# Online Appendix to: Labor Market Pooling under Demographic Decline: Evidence from Japan's Tokyo Concentration

Tomoya Mori[*] Miki Ogawa[†]

August 27, 2026

This Online Appendix complements the main paper (referenced as "main paper" below). It contains supplementary material on long-term cohort evidence, data construction, the demand-constrained indicator, the education margin, hinterland figures, weighting sensitivity, the family-driven-mover noise-removal robustness check, and detailed country-by-country external-validation evidence. Cross-references prefixed with "main:" point into the main paper.

## Contents



[*]Institute of Economic Research, Kyoto University, Yoshida-honmachi, Sakyo-ku, Kyoto 606-8501, Japan; and Research Institute of Economy, Trade and Industry (RIETI). Email: mori@kier.kyoto-u.ac.jp.

[†]Graduate School of Economics, Kyoto University, Yoshida-honmachi, Sakyo-ku, Kyoto 606-8501, Japan. Email: ogawa.miki.26v@st.kyoto-u.ac.jp.

## OA.1 Supplementary descriptive evidence

This appendix collects supplementary descriptive evidence, each piece of which appears as a single sentence in Section 2: (i) the long-run national contraction of the mobile young workforce, (ii) the age-band breakdown of the inter-prefectural migration rate, (iii) the post-2010 net in-migration to the four major MAs by age, (iv) the Tokyo MA migration rate in the JPSED panel, and (v) the single-age decomposition of Tokyo MA's 20–29 net inflow.

**The shrinking mobile young workforce.** Figure OA.1 shows the national population aged 20–29 and 30–39 from 1954 through 2025. The 20–29 cohort peaked at roughly 18 million in the early 1970s and now stands at about 11 million—a 40% decline over five decades. The 30–39 cohort followed a similar trajectory with a delay.

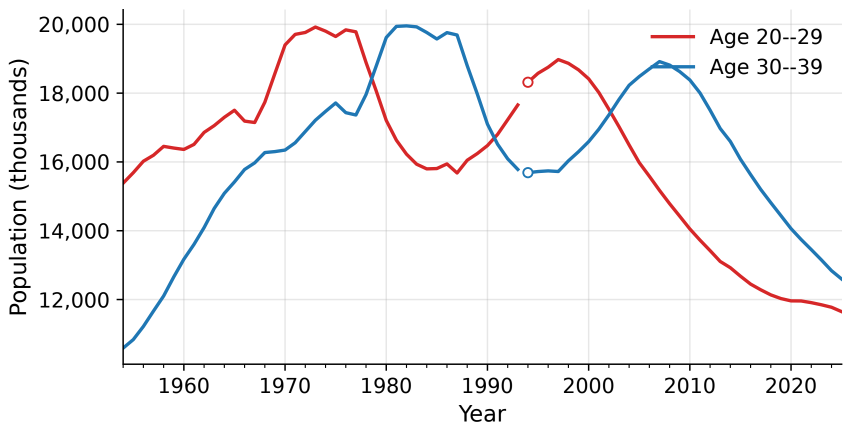



**Figure OA.1:** National population aged 20–29 and 30–39, 1954–2025.

*Notes:* The 20–29 cohort peaked near 18 million in the early 1970s and stands at about 11 million by 2025 — a 40% fall over five decades — with the 30–39 cohort following at a lag. Sources: National Institute of Population and Social Security Research (IPSS) demographic series (1-year ages aggregated to 5-year bands) for 1954–2024; Basic Resident Registry (Japanese residents only, 5-year age bands) for 2025 (Ministry of Internal Affairs and Communications, 1954–2025). A small level break appears at 2025 because JBRR excludes foreign residents (about 5% of working-age population), but the long-run trend is preserved.

**Inter-prefectural migration is dominated by the young.** Table OA.1 reports inter-prefectural in-migration rates by age band (Basic Resident Registry; methodology in table notes). Two facts dominate. First, the 20–29 cohort migrates at roughly 2–2.5× the rate of cohorts aged 30 and above (9.6%/yr in 2025 vs. 4.0% at 30–39, 1.4% at 40–49, and 0.9% at 50–64). Second, the 20–29 share of all inter-prefectural moves has risen from ≈ 37% (2010) to 45% (2025), despite the shrinking absolute size of the young population. The migration patterns we analyze in the JPSED panel are therefore overwhelmingly young-worker patterns.

**Age-asymmetric Tokyo MA in-migration accelerated post-2010.** Figure OA.2 disaggregates net in-migration into the three metropolitan areas and Fukuoka prefecture by age band, 2010–2025, from the Basic Resident Registry. The pattern is striking. Panel A (ages 20–29, the first-job-entry and early-career cohort): Tokyo MA's net in-migration rose from approximately +60,000 in 2010 to over +110,000 per year by 2024–25, while Osaka MA stayed essentially flat near zero, Nagoya MA turned sharply negative, and Fukuoka prefecture remained slightly negative. Panel

**Table OA.1:** Inter-prefectural in-migration rate by age band, percent of age-band population per year.

| Age | 2010 | 2015 | 2020 | 2025 | Change |
|---|---|---|---|---|---|
| 20–29 | 6.1 | 6.9 | 8.8 | 9.6 | +3.5 |
| 30–39 | 3.0 | 3.3 | 3.6 | 4.0 | +1.0 |
| 40–49 | 1.4 | 1.4 | 1.4 | 1.4 | +0.0 |
| 50–64 | 0.7 | 0.7 | 0.8 | 0.9 | +0.2 |
| 20–29 share (%) | 36.8 | 37.4 | 42.7 | 44.7 | +7.9 |

*Notes:* Both numerator and denominator are from the Basic Resident Registry (Japanese residents). Numerator: JBRR single-year-age national in-migration totals summed within each band. Denominator: JBRR national population at the same 5-year bands. The bottom row is the share of all inter-prefectural moves accounted for by the 20–29 cohort.

B (ages 30–39): Tokyo MA's net in-migration is modestly positive but an order of magnitude smaller than the 20–29 inflow, and the other three destinations are flat or marginally positive. Panel C (ages 40 and above): Tokyo MA is actually a *net loser* of older workers in most years (net −1 to −13K per year combining 40–49, 50–59, and 60+), and Osaka MA, Nagoya MA, and Fukuoka prefecture all sit near or below zero as well. The contrast across the panels is the point: net in-migration is now almost entirely a phenomenon of the 20s—the ages of labor-market entry and early career, and by far the most mobile band (Table OA.1)—and within that band Tokyo MA is the only major center that absorbs. Osaka and Nagoya, the major secondary destinations of the late twentieth century, no longer attract any age band, and the older cohorts are drawn to none of the four major centers.

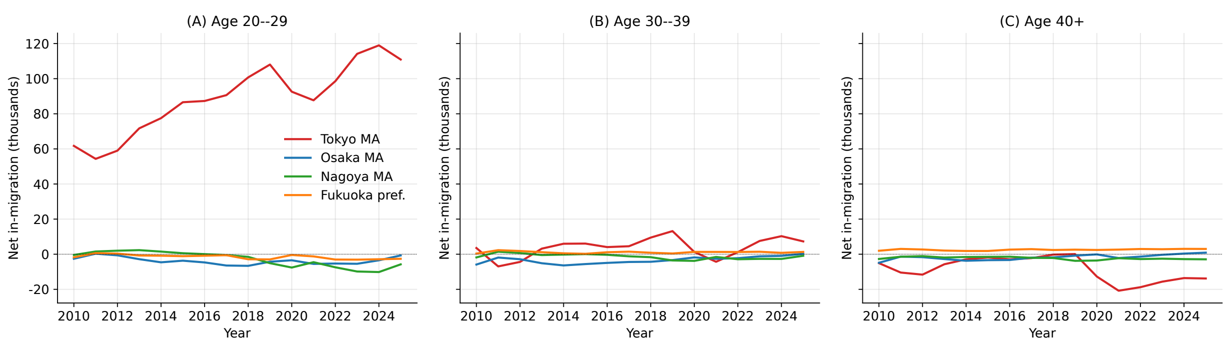



**Figure OA.2:** Net in-migration to the three major metropolitan areas and Fukuoka prefecture, by age band, 2010–2025 (thousands).

*Notes:* The three major MAs are Tokyo, Osaka, and Nagoya. Panel A: ages 20–29; Panel B: ages 30–39; Panel C: ages 40 and above. Tokyo MA's 20–29 net in-migration has roughly doubled (from ∼60K to ∼110K per year) while Osaka MA, Nagoya MA, and Fukuoka prefecture remain near zero or turn negative; the 30–39 band shows much smaller flows for all four, with Tokyo MA modestly positive; and Tokyo MA is a net loser of the 40+ cohort, with the other three also at or below zero for that older band. Source: Basic Resident Registry single-year-age in/out flows aggregated to the three bands (Ministry of Internal Affairs and Communications, 1954–2025).

**JPSED Tokyo MA migration rate.** The JPSED panel reproduces the JBRR pattern: across the nine annual transitions (2016/17–2024/25), the young (20–29) Tokyo MA inflow rate exceeds

1%/yr while the 30+ inflow and Tokyo-MA outflow rates stay flat around 0.2–0.35%, and the same age asymmetry holds across the four major MAs, with Tokyo's young inflow several times the other three destinations' (Figure OA.3).

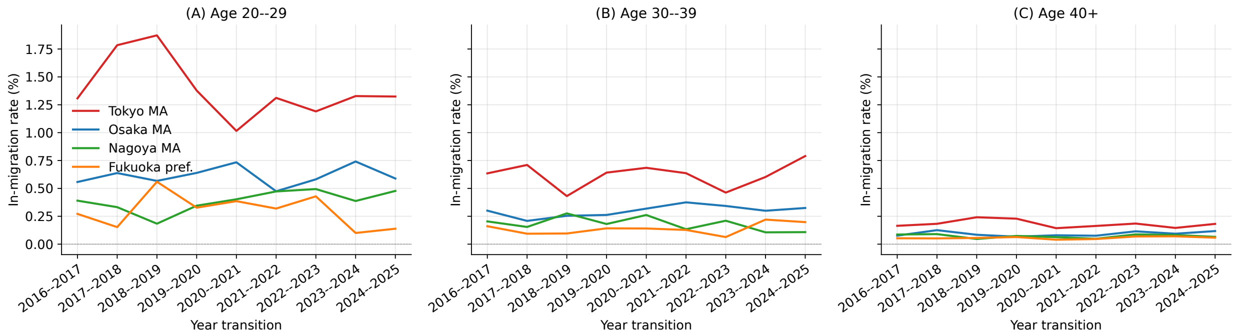



**Figure OA.3:** In-migration rate to the three major metropolitan areas and Fukuoka prefecture by age band, JPSED 2016/17–2024/25 (%).

*Notes:* The three major MAs are Tokyo, Osaka, and Nagoya; annual transitions. Panel A: ages 20–29; Panel B: ages 30–39; Panel C: ages 40 and above. Tokyo MA's 20–29 inflow rate is several times higher than at the other three destinations throughout; the 30–39 and 40+ bands sit at much lower levels for all four, with Tokyo MA still leading but the gap narrower. JPSED cross-sectional weights (Recruit Works Institute, 2016–2025).

**Single-age decomposition of Tokyo MA's 20–29 net inflow.** Decomposing the Tokyo MA 20–29 *net* in-migration (gross in minus gross out, single-year-age aggregated over the four Tokyo MA prefectures, Basic Resident Registry) by single year of age, age 22 accounts for about 28–33% of the net 20s inflow throughout 2010–2025 (31.2% in 2010, 28.2% in 2015, 30.2% in 2019, 32.8% in 2025). Ages 20–21 contribute another ∼18–31%, declining over time as gross-out at these ages has risen, while the post-graduate ages 23–29 account for the remaining ∼38–52%, rising over time (Figure OA.4). In absolute terms, age 22 net in-migration grew 89% between 2010 and 2025 (from 19.3K to 36.4K per year), ages 20–21 grew only 5% (19.0K to 20.0K, essentially flat), and ages 23–29 grew 132% (23.5K to 54.5K)—the post-graduate ages are the fastest-growing component of Tokyo MA's net 20s absorption. Counted by gross inflows alone — ignoring who leaves — the shares look different: age 22 ∼15%, ages 20–21 ∼12%, ages 23–29 ∼72%. The reason is that at the post-graduate ages many people also leave Tokyo MA: the gross inflow at 23–29 is by far the largest, but so is the outflow, leaving a net inflow much smaller than the gross.

## OA.2 Net migration flows by destination: visual evidence of hierarchical consolidation

This appendix maps bilateral net migration flows to each of the five metropolitan centers (Tokyo MA and the four regional cores) over the four postwar epochs used throughout the paper: (A) the high-growth era, 1955–1973; (B) stable growth, 1974–1990; (C) the post-bubble period, 1991–2010; and (D) the population-decline era, 2011–2024. The maps give the most direct view of the

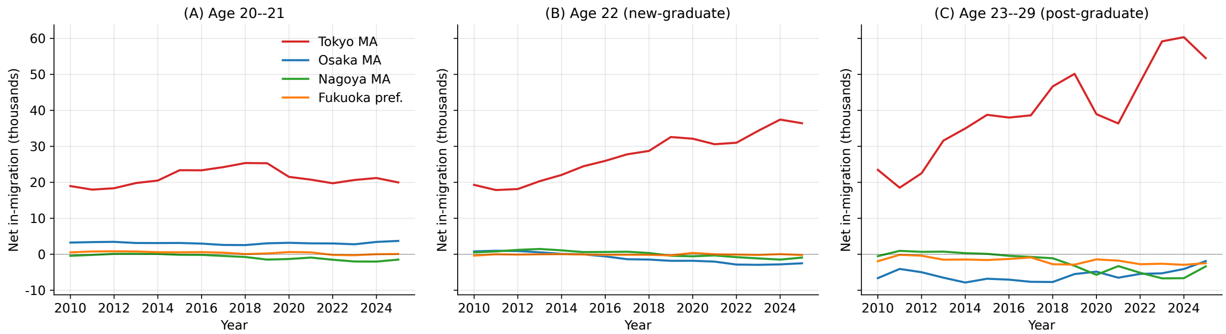



**Figure OA.4:** Net in-migration to the three major metropolitan areas and Fukuoka prefecture by single-year-age sub-band of the 20–29 cohort, 2010–2025 (thousands).

*Notes:* The three major MAs are Tokyo, Osaka, and Nagoya. Panel A: ages 20–21 (pre-graduation entrants and short-college / vocational-school graduates); Panel B: age 22 (new-graduate spike); Panel C: ages 23–29 (post-graduate movers). Among Tokyo MA's 20–29 net inflow, ages 23–29 are the fastest-growing component (+132% over 2010–2025), age 22 grew +89%, and ages 20–21 are essentially flat (+5%); the other three centers remain near or below zero across all three sub-bands. Source: Basic Resident Registry single-year-age in/out flows aggregated to each destination (Ministry of Internal Affairs and Communications, 1954–2025).

framework's upward-consolidation prediction: as large regional cores decline, their workers move on net to Tokyo MA. Each figure shows one destination in four panels, one per epoch (A)–(D), coloring origin prefectures by their annual net contribution to the destination MA in that epoch, with the same red–white–blue scale (red = origin sending net workers *to* the destination; blue = destination sending net workers back *to* the origin). The regions are the paper's standard seven centers (Figure 7 in Section 5): the three major MAs drawn as single polygons — Tokyo (Tokyo, Saitama, Chiba, Kanagawa), Osaka (Osaka, Kyoto, Hyogo, Nara), Nagoya (Aichi, Gifu, Mie) — plus Fukuoka, Miyagi (Sendai), Hiroshima, and Hokkaido (Sapporo) at the prefecture level. Unlike the single-region net rates of Appendix OA.1, the flows here are *bilateral*: net in-migration to center $M$ from origin $i$ is the gross flow $i \to M$ minus $M \to i$, with flows inside each MA excluded.

The four epoch panels document the transition quantitatively. In the high-growth era (panels A) the three major MAs jointly absorbed $\approx$ 525K migrants per year at the 1962 peak (Tokyo $\sim$ 280K, Osaka $\sim$ 165K, Nagoya $\sim$ 60K), Osaka anchoring the Hanshin industrial belt (heavy chemicals, steel, textile machinery) and Nagoya the Chukyo/Tokai belt (Toyota-centered automotive manufacturing); the western- and central-Japan hinterlands of Osaka MA and Nagoya MA appear in red in panels (A) of Figures OA.6 and OA.7 — the regional cores were then absorbing centers themselves, not stops on the way to Tokyo. By 1980 (panels B) the Osaka, Nagoya, and Fukuoka inflows had collapsed by 90–100%, and only Tokyo (oscillating 25–70K through 1980–2010, panels B–C) sustained meaningful absorption, while the regional cores' hinterlands faded to near-white (panels C–D). In the population-decline era (panels D) Tokyo dominates — it alone absorbs 50–70K/yr, roughly 5$\times$ Fukuoka prefecture (10–15K, a revived Kyushu anchor) — and it absorbs not only from small rural origins but from the regional cores themselves: Osaka MA, Nagoya MA, and Fukuoka prefecture all appear in red in panel (D) of Figure OA.5, and the same flows seen from the destination side — panels (D) of Figures OA.6–OA.9 — show Tokyo MA in deep blue. Throughout, Osaka MA is the single deepest-red origin in every panel (A)–(D) of Figure OA.5 — no origin has sent Tokyo MA more net migrants than Osaka in any postwar epoch. The regional cores have thus lost their role as destinations while continuing to send net migrants to Tokyo MA: they are now intermediate stops in a system that was once multi-polar and is now Tokyo-dominated, with Fukuoka the one surviving — much smaller — secondary anchor.

## OA.3 Long-term cohort evidence: from collective-recruitment to two-stage migration

The retrospective long-term analysis referenced in Section 4 of the main paper uses *only the JPSED 2025 wave*, the single most recent cross-section of the survey. Each 2025 respondent answers a fixed retrospective module—residence at age 15, prefecture of first job, and year of entry into first job—about her own past, regardless of when that past occurred. A 2025 respondent currently aged 70 reports first-job information from around 1973; a 2025 respondent aged 25, from around 2018. The single 2025 cross-section therefore spans first-job entry years from 1955 to 2025, which we group into four economic epochs: high-growth (1955–73), stable growth (1974–90), post-bubble (1991–2010), and population-decline (2011–25). Using a single wave is appropriate here because (i) every 2025 respondent answers the same retrospective questions, so the measurement is uniform across cohorts, (ii) the cross-sectional weight adjusts the 2025 sample to match the Japanese working-age population on demographic and occupational margins, and (iii) even the earliest epoch retains enough respondents — no epoch's cohort is thin. The principal limitation is survivor and emigration bias: for the cohorts whose first job was before 1980 (aged $\geq 65$ in 2025), the cross-section contains only the members still alive and residing in Japan in 2025, so if survival or emigration is correlated with migration history, their retrospective answers need not represent the original cohort. We therefore treat the long-term series as *descriptive context* rather than as causal identification. Two facts emerge: the destination mix of *first* jobs has shifted away from the major MAs, while migration to Tokyo MA *after* the first job has accelerated sharply.

**First-job destination: a declining role of direct rural-to-MA migration.** Table OA.2 reports the destination distribution of first jobs taken by young workers from rural origins (i.e., 15-year-old residence outside the four major MAs). The fraction of rural-origin youth whose first job is in any of the four major MAs has declined steadily, from 33% in the high-growth era to 23% in the population-decline era. The Tokyo MA share alone fell from 22% to 15%. This decline reflects the historical demise of *collective recruitment* (*shudan-shushoku*), the postwar pipeline that channeled rural high-school graduates directly to Tokyo and Osaka factories, as universities expanded and rural youth increasingly went on to college, entering the labor force at university graduation rather than straight out of high school. The channel itself has not closed—about a quarter of rural-origin youth still take their first job in a major MA—but it is no longer where the growth in Tokyo-bound migration occurs; that growth has moved to a later career stage, documented next.

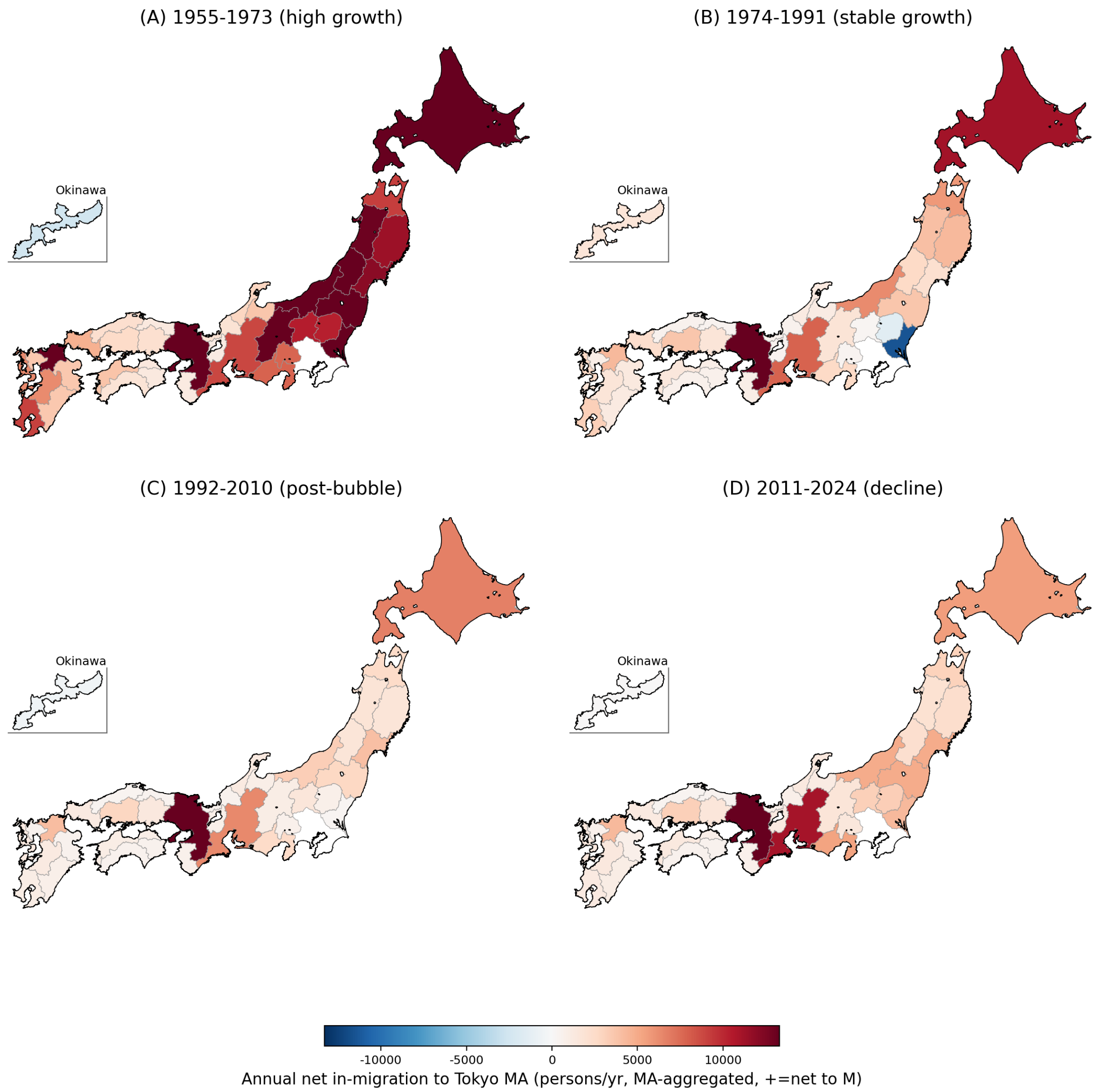



**Figure OA.5:** Annual net in-migration to *Tokyo MA* from each origin, four canonical epochs.

*Notes:* Persons per year; each non-MA origin is a prefecture, and the other MAs are shown as single MA polygons. Epochs: (A) 1955–1973, (B) 1974–1990, (C) 1991–2010, (D) 2011–2024. Red = the origin sends net workers to Tokyo MA; blue = Tokyo MA sends net workers to the origin. In panel (D), the population-decline era, Osaka MA, Nagoya MA, and Fukuoka prefecture all become net senders to Tokyo MA — the upward consolidation the framework predicts. Source: JBRR bilateral origin–destination migration, 1955–2024 (Ministry of Internal Affairs and Communications, 1954–2025).

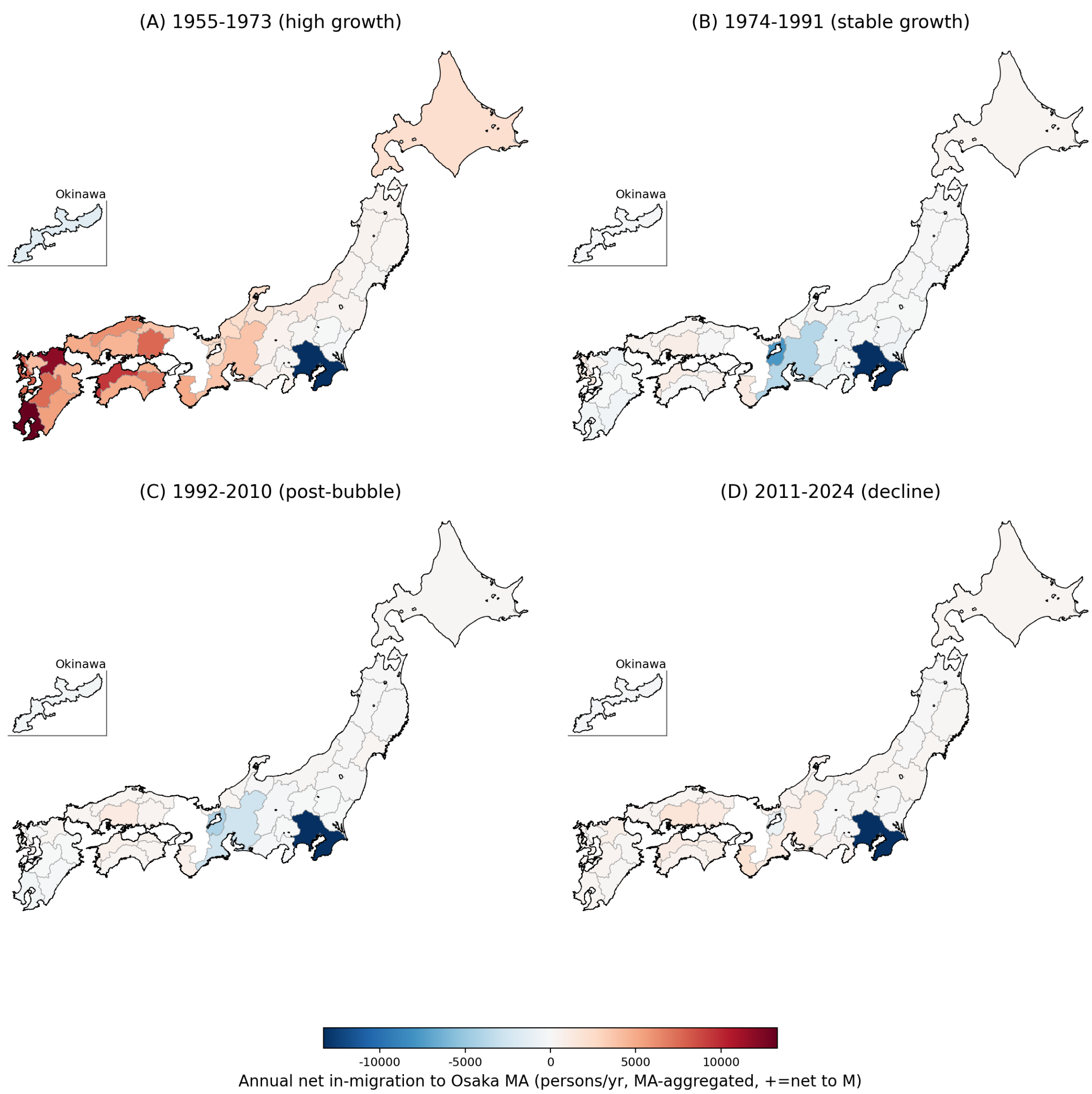



**Figure OA.6:** Annual net in-migration to *Osaka MA* from each origin, four canonical epochs.

*Notes:* Colour scale and source as in Figure OA.5. Tokyo MA appears in deep blue across panels (B)–(D), reflecting persistent net outflow from Osaka MA to Tokyo MA. The historical reversal is visible between (A) (1955–1973: Osaka MA absorbed western Japan) and (D) (2011–2024: Osaka MA absorbs essentially nothing, while losing to Tokyo MA).

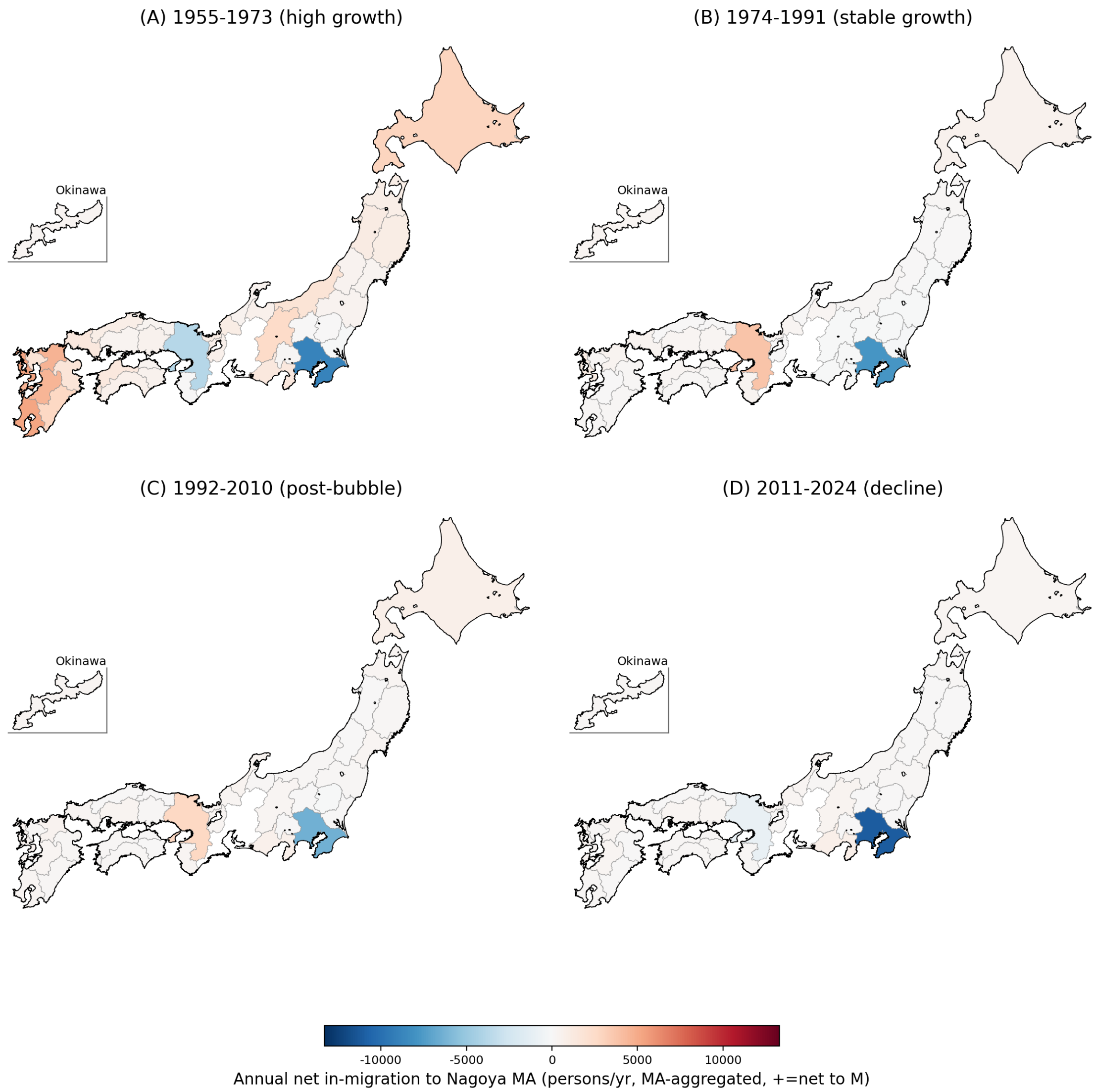



**Figure OA.7:** Annual net in-migration to *Nagoya MA* from each origin, four canonical epochs.

*Notes:* Colour scale and source as in Figure OA.5. Tokyo MA in deep blue across panels (B)–(D); the Tokai region's historical anchor role, strong in panel (A), has faded by panel (B) (its hinterland inflows collapsed by 1980).

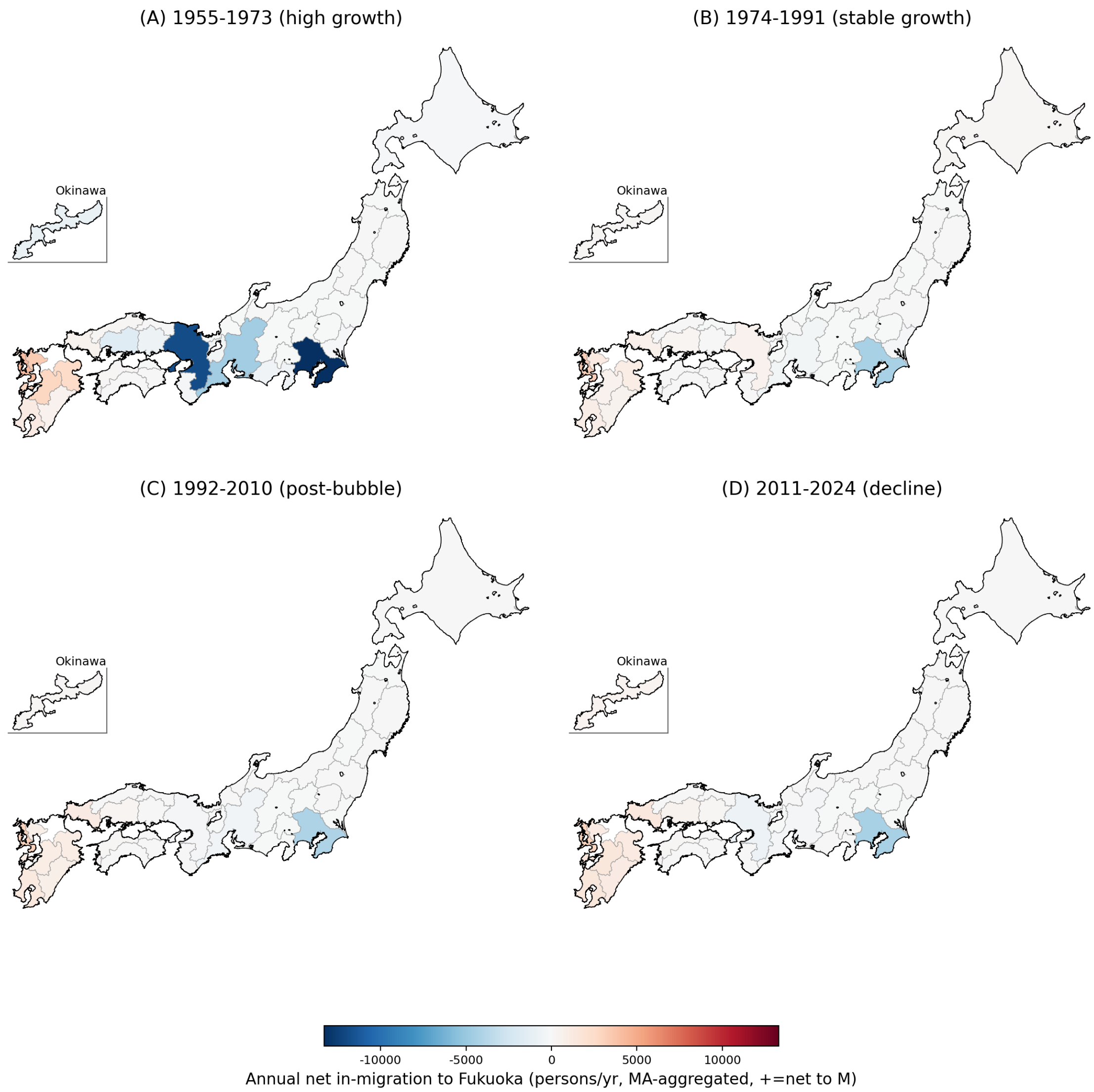



**Figure OA.8:** Annual net in-migration to *Fukuoka prefecture* from each origin, four canonical epochs.

*Notes:* Colour scale and source as in Figure OA.5. Fukuoka stands out among the regional cores as the only one that has rebuilt a regional hinterland in the population-decline era (panel D shows mild red around the rest of Kyushu), while still net-losing to Tokyo MA.

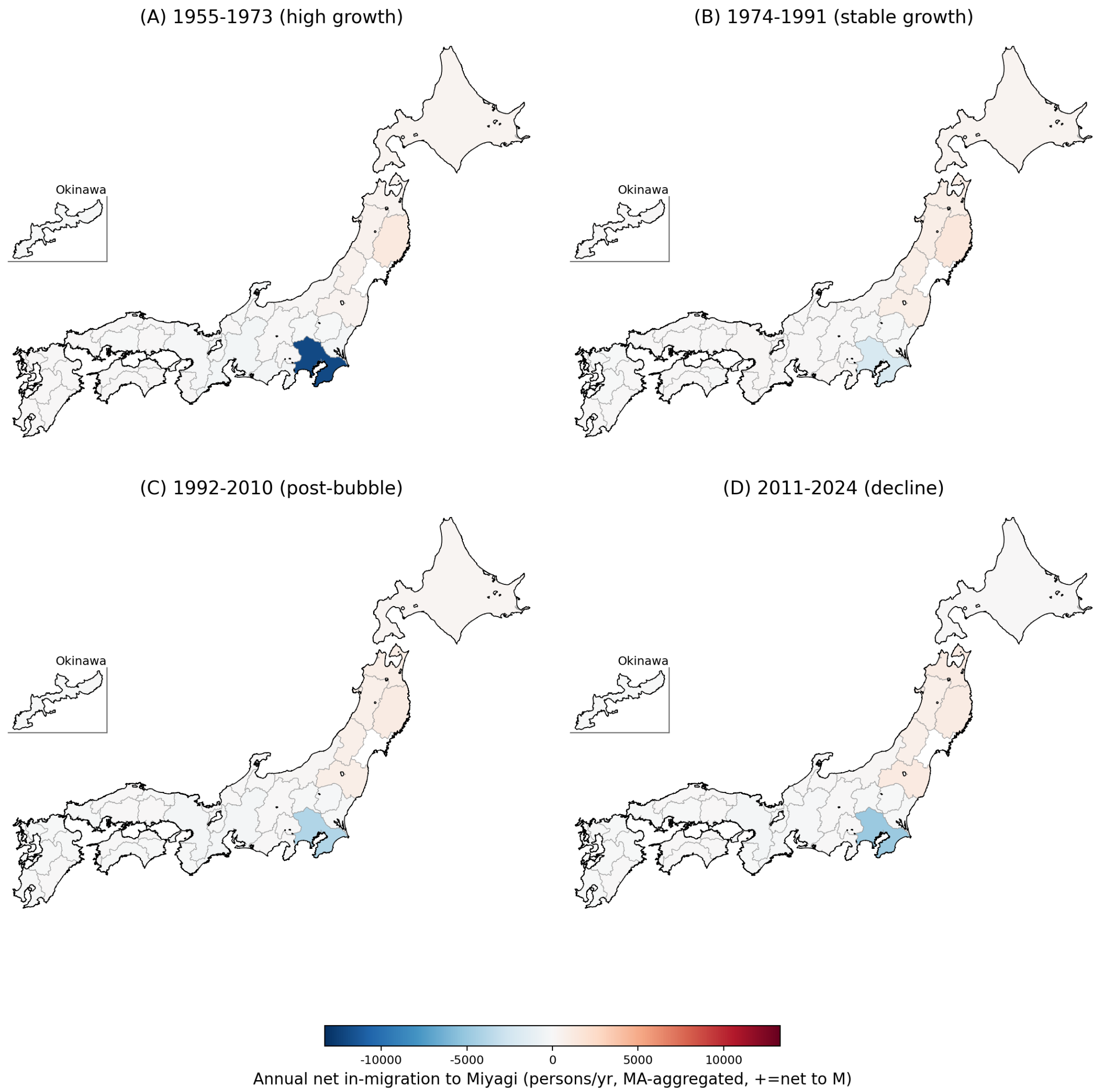



**Figure OA.9:** Annual net in-migration to *Miyagi prefecture* (Sendai) from each origin, four canonical epochs.

*Notes:* Colour scale and source as in Figure OA.5. Miyagi's role as the regional anchor for Tohoku is visible in panels (C)–(D), but is much fainter than Fukuoka's anchor role for Kyushu.

**Table OA.2:** First-job destination distribution by epoch.

| First-job epoch | N | → Tokyo MA | → Osaka MA | → Nagoya MA | → Fukuoka | → All four MAs | → Other (rural) |
|---|---|---|---|---|---|---|---|
| High-growth (1955–73) | 1,835 | 22.3% | 8.0% | 1.9% | 0.9% | **33.1%** | 66.9% |
| Stable growth (1974–90) | 6,190 | 20.2% | 5.1% | 1.4% | 1.3% | **28.0%** | 72.0% |
| Post-bubble (1991–2010) | 9,075 | 15.7% | 4.3% | 1.9% | 1.9% | **23.8%** | 76.2% |
| Pop. decline (2011–25) | 6,317 | 15.3% | 4.5% | 1.5% | 1.7% | **22.9%** | 77.1% |

*Notes:* Each cell is the survey-weighted share of the epoch's rural-origin youth whose first job was in the column's destination; the *All four MAs* column is the share whose first job was in *any* of the four metropolitan destinations (Tokyo, Osaka, or Nagoya MA, or Fukuoka), and *Other (rural)* is the rest. *Sample*: rural-origin respondents in the single JPSED 2025 wave — residence at age 15 in one of the 35 rural prefectures (outside the Tokyo, Osaka, and Nagoya MAs and Fukuoka), with a first-job year in 1955–2025 — weighted by the JPSED 2025 cross-sectional weight. *Caveat*: no survival correction is applied. For first jobs before 1980 (respondents aged $\geq 65$ in 2025), the 2025 wave observes only those still alive and resident in Japan, so these cohorts are survivors and should be read as descriptive only.

**Second-stage Tokyo MA migration: a near-sixfold acceleration.** We next look at migration to Tokyo MA *after* the first job. Table OA.3 reports, for each first-job epoch, the cumulative share of rural-origin workers whose first job was *not* in Tokyo MA but who currently reside there—computed as the survey-weighted mean of the current-Tokyo-MA-residence indicator within the epoch—normalized by the average years elapsed since first job. The per-year rate of post-first-job Tokyo migration rose from 0.14%/yr in the high-growth era to **0.83%/yr** in the population-decline era—a near-sixfold acceleration. Figure OA.10 plots the same object at a finer grain: within each epoch, respondents are binned into (basically) five-year intervals of years since first job, and each marker shows the bin's weighted share currently residing in Tokyo MA, placed at the bin midpoint.[1] The slope of each cohort's curve corresponds to its per-year rate, and the population-decline cohort (red) reaches within 10–15 years post first job a Tokyo MA share that the high-growth cohort (blue) only accumulated after 50 years.

## OA.4 Robustness to industry classification: NTT Town Pages

The main-text Figure 4 (Section 2.2) documents the concavity-on-jobs primitive at the UA grain using the harmonized 3-digit JSIC classification of the Economic Census (2009 and 2021). Figure OA.11 re-runs its static and dynamic diagnostics — the industry-count-vs-population cross-section and the 2010 → 2020 industry-exit rate vs. population — as a two-panel figure on the NTT Town Pages business-category data across 431 UAs (2020 polygons, fixed across years). NTT classifies establishments at a finer business-category grain (1,859 NTT classification codes), with no formal correspondence to the JSIC. The 2010 NTT establishment file is geocoded to WGS84 lat/lon

[1] Construction of Figure OA.10: years since first job = 2025− first-job year, so the bins correspond to the five-year first-job blocks 2021–25, 2016–20, ..., cut where they cross an epoch boundary: 2021–25 / 2016–20 / 2011–15 (population decline); 2006–10 / 2001–05 / 1996–2000 / 1991–95 (post-bubble); 1986–90 / 1981–85 / 1976–80 / 1974–75 (stable growth); 1971–73 / 1966–70 / 1961–65 / 1956–60 / 1955 (high growth). Bins cut by a boundary hold fewer than five years (1974–75; 1971–73). Bins with fewer than 30 respondents are dropped, which removes only the 1955 bin (a single respondent). Because every respondent is observed only in the 2025 wave, each curve is a synthetic-cohort profile, not a panel trajectory: points further to the right are respondents whose first job was longer ago (older people today), not the same individuals followed as they age. Same sample and survival caveat as Table OA.3.

**Table OA.3:** Second-stage Tokyo MA migration, by first-job epoch.

| First-job epoch | N | Avg. years since FJ | Cum. % now in Tokyo MA | Per-year rate |
|---|---|---|---|---|
| High-growth (1955–73) | 1,345 | 56.2 | **8.11%** | **0.144%/yr** |
| Stable growth (1974–90) | 4,946 | 43.2 | **6.03%** | **0.140%/yr** |
| Post-bubble (1991–2010) | 7,734 | 24.6 | **6.81%** | **0.277%/yr** |
| Pop. decline (2011–25) | 5,301 | 7.8 | **6.50%** | **0.831%/yr** |

*Notes: Sample*: rural-origin respondents in the single JPSED 2025 wave whose first job (FJ) was *not* in Tokyo MA. *Columns*: $N$ is the number of such respondents in the epoch; *Avg. years since FJ* is the mean of 2025 − (first-job year) within the epoch; *Cum. % now in Tokyo MA* is the survey-weighted share currently residing in Tokyo MA; and *Per-year rate* is that cumulative share divided by the average years since first job. Younger cohorts (later first-job year) are observed at fewer years post-first-job, so their cumulative share is mechanically lower; the per-year rate is the appropriate cross-cohort comparison, and it shows a striking acceleration. Same data source and survival caveat as Table OA.2.

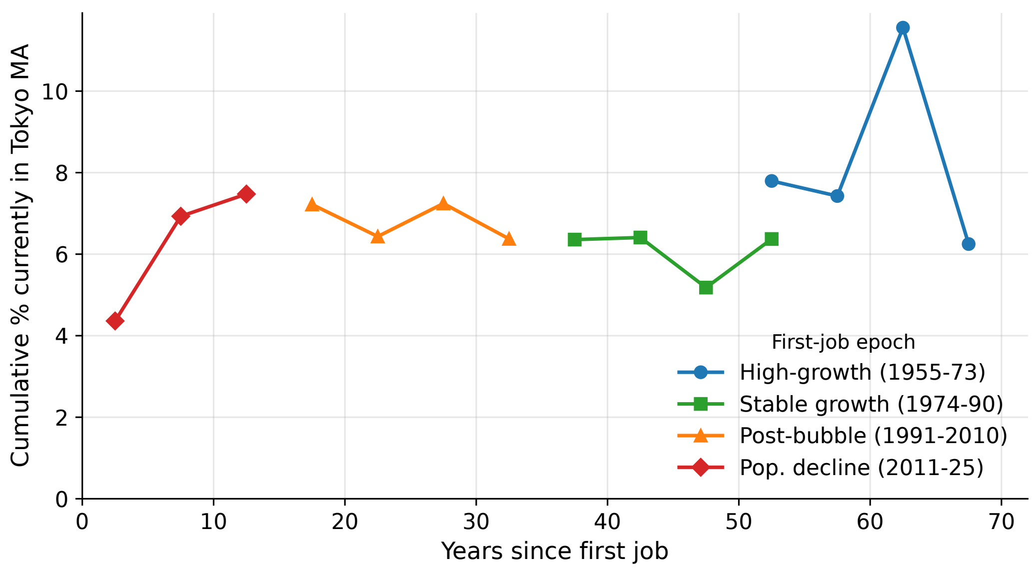



**Figure OA.10:** Cumulative share of rural-origin workers currently residing in Tokyo MA, by years since first job and first-job epoch.

*Notes:* Rural-origin workers whose first job was not in Tokyo MA; the share is plotted separately for each first-job epoch. The slope of each cohort's curve gives the per-year rate of post-first-job migration to Tokyo MA. The population-decline cohort (red) reaches 7.5% within 10–15 years post first job—a level the high-growth cohort (blue) accumulated only after 50+ years. Source: single JPSED 2025 wave (Recruit Works Institute, 2016–2025).

(LLM-assisted address resolution, 98.9% exact / fallback match across 6.82M records), aggregated to the JGD2011 3rd-mesh (1 km) grain, and then matched to the 2020 UA polygon assignment used throughout the paper (21,877 mesh3 cells in 2020 UA polygons; 73.0% of geocoded records fall within the UA polygon set; the remainder are non-UA establishments outside the urban-agglomeration definition).

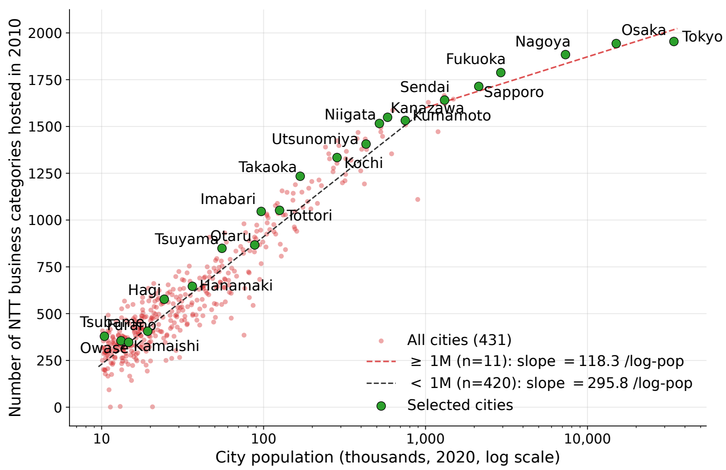
Panel (a) shows the static 2010 cross-section. The two-segment structure recurs: the slope of industry count w.r.t. log-population is 118.3 for $\geq$ 1M UAs (top 11) and 295.8 for $<$ 1M UAs (remaining 420), a small-UA-to-top-tier ratio of 2.5×—close to the JSIC ratio (2.8×). Panel (b) shows the dynamic 2010 → 2020 industry-exit rate. The exit rate falls monotonically from approximately 50–65% at the smallest UAs to under 10% at the top-tier metropolitan areas, with two-band linear-fit slopes of −3.18 percentage points per log-unit of population at the top tier ($\geq$ 1M, $n = 11$) and −3.34 at the rest ($<$ 1M, $n = 417$ after dropping 3 small UAs with thin 2010 NTT coverage, $n_{2010} < 50$). Unlike the JSIC counterpart (Figure 4(b), where the slope above 1M, −1.57, is much shallower than the −3.78 below), the NTT fits are nearly the same in the two bands. The reason is the fineness of the classification: with 1,859 categories even the largest UAs host far from all of them, so they too have marginal categories left to lose, and their exit rate falls with size at much the same rate as smaller UAs'. Under the coarse JSIC classification, by contrast, the top tier hosts nearly every category and rarely loses one outright — which is what makes its slope shallow. Both versions show the same qualitative concavity-on-jobs pattern; the finer grain makes exits more frequent everywhere, steepening the slopes overall, while narrowing the difference between the two size bands.

## OA.5 Characteristic industries by city-size class: NTT vs. JSIC

This appendix puts concrete industry names on the hierarchy property: it lists which industries a city can be expected to host once its population clears each threshold. We build the lists in both classifications used in the paper — the fine-grain NTT Town Pages business categories (Appendix OA.4) and the harmonized JSIC census categories of the main text — and then document why the two classifications cannot be mapped onto each other, which matters whenever an NTT-based industry example needs a counterpart in the Economic Census.

**Method.** Fix a descending ladder of population thresholds $T_1 > T_2 > \cdots > T_K$ (here 10M, 1M, 500k, 100k, 50k, 30k), and call an industry *ubiquitous at* $T$ if at least 95% of the UAs with population $\geq T$ host it. An industry is *characteristic of size class* $T_k$ if it is ubiquitous at $T_k$ and every larger threshold but no longer at $T_{k+1}$: reliably present once a city clears $T_k$, and not below — so $T_k$ estimates its hosting threshold at the ladder grain. Industries that fail the bar even at $T_1 = 10\text{M}$ — not hosted by both Tokyo and Osaka — remain unclassified: no size class fits them under a 95%-presence rule. For a few industries the 95% test is not monotone across thresholds, because few UAs are large: only eleven exceed 1M, so an industry missing from just one of them is

**(a)** Static 2010 snapshot




**(b)** Dynamic 2010 $\rightarrow$ 2020 industry-exit rate

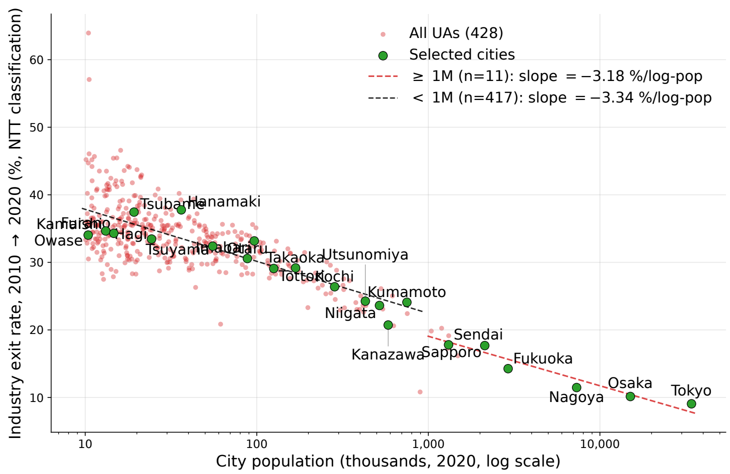



**Figure OA.11:** Concavity on jobs at the Urban Agglomeration grain, NTT Town Pages classification.

*Notes:* Both panels mirror the JSIC-based Figure 4, on the 431 UAs of the fixed 2020 definition against 2020 UA population (log scale). (a) Count of NTT Town Pages business categories each UA hosted in 2010; the count is concave in size — band slopes of 295.8 categories per log-unit of population below 1M versus 118.3 for the top 11 UAs, a 2.5× steeper gradient at small UAs. (b) Share of each UA's 2010-hosted categories no longer hosted in 2020 (dropping 3 UAs with fewer than 50 in 2010); smaller UAs lose a larger share, with two-band slopes of −3.34 and −3.18 percentage points per log-population. The comparison with the JSIC version is discussed in the text. Source: NTT Town Pages classification and the Economic Census (2009 and 2021); UA population from the 2020 Population Census (Statistics Bureau, Ministry of Internal Affairs and Communications).

at 91% and fails the bar at $T = 1\text{M}$, yet passes at 500k, where one missing city weighs less. Such an industry is assigned the smallest threshold down to which it passes the bar at every $T$: e.g., one that passes at 10M, fails at 1M, and passes again at 500k is classed at 10M — the later pass does not count, because a 500k class would require ubiquity at 1M as well (23 of 1,735 classified NTT categories, 2 of 499 JSIC). Both classifications are evaluated on the same 431 2020-definition UAs, so any difference between the two panels comes from the classification grain alone. One caveat: only Tokyo (34.2M) and Osaka (15.1M) exceed 10M, so "ubiquitous at 10M" just means "hosted by both" — read the 10M class as the categories reliably present only in Tokyo and Osaka, not as a calibrated threshold estimate.

**NTT Town Pages (Table OA.4).** The 1,859 NTT categories resolve the hierarchy finely, and the classes read as everyday experience: a 30k city keeps its internists, ramen shops, eyeglass stores, funeral services, and dry cleaners; obstetrics, lawyers, wedding halls, and the town's own newspaper office arrive at 50k; surgery, general hospitals, and ryotei at 100k; at 500k the specialty tier appears — fugu and kaiseki restaurants, import-car dealers, traditional-instrument shops, allergy and psychosomatic clinics, cinemas, and live-music venues; at 1M automobile manufacturing, TV-program production, specialized-book publishers, and the machinery and food trade lines; and the 822 Tokyo–Osaka-only categories are dominated by highly specialized manufacturing–wholesale lines. Reading down a single field traces the specialization ladder step by step: orthopedics (30k) → surgery (100k) → neurosurgery (500k) → dental–oral surgery (1M) → cardiovascular surgery (10M); newspaper delivery (30k) → newspaper offices (50k) → magazine publishers (500k) → specialized-book publishers (1M) → trade-journal publishers (10M); sushi (30k) → takeout sushi (500k) → delivery-only sushi (1M); dry cleaning (30k) → carpet cleaning (500k) → fur cleaning (10M); funeral services (30k) → wedding halls (50k) → bridal producers (500k); general trading (500k) → machinery and food trade (1M) → ten commodity-specific trade lines (10M).

**JSIC census categories (Table OA.5).** Re-running the identical ladder on the census-side hosting data (501 harmonized categories, same UAs) compresses the picture. The bulk of the service hierarchy lands in the bottom class: all clinic and specialty-restaurant categories, most retail, and the commodity wholesale lines are ubiquitous already at 30k, because the JSIC category pools what NTT separates. What survives higher up are categories whose entire content is specialized — cinemas (801), newspapers (413), and textile wholesale (511) at 500k; golf courses and racing venues at 1M; silk reeling and the large general trading companies (50A) at 10M.

**Comparability caveats (Table OA.6).** No official concordance between the NTT and JSIC classifications exists, and a usable one is hard to build. The basic problem is one-to-many: a single NTT category often bundles the manufacturing, wholesale, and retail of the same product under one heading ("weaving," "kamaboko"), so even whether a category is secondary or tertiary can be ambiguous. Beyond this, there are three distinctions that NTT draws and the JSIC does not; in each case the JSIC pools a high-threshold NTT category together with low-threshold ones, so that part of the ladder disappears on the census side. (i) *Medical specialties*: the JSIC sorts medical establishments by institution type — general clinics (832), hospitals (831), dental clinics (833), all in the 30k class — so the NTT ladder of clinical departments, from internal medicine (30k) up to cardiovascular surgery (10M), disappears entirely. (ii) *Cuisine and content genres*: fugu and kaiseki restaurants (500k in NTT) are pooled into Japanese-cuisine restaurants (76A, 30k), and specialized-book publishing and TV-program production (1M) become simply publishing (414) and video production (411), both 100k. (iii) *International trade vs. domestic wholesale*: the JSIC does not separate international traders from domestic wholesalers, so the NTT commodity-trade lines (1M–10M) land in commodity wholesale categories at 30k–500k. (The one exception: the paper splits general-merchandise wholesale by employment size, and its 100-plus-employee establishments — essentially the general trading companies — are indeed found only in Tokyo and Osaka, matching the NTT trade lines.) Table OA.6 lists, for each NTT example above, its nearest JSIC category and whether the two end up in the same size class. Usually they do not, for the reasons (i)–(iii) — and in those cases it is the NTT class that carries the economically meaningful reading; they match only when the JSIC category is just as narrow as the NTT one — the specialized activity alone, with nothing more ubiquitous pooled into it (cinemas, silk reeling, marine-food manufacturing). Newspapers are the one mismatch that runs the other way — here the JSIC class is the meaningful one: NTT puts them at 50k, the JSIC (413) at 500k, not because of pooling but because the two count different establishments — NTT lists every local bureau and sales office, the delivery network, while the JSIC assigns the industry to the publishing head office, where the publishing itself sits. The practical rule is a division of labor: the NTT table supplies the interpretable industry content — the concrete examples the paper quotes — while any analysis run on census categories takes its classes from Table OA.5, estimated directly on the census hosting data. A size class from the NTT table is never carried over to a JSIC category.

## OA.6 Extensive-margin hazards: entry composition

**Where entries land within the UA.** Figure 5 of the main text showed which industries a UA *loses*: exits concentrate at the specialized (Q1) end of its hosted-industry set. Industries also *enter*, and Figure OA.12 shows which ones a UA gains. We measure entries as a rate, not a raw count, because a UA that already hosts most industries has little room left to gain any — a raw count would mostly measure that room. The rate is built as follows. For each UA, take the industries it did *not* host in 2009: these are the industries it could gain, its *at-risk* set. Place each of them on the UA's own 2009 specialization scale, using its national hosting count: Q1 is the rarest end, Q5 the most ubiquitous, and an industry rarer than anything the UA hosted goes to Q1. For each class,

$$\text{EntryHz} \;=\; 100 \times \frac{\#\{\text{at-risk industries of the class that the UA newly hosts by 2020}\}}{\#\{\text{at-risk industries of the class}\}},$$

**Table OA.4:** Characteristic industries by city-size class: NTT Town Pages categories.

| Size class $T$ | UAs | Inds. | Medical | Restaurants | Retail | Personal services | Media, entertainment | Manufacturing, construction, trade |
|---|---|---|---|---|---|---|---|---|
| 30k | 198 | 143 | internal medicine; pediatrics; dentistry; ophthalmology; orthopedics | ramen; sushi; soba–udon | clothing stores; eyeglass stores; florists; bakeries | funeral services; dry cleaning; bonesetters–seitai; real-estate agencies | karaoke boxes; newspaper delivery | general construction; plumbing |
| 50k | 143 | 61 | obstetrics–gynecology; dermatology | okonomiyaki; Japanese cuisine; Italian | men's clothing; stationery; garden shops; delicatessens | lawyers; wedding halls; acupuncture–moxibustion | newspaper offices | interior finishing; HVAC installation |
| 100k | 83 | 144 | surgery; psychiatry; general hospitals | ryotei–kappo; Indian; curry houses; steak houses | Western-style clothing; toy stores; jewelers | house cleaning; care-transport taxis | game arcades; broadcast stations | confectionery mfg.; noodle mfg.; industrial chemicals |
| 500k | 21 | 405 | allergy; respiratory medicine; psychosomatic medicine; neurosurgery | fugu; kaiseki; Korean; eel | import-car dealers; traditional instruments (wagakki); tea-ceremony utensils | public baths (sento); bridal producers; cemeteries; after-school childcare | cinemas; live-music venues; magazine publishers | chemical plants; scientific instruments; kamaboko; frozen foods; general trading |
| 1M | 11 | 160 | dental–oral surgery; gynecology | Sichuan; Taiwanese; beef tongue | regional-specialty goods; black-tea specialty; anime and trading-card shops | interpreters; super-sento bath complexes | TV-program production; specialized-book and textbook publishers | automobile mfg.; chemical-industry machinery; machinery trade; food trade |
| 10M | 2 | 822 | cardiovascular surgery; anesthesiology; radiology; rheumatology | soft-shell turtle (suppon); champon; regional-cuisine restaurants | festival supplies; travel goods; lacquerware | fur cleaning; pet-grooming schools | film production–distribution; theaters; trade-journal publishers | weaving; raw-silk production; commodity-specific trade (textiles, metals, chemicals, ...) |

*Notes:* NTT Town Pages 2020 business categories (1,859) across the 431 2020-definition UAs. A category is characteristic of size class $T$ when at least 95% of the UAs with population $\geq T'$ host it for every threshold $T' \geq T$ on the ladder, but not at the next threshold down. *Size class* is that hosting threshold $T$ (the 10M class means present in Tokyo and Osaka only — the sole two UAs above 10M); *UAs* is the number of UAs clearing $T$; *Inds.* is the number of categories in the class (the 30k row is the necessity floor — categories ubiquitous at every threshold). A further 124 categories are *unclassified*: not reliably present even in Tokyo and Osaka (they fail the 95% bar at 10M), so no size class fits them. Within each class, the cells list only a few representative categories per field — not its full membership, which the *Inds.* column counts — with the names translated from the Japanese originals and sorted into the broad-field columns (Medical, Restaurants, Retail, and so on) for readability.

**Table OA.5:** Characteristic industries by city-size class: harmonized JSIC census categories.

| Size class $T$ | UAs | Inds. | Characteristic JSIC categories (code) |
|---|---|---|---|
| 30k | 198 | 137 | general clinics (832); hospitals (831); dental clinics (833); sushi restaurants (764); coffee shops (767); every specialty-restaurant split incl. Japanese cuisine (76A) and okonomiyaki–takoyaki (76F); automobile retail (591); eyeglass retail (608); books–stationery retail (606); food–beverage wholesale (522); industrial-machinery wholesale (541); chemical wholesale (532); bread–confectionery mfg. (097); funeral services (79A); barbershops (782); beauty salons (783); ordinary laundries (78A); karaoke boxes (80N); fitness clubs (80H); pachinko halls (80K) |
| 50k | 143 | 26 | iron–steel wholesale (534); nonferrous-metal wholesale (535); game arcades (80L); mahjong clubs (80J) |
| 100k | 83 | 55 | performance halls and troupes (802); video production–distribution (411); publishing (414); private broadcasting (382); apparel wholesale (512); personal-goods wholesale (513); paper wholesale (553); livestock-food mfg. (091); sports-facility operators (80A); musical-instrument retail (60C); wedding halls (79B) |
| 500k | 21 | 140 | cinemas (801); newspapers (413); public broadcasting (381); textile wholesale (511); motor-vehicle mfg. (311); basic-material-industry machinery mfg. (265); weaving (112); dyeing and finishing (114); marine-food mfg. (092); bowling alleys (80E); amusement parks (805); gymnasiums (80B); public bathhouses (784); crematoria and cemetery management (795); ceremonial mutual-aid associations (79C); washing and dyeing (78C) |
| 1M | 11 | 33 | golf courses (80C); racing venues (803); tennis facilities (80F); audio-information production (412) |
| 10M | 2 | 108 | silk reeling and spinning (111); general-merchandise wholesale with 100+ employees (50A, ≈ the general trading companies) |

*Notes:* Same ladder, same 431 UAs, and same 95% rule as Table OA.4, applied to the 501 harmonized JSIC categories. A few of these are finer than a standard 3-digit JSIC industry: where the paper divides a 3-digit industry using its 4-digit detail, each piece keeps the 3-digit number with a letter added (for example, code 76 splits into 76A and 76B). Two categories are unclassified, as in Table OA.4. The compression relative to Table OA.4 — 137 of 499 classified categories already ubiquitous at 30k, and the entire clinic and specialty-restaurant hierarchy in the bottom class — is a pure classification-grain effect.

**Table OA.6:** NTT vs. JSIC: approximate correspondence and where the hierarchy collapses.

| NTT category (class) | Nearest JSIC 3-digit category (class) | Verdict |
|---|---|---|
| *(a) Clinical-department axis (absent from JSIC)* | | |
| internal medicine (30k) → surgery (100k) → neurosurgery (500k) → cardiovascular surgery (10M) | general clinics 832 / hospitals 831 (both 30k) | collapsed |
| dentistry (30k) → dental–oral surgery (1M) | dental clinics 833 (30k) | collapsed |
| *(b) Cuisine / genre axis (absent from JSIC)* | | |
| sushi restaurants (30k) | sushi restaurants 764 (30k) | match |
| fugu, kaiseki, eel restaurants (500k) | Japanese-cuisine restaurants 76A (30k) | collapsed |
| ramen (30k); Korean (500k); Sichuan (1M) | Chinese restaurants 76B / other specialty 76D (30k) | collapsed |
| specialized-book publishers (1M) | publishing 414 (100k) | collapsed |
| TV-program production (1M) | video production–distribution 411 (100k) | collapsed |
| *(c) Trade vs. domestic wholesale (distinction absent from JSIC)* | | |
| machinery trade; food trade (1M) | industrial-machinery wholesale 541; food–beverage wholesale 522 (30k) | collapsed |
| textile trade (10M); chemical-product trade (10M) | textile wholesale 511 (500k); chemical wholesale 532 (30k) | collapsed |
| general trading (500k) | general-merchandise wholesale, 100+ empl. 50A (10M); other 50B (100k) | split by size |
| *(d) Retail and entertainment* | | |
| eyeglass stores (30k) | camera–watch–eyeglass retail 608 (30k) | match |
| import-car dealers (500k) | automobile retail 591 (30k) | collapsed |
| traditional instruments, wagakki (500k) | musical-instrument retail 60C (100k) | absorbed |
| cinemas (500k) | cinemas 801 (500k) | exact match |
| live-music venues (500k) | performance halls and troupes 802 (100k) | near |
| karaoke boxes (30k) | karaoke-box operators 80N (30k) | match |
| newspaper offices (50k) | newspapers 413 (500k) | inverted |
| *(e) Personal services* | | |
| funeral services (30k) | funeral services 79A (30k) | match |
| wedding halls (50k) | wedding-hall operators 79B (100k) | near |
| dry cleaning (30k) → carpet cleaning (500k) → fur cleaning (10M) | ordinary laundries 78A (30k); washing and dyeing 78C (500k) | partly collapsed |
| *(f) Manufacturing (NTT often merges manufacturing with wholesale/retail)* | | |
| automobile manufacturing (1M) | motor vehicles and parts 311 (500k) | near |
| chemical-industry machinery (1M) | basic-material-industry machinery 265 (500k) | near |
| raw-silk production (10M) | silk reeling and spinning 111 (10M) | match |
| kamaboko (500k) | marine-food manufacturing 092 (500k) | match |

*Notes:* Hand-assigned nearest JSIC 3-digit categories for the NTT examples of Table OA.4; no official concordance exists. The size class in parentheses after each category name is its measured ladder class under that classification. *Verdict*: *match* = same size class under both classifications; *near* = one class apart; *collapsed* = the NTT distinction has no JSIC counterpart, so the category falls into a broader JSIC category sitting at a lower class; *absorbed* = pooled into a broader JSIC category at an intermediate class; *inverted* = the two classifications count different establishments, which flips the ordering (the JSIC files newspapers at the publishing head office, whereas Town Pages lists each local bureau); *split by size* = the JSIC separates by employment size what NTT keeps as a single category.

aggregated within size band × class — the bars of the figure. Within every size band the entry rate rises steeply from the specialized end to the ubiquitous end (at sub-100k UAs, 6.3% in the Q1 range against ≈ 50% in the Q4–Q5 ranges). A missing bar means the class had no at-risk industries — the band's UAs already hosted every industry of that class in 2009, so there was nothing left to enter. This happens in the ≥1M band for the Q3, Q4, and Q5 ranges, and in the 100k–1M band for the Q5 range. In short: the industries a UA gains are almost all common ones — types most cities already host — and almost never the specialized industries that only much larger cities host. One question the figure cannot answer is the one that matters here: does a UA gain industries *because* it is declining? If it did, the entries would offset the decline-driven exits, and the two would have to be netted. Decline is not on the figure's axes, so the answer comes from the margin-response regression of Section 5.1: the entry hazard does not respond to decline; it responds to size alone. The list of actual entrants says the same: guest houses, nursing care, solar generation, e-commerce — the era's nationwide growth industries. They follow national trends, not the origin's population decline. This is why the main text measures the extensive margin with gross exits, without netting entries against them.

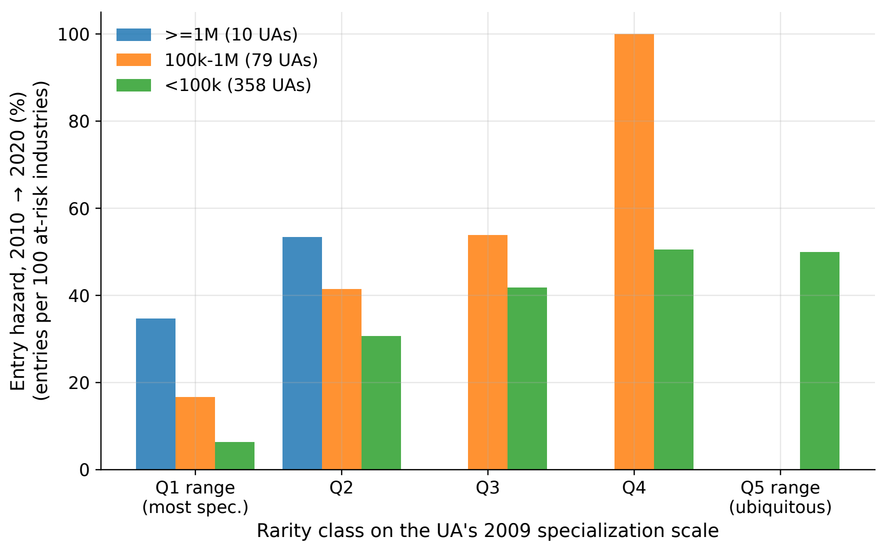



**Figure OA.12:** Entry hazard by specialization (rarity) class, 2010–2020

*Notes:* The entry-side mirror of the main-text exit figure. Industries are ranked by *rarity* — how many of the country's UAs host them — and split into five classes, Q1 (rarest, most specialized) to Q5 (most common), with the cut-offs set from the UA's own 2009 industries. Each bar is an entry rate: of the industries a UA did *not* host in 2009 — those it could still gain in a given rarity class — the share it newly hosts by 2020, shown by size band. A missing bar means the band's UAs already hosted every industry of that class in 2009, so there was nothing left to gain. We report a rate, not a raw count, because a UA has far more rare industries left to gain than common ones; a raw count would therefore be largest simply where that pool is largest (the rare end), not where entry is actually most likely. Source: Economic Census, 2009 and 2021, harmonized universe; 2010 UA polygons fixed.

## OA.7 Data construction details

This appendix collects the data construction details summarized in Section 4 and the weighting details summarized in Section 5.2.

**Economic Census hosting panel: survey universe and industry harmonization.** The industry-hosting panels behind Figures 4, 5, and 8 and Table 4 compare establishment locations between the 2009 Economic Census for Business Frame and the 2021 Economic Census for Business Activity, both conducted jointly by the Ministry of Internal Affairs and Communications and the Ministry of Economy, Trade and Industry (we label the latter's reference year 2020 throughout). Two harmonization steps make the two censuses comparable.

First, which establishments to count (the *survey universe*). The raw 2021 microdata still list some establishments that had already closed by the census date: they remain on the register but had shut down (transition-status code 3, *haigyo*), and their employment is imputed rather than observed. The official published tables drop these records using an aggregation-target flag, and we do the same. Restricting the 2020 side to the flagged establishments reproduces the officially published establishment counts exactly — for construction (division D), for example, the restricted microdata give precisely the published 485,135. If instead we kept the closed establishments, the 2020 hosting count would be overstated: an industry that has actually left a city would still appear to be present—a hidden exit—and in some cases an absent industry would appear to have newly arrived. This bias falls hardest on small UAs, where whether an industry is hosted can turn on a single establishment. The 2009 microdata contain no such closed-establishment records, so the 2009 side needs no equivalent filter.

Second, *industry codes*: the 2009 census codes some industries at the sub-class level, where the third character is a letter (e.g. 07A, 07B), while the 2021 census uses the standard numeric 3-digit minor group (078). The substantive classification change between the two JSIC revisions (rev. 12, 2007 / rev. 13, 2013) is minimal at the minor-group level; the difference is coding granularity. We collapse every letter-coded sub-class to its parent 3-digit minor group using the hierarchical structure of the two official classification lists, verified name-by-name (e.g. 07A floor work + 07B interior finishing → 078 floor and interior finishing). Table OA.7 reports the complete crosswalk: 99 letter codes in 2009 and 3 in 2020 collapse into 36 unified minor groups; all other codes coincide across the two censuses, yielding 494 unified 3-digit industries common to both years.

**Basic Survey on Wage Structure (BSS).** The BSS wage facts of Section 2.2 (Table 1) are computed from the BSS individual records — the Ministry of Health, Labour and Welfare's annual establishment-side survey of contracted wages — accessed through the government's secondary-use microdata channel. We pool the 2015–2019 surveys: this matches the JPSED sample window and keeps the pool under a single JSIC vintage (the 2013 revision). We retain records with prefecture code 1–47, age 15–70 (the JPSED worker frame), and a positive contracted wage — about 1.21 million observations per year; the 20–29 sub-sample used for the wage–hosting-count gradient

**Table OA.7:** JSIC crosswalk between the 2009 and 2021 Economic Census industry codes.

| Unified 3-digit | Descriptor | Collapsed source codes |
|---|---|---|
| 078 | Floor and interior finishing work | 07A, 07B (2009) |
| 325 | Sporting and athletic goods manufacturing | 32A, 32B (2009) |
| 329 | Miscellaneous manufacturing, n.e.c. | 32C, 32D (2009) |
| 382 | Private-sector broadcasting, except cablecasting | 38X (2020) |
| 392 | Other information processing and providing services | 39A, 39B, 39C (2009) |
| 416 | Services incidental to video, sound and text information production | 41A, 41B (2009) |
| 501 | Other general merchandise wholesale | 50A, 50B (2009) |
| 521 | Other agricultural, livestock and aquatic products wholesale | 52A, 52B, 52C, 52D, 52E (2009) |
| 559 | Miscellaneous wholesale trade, n.e.c. | 55A, 55B (2009) |
| 589 | Food and beverage retail, n.e.c. | 58A, 58B (2009) |
| 607 | Musical instrument retail | 60A, 60B, 60C (2009) |
| 609 | Other retail trade, n.e.c. | 60D, 60E, 60F, 60G (2009) |
| 622 | Banks, except central bank | 62X (2020) |
| 709 | Goods rental and leasing, n.e.c. | 70A, 70B (2009) |
| 721 | Patent attorneys' offices | 72A, 72B (2009) |
| 724 | Certified tax accountants' offices | 72C, 72D (2009) |
| 728 | Pure holding companies | 72E, 72F (2009) |
| 729 | Professional services, n.e.c. | 72G, 72H (2009) |
| 742 | Other civil engineering and architectural services | 74A, 74B, 74C (2009) |
| 759 | Accommodation, n.e.c. | 75A, 75B (2009) |
| 762 | Other specialty restaurants | 76A, 76B, 76C, 76D (2009) |
| 769 | Eating and drinking places, n.e.c. | 76E, 76F, 76G (2009) |
| 781 | Linen supply | 78A, 78B (2009) |
| 796 | Ceremonial mutual-aid associations | 79A, 79B, 79C (2009) |
| 799 | Other life-related services, n.e.c. | 79D, 79E (2009) |
| 804 | Fitness clubs | 80A, 80B, 80C, 80D, 80E, 80F, 80G, 80H (2009) |
| 806 | Other amusement halls | 80J, 80K, 80L, 80M (2009) |
| 809 | Amusement and recreation, n.e.c. | 80N, 80P (2009) |
| 821 | Other social education | 82A, 82B, 82C, 82D, 82E (2009) |
| 824 | Other instruction for arts, culture and practical skills | 82F, 82G, 82H, 82J, 82K, 82L, 82M (2009) |
| 834 | Nursing services | 83A, 83B (2009) |
| 836 | Other services incidental to medical care | 83C, 83D (2009) |
| 853 | Other child welfare services | 85A, 85B (2009) |
| 854 | Other welfare and care services for the aged | 85C, 85D, 85E, 85F, 85G, 85H, 85J (2009) |
| 859 | Social insurance, social welfare and care services, n.e.c. | 85K, 85L (2009) |
| 862 | Contracted post office services | 86X (2020) |

*Notes:* Each row is one of the 36 unified 3-digit minor groups that required collapsing; the remaining 458 codes already match across the two censuses and are not listed. The *Collapsed source codes* column gives the finer sub-classes we merged into the group — those whose third character is a letter (e.g. 07A, 07B) — with the census that coded them that way (2009 or 2020) shown in parentheses. Descriptors are informal English translations of the official Japanese titles.

pools to $N = 932{,}992$. The wage concept is the contracted monthly wage (excluding overtime and bonuses), deflated with the same prefecture-by-year housing-inclusive deflator that we apply to the JPSED wage outcome (Appendix OA.20). Each worker's industry is the establishment's JSIC 3-digit code; the industry's hosting count is the number of the 448 fixed 2010 UAs hosting it in the 2021 Economic Census for Business Activity (reference year 2020), from the harmonized hosting panel above.

**JPSED variables and source codes.** The JPSED is administered by the Recruit Works Institute and distributed via the Social Science Japan Data Archive (SSJDA). The ten waves we use are SSJDA datasets 1088 (2016), 1164 (2017), 1227 (2018), 1279 (2019), 1349 (2020), 1429 (2021), 1523 (2022), 1598 (2023), 1730 (2024), 1775 (2025). For each consecutive-wave pair we extract seven content variables: sex, age, prefecture of current residence, educational attainment, employment type (regular / part-time), industry and occupation of current main employment, and annual income from main employment (in 10,000-yen units, *man*-yen, ≈ \$70), together with the longitudinal-pair weight described in the "Sample weights" paragraph below.

The JPSED question numbers attached to these variables *differ across waves*: the questionnaire was reorganized three times — between the 2016 and 2017, the 2017 and 2018, and the 2018 and 2019 waves — so the first (2016) wave's codes match none of the later waves, and the numbering settles only from the 2019 wave onward. Three variables are easy to get wrong in the 2016 wave: `Q5` codes *marital status* (`Q12` codes educational attainment); the employment-type question is `Q18` (renumbered `Q19` from 2017 onward); and the income question is `Q85_1` rather than the later waves' variants (`Q91_1`/`Q98_1` in 2017–2018, `Q94_1`/`Q99_1`/`Q100_1` from 2019). Table OA.8 reports the full per-wave map; only Q1 (sex), Q2 (age), and Q4 (prefecture of current residence) carry the same code across all ten waves.

**Table OA.8:** JPSED question-number map for the seven content variables we extract, per wave.

| Variable | 2016 | 2017 | 2018 | 2019–2025 |
|---|---|---|---|---|
| sex | Q1 | Q1 | Q1 | Q1 |
| age | Q2 | Q2 | Q2 | Q2 |
| prefecture of current residence | Q4 | Q4 | Q4 | Q4 |
| educational attainment | Q12 | Q5 | Q5 | Q5 |
| employment type (regular / part-time) | Q18 | Q19 | Q19 | Q19 |
| industry of current main employment | Q28 | Q29 | Q33 | Q30 |
| occupation | Q30 | Q31 | Q35 | Q32 |
| annual income from main employment | Q85_1 | Q91_1 | Q98_1 | varies[a] |

*Notes:* Only Q1 (sex), Q2 (age), and Q4 (prefecture of current residence) are stable across all ten waves; the other variables shift by one to several question numbers at each of the 2016/17, 2017/18, and 2018/19 questionnaire reorganizations. [a] Income question codes from 2019 onwards: Q100_1 (2019), Q94_1 (2020), Q100_1 (2021–2022), Q99_1 (2023–2025).

**Sample restrictions in detail.** We construct the wage-growth analytical sample as follows.

1. Restrict to consecutive-wave pairs of the same individual in successive years.

2. Restrict to consecutive pairs with valid (positive) $w_t$ and $w_{t-1}$, i.e., main-employment income observed both before and after.

3. Trim the top and bottom 1% of $\Delta \log w \equiv \log w_t - \log w_{t-1}$ to reduce the influence of outliers.

4. Define *move to Tokyo MA* as having current prefecture in {Saitama, Chiba, Tokyo, Kanagawa} (Statistics Bureau prefecture IDs 11, 12, 13, 14) but previous prefecture outside this set.

5. Define *rural origin* as origin prefecture not in the four major MAs (Tokyo MA, Osaka MA, Nagoya MA, Fukuoka prefecture). Robustness checks (Appendix OA.15) consider alternative metro definitions in which Miyagi (Sendai) and/or Hiroshima are also classified as MAs.

6. Restrict to COVID-clean transitions. Wave $t$ reports on calendar (reference) year $t-1$, so the transition into wave $t$ covers reference years $t-2$ and $t-1$; it is COVID-clean when both lie outside 2020–2022. This keeps waves $t \in \{2017, 2018, 2019, 2020\}$ — reference-year transitions 2015/16 through 2018/19, all pre-COVID (the 2020 *wave* covers 2018/19) — and $t = 2025$ (reference years 2023/24). The baseline of the paper further restricts to the four pre-COVID transitions (Section 4).

7. Drop pairs whose $t-1$ status is *both* in school (JPSED education code in $\{9, 10, 11, 12, 13, 14, 15\}$) *and* in part-time employment (employment-type code equal to 2; `Q18` in the 2016 wave, `Q19` from 2017), which for an in-school respondent is essentially a student side job. Including these pairs would mechanically embed the student-to-full-time labor-market transition into the wage-growth measurement (the $t-1$ wage is a side-job hourly wage, the $t$ wage is the post-graduation full-time entry wage), which is not the framework's Tokyo MA wage premium. The exclusion applies to the full sample.

The resulting wage-growth sample over all COVID-clean transitions contains 59,819 rural-origin person-wave pairs, of which 439 are migrants to the Tokyo MA; the $\leq 29$ sub-sample contains 5,657 pairs and 128 migrants. The pre-COVID baseline of the paper (Section 4) contains 47,536 pairs (358 migrants; 4,617 pairs and 95 migrants in the $\leq 29$ sub-sample).

**Sample weights.** JPSED provides two kinds of weights. *Cross-sectional* weights, built by the Recruit Works Institute, make each single year's respondents match the working-age population on demographic and occupational margins. *Longitudinal* (attrition) weights do the same for respondents observed in two consecutive waves; the recommended adjacent-pair version covers ages 20–64 and is released in the special-survey datasets 1165, 1228, 1280, 1350, 1524, 1599, 1731, and 1776. Our target is a population-level statement—how origin decline rates and sizes shape the Tokyo MA wage premium for the panel of consecutive-wave movers—so our primary estimates apply the longitudinal pair weight at each transition (Specs 1–5 in Table 5). Not every pair receives a panel weight: the weights are defined only for ages 20–64, so pairs outside that frame — notably workers aged 15–19 — have none. Those pairs get the wave-$t$ cross-sectional weight instead.

We report both weighted and unweighted estimates. Solon et al. (2015) note that no single weighting is "correct": under heterogeneous effects, WLS and OLS recover different weighted averages of the treatment effect, so the choice should follow the parameter of interest, and comparing the two is itself a diagnostic for misspecification or sample–population heterogeneity. Our descriptive numbers in Section 2 are weighted, because they are meant to describe migration in the working-age population as a whole rather than in the raw JPSED sample; for consistency our primary regressions are weighted too, and we report the corresponding unweighted OLS estimates (Specs 4′–5′) in Online Appendix OA.19 (Table OA.33) as a check. The framework cells come out qualitatively unchanged without weights; where individual coefficients do shift, we read the gap as telling us *which* subpopulations carry the heterogeneity — the Solon et al. (2015) diagnostic — not as evidence against the qualitative pattern.

**Origin decline series (young population).** This paragraph concerns only the *decline* coordinate $g_o$; the *size* coordinate of the analysis is the origin's *total* (all-age) 2010 population, defined in Section 4. The young (15–29) population of prefecture $o$ in year $t$, $S^{\mathrm{y}}_{o,t}$, sums the three 5-year age bands 15–19, 20–24, and 25–29; the 2010 value comes from the Population Census (Statistics Bureau, Ministry of Internal Affairs and Communications), and the 2019 and 2025 values from the Basic Resident Registry population tables. The decline rate $g_{o,t_1,t_2} \equiv (S^{\mathrm{y}}_{o,t_2}/S^{\mathrm{y}}_{o,t_1} - 1) \cdot 100$ is the signed percentage change of this young population over the window $[t_1, t_2]$. The main specification uses the long-run window $[2010, 2019]$: the cross-origin variation in cumulative decline that the framework needs builds up over the long pre-COVID horizon. Robustness varies the window (Section 7, rows R6a–R6b of Table OA.28): R6a extends it to $[2010, 2025]$ holding the wage panel fixed; R6b additionally extends the panel with the COVID-clean 2023/24 transition.

## OA.8 The margin responses of Section 5.1 at the UA grain

Section 5.1 measures the extensive margin at the *origin* scale — the prefecture: an industry exits when no UA of the origin prefecture hosts it any more. The hosting thresholds themselves, however, live at the UA grain, so this appendix examines what the extensive margin looks like at the UA level. Two findings. First, the extensive margin. The naive version — each UA's count of dropped industries, regressed on the UA's own young decline — returns a wrong-signed coefficient. This is not because exits are rare at a single UA. It is because the market an industry serves is larger than the UA: hierarchy position decides which UA of a market drops a shared industry, and the market area's demand decides whether the industry disappears altogether; the UA's own decline decides neither. The design that works builds the market area in. We keep only the holdings for which the UA is the only host within $R$ km. For these holdings the market area keeps the industry only as long as that one UA keeps it, so whether the UA keeps the industry and whether the market area keeps it are one and the same. And because the sole host is where the market area's demand is centered, the UA's own young decline now stands in for the market area's. In this sample, exit responds to it with the predicted sign (OA.8.1, Table OA.10). Note what this design used: distance circles only — no pooling of UAs into a prefecture union, and no administrative boundary anywhere. Second, the intensive margin mirrors the picture: survivor contraction responds to all-age rather than young decline (OA.8.2).

### OA.8.1 The extensive margin at the UA grain: the failing count, and the sole-host exit that replaces it

The exit event of Section 5.1 is a market-level event: an industry exits when no UA of the origin hosts it any more. Seen from the UA side, the event lands on one particular UA — the only one still hosting the industry; when that UA drops it, the origin loses the industry. Counted at the origin level, these events deliver the margin response of the main text ($-0.654$, $t = -2.9$). What happens if the same count is instead run at the UA grain — each UA's own dropped industries, on its own young decline and size? Table OA.9 shows the result: the coefficient on decline comes out *wrong-signed* (+0.10).

The count fails because an industry's customers do not stop at the UA boundary. An industry close to its hosting threshold needs more customers than its host UA alone can provide: it survives on the demand of a market area wider than the UA. Its exit therefore depends on the population of that wider area, not on the population of the host UA — so the host UA's own decline is the wrong regressor from the start. The data show the mismatch directly: a UA drops on average 32 industries over the decade, but about nine in ten of them remain hosted elsewhere in the origin prefecture. The industry did not disappear; it simply stopped being hosted in this UA and remained hosted in a larger one nearby — the upward retreat of Figure 5 (industries consolidating into bigger cities), seen from the UA that lost it. Which UA of the market loses a shared industry is decided by hierarchy position: the smaller host drops it. The remaining one in ten of the drops are the exits of Section 5.1: the industry disappears from the whole origin, and whether that happens is decided by the demand of the whole market area. Neither of these two outcomes — which UA drops the industry, or whether it exits at all — turns on the host UA's own young decline. Yet the UA's own young decline is exactly what the naive count regresses on, which is why its sign comes out wrong. The remedy is to build the market area into the measurement.

We draw a *buffer* around each UA — a circle of radius $R$ (10–50 km) around its population-weighted centroid, metropolitan UAs included — and treat it as the market area. The sample is the 2009 holdings for which the focal UA is the *only* host within the buffer. This restriction buys two things. First, the industry has no neighboring host to fall back on: if the UA drops it, the market area loses it, so a drop and an exit are the same event. Second, the customers who keep the industry alive are the people in and around the focal UA, so the UA's own young decline measures the demand loss the industry faces. The outcome is the union definition of the main text applied to the buffer: the industry counts as exited when no UA within it hosts the industry in 2020. Table OA.10 reports the regression of this indicator on the UA's own young decline and log population, with unified 3-digit industry fixed effects — included because each industry has its own hosting threshold, so which industries sit at the exit margin differs by UA size. A percentage point of young decline raises the exit probability by 0.14 points at $R = 20$ km, 0.21 at 30 km, and 0.27 at 50 km, significant throughout ($t$ between $-2.5$ and $-4.6$); at 10 km the estimate attenuates ($-0.04$, $t = -1.2$): true market areas are wider than a 10 km circle, so another host of the same industry often sits just beyond the circle — a holding that looks like a sole host within 10 km is then not the sole host of its actual market, and the design's logic no longer applies to it. Because nearby UAs' buffers overlap, each estimate carries two $t$ statistics — clustered by prefecture, and a Conley spatial HAC (Bartlett kernel, 200 km cutoff) that lets correlation decay with distance regardless of boundaries; both support the inference. The response is not a prefectural-capital phenomenon — sole-host holdings concentrate there, so one might worry the capitals drive the estimate — but dropping the capital UAs from the sample leaves it essentially unchanged (at 30 km: $-0.21$, $t = -2.5$). In sum, the buffer brings the market area into the design, and with it the decline response is confirmed at the UA level as well: population decline drives industry exits.

### OA.8.2 The intensive-margin mirror

Survivor contraction, being a continuous margin, carries variation at every aggregation level — what it lacks is any response to the *young* measure. Table OA.11 shows it responds instead to *all-age* decline, at both grains: the prefecture-union contraction of Section 5.1 loads on all-age decline ($-0.85$, $t = -2.5$) while remaining null on the young measure ($-0.12$, $t = -0.8$), and the per-UA survivor contraction responds to the host UA's own all-age decline ($-0.40$, $t = -2.1$; strongest among the 100k–1M UAs, $-0.71$, $t = -4.2$) while null on the UA's young measure ($t = -0.3$). This contrast says which demographic measure each margin needs; it does not say the young do not matter for demand. Young decline is of course one component of all-age decline (the two correlate at $+0.44$ across the rural prefectures), and the young who remain are local customers like everyone else. But the demand facing a surviving establishment is formed by the *entire* resident population, of which the young are a modest and shrinking slice, so the young component alone captures only a fraction of the demand change — the intensive margin requires the all-age measure, to which survivor employment adjusts roughly proportionally. The extensive margin requires the opposite choice, for the reason given in Section 4: to capture the demographic contraction that pushes origins across hosting thresholds one must look at the *young* population's variation, because the all-age total — buffered by a growing elderly population — moves late and understates the decline under way, while the young cohort is its leading, migration-loaded edge. The two margins therefore call for different demographics at different scales: threshold exits respond to the system-wide *young* decline, while survivor contraction responds to the host UA's *all-age* demand. And it is the exit side that moves workers to Tokyo: when the exit count and survivor contraction are entered together in the wage regression, the exit count keeps its coefficient while the contraction term comes out null (Appendix OA.9). Since the mechanism thus runs through the exits, the wage analysis measures origin decline by the young population.

**Table OA.9:** The exit-count regression run at the UA grain instead of the prefecture.

| Grain | Exit count | mean | $\hat{\beta}_g$ [$t$] |
|---|---|---|---|
| UA (no aggregation) | industries the UA drops | 32.4 | $+0.102^{*}$ $(+1.9)$ |
| Prefecture (Table 3) | union exits | 24.6 | $-0.654^{***}$ $(-2.9)$ |

*Notes:* The Section 5.1 exit-count regression Exits $\sim g + \log S$, run at two spatial units. "UA (no aggregation)": for each of the 371 rural-prefecture UAs, the count of industries hosted in 2009 but not 2020, on the UA's own young (15–29) decline 2010–2020 and log 2010 UA population, clustered by prefecture. "Prefecture": the origin-level exit count of Table 3, where an industry counts as lost only when no UA of the prefecture still hosts it. The per-UA count comes out wrong-signed because an industry's market area is wider than a single UA (appendix text). $t$ in parentheses; $^{***}/^{**}/^{*}$: 1/5/10%.

**Table OA.10:** Sole-host exits at the UA grain.

| | All sole-host holdings | | | | | Excluding prefectural seats | | | |
|---|---|---|---|---|---|---|---|---|---|
| $R$ | $N$ | mean dep. var. | $\hat{\beta}_g$ | $t_{\text{pref}}$ | $t_{\text{Conley}}$ | $N$ | $\hat{\beta}_g$ | $t_{\text{pref}}$ | $t_{\text{Conley}}$ |
| 10 km | 61,753 | 15.3 | $-0.041$ | $-1.2$ | $-1.2$ | 50,377 | $-0.033$ | $-0.9$ | $-0.9$ |
| 20 km | 29,329 | 17.6 | $-0.141^{***}$ | $-2.9$ | $-2.9$ | 23,502 | $-0.140^{***}$ | $-2.7$ | $-2.9$ |
| 30 km | 16,571 | 19.9 | $-0.211^{**}$ | $-2.5$ | $-2.8$ | 12,666 | $-0.213^{**}$ | $-2.5$ | $-2.6$ |
| 50 km | 5,496 | 23.1 | $-0.274^{***}$ | $-4.6$ | $-4.1$ | 3,864 | $-0.235^{***}$ | $-3.9$ | $-3.9$ |

*Notes:* The sample is the 2009 UA×industry holdings (fixed 2010 polygons) for which the focal UA is the *only* host within $R$ km of the UA's population center (the average location of its residents); there, an industry's exit from the $R$-km market area responds to the UA's own young decline. Dependent variable: $100 \times \mathbf{1}$[no UA within $R$ km hosts the industry in 2020] — the industry has left the whole circle — so its average (the *mean dep. var.* column) is the exit rate in that $R$-km sample. Regressors: UA young (15–29) growth 2010–2020 and log 2010 UA population, with 3-digit industry fixed effects (on the harmonized codes of Table OA.7). Each estimate carries two $t$'s — clustered by the focal UA's prefecture, and a Conley spatial HAC (Bartlett kernel, 200 km) — with stars using the smaller $|t|$. The right panel drops each prefecture's seat-of-government UA. $^{***}$, $^{**}$, $^{*}$: 1%, 5%, 10%.

**Table OA.11:** The intensive-margin mirror.

| Grain | Contraction measure | Young $g$ | All-age $g$ |
|---|---|---|---|
| Prefecture union | −shrink index (Table 3) | $-0.121$ $(-0.8)$ | $-0.851^{**}$ $(-2.5)$ |
| UA grain | per-UA survivor contraction | $-0.072$ $(-0.3)$ | $-0.396^{**}$ $(-2.1)$ |

*Notes:* Survivor contraction responds to all-age decline, not young decline, and the point holds at both spatial grains. Contraction regressed on young (15–29) versus all-age decline, with log origin size. "Prefecture union": the Section 5.1 contraction measure over the 34 rural prefectures (young and all-age 2010–2019). "UA grain": for each of the 371 rural-prefecture UAs, the mean log employment change across its surviving (continuing) industries, weighted by each industry's employment, on the UA's own young and all-age decline 2010–2020, clustered by prefecture. $t$ in parentheses; $^{***}/^{**}/^{*}$: 1/5/10%.

## OA.9 The origin-level exit count: construction and what it identifies

In the framework, origin size and decline $(S_o, g_o)$ are the *drivers* and industry exits are the induced *response*: the same decline crosses few, widely-spaced thresholds at a large origin and many, densely-packed ones at a small origin (Section 2.2). The exit count is therefore a mediator, not a treatment. Still, one can put the count itself into the wage regression and ask what it identifies. This appendix does that: it states the specification and why we do not build cells on the count, gives the construction of the measure, and reports the results.

**Specification.** The baseline works with the population drivers $(S_o, g_o)$ — the framework's sufficient statistics for the hierarchy — while the exit count is the *direct* measure of the extensive-margin change itself. The exercise is the analog of the linear Spec 3 of Table 5: we replace the continuous decline interaction Move $\times\, \tilde{g}_o$ with Move $\times$ Exits$_o$, keeping Move $\times \log \tilde{S}_o$ as the size control — the question is whether, *at the same origin size*, a worker from an origin that lost more industries earns a different Tokyo premium (the regression carries origin fixed effects, which absorb anything constant within an origin, so an origin-level variable is identified only through its interaction with Move). Everything else — sample, controls, weights, origin and wave fixed effects, inference — is the main estimation's. We do not build cells on the count, whose cross-prefecture variation is too thin to split (half the rural origins sit between 20 and 30 industries lost); the only refinement is letting the count's slope differ on the two size sides. The count enters raw, so coefficients read per additional industry lost.

**Measure.** Exits$_o$ is exactly the measure of Eq. (1) of Section 5.1; here we give the construction details deferred there. Each of the 448 fixed 2010 UAs is assigned to the prefecture holding the largest share of its 1km cells (21 UAs straddle a border; a population-weighted assignment would flip only 2). Exits$_o$ then counts the unified 3-digit industries that at least one of $o$'s UAs hosted in 2009 and none hosts in 2020 (range 10–38, median 25 over the rural origins), from the UA-level hosting panel of Appendix OA.7, which aligns the 2009 and 2021 censuses to a common industry classification and to the same coverage of operating establishments. Okinawa hosts no UA in the 2010 delineation, so it drops; the 2SLS sample is then identical to the baseline IV sample, which already excludes Okinawa through the TFR instrument ($N = 46{,}500$; $\leq 29$: $N = 4{,}528$, 94 movers).

**Results.** The exit-count slope should not be read causally: the negative coefficient does not mean that losing industries lowers the premium. Rather, origins that lost more industries tended to lose cheaper, more-ubiquitous ones (the exit count correlates +0.66 with the average national ubiquity of the industries lost), and ubiquitous industries carry small wage steps, so the negative sign reflects the *composition* of what is lost, not the number.

With that reading, in WLS the exit count enters at −0.011 per industry lost in the full rural sample (Table OA.12): at the same origin size, a worker from an origin that lost one more industry

**Table OA.12:** The premium on the population-independent exit count.

| | Full rural sample | | $\leq 29$ sub-sample | |
|---|---|---|---|---|
| | WLS | 2SLS (df=2/4) | WLS | 2SLS (df=2/4) |
| *(1) Linear* | | | | |
| Move $\times$ Exits (industries lost) | $-0.0107^{*}$ | $+0.0212$ | $-0.0112$ | $+0.0683$ |
| | (0.0060) | (0.0220) | (0.0129) | (0.0576) |
| *(2) Horse race against survivor shrinkage (WLS)* | | | | |
| Move $\times$ Exits | $-0.0118^{*}$ | — | $-0.0110$ | — |
| | (0.0066) | | (0.0130) | |
| Move $\times$ Contr | $+0.0090$ | — | $-0.0019$ | — |
| | (0.0074) | | (0.0198) | |
| *(3) Size-side split, exit count continuous* | | | | |
| Move $\times$ $\mathbf{1}[S^{-}]$ $\times$ Exits (small origins) | $-\mathbf{0.0278}^{**}$ | $-0.0283$ | $-0.0166$ | $+0.0180$ |
| | (0.0116) | (0.0177) | (0.0216) | (0.0454) |
| Move $\times$ $\mathbf{1}[S^{+}]$ $\times$ Exits (large origins) | $-0.0037$ | $+0.0075$ | $-0.0125$ | $+0.0042$ |
| | (0.0058) | (0.0189) | (0.0140) | (0.0366) |
| Hansen $J$ ($p$): linear / S-side | — | 0.988 / 0.427 | — | 0.526 / 0.282 |
| $N$ | 46,500 | 46,500 | 4,528 | 4,528 |
| Movers | 348 | 348 | 94 | 94 |

*Notes:* The three panels enter the exit count (1) on its own, (2) alongside survivor contraction — to see which margin carries the premium — and (3) with its slope allowed to differ between large and small origins. Dependent variable, sample, weights, origin fixed effects, and inference (score-based wild cluster bootstrap at the origin prefecture, $B = 999$) as in the main estimation; all include the continuous Move $\times \log \tilde{S}$ size gradient, and the split adds Move $\times\, \mathbf{1}[S^{+}]$. Exits is the raw industry-exit count; Contr is the survivor contraction. The 2SLS columns instrument the endogenous decline term with the paper's three demographic shocks — a demographic Bartik, the 1965–85 fall in the total fertility rate, and 1980–85 capital deepening — entered as Move interactions (the split spec adds their interactions with origin size as well). $^{***}$, $^{**}$, $^{*}$: 1%, 5%, 10%.

— on average a cheaper, more ubiquitous one — earns a Tokyo premium about one percentage point lower. The estimate is unchanged when survivor contraction is entered alongside it ($-0.012^{*}$, with the contraction term carrying nothing — the loading belongs to the extensive margin, not the intensive one), and, when the slope is split by size side, concentrated at small origins ($-0.028^{**}$; large origins null), which is where concavity on jobs bunches the hosting thresholds so that decline crosses many of them at once. The $\leq 29$ estimates are sign-consistent but imprecise, reflecting the count's compressed variation. The 2SLS columns, by contrast, we read as a caveat rather than as evidence: the three instruments are demographic shocks constructed to shift the decline rate $g_o$. The count does respond to decline, but *conditional on size* (Section 5.1); most of its variation comes from origin size, so the raw correlation between the exit count and $g_o$ is only $+0.19$. The instruments were built to shift $g_o$; since the count barely moves with $g_o$, the instruments barely move the count either, and when Move $\times$ Exits replaces Move $\times$ $\tilde{g}$ as the endogenous regressor the first stage no longer matches — the point estimates flip sign with large standard errors ($+0.02$ full rural, $+0.07$ at $\leq 29$, both n.s.). The count regression is thus identified descriptively (WLS); our instruments were not built to identify it causally.

Taken together, the count validates the mechanism but is not a usable replacement for the population drivers in the wage design: its variation is too compressed to support cells, its continuous slopes are imprecise ($\leq$ 29 sign-consistent but insignificant), it cannot be instrumented (our demographic shocks shift $g_o$, not the count), and its negative slope is compositional rather than causal, as set out above. This is why the baseline works with the population drivers and treats the count as the margin-response outcome it is.

## OA.10 The UAs behind the prefecture cells: origin composition

Table OA.13 shows, for each rural prefecture origin, which UAs stand behind its cell classification (Section 5.2). Each row gives the origin's cell and the two cell coordinates (total 2010 population and young 2010–2019 change); the number of UAs the prefecture contains; its *core* (the largest UA) with the core's population and young 2015–2020 change; and the same two numbers for the median periphery UA.

Three patterns stand out. First, a large-origin prefecture (cells I/II) pools one large core with many small UAs — Hiroshima (cell I) pairs the Hiroshima core (1,268k) with 11 further UAs, and Hokkaidō (cell II) pairs Sapporo (2,072k) with 35 further UAs — the mixture that prevents a clean I/II split at this grain. Second, among the small origins the cores of cells III and IV are of similar size; what separates the two cells is the core's young change: the cell III cores retain their young population (median $-2.2\%$, with Toyama $+1.0$, Matsue $+0.9$, and Fukui $+0.7$ gaining) while the cell IV cores lose it (median $-8.4\%$; Nagasaki $-15.3$, Tokushima $-14.3$, Kagoshima $-13.1$, Aomori $-12.9$). Third, the periphery columns look alike across cells — small UAs losing young population at broadly similar rates everywhere — so the origin-level distinction is carried by the core, which is exactly what the population-weighted prefecture aggregate $\tilde{g}_o$ registers.

**Table OA.13:** Origin composition: each rural prefecture's UAs — cell, prefecture coordinates, core UA, and periphery.

| Origin | Cell | Prefecture | | #UAs | Core UA | | | Periphery (median) | |
|---|---|---|---|---|---|---|---|---|---|
| | | Pop (k) | Δyoung (%) | | Name | Pop (k) | Δyoung (%) | Pop (k) | Δyoung (%) |
| Shizuoka | I | 3,743 | -7.8 | 17 | Hamamatsu | 883 | -2.0 | 30 | -5.3 |
| Hiroshima | I | 2,829 | -7.2 | 12 | Hiroshima | 1,268 | -4.0 | 48 | -7.7 |
| Nagano | I | 2,147 | -3.6 | 17 | Nagano | 292 | -4.4 | 25 | -2.8 |
| Gunma | I | 1,997 | -6.5 | 5 | Takasaki | 1,495 | -2.6 | 31 | +0.7 |
| Okayama | I | 1,928 | -8.7 | 8 | Okayama | 999 | -4.2 | 37 | -7.5 |
| Hokkaidō | II | 5,498 | -15.7 | 36 | Sapporo | 2,072 | -5.4 | 21 | -9.0 |
| Ibaraki | II | 2,956 | -11.7 | 26 | Tsuchiura | 365 | -2.6 | 22 | -5.0 |
| Niigata | II | 2,364 | -11.7 | 23 | Niigata | 525 | -10.0 | 21 | -9.4 |
| Miyagi | II | 2,331 | -15.5 | 9 | Sendai | 1,237 | -4.8 | 36 | -7.4 |
| Fukushima | II | 2,017 | -11.5 | 14 | Koriyama | 274 | -6.3 | 33 | -6.1 |
| Tochigi | II | 1,989 | -10.4 | 17 | Utsunomiya | 429 | -6.3 | 25 | -3.1 |
| Ehime | III | 1,423 | -8.8 | 10 | Matsuyama | 528 | -8.0 | 41 | -7.7 |
| Okinawa | III | 1,385 | -6.2 | — | — | — | — | — | — |
| Ishikawa | III | 1,161 | -7.0 | 9 | Kanazawa | 569 | -2.9 | 20 | -4.4 |
| Toyama | III | 1,089 | -4.1 | 7 | Toyama | 284 | +1.0 | 27 | -0.4 |
| Wakayama | III | 993 | -4.3 | 9 | Wakayama | 424 | -6.7 | 23 | -13.6 |
| Kagawa | III | 980 | -1.3 | 5 | Takamatsu | 384 | -1.5 | 25 | -2.4 |
| Yamanashi | III | 858 | -10.0 | 4 | Kofu | 445 | -4.7 | 13 | -19.7 |
| Fukui | III | 799 | -4.6 | 7 | Fukui | 237 | +0.7 | 16 | -10.7 |
| Shimane | III | 714 | -7.2 | 6 | Matsue | 138 | +0.9 | 24 | -0.4 |
| Kumamoto | IV | 1,806 | -13.3 | 11 | Kumamoto | 720 | -6.6 | 20 | -8.8 |
| Kagoshima | IV | 1,699 | -16.0 | 10 | Kagoshima | 535 | -13.1 | 19 | -5.8 |
| Yamaguchi | IV | 1,447 | -11.0 | 12 | Ube | 191 | -8.7 | 32 | -6.7 |
| Nagasaki | IV | 1,420 | -12.9 | 7 | Nagasaki | 378 | -15.3 | 51 | -8.5 |
| Shiga | IV | 1,397 | -10.6 | 11 | Otsu | 579 | -1.6 | 25 | -4.6 |
| Aomori | IV | 1,368 | -15.4 | 8 | Aomori | 246 | -12.9 | 34 | -11.3 |
| Iwate | IV | 1,325 | -14.4 | 13 | Morioka | 267 | -10.3 | 15 | -10.3 |
| Oita | IV | 1,190 | -12.8 | 9 | Oita | 419 | -7.2 | 23 | -9.4 |
| Yamagata | IV | 1,166 | -13.5 | 12 | Yamagata | 203 | -8.4 | 23 | -7.8 |
| Miyazaki | IV | 1,131 | -12.2 | 9 | Miyazaki | 311 | -7.7 | 23 | -9.7 |
| Akita | IV | 1,084 | -16.4 | 7 | Akita | 266 | -6.8 | 26 | -4.6 |
| Saga | IV | 847 | -10.2 | 8 | Saga | 211 | -9.5 | 21 | -2.7 |
| Tokushima | IV | 778 | -11.3 | 4 | Tokushima | 371 | -14.3 | 18 | -8.1 |
| Kochi | IV | 758 | -12.5 | 6 | Kochi | 297 | -7.5 | 13 | -5.8 |
| Tottori | IV | 584 | -10.6 | 3 | Tottori | 130 | -5.1 | 75 | -5.3 |

*Notes:* One row per rural prefecture origin, sorted by cell and, within cell, by prefecture population. *Prefecture* columns: total population 2010 (thousands) and young 15–29 change 2010–2019 (%), the cell coordinates of Section 5.2 (splits at the 39-region means: 1,810k and −10.1%). *Core UA* = the origin's largest UA by 2010 population; *periphery* entries are medians over the origin's remaining UAs. Three measurement caveats. (i) *Windows differ*: the cells are defined on the prefecture young change over 2010–2019, while UA-grain young data exist only for 2015–2020; the UA columns therefore indicate direction and relative magnitude, not a same-window decomposition of the prefecture figure. (ii) *Sources differ*: prefecture columns are Population Census (Statistics Bureau, Ministry of Internal Affairs and Communications; registry-augmented 2019); UA columns are mesh (grid-square) census population aggregated to UA polygons. (iii) *UA set*: fixed 2010 delineation; each UA is assigned to the single prefecture with the largest population-weighted overlap, and #UAs counts the UAs so assigned. The UA files do not cover Okinawa, whose UA columns are blank; UA-based samples elsewhere in the paper accordingly count 34 origins while the wage sample's 35 include Okinawa.

## OA.11 Prefecture aggregation and the cell I/II split

Building on the origin compositions of Appendix OA.10, this appendix takes up the one place where the prefecture grain does not separate the cells cleanly: the I/II boundary on the large-origin side.

We first show the problem, then resolve it. Section OA.11.1 looks inside the large origins at their cores — regional capital cities such as Sapporo and Sendai: all of them lose their own specialized industries upward, but they differ sharply in whether they absorb the surrounding region's losses — a difference the single prefecture number averages away. Section OA.11.2 shows the large-origin cells each rest on just one or two high-leverage prefectures, while the small-origin cells survive the removal of any single origin — the cell III null, in particular, never turns significantly positive. Section OA.11.3 re-measures the cell-II prefectures at the UA grain, where the "fast-declining" label mostly dissolves — at the grain where their young actually live, most are not fast-declining. The conclusion is that cells I and II should be read as one large-origin finding, not two — as the main text does in finding (ii) (Section 6.1), by interpretation rather than by merging the cells in estimation.

### OA.11.1 What a large rural origin contains: the cores in the Economic Census

Table OA.14 documents what consolidation looks like at the largest rural UAs over 2009–2020. One classification note first: JSIC files the headquarters and branch offices of multi-site firms under separate *management codes* (*kanri-hojo*, "administrative and ancillary economic activities") — 88 of the 494 unified 3-digit codes. These are establishment-function categories rather than industries, so the table counts them separately; *industry* below always means one of the remaining 406 codes. For each UA the table reports: the industries lost and gained; the share of the losses for which the industry's national hosting count also fell — the industry was becoming rarer nationwide ("Consol. %"); how many of the gains came from UAs of the same region ("From region"); the net change in the number of management codes the UA hosts ("Mgmt. net"); and, of the gained management codes, how many were lost by another UA of the same region ("Mgmt. from region").

The sharpest contrast in the table is between two regional capitals, Sendai and Sapporo: they embody two different configurations of a core, distinguished not by their losses but by their *gains*. On the loss side the two look alike: both lose their own specialized industries upward. Sendai lost 16 industries over the period, Sapporo 13. These losses are not local accidents: for 88 percent of Sendai's and 92 percent of Sapporo's, the number of UAs hosting the industry nationwide also fell — the industry was consolidating into fewer cities across the country, and the core was one of the cities it left. Sapporo's losses include petroleum refining, steelmaking, pulp, and recorded media: each was still hosted in Tokyo in 2020 while its nationwide hosting count fell.

The difference is what comes in from below. Sendai *absorbs* its region's consolidation: of the 19 industries it gained, 15 were lost by another Tōhoku UA over the same period. And on the management side — counted separately from those 19 — Sendai gained 20 management codes, of which 14 were lost by another Tōhoku UA: the regional headquarters and branch offices that national firms place in Sendai to run their Tōhoku operations, the classic *branch-office economy* (*shiten keizai*) pattern.

Sapporo absorbs almost nothing on either side. Of its 8 industry gains, 3 came from Hokkaidō's peripheral losses; and while Sapporo too adds management codes (net +16, more than Sendai), only 2 of the 19 it gained were lost elsewhere in Hokkaidō — these functions arrive from outside the region, not as absorption of its own periphery. What Hokkaidō's periphery loses does not reappear in Sapporo: it consolidates to the national apex.

These are sharply different objects to be labeled with one prefecture-level number. A "large declining origin" may contain a core that collects its region's industries even as it sheds its own specialized ones (Miyagi), or a core that the region's losses bypass, going straight to Tokyo (Hokkaidō) — and the prefecture aggregate, averaging core and periphery, represents neither. We do not push these census facts further into the wage data: the JPSED origin is the prefecture, so the wage design cannot price a Sendai-type core differently from a Sapporo-type one. Their role here is diagnostic — they show what the prefecture grain averages away, and the subsections below quantify the consequences for the cell classification and the coefficients.

**Table OA.14:** Upward consolidation at the top of the hierarchy: industry losses and gains of the regional cores.

| UA | Hosted 2009 | Lost | Gained | Net | Consol. (%) | From region | Mgmt. net | Mgmt. from region |
|---|---|---|---|---|---|---|---|---|
| Tokyo | 406 | 3 | 0 | −3 | 33 | — | +1 | — |
| Sapporo | 372 | 13 | 8 | −5 | 92 | 3/8 | +16 | 2/19 |
| Sendai | 358 | 16 | 19 | +3 | 88 | 15/19 | +10 | 14/20 |
| Hiroshima | 371 | 12 | 11 | −1 | 92 | 8/11 | +3 | 5/8 |
| Hamamatsu | 370 | 12 | 12 | +0 | 92 | 5/12 | −1 | 3/10 |
| Numazu | 362 | 14 | 15 | +1 | 71 | 8/15 | −1 | 6/10 |
| Shizuoka | 366 | 12 | 12 | +0 | 83 | 5/12 | +2 | 6/11 |
| Niigata | 342 | 13 | 19 | +6 | 85 | 12/19 | +2 | 4/13 |
| Okayama | 369 | 9 | 14 | +5 | 78 | 3/14 | +2 | 2/12 |
| Takasaki | 383 | 16 | 3 | −13 | 88 | 2/3 | +5 | 5/11 |
| Utsunomiya | 340 | 23 | 12 | −11 | 91 | 5/12 | +13 | 5/20 |
| Asahikawa | 314 | 27 | 14 | −13 | 85 | 9/14 | +7 | 9/12 |
| Hakodate | 312 | 30 | 22 | −8 | 83 | 18/22 | +1 | 6/10 |

*Notes:* From the cleaned 494-industry hosting panel (Section 6); fixed 2010 UA polygons, Economic Census 2009 and 2021 (cross-sections dated 2009 and 2020). *Lost* / *Gained*: the industries the UA stopped / started hosting over 2009–2020, among the 406 product 3-digit codes (the 88 management/auxiliary *kanri-hojo* codes — headquarters and administrative units — are counted separately, in the last two columns). *Consol. (%)*: of the lost industries, the share that also became rarer nationwide over the period — fewer UAs host them in 2020 than in 2009 — signalling that the industry is retreating to fewer, larger cities, not merely leaving this one. *From region*: of the gained industries, how many were lost by another UA of the same region — gains absorbed from the UA's own hinterland (Hokkaidō for Sapporo/Asahikawa/Hakodate, Tōhoku for Sendai, Chūgoku for Hiroshima, Shizuoka for Hamamatsu/Shizuoka/Numazu, northern Kantō for Takasaki/Utsunomiya). *Mgmt. net* / *Mgmt. from region*: the same two measures computed on the 88 management (*kanri-hojo*) codes held out above — the categories for headquarters and administrative branch offices. The *from-region* share — how many of the gained management codes came from the UA's own region — tells an absorbing core (Sendai, 14/20) from a pass-through one (Sapporo, 2/19).

### OA.11.2 Single-prefecture leverage

The cell coefficients are triple interactions on the *continuous* products $S^{\pm}g^{\pm}$, so a single origin's influence on a cell coefficient is governed by its *leverage* — the absolute size of its own regressor value, $|\log \tilde{S}_o \cdot \tilde{g}_o|$, i.e. how far the origin sits from the demeaning center in the (size, decline) plane. What moves a cell coefficient is thus not how many movers an origin contributes but how extreme its product is, because a fitted slope pivots around the regressor's mean: observations near the mean are consistent with almost any slope and barely constrain it, while a far-out origin is a lever arm — a small difference in its outcome tilts the whole line. On the large-origin side this leverage is extremely concentrated: Hokkaidō carries $\log \tilde{S} \cdot \tilde{g} = -6.18$ in the cell-II regressor, 4.5 times the second-largest cell-II prefecture (Miyagi, $-1.36$), while cell II's 148 movers are spread evenly across its six prefectures; on the cell-I side Shizuoka combines the largest leverage (+1.68) with the largest mover count.

Table OA.15 reports the resulting leave-one-out pattern for the headline $\leq 29$ specification (Spec 5). Dropping Shizuoka eliminates the cell-I coefficient ($+0.615^{**} \rightarrow +0.003$, $t = 0.01$); dropping Hokkaidō eliminates the cell-II coefficient ($-0.336^{**} \rightarrow -0.495$, $t = -1.02$); dropping Ibaraki — the third-largest $\tilde{S} > 0$ origin — destabilizes both large-origin cells at once (cell I $t = 0.62$, cell II $t = -0.81$). The entire $\tilde{S} > 0$ half of the design therefore rests on three prefectures. By contrast, the cell-IV coefficient is identified by fifteen prefectures and survives *every* single-prefecture deletion: dropping any one of its own fifteen leaves it between +0.484 and +0.585 (minimum $t = 3.14$). It is also untouched by the deletions that overturn the large-origin cells: without Hokkaidō it stands at +0.570 ($t = 2.12$), without Shizuoka at +0.489 ($t = 3.37$).

**Table OA.15:** Leave-one-out over the large-origin prefectures (Spec 5, $\leq 29$): cells I and II each rest on a single high-leverage prefecture.

| Dropped origin | Cell | Leverage | Movers$_{\leq 29}$ | $\hat{\beta}^{++}_{Sg}$ (I) | $\hat{\beta}^{+-}_{Sg}$ (II) | $\hat{\beta}^{--}_{Sg}$ (IV) |
|---|---|---|---|---|---|---|
| (none: baseline) | — | — | — | $+0.615^{**}$ (+2.3) | $-0.336^{**}$ (-2.2) | $+0.503^{***}$ (+3.2) |
| – Gunma | I | +0.36 | 6 | $+0.768^{***}$ (+3.8) | $-0.350^{**}$ (-2.4) | $+0.478^{***}$ (+3.1) |
| – Nagano | I | +1.11 | 4 | $+0.552^{**}$ (+2.4) | $-0.329^{**}$ (-2.4) | $+0.444^{***}$ (+3.2) |
| – Shizuoka | I | +1.68 | 9 | +0.003 (+0.0) | $-0.373^{**}$ (-2.6) | $+0.489^{***}$ (+3.4) |
| – Okayama | I | +0.09 | 2 | $+0.587^{**}$ (+2.2) | $-0.304^{*}$ (-1.9) | $+0.473^{***}$ (+3.0) |
| – Hiroshima | I | +1.30 | 2 | $+0.680^{**}$ (+2.6) | $-0.355^{**}$ (-2.3) | $+0.512^{***}$ (+3.3) |
| – Hokkaido | II | −6.18 | 6 | +0.712 (+1.6) | −0.495 (-1.0) | $+0.570^{**}$ (+2.1) |
| – Miyagi | II | −1.36 | 5 | $+0.591^{**}$ (+2.4) | $-0.385^{**}$ (-2.0) | $+0.675^{**}$ (+2.5) |
| – Fukushima | II | −0.15 | 5 | $+0.612^{**}$ (+2.0) | $-0.335^{*}$ (-1.8) | $+0.503^{***}$ (+3.2) |
| – Ibaraki | II | −0.79 | 7 | +0.392 (+0.6) | −0.244 (-0.8) | $+0.493^{***}$ (+3.1) |
| – Tochigi | II | −0.03 | 4 | $+0.849^{***}$ (+3.5) | $-0.491^{***}$ (-3.6) | $+0.632^{***}$ (+4.2) |
| – Niigata | II | −0.42 | 8 | $+0.553^{**}$ (+2.0) | $-0.314^{**}$ (-2.0) | $+0.505^{***}$ (+3.3) |

*Notes:* Each row re-estimates Spec 5 of Table 5 (same controls, origin-prefecture fixed effects, wild cluster bootstrap $B = 999$) after dropping all observations with the indicated origin prefecture; origin measures and demeaning centers are held at their baseline values. "Leverage" is the origin's value of the interaction regressor $\log \tilde{S}_o \cdot \tilde{g}_o$ entering its own cell's coefficient. $t$-statistics in parentheses. $^{***}$, $^{**}$, $^{*}$: significance at the 1%, 5%, 10% level.

The small-origin cells pass the same exercise: Table OA.16, the small-side counterpart of

Table OA.15, re-estimates Spec 5 dropping each of the 24 small origins in turn. Cell IV remains significant at every deletion (+0.31 to +0.59, minimum $t = +2.6$). Cell III stays negative and insignificant at every deletion but one: dropping Toyama — the origin anchoring the slow-decline corner ($\tilde{g}_o = +3.7$) — moves it to borderline significance on the *negative* side, the direction the mechanism itself predicts. At no deletion does cell III turn positive.

**Table OA.16:** Cells III and IV under single-origin deletion (Spec 5, $\leq 29$): the small-side counterpart of Table OA.15.

| Dropped origin | Cell | $\hat{\beta}^{-+}_{Sg}$ (III) | $\hat{\beta}^{--}_{Sg}$ (IV) |
|---|---|---|---|
| (none: baseline) | — | −0.180 (-1.0) | +0.503*** (+3.3) |
| – Toyama | III | −0.374** (-2.2) | +0.502*** (+3.3) |
| – Ishikawa | III | −0.176 (-1.0) | +0.504*** (+3.3) |
| – Fukui | III | −0.166 (-0.9) | +0.506*** (+3.3) |
| – Yamanashi | III | −0.050 (-0.3) | +0.310** (+2.6) |
| – Wakayama | III | −0.114 (-0.7) | +0.503*** (+3.2) |
| – Shimane | III | −0.180 (-1.0) | +0.504*** (+3.3) |
| – Kagawa | III | −0.180 (-1.0) | +0.504*** (+3.3) |
| – Ehime | III | −0.128 (-0.8) | +0.458*** (+3.7) |
| – Okinawa | III | −0.184 (-1.0) | +0.504*** (+3.3) |
| – Aomori | IV | −0.181 (-1.0) | +0.501*** (+3.2) |
| – Iwate | IV | −0.180 (-1.0) | +0.504*** (+3.3) |
| – Akita | IV | −0.180 (-1.0) | +0.585*** (+3.6) |
| – Yamagata | IV | −0.179 (-1.0) | +0.490*** (+3.1) |
| – Shiga | IV | −0.190 (-1.0) | +0.510*** (+3.2) |
| – Tottori | IV | −0.186 (-1.0) | +0.508*** (+3.2) |
| – Yamaguchi | IV | −0.182 (-1.0) | +0.504*** (+3.3) |
| – Tokushima | IV | −0.196 (-1.1) | +0.502*** (+3.5) |
| – Kochi | IV | −0.180 (-1.0) | +0.503*** (+3.3) |
| – Saga | IV | −0.199 (-1.1) | +0.523*** (+3.4) |
| – Nagasaki | IV | −0.185 (-1.0) | +0.505*** (+3.3) |
| – Kumamoto | IV | −0.189 (-1.0) | +0.484*** (+3.4) |
| – Oita | IV | −0.178 (-1.0) | +0.501*** (+3.3) |
| – Miyazaki | IV | −0.181 (-1.0) | +0.503*** (+3.3) |
| – Kagoshima | IV | −0.181 (-1.0) | +0.500*** (+3.3) |

*Notes:* Each row re-estimates Spec 5 of Table 5 ($\leq 29$, WLS; same controls, origin-prefecture fixed effects, wild cluster bootstrap $B = 999$) after dropping the indicated origin, with origin measures and demeaning centers held at their baseline values. Columns: the cell III interaction (Move $\times S^- \times g^+$) and the cell IV interaction (Move $\times S^- \times g^-$). Rows: each of the nine cell III origins in turn, then each of the fifteen cell IV origins. $t$-statistics in parentheses. ***/**/*: 1/5/10%.

**The 2SLS counterpart.** The same leverage geometry governs the instruments. Table OA.17 reports per-regressor first stages for the six endogenous regressors of the Spec 4/5 system, separating instrument strength from instrument *construction*: because the triple-interaction instruments are built as $M \times S^{\pm} \times \hat{g}^B$, a large common $S^+$ factor can manufacture a large nominal $F$ from few origins. That is exactly the cell-II pattern: its nominal joint $F = 197.7$ rests on six active origins with a raw corr$(\tilde{g}, \hat{g}^B)$ of only +0.41, and collapses to 3.1 when Hokkaidō is left out — the instrument inherits the single-prefecture leverage of the regressor it targets. The cell-IV first stage, by contrast, is genuinely strong: fifteen active origins, raw correlation +0.64, and a leave-one-out joint-$F$ range of 7.9–17.0 that never approaches collapse. The 2SLS amplification of the cell-II coefficient reported in the main text should therefore be read with the same joint-large-origin caution as its WLS counterpart, while the cell-IV 2SLS rests on a first stage that is robust in the same way the WLS coefficient is.

**Table OA.17:** First-stage diagnostics by cell: the cell-II first stage is a single-origin artifact, cell IV genuinely strong.

| Endogenous regressor | $F$ (Bartik) | $F$ (TFR) | $F$ ($K/L$) | joint $F$ | corr($\tilde{g}, \hat{g}^B$) among active | Active prefs | LOO joint $F$ (min–max) |
|---|---|---|---|---|---|---|---|
| *Panel A: decline side ($g^-$) — carries the framework cells:* | | | | | | | |
| $M \times g^-$ | 40.9 | 5.3 | 3.3 | 13.1 | +0.56 | 21 | 9.3–24.7 |
| $M \times S^+ g^-$ (cell II) | 488.3 | 252.1 | 6.7 | 197.7 | +0.41 | 6 | 3.1–8,205 |
| $M \times S^- g^-$ (cell IV) | 48.8 | 0.5 | 0.5 | 15.3 | +0.64 | 15 | 7.9–17.0 |
| *Panel B: growth side ($g^+$):* | | | | | | | |
| $M \times g^+$ | 8.1 | 2.4 | 0.6 | 3.2 | +0.25 | 13 | 2.3–4.0 |
| $M \times S^+ g^+$ (cell I) | 36.2 | 15.1 | 2.8 | 18.3 | +0.80 | 5 | 10.8–25.5 |
| $M \times S^- g^+$ (cell III) | 11.8 | 3.9 | 0.3 | 6.8 | +0.25 | 8 | 2.9–12.2 |
| N (rural prefectures) | | | | 34 | | | |

*Notes:* Prefecture-level first stages over the 34 rural origins with non-missing instruments. "Active prefs": the number of origins that belong to the cell (its endogenous regressor is non-zero there and zero elsewhere) — e.g. only the six large, fast-declining prefectures for cell II. "corr($\tilde{g}, \hat{g}^B$) among active": the correlation between demeaned young decline $\tilde{g}$ and its Bartik prediction $\hat{g}^B$ over those active origins — a direct check that the instrument tracks decline itself. "LOO joint $F$": range of the joint first-stage $F$ when each active origin is dropped in turn.

This asymmetry is structural rather than accidental: an urban hierarchy is thin at the top by construction, so the number of prefectures that can identify the large-origin cells is inherently small, while the bottom of the hierarchy supplies many origins. The diagnosis also says where to look next: since the large side rests on two or three prefectures, the useful question is what *those* prefectures look like when measured at the unit the theory is written on — the urban area. The next subsection re-measures them there.

### OA.11.3 The cell-II prefectures at the UA grain

Table OA.18 decomposes the six cell-II prefectures into their urban areas (fixed 2010 polygons, the 448 UAs of Section 2.2). Each UA is classified by the same sign rule applied at the UA grain: log UA population (2010) and UA young (15–29) growth (2015–2020; the 1 km census mesh carries five-year age classes only from 2015, so this is the only mesh-feasible young window), each demeaned over all 448 UAs. The middle columns report the share of the prefecture's census young population (2015) living in each UA type.

Two patterns emerge. First, four of the six cell-II prefectures are *mixtures dominated by a cell-I-type core*: in Hokkaidō 54% of the young live in UAs whose young decline is *slower* than the all-UA mean — Sapporo (young change −5.4%, milder than the all-UA mean of −6.8%; its total population even *grew* over 2010–2020) — with the fast-declining smaller UAs (Hakodate, Kushiro, Otaru) holding under a quarter; Miyagi (Sendai) is even more concentrated at 61%. So the prefecture's "large × fast-declining" (cell-II) label is an averaging artifact: the label is true of the prefecture's aggregate numbers, but no single UA inside it is both large and fast-declining — the large UA (Sapporo) declines slowly, and the fast-declining UAs are small. Second, Niigata and Fukushima are *coherent* cell-II cases — their principal UAs themselves decline fast (Niigata UA: −10.0%) — yet they barely move the cell-II coefficient. The reason is the leverage logic of Section OA.11.2: they sit close to the demeaning center, so their products are small ($\log \tilde{S}_o \cdot \tilde{g}_o = -0.42$ and $-0.15$, a fraction of Hokkaidō's $-6.18$) and the two genuine cell-II cases contribute almost nothing — the coefficient is driven by Hokkaidō, the artifact case, which the next paragraph re-measures at the UA grain.

The last columns then reclassify each prefecture at the UA grain. We recompute its two origin coordinates as the young-population-weighted average of its UAs' log 2010 size and 2015–2020 young growth (weighting by 2015 young population, so the averaged decline is exactly the decline of the prefecture's UA-resident young), demean by the same region-39 convention at this grain, and read off the resulting cell (last column of Table OA.18). The cell-II label largely evaporates: of the six, only Hokkaidō stays in cell II, and only marginally — the others move to cells I, III, or IV.

Its leverage falls accordingly. Hokkaidō's cell-II leverage $\log \tilde{S}_o \cdot \tilde{g}_o$ was $-6.18$ ($= 1.11 \times -5.6$), the single largest in the design and the value that drove the entire cell-II coefficient (Section OA.11.2); at the UA grain it becomes $-0.63$ ($= 0.65 \times -0.97$), an order of magnitude smaller. The reason is concrete: Hokkaidō's prefecture-total young decline of $-5.6$ is set by its thinly-populated rural periphery, but the UAs where most of its young actually live — Sapporo above all, holding 54% of them — decline far more mildly, so measuring decline where the young live moves Hokkaidō back toward the boundary.

The "large fast-declining" cell-II label was thus a prefecture-aggregation artifact — and it was carrying the entire cell, since the coefficient rests on Hokkaidō alone. At the grain where its young actually live, Hokkaidō is a large, slowly-declining high-order core, the same object as the cell-I prefectures Shizuoka and Hiroshima and not a distinct mechanism. Cells I and II are therefore one finding, not two; their split is an aggregation accident. The main text draws the consequence in finding (ii) (Section 6.1): the two large cells are read jointly, with no separate I/II interpretation — and the joint reading stays at the level of interpretation: we do not pool cells I and II into a single regressor, because doing so would discard the size-graded structure — where a larger origin amplifies the periphery's decline response — that the design rests on, and such a merged specification is misspecified.

## OA.12 Marginal effects of origin size by cell

Table OA.19 reports the marginal effect of origin size on the Tokyo MA move premium, $\widehat{\mathrm{ME}}(S) = \hat{\beta}_3^{\pm} + \hat{\beta}_{Sg}^{\pm\pm}\, g$, by cell and at three decline-depth points ($g \in \{\pm 1.0, \pm 1.5, \pm 2.0\}\, \sigma_g$); the framework

**Table OA.18:** The six cell-II prefectures decomposed at the UA grain.

| | Prefecture grain | | | Young-2015 share (%) by UA-grain cell type | | | | | UA-weighted measures | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Origin | $\log \tilde{S}_o$ | $\tilde{g}_o$ | Lev. | I | II | III | IV | non-UA | $\log \tilde{S}_o^{UA}$ | $\tilde{g}_o^{UA}$ | Cell |
| Hokkaido | +1.11 | −5.6 | −6.18 | 54 | 19 | 3 | 4 | 20 | +0.65 | −0.97 | II |
| Miyagi | +0.25 | −5.4 | −1.36 | 61 | 7 | 1 | 3 | 28 | +0.90 | +0.43 | I |
| Fukushima | +0.11 | −1.4 | −0.15 | 22 | 26 | 6 | 3 | 44 | −0.82 | −1.63 | IV |
| Ibaraki | +0.49 | −1.6 | −0.79 | 33 | 9 | 7 | 6 | 45 | −0.72 | +1.54 | III |
| Tochigi | +0.10 | −0.3 | −0.03 | 46 | 0 | 10 | 3 | 42 | −0.84 | +1.24 | III |
| Niigata | +0.27 | −1.6 | −0.42 | 17 | 33 | 3 | 9 | 38 | −0.68 | −3.03 | IV |

*Notes:* For the young-share columns, each UA is typed by the sign of its (demeaned log 2010 population, demeaned young 15–29 growth 2015–2020), the demeaning centers being the means over all 448 UAs (each UA counting once). The middle columns give the share of the prefecture's 2015 census young population living in each UA type; "non-UA" is the young living outside every 2010 UA polygon. The last columns re-measure the prefecture's own size and decline as young-2015-weighted means over its UAs (demeaned by the same region-39 centers), and "Cell" is the cell that classification implies.

predictions and their reading are in Section 6.1. These are the *unweighted OLS* estimates; the weighted WLS headline versions are in the main paper's Appendix D, and the two agree in sign at every cell and evaluation point, so weighting changes neither the results nor their interpretation.

The gradients carry the framework's signs on both sides: positive on the large side — the premium grows with origin size — and negative on the small fast-declining side — it grows as size falls (Section 6.1 reports the headline values). On the large side we do not read how the gradient moves with decline: the informative dimension there is size, the I-versus-II split is not separately identified at the prefecture grain, and the two large cells are read jointly (finding (ii)). On the small side, by contrast, the $g$-dependence *is* read, because cell IV pools prefectures made of small UAs throughout — no large core, so no core–periphery mixture contaminates its $\tilde{g}$: deeper decline crosses more necessity thresholds at once, and ME($S$) deepens accordingly, from $-0.90^{***}$ at $-1.0\sigma_g$ to $-2.46^{***}$ at $-2.0\sigma_g$ (Table OA.19). Cell III, where neither mechanism is active, is null throughout.

## OA.13 Cell assignment and the predetermined component of decline

The four cells are defined on the *observed* demeaned drivers, while the 2SLS identifies from the *predetermined* component of decline; the two can disagree on sign. This appendix documents where they disagree, and verifies that the cell estimates survive dropping those origins. Regressing the observed $\tilde{g}_o$ on the three predetermined shocks over the 34 rural origins (a composite origin-level first stage, $R^2 = 0.50$) yields a prediction $\hat{g}_o$ whose sign differs from the observed cell side in six origins (Table OA.20). Two are boundary cases: Tochigi and Shiga have observed decline within half a percentage point of the demeaning center — essentially sitting on the cell boundary $\tilde{g} = 0$ — so the tiny difference between their realized and predetermined decline is enough to put them on opposite sides of it, and their sign-disagreement is mechanical rather than substantive. For the remaining four origins, the predetermined prediction and the realized decline truly differ; the clearest

**Table OA.19:** Marginal effect of origin size on the Tokyo MA wage premium, by cell and decline depth (unweighted OLS).

| Cell (g evaluation point) | Unweighted (OLS) Full rural (Spec 4′) | Unweighted (OLS) 20–29 (Spec 5′) |
|---|---|---|
| Cell I (slow-decl, g = $+1.0\sigma_g$ = +3.95) | +0.4602** (0.2184) | +0.6462*** (0.2401) |
| Cell I (slow-decl, g = $+1.5\sigma_g$ = +5.93) | +0.8796** (0.3987) | +1.4247*** (0.5427) |
| Cell I (slow-decl, g = $+2.0\sigma_g$ = +7.91) | +1.2991** (0.5899) | +2.2033** (0.8845) |
| Cell II (g = $-1.0\sigma_g$ = -3.95) | +0.1148 (0.0819) | +0.0494 (0.1352) |
| Cell II (g = $-1.5\sigma_g$ = -5.93) | +0.3615*** (0.0788) | +0.5295*** (0.1712) |
| Cell II (g = $-2.0\sigma_g$ = -7.91) | +0.6083*** (0.1499) | +1.0096** (0.3921) |
| Cell III (slow-decl, g = $+1.0\sigma_g$ = +3.95) | -0.1809 (0.2576) | -0.1540 (0.3080) |
| Cell III (slow-decl, g = $+1.5\sigma_g$ = +5.93) | -0.4975 (0.3934) | -0.5622 (0.4355) |
| Cell III (slow-decl, g = $+2.0\sigma_g$ = +7.91) | -0.8141 (0.5369) | -0.9703 (0.6421) |
| Cell IV (g = $-1.0\sigma_g$ = -3.95) | -0.0023 (0.1900) | **-0.8991***** (0.2522) |
| Cell IV (g = $-1.5\sigma_g$ = -5.93) | -0.2296 (0.2821) | **-1.6799***** (0.4727) |
| Cell IV (g = $-2.0\sigma_g$ = -7.91) | -0.4569 (0.3851) | **-2.4607***** (0.7516) |

*Notes:* Unweighted OLS marginal effects (the weighted WLS counterparts are in the main paper's Appendix D). Standard errors in parentheses are score-based wild cluster bootstrap ($B = 999$, Rademacher) at the origin prefecture. Columns are the full-rural (Spec 4′) and 20–29-only (Spec 5′) sub-samples. Bold cells in the 20–29 (Spec 5′) column are the framework-predicted ME(S) for cell IV (< 0), evaluated at three fast-decline points on the negative-$g$ axis. ***, **, * denote significance at the 1%, 5%, 10% level. $\sigma_g$ is the within-sample standard deviation of demeaned origin decline rate $\tilde{g}$ in percentage points. The bold cell-IV ME($S$) is negative and steepens across the three decline points: the premium rises the smaller and faster-declining the origin.

cases have documented sources. Miyagi's cohort fundamentals predict essentially average decline ($\hat{g} = +0.36$) while its realized young decline is $-5.35$: the gap is the non-demographic outflow after the 2011 Tōhoku earthquake, exactly the component a predetermined instrument should not contain. Shimane and Ehime show the reverse timing: their heavy young outflow came *before* 2010 and had largely run its course by the observation window, so the realized 2010s decline looks mild even though the cohort fundamentals still predict deep decline.

**Table OA.20:** Origins where the composite predetermined prediction $\hat{g}_o$ and the observed $\tilde{g}_o$ disagree in sign.

| | Observed $\tilde{g}_o$ | Composite $\hat{g}_o$ | Raw Bartik |
|---|---|---|---|
| Miyagi | −5.35 | +0.36 | −8.31 |
| Ibaraki | −1.62 | +2.12 | −5.99 |
| Tochigi | −0.29 | +0.26 | −8.11 |
| Shiga | −0.51 | +5.83 | −1.88 |
| Shimane | +2.92 | −1.56 | −11.47 |
| Ehime | +1.29 | −0.61 | −10.12 |
| Rural mean (raw Bartik) | | | −9.08 |

*Notes:* $\hat{g}_o$ is the fitted value from regressing the observed demeaned young-population decline $\tilde{g}_o$ (2010–2019, 39-region demean) on the demographic Bartik, the 1965–85 TFR change, and 1980–85 capital deepening over the 34 rural origins ($R^2 = 0.50$; sign agreement 28/34). The *Raw Bartik* column is each origin's demographic Bartik prediction, which averages −9.08 over the rural origins.

Table OA.21 re-estimates the observed-cell (cell-fixed) piecewise without these six origins — those whose predetermined decline component $\hat{g}_o$ carries the opposite sign to the observed $\tilde{g}_o$, i.e. exactly the origins whose cell assignment would flip under the decline variation the instruments identify from. The framework cells survive: cell II keeps its sign and magnitude on half of its original membership ($-0.336^{**}$ to $-0.402^{*}$), and cell IV *strengthens* ($+0.503^{***}$ to $+0.806^{***}$). This is what the specification implies. Because the interaction regressors are the continuous products $S^{\pm}g^{\pm}$, each origin's weight in a cell coefficient rises with its distance from the center, $|\tilde{g}|$ (and $|\log \tilde{S}|$): the deep, unambiguous origins — deeply declining small ones, clearly large fast-declining ones — carry the coefficient, while origins near the $\tilde{g} = 0$ boundary contribute little because their regressors are small there. The six dropped origins are exactly the cases above — the two boundary cases, and the four whose realized decline genuinely departs from the demographic prediction — so removing them leaves the estimates essentially intact — the sense in which the cell-fixed estimates are robust. (Three of the eleven large origins — Miyagi, Ibaraki, Tochigi — are among the dropped six, so the large side thins: in Table OA.21 cell I keeps its point estimate ($+0.615 \to +0.545$) but loses significance (finding (ii)). The framework cells strengthen in the $\leq 29$ columns (cell II $-0.336 \to -0.402$, cell IV $+0.503 \to +0.806$), while the full-rural columns fade. That fading is not a sample-size effect — the same six origins are dropped from both samples, removing about a fifth of the observations in each — but reflects that the premium is a young-worker phenomenon: the all-age columns are diluted by older movers who carry no premium, so they were never the framework test. The $\leq 29$ columns are the headline.)

**Table OA.21:** Cell-fixed piecewise without the sign-disagreeing origins.

| | $\leq 29$ | | Full rural | |
|---|---|---|---|---|
| | Baseline | Flips dropped | Baseline | Flips dropped |
| Cell I (large × slow) | $+0.615^{**}$ | $+0.545$ | $+0.220^{**}$ | $+0.060$ |
| | (0.265) | (0.497) | (0.107) | (0.285) |
| Cell II (large × fast) | $-0.336^{**}$ | $-0.402^{*}$ | $-0.172^{***}$ | $-0.079$ |
| | (0.156) | (0.213) | (0.054) | (0.124) |
| Cell III (small × slow) | $-0.180$ | $-0.254$ | $-0.145$ | $-0.134$ |
| | (0.181) | (0.167) | (0.105) | (0.136) |
| Cell IV (small × fast) | $+0.503^{***}$ | $+0.806^{***}$ | $+0.140^{**}$ | $+0.076$ |
| | (0.155) | (0.196) | (0.059) | (0.086) |
| $N$ | 4,617 | 3,711 | 47,536 | 38,184 |

*Notes:* WLS with JPSED longitudinal weights, origin and wave fixed effects, demographic controls as in the main specification (Table 5, Specs 4–5); score-based wild cluster bootstrap standard errors ($B = 999$, Rademacher) at the origin prefecture in parentheses. "Flips dropped" excludes the six origins of Table OA.20. $^{***}$, $^{**}$, $^{*}$: significance at the 1%, 5%, 10% level.

## OA.14 Bartik instrument for the decline rate: construction and 2SLS

This appendix details the construction of the Bartik-style instrument for the prefecture-level young-population decline rate $g_o$, the two further predetermined shocks, and the first-stage diagnostics and 2SLS estimates (linear and 4-way piecewise) behind Section 6.2.

**Identification logic.** The actual decline rate $g_o$ over 2010–2019 has two components: (a) a *mechanical* component driven by past birth rates and demographic momentum (cohorts born in 1981–2004 entering the 15–29 age range over the window), and (b) an *endogenous* component driven by net migration over 2010–2019, which itself responds to within-window labor-market conditions. Component (b) generates the reverse-causation concern in the WLS Move × Decline regression: workers who anticipate weakening local labor markets migrate out, accelerating observed decline; the WLS estimate of $\beta_4$ then conflates the migration response to hosting-threshold-driven industry exit with the migration-driven component of measured decline itself. The Bartik instrument isolates component (a) by using only predetermined birth records and national-level survival rates.

**Construction.** Let $B_{o,b}$ denote prefecture-$o$ birth count in year $b$, observed in the published prefecture-by-year birth series of the Vital Statistics (Ministry of Health, Labour and Welfare), and $\sigma_a^{\text{nat}}$ the national survival rate from birth to age $a$. For young (15–29) population in prefecture $o$ at year $T$, the cohort birth-years span $[T-29, T-15]$. The published series provides 5-year birth observations 1985, 1990, 1995, 2000, 2005, each representing the central year of a nominal 5-year birth period; we therefore approximate

$$\widehat{S}_{o,T}^{\text{young}} = \sum_{b \in \{1985,1990,\ldots\}: T-29 \le b \le T-15} 5 \cdot B_{o,b} \cdot \sigma_{T-b}^{\text{nat}}, \tag{OA.1}$$

$$\widehat{g}_o^{\text{Bartik}} = 100 \cdot \log\big(\widehat{S}_{o,2019}^{\text{young}} / \widehat{S}_{o,2010}^{\text{young}}\big). \tag{OA.2}$$

Each 5-year birth observation is multiplied by 5 to convert to a total cohort size, and $\sigma_a^{\text{nat}}$ is set to a flat 0.99 across ages 15–29 (modern Japanese 0-to-15-29 survival is approximately 99% across cohorts; using age-specific survival rates instead changes almost nothing: survival is near one at every age, so virtually all of the instrument's cross-prefecture variation comes from the birth counts $B_{o,b}$ themselves). The 2010 young population draws on births in 1985, 1990, 1995 (cohorts aged 25, 20, 15 in 2010); the 2019 young population draws on births in 1990, 1995, 2000 (cohorts aged 29, 24, 19 in 2019, with 2000 capturing the 19-year-olds; the 2005 birth observation captures the 14-year-olds, just outside the young-population age range, and is therefore not used). The data span covered is exactly the cohorts whose mechanical aging composes the 2010 and 2019 young populations, with no overlap with the JPSED 2016–2025 wage panel.

**Why 5-year birth observations rather than annual.** The instrument could instead be built from annual birth counts, and the annual data exist for every cohort year of both windows: prefecture-level annual births are machine-readable for 1989–2020, and we digitized the earlier years from the printed *Vital Statistics of Japan* volumes (Ministry of Health, Labour and Welfare) (Table 4.3, "live births by prefecture and year"): 1987–88 from the 1992 edition and 1981–86 from the 1986 edition. The digitized years are fully validated — in every year the 47 prefecture counts sum exactly to the volume's printed national total, and the 1980 and 1985 columns reproduce the quinquennial series of equation (OA.1) prefecture by prefecture. The annual construction then sums each cohort year's observed births over $[T-29, T-15]$, with no interpolation anywhere.

It is nevertheless *weaker* than the 5-year construction: on the same standalone rural-sample regression, the annual first stage gives $F = 19.3$ (Pearson correlation +0.60) against $F = 28.9$ (+0.67) for the 5-year construction. Call the birth years that construction uses its *anchors*: $\{1985, 1990, 1995\}$ for the 2010 target and $\{1990, 1995, 2000\}$ for 2019 (the 1980 and 2005 observations of the quinquennial series fall outside both age windows and are not used).

Why does the coarser construction win? Two easy explanations can be ruled out first. The first is that the annual series is simply noisier — that year-to-year fluctuations in the recorded births drag the annual instrument down. If so, smoothing those fluctuations out should recover the lost relevance; but smoothing the annual counts with a centered 5-year moving average before summing leaves the annual first stage essentially unchanged ($F = 19.0$). The second is that the anchors happen to fall on especially informative years. If so, the true annual births summed over just the years the anchors span should do as well; instead they do worse still ($F = 9.5$). Neither the noise nor the choice of years accounts for the gap.

What remains is that the two constructions measure different things. The annual construction records the exact size of every birth cohort that ages into and out of the young population — fine-grained and accurately measured. The anchor construction is coarser. Its predicted decline is the difference between the two target young populations, 2019 minus 2010, and each is built from three 5-year birth anchors: 1985, 1990, 1995 for the 2010 target, and 1990, 1995, 2000 for the 2019 target. The two sets share 1990 and 1995, so those cancel in the difference; what remains is a single number per prefecture — the 2000 birth count minus the 1985 birth count, i.e. how far the prefecture's births fell from 1985 to 2000, its long-run birth trend. That trend is the better predictor: realized young decline, which includes migration, follows the prefecture's persistent trajectory, not the year-to-year fluctuations in cohort size. So the extra detail the annual construction adds — exactly how many were born in 1981–84 or 2001–04 — is not measurement noise (smoothing it away did not help, $F$ stays 19), but real, accurately measured cohort variation that realized decline simply does not follow; adding it only dilutes the instrument's relevance. We therefore retain the 5-year construction in equation (OA.1): it keeps the persistent trend that realized decline follows and drops the cohort detail that realized decline does not follow.

**Exogeneity.** The instrument satisfies the exclusion restriction in two complementary senses. First, the birth records $B_{o,1985}, \ldots, B_{o,2005}$ are observed long before the JPSED wage panel begins (the most recent birth observation is in 2005, 11 years before the first wage-growth pair in 2016/17), and are mechanically determined by past total fertility rates and prefecture-level demographic conditions in those birth years. They cannot be affected by 2016–2025 labor-market shocks. Second, the national survival rate $\sigma_a^{\text{nat}}$ is by construction prefecture-invariant and does not capture prefecture-specific outmigration or mortality shocks. The instrument therefore predicts the *counterfactual* 2010 and 2019 young populations under the assumption of zero net migration after 1985, isolating the demographic-momentum component of decline.

**First-stage diagnostics.** First-stage diagnostics for the Bartik are collected in Table OA.22: across the 35 rural origins the Bartik alone has a first-stage slope of +0.90 with $F = 27.3$ and Pearson correlation +0.67 with the actual $g_o$, comfortably clearing the conventional weak-instrument benchmark; the residual variation reflects the migration-driven endogenous component that the instrument is designed to strip out.

**4-way piecewise under IV.** The 4-way piecewise specification (Spec 4 / Spec 5 in Table 5) has six endogenous regressors: the four cell interactions $M \times S^{\pm} g^{\pm}$ (one per size-sign $\times$ $g$-sign cell) and the two move-interacted decline pieces $M \times g^{+}$ and $M \times g^{-}$ (notation as in equation (6)). A single Bartik IV cannot separately identify all six coefficients: in the first stage, every endogenous regressor would be projected onto the same one-dimensional fitted $\widehat{g}_o$, so all four cell interactions end up as algebraic transformations of the *same* prediction and lose any independent cross-prefecture variation that could distinguish one cell from another in the second stage. The same logic applies to the TFR or K-deepening IV alone. We therefore combine three independent IV sources: together

**Table OA.22:** First-stage diagnostics for three predetermined IVs of $g_o$, rural prefectures.

| | $N$ | slope | first-stage $F$ | corr(IV, $g_o$) |
|---|---|---|---|---|
| $\widehat{g}_o^{\text{Bartik}}$ alone | 35 | +0.896 | 27.29 | +0.673 |
| $\widehat{\text{TFR}}_o^{\Delta}$ alone | 34 | +11.31 | 6.05 | +0.399 |
| $\widehat{g}_o^{K/L}$ alone | 35 | −1.156 | 3.68 | −0.317 |
| demo Bartik + TFR | 34 | — | 14.27 | — |
| demo Bartik + $\widehat{g}_o^{K/L}$ | 35 | — | 14.78 | — |
| TFR + $\widehat{g}_o^{K/L}$ | 34 | — | 4.44 | — |
| demo Bartik + TFR + $\widehat{g}_o^{K/L}$ | 34 | — | 10.12 | — |

*Pairwise rural-sample correlations among IVs:*
corr(demo Bartik, TFR) = +0.348, corr(demo Bartik, $K/L$) = −0.230, corr(TFR, $K/L$) = −0.201.

*Notes:* First-stage WLS regressions of demeaned $g_o$ on the indicated IV(s) across the 35 rural-prefecture origins (34 when $\widehat{\text{TFR}}_o^{\Delta}$ is included, dropping Okinawa). Single-IV rows report the slope, $F$, and Pearson correlation with $g_o$; a multi-IV first stage has no single slope, so those rows report only the joint $F$.

they furnish enough independent cross-prefecture variation to identify the six endogenous regressors separately at the cell level.

The two additional predetermined instruments — the 1965–85 TFR change $\widehat{\text{TFR}}_o^{\Delta}$ and the 1980–85 political capital deepening $\widehat{g}_o^{K/L}$ — are constructed in Appendix C of the main paper (equations C.1 and C.2); this appendix documents why they tighten identification, their economic independence, and the resulting first stages.

**Why the additional instruments tighten the exclusion restriction.** The main residual concern with the Bartik alone is a confounder in the exclusion-restriction sense: a prefecture trait that drives both the instrument and today's wages directly. The concrete candidate is persistent rural/agricultural character — the prefectures with high 1981–2004 fertility may have low wage levels and growth today for reasons unrelated to decline-induced consolidation. Requiring a common Move × Decline effect across three IVs that operate through different historical channels addresses this: a confounder would have to thread through all three channels at once, far less plausible than one acting on the Bartik alone.

The capital-deepening instrument $\widehat{g}_o^{K/L}$ is the 1980–85 rate of capital deepening, and its identifying economics runs through Japan's 1980s public-investment cycle. That cycle channeled capital into the peripheral prefectures outside the Pacific industrial belt,[2] but the capital was politically allocated and raised little productivity: Mori (2026)'s Solow decomposition over 1980–85 attributes about 60% of $g(Y/L)$ in the non-belt prefectures to capital deepening and only 17% to TFP, against 33% and 52% in the Tokyo MA. This is what makes the instrument *relevant* to $g_o$:

[2]The *Pacific industrial belt* is the coastal manufacturing zone running from Tokyo through Nagoya and Osaka to northern Kyushu; in Mori (2026) it comprises the 15 prefectures along this coast (Saitama, Chiba, Tokyo, Kanagawa, Shizuoka, Aichi, Mie, Shiga, Kyoto, Osaka, Hyogo, Okayama, Hiroshima, Yamaguchi, Fukuoka), and *non-belt* means the remaining 31 prefectures, Okinawa excluded.

because the injected capital did not raise the marginal product of labor, the heavily over-invested prefectures gained no lasting labor-market advantage, and thinned out faster over 2010–2019 as the artificial capital depreciated. The same fact secures the *exclusion restriction*, for two reasons: the 25–40 year lag to the JPSED wage panel rules out any direct contemporaneous channel, and, because the capital demonstrably failed to raise TFP, it cannot reach today's wages through a lasting productivity advantage — the one route that a long lag by itself cannot close, since a permanent productivity gain would still shape wages decades later. The 1980–85 window is chosen to preserve exactly this: it is the period over which Mori (2026) documents the politically-allocated, non-productivity-matched capital, whereas extending into the late-1980s bubble would add privately financed speculative capital concentrated in the three major metropolitan areas (Stone and Ziemba, 1993) — reconnecting the instrument to metropolitan demand conditions, exactly the channel the short window is chosen to exclude.

**Economic independence of the three instruments.** The three instruments target the same endogenous variable $g_o$ but pick up cross-prefecture variation through economically distinct channels.

The Bartik instrument loads on 1981–2004 birth *counts*. By the identity $\text{Births}_{o,b} = \text{TFR}_{o,b} \times W_{o,b}$ (with $W_{o,b}$ the prefecture-$o$ stock of women of childbearing age in year $b$), its cross-prefecture variation combines two sources: the per-woman fertility rate $\text{TFR}_{o,b}$, and the stock $W_{o,b}$, which itself reflects earlier migration and earlier births. A high-Bartik prefecture may therefore be high because of contemporaneous fertility or because a large $W$ stock was already in place, and these two sources are not separable from the Bartik alone.

The TFR-decline instrument $\widehat{\text{TFR}}_o^{\Delta} = \text{TFR}_{o,1985} - \text{TFR}_{o,1965}$ shuts off the second source. It is built from the per-woman fertility rate (the $W$ stock is already divided out by the rate's definition) and is computed over a 1965–85 window predetermined to the Bartik birth window. The TFR-decline instrument therefore isolates the cultural–behavioral component of fertility, independently of any stock-of-women channel that could correlate the Bartik with the current labor market. Two prefectures with the same Bartik value can have very different TFR-decline values—a high-TFR-decline prefecture whose Bartik nonetheless looks high because a large $W$ stock remained, or a moderate-TFR-decline prefecture whose Bartik looks low because earlier out-migration thinned $W$. The rural-sample correlation between the two IVs is accordingly only $+0.35$.

The K-deepening instrument operates through a third channel entirely: spatial-economic structure inherited from Japan's 1980s political-investment cycle, which redistributed capital without a corresponding productivity response (Mori, 2026). It is empirically negatively correlated with both demographic IVs ($-0.23$ with Bartik, $-0.20$ with TFR decline; Table OA.22).

The Hansen $J$ test asks whether the three instruments point to the same Move $\times$ Decline coefficient. If a confounder biased one channel only — say, a hidden $W$-stock channel affecting the Bartik but not the TFR-decline IV — the Bartik-implied coefficient would drift away from the others and the $J$ test would reject. Non-rejection therefore supports the reading that no single-channel confounder is at work: to bias the estimate it would have to run through 1981–2004 demographic momentum, 1965–85 per-woman fertility behavior, *and* 1980s political capital allocation at once.

Table OA.23 documents combination stability at the linear Spec 3 level: every instrument combination involving the demographic Bartik gives a Move $\times$ Decline coefficient in the narrow band $-0.011$ to $-0.013$ with non-rejecting Hansen $J$ ($p \geq 0.46$), while the TFR and K-deepening instruments alone are too imprecise to identify anything — the identifying power comes from the Bartik, and the three channels agree. At this linear level the coefficient is small and insignificant, matching the main text: the pooled linear aggregate carries little signal (Table 5, Specs 1–3), and the content appears at the cell level (Table OA.24 below).

**Table OA.23:** Spec 3 Move $\times$ Decline coefficient: WLS, just-identified, and over-identified 2SLS.

| Specification | coefficient | SE | $t$ | Hansen $J$ ($p$) |
|---|---|---|---|---|
| WLS | +0.0051 | 0.0098 | +0.52 | — |
| *Just-identified (1 IV):* | | | | |
| $\widehat{g}_o^{\text{Bartik}}$ alone | −0.0114 | 0.0092 | −1.24 | — |
| $\widehat{\text{TFR}}_o^{\Delta}$ alone | −0.0247 | 0.0207 | −1.19 | — |
| $\widehat{g}_o^{K/L}$ alone | +0.0707 | 0.1259 | +0.56 | — |
| *Over-identified (2 IVs, $J$ df $= 1$):* | | | | |
| demo Bartik + TFR | −0.0126 | 0.0089 | −1.42 | 0.309 (0.578) |
| demo Bartik + $\widehat{g}_o^{K/L}$ | −0.0112 | 0.0092 | −1.22 | 0.556 (0.456) |
| TFR + $\widehat{g}_o^{K/L}$ | −0.0203 | 0.0167 | −1.22 | 0.739 (0.390) |
| *Over-identified (3 IVs, $J$ df $= 2$):* | | | | |
| demo Bartik + TFR + $\widehat{g}_o^{K/L}$ | −0.0125 | 0.0089 | −1.41 | 1.105 (0.575) |

*Notes:* WLS / 2SLS estimates of the Spec 3 Move $\times$ Decline coefficient on the baseline COVID-clean wage-growth sample (pre-COVID transitions, JPSED longitudinal pair weights). The endogenous regressor is Move $\times$ $g_o$, instrumented by the listed shock(s) interacted with Move. Standard errors are score-based wild cluster bootstrap ($B = 999$, Rademacher) at the origin prefecture. Hansen $J$ tests the over-identifying restrictions: $H_0$ is that all listed instruments identify the same coefficient. TFR-using specifications drop Okinawa ($N = 46{,}500$; otherwise $N = 47{,}536$). The three-IV row is the linear entry of Table 6 in the main text. ***, **, * denote significance at the 1%, 5%, 10% level.

**Three-IV combination resolves the 4-way under-identification.** Each endogenous regressor gets one matched instrument from each shock, built by putting the shock in the place of $g_o$. For example, cell II's regressor is $M \times S^+ g^-$; its Bartik-based instrument is $M \times S^+ \times \text{Bartik}^-$ (the Bartik split at zero, entering through the same Move and size-side interactions), and the TFR and K-deepening shocks contribute the analogous instruments. Six endogenous regressors $\times$ three shocks give 18 excluded instruments, over-identified by 12. The resulting 2SLS estimates of the framework-key cells are reported in Table OA.24.

All four cells retain their WLS signs under 2SLS, and instrumenting moves the framework cells *away* from WLS rather than toward it — the direction opposite to what weak-instrument bias would produce (Section 6.2). In the $\leq 29$ sample every cell strengthens (cell I $+0.615 \to +0.910$,

**Table OA.24:** 4-way piecewise cell coefficients: WLS vs. over-identified 2SLS (18 instruments), full rural sample and $\leq 29$ sub-sample.

| | Full rural | | $\leq 29$ | |
|---|---|---|---|---|
| | WLS | 2SLS | WLS | 2SLS |
| Cell I ($\beta_{Sg}^{++}$) | $+0.220^{**}$ | $+0.288^{***}$ | $+0.615^{**}$ | $+0.910^{***}$ |
| | (0.107) | (0.112) | (0.265) | (0.197) |
| Cell II ($\beta_{Sg}^{+-}$) | $-0.172^{***}$ | $-0.200^{***}$ | $-0.336^{**}$ | $-0.461^{***}$ |
| | (0.054) | (0.058) | (0.156) | (0.131) |
| Cell III ($\beta_{Sg}^{-+}$) | $-0.145$ | $-0.156^{*}$ | $-0.180$ | $-0.321^{*}$ |
| | (0.105) | (0.092) | (0.181) | (0.168) |
| Cell IV ($\beta_{Sg}^{--}$) | $+0.140^{**}$ | $+0.138^{**}$ | $+0.503^{***}$ | $+0.622^{***}$ |
| | (0.059) | (0.054) | (0.155) | (0.171) |
| $N$ (pairs) | 47,536 | 46,500 | 4,617 | 4,528 |
| Excluded IVs | — | 18 | — | 18 |
| Hansen $J$ (df = 12) | — | 13.53 | — | 16.96 |
| $p$-value | — | 0.332 | — | 0.151 |

*Notes:* WLS columns are Specs 4 and 5 of Table 5. 2SLS: the endogenous regressors are the six interactions containing the (endogenous) decline rate — Move $\times\, g^{\pm}$ and the four cell terms Move $\times\, S^{\pm} \times g^{\pm}$. Each is instrumented by putting each of the three predetermined shocks (demographic Bartik, 1965–85 TFR change, 1980–85 capital deepening) into that same interaction in place of $g$ — $3 \times 6 = 18$ instruments. TFR-using 2SLS drops Okinawa. Standard errors in parentheses are score-based wild cluster bootstrap ($B = 999$, Rademacher) at the origin prefecture. Hansen $J$ tests the 12 over-identifying restrictions. The $\leq 29$ columns reproduce Table 6 of the main text. ***, **, *: significance at the 1%, 5%, 10% level.

cell II $-0.336 \rightarrow -0.461$, cell IV $+0.503 \rightarrow +0.622$, all significant at 1%; cell III $-0.180 \rightarrow -0.321$, borderline); in the full rural sample the pattern is the same in milder form. The Hansen $J$ test does not reject the 12 over-identifying restrictions in either sample ($p = 0.332$ full, $p = 0.151$ at $\leq 29$). The three-source combination thus delivers a separate causal coefficient for each cell, rather than only the single average Move $\times$ Decline effect a single-IV linear specification could deliver.

**What weak-instrument-robust inference does and does not certify.** The joint 3-IV first stage is $F = 10.12$ (Table OA.22): strong enough for the point estimates, not strong enough for the usual confidence intervals. Weak instruments cause two distinct problems — bias in the coefficients, and under-coverage of the $t$-intervals — and the two have different strength thresholds. We state what each half means.

*The point estimates are secure.* $F$ clears the Stock–Yogo 10%-maximal-bias threshold ($\approx 9.1$ for three instruments), so the 2SLS cell coefficients carry little weak-instrument bias. Two further signs point the same way: instrumenting moves the coefficients *away* from WLS, whereas weak instruments would pull them *toward* WLS; and three economically distinct shocks arrive at the same coefficients (the non-rejecting Hansen $J$, Table OA.24).

*The intervals on individual cells carry a caveat.* $F$ falls short of the stricter threshold that guarantees correct interval coverage (the 10%-maximal-size threshold, $\approx 22.3$), so the reported $t$-intervals on single cell coefficients may be somewhat too narrow. The textbook remedy — the Anderson–Rubin (AR) confidence set, valid at *any* instrument strength — cannot be computed cell by cell here: inverting the AR test requires searching over all endogenous coefficients jointly, which is infeasible with six endogenous regressors and only 34 clusters.

*What AR does certify are two qualitative claims* (Table OA.25). (i) *A decline-driven response exists at any instrument strength.* The four-cell system has too many endogenous dimensions to invert, so we apply AR to Spec 3, the simplest version of the model, whose only endogenous regressor is the linear Move $\times \tilde{g}$. Its AR confidence set for the young sample is $[-0.051, -0.016]$: zero is excluded. The AR test at $\beta_0 = 0$ is, equivalently, the reduced-form test of whether the predetermined shocks move the wage outcome at all, so excluding zero says that they do — and since the shocks can reach wages only through decline (the exclusion restriction), it follows that the shocks move $g$ *and* that $g$ moves the premium: neither can be zero, however weak the first stage is. From the set we use only this one bit — zero is outside it. We do not interpret where the set lies or how wide it is: the linear slope packs into a single number a response that the four-cell design decomposes into opposite-signed cell coefficients (finding (iv), Section 6.1), so the set's location and width are properties of that mixture, not of a structural parameter. If anything, the mixture makes the test conservative. In the four-cell design the two fast-decline cells imply a premium that rises as decline deepens, while cell I implies a slope of the opposite direction on its growth side; a single linear slope averages the two, so the cancellation pushes the reduced form toward zero and the test toward non-rejection. The young sample rejects despite it. (Full rural, where the premium is further diluted by older movers, does not exclude zero: $[-0.042, +0.013]$; its large side alone does:

[−0.066, −0.010].)

(ii) *The four-cell structure is needed.* Force instead a single linear-in-$g$ slope on each size side: for the young sample the AR set is *empty* — no pair of slopes is consistent with the instruments at 95% . Yet the same instruments, applied to the four-cell piecewise model, pass the Hansen $J$ test — they agree with one another once the model allows four separate slopes. The rejection therefore falls on the linear model, not on the instruments. These two claims hold regardless of instrument strength. The exact magnitude of each cell coefficient is a different kind of statement: it comes from the over-identified 2SLS point estimates, supported by the three signs above — the bias threshold is cleared, instrumenting moves the coefficients away from WLS, and the three channels agree.

**Table OA.25:** Anderson–Rubin weak-instrument-robust 95% confidence sets.

| | $\leq 29$ | | Full rural | |
|---|---|---|---|---|
| | 2SLS | AR 95% CI | 2SLS | AR 95% CI |
| Move × $\tilde{g}$ (linear) | −0.031 | [−0.051, −0.016] | −0.012 | [−0.042, +0.013] |
| Move × $S^{+}\tilde{g}$ (large side) | +0.017 | empty$^{\dagger}$ | −0.029 | [−0.066, −0.010] |
| Move × $S^{-}\tilde{g}$ (small side) | +0.081 | empty$^{\dagger}$ | −0.019 | [−0.132, +0.093] |

*Notes:* Anderson–Rubin (AR) confidence sets, valid however weak the instruments: the reported interval is the set of coefficient values the instruments do not reject at 95% (cluster-robust, clustered by origin prefecture, $G = 34$). WLS with JPSED longitudinal weights, demographic controls, and origin and wave fixed effects; instruments demeaned at the rural mean and interacted with Move (and with $S^{\pm}$ for the size-side split). With two endogenous terms the set is a region in the plane of the two coefficients; each row reports the range it spans for one coefficient. $^{\dagger}$Empty region: no coefficient pair is accepted at 95% — the single-slope (continuous-in-$g$) aggregation is rejected for the young sample (see text).

**Robustness: controlling for 1985 prefecture productivity.** The exclusion concern for the Bartik, discussed above, is that the high-fertility prefectures it loads on are the rural, agricultural ones, whose persistent rural character could depress current wages independently of decline-induced consolidation. We test it directly: if the Bartik merely proxies this persistent rural backwardness, controlling for how productive the prefecture already was in 1985 should destroy the result, whereas if it identifies decline-driven variation the result should survive. We therefore add a control for the prefecture's 1985 output level: $\log Y_{o,1985}$, the log of real value-added summed over all industries (R-JIP 2017, deflated to 2000 prices), demeaned like the other origin variables. It enters through its Move interaction, Move × $\log Y_{o,1985}$ (the level is absorbed by the origin fixed effects), as an exogenous control in both the structural equation and the first stage. The control is predetermined relative to the 2016–2025 wage panel (1985 is 31 years before the first wage-growth pair) and absorbs the time-invariant prefecture productivity scale; if the cell coefficients remain significant and of the same sign, the Bartik is identifying decline-induced wage variation that persistent productivity cannot account for.

At the $\leq 29$ headline the framework cells are essentially unchanged under the control — if anything slightly larger (cell II −0.461 → −0.521, cell IV +0.622 → +0.694, both significant at 1%; Hansen $J$ $p = 0.18$) — so the Bartik is not merely proxying persistent productivity. In the

**Table OA.26:** 4-way piecewise cell coefficients (2SLS, 18 instruments) with vs. without the 1985 productivity control.

| | Full rural | | $\leq 29$ | |
|---|---|---|---|---|
| | 2SLS | $+\log Y_{1985}$ | 2SLS | $+\log Y_{1985}$ |
| Cell I ($\beta_{Sg}^{++}$) | $+0.288^{***}$ | $+0.245$ | $+0.910^{***}$ | $+1.020^{***}$ |
| | (0.112) | (0.193) | (0.197) | (0.276) |
| Cell II ($\beta_{Sg}^{+-}$) | $-0.200^{***}$ | $-0.181^{*}$ | $-0.461^{***}$ | $-0.521^{***}$ |
| | (0.058) | (0.101) | (0.131) | (0.185) |
| Cell III ($\beta_{Sg}^{-+}$) | $-0.156^{*}$ | $-0.136$ | $-0.321^{*}$ | $-0.397^{*}$ |
| | (0.092) | (0.128) | (0.168) | (0.234) |
| Cell IV ($\beta_{Sg}^{--}$) | $+0.138^{**}$ | $+0.126$ | $+0.622^{***}$ | $+0.694^{***}$ |
| | (0.054) | (0.098) | (0.171) | (0.251) |
| Hansen $J$ $p$-value (df $= 12$) | 0.332 | 0.271 | 0.151 | 0.183 |

*Notes:* The 2SLS columns reproduce Table OA.24; the $+\log Y_{1985}$ columns add Move $\times \log Y_{o,1985}$ (demeaned; 1985 real value-added, all industries, 2000 prices; R-JIP 2017) as an exogenous control in both the wage equation and the first stage. Standard errors in parentheses are score-based wild cluster bootstrap ($B = 999$, Rademacher) at the origin prefecture. $^{***}$, $^{**}$, $^{*}$: significance at the 1%, 5%, 10% level.

full rural sample the control mainly widens the standard errors, leaving every sign unchanged but reducing significance; the framework-consolidation interpretation rests on the young sample, where the premium is concentrated and the cells are unaffected.

## OA.15 Detailed robustness checks: destination, origin, window, and outlier treatment

This appendix reports the robustness analysis summarized in Section 7: the destination placebo (R1), the cell coefficients under alternative Tokyo and origin definitions (R2–R5), decline windows (R6), and outlier treatment (R7).

**R1: Destination placebo (all four major MAs).** The baseline defines the move treatment as a rural-origin worker arriving in the *Tokyo* MA, with movers to the other three major MAs (Osaka, Nagoya, Fukuoka) dropped from the sample. This restriction implements the terminal-destination argument of Section 5.3: Tokyo is the one destination hosting the full industry range, so whatever industry a worker has lost, Tokyo has it — moves there share the single decline-driven motive, while moves to intermediate destinations mix in idiosyncratic ones. As a direct placebo we re-estimate both the WLS Spec 5 and the headline triple-IV with the treatment broadened so that Move $= 1$ for a rural-origin arrival in *any* of the four major MAs (Tokyo, Osaka, Nagoya, or Fukuoka), retaining rather than dropping the Osaka/Nagoya/Fukuoka movers and leaving every other ingredient of the specification (origin set, cell split, instruments, weights, window, trim) unchanged. Broadening the treatment nearly doubles the $\leq 29$ mover count, from 95 to 181.

**Table OA.27:** Destination placebo (R1): Tokyo-only treatment versus a treatment broadened to all four major MAs.

| | Tokyo MA (headline) | All four MAs (placebo) |
|---|---|---|
| *Panel A. WLS (Spec 5, $\leq 29$)* | | |
| Cell I ($\beta_{Sg}^{++}$, large $\times$ slow decline) | $+0.615^{**}$ | $+0.133$ |
| | (0.265) | (0.329) |
| Cell II ($\beta_{Sg}^{+-}$, large $\times$ fast decline) | $-0.336^{**}$ | $-0.021$ |
| | (0.156) | (0.155) |
| Cell III ($\beta_{Sg}^{-+}$, small $\times$ slow decline) | $-0.180$ | $-0.126$ |
| | (0.181) | (0.156) |
| Cell IV ($\beta_{Sg}^{--}$, small $\times$ fast decline) | $+0.503^{***}$ | $+0.267$ |
| | (0.155) | (0.171) |
| *Panel B. 2SLS (18 IV: demo + TFR + K-deep, $\leq 29$)* | | |
| Cell I ($\beta_{Sg}^{++}$, large $\times$ slow decline) | $+0.910^{***}$ | $+0.253$ |
| | (0.197) | (0.363) |
| Cell II ($\beta_{Sg}^{+-}$, large $\times$ fast decline) | $-0.461^{***}$ | $-0.065$ |
| | (0.131) | (0.158) |
| Cell III ($\beta_{Sg}^{-+}$, small $\times$ slow decline) | $-0.321^{*}$ | $-0.315$ |
| | (0.168) | (0.190) |
| Cell IV ($\beta_{Sg}^{--}$, small $\times$ fast decline) | $+0.622^{***}$ | $+0.431^{*}$ |
| | (0.171) | (0.207) |
| Hansen $J$ (df = 12) | 16.96 | 27.81 |
| $p$-value | 0.15 | **0.006** |
| $N$ (person-wave pairs, WLS) | 4,617 | 4,703 |
| Movers | 95 | 181 |

*Notes:* Headline baseline: $\leq 29$ sub-sample, origin fixed effects, real housing-inclusive wage, drop-imputation sample, $S$ = total population 2010, $g$ = young 15–29 decline 2010–2019, region-39 cell split. The two columns differ *only* in the destination treatment set: "Tokyo MA" is the headline (movers to Osaka/Nagoya/Fukuoka dropped); "All four MAs" marks Move = 1 for a rural-origin arrival in any of the four major MAs (those movers retained). Panel A is the 4-way piecewise WLS (Spec 5); Panel B is the over-identified 2SLS (18 instruments, as in Table OA.24). Standard errors in parentheses are score-based wild cluster bootstrap ($B = 999$, Rademacher) at the origin prefecture. The 2SLS sample is marginally smaller than the WLS one because the TFR instrument is missing for Okinawa. $^{***}/^{**}/^{*}$ significant at 1/5/10%.

The result is a clean placebo. Every cell keeps its framework-predicted sign under the broadened treatment, so nothing flips; but the signal hollows out. Under WLS all four cells become statistically insignificant and their magnitudes contract toward zero (cell I $+0.615^{**} \rightarrow +0.133$; cell II $-0.336^{**} \rightarrow -0.021$; cell III $-0.180 \rightarrow -0.126$; cell IV $+0.503^{***} \rightarrow +0.267$). Under 2SLS the two large-origin cells hollow out the same way (cell I $+0.910^{***} \rightarrow +0.253$, n.s.; cell II $-0.461^{***} \rightarrow -0.065$, n.s.), while the small-origin cells degrade less — cell III barely moves ($-0.321^{*} \rightarrow -0.315$) and cell IV weakens but keeps marginal significance ($+0.622^{***} \rightarrow +0.431^{*}$). The decisive diagnostic is the over-identification test: the headline triple-IV comfortably passes ($J = 16.96$, df $= 12$, $p = 0.15$), whereas the broadened-destination version *rejects* ($J = 27.81$, $p = 0.006$).

The economics of the rejection is the point of the placebo. The framework predicts a wage premium for workers who move because their industry has consolidated upward to the apex under sustained origin decline; the predetermined demographic shocks instrument exactly that decline-driven push toward the destination that hosts the full industry range — Tokyo. The premium is an *apex* premium: when specialized industries leave a large declining origin, they consolidate all the way to Tokyo, not to a mid-tier regional core. A large-origin worker who instead stops at an intermediate-tier MA — Osaka or Nagoya rather than Tokyo — has therefore generally not gone where his industry went, so his move is not the decline-driven, follow-your-industry path to the apex. A substantial share of such intermediate-tier moves are instead driven by idiosyncratic, decline-orthogonal motives (sector-specific job matches, family ties, pre-existing regional networks). Folding these movers into the treatment mixes a partly decline-unrelated population into a treatment the instruments were built to track for the Tokyo channel alone. The instruments then no longer satisfy the exclusion restriction for every mover in this mixed treatment group, and the over-identifying restrictions are rejected. The Hansen $J$ rejection is thus not a failure of the instruments but evidence that the instruments are doing precisely what the framework requires—isolating the decline-driven moves toward the apex, the destination holding every industry a worker may have lost — and that they correctly detect the contamination introduced by the intermediate-tier destinations.

The asymmetry in how the cells degrade is also informative. At large origins (cells I and II) the premium collapses essentially to zero under the broadened treatment in both WLS and 2SLS: the signal there is a destination-specific consolidation premium that exists only at the apex, so adding non-apex destinations dilutes it directly. Cell IV (small origins, concavity-on-jobs) degrades least and even retains marginal significance in 2SLS, for a reason specific to small origins. Because the origin is small, *any* destination larger than it already hosts the ubiquitous necessity industries it has lost; a worker leaving a small fast-declining origin can therefore recover what the origin lost by moving to Osaka, Nagoya, or Fukuoka just as to Tokyo, since all four lie above the origin in the hierarchy and host what it can no longer support. The cell-IV movers added by the broadened treatment are thus largely on-mechanism rather than contaminating, so part of the cell-IV signal survives the broadened treatment and still loads on the predetermined decline shocks. The placebo therefore supports both the apex restriction used for identification and the framework's distinction between the two origin types: at large origins the lost industries survive only at the apex, so the premium is destination-specific; at small origins the lost necessities are hosted by any larger destination, so it is not — the premium can be realized at any larger destination.

**Table OA.28:** Robustness results: the four cell coefficients and the marginal effects of origin size.

| Specification | $\beta^{++}_{Sg}$ (cell I) | $\beta^{+-}_{Sg}$ (cell II) | $\beta^{-+}_{Sg}$ (cell III) | $\beta^{--}_{Sg}$ (cell IV) | $\widehat{\mathrm{ME}}(S)_{II}$ at $g=-\sigma_g$ | $\widehat{\mathrm{ME}}(S)_{IV}$ at $g=-\sigma_g$ | N | N (movers) |
|---|---|---|---|---|---|---|---|---|
| *Panel A: Spec 4 cell coefficients, full rural sample:* | | | | | | | | |
| Main | +0.220** | −0.172*** | −0.145 | +0.140** | | | 47,536 | 358 |
| | (0.107) | (0.054) | (0.105) | (0.059) | | | | |
| R2 destination = Tokyo Prefecture | −0.297* | −0.138* | +0.035 | +0.119 | | | 47,536 | 152 |
| | (0.155) | (0.079) | (0.139) | (0.113) | | | | |
| R3 Fukuoka reclassified as rural origin | +0.128 | −0.144*** | −0.133 | +0.051 | | | 51,914 | 350 |
| | (0.089) | (0.046) | (0.092) | (0.062) | | | | |
| R5 origins = all non-Tokyo-MA prefectures | +0.048 | −0.064*** | −0.120 | +0.042 | | | 77,832 | 503 |
| | (0.039) | (0.024) | (0.079) | (0.058) | | | | |
| R6a $g$ window 2010→2025, panel fixed | +0.154* | −0.118** | −0.027 | +0.028 | | | 47,536 | 358 |
| | (0.082) | (0.047) | (0.070) | (0.061) | | | | |
| R6b $g$ window 2010→2025, panel + 2023/24 | +0.166*** | −0.102*** | −0.077 | +0.027 | | | 59,819 | 439 |
| | (0.063) | (0.038) | (0.054) | (0.066) | | | | |
| R7a winsorised at 1%/99% (caps, not drops) | +0.103 | −0.164* | −0.118 | +0.228** | | | 48,573 | 370 |
| | (0.170) | (0.090) | (0.134) | (0.100) | | | | |
| R7b winsorised at 5%/95% | +0.058 | −0.075** | −0.063 | +0.062 | | | 48,573 | 370 |
| | (0.076) | (0.035) | (0.065) | (0.041) | | | | |
| *Panel B: Spec 5 cell coefficients and marginal effects, ≤ 29 sub-sample:* | | | | | | | | |
| Main | +0.615** | −0.336** | −0.180 | +0.503*** | +0.216 | −1.110*** | 4,617 | 95 |
| | (0.265) | (0.156) | (0.181) | (0.155) | (0.179) | (0.267) | | |
| R2 destination = Tokyo Prefecture | +0.107 | −0.398 | +0.028 | +0.922*** | +0.231 | −2.064*** | 4,617 | 42 |
| | (0.454) | (0.257) | (0.187) | (0.228) | (0.378) | (0.470) | | |
| R3 Fukuoka reclassified as rural origin | +0.699** | −0.289* | −0.243 | +0.342*** | +0.130 | −0.953*** | 4,983 | 91 |
| | (0.279) | (0.156) | (0.159) | (0.125) | (0.204) | (0.209) | | |
| R5 origins = all non-Tokyo-MA prefectures | +0.136* | −0.155** | −0.081 | +0.322*** | +0.347*** | −0.993*** | 7,565 | 136 |
| | (0.080) | (0.064) | (0.116) | (0.098) | (0.075) | (0.199) | | |
| R6a $g$ window 2010→2025, panel fixed | +0.499** | −0.241 | −0.090 | +0.202 | +0.144 | −0.327 | 4,617 | 95 |
| | (0.250) | (0.148) | (0.192) | (0.132) | (0.166) | (0.289) | | |
| R6b $g$ window 2010→2025, panel + 2023/24 | +0.567*** | −0.312*** | −0.181 | +0.148 | +0.326*** | −0.328 | 5,657 | 128 |
| | (0.189) | (0.102) | (0.153) | (0.092) | (0.118) | (0.293) | | |
| R7a winsorised at 1%/99% (caps, not drops) | +0.802** | −0.689*** | −0.187 | +0.846*** | +0.979** | −2.081** | 4,767 | 100 |
| | (0.392) | (0.242) | (0.226) | (0.319) | (0.446) | (0.885) | | |
| R7b winsorised at 5%/95% | +0.387** | −0.273*** | −0.086 | +0.325*** | +0.291* | −0.930*** | 4,767 | 100 |
| | (0.190) | (0.105) | (0.122) | (0.123) | (0.174) | (0.320) | | |

*Notes:* Each row varies one ingredient relative to the Main baseline (COVID-clean pre-COVID panel, 1% trim, 2010–2019 decline window, 39-region demeaned cells, origin and wave fixed effects, JPSED weights). Panel A reports the Spec 4 cell coefficients (all four cells) on the full rural sample; Panel B reports the same coefficients on the ≤ 29 sub-sample (Spec 5) together with the Spec-5 marginal effects of origin size at fast decline. The $N$ and $N$ *(movers)* columns give each panel's total sample size and its number of Tokyo-MA movers. R2: destination narrowed to Tokyo Prefecture. R3: Fukuoka reclassified as a rural origin. R5: origin set expanded to all non-Tokyo-MA prefectures (per-prefecture origin coordinates demeaned at the fixed 39-region centers; moves from the added prefectures to Tokyo MA count as moves). R6a: 2010–2025 decline window with the wage panel fixed. R6b: 2010–2025 window with the panel also extended by the COVID-clean 2023/24 transition. R7a/R7b: winsorisation — wage growth is capped at the stated percentiles instead of trimmed, so all observations are kept (the baseline 1%/99% trim, by contrast, drops the tails; Appendix OA.7). Standard errors in parentheses are wild cluster bootstrap ($B = 999$, Rademacher) at origin prefecture; ***/**/* significant at 1/5/10%.

**R2: Tokyo destination definition.** Restricting the destination to Tokyo Prefecture alone, rather than the full four-prefecture Tokyo MA, drops the mover count from 358 to 152 in the full sample (and from 95 to 42 at ≤ 29). The consolidation footprint spans the full MA — many consolidated establishments locate in Kanagawa, Saitama, and Chiba — so the single-prefecture destination discards part of the premium and more than halves the treatment group; the check asks what survives (Table OA.28, rows R2). At $\leq 29$, the headline, the small-fast signal does: cell IV stays at $+0.922^{***}$ with $\widehat{\mathrm{ME}}(S)_{IV} = -2.064^{***}$. The large side loses significance along with its movers (cell II $-0.398$, n.s.; cell I $+0.615^{**} \rightarrow +0.107$, null): with 42 movers left, the large cells — which even at baseline rest on a handful of movers and single high-leverage origins (finding (ii)) — lose their identification, not their signs. In the full sample cell II holds at $-0.138^{*}$ while cell I turns weakly negative ($-0.297^{*}$), plausibly reflecting the older movers a Tokyo-Prefecture-only destination picks up — employer-directed transfers rather than hierarchy-driven migration. Cell III is null throughout.

**R3: Fukuoka reclassified as rural origin.** Re-classifying Fukuoka as a rural origin on the baseline COVID-clean panel ($N$ = 51,914, 350 movers) leaves the estimates essentially unchanged — in particular in the $\leq 29$ panel (Table OA.28, rows R3). The cells are therefore not driven by whether Fukuoka is classified as the lowest-tier major MA or as a regional core.

**R4: Miyagi reclassified as a metropolitan area.** Reclassifying Miyagi (Sendai) as a metropolitan center is, operationally, the deletion of Miyagi from the rural origin set (the destination stays the Tokyo MA) — and that deletion is already reported in the single-prefecture leave-one-out analysis: dropping Miyagi leaves the cells essentially unchanged ($\leq 29$: $+0.591^{**}$, $-0.385^{**}$, $-0.196$, $+0.675^{**}$; Appendix OA.11, Table OA.15). We therefore do not repeat it as a separate table row.

**R5: origins expanded to all non-Tokyo-MA prefectures.** R5 simply adds the remaining non-Tokyo-MA prefectures — the Osaka and Nagoya MA prefectures and Fukuoka — to the origin set, keeping everything else fixed: origin coordinates remain per-prefecture values demeaned at the baseline 39-region centers (no re-demeaning), and arrivals in the Tokyo MA from the added prefectures count as Move = 1 ($N$ = 77,832, 503 movers; 7,565 and 136 at $\leq 29$). The cell pattern survives (Table OA.28, rows R5): at $\leq 29$ the four cells read $+0.136^{*}$, $-0.155^{**}$, $-0.081$, $+0.322^{***}$, with the two framework gradients intact ($\widehat{\mathrm{ME}}(S)$: large side $+0.347^{***}$, small side $-0.993^{***}$). Magnitudes attenuate relative to the rural-only baseline, as expected. The origins R5 adds are themselves major metros (Osaka, Nagoya, Fukuoka), and their residents move to Tokyo mostly through ordinary present-day metro-to-metro migration — career and corporate transfers, the pull of Tokyo's economy — not the decline-driven upward consolidation the framework describes. Pooling these off-mechanism origins with the declining rural ones dilutes the cell coefficients, so the pattern holds but the magnitudes shrink.

**R6: Decline window.** The baseline measures origin decline over 2010–2019, the long pre-COVID structural window. R6a extends the window to 2010–2025 while holding the wage panel fixed; R6b also extends the panel with the COVID-clean 2023/24 transition ($N$ = 59,819, 439 movers). The cells keep their signs, with significance weakening mainly on the small-origin side (Table OA.28, rows R6a–R6b). The reason lies outside the mechanism: Tokyo-bound migration collapsed in 2020 and has not recovered to its pre-pandemic level (Section 4), so the post-2019 years add a period in which the migration response was suspended, diluting the variation the mechanism operates on.

**R7: Outlier treatment (winsorisation).** The premium lives in the tails of the wage-growth distribution: a displaced mover's wage jump is large. Trimming deletes exactly these observations, so it mechanically weakens the cells — and the upper-tail cut falls hardest on the deep-decline origins, whose movers jump most, so a heavy trim does not prune noise evenly — it deletes precisely the observations that carry the effect. The appropriate sensitivity keeps the observations and caps the values (winsorisation). Winsorised at the baseline 1%/99% cutoffs, the cells strengthen relative to the trimmed baseline (Table OA.28, row R7a) — the baseline trim, which deletes a handful of signal-bearing movers, is the conservative choice. Even winsorised at $\pm 0.7$ log points (the 5%/95% cutoffs), the $\leq 29$ result is unchanged (row R7b): all four cells keep their signs, with cell I ($+0.387^{**}$), cell II ($-0.273^{***}$), and cell IV ($+0.325^{***}$) significant and cell III null ($-0.086$) as in the baseline. The signal is not an outlier artifact: any treatment of the tails that retains the observations reproduces the result.

## OA.16 Wage premium by age band

The main estimation reports the full rural sample and the $\leq 29$ cohort (Section 6.1). Here we additionally restrict to workers aged 30–64 — the sample excluding the young — and re-run the 4-way piecewise specification. Table OA.29 reports the three age bands side by side under the baseline pre-COVID decline window (2010–2019), with the same controls, weights, and wild-cluster-bootstrap inference as the main estimation. (The alternative 2010–2025 decline window is part of the robustness set, R6; Online Appendix OA.15.)

At 30–64 essentially nothing of the young structure survives. The small-origin side disappears outright — the small×fast-decline interaction is $+0.020$ against $+0.482^{***}$ at 20–29, and the implied $\widehat{\mathrm{ME}}(S)$ at fast decline is wrong-signed and null ($+0.31$, $t = 1.2$) — while the large-origin side retains only a weakened version of the 20–29 coefficient ($-0.109^{**}$, a third of its magnitude). The reading is composition, not mechanism: inter-prefectural migration concentrates in the 20s (Section 2), the 30–64 Tokyo movers are proportionally a third as frequent (255 of 38,999 versus 95 of 4,617), and the migration wage premium is absent where migration itself is rare.

## OA.17 Education margin: Move × University-or-higher

This appendix adds Move × Univ-or-higher to the baseline specifications — the $\leq 29$ sub-sample (Spec 5) and the full-rural sample (Spec 4) — as referenced in Section 6.1. Table OA.30 reports the coefficients side-by-side with the corresponding baseline specifications.

**Result.** The education interaction is positive but imprecise: Move × Univ = $+0.104$ (SE 0.131) at $\leq 29$ (Spec 5U) and $+0.103$ (SE 0.072) in the full rural sample (Spec 4U). And its inclusion

**Table OA.29:** Wage premium 4-way piecewise interaction and marginal effects, by age band.

| | Full rural | 20–29 | 30–64 |
|---|---|---|---|
| Move $\times \log \tilde{S}^{+}$ (large) | $-0.597^{**}$ | $-1.072$ | $-0.311$ |
| | (0.291) | (0.758) | (0.236) |
| Move $\times \log \tilde{S}^{-}$ (small) | $+0.497^{***}$ | $+0.821$ | $+0.387$ |
| | (0.192) | (0.530) | (0.306) |
| Move $\times g^{+}$ (slow decl) | $-0.066^{**}$ | $-0.125$ | $-0.057^{**}$ |
| | (0.032) | (0.092) | (0.023) |
| Move $\times g^{-}$ (fast decl) | $+0.083^{***}$ | $+0.184^{**}$ | $+0.046^{**}$ |
| | (0.028) | (0.084) | (0.022) |
| Cell I: $\log \tilde{S}^{+} \times g^{+}$ (large $\times$ slow) | $+0.223^{**}$ | $+0.587^{**}$ | $+0.278^{**}$ |
| | (0.109) | (0.271) | (0.126) |
| Cell II: $\log \tilde{S}^{+} \times g^{-}$ (large $\times$ fast) | $-0.178^{***}$ | $-0.321^{**}$ | $-0.109^{**}$ |
| | (0.056) | (0.156) | (0.044) |
| Cell III: $\log \tilde{S}^{-} \times g^{+}$ (small $\times$ slow) | $-0.147$ | $-0.151$ | $-0.201$ |
| | (0.102) | (0.183) | (0.140) |
| Cell IV: $\log \tilde{S}^{-} \times g^{-}$ (small $\times$ fast) | $+0.145^{**}$ | $+0.482^{***}$ | $+0.020$ |
| | (0.062) | (0.157) | (0.095) |
| ME($S$) at cell IV (fast decl, predicted $< 0$) | $-0.075$ | $-1.060^{***}$ | $+0.308$ |
| | (0.193) | (0.310) | (0.248) |
| ME($S$) at cell II (fast decl, predicted $> 0$) | $+0.107$ | $+0.180$ | $+0.119$ |
| | (0.084) | (0.190) | (0.100) |
| $N$ | 47,536 | 4,617 | 38,999 |
| Moves to Tokyo MA | 358 | 95 | 255 |

*Notes:* The wage-growth sample is the rural-origin COVID-clean consecutive-wave transitions (2016/17, 2017/18, 2018/19, 2023/24, 2024/25), with origin decline measured over 2010–2019. Each column is a separate WLS regression on the COVID-clean wage-growth sample (rural origins), with longitudinal pair weights and score-based wild cluster bootstrap standard errors ($B = 999$, Rademacher) at the origin prefecture. The destination is the Tokyo MA. $^{***}$, $^{**}$, $^{*}$ denote significance at the 1%, 5%, 10% level.

**Table OA.30:** Wage-growth regression with Move × Univ-or-higher added

| | $\leq$ 29 sub-sample | | Full rural | |
|---|---|---|---|---|
| | Spec 5 (baseline) | Spec 5U (+ Move×Univ) | Spec 4 (baseline) | Spec 4U (+ Move×Univ) |
| Moved | +0.325 | +0.277 | +0.025 | -0.016 |
| | (0.328) | (0.307) | (0.110) | (0.115) |
| Moved × Young | | | +0.156* | +0.142* |
| | | | (0.088) | (0.082) |
| Moved × Univ | | +0.104 | | +0.103 |
| | | (0.131) | | (0.072) |
| Moved × $\text{Size}^+$ | -1.094 | -1.170 | -0.547** | -0.587** |
| | (0.755) | (0.784) | (0.276) | (0.293) |
| Moved × $\text{Size}^-$ | +0.854 | +0.879* | +0.486*** | +0.498*** |
| | (0.538) | (0.525) | (0.186) | (0.189) |
| Moved × $\text{Decl}^+$ | -0.130 | -0.136 | -0.063* | -0.067** |
| | (0.092) | (0.091) | (0.033) | (0.034) |
| Moved × $\text{Decl}^-$ | +0.180** | +0.177** | +0.081*** | +0.082*** |
| | (0.083) | (0.080) | (0.027) | (0.028) |
| Moved × $\text{Size}^+$ × $\text{Decl}^+$ (I) | +0.615** | +0.640** | +0.220** | +0.243** |
| | (0.265) | (0.267) | (0.107) | (0.122) |
| Moved × $\text{Size}^+$ × $\text{Decl}^-$ (II) | -0.336** | -0.339** | -0.172*** | -0.177*** |
| | (0.156) | (0.156) | (0.054) | (0.057) |
| Moved × $\text{Size}^-$ × $\text{Decl}^+$ (III) | -0.180 | -0.200 | -0.145 | -0.162 |
| | (0.181) | (0.173) | (0.105) | (0.108) |
| Moved × $\text{Size}^-$ × $\text{Decl}^-$ (IV) | +0.503*** | +0.502*** | +0.140** | +0.141** |
| | (0.155) | (0.151) | (0.059) | (0.060) |
| Year FE | Yes | Yes | Yes | Yes |
| Demographic ctrls | Yes | Yes | Yes | Yes |
| JPSED weight | Yes | Yes | Yes | Yes |
| N | 4,617 | 4,617 | 47,536 | 47,536 |

*Notes:* Each column is a separate WLS regression on the COVID-clean wage-growth sample (rural origins; transitions 2016/17–2018/19 + 2023/24–2024/25), with longitudinal pair weights (substituting wave-$t$ cross-sectional weights for pairs without a panel weight) and score-based wild cluster bootstrap standard errors ($B$ = 999, Rademacher) at the origin prefecture. Univ is the indicator for university-or-higher education. Spec 4 and Spec 5 reproduce the corresponding columns of Table 5; Spec 4U and Spec 5U add the Move × Univ interaction. ***, **, * denote significance at the 1%, 5%, 10% level.

leaves the cell coefficients essentially unchanged: at $\leq 29$, cell I moves from $+0.615$ to $+0.640$, cell II from $-0.336$ to $-0.339$, cell III from $-0.180$ to $-0.200$, and cell IV from $+0.503$ to $+0.502$, with the full-rural columns moving equally little. The positive sign on Move $\times$ Univ — university movers gaining somewhat more — is the expected direction, matching the higher return to skill in the thicker labor markets of large cities (Combes et al., 2008; Dauth et al., 2022). But any premium difference between university and non-university movers is absorbed by this Move $\times$ Univ interaction term and leaves the cell-level estimates intact, so the size $\times$ decline structure is not an education-composition effect in disguise (main text finding (ii), Section 6.1).

## OA.18 Robustness: removing family-driven moves

The framework's prediction covers any move toward the apex, whether the worker chose to leave or was pushed out, so voluntary and involuntary moves do not need to be separated. The potential confounder is the small fraction of moves driven by *family events independent of the labor-market mechanism*—marriage, pregnancy, childbirth, childcare, caregiving, spouse transfer, or family-business succession. We test that this family-driven noise does not drive the baseline estimates.

**Family-driven mover identification.** JPSED asks every worker who resigned in the past year to indicate the reason as a multi-answer item. We define a *family-driven mover* as a rural-to-Tokyo MA migrant whose stated resignation reasons include at least one purely family-driven reason (marriage, pregnancy, childbirth, childcare, caregiving, spouse transfer, family business) and *none* of the labor-market-related reasons. Of the 358 rural-to-Tokyo MA wage-sample movers, 18 (5.0%) qualify; the corresponding count for the 20–29 sub-sample is 5 of 95 (5.3%). The *noise-removed* move indicator drops these family-driven movers from Move = 1 and reclassifies them as stayers; all other movers (career-driven quits, involuntary displacements, and movers with no resignation in the past year) remain Move = 1.

**Result.** Tables OA.31 and OA.32 report the 4-way piecewise specification (baseline 2010–2019 decline window) under the baseline and noise-removed move definitions, for the 20–29 sub-sample and the full rural sample respectively. The noise-removed estimates stay close to the baseline, which here reproduces Specs 4–5 of Table 5. In the $\leq 29$ headline every cell keeps its sign and both framework cells stay significant: cell I $+0.615 \to +0.631$, cell II $-0.336 \to -0.285$, cell III $-0.180 \to -0.157$, cell IV $+0.503 \to +0.491$ (Table OA.31). In the full-rural sample (Table OA.32) the shifts are larger: every cell keeps its sign (cell I $+0.220 \to +0.397$, cell II $-0.172 \to -0.128$, cell III $-0.145 \to -0.188$, cell IV $+0.140 \to +0.087$), though cell IV loses significance — consistent with the premium's concentration in the young, whose headline above is unaffected. The cell structure is therefore robust to removing the small fraction of moves driven by family events independent of the labor-market mechanism, and we report baseline figures throughout the main paper.

**Table OA.31:** 4-way piecewise regression, 20–29 sub-sample: baseline versus noise-removed move definitions.

| Subset | Cell I (large×slow) | Cell II (large×fast) | Cell III (small×slow) | Cell IV (small×fast) | N (pairs) | N (movers) |
|---|---|---|---|---|---|---|
| Baseline | $+0.615^{**}$ | $-0.336^{**}$ | $-0.180$ | $+0.503^{***}$ | 4,617 | 95 |
| | (0.265) | (0.156) | (0.181) | (0.155) | | |
| Noise-removed | $+0.631^{**}$ | $-0.285^{*}$ | $-0.157$ | $+0.491^{***}$ | 4,617 | 90 |
| | (0.282) | (0.160) | (0.179) | (0.153) | | |

*Notes:* Each row reports the same WLS 4-way piecewise regression (origin decline measured over 2010–2019) with score-based wild cluster bootstrap standard errors at the origin prefecture ($B = 999$ Rademacher replications); the rows differ only in the set of movers treated as Move = 1. *Noise-removed* drops the 5 family-driven movers from the 95 baseline movers, retaining 90. Standard errors in parentheses; $^{***}$, $^{**}$, $^{*}$ denote significance at the 1%, 5%, 10% level. Companion full-rural sample in Table OA.32.

**Table OA.32:** 4-way piecewise regression, full rural sample: baseline versus noise-removed.

| Subset | Cell I (large×slow) | Cell II (large×fast) | Cell III (small×slow) | Cell IV (small×fast) | N (pairs) | N (movers) |
|---|---|---|---|---|---|---|
| Baseline | $+0.220^{**}$ | $-0.172^{***}$ | $-0.145$ | $+0.140^{**}$ | 47,536 | 358 |
| | (0.107) | (0.054) | (0.105) | (0.059) | | |
| Noise-removed | $+0.397^{**}$ | $-0.128^{*}$ | $-0.188^{*}$ | $+0.087$ | 47,536 | 340 |
| | (0.170) | (0.066) | (0.110) | (0.056) | | |

*Notes:* Same WLS specification, weighting, and inference as Table OA.31; sample is the full rural cohort (all ages). *Noise-removed* drops the 18 family-driven movers from the 358 baseline movers of the regression sample, retaining 340. Standard errors in parentheses; $^{***}$, $^{**}$, $^{*}$ denote significance at the 1%, 5%, 10% level.

## OA.19 Weighting sensitivity: unweighted OLS

The main paper estimates the wage-premium specifications by weighted least squares with JPSED longitudinal pair weights. The object of interest is a population-level statement about Japan's wage-growth panel, and the longitudinal pair weight is the design-correct weighting for it: the Recruit Works Institute calibrates these weights so that the weighted sample reproduces the sex $\times$ age $\times$ employment-status composition of the Japanese working-age population in the Statistics Bureau's Labour Force Survey (*Rōdoryoku Chōsa*); the pair weight further corrects for panel drop-out, up-weighting the respondents who stay across both waves so they stand in for those who leave. The WLS coefficients therefore describe what the cell-level wage premia look like in that population. Following the diagnostic recommendation of Solon et al. (2015), we report the corresponding unweighted OLS estimates as a sensitivity to read the WLS–OLS comparison as a diagnostic for treatment-effect heterogeneity across the weighting distribution. Table OA.33 reports the unweighted OLS estimates (Spec $4'$ = full rural, Spec $5' = \leq 29$) alongside the corresponding WLS Specs 4–5 of Table 5; this appendix discusses the WLS–OLS comparison directly.

**Reading the WLS–OLS comparison.** The cell pattern is robust to dropping the weights: every cell keeps its sign in the unweighted columns, with moderately attenuated magnitudes ($\leq 29$: cell I $+0.615 \rightarrow +0.399$, cell II $-0.336 \rightarrow -0.246$, cell III $-0.180 \rightarrow -0.209$, cell IV $+0.503 \rightarrow +0.400$; the full-rural columns move similarly). The mild attenuation under OLS says the premia are somewhat stronger in the population-weighted composition than in the raw JPSED sample composition—a heterogeneity reading in the spirit of Solon et al. (2015), not evidence against the pattern. The unweighted estimates serve as a transparency check; they are not an alternative population-level estimate.

## OA.20 CPI deflation: deflator construction and full table

This appendix documents the deflator construction and reports the full nominal-vs-real results table (Table OA.34) for the CPI-deflation robustness summarized in Section 7. The body section gives the substantive conclusion: the cell coefficients are essentially unchanged under CPI-deflated wage growth, including the housing-inclusive deflator.

**Deflator construction.** We re-estimate Specs 4 and 5 with prefecture-by-year CPI-deflated wage growth, using $\mathrm{CPI}(p,t) = \mathrm{NationalCPI}(t) \times \mathrm{RegionalIndex}(p)/100$: $\mathrm{NationalCPI}(t)$ is the Statistics Bureau's national all-items CPI (2020 base) and $\mathrm{RegionalIndex}(p)$ is its cross-prefecture Regional Difference Index (2023 release, national = 100), the prefecture's price level relative to the national mean (Statistics Bureau, Ministry of Internal Affairs and Communications, 2024). We use two versions of the regional index, differing only in how owner-occupied housing is priced. The CPI is built on a fixed *basket*: a list of goods and services, each weighted by its share of household spending. In the *published* index, housing enters this basket only through the rents

**Table OA.33:** Weighting sensitivity: the piecewise specifications under WLS (Specs 4–5, repeated from the main text) and unweighted OLS (Specs 4′–5′).

| | Weighted (WLS) | | Unweighted (OLS) | |
|---|---|---|---|---|
| | (4) Full rural | (5) $\leq 29$ only | (4′) Full rural | (5′) $\leq 29$ only |
| Moved → MA | +0.025 | +0.325 | +0.010 | +0.194 |
| | (0.110) | (0.328) | (0.087) | (0.267) |
| Moved × Young | +0.156* | | +0.040 | |
| | (0.088) | | (0.059) | |
| Moved × Size$^{+}$ (large) | -0.547** | -1.094 | -0.379 | -0.911 |
| | (0.276) | (0.755) | (0.241) | (0.580) |
| Moved × Size$^{-}$ (small) | +0.486*** | +0.854 | +0.452*** | +0.662 |
| | (0.186) | (0.538) | (0.158) | (0.537) |
| Moved × Decline$^{+}$ (slow decl) | -0.063* | -0.130 | -0.061** | -0.111* |
| | (0.033) | (0.092) | (0.026) | (0.066) |
| Moved × Decline$^{-}$ (fast decl) | +0.081*** | +0.180** | +0.055** | +0.128* |
| | (0.027) | (0.083) | (0.024) | (0.070) |
| Moved × Size$^{+}$ × Decl$^{+}$ (cell I: large × slow decl) | +0.220** | +0.615** | +0.212** | +0.399** |
| | (0.107) | (0.265) | (0.100) | (0.182) |
| Moved × Size$^{+}$ × Decl$^{-}$ (cell II: large × fast decl) | -0.172*** | -0.336** | -0.125*** | -0.246** |
| | (0.054) | (0.156) | (0.046) | (0.122) |
| Moved × Size$^{-}$ × Decl$^{+}$ (cell III: small × slow decl) | -0.145 | -0.180 | -0.160** | -0.209 |
| | (0.105) | (0.181) | (0.076) | (0.130) |
| Moved × Size$^{-}$ × Decl$^{-}$ (cell IV: small × fast decl) | +0.140** | +0.503*** | +0.115** | +0.400*** |
| | (0.059) | (0.155) | (0.057) | (0.155) |
| Year FE | Yes | Yes | Yes | Yes |
| Demographic ctrls | Yes | Yes | Yes | Yes |
| JPSED weight | Yes | Yes | No | No |
| N | 47,536 | 4,617 | 47,536 | 4,617 |

*Notes:* Same sample, controls, fixed effects, and inference as Table 5 (score-based wild cluster bootstrap, $B = 999$, Rademacher, at origin prefecture); the only difference is weighting: the WLS columns apply the JPSED longitudinal pair weight, the OLS columns are unweighted. ***, **, *: significance at the 1%, 5%, 10% level.

that renters actually pay, and those carry a small weight; the housing costs of owner-occupiers do not appear at all. The resulting Tokyo MA–rural price gap is small ($\approx 0.031$ log points). The *housing-inclusive* index adds owner-occupied housing back with the weight it would carry in the basket: $\text{RegionalIndex}(p) + 0.16 \cdot (\text{RentIndex}(p) - 100)$, where 0.16 is the weight of imputed rent in the 2020-base basket and RentIndex is the same survey's rent sub-index (Tokyo $\approx 155$ vs. rural $\approx 70$–90). With housing priced where residents actually pay for it, the Tokyo MA–rural log price gap widens from 0.031 to $\approx 0.109$ log points (roughly 3% to 11%). The housing-inclusive index is the deflator behind the paper's baseline outcome (not one option among several).

**Table OA.34:** CPI-deflation robustness: nominal vs. prefecture-by-year real wage growth.

| | Spec 4 (Full rural) | | | Spec 5 ($\leq 29$) | | |
|---|---|---|---|---|---|---|
| | Nominal | Real (excl. IR) | Real (incl. IR) | Nominal | Real (excl. IR) | Real (incl. IR) |
| $\beta_{Sg}^{++}$ (Cell I: large × slow decl) | $+0.146^{*}$ | $+0.156^{**}$ | $+0.156^{**}$ | $+0.507^{***}$ | $+0.520^{***}$ | $+0.520^{***}$ |
| | (0.078) | (0.078) | (0.078) | (0.162) | (0.162) | (0.162) |
| $\beta_{Sg}^{+-}$ (Cell II: large × fast decl) | $-0.134^{***}$ | $-0.140^{***}$ | $-0.132^{***}$ | $-0.288^{**}$ | $-0.295^{**}$ | $-0.288^{**}$ |
| | (0.051) | (0.051) | (0.049) | (0.116) | (0.116) | (0.115) |
| $\beta_{Sg}^{-+}$ (Cell III: small × slow decl) | $-0.249$ | $-0.270$ | $-0.271$ | $-0.458^{*}$ | $-0.481^{*}$ | $-0.483^{*}$ |
| | (0.178) | (0.177) | (0.178) | (0.257) | (0.256) | (0.255) |
| $\beta_{Sg}^{--}$ (Cell IV: small × fast decl) | $+0.227^{*}$ | $+0.232^{*}$ | $+0.215^{*}$ | $+0.858^{***}$ | $+0.865^{***}$ | $+0.851^{***}$ |
| | (0.119) | (0.121) | (0.117) | (0.264) | (0.267) | (0.265) |
| $N$ | 47,855 | 47,855 | 47,855 | 4,702 | 4,702 | 4,702 |

*Notes:* "Real" columns deflate each wage observation by $\text{CPI}(p,t) = \text{NationalCPI}(t) \times \text{RegionalIndex}(p)/100$, where NationalCPI is the Statistics Bureau of Japan's national all-items CPI (2020 base) and RegionalIndex is the cross-prefecture Regional Difference Index (2023 release; national = 100). "excl. IR" (excluding owner-occupier imputed rent) uses the published total RegionalIndex, in which housing enters only through the 3%-weight rented-housing sub-component. "incl. IR" (including imputed rent) adds the imputed-rent profile via $\text{RegionalIndex}_{\text{incl}}(p) = \text{RegionalIndex}_{\text{excl}}(p) + 0.16 \cdot (\text{RentIndex}(p) - 100)$, where 0.16 is the imputed-rent weight in the 2020-base CPI basket and RentIndex is the rented-housing sub-component of the regional difference index (Tokyo $\approx 155$, rural prefectures $\approx 70$–90). The Tokyo MA – rural-origin log gap is 0.031 under "excl. IR" and 0.109 under "incl. IR". All other sample restrictions and inference (rural-origin COVID-clean panel, wild cluster bootstrap at the origin prefecture, $B = 999$, Rademacher) match Table 5. $^{***}$, $^{**}$, $^{*}$ denote significance at the 1%, 5%, 10% level.

**Why the robustness is tight.** Within-year price-level changes are common to all observations of a given year and are absorbed by wave fixed effects, so the deflation enters the regression primarily through the cross-prefecture component, which the Move main effect partially absorbs as a destination-specific shift; the cell-specific interactions are identified by the smaller cross-origin variation in cost of living within the rural sample. The framework's cell predictions are therefore about what movers *earn* — the pay difference the consolidation mechanism generates — not about what living *costs* where they land: if the cell pattern were merely compensation for Tokyo's higher cost of living, removing that cost by deflation would shrink it. In the data, deflating leaves the cells essentially unchanged (Table OA.34), which confirms this directly.

# OA.21 External validation: detailed cross-country evidence

This appendix provides the detailed evidence summarized in Section 8 and Table 7.

## OA.21.1 Decline rate vs. net migration at the regional level

Figure OA.13 reports the cross-regional correlation between young (15–29) population decline 2014–2024 and net migration rate (per 1,000 of population) averaged over 2020–2024, in five panels. To make the cross-country comparison consistent at the metropolitan-agglomeration scale, we aggregate the geographic units so that each major MA appears as a single observation: Japan (39 units—four MAs (Tokyo, Osaka, Nagoya, Fukuoka) plus 35 non-MA prefectures), Italy NUTS-2 (21 regions), Spain NUTS-2 (19 autonomous communities with non-missing data), Germany NUTS-2 (38 regions), and Korea (15 units—Seoul MA aggregated from three sido plus the 12 remaining single-sido units).

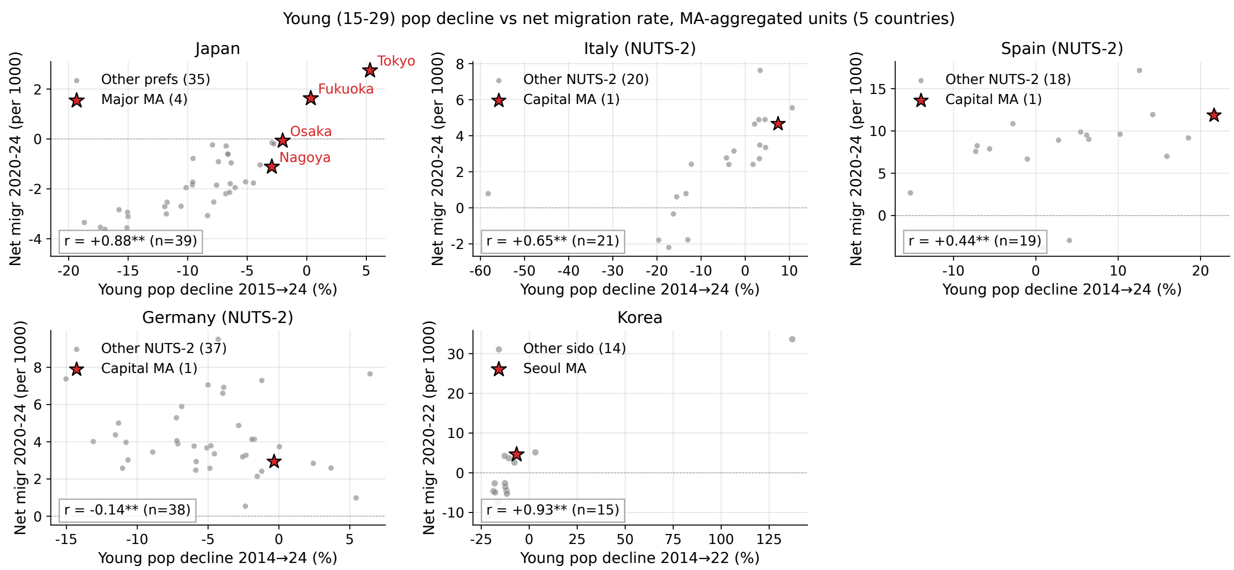



**Figure OA.13:** Young (15–29) population decline vs. net migration rate at the MA-aggregated regional level.

*Notes:* Young (15–29) population decline 2014–24 vs. net migration rate 2020–24 (per 1,000) at the MA-aggregated regional level. Japan: prefectures with the four major MAs (Tokyo, Osaka, Nagoya, Fukuoka) collapsed into single weighted observations. Korea: sido with Seoul MA collapsed from three sido. Italy, Spain, and Germany: Eurostat NUTS-2 regions, which already sit at roughly the metropolitan scale for the principal metros (e.g. Lombardia for Milano, Lazio for Roma, and the Madrid and Berlin regions). Sources: Eurostat (Italy, Spain, Germany), OECD Regional Database (Korea), and Census and Basic Resident Registry (Japan). The four Japanese MAs are individually labeled.

In every country the correlation carries the framework's predicted sign; its strength, however, varies sharply with the country's demographic-decline stage. The $r$ reported in each panel is the correlation, across the country's regions, between the young (15–29) decline rate over 2014–2024 and the average annual net migration rate per 1,000 over 2020–2024. A more positive $r$ means that regions declining faster also lose more residents to migration — consolidation pulling workers out of declining origins toward the apex MA.

- **Korea** (TFR $\approx$ 0.7, the most demographically advanced): all-unit correlation $r = +0.93$ ($n = 15$, $p < 0.001$), the strongest among the five. The visible upper-right outlier is Sejong (*Sejong Special Self-Governing City*, KR054), a planned administrative city approximately 120 km south of Seoul (in the Chungcheong region, *not* part of the Seoul metropolitan area) that was elevated to special-administrative status in 2012. Sejong's young-population growth of +137% over 2014–2022 reflects exogenous policy-driven inflows from the staged relocation of central-government ministries—the state moved its ministries and their employees there by decision, so the inflow records where the government placed workers, not workers responding to industry exits. Excluding Sejong, the all-unit correlation falls to $r = +0.77$ ($n = 14$, $p = 0.001$)—still the strongest among the five countries. This aggregate scatter is the descriptive counterpart of the framework's prediction (declining regions lose workers to the consolidation apex), but it does not by itself identify the industry-exit channel; the worker-level KLIPS regression in Appendix OA.21.3 is the direct test.

- **Japan** (TFR $\approx$ 1.15): all-unit $r = +0.88$ ($n = 39$). Among the four MAs, Tokyo (decline $g = +5.3\%$, net migration = +2.8 per 1,000 population per year) and Fukuoka ($g = +0.3\%$, net migration = +1.6/1,000/yr) sit in the upper-right quadrant (still expanding young pool, still attracting net inflow), whereas Osaka ($g = -2.0\%$, −0.1/1,000/yr) and Nagoya ($g = -2.9\%$, −1.1/1,000/yr) sit near the boundary of decline.

- **Italy** (TFR $\approx$ 1.21): all-NUTS-2 $r = +0.65$ ($n = 21$). Lazio (Roma MA) is a positive outlier; the southern regions (Mezzogiorno) cluster in the lower-left.

- **Spain**: all-NUTS-2 $r = +0.44$ ($n = 19$, $p = 0.06$). The Madrid community is a strong positive outlier; the depopulating interior communities ("España vaciada") sit in the lower-left.

- **Germany**: all-NUTS-2 $r = -0.14$ ($n = 38$, n.s.); the relationship is essentially flat, reflecting the fact that German labor markets have remained roughly stable in aggregate thanks to large international immigration.

### OA.21.2 Spatial structure of internal migration: dominant destination and net inflow to the apex MA

The demographic setting fixes the reading before any map is drawn (Section 8): Korea's working-age population is already contracting, as Japan's is, while in Italy, Spain, and Germany international immigration has so far prevented contraction (Figure 1 in the Introduction). The maps then show where internal migration concentrates under each setting, and the expectation is asymmetric: apex-ward absorption where the trigger is met (Korea), and at most the ordinary attraction of a large city where it is not (the three European cases).

On this premise, we ask where internal migration actually concentrates in each country. Each country gets two visualizations, stacked vertically as panels (A) and (B) of one combined figure (Korea: Figure OA.14; Italy: Figure OA.15; Spain: Figure OA.16; Germany: Figure OA.17):

*Panel (A)—dominant destination per origin region* (the bilateral-OD analog of Japan's net-flow destination maps, Appendix OA.2). Origin regions whose largest single bilateral flow goes to the apex MA are colored red; origins whose largest flow goes to some other region are colored orange — a regional hub, not the apex, absorbs them; origins below the 1% cumulative-flow threshold are masked gray. Source: bilateral migration data per country (ISTAT for Italy, INE for Spain, KOSIS for Korea, Destatis for Germany), aggregated into three five-year epochs (2010–2014, 2015–2019, 2020–2024).

*Panel (B)—annual net in-migration to the apex MA per origin region* (the bilateral-flow analog of Japan's Figure OA.5). Same three epochs, with a red–white–blue scale (red = origin sends net workers to the apex MA; blue = apex MA sends net workers to the origin; white = approximate balance). Panel (A) shows only where each origin's largest flow goes, not how much moves; panel (B) adds the magnitudes.

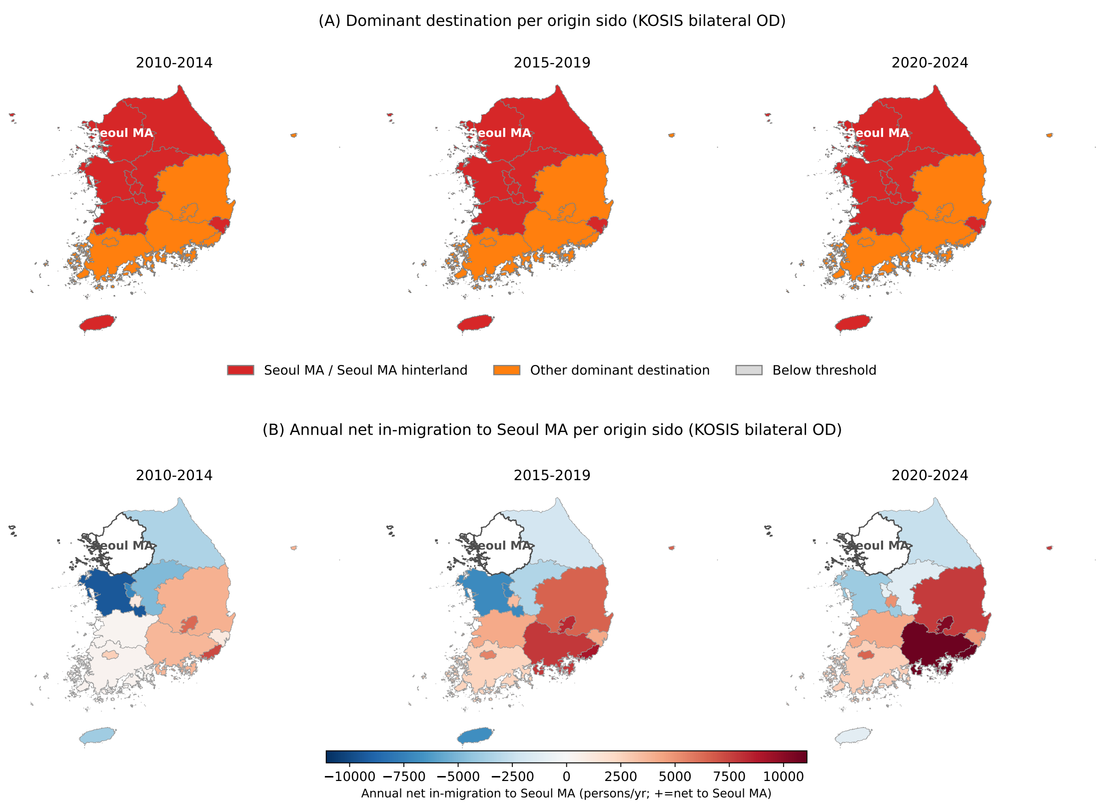



**Figure OA.14:** Dominant destination and net in-migration to the apex MA, Korea (full trigger).

*Notes:* Panel (A): dominant destination per origin sido (KOSIS bilateral OD). Seoul MA absorbs $\sim$ 79% of non-Seoul-MA sidos as their dominant destination; the remaining 3 sidos (Daegu, Busan, Gwangju) form a south-Korean regional-anchor cluster with their surrounding provinces — the analog of Fukuoka's Kyushu anchor role in Japan. Panel (B): annual net in-migration to Seoul MA per origin sido. Nearly all sidos appear in deep red across all three epochs.

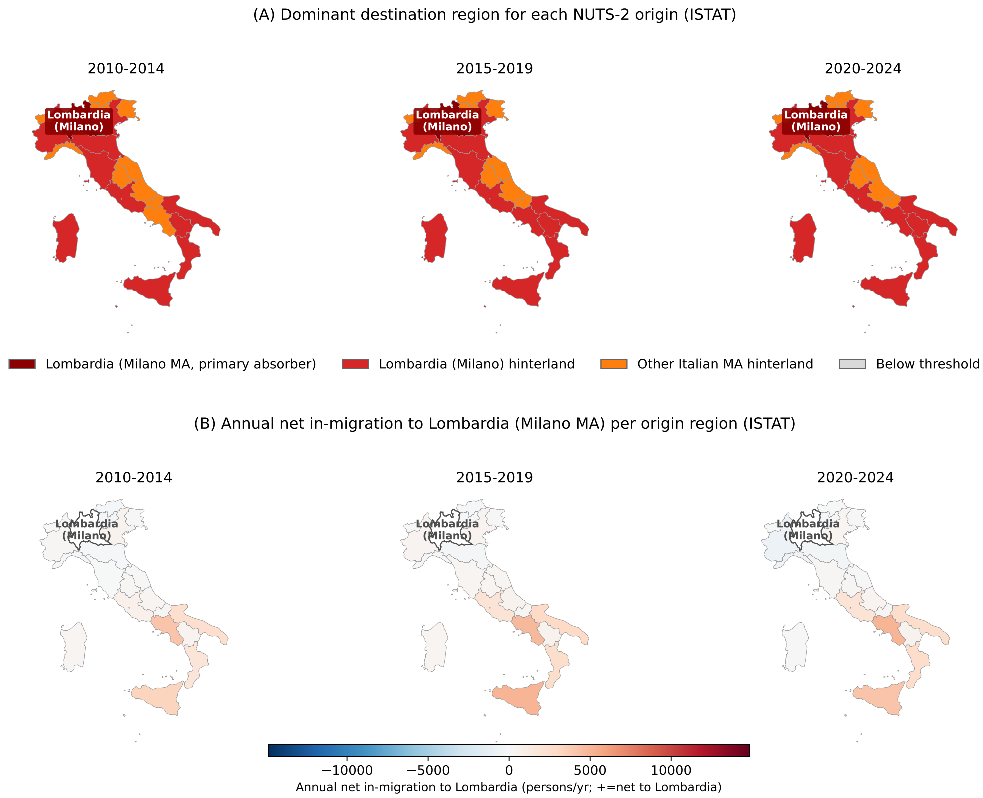



**Figure OA.15:** Dominant destination and net in-migration to the apex MA, Italy (partial trigger).

*Notes:* Panel (A): dominant destination NUTS-2 region. Lombardia (the region containing Milano) is the dominant destination for 11 of 21 NUTS-2 regions in 2020–2024, including Lazio (Roma's region); the pattern is stable across all three epochs. Panel (B): annual net in-migration to Lombardia per origin. Mezzogiorno regions appear in light red (sending to Lombardia); the Centre-North is in approximate balance. Source: ISTAT.

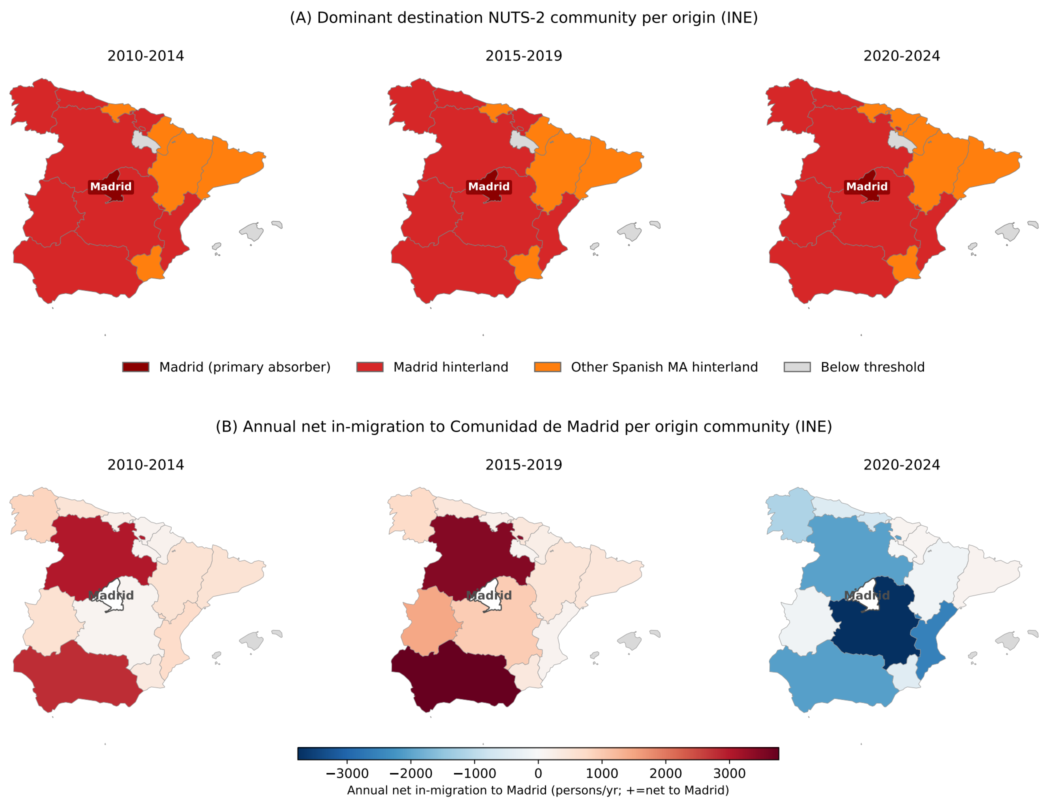



**Figure OA.16:** Dominant destination and net in-migration to the apex MA, Spain (trigger not yet met).

*Notes:* Panel (A): dominant destination NUTS-2 community. Madrid is the dominant destination for 7 of 17 NUTS-2 communities, including the depopulating interior (Castilla y León, Castilla-La Mancha, Extremadura), Galicia, Asturias, and the Comunitat Valenciana. Panel (B): annual net in-migration to Comunidad de Madrid per origin community; the 2020–2024 panel shows a post-COVID reversal, with most communities *net-receiving* from Madrid. Source: INE.

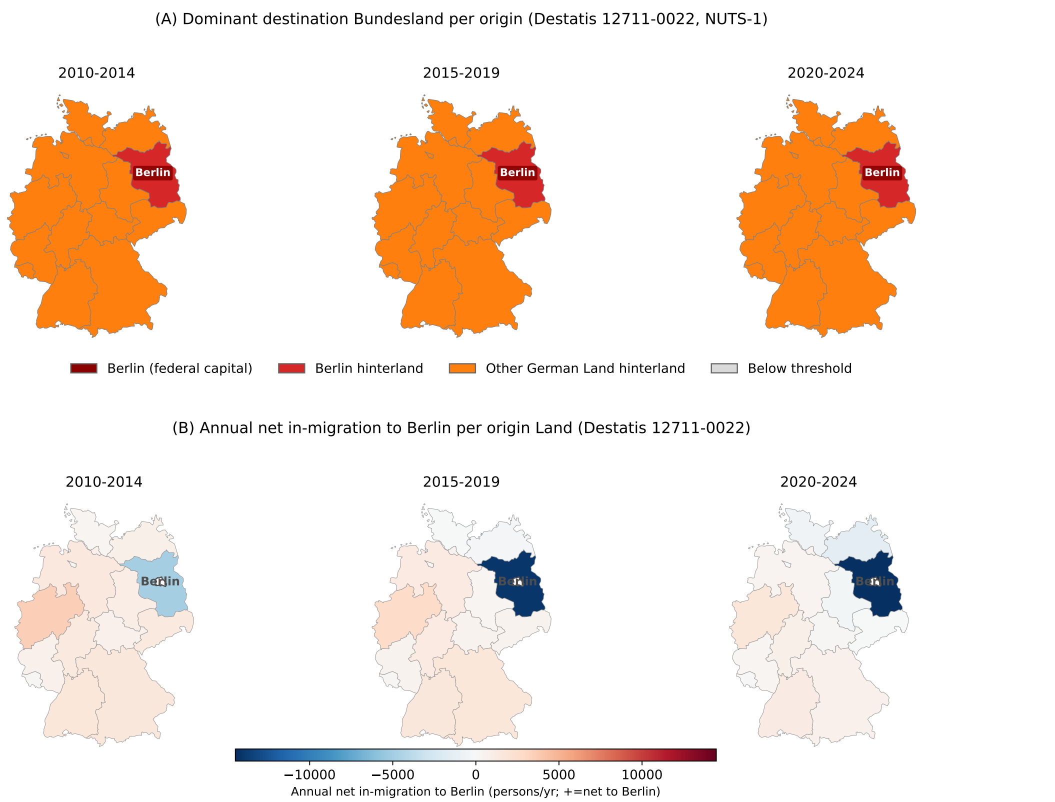



**Figure OA.17:** Dominant destination and net in-migration to the apex MA, Germany (no trigger).

*Notes:* Panel (A): dominant destination Bundesland (NUTS-1, 16 units) per origin. Only Brandenburg appears as Berlin's hinterland in any epoch (reflecting suburbanization); the remaining 14 Bundesländer each have a non-Berlin dominant destination, mostly a geographic neighbor (Hamburg↔Schleswig-Holstein, Bayern↔Baden-Württemberg, Bremen→Niedersachsen, Saarland→Rheinland-Pfalz). Panel (B): annual net in-migration to Berlin per origin Land. Almost all Bundesländer appear in white (approximate balance); only Brandenburg is in deep blue (Berlin's suburbanization). The visual null contrasts sharply with Korea and Italy (Spain reads the same way as Germany). Source: Destatis.

Read together, the four figures order the countries exactly as the trigger condition of Section 8 predicts. Korea (full trigger; Figure OA.14) shows nearly all sidos in deep red in panel (B) and 11/14 sidos as Seoul MA hinterland in panel (A). Italy (partial trigger; Figure OA.15) shows the Mezzogiorno regions in light red and Lombardia (Milano) absorbing roughly half the country as its hinterland, stable across the three epochs — the same absorption pattern as in Korea and Japan, but covering only part of the country. Spain and Germany — where the demographic trigger of Section 8 is not met — show what that classification predicts (Figures OA.16 and OA.17). Germany is a clean null: panel (B) is uniformly white, with only Brandenburg in deep blue (Berlin suburbanization). Spain is noisier but contradicts nothing: several communities are red in the pre-COVID epochs, yet the underlying net rates are an order of magnitude below Korea's (invisible in the maps, whose color scales are per-country), and in 2020–24 the direction reverses, most communities receiving net workers *from* Madrid. With $r = +0.44$ and flows this small and unstable, Madrid's dominance in panel (A) reads as a large city's ordinary pull, not decline-driven consolidation.

The hinterland counts from these figures — together with Japan's, where Tokyo MA is the dominant destination for roughly 28–30 of the 47 prefectures — are what the apex-MA-dominance column of Table 7 summarizes (Section 8).

### OA.21.3 Korea worker-level wage premium (KLIPS)

We upgrade the Korea regional-aggregate evidence (Section 8) to a worker-level test analogous to the JPSED specification of Section 6.1, using all 27 waves (1998–2024) of the Korean Labor and Income Panel Study (KLIPS; Korea Labor Institute, open access). Korea cannot support the four-cell design itself — its origin sizes are too alike for a size split to carry information — but the decline axis is identifiable, and on it the framework's signature appears: in the era after Korea's tertiarization was complete, when the relation between industry location and population had stabilized (2006 onward, defined below), movers to Seoul MA from fast-declining origins earn about 9 log points more than movers from slow-declining origins, and the gap keeps its sign under every single-origin deletion.

**Sample definition.** Seoul MA is the aggregate of Seoul, Incheon, and Gyeonggi-do: the three sido form a single commuting-integrated metropolitan labor market, so they enter as one destination — the Korean analog of the Tokyo MA. The outcome is the within-individual log change in real monthly wage-worker pay between consecutive KLIPS waves (each endpoint deflated by its sido-year CPI); the treatment is a move into Seoul MA; the sample is the $\leq 29$ young sub-sample with the same student part-time exclusion, trimming, longitudinal weights, and wave effects as the JPSED baseline, and standard errors are score-based wild cluster bootstrap ($B = 999$, Rademacher) at the origin sido. The origin pool holds ten origins: the five non-Seoul metropolitan cities (Busan, Daegu, Gwangju, Daejeon, Ulsan) and five provinces (Gangwon, Jeonbuk, Jeonnam, Gyeongbuk, Gyeongnam). All ten are sido — Korea's first-level administrative units, the grain at which KLIPS records residence — but sido come in two kinds: a metropolitan city is a single city that was carved out of the surrounding province and given the same rank as a province, while a province is the remainder, a region of many small cities and counties. Two provinces bordering Seoul MA — Chungbuk and Chungnam — are excluded. They function as Seoul MA's outer ring, not as its periphery. Korean law caps new plants and offices inside Seoul MA, so firms that want to be near Seoul build just outside its boundary — which is exactly where these two provinces lie. Jobs accumulate there (commuter rail reaches Cheonan in northern Chungnam). Their workers therefore already sit inside Seoul's extended labor market, and a move from there into Seoul MA is a short move within one metropolis, not the periphery-to-apex move the framework prices. Sejong is excluded with them: the planned administrative city was created in 2012 inside this same border area, and its inflows are government relocation by decision rather than labor-market moves; in the 1992 base, its predecessor, Yeongi-gun (then a county of Chungnam), is subtracted from Chungnam's count, so the window compares the same territory at both ends. The five metropolitan cities, by contrast, stay in the origin pool. The Japanese design excludes the Osaka and Nagoya MAs because they compete with Tokyo as destinations; the Korean metros hold no such role — they are themselves declining origins (Busan's young population fell 49% over 1992–2019, the deepest decline among the twelve). Jeju is excluded throughout: a tourism-economy island whose young population grew over the window. The accounting is exhaustive: of Korea's seventeen sido, three form the destination, ten the origin pool, and four are excluded for the reasons just given. Figure OA.18 maps the assignment.

**Why the four-cell design does not transfer.** The Japanese design reads two origin coordinates, size and decline. In Korea the size coordinate has no identifying variation: the ten origins span total populations of 1.15–3.41 million — a log range of 1.09, with standard deviation 0.37 — against Japan's rural-origin range of 0.58–5.50 million (log range 2.24, standard deviation 0.50, over 35 origins). Japan's *small-origin side alone* (24 prefectures below the size split) spans 1.13 log units — wider than the whole range of the Korean origins. Every Korean origin sits in what the Japanese design treats as the middle of the size distribution; the very small and very large origins that identify the size dimension are missing among the Korean origins — the one very large agglomeration is Seoul MA itself, and it is the destination. This is the imprint of Korea's spatial structure: half the national population lives in the single apex, and no second tier of large cities remains below it — the regional metros stopped absorbing migrants decades ago. The industry side shows the same hierarchy as in Japan: hosted counts rise with size along a tight relation, and industry exit rates fall with it — the direction of Japan's concavity on jobs (Figure OA.19).[3] But the ten origins all sit close together on the size axis, with no gap at which a large/small split could be drawn. The decline coordinate, by contrast, carries variation comparable to Japan's: $\tilde{g}$ spans 19 percentage

[3]The figure reads the five-digit level, the finest grain of the Korean classification (1,128 categories hosted nationally in 2006), rather than the four-digit grain (476 categories, the closest match in count to the harmonized three-digit JSIC of the Japanese panels). The choice follows the same counting logic as the exit measures of Section 5.1: industry exits are rare events, so a grain must be fine enough to record enough of them per origin to measure. At the four-digit grain each origin loses only 3–11 industries over the window — too few to tell a real difference between origins from the randomness of counting so few events. The five-digit grain lifts this to 14–44 exits per origin, and the differences between origins accordingly stand well clear of that counting randomness.

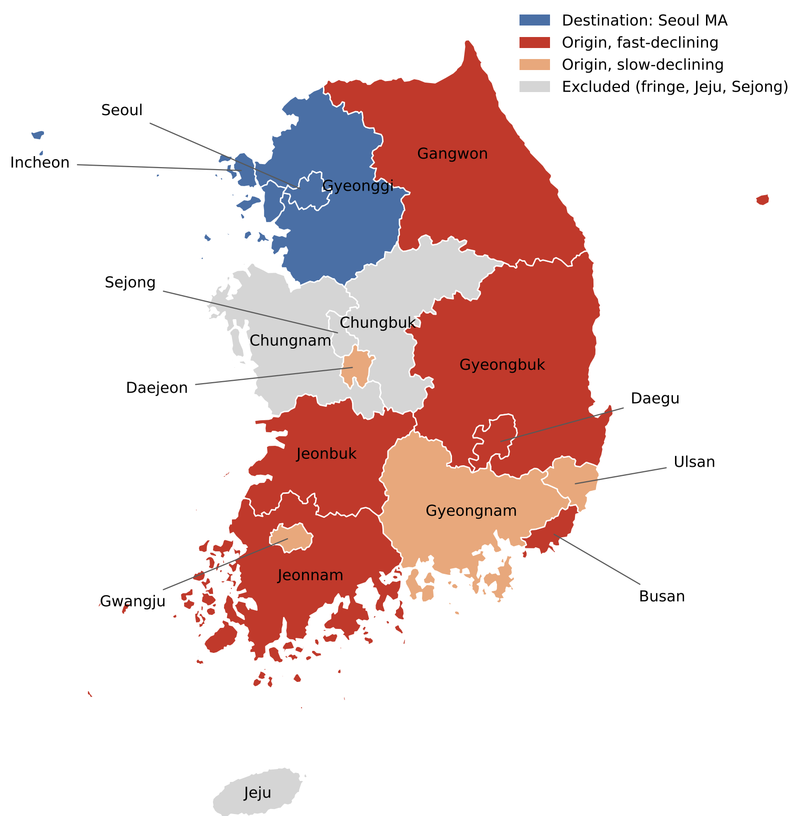



**Figure OA.18:** The KLIPS design across Korea's seventeen sido: destination, origins, and exclusions.

*Notes:* Blue: the destination (Seoul MA = Seoul, Incheon, Gyeonggi). Red and light orange: the ten origins, split into the fast- and slow-declining sides of the baseline era (Table OA.35). Gray: excluded — Chungbuk and Chungnam (the Seoul MA fringe), Jeju (tourism island), and Sejong (the planned administrative city). Boundaries: kostat 2013 sido polygons.

points across the ten origins over the era window (Table OA.35), against 15 points across Japan's thirty-five rural origins over its 2010–2019 window. We therefore estimate the decline axis alone — a fast/slow contrast and a continuous $M \times \tilde{g}$ gradient — and compare with Japan on signs only.

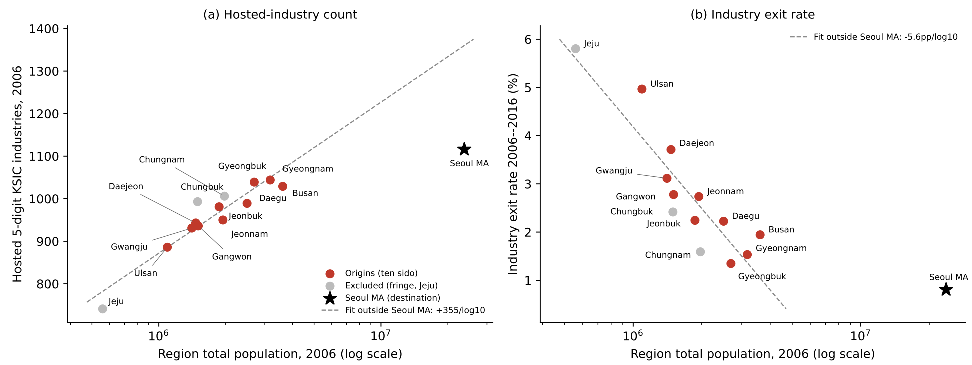



**Figure OA.19:** Hosted industries and industry exits against region size in Korea.

*Notes:* Hosted industries and industry exits against region size in Korea (sido grain; five-digit KSIC; Establishment Census). Panel (a): industries hosted at the 2006 window start. The relation the Japanese design reads is present — hosted counts rise with size — but the origins occupy a narrow band of both size and count; only Seoul MA stands apart, so a size split among the origins has almost no variation to read. Panel (b): the share of those industries no longer hosted by 2016. The exit rate falls with size — the direction of Japan's concavity on jobs — and is smallest at Seoul MA. The establishment windows end in 2016, not 2019, to stay inside a single KSIC revision: the classification changes in 2017, and crossing it would count reclassifications as exits.

**Regime condition: why the baseline era starts in 2006.** The Korean young (15–29) stock peaked around 1991–92 and has fallen monotonically since, from 13.2 million in 1992 to 9.5 million in 2019, so we measure decline from 1992.[4]

But the mechanism the paper tests — population decline removing the industries a place can host, and movers from the depleted places carrying a premium at the apex — needs a window in which that response can be seen, and two things decide where such a window lies. First, moves driven by an economy-wide industrial transformation have nothing to do with regional decline, so an era of ongoing transformation contaminates the decline effect. Second, the hierarchy between population size and industry location — the relation the hosting-threshold mechanism runs on — holds more tightly for tertiary industries than for manufacturing: tertiary industries serve customers where they live, so they follow population where it accumulates and exit where it drains away, while manufacturing also answers to ports, industrial estates, factor costs, and siting policy, and so tracks population only loosely. Japan's observation window raises neither problem: by the time its decline phase began, tertiarization had been complete for decades. Korea's window since 1992 splits into

[4]Reaching back to 1992 takes two mechanical data steps, recorded here for replication. First, KOSIS provides the age-by-region counts as two separate data series — an older one that ends in 2010 and its successor that starts in 2011 — and we join them at 2010/2011. The join is smooth: in every sido, the 2011 value of the new series is only 0.3–3.4% below the 2010 value of the old one — roughly one year's worth of young decline — so switching from one series to the other does not create an artificial jump in the data. Second, the regions of 1992 are not the regions of today: Ulsan was then part of Gyeongnam (it became a metropolitan city in 1997), and today's Sejong was then Yeongi-gun, a county of Chungnam. We therefore recompute the 1992 young counts as if today's regions had already existed.

two eras on exactly these margins, and we can measure the split directly in the Establishment Census (all establishments, sido grain, five-digit KSIC; windows 1993–2005 and 2006–2016 chosen to sit inside single KSIC revisions). In the *transition era* (1993–2005) Korea was still tertiarizing — the manufacturing share of establishments fell in every region ($-1.6$pp nationally), a compositional shift uncorrelated with regional decline (correlation with $g$: $-0.03$) — and the relation between industry location and population was still weak: regressing regional establishment growth on population growth gives an elasticity of 0.39 ($R^2 = 0.30$) for non-manufacturing and essentially zero fit for manufacturing ($R^2 = 0.03$). In the *post-transition era* (2006 onward) the transition is complete and the relation is tight: the non-manufacturing elasticity reaches 0.96 ($t = 4.0$, $R^2 = 0.57$) — establishment growth now moves almost exactly with population growth — while manufacturing's fit stays low ($R^2 = 0.04$).

The baseline sample is therefore the post-transition era, with the decline window and the wage transitions aligned inside it: $g$ = young growth 2006–2019, transitions 2006/07–2018/19 (pre-COVID). Size is dated by the same rule: $\log \tilde{S}$ is measured at each era's own window start — the 1992 base for the transition era, the 2006 base for the baseline era. The transition era ($g$ = 1992–2005, transitions 1998/99–2005/06) is reported for comparison: its moves track the economy-wide transformation, which the result above shows to be unrelated to origin decline. The baseline era is thus the window that clears both conditions: the transformation is over, and the economy's weight sits in the sector whose location is tightly bound to population.

**Specification.** Within each era we estimate, on the ten-origin pool,

$$\Delta \log w_{i,t} \;=\; \beta_1 M_{i,t} \;+\; \beta_2\, M_{i,t}\, \mathbf{1}[\tilde{g}_{o(i)} < 0] \;+\; \gamma' X_{i,t} \;+\; \alpha_{o(i)} \;+\; \lambda_t \;+\; \varepsilon_{i,t}, \tag{OA.3}$$

Columns (2) and (4) of Table OA.36 additionally report a continuous variant, replacing $\beta_2 M \times \mathbf{1}[\tilde{g} < 0]$ with $M \times \log \tilde{S}$ and $M \times \tilde{g}$; it documents that the contrast is not an artifact of the binary split and that the size interaction is empty, but the claim rests on equation (OA.3): with ten origins, a continuous slope depends on exactly how deep the deepest origin's decline is measured to be, while the binary contrast uses only which side of the boundary each origin lies on. All origin coordinates are demeaned over the 14-region grid (Seoul MA aggregated, five metro cities, seven provinces, Jeju). $\tilde{g}$ is measured over the era's own window, and $\log \tilde{S}$ at the era's own window start (1992; 2006) on the city/county (unit) grain — the young-weighted average over each origin's municipalities at that start. Under this convention a metro city is measured by its own size, and a province by the young-weighted average size of its municipalities — so that a province made of small cities is not read as one large city. $\alpha_o$ are origin (sido) fixed effects — the baseline convention of the Japanese specification — so the origin-level size and decline shifters are absorbed, and the Move-side contrasts are identified from within-origin mover–stayer comparisons compared across origins. $\mathbf{1}[\tilde{g} < 0]$ marks the origins declining faster than the 14-region mean; Table OA.35 lists the classification. Only $\beta_2$ (and the continuous $M \times \tilde{g}$) is interpreted: as in Japan, the level is not read — the object is the fast–slow difference (Section 5.2).

**Table OA.35:** Korean origin pool and fast/slow classification.

| Origin | $g$ 1992–2019 | $\tilde{g}$ 2006–19 | side | movers |
|---|---|---|---|---|
| Busan | −48.7% | −14.6 | fast | 7 |
| Daegu | −34.7% | −5.0 | fast | 0 |
| Gangwon | −43.0% | −0.1 | fast | 3 |
| Jeonbuk | −49.1% | −4.5 | fast | 3 |
| Jeonnam | −55.7% | −5.2 | fast | 4 |
| Gyeongbuk | −49.3% | −8.8 | fast | 7 |
| Gwangju | −22.0% | +4.5 | slow | 12 |
| Daejeon | −11.3% | +0.5 | slow | 1 |
| Ulsan | −17.0% | +0.9 | slow | 2 |
| Gyeongnam | −34.0% | +0.3 | slow | 4 |

*Excluded*: Chungbuk, Chungnam (Seoul MA fringe); Jeju (tourism island); Sejong.
*Destination*: Seoul MA (young share of nation 44.1% → 52.2%, 1992–2019).

*Notes:* The ten origins of the regime-condition design (Seoul MA is the destination; Chungbuk, Chungnam, Jeju, and Sejong excluded). $g$ (1992–2019) is each origin's young (15–29) population change over the full window — negative everywhere, i.e. decline. $\tilde{g}$ (2006–19) is the young growth over the baseline era, demeaned across the 14 regions, and it is $\tilde{g}$ that sets the *side*: faster- versus slower-than-average decline ($\tilde{g} < 0$ vs. $\tilde{g} > 0$). *Movers* = Seoul-MA movers in the baseline (2006/07–2018/19) transitions, $\leq 29$ sample (Daegu contributes none). Gangwon and Gyeongnam lie within half a point of the boundary, one on each side; leaving either out changes nothing (Table OA.37). Source: KOSIS resident registration (legacy/current tables spliced, 1992 base boundary-adjusted); KLIPS waves 1–27.

**Results.** Table OA.36 reports both eras. In the post-transition era the framework's signature appears. Moving to Seoul as such carries no premium: the linear Move coefficient is −0.019 ($t = -0.5$). What exists is a difference across origins. Movers from fast-declining origins earn +0.089 more than movers from slow-declining origins ($t = +2.1$); in the continuous variant the premium grows with the depth of the origin's decline ($M \times \tilde{g}$: −0.0094 per percentage point, $t = -7.1$), and the size interaction is the predicted null ($t = -0.0$). Dropping origins one at a time (Table OA.37), the gap keeps its sign in all ten deletions and is significant in eight. The two exceptions are the two poles of the comparison: dropping Gwangju ($t = +0.1$) removes 12 of the 19 slow-side movers — the comparison group — and dropping Busan (+0.065, $t = +1.6$) removes the deepest-declining origin, on which the fast side leans. The continuous gradient behaves the same way (last column of the table): sign-stable in all ten deletions, significant in eight, weakest under the same two — without Busan its standard error inflates about eightfold (−0.014, $t = -1.4$). In the transition era nothing is identified: the fast/slow contrast is −0.006 ($t = -0.1$) and the continuous gradient −0.003 ($t = -0.5$). The transition-era estimates are noisy — the era holds only 25 movers, and their wage changes also pass through the 1997–98 financial crisis — so they show no decline-linked premium without proving an exact zero. A null is nevertheless what this era should produce: its moves tracked the transformation, and decline did not yet translate into lost industries at the origin. The decline-linked premium thus appears only in the later era.

**Table OA.36:** Korea worker-level wage premium for Seoul MA migration.

| | Transition era $g$: 1992–2005; moves 1998/99–2005/06 | | Population-following era $g$: 2006–2019; moves 2006/07–2018/19 | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Move ($M$) | $+0.057^{*}$ | $+0.066$ | $-0.069^{*}$ | $-0.053^{***}$ |
| | (0.030) | (0.060) | (0.038) | (0.020) |
| $M \times \mathbf{1}$[fast decline] | $-0.006$ | | $+0.089^{**}$ | |
| | (0.074) | | (0.043) | |
| $M \times \log \tilde{S}$ | | $-0.021$ | | $-0.000$ |
| | | (0.029) | | (0.012) |
| $M \times \tilde{g}$ | | $-0.0026$ | | $-0.0094^{***}$ |
| | | (0.0052) | | (0.0013) |
| Person-wave pairs | 2,020 | 2,020 | 2,922 | 2,922 |
| Movers | 25 | 25 | 43 | 43 |

*Notes:* Worker-level wage premium for a Seoul-MA move, ten origins (Table OA.35), KLIPS waves 1–27. Columns (1)–(2) are the transition era, (3)–(4) the post-transition (baseline) era; within each era the two columns pair the fast/slow contrast (equation (OA.3)) with its continuous variant, with $\mathbf{1}$[fast decline] $= \mathbf{1}[\tilde{g}_o < 0]$ on that era's own axes. Dependent variable: the within-worker log change in real monthly wage-worker pay between consecutive waves (endpoints deflated by sido-year CPI); $\leq 29$ sub-sample, student part-timers excluded, growth trimmed at 1/99%. Controls: age, age$^2$, university-or-higher, sex; origin (sido) and wave fixed effects; KLIPS longitudinal weights. Standard errors: score-based wild cluster bootstrap ($B = 999$, Rademacher) at the origin sido. $^{***}/^{**}/^{*}$: 1/5/10%. Source: KLIPS waves 1–27; KOSIS.

**Table OA.37:** Leave-one-out over origins: the baseline fast/slow contrast and the continuous gradient (columns (3) and (4) of Table OA.36).

| Dropped origin | side | movers left | $M \times \mathbf{1}$[fast] | $M \times \tilde{g}$ |
|---|---|---|---|---|
| Busan | fast | 36 | $+0.065$ $(+1.57)$ | $-0.0144$ $(-1.39)$ |
| Daegu | fast | 43 | $+0.086^{*}$ $(+1.89)$ | $-0.0092^{***}$ $(-6.45)$ |
| Gwangju | slow | 31 | $+0.006$ $(+0.09)$ | $-0.0054$ $(-0.98)$ |
| Daejeon | slow | 42 | $+0.100^{***}$ $(+2.68)$ | $-0.0096^{***}$ $(-7.69)$ |
| Ulsan | slow | 41 | $+0.088^{*}$ $(+1.75)$ | $-0.0093^{***}$ $(-6.20)$ |
| Gangwon | fast | 40 | $+0.106^{***}$ $(+2.72)$ | $-0.0086^{***}$ $(-5.42)$ |
| Jeonbuk | fast | 40 | $+0.094^{**}$ $(+2.10)$ | $-0.0094^{***}$ $(-7.00)$ |
| Jeonnam | fast | 39 | $+0.083^{*}$ $(+1.91)$ | $-0.0095^{***}$ $(-8.21)$ |
| Gyeongbuk | fast | 36 | $+0.094^{*}$ $(+1.69)$ | $-0.0105^{***}$ $(-12.39)$ |
| Gyeongnam | slow | 39 | $+0.103^{***}$ $(+2.89)$ | $-0.0098^{***}$ $(-9.58)$ |

*Notes:* Each row re-estimates the two specifications with one origin removed (9 clusters after each deletion); wild-cluster-bootstrap $t$ in parentheses. Dropping Gwangju removes 12 of the 19 movers on the slow-declining side; even so, both coefficients keep their sign. $^{***}/^{**}/^{*}$: 1/5/10%. Source: KLIPS waves 1–27; KOSIS.

### OA.21.4 Caveats

Two caveats close the appendix. First, on scope. The framework's mechanism switches on only where the working-age population is actually contracting, and that trigger is met today only in Japan and Korea: in Italy, Spain, and Germany, international immigration has so far kept the working-age population from contracting (Section 8). The paper's cross-country claims therefore rest on Japan (the body of the paper) and Korea (the worker-level replication above). The three European cases enter with a different job: they check that where the trigger is not met, the framework's pattern is absent. Germany shows exactly that predicted absence — a statement about regional aggregates only; no worker-level analysis is attempted for Germany. Italy and Spain do show concentration toward Milano and Madrid in the same aggregates, but since the trigger is not met there, the mechanism cannot be its source, and we do not count that concentration as evidence for the framework. Second, on units: each country is measured on its own regional definition (NUTS-2 for Italy and Spain, NUTS-1 for Germany, sido for Korea, MA-aggregated prefectures for Japan). This is workable for two reasons: the definitions produce similar numbers of units per country (15–39), of broadly similar economic size; and the cross-country comparison is qualitative — we ask whether each country's correlation is positive and which countries show the stronger pattern, and we never compare the numbers themselves. Robustness to alternative unit aggregations would still be a useful refinement.

## References


**Combes, Pierre-Philippe, Gilles Duranton, and Laurent Gobillon**, "Spatial wage disparities: Sorting matters!," *Journal of Urban Economics*, 2008, *63* (2), 723–742.

**Dauth, Wolfgang, Sebastian Findeisen, Enrico Moretti, and Jens Suedekum**, "Matching in Cities," *Journal of the European Economic Association*, 2022, *20* (4), 1478–1521.

**Ministry of Internal Affairs and Communications**, "Basic Resident Registry (Jumin Kihon Daicho): Report on Internal Migration, and Population by Age, Sex, and Prefecture," https://www.stat.go.jp/data/idou/index.html; https://www.soumu.go.jp/main_sosiki/jichi_gyousei/daityo/jinkou_jinkoudoutai-setaisuu.html 1954–2025.

**Mori, Tomoya**, "Demography in Disguise: Comments on "Demographic Challenges and Economic Stagnation in Japan" by Takeo Hoshi," 2026. SSRN Working Paper, https://doi.org/10.2139/ssrn.6655558.

**Recruit Works Institute**, "Japanese Panel Study of Employment Dynamics (JPSED)," 2016–2025. Data accessed via SSJDA, University of Tokyo.

**Solon, Gary, Steven J. Haider, and Jeffrey M. Wooldridge**, "What are we weighting for?," *Journal of Human Resources*, 2015, *50* (2), 301–316.

**Statistics Bureau, Ministry of Internal Affairs and Communications**, "Consumer Price Index, National 2020-Base Series, and Regional Difference Index of Consumer Prices (Shouhisha Bukka Chiiki-sa Shisuu) 2023 Release," https://www.stat.go.jp/data/cpi/ 2024.

**Stone, Douglas and William T. Ziemba**, "Land and Stock Prices in Japan," *Journal of Economic Perspectives*, 1993, *7* (3), 149–165.