-0.0450	+0.0038
Experience 14	-0.0540	-0.0688	-0.0148
G. Local high share change (rel. to entry)			
Experience 3	+0.0074	+0.0000	-0.0074
Experience 5	+0.0111	+0.0001	-0.0110
Experience 10	+0.0195	+0.0083	-0.0112
Experience 14	+0.0273	+0.0157	-0.0116
E. High share gap (migrant - local)			
Experience 0	+0.0056	+0.0000	-0.0056
Experience 5	-0.0071	-0.0001	+0.0070
Experience 10	-0.0315	-0.0078	+0.0237
F. Local mid-and-high share change (rel. to entry)			
Experience 3	+0.0154	+0.0004	-0.0150
Experience 5	+0.0238	+0.0005	-0.0233
Experience 10	+0.0338	+0.0172	-0.0166
Experience 14	+0.0400	+0.0268	-0.0132
D. Mid-and-high share gap (migrant - local)			
Experience 0	+0.0029	+0.0000	-0.0029
Experience 5	-0.0259	-0.0005	+0.0254
Experience 10	-0.0506	-0.0158	+0.0348
B. Local wage-experience profile (rel. to entry)			
Experience 3	+0.1523	+0.0899	-0.0624
Experience 5	+0.2076	+0.1482	-0.0594
Experience 10	+0.3008	+0.3034	+0.0026
Experience 14	+0.3559	+0.4232	+0.0673
Weighted SSE (objective): 1145.1			

Notes: Panels sorted by goodness of fit (best first). Deviations computed as Model − Data. Weighted SSE is the calibration objective (sum of squared deviations weighted by moment-specific weights; see Appendix D). Moments estimated from administrative matched employer-employee data. Model

D.4 Moment Weights

The baseline weight for each moment is inverse-variance, $w_k = 1/\mathrm{SE}_k^2$, with block-specific scaling factors. Exit moments (block E) are upweighted by a factor of 10, promotion dynamics (blocks C, D, G, I) by 2–3, and the stagnation–exit gradient (block F) by 3. The heavier weight on exit reflects a practical difficulty: because selective exit of low-m migrants narrows the wage gap among survivors, the optimizer can reduce the wage-gap loss by suppressing exit. The upweighting forces it to maintain realistic exit rates.

D.5 Descriptive Statistics

Table 14 reports unconditional transition rates and complexity-rung distributions by group and experience. These statistics are not calibration targets but provide context for the pre-set parameters and the structural interpretation of the model.

Common arrival rate. The J2J transition rates in Panel A are 7.9% for locals and 10.1% for migrants. This modest gap is consistent with the assumption of a common competitor arrival rate λ across groups. In the model, a higher migrant J2J rate can arise endogenously even under equal λ: noisier signals about migrant ability produce more frequent mispricing, which raises the probability that a poacher's offer exceeds the incumbent's posted wage. The observed gap thus reflects differential acceptance behaviour rather than differential search intensity, and imposing a common λ avoids introducing a free parameter that would be poorly identified separately from the noise structure.

Permanent exit and rung dynamics. Panel A also shows that migrants' permanent exit rates are substantially higher at entry (8.5% vs. 2.2%) but converge toward locals by experience 10 (1.3% vs. 0.9%). This declining excess exit motivates the Roy-selection exit mechanism in the model. Panel B documents the complexity-rung reversal: migrants start at slightly *higher* rung shares (24.9% vs. 21.9% at Spezialist+; 17.4% vs. 11.1% at Experte), but

Table 14: Descriptive Statistics

	Locals		Migrants	
	Rate	N	Rate	N
Panel A: Transition Rates				
J2J rate	0.079	1,107,725	0.101	176,766
Perm. exit, exp 0	0.022	19,885	0.085	12,310
Perm. exit, exp 3	0.012	36,918	0.029	11,002
Perm. exit, exp 5	0.011	40,093	0.022	9,007
Perm. exit, exp 10	0.009	38,746	0.013	5,842
Panel B: Complexity Rung Distribution				
	P(Spezialist+)		N	
Exp	Locals	Migrants	Locals	Migrants
0	0.219	0.249	19,885	12,310
5	0.403	0.330	40,093	9,007
10	0.438	0.308	38,746	5,842
14	0.448	0.279	38,143	5,158
	P(Experte)		N	
Exp	Locals	Migrants	Locals	Migrants
0	0.111	0.174	19,885	12,310
5	0.223	0.215	40,093	9,007
10	0.244	0.184	38,746	5,842
14	0.254	0.158	38,143	5,158
Observations				
Total person-year obs				1,284,491
Locals				1,107,725
Migrants				176,766

Notes: Source: IEB 2% sample, 2012–2022. Full-time male workers, age 18–55. Panel A rates are unconditional annual means. Panel B shows raw (unconditional) rung shares.

fall behind locals by experience 5 and the gap widens through experience 14. This pattern disciplines the promotion thresholds and the interaction between learning speed and the option value of waiting for better information.

D.6 Optimization

The 14-dimensional parameter space is explored using differential evolution (DE), a population-based global optimizer that does not require gradient information and handles the non-convexities inherent in the mapping from parameters to simulated moments. The population consists of 25 candidate vectors. Five are hand-selected seeds that span economically distinct regions of the parameter space (low vs. high noise, active vs. suppressed exit); the remaining 20 are drawn uniformly from the bounds listed in Table 11. I use the "plateau" stopping rule: the algorithm terminates after 30 consecutive generations without an improvement exceeding 1.0 in the objective $Q(\theta)$, with a minimum of 30 generations enforced before the rule activates. Four parallel workers evaluate the objective concurrently.

After the global search converges, I refine the best candidate with L-BFGS-B, a quasi-Newton method that respects box constraints. This polishing step is capped at 1,800 function evaluations, 200 iterations, or 20 minutes of wall-clock time, whichever binds first, with early termination if no improvement above 10^{-3} is found over 250 consecutive evaluations.

D.7 Computational Details

Each evaluation of $Q(\theta)$ proceeds in three steps. First, the dynamic programming problem is solved by backward induction on a grid of 400 belief-mean points, 25 log-spaced variance points, 3 rungs, and 40 experience levels, separately for each group. The expectation over next-period beliefs—which enters both the surplus and the worker value—is computed by Gauss–Hermite quadrature with 15 nodes. The inner loop is accelerated with just-in-time compilation. Second, a panel of $N = 50{,}000$ workers per group is simulated forward for $T = 15$ experience periods, drawing ability realizations, signals, exit shocks, and separations

from the model's distributions. Third, the 32 moments are computed from the simulated panel and compared to the data targets. The large simulation sample reduces the Monte Carlo variance of the objective function, ensuring that the optimizer responds to structural features of the parameter space rather than simulation noise.