DEMOGRAPHY / REGIONAL GROWTH / GERMANY

Composition Is Not Adjustment

Regional Aging, Convergence, and Growth in Germany

Michael Schymura

23 August 2026 · Long research edition 0.9.3

ABSTRACT

Aging–growth regressions can mix adjustment to demographic change with persistent regional composition. We examine that distinction using three decades of German Länder data and annual output, hours, employment, and population accounts. A 10% higher period-average old-to-young ratio is associated with 0.317 percentage points lower annual growth in real workplace output per resident aged 20–79. Small-cluster inference is not unanimous: exhaustive WCR11 enumeration gives $p = 0.010$, whereas the CRV3 jackknife gives $p = 0.070$. The descriptive estimate attenuates to -0.096 after 2000. A 15-Land additive East–West model and a separate all-16 period-only sensitivity do not reject slope equality, although their period contrasts differ in sign. Direct tests show that within- and between-Land associations differ for output per resident, output per hour, and workplace relative to resident employment after five-outcome adjustment. The positive output-per-hour association is accounted for by declining hours and approximately flat raw output in the point estimates, providing no evidence of raw-output expansion. Official 2026 boundary data yield no significant residual spatial autocorrelation, and prespecified panel-Conley sensitivities do not widen benchmark uncertainty. The evidence documents conditional regional associations rather than a causal law: composition, transition, convergence, and adjustment are distinct empirical objects.

JEL: J11, O40, R11, R23

Keywords: population aging; regional convergence; initial conditions; accounting; Germany

Data statement: official population, regional accounts, ETR employment and hours, and BKG boundaries underlie the evidence. Redistributable derivatives, source vintages, URLs, code, and checksums are included. ETR resident-employment workbooks carry an artifact-specific redistribution restriction; exact retrieval instructions and hashes are supplied.

Reproducibility statement: one locked environment and evidence renderer feed all editions. Verification covers direct restrictions, spatial inputs, tables, metadata, anonymity, page rasters, and deterministic rebuilds.

Contents

SECTION 1

The Question Behind the Coefficient

A large coefficient can be real without being portable. Across three decades and 16 German Länder, regions with older average population structures record slower contemporaneous growth in real workplace output per resident aged 20–79. The magnitude is close to the cross-country estimate reported by Hayashi (2026). Yet the resemblance weakens when the historical window, geographic comparison, demographic object, or accounting denominator changes. The central question is therefore not whether one German regression can reproduce a negative aging–growth coefficient. It can. The question is what variation that coefficient represents.

This paper separates four objects that are often treated as synonyms. *Composition* is the average demographic state of a region during an outcome window. *Transition* is the change in that state. *Convergence* is growth conditional on initial economic position. *Adjustment* is the movement of output, hours, employment, commuting, and population through which regional accounts respond. A composition coefficient may reflect demographic pressure. It may also reflect selective migration, commuting, sector structure, persistent regional type, or post-unification history. The coefficient alone cannot decide among those explanations.

Germany makes this distinction unusually visible. National law, currency, and many institutions are common, but migration, workplace location, industrial structure, and the East–West convergence process are not. Younger adults can move toward universities and expanding labor markets. Workers can produce outside their Land of residence. Initial output levels differ sharply across the post-unification geography. As a result, demographic composition is partly an equilibrium outcome rather than a treatment assigned to otherwise comparable regions.

The paper begins with the closest German regional analogue to Hayashi's cross-country benchmark. The old-to-young ratio is defined among residents aged 20–79, and the outcome is real workplace output per resident in the same age range. A 10% higher period-average ratio is associated with -0.317 percentage points lower annual growth. The effect is economically substantial. Its statistical interpretation is less singular: exhaustive WCR11 enumeration gives $p = 0.010$, while the leave-one-cluster CRV3 reference gives $p = 0.070$. Both results belong in the same sentence. Exhaustive enumeration removes Monte Carlo error from the sign distribution; it does not create generally exact finite-sample inference.

The full-sample coefficient is not the paper's strongest result. Once the 1991–2001 window is removed, the point estimate attenuates to -0.096 percentage points and is imprecise. Estimates that delete the five eastern non-city Länder, retain only western Länder, or add East-by-window intercepts are also much smaller. These comparisons show descriptive concentration, not by themselves a structural break. Paper 0.9.3 therefore estimates a registered pooled interaction model. Berlin is retained in the nationwide estimates but excluded from the formal East–West slope contrast because its divided-city history does not map cleanly to either non-city group. The post-2000 minus pre-2000 slope difference is 0.304 points ($p = 0.198$), and the eastern minus western slope difference is 0.397 points ($p = 0.151$). A joint two-restriction test gives $p = 0.220$. The point estimates attenuate sharply across period and geography, but the pooled tests do not resolve those slope differences with 15 clusters.

The annual accounts reveal a different and more consequential distinction. Persistent between-Land age composition is negatively associated with output per resident and output per hour. Within-Land and first-difference associations often have the opposite sign. The direct within–between routine in the preceding revision mixed the studentization of separately restricted coefficient draws with draws generated under the direct null. Paper 0.9.3 corrects that procedure: each contrast now uses one restriction, one null-imposed bootstrap distribution, and one studentization throughout. The five direct-null draw series then form one single-step maximum absolute statistic. Under that corrected contract, the within–between differences for output per resident, output per hour, and workplace relative to resident employment reject after five-outcome adjustment. Their adjusted p-values are 0.030, 0.021, and 0.037. The source of variation is not merely visually different; it is statistically distinguishable for selected components.

That correction strengthens the estimand argument without identifying a mechanism. In the point estimates, the positive output-per-hour association combines declining actual hours with approximately flat raw output. It is therefore described as labor-input compression, not productivity-enhancing adaptation. Land-specific trends make the central annual within coefficients imprecise. Nor can the accounting residual be renamed total factor productivity, gross capital per hour be renamed investment or automation, or aggregate workplace concentration be interpreted as a bilateral commuting probability. The accounts identify identities and associations. They do not identify technology.

The paper also addresses spatial dependence at the same regional level as the benchmark. Official BKG administrative boundaries dated 1 January 2026 generate a reproducible 16-Land queen-contiguity matrix and centroid distances. Window-specific residual Moran statistics are insignificant; the smallest permutation p-value is 0.301. Panel-Conley covariance estimates at the preregistered 300 km and 500 km cutoffs yield standard errors of 0.103 and 0.098 percentage points. This sensitivity does not widen the benchmark uncertainty. With only 16 Länder, however, it remains a bounded covariance diagnostic rather than a spatial model of spillovers.

The contribution is a German external-validity and estimand audit. It is not a claim that the German and cross-country designs estimate an identical parameter, and it is not a contest over the sign of one coefficient. The cross-country benchmark compares national growth over a long horizon. The German design compares regional windows, adds common period effects and initial conditions, and exposes within-region accounts. Their coefficients are comparable enough to motivate the audit but different enough that the estimand distinction must be part of the result.

This focus complements the recent economics-of-ageing literature. Firm- and worker-level studies ask how employee age relates to measured labor productivity (Börsch-Supan et al., 2021; Kim and Lee, 2023). Aggregate work studies entrepreneurship, employment, structural change, and productivity in ageing societies (Hoshi et al., 2025; André et al., 2026). Regional evidence connects ageing to German productivity growth and Japanese public-capital productivity (Bode et al., 2023; Kameda et al., 2026). Those papers examine substantive margins through which ageing may affect economic performance. The present paper asks a prior question: when an aggregate regional coefficient appears, which demographic and accounting variation has produced it?

The analysis proceeds in five steps. First, it aligns Hayashi's demographic ratio and economic denominator with German regional data. Second, it separates period-average composition, observed demographic change, and predetermined cohort pressure. Third, it tests period and geographic slope equality directly rather than inferring instability from separate confidence intervals. Fourth, it decomposes output per resident into output, hours, workplace employment, resident employment, and population on common annual samples. Fifth, it triangulates small-cluster and spatial uncertainty while keeping the interpretation explicitly noncausal.

The resulting conclusion is bounded but economically useful. The full German benchmark is negative and substantial, yet it is not a portable estimate of adjustment to aging. Its magnitude attenuates after 2000 and in western comparisons; formal pooled tests remain underpowered. Annual output-per-hour associations primarily reflect fewer hours in the point estimates, not output expansion. Regional sorting, migration, commuting, convergence, and shared shocks remain alternative explanations. German regional evidence therefore supports an estimand warning, not a causal law linking aging to growth.

SECTION 2

Why Composition and Adjustment Differ

2.1 Composition, Transition, and Pressure

Let P_{rat} denote the population of region r, age a, in year t. The annual log old-to-young ratio is

$$a_{rt} = \log\left(\frac{\sum_{a=50}^{79} P_{rat}}{\sum_{a=20}^{49} P_{rat}}\right). \tag{2.1}$$

For a horizon of H years, the period-average level uses exactly H annual observations:

$$\bar{a}_{r,t,H} = \frac{1}{H}\sum_{s=0}^{H-1} a_{r,t+s}. \tag{2.2}$$

This is the logarithm of the geometric mean of annual ratios. It is not the log ratio of arithmetic-average population stocks.

The frozen realized-aging field compares adjacent equal-length period averages:

$$\Delta\bar{a}_{r,t,H} = \bar{a}_{r,t,H} - \bar{a}_{r,t-H,H}. \tag{2.3}$$

It is not the endpoint change $a_{r,t+H} - a_{rt}$. The two objects load different years, smooth measurement error differently, and can rank regions differently. Paper 0.9.3 retains Equation (2.3) for the original realized-transition grid because it matches the governed field `observed_aging_change`. Endpoint change is estimated only inside the separate conditional-convergence system.

A third object is predetermined cohort-transition pressure. For horizon h, a leave-focal-region-out stock-transition factor $\widehat{\tau}_{-r,t,h,a}$ applies historical donor transitions to the focal region's baseline cohorts. The notation is deliberate. The factor can contain mortality, donor migration, census revisions, and unresolved boundary changes; it is neither a pure survival probability nor automatically an excluded instrument.

2.2 The Accounting Backbone

Let Q^W denote real workplace GDP, $\mathcal{H}^W$ workplace actual hours, E^W workplace employment, E^R resident employment, and N^R residents aged 20–79. The current data support the exact identity

$$\frac{Q^W}{N^R} = \frac{Q^W}{\mathcal{H}^W} \times \frac{\mathcal{H}^W}{E^W} \times \frac{E^W}{N^R}. \tag{2.4}$$

Its components are hourly labor productivity, hours per workplace-employed person, and workplace employment relative to the resident broad-adult population. The final term is a mixed utilization-and-commuting margin. Once resident employment is observed on a common concept and geography, it can be split further:

$$\frac{E^W}{N^R} = \frac{E^W}{E^R} \times \frac{E^R}{N^R}. \tag{2.5}$$

The annual accounting extension admits resident employment for all 16 Länder from 2000 through 2021 at the published precision of the compact ETR revision. The annual accounting system can therefore split the final term in Equation (2.4). The first factor is labelled workplace concentration; it is not an individual commuting rate. The second is resident-employment intensity; its numerator covers all resident employment and is not restricted to residents aged 20–79.

Taking log changes makes Equation (2.4) additive. If every component is estimated on the same observations with the same demographic regressor, predetermined controls, fixed effects, weights, and resampling sequence, linearity also implies

$$\widehat{\beta}_{Q/N} = \widehat{\beta}_{Q/\mathcal{H}} + \widehat{\beta}_{\mathcal{H}/E^W} + \widehat{\beta}_{E^W/N}. \tag{2.6}$$

The restriction does not apply to the original headline coefficients because their lagged outcome controls differ by denominator. A subtraction of those coefficients is therefore a diagnostic contrast, not a common-X mechanism result. A separate system holds observations, regressors, effects, weights, parameter order, and bootstrap draws fixed across all seven accounting outcomes.

The growth-accounting layer is

$$g^{Q/\mathcal{H}} = \alpha g^{K/\mathcal{H}} + g^R, \tag{2.7}$$

where K is the frozen gross-fixed-asset stock chain index, α is the endpoint-average capital share from the adjusted labor-share rule, and g^R is the accounting residual. Moreover,

$$g^{K/\mathcal{H}} = g^{K/N^R} - g^{\mathcal{H}/N^R}. \tag{2.8}$$

Thus, capital deepening can rise because capital grows, hours fall, or both. The residual can contain utilization, labor quality, allocation, markups, sector composition, and measurement error. The accounting backbone does not identify causal mediation, technology, or automation.

2.3 Regional Equilibrium and Testable Propositions

A parsimonious effective-labor block is sufficient to organize the regional setting:

$$L^e_{rt} = e^y_{rt} N^y_{rt} + e^o_{rt} N^o_{rt} + M_{rt}, \tag{2.9}$$
$$Q^W_{rt} = A_{rt} K^\alpha_{rt} (L^e_{rt})^{1-\alpha}, \tag{2.10}$$

where e^y and e^o collect age-specific participation, employment, hours, education, experience, occupation, and task mix. The term M_{rt} represents labor supplied across residence boundaries. In a spatial equilibrium, productivity and job opportunities can attract younger residents and commuters, lowering the observed old-to-young ratio while raising workplace output. The same negative composition coefficient is therefore compatible with a direct demographic response, selective sorting, commuting, or persistent regional type (Mitze and Reinkowski, 2011; Monte et al., 2018).

The framework yields four testable propositions. First, if aging pressure operates mainly through utilization, its association should be more negative for employment and hours than for output per hour. Second, a capital-deepening association should be decomposed into capital and hours before it is interpreted as investment. Third, E^W/N^R should be split into workplace concentration and resident-employment intensity before it is assigned to either margin. Finally, predetermined cohort pressure should be separated from contemporaneous composition before selective sorting is invoked as an explanation. The common-sample accounting systems address the first two propositions. The annual identity addresses the third, while the cohort design supplies the fourth as a reduced form. None identifies a causal mechanism.

2.4 From Hayashi to the German Design

The August 2026 paper supplies both a canonical average-level equation and a transformed aging-speed projection. Table 1 prevents the two from acquiring one ambiguous label.

Table 1 Version and estimand crosswalk

Source object	Outcome	Demographic object	German role
Hayashi (2026), eq. (2)	Growth in GDP per population aged 20–79	Period-average log ratio	Canonical benchmark
Hayashi (2026), eq. (8)	Momentum-adjusted growth	Realized aging speed	Horizon adaptation
Hayashi (2026), eq. (17)	Projected growth differential	Forecast-vintage aging speed	Information-set benchmark
Acemoglu et al. (2026), App. C	Growth in GDP per capita	Average level versus end-point change	Estimand audit

The complete machine-readable mapping is stored in *(see replication package)*.

Source: Hayashi (2026), supplied August 2026 paper; Acemoglu et al. (2026).

Sample: Source equations retained in the foundation crosswalk.

The projection design introduces the convergence-weighted momentum coefficient $m(H,\kappa) = \exp(-H\kappa)$. The prespecified grid is $\kappa \in \{0.02, 0.04, 0.06\}$. An authentic 12th coordinated projection package published in 2009 has base 31 December 2008, variants W1/W2, and annual totals for all 16 Länder in 2011 and 2021. Its exact-age sheets, however, cover coordinated aggregate groups rather than individual Länder; individual-Land ages 20–49 and 50–79 cannot be reconstructed from the surviving broad bands. The live GENESIS table now contains the 16th projection and cannot be backdated. Forecast-vintage momentum therefore remains unavailable. The paper instead constructs a distinct 2001–2011 cohort-predicted transition from baseline cohorts and leave-focal-Land-out historical transitions. That exposure does not recreate the missing 2009 forecast information set and is not interpreted as a forecast or instrument.

2.5 How the Framework Changes the Literature

2.5.1 Age composition is not one treatment

The empirical literature does not imply one mechanical mapping from age structure to economic performance. Cross-country work relates population composition to growth and productivity while emphasizing age profiles, initial conditions, and the horizon over which adjustment occurs (Lindh and Malmberg, 1999; Feyrer, 2007). Regional evidence adds another complication: demographic composition, human capital, sector structure, and spatial sorting vary together, and the resulting associations are heterogeneous across places (Daniele et al., 2019).

German studies make that regional problem concrete. Brunow and Hirte (2006) study age structure and regional growth; Brunow and Hirte (2009) connect workforce age and human capital to productivity across labor-market regions; Bönte et al. (2009) examine entrepreneurship; and Gregory and Patuelli (2015) document persistent spatial polarization in demographic and economic conditions. The closest outcome study, Bode et al. (2023), relates workforce aging to county productivity growth and finds an aggregate relationship that is negative but imprecise. These papers do not estimate one common object. Workforce age differs from the resident old-to-young ratio; administrative counties differ from functional labor markets; and GDP per resident aged 20–79 differs from GDP per employed person. The present paper constructs Hayashi's exact resident ratio for the Länder and reports both economic denominators under one registered estimand map.

2.5.2 Migration makes regional demography endogenous

Regional age structure is partly an equilibrium outcome. Age-specific migration responds to labor-market conditions in Germany (Mitze and Reinkowski, 2011); a productive region can attract younger adults, expand the denominator of Equation (2.1), and appear demographically younger for reasons that also predict subsequent output. This makes a negative level association compatible with selective sorting rather than a demographic mechanism.

Cohort-shift designs offer one way to remove contemporaneous local migration mechanically. Maestas et al. (2023) use baseline cohort structure and national transitions to study aging across U.S. states. Yet baseline exposure can proxy for historical fertility, industry, education, reunification, or prior migration. The shift-share literature therefore requires explicit exposure diagnostics, leave-out construction, and an account of which shocks carry identifying content (Goldsmith-Pinkham et al., 2020; Borusyak et al., 2022; Adão et al., 2019). The German cohort design executes those diagnostics for one eligible window. The leave-out construction passes its computational checks, but the exposure has little cross-Land spread and its first-stage and reduced-form estimates are imprecise. The result remains a predetermined-exposure reduced-form association, not an instrumental-variable design.

2.5.3 Adjustment can run through labor and technology

The mechanism literature also cautions against reading a demographic coefficient as a single productivity channel. Labor scarcity can reduce employment and entrepreneurship, but it can also induce capital deepening, automation, and organizational adjustment (Acemoglu and Restrepo, 2017, 2022; Hoshi et al., 2025). German robot exposure shows that technology shocks reallocate tasks and employment across workers and regions rather than operating as a frictionless productivity increment (Dauth et al., 2021). The Japanese discussion has likewise moved below the national aggregate and toward regional migration and adjustment (Hoshi, 2026).

This literature motivates the paper's denominator discipline. GDP per resident aged 20–79 combines workplace output, hours, workplace concentration, resident-employment intensity, and the size of the resident broad-adult population. Divergence between denominators is therefore informative, not an invitation to choose the more convenient outcome. The accounting systems observe output per hour and labor aggregates, but they do not observe bilateral commuting, automation, ICT adoption, or causal mediation. The stronger technology literature likewise requires direct adoption or numerator evidence rather than a convenient ratio (Acemoglu and Restrepo, 2022; Dauth et al., 2021).

Few-cluster inference imposes a related discipline. WCR procedures, influence diagnostics, and complete specification maps are useful precisely because 16 Länder do not support casual asymptotics (MacKinnon et al., 2023; Canay et al., 2021). Yet separate intervals still do not test a coefficient difference, and many correlated targets require a prespecified familywise procedure (Romano and Wolf, 2005). Paper 0.9.3 therefore reports directly imposed and consistently studentized WCR-t tests, retains cross-sample and cross-model comparisons as descriptive diagnostics, and treats every East restriction as five-East-cluster inference.

Taken together, the literature makes the contribution narrower than a new regional-aging correlation and more demanding than a sign comparison. The paper asks whether the same demographic pattern survives a change in estimand, denominator, sample, weighting rule, and horizon. Germany holds many national institutions constant, but it does not resolve sorting or identify a technology channel by itself. The complete specification map is therefore evidence about where the association changes, not a menu from which one preferred coefficient is selected.

SECTION 3

Data and the German Regional Setting

The empirical core is a balanced annual panel of all 16 German federal states (Bundesländer). The unit of observation is a Land-year, not Germany as a whole and not a county. The panel runs from 1991 through 2024 and contains $16 \times 34 = 544$ observations. Berlin, Bremen, and Hamburg remain in the main sample. Because their city-state boundaries make workplace output especially sensitive to commuting, every result is also reported for the 13 non-city Länder.

3.1 Sources and Measurement

Three official source families supply the core outcome panel. Destatis GENESIS table 12411-0012 provides year-end resident population by Land and exact age. VGRdL provides nominal and chain-linked real GDP by place of work. ETR provides employed persons and actual hours by place of work. A compact ETR revision also provides resident employment at one-decimal published precision for 2000–2021. The accounting extension additionally uses the VGRdL gross-fixed-asset stock chain index and nominal manufacturing and total gross value added. Every object retains its literal statistical concept; in particular, resident employment is not age matched and a capital-stock index is not capital services. Official BKG VG250 boundaries dated 1 January 2026 supply the Länder geometry, queen-contiguity weights, and centroid distances used only for the spatial covariance sensitivity.

3.2 Recovered Historical Inputs and Their Boundary

The evidence-recovery release adds 32 manufacturing-share controls: one same-vintage VGRdL ratio for each Land in 2000 and 2010. Each ratio retains the literal manufacturing numerator and total-GVA denominator source cells. The control is therefore fixed before its corresponding ten-year outcome window; no contemporaneous sector composition is inserted after the outcome is known.

Historical exact-age recovery is deliberately partial. Across 496 Land-years from 1981 through 2011, 451 contain complete single-age observations from 0 to 89, yielding 40,590 supported cells. The remaining 45 eastern Land-years in 1981–1989 contain 4,050 unsupported cells. They are recorded as unavailable, not converted from zeros and not backfilled from totals, births, or later age distributions. Consequently, the complete 1991–2024 core panel remains available, while an all-Länder exact-age design beginning in 1981 remains blocked. More historical rows do not remove a geographical hole.

The annual population series also crosses census and revision regimes. The 2011 census affects the population basis inside the main windows, while the 2022 census anchors a later retrospective Kreis bridge and the 2022–2024 Länder observations. The frozen annual and window products retain revision flags rather than smoothing the breaks away. Paper 0.9.3 treats those flags as measurement boundaries. The stable-vintage recovery obtains 384 of 672 required age cells but finds no compatible 2001–2009 detail for the remaining 288. The stable-Census branch therefore emits no exposure or estimate. The 2022 census change lies outside the registered 2001–2021 windows.

The demographic variable is the number of residents aged 50–79 divided by the number aged 20–49. A ratio of one means that the two groups are equally large. The empirical models use its natural logarithm, either as an annual level, as the average log ratio within a period, or as the realized change between two periods. This is

Table 2 Observed Länder data used in the empirical analysis

Object	Official source and concept	Construction	Coverage
Population	Destatis, resident population at year-end	Exact ages summed to 0–19, 20–49, 50–79, and 20–79	16 Länder, 1991–2024
Real GDP	VGRdL, workplace GDP	Official chain-volume index anchored to nominal GDP in 2020	16 Länder, 1991–2024
Workplace employment	ETR, workplace employed persons	Official employment count by place of work; not hours	16 Länder, 1991–2024
Resident employment	ETR, resident employed persons	Compact revision reconstructed and verified at published one-decimal precision	16 Länder, 2000–2021
Actual hours	ETR, workplace hours worked	Annual hours used to construct output per actual hour	Registered 2001–2021 windows
Capital stock	VGRdL, gross fixed assets	Chain index used as a stock proxy; not capital services	Registered 2001–2021 windows
Sector control	VGRdL, nominal gross value added	Manufacturing GVA divided by total GVA in 2000 and 2010	32 Land-baseline values
Spatial geometry	BKG, VG250 administrative areas	Canonical Land-area records ($GF = 4$); derived queen weights and centroids	16 Länder, 1 January 2026

Source: Destatis GENESIS 12411-0012, VGRdL Länder results, and ETR Länder results; frozen source identities are recorded in the evidence-recovery release.
Sample: Balanced core panel of 544 Land-year observations plus the registered accounting windows and predetermined controls.

an age-composition measure. It is not the usual old-age dependency ratio, which would use a retirement-age numerator and a different denominator.

3.3 Residence and Workplace Denominators

Both headline outcomes use the same chain-linked real-GDP numerator. They differ only in the denominator:

$$y^N_{rt} = \frac{\text{real workplace GDP}_{rt}}{\text{resident population aged 20–79}_{rt}}, \tag{3.1}$$

$$y^E_{rt} = \frac{\text{real workplace GDP}_{rt}}{\text{workplace employed persons}_{rt}}. \tag{3.2}$$

GDP per resident aged 20–79, y^N_{rt}, captures real workplace output relative to a broad resident population. It can change because output, employment, commuting, or the resident denominator changes. It is neither GDP per employed person nor a residence-based welfare measure.

Hayashi describes ages 20–79 as working age, but this is not an official German labor-force category. The band includes people above statutory retirement age and other residents who are not employed. Paper 0.9.3 therefore uses "resident aged 20–79" or "broad adult population" except when the inherited terminology itself is under discussion.[1]

GDP per employed person, y^E_{rt}, divides the same output by realized workplace employment. It is a labor-productivity proxy, although it does not adjust for hours worked. Comparing the two outcomes reveals whether a regional pattern is tied to output per potential worker or remains after conditioning on the number of people actually employed.

For 2000–2021, resident employment makes the denominator bridge in Equation (2.5) observable. The first ratio records workplace concentration and the second resident-employment intensity. Aggregate workplace con-

[1]This terminology change does not alter the frozen variable or any estimate. It prevents a demographic denominator from being mistaken for an observed labor-force count.

centration is not a bilateral flow, an individual commuting rate, or a migration measure. Resident-employment intensity is likewise not an age-matched employment or participation rate.

The resident-employment source is publicly retrievable but not redistributed in the package. The official ETR series page is `https://www.statistikportal.de/de/etr/ergebnisse/erwerbstaetige-personen/erwerbstaetige-jahresdurchschnitt-0`. The compact workbook is available at `https://www.statistikportal.de/sites/default/files/2025-11/Erwerbst%C3%A4tige%20am%20Wohnort%20in%20Deutschland%20nach%20L%C3%A4ndern.xlsx`; the combined workbook is `https://www.statistikportal.de/sites/default/files/2025-11/Inlaenderrechnung_2024_Jahr%201991-2024.xlsx`; and the Revision-2024 methods handbook is `https://www.statistikportal.de/sites/default/files/2026-05/Methodenhandbuch_2026.pdf`. At retrieval, their SHA-256 digests began `f5f755d65ffc`, `00bac93621bd`, and `efdda7c26522`, respectively; the replication instructions record the complete 64-character values. The two workbooks agree for all 544 annual residence cells from 1991 through 2024. Their artifact-specific notice permits attributed noncommercial copying but requires consent for electronic redistribution and limits published absolute values to one decimal thousand persons. The package therefore records retrieval, selectors, transformations, and hashes while requiring researchers to obtain the official bytes themselves.

Table 3 Basic descriptive statistics for the annual Länder panel

Variable	Unit	Mean	Std. dev.	Minimum	Maximum
Residents aged 50–79 / residents aged 20–49	ratio	0.879	0.192	0.587	1.373
Real GDP per resident aged 20–79	2020 euros	47,449	14,564	16,450	94,781
Real GDP per employed person	2020 euros	70,540	13,080	24,823	106,640

Source: Frozen Länder pre-analysis panel; unweighted calculations from registered annual values.

Sample: N = 544 Land-year observations, 16 Länder, 1991–2024. Minima and maxima pool regional and time variation.

3.4 What the Regional Profiles Show

The profiles reveal a distribution, not one representative German region. By 2024, the observed old-to-young ratio ranges from 0.73 in Hamburg to 1.32 in Mecklenburg-Vorpommern. The endpoints in Figure 1 also show that this dispersion emerged along different paths rather than through one common demographic shift.

Among the five eastern non-city Länder, the mean increase in the old-to-young ratio is 0.59, compared with 0.37 among the eight western non-city Länder. Their observed 1991–2021 growth averages are 2.66 versus 0.93 percentage points for GDP per resident aged 20–79 and 2.71 versus 0.47 for GDP per employed person. These contrasts span post-unification convergence and large differences in initial economic levels. They are descriptive, not causal: faster growth in the east cannot be assigned to its simultaneous demographic change.

Among the city states, Hamburg records the highest 2024 GDP per resident aged 20–79, at 94.0 thousand euros; its GDP per employed person is 96.6 thousand euros. The comparison is especially sensitive to commuting because workplace output is paired with a resident denominator in the first measure. A high level therefore describes the location of production relative to residents, not a city-state demographic premium.

The two main scatter panels contain one aligned 1991–2021 observation for each of the 16 Länder in Figure 2, while the separate appendix edition retains all 544 annual observations for each displayed series. The main panels expose the raw long-run association without allowing common time trends to masquerade as additional regional evidence. The appendix then shows how each Land reached its endpoints. Neither display isolates an effect of aging or fits a new model.

3.5 Kreis Evidence and Remaining Limits

The Kreis layer crosses a descriptive, but not an estimative, boundary. It contains 396 stable Kreise and kreisfreie Städte from 2012 through 2021 and supports levels, changes, ranks, distributional overlap, and within-Land variance shares. Output remains current price, hours remain standard hours, and the paper does not estimate a Kreis spatial model. The results section reports the decision-relevant distributional evidence. The separate

Table 4 Regional profiles: observed demography, output levels, and growth

Land	Group	O/Y 1991	O/Y 2024	Δ O/Y	GDP/A	GDP/E	Growth A	Growth E
Brandenburg	East	0.69	1.28	0.59	42.3	69.3	2.41	2.69
Mecklenburg-Vorpommern	East	0.65	1.32	0.68	41.8	65.0	2.32	2.32
Sachsen	East	0.78	1.13	0.34	45.8	65.1	2.85	2.72
Sachsen-Anhalt	East	0.74	1.32	0.58	39.7	63.5	2.63	2.66
Thüringen	East	0.72	1.28	0.57	41.3	63.4	3.11	3.16
Baden-Württemberg	West	0.64	0.96	0.32	66.2	85.7	1.01	0.69
Bayern	West	0.66	0.97	0.31	67.8	85.0	1.31	0.93
Hessen	West	0.67	0.98	0.31	68.0	87.8	0.85	0.44
Niedersachsen	West	0.71	1.09	0.38	53.3	74.4	0.93	0.34
Nordrhein-Westfalen	West	0.71	1.01	0.30	54.5	74.2	0.87	0.29
Rheinland-Pfalz	West	0.72	1.09	0.37	49.0	73.1	0.94	0.52
Saarland	West	0.74	1.16	0.42	45.8	67.0	0.87	0.22
Schleswig-Holstein	West	0.70	1.13	0.43	45.7	67.6	0.64	0.34
Berlin	City	0.59	0.73	0.14	64.8	81.5	1.20	0.76
Bremen	City	0.72	0.85	0.13	67.5	78.9	0.78	0.33
Hamburg	City	0.70	0.73	0.03	94.0	96.6	0.88	0.33

Source: Frozen Länder pre-analysis annual and long-window observations.
Sample: All 16 Länder. O/Y is residents aged 50–79 divided by residents aged 20–49; GDP levels are thousands of 2020 euros in 2024; growth is annualized 1991–2021 log growth in percentage points.

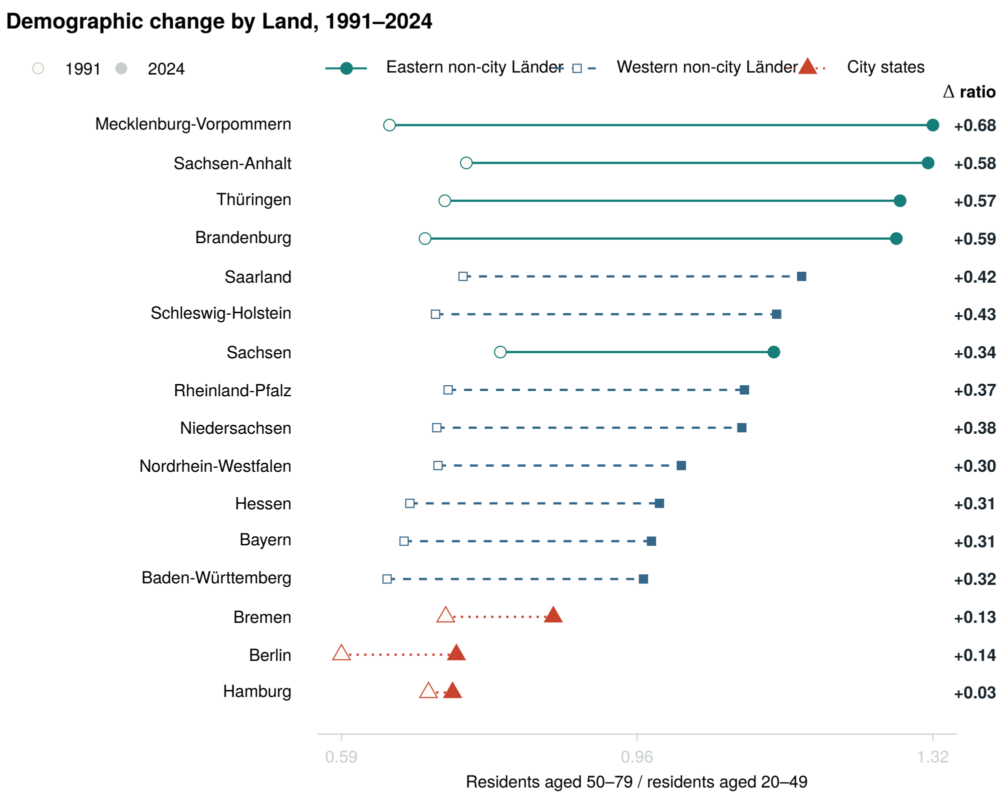

Figure 1 Regional demographic change. Each row connects a Land's observed 1991 and 2024 old-to-young ratios; rows are ordered by the 2024 endpoint and the right column reports the signed change.

Source: Frozen Länder pre-analysis annual panel.

Sample: All 16 Länder. Open and filled endpoints identify 1991 and 2024; shape and line style identify regional groups without relying on color.

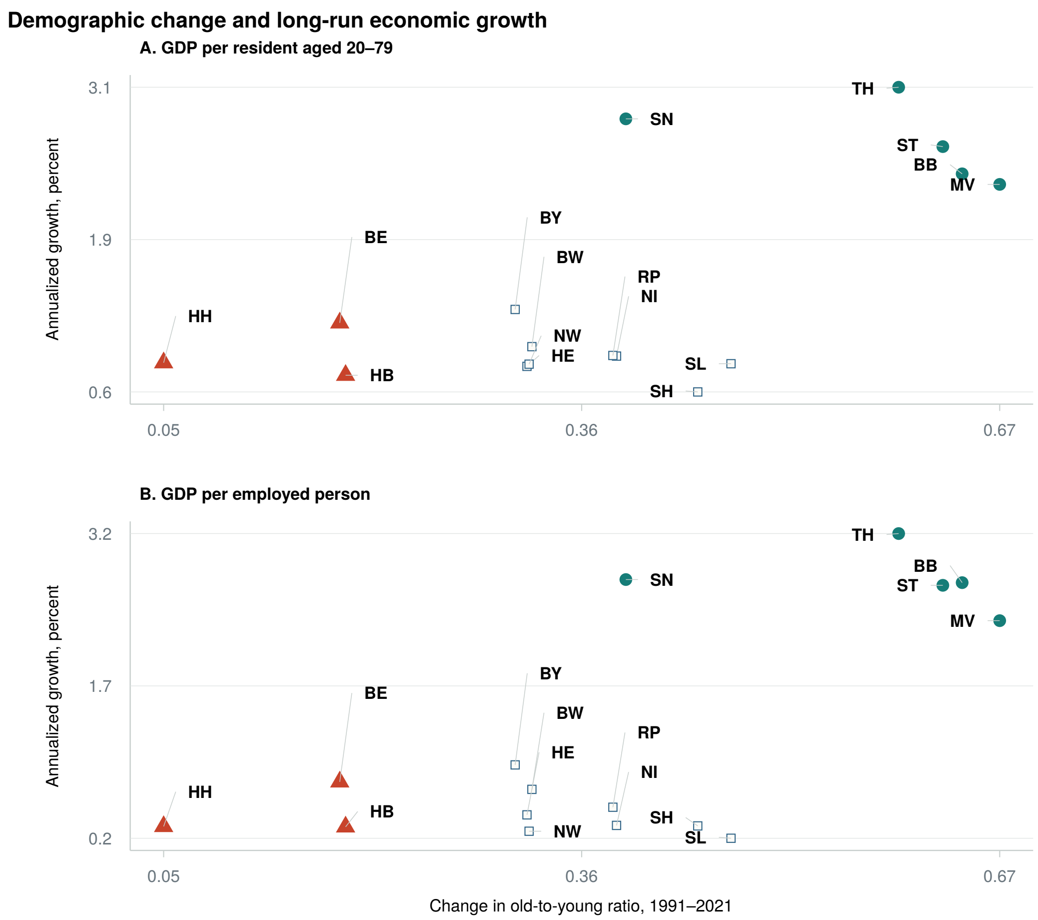

Figure 2 Observed matched-window changes in demography and economic growth. Each panel contains one 1991–2021 observation per Land on a shared demographic-change scale. Points and labels are observed values, not fitted relationships.

Source: Frozen Länder pre-analysis annual and long-window panels.

Sample: All 16 Länder. GDP is workplace based; the ages-20–79 denominator is residence based. Marker shape and direct labels preserve identity without relying on color.

appendix retains the full descriptive table, source corrections, and conditions under which smaller-area analysis could be extended.

Patent, broadband, automation, bilateral commuting, functional geography, stable-Census, and Kreis reduced-form evidence remain unavailable or incompatible with the current estimands. None is recoded as zero or used to support a mechanism claim. The new BKG product resolves Länder covariance for the headline regression; it does not manufacture a 396-Kreis historical topology or a structural spillover design.

3.6 From Annual Levels to Growth Observations

The growth outcomes are annualized log changes between a window's endpoints, as defined formally in Equation (4.1). Thus, 2.0 means two percentage points of annualized log growth (approximately 2% per year); it does not mean a two-percent level difference. The main adjacent-window data contain three decade comparisons for each Land: 1991–2001, 2001–2011, and 2011–2021. The matched estimand audit uses the latter two decades because they also have a preceding period from which realized aging can be calculated. Two additional comparisons use one observation per Land: the 1991–2021 long window and the 2009–2019 pre-pandemic window.

The 2011–2021 comparison spans the pandemic years and is retained in the main design; the 2009–2019 window shows the corresponding pre-pandemic boundary. The 2022–2024 observations enter the annual fixed-effects models but not the registered endpoint-growth windows, which stop in 2021.

Table 5 Observed annualized real-GDP growth by Land-window

Registered window	Endpoints	N	GDP / ages 20–79	GDP / employed
First adjacent decade	1991–2001	16	2.10 [0.16, 5.63]	2.48 [−0.08, 7.16]
Second adjacent decade	2001–2011	16	1.20 [0.43, 1.90]	0.66 [−0.12, 1.28]
Third adjacent decade	2011–2021	16	1.12 [0.04, 1.88]	0.38 [−0.35, 1.10]
Long comparison	1991–2021	16	1.47 [0.64, 3.11]	1.17 [0.22, 3.16]
Pre-pandemic	2009–2019	16	1.96 [1.09, 2.94]	0.91 [0.03, 1.69]

Source: Frozen Länder pre-analysis windows; unweighted means with minima and maxima in brackets.

Sample: Values are observed annualized growth rates in percent, not regression coefficients.

These observed growth rates must be separated from the estimates reported in Section 5. A result such as −0.317 percentage points is not a Land's growth rate. It is the estimated difference in annual growth associated with a 10% higher old-to-young ratio, after the model's controls and rescaling. The annual fixed-effects rows use euros rather than percentage points because they relate demographic composition to annual GDP levels, not endpoint growth.

3.7 Boundary of the Current Evidence

The current evidence layer is therefore narrower than the full research architecture. The Kreis product supports frozen descriptive profiles and distributions, but Specs 022–024 establish that no regression or spatial claim is currently eligible. A national deflator can remove common inflation from current-price output, but it cannot recover relative regional prices.

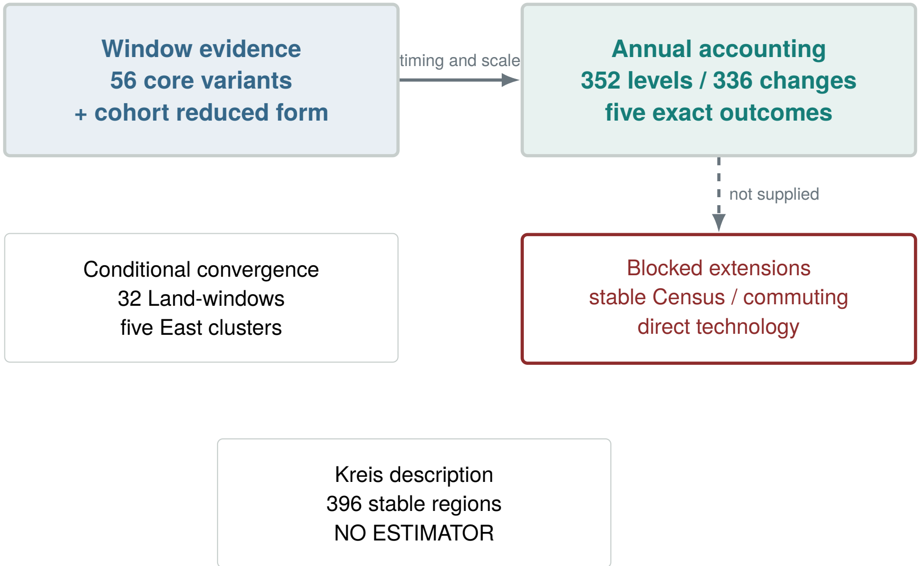

Figure 3 Evidence architecture through the 0.9.1 integration

Source: Frozen Länder release chain through Specs 041, 042A, 042B, 042C, and 044.

Sample: Window, annual, convergence, blocked-extension, and Kreis claims remain separate evidence classes. Arrows denote design progression, not causality.

Table 6 Measurement contracts by regional layer

Dimension	Länder	Kreise / kreisfreie Städte
Demography	Exact single ages	Exact ages with census and territorial controls
Output	Nominal and chain-linked real	Current-price workplace output
Labor input	Employed persons and actual hours	Employed persons and standard hours
Capital	Official investment and capital-stock series	No official capital stock
Primary claim	Real growth and bounded accounting associations	Description; reduced-form association blocked
Prohibited inference	Direct technology adoption, causality, or large-sample certainty with 16 clusters	Official real output or TFP

Source: GENESIS, Regionalstatistik, VGRdL, and the frozen source-release registry.

Sample: Länder contracts are populated through 2024 for the real-output core and through the registered accounting windows; the Kreis descriptive contract is populated and the reduced-form contract is blocked.

SECTION 4

Empirical Design and Identification Ladder

4.1 Growth and Timing

For any positive outcome k_{rt}, annualized endpoint growth is

$$g^k_{r,t,H} = \frac{100}{H}\left[\log k_{r,t+H} - \log k_{rt}\right]. \tag{4.1}$$

The period-average composition $\bar{a}_{r,t,H}$ is defined in Equation (2.2). Realized aging compares adjacent equal-length period averages as defined in Equation (2.3). It is not the endpoint change $a_{r,t+H} - a_{rt}$, which remains a separate registered sensitivity without a frozen estimate.

The canonical adjacent-window equation for outcome k is

$$g^k_{r,w} = \alpha^k + \beta^k \bar{a}_{r,w} + \gamma^k \log k_{r,w,0} + \delta^k \bar{c}_{r,w} + \lambda^k_w + u^k_{r,w}, \tag{4.2}$$

where $\log k_{r,w,0}$ is the start-of-window outcome level and $\bar{c}_{r,w}$ is the period-average log child-dependency ratio. The primary sample contains 48 observations: 16 Länder across 1991–2001, 2001–2011, and 2011–2021. Window effects absorb common decade shocks.

The matched transition audit uses the latter two windows and replaces $\bar{a}_{r,w}$ with $\Delta\bar{a}_{r,w}$:

$$g^k_{r,w} = \tilde{\alpha}^k + \tilde{\beta}^k \Delta\bar{a}_{r,w} + \tilde{\gamma}^k \log k_{r,w,0} + \tilde{\delta}^k \bar{c}_{r,w} + \tilde{\lambda}^k_w + \tilde{u}^k_{r,w}. \tag{4.3}$$

The 32 observations are identical to the corresponding matched composition runs. Sample alignment therefore removes one source of disagreement, but the two treatments remain different empirical objects.

Because this restriction removes the 1991–2001 window while preserving the headline treatment, outcome, controls, and inference contract, it is also the direct regime check for the central historical concern. The comparison is not between an old and a newly selected model. Both equations were registered in the original design grid, and both remain reported regardless of their result.

Paper 0.9.3 adds one pooled test of the period and geographic differences:

$$\begin{aligned} g^{Q/N}_{rw} &= \alpha + \beta\bar{a}_{rw} + \kappa_P(\bar{a}_{rw} \times P_w) + \kappa_E(\bar{a}_{rw} \times E_r) + \gamma \log(Q^W/N^R)_{r,w,0} + \delta\bar{c}_{rw} \\ &\quad + \lambda_w + \xi_w E_r + u_{rw}, \end{aligned} \tag{4.4}$$

where P_w equals one for the two post-2000 windows and E_r identifies the five eastern non-city Länder. Berlin remains in the nationwide descriptive estimates but is omitted from Equation (4.4); its divided-city history is not assigned to either non-city group. The direct period restriction is $H_0 : \kappa_P = 0$, the geographic restriction is $H_0 : \kappa_E = 0$, and the joint test imposes both. East-by-window intercepts ξ_w prevent average East-level differences in each decade from being forced into the slope interactions.

The annual fixed-effects audit estimates levels rather than growth:

$$y^k_{rt} = \mu^k_r + \tau^k_t + \beta^k_{FE} a_{rt} + \delta^k_{FE} c_{rt} + \varepsilon^k_{rt}, \tag{4.5}$$

where y^k_{rt} is measured in 2020 euros, c_{rt} is the annual log child-dependency ratio, and Land and year effects use all 544 observations. This is a within-Land level association, not the endpoint-growth equation.

4.2 Predetermined Cohort-Transition Reduced Form

The paper executes one timing-eligible cohort design for 2001–2011. Let Z^C_r denote the log change in the old-to-young ratio obtained by applying leave-focal-Land-out historical cohort transitions to Land r's 2001 age stocks. The reduced form is

$$g^{Q/N}_{r,2001,10} = \alpha + \pi Z^C_r + \rho \log(Q^W/N^R)_{r,2001} + \psi_y \log N^y_{r,2001} + \psi_o \log N^o_{r,2001} + u_r. \tag{4.6}$$

The sample contains one observation for each of the 16 Länder. Unweighted estimation is primary; start-population weighting is secondary. The donor construction contains 960 leave-focal-Land-out transitions, with a maximum donor share of 0.285 and a minimum effective donor count of 6.82. First-stage relevance and 1991–2001 pre-outcome placebo equations are diagnostics, not selection gates. The transition factors may contain mortality, migration, revision, and boundary components. Accordingly, Z_r^C is predetermined exposure and Equation (4.6) is a reduced form, not an instrumental-variable equation or a migration-free treatment.

4.3 Annual Within, Between, and First-Difference System

The admitted resident-employment series supports a five-outcome annual system from 2000 through 2021. For log accounting outcome m_{rt}^k, the Mundlak level equation is

$$m_{rt}^k = \alpha^k + \beta_W^k (a_{rt} - \bar{a}_r) + \beta_B^k \bar{a}_r + \tau_t^k + u_{rt}^k, \tag{4.7}$$

and the adjacent first-difference equation is

$$\Delta m_{rt}^k = \widetilde{\alpha}^k + \beta_\Delta^k \Delta a_{rt} + \widetilde{\tau}_t^k + v_{rt}^k. \tag{4.8}$$

The outcomes are $\log(Q^W/N^R)$, $\log(Q^W/\mathcal{H}^W)$, $\log(\mathcal{H}^W/E^W)$, $\log(E^W/E^R)$, and $\log(E^R/N^R)$. Equation (4.7) uses 352 Land-years; equation (4.8) uses 336 adjacent annual changes. The five outcomes reconcile exactly at the observation and coefficient levels. Direct tests impose $\beta_W^k = \beta_B^k$ within each system. Paper 0.9.3 corrects the direct-test wrapper so the estimate, standard error, observed statistic, p-value, confidence interval, and bootstrap draws all use the same restriction $R = e_W - e_B$. The five properly studentized direct-null draw series form one single-step maximum absolute statistic. A pointwise interval and a familywise decision therefore answer different questions.

4.4 Conditional Convergence and East Timing

The conditional-convergence system uses the fixed 32-row panel formed by the 2001–2011 and 2011–2021 windows. For each of the same five accounting outcomes, the unconditional model relates annualized growth to centered initial log GDP per resident and a late-window indicator. The conditional model adds centered baseline aging, centered endpoint aging transition, an indicator for the five eastern non-city Länder, and its interaction with the late window:

$$\begin{aligned} g_{rw}^k = \alpha^k &+ \rho^k \widetilde{\log(Q^W/N^R)}_{r,w,0} + \theta^k \widetilde{a}_{r,w,0} + \phi^k \widetilde{\Delta a}_{rw} \\ &+ \eta^k E_r + \lambda^k L_w + \omega^k (E_r \times L_w) + u_{rw}^k. \end{aligned} \tag{4.9}$$

The direct-test layer reports initial-level slopes, aging coefficients, East coefficients in each window, the East window shift, the common late-window shift, and their sum. Each restriction belongs to a five-outcome max-t family. Unweighted estimation is primary and start-population weighting is secondary. East-related inference rests on five East clusters. Comparisons between the unconditional and conditional models are descriptive because no joint estimator supplies cross-model covariance.

4.5 Weights, Scaling, and Inference

Unweighted estimation is primary. The secondary window weight is resident population aged 20–79 at the start of the outcome window; the secondary annual weight is the contemporaneous population aged 20–79. Weights are normalized to mean one within each run. Every model is also repeated after jointly excluding Berlin, Bremen, and Hamburg.

Treatment coefficients are estimated per one log point. The original growth tables and figures multiply coefficients and intervals by $\log(1.1)$, so their displayed unit is percentage points of annual growth associated with a 10% higher old-to-young ratio. The cohort, annual, and convergence tables retain native coefficients: annual log percentage points per log point for endpoint growth, log outcome per log aging for annual levels and differences, and annual log percentage points for East/window restrictions.

All Länder inference clusters by Land and uses restricted, null-imposed WCR11 bootstrap-t inference with Rademacher weights. Preserved parent products retain their registered 9,999 draws. The Paper 0.9.3 headline, pooled interactions, annual direct tests, and deletion sensitivities enumerate all 2^G sign patterns. Exhaustive enumeration removes Monte Carlo error but does not create generally exact finite-sample inference. Leave-one-Land and observation-level influence diagnostics are reported for every admitted run (MacKinnon et al., 2023; Canay et al., 2021).

4.6 Submission-Revision Sensitivity Layer

Paper 0.9.3 adds diagnostic checks without changing the preserved primary models. The headline composition equation is re-estimated on the matched post-2000 sample, after deleting the five eastern non-city Länder, on the ten western Länder, and after adding East-by-window intercepts. Every one-Land deletion is also refitted and tested. Equation (4.4) then turns the two central descriptive comparisons into direct restrictions. Rademacher inference enumerates all 2^G sign patterns: 65,536 for the 16-Land models and 32,768 for the 15-Land formal heterogeneity system. Leave-one-cluster jackknife covariance (CRV3) with a t_{G-1} reference distribution supplies a deliberately different small-sample check. Neither method selects a preferred model.

The annual system receives three boundary tests. The first ends the level and first-difference samples in 2019. The second removes the adjacent change ending in 2011, where the population series crosses the census transition. The third adds Land-specific linear trends to the full level equation. The trend model is intentionally demanding: with 16 trends and year effects, it asks whether the within coefficient survives after each Land receives a separate smooth trajectory. All five primitive log levels—workplace output, actual hours, resident population aged 20–79, workplace employment, and resident employment—are estimated on the same annual matrix. Their differences must reproduce the five ratio coefficients within 10^{-10}.

The spatial sensitivity uses the official BKG VG250 Länder layer dated 1 January 2026. Canonical Land-area records ($GF = 4$) generate 16 valid geometries, row-standardized queen-contiguity weights, and great-circle distances between polygon centroids. For each window, Moran's I is computed from the central-model residuals with 99,999 fixed-seed permutations. The panel Conley covariance combines unrestricted serial dependence within each Land with same-window cross-Land covariance weighted by the Bartlett kernel $K(d/c) = 1 - d/c$ for $d < c$. The cutoffs $c = 300$ km and $c = 500$ km were registered before estimation. Reference p-values use t_{15}. With only 16 Länder, the procedure is a covariance sensitivity, not a structural spatial model (Conley, 1999).

4.7 Accounting Associations

For accounting outcome m and Land-window w, the frozen design is

$$m_{rw} = \alpha_m + \beta_m \Delta\bar{a}_{rw} + \rho_m m_{r,w-1} + \theta_m s_{r,w-1} + \lambda_{m,w} + u_{m,rw}, \tag{4.10}$$

where $m_{r,w-1}$ is the outcome-specific pre-window level and $s_{r,w-1}$ is the strictly predetermined manufacturing share. The outcomes are annualized growth in real output per actual hour, the gross-fixed-asset stock chain index per actual hour, and the growth-accounting residual defined by Equation (2.7). The pooled/full variant contains both 2001–2011 and 2011–2021 and includes a period effect. The early 2001–2011 window contains 16 observations and no period effect. Both are unweighted and use the same WCR11 policy.

Because $m_{r,w-1}$ differs across outcomes, these six regressions do not form the common-X system in Equation (2.6). They locate conditional covariation in accounting outcomes; they do not identify adaptation, causal mediation, compensating innovation, direct adoption, or an identified aging effect.

4.8 Common-X Accounting and Joint Inference

For each common-X accounting outcome M, paper 0.8.1 uses the annualized ten-year change

$$g^M_{r,w} = \frac{100}{H}\left[\log M_{r,w,H} - \log M_{r,w,0}\right], \qquad H = 10. \tag{4.11}$$

Displayed effects multiply the resulting coefficient by $\log(1.1)$. The common-X analysis estimates a separate seven-outcome system on the two matched 2001–2011 and 2011–2021 windows. Every equation uses the same 32 Land-window observations, adjacent-average aging treatment, initial manufacturing share, period effect, parameter order, Land clusters, and restricted WCR11 draws. The primary system is unweighted; population weighting and exclusion of the three city states are pre-specified secondary systems. This design removes outcome-specific lagged controls so that the coefficient restrictions in Equations (2.6) and (2.8) hold exactly. All 64 observation-level identities and all six coefficient identities reconcile within 10^{-10}.

The corrected inference layer leaves those accounting products unchanged. Each component or same-system linear contrast is represented as a transformed outcome and tested under its own null with studentized restricted WCR-t draws. Symmetric pointwise intervals invert the pointwise test. Single-step max-t p-values invert the displayed simultaneous intervals. Romano–Wolf stepdown p-values are reported separately because they do not invert the single-step intervals. Minimum detectable effects are simulation-calibrated from the same max-t draws at 80% power. Sample and window variants report literal coefficients as descriptive concentration diagnostics; no corrected cross-sample difference inference is claimed. The annual dynamics design is executed separately; the stable-Census sensitivity remains unavailable where compatible pre-2010 exact-age inputs do not exist.

4.9 Identification Ladder

Table 7 Demographic objects, evidence states, and claim ceilings

Object	Treatment	Current state	Interpretation
Period-average composition	$\bar{a}_{r,t,H}$	Frozen estimates	Conditional older-composition association
Baseline composition	a_{rt} at window start	Frozen conditional coefficient	Conditional regional-state association
Adjacent-average transition	$\Delta\bar{a}_{r,t,H}$	Frozen estimates	Realized transition correlation
Endpoint transition	$a_{r,t+H} - a_{rt}$	Frozen conditional coefficient	Conditional convergence association
Cohort transition	$Z_{r,t,h}(\widehat{\tau}_{-r})$	Frozen 2001–2011 reduced form	Predetermined-exposure association
Annual within/between change	$a_{rt} - \bar{a}_r$, $\bar{a}_r$, Δa_{rt}	Frozen five-outcome systems	Conditional accounting association
Kreis distribution	One fixed 2012–2021 span	Frozen descriptive aggregation	Within-Land heterogeneity; no estimator
Capital numerator and hours	Baseline composition or endpoint transition	Frozen three-member identity families	Stock-hours association; no mechanism
Common-X components	Exact identities in Equation (2.4)	Frozen in Specs 035–036	Accounting location of an association

Source: Governed estimand definitions and accounting systems documented in the technical appendix.

Sample: Länder and Kreise remain separate regional layers. Registered but unestimated objects carry no numerical result.

Migration is both confounder and possible adjustment margin and is not inserted automatically as a contemporaneous control. Cohort execution and the aggregate resident/workplace split are now admitted. Forecast-vintage momentum, bilateral commuting, direct technology outcomes, the stable-Census branch, and Kreis reduced forms remain separate blocked gates. A sign difference or one interval excluding zero is not a covariance-aware difference test. Specs 035–036 and 042B provide direct tests only within their admitted systems; The convergence system does not supply a cross-model significance test.

PRESERVED The original model grid and Papers 0.8.2, 0.9.0, and 0.9.1 remain unchanged. Paper 0.9.3 adds a corrected direct-test layer, registered pooled period/geography restrictions, official Länder spatial inputs, and submission-specific manuscript editions.

SECTION 5

Composition, Transition, and Denominators

The German evidence begins with a result that resembles Hayashi's international benchmark. It does not survive every comparison that matters. The primary comparison relates annualized growth in GDP per resident aged 20–79 to the period-average log old-to-young ratio across three adjacent decades and all 16 Länder. The model contains 48 Land-window observations, initial log output, period-average child dependency, and window effects. Länder are the clusters; the primary estimate is unweighted. Paper 0.9.3 reports exhaustive null-imposed WCR11 enumeration beside CRV3 leave-one-Land inference. Exhaustive enumeration removes Monte Carlo error but does not create generally exact finite-sample inference. Population weighting and removal of the three city states are registered secondary views rather than selection rules. Each model estimates a common conditional slope, not 16 Land-specific effects.

The answer to the title question depends on what "aging" means. Average age structure produces a stable negative association. Faster realized aging does not reproduce it with comparable precision, and annual within-Land estimates change sign when the GDP denominator changes. The complete grid is therefore an ordered audit: first the closest German version of Hayashi's level equation, then changes in weighting, demographic object, denominator, horizon, and accounting outcome. None of those changes is a cosmetic robustness check. A sign comparison is not a covariance-aware difference test.

5.1 The German Benchmark and the Hayashi Scale

For GDP per resident aged 20–79, all 4 adjacent-window point estimates are negative. In the primary all-Länder unweighted run, a 10% higher old-to-young ratio corresponds to 0.317 percentage points lower annual growth, with a 95% WCR11 interval of [−0.565, −0.068]. Across all registered variants, the estimates range from -0.573 to -0.317 percentage points; 3 intervals lie below zero and 1 spans it.

Hayashi's column (C) estimate for OECD post-transitioners, the Asian Tigers, and China is −4.17 on the log old-to-young ratio. Multiplication by $\log(1.1)$ places it on the paper's reporting scale: −0.397 percentage points for a 10% higher ratio. The German population-weighted estimate is −0.376, while the primary unweighted estimate is −0.317. The proximity is informative because the demographic ratio and GDP denominator are aligned. It is not a replication identity: Hayashi compares countries over 1990–2020, whereas the German estimate pools three Land-level decades, adds period effects and child dependency, and relies on 16 clusters.

Table 8 Hayashi and the primary German level benchmark on a common scale

Estimate	Unit of comparison	Effect of 10% higher ratio	Statistical contract
Hayashi (2026), column (C)	37 mature economies, 1990–2020	−0.397 pp	OLS slope −4.17 (SE 0.68)
Germany, all Länder, unweighted	48 Land-decades, 1991–2021	−0.317 pp	Exhaustive WCR11 95% CI [−0.565, −0.068]
Germany, all Länder, population-weighted	48 Land-decades, 1991–2021	−0.376 pp	Exhaustive WCR11 95% CI [−0.672, −0.084]

Source: Hayashi (2026), Table 1, column (C); frozen Länder target estimates documented in the replication package.

Sample: Point estimates are multiplied by $\log(1.1)$. Different samples, controls, and inference prevent a pooled equality test.

For GDP per employed person, all 4 adjacent-window point estimates are also negative. The primary all-Länder unweighted estimate is −0.367 percentage points with interval [−0.616, −0.116]; the registered range is -0.585 to -0.367. Again, 3 intervals lie below zero and 1 spans it.

Thus, the closest German counterpart to Hayashi's level equation yields a negative full-period association under both denominators. That is the benchmark result, but not a general German aging coefficient. It compares periods with different average age structures, conditional on initial output and child dependency, and it gives the post-unification decade the same coefficient as the later windows. It does not isolate the adjustment produced by a demographic change.

Table 9 Regime and geography sensitivity of the composition benchmark

Sample or specification	N	Effect	WCR11 95% CI	WCR11 p	CRV3 p
All Länder, three windows	48	-0.317	[-0.565, -0.069]	0.010	0.070
All Länder, post-2000	32	-0.096	[-0.394, 0.202]	0.460	0.419
Eastern group deleted	33	-0.093	[-0.564, 0.378]	0.574	0.664
Western Länder only	30	0.083	[-0.957, 1.122]	0.814	0.885
East-by-window controls	48	-0.084	[-0.361, 0.193]	0.447	0.463

Source: Governed three-window Länder panel and submission-specific re-estimation.
Sample: Effects are annual percentage points for a 10% higher period-average old-to-young ratio. WCR11 exhaustively enumerates every Rademacher sign pattern; this removes Monte Carlo error but does not create generally exact finite-sample inference. CRV3 is leave-one-cluster jackknife inference with a t_{G-1} reference distribution. The East-by-window row adds an East indicator and its interactions with the second and third windows.

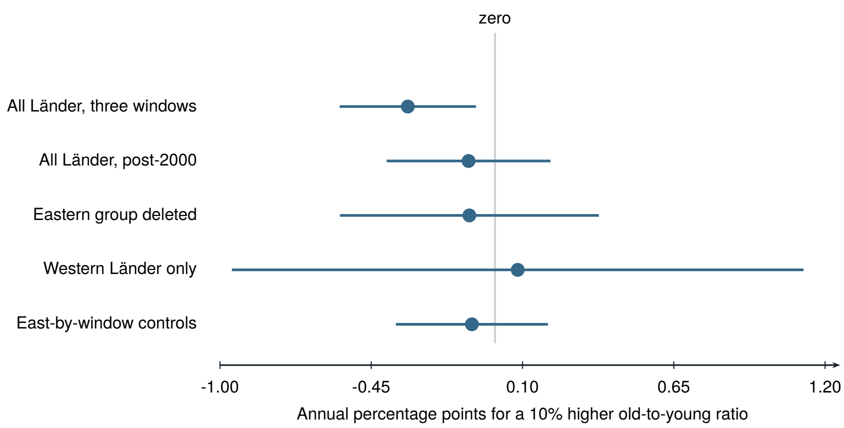

Figure 4 Composition coefficient under the referee-requested regime and geography checks. Horizontal lines are 95% intervals from exhaustive WCR11 enumeration; the vertical zero axis is labeled explicitly.

Source: Governed three-window Länder panel and submission-specific re-estimation.

Sample: The plot compares prespecified and requested specifications; it is not a model-selection device.

The regime and geography checks change the reading of the benchmark. On the same two post-2000 windows used by the matched transition audit, the point estimate falls from −0.317 to -0.096 percentage points and has exhaustive WCR11 $p = 0.460$. Deleting the five eastern non-city Länder yields −0.093; restricting the sample to the ten western Länder yields 0.083; and adding East-by-window intercepts yields −0.084. All three intervals span zero. These are descriptive attenuations, not formal tests of coefficient equality.

The pooled model tests those differences directly. Berlin remains in the nationwide benchmark but is excluded from the East–West slope comparison because its divided-city history does not belong cleanly to either non-city group. The post-2000 slope contrast is 0.304 points with $p = 0.198$; the eastern–western contrast is 0.397 points with $p = 0.151$. The joint two-restriction p-value is 0.220. Wide intervals and only five eastern clusters leave substantial uncertainty. The full coefficient is descriptively concentrated in the 1990s and East–West geography, but the pooled model does not establish a structural break.

No single Land alone creates the pooled association. Across all 16 deletion refits, coefficients range from -0.457 to -0.265 percentage points, and exhaustive WCR11 p-values range from 0.005 to 0.061. Omitting Hamburg produces the only deletion p-value above 0.05. The separate appendix reports the full pattern. Stability to one deletion and instability across historical regimes are different properties.

The official-boundary sensitivity does not overturn the benchmark. Neither queen-contiguity residual diagnostic rejects spatial randomness, and the 300 km and 500 km panel-Conley standard errors are 0.103 and 0.098 percentage points. These calculations address cross-Land covariance under two preregistered distance bands. They do not identify spatial spillovers or remove sorting from the conditional mean.

Table 10 Pooled tests of period and geographic slope instability

Slope or restriction	Effect	Pointwise 95% CI	WCR11 p	CRV3 p
Panel A: implied slopes				
Western Länder, 1991–2001	-0.477	[-1.177, 0.223]	0.224	0.403
Western Länder, post-2000	-0.173	[-1.112, 0.766]	0.657	0.767
Eastern non-city Länder, 1991–2001	-0.080	[-0.794, 0.634]	0.785	0.865
Eastern non-city Länder, post-2000	0.224	[-0.414, 0.862]	0.463	0.633
Panel B: direct restrictions				
Post-2000 minus 1991–2001 slope	0.304	–	0.198	0.344
Eastern minus western slope	0.397	–	0.151	0.339
Period and geography slopes jointly equal	–	–	0.220	0.365
All-16 post-2000 minus 1991–2001 slope	-0.510	–	0.163	0.096

Source: Prespecified pooled interaction family and all-16 period-only sensitivity using the authenticated three-window panel.
Sample: Effects are annual percentage points for a 10% higher period-average old-to-young ratio. The additive East–West model excludes Berlin and constrains the period shift to be common across East and West and the geographic slope gap to be constant across periods. Its joint row uses a two-restriction maximum absolute WCR11 statistic over $2^{15} = 32{,}768$ sign patterns. The final row is a separate nationwide period-only model with all 16 Länder and $2^{16} = 65{,}536$ patterns; it does not impose East–West terms.

Table 11 Governed Länder spatial-dependence sensitivity

Inference or window	Effect / Moran's I	SE / $E[I]$	p
Panel A: benchmark coefficient			
Exhaustive WCR11 enumeration	-0.317	0.109	0.010
CRV3 cluster jackknife	-0.317	0.163	0.070
Panel Conley, 300 km	-0.317	0.103	0.008
Panel Conley, 500 km	-0.317	0.098	0.006
Panel B: residual spatial diagnostics			
1991–2001	-0.076	-0.067	0.959
2001–2011	0.121	-0.067	0.301
2011–2021	0.030	-0.067	0.579

Source: BKG VG250 Länder boundaries, 1 January 2026; ; author derivation.
Sample: Panel Conley covariance combines unrestricted within-Land serial dependence with same-window cross-Land Bartlett weights at the prespecified centroid cutoffs and reports t_{15} reference p-values. Moran p-values use 99,999 fixed-seed permutations of row-standardized queen-contiguity weights. With only 16 Länder these diagnostics are bounded sensitivities, not a spatial structural model.

5.2 Same Sample, Different Demographic Object

On the matched resident-aged-20–79 sample, all 8 canonical and adjacent-average transition point estimates are negative, between -0.202 and -0.015 percentage points. However, all 8 WCR11 intervals span zero.

The matched employed-person audit does not preserve even the common sign: 4 point estimates are negative and 4 are positive, across a range from -0.178 to 0.121 percentage points. All 8 intervals span zero.

Restricting the composition and adjacent-average transition estimands to the same two decades removes sample composition as an explanation. For the working-age-adult outcome, all point estimates remain negative but every interval spans zero. For the employed-person outcome, the realized-change estimates are positive while the matched level estimates are negative. An older average age structure and a faster change in age structure are therefore not two labels for the same empirical object. The joint-inference record retains the direct period-average versus adjacent-average contrast as blocked because the two targets do not share an admitted joint estimator. The separate intervals do not establish that the coefficients differ.

5.3 Predetermined Cohort Pressure

The cohort design changes both the source and scale of demographic variation. Across the 16 Länder, predicted aging ranges only from -0.107 to -0.094 log point, a spread of 0.013. The native one-log-point coefficient is consequently large relative to the observed support and should not be read as a feasible one-unit intervention.

Table 12 Predetermined cohort-transition reduced form, 2001–2011

Weighting	Estimate	95% CI	p	First-stage p	Placebo p	MDES
Unweighted	-58.220	[-179.630, 62.533]	0.294	0.214	0.127	131.052
Start-population weighted	-93.878	[-189.672, 2.988]	0.054	0.089	0.496	92.462

Source: Frozen Spec 041 release *(see replication package)*.

Sample: $N = 16$ Länder. Estimates and MDES are annual log percentage points per log point of cohort-predicted aging. Intervals and p-values use null-imposed WCR11 with 9,999 draws.

The unweighted reduced-form estimate is -58.220 annual log percentage points per log point, with interval $[-179.630, 62.533]$ and $p = 0.294$. The start-population-weighted estimate is more negative at -93.878, but its interval $[-189.672, 2.988]$ still includes zero and its p-value is 0.054. First-stage relevance is not established in the primary specification ($p = 0.214$), and the unweighted MDES of 131.052 exceeds twice the absolute point estimate. The pre-outcome placebo also remains imprecise. These diagnostics neither validate an instrument nor establish a null. They show that one eligible cohort window does not supply enough independent exposure variation to separate a predetermined-pressure association from zero with useful precision.

5.4 Within-Land Evidence and the Denominator

Within Länder over time, the four resident-aged-20–79 level estimates are negative. On the 10% demographic scale they range from -1220.583 to -431.240 euros, and every interval spans zero.

For GDP per employed person, the four corresponding fixed-effects estimates are positive, between 458.576 and 1091.384 euros, while every interval again spans zero.

Land and year effects remove persistent Land levels and common annual shocks, but they do not resolve reverse migration or other time-varying confounding. The result is imprecise under both outcomes and changes sign with the denominator. GDP per resident aged 20–79 combines workplace output, employment, commuting, and a resident demographic denominator; GDP per employed person conditions on realized workplace employment. Their opposite point-estimate signs show that the within-Land result is not a denominator-invariant productivity relationship. They do not by themselves prove a statistically resolved denominator difference. In the later common-X system, $Q/N - Q/H = H/N$ is an exact accounting identity rather than a separate denominator-mechanism test. The corrected contrasts in Section 6 therefore test substantive transformed outcomes instead of relabelling that identity.

5.5 Annual Within, Between, and First-Difference Accounting

The 2000–2021 system replaces the two-outcome annual audit with five exact accounting outcomes and separates persistent regional differences from annual within-Land change. Table 13 reports native log–log coefficients. Its intervals are pointwise; p_F is the single-step five-outcome decision for the same target class.

Table 13 Annual within, between, and first-difference aging associations

Outcome	Within Land	Between Länder	First difference
$\log(Q^W/N^R)$	0.245 [-0.017, 0.502]; $p_F = 0.261$	-1.863 [-2.810, -0.898]; $p_F = 0.010$	0.174 [0.016, 0.331]; $p_F = 0.272$
$\log(Q^W/\mathcal{H}^W)$	0.354 [0.110, 0.595]; $p_F = 0.031$	-1.166 [-1.769, -0.550]; $p_F = 0.010$	0.327 [0.151, 0.498]; $p_F = 0.018$
$\log(\mathcal{H}^W/E^W)$	-0.065 [-0.161, 0.031]; $p_F = 0.342$	0.081 [-0.125, 0.282]; $p_F = 0.838$	-0.066 [-0.136, 0.007]; $p_F = 0.122$
$\log(E^W/E^R)$	0.052 [-0.059, 0.159]; $p_F = 0.677$	-0.633 [-1.001, -0.256]; $p_F = 0.012$	-0.022 [-0.080, 0.035]; $p_F = 0.882$
$\log(E^R/N^R)$	-0.095 [-0.242, 0.049]; $p_F = 0.537$	-0.144 [-0.419, 0.134]; $p_F = 0.769$	-0.065 [-0.131, -0.001]; $p_F = 0.150$

Source: Frozen Spec 042B release *(see replication package)*.
Sample: Unweighted 16-Land systems. Level columns use 352 Land-years (2000–2021); first differences use 336 adjacent changes (2001–2021). Entries are coefficients with pointwise WCR11 95% intervals and single-step five-outcome p_F.

For GDP per resident, the between-Land coefficient is -1.863 with interval $[-2.810, -0.898]$ and $p_F = 0.010$. The within-Land coefficient is 0.245 and the first-difference coefficient is 0.174. Their pointwise p-values are 0.060 and 0.030, but their familywise values are 0.261 and 0.272. The inherited annual release reported a direct restriction, but its displayed studentization combined coefficient draws generated under separate within and between nulls. Paper 0.9.3 re-estimates that contrast under the direct null throughout. For output per resident, within minus between is 2.108 and the corrected five-outcome adjusted p-value is 0.030. The source-of-variation difference is therefore statistically resolved under the corrected exhaustive WCR11 family.

Output per hour is more precise in the registered annual system. Its within-Land coefficient is 0.354 ($p_F = 0.031$), its between-Land coefficient is -1.166 ($p_F = 0.010$), and its first-difference coefficient is 0.327 ($p_F = 0.018$). Thus, the positive within and first-difference associations and the negative between association each survive their respective five-outcome corrections. This does not turn the three coefficients into one treatment response; they describe different sources of variation.

Table 14 Direct within-versus-between equality tests

Outcome	Difference	Simultaneous 95% CI	Pointwise p	Adjusted p	CRV3 p
Output / residents aged 20–79	2.108	[0.388, 3.827]	0.018	0.030	0.001
Output / actual hour	1.520	[0.400, 2.640]	0.018	0.021	< 0.001
Actual hours / workplace worker	-0.147	[-0.538, 0.244]	0.205	0.527	0.245
Workplace / resident employment	0.685	[0.079, 1.291]	0.015	0.037	0.002
Resident employment / residents aged 20–79	0.049	[-0.581, 0.679]	0.801	0.992	0.800

Source: Direct re-estimation of the five inherited annual restrictions.
Sample: Native log–log coefficients. Each pointwise test uses one direct restriction, one null-imposed exhaustive WCR11 distribution, and its own studentization. Adjusted p-values and simultaneous intervals use the single-step maximum absolute statistic across the five aligned direct-null draw series. Exhaustive enumeration removes Monte Carlo error but is not generally exact finite-sample inference. CRV3 uses a t_{15} reference distribution.

The complete direct-test table prevents one favorable comparison from standing in for the system. Output per resident, output per hour, and workplace relative to resident employment reject after five-outcome adjustment. Their corrected adjusted p-values are 0.030, 0.021, and 0.037. Hours per workplace worker and resident employment intensity remain unresolved. Pointwise, familywise, and CRV3 columns answer different inferential questions; none is selected after seeing the results.

The primitive equations explain the ratio result. The within coefficient is -0.098 for raw workplace output and -0.452 for actual hours. Their difference is the positive 0.354 output-per-hour coefficient. Resident population, workplace employment, and resident employment also decline with the within-Land demographic regressor in the point estimates. These movements are compatible with labor-input compression, selection, or restructuring; they do not identify productivity-enhancing adaptation.

The extended identity shows where the GDP-per-resident coefficients sit. In the within equation, 0.245 is the exact sum, before rounding, of output per hour, hours per worker, workplace concentration, and resident-employment intensity. The same reconciliation holds for the between and first-difference columns. The between-Land workplace-concentration coefficient is negative and rejects its registered five-outcome null, whereas resident-employment intensity does not reject in any of the three estimands. Aggregate ratios cannot identify bilateral commuting, migration, or age-specific labor utilization.

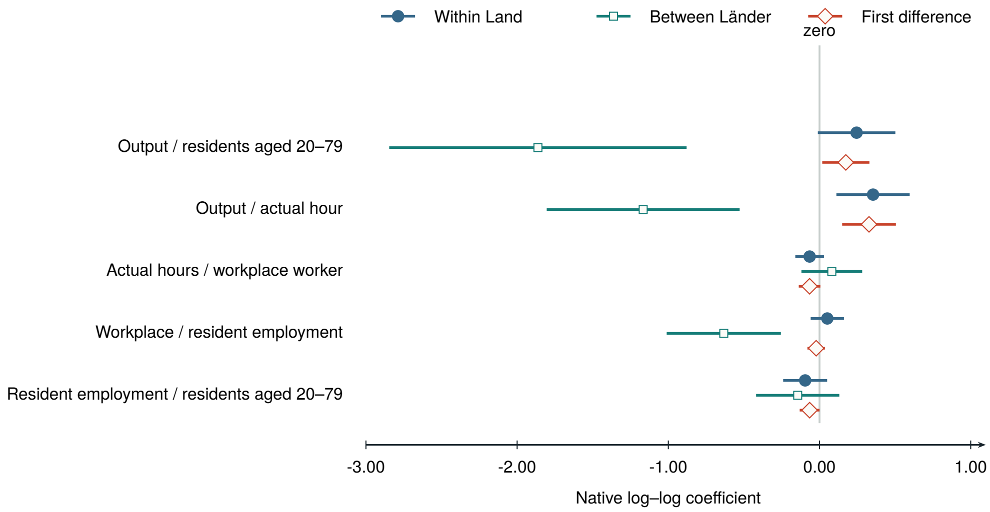

Figure 5 Annual within-Land, between-Land, and first-difference accounting coefficients. Horizontal lines are 95% intervals from exhaustive WCR11 enumeration; the vertical zero axis is labeled explicitly.

Source: Exhaustive re-inference from the authenticated annual panels.

Sample: Five accounting outcomes on common level and difference samples. The figure displays pointwise intervals and does not replace the prespecified five-outcome familywise decisions.

Table 15 Primitive annual accounting coefficients

Primitive	Within	WCR11 p	Between	WCR11 p
Raw workplace output	-0.098	0.491	-3.378	0.018
Actual hours	-0.452	< 0.001	-2.212	0.139
Resident population aged 20–79	-0.344	0.002	-1.516	0.371
Workplace employment	-0.387	0.003	-2.293	0.140
Resident employment	-0.439	< 0.001	-1.660	0.339

Source: Governed 2000–2021 annual Länder accounting panel.

Sample: Native log–log coefficients. The five primitive equations use the same 352 observations, year effects, and exhaustive 16-cluster WCR11 sign distribution. Their differences reproduce all five frozen ratio coefficients within 10^{-10}.

The annual boundary checks preserve some signs but remove any simple robustness claim. Ending the difference panel in 2019 yields coefficients of 0.175 for output per resident ($p = 0.018$) and 0.326 for output per hour ($p = 0.006$). Omitting the change ending in the 2011 census year gives 0.186 and 0.353, respectively. By contrast, adding Land-specific trends leaves positive point estimates of 0.372 and 0.274 but raises their exhaustive WCR11 p-values to 0.368 and 0.228. The separate appendix reports the full exhaustive-WCR11 and CRV3 grid. The annual association is not a pandemic or single-census artifact, but it is not robust to flexible regional trends.

5.6 Horizon and Pandemic Boundaries

All four long-period resident-aged-20–79 estimates are negative, from -0.785 to -0.263 percentage points. Two intervals lie below zero and two span it.

All four long-period employed-person estimates are negative, between -0.532 and -0.250 percentage points, and all four intervals lie below zero.

The long comparison resembles Hayashi's cross-sectional horizon most closely. It also compresses each Land into one observation. The uniformly negative point estimates are economically legible, especially for GDP per employed person, but persistent regional differences remain inseparable from average demographic composition.

Before the pandemic endpoint, all eight resident-aged-20–79 composition and adjacent-average transition estimates remain negative, ranging from -1.224 to -0.017 percentage points. Two intervals lie below zero and six span it.

The corresponding employed-person estimates split evenly: 4 are negative and 4 are positive, with a range from -0.684 to 0.155 percentage points. One interval lies below zero and seven span it.

Removing the 2020–2021 endpoint does not erase the negative resident-denominator pattern, but it does not reconcile the employed-person change estimates either. The pandemic is a relevant boundary condition. It is not a general explanation for the difference between composition and change.

5.7 Conditional Convergence and East Timing

The two-window system asks a narrower question: whether lower initial output, baseline aging, the endpoint aging transition, and East timing are associated with subsequent growth on one fixed 32-row sample. The unweighted GDP-per-resident restrictions appear in Table 16.

Table 16 Initial conditions, aging, and East–West convergence

Target for $g^{Q/N}$	Estimate	95% CI	Pointwise p	Max-t p_F
Unconditional initial output	-1.187	[-1.706, -0.668]	0.005	0.010
Conditional initial output	-1.829	[-3.448, -0.210]	0.036	0.064
Baseline aging	-1.963	[-4.881, 0.954]	0.154	0.623
Aging transition	-2.755	[-5.761, 0.250]	0.065	0.348
East, early window	0.118	[-0.487, 0.724]	0.690	0.994
East, late window	0.665	[-0.003, 1.333]	0.052	0.201
East × late window	0.547	[-0.076, 1.170]	0.096	0.518

Source: Frozen Spec 042C release *(see replication package)*.

Sample: $N = 32$ Land-windows (2001–2011 and 2011–2021), 16 Land clusters, five East clusters. Slopes are annual log percentage points per log point; East/window terms are annual log percentage points. Each p_F belongs to a predeclared five-outcome family.

The unconditional initial-level coefficient is -1.187 with pointwise interval $[-1.706, -0.668]$ and $p_F = 0.010$. Lower initial output is therefore associated with faster subsequent growth in this registered system. The conditional initial-level coefficient is -1.829; its pointwise p-value is 0.036, but its familywise value is 0.064. Because the models do not share a joint cross-model estimator, the change in coefficients has no valid cross-model p-value or confidence interval.

Neither demographic coefficient supplies a familywise result. Baseline aging is -1.963 ($p_F = 0.623$), while the endpoint aging-transition coefficient is -2.755 ($p = 0.065$, $p_F = 0.348$). The early East coefficient, late

East coefficient, and East window shift for GDP per resident likewise do not clear their five-outcome gates. Conditional convergence is therefore not an aging channel, and the model does not establish a distinct East GDP-per-resident trajectory after familywise inference.

The component system is more specific. In the late window, the conditioned East coefficient for hours per worker is -0.236 ($p_F = 0.016$). The East late-minus-early shift is -0.736 for hours per worker ($p_F = 0.004$) and 0.345 for workplace concentration ($p_F = 0.012$). The East–West change in the workplace-concentration gap is 0.228 ($p_F = 0.010$). These offsetting accounting associations locate East timing in labor-market ratios; they do not identify commuting, migration, or a structural adjustment mechanism. Every East result rests on five East clusters.

5.8 Kreis Heterogeneity and the Capital Numerator

Table 17 Kreis heterogeneity in the single 2012–2021 span

Quantity	Value	Literal concept
Growth within-Land share	0.872	current price
Initial-level within-Land share	0.809	current price
East median annualized growth	3.056	current price
West median annualized growth	2.567	current price
East–West empirical overlap	0.627	descriptive distribution
Output per standard hour within-Land share	0.763	current price / standard hours

Source: Authenticated frozen Spec 043C aggregation products.

Sample: 396 stable regions, one 2012–2021 span over 9 elapsed years. The overlap is empirical and has no sampling interpretation.

The finer regional layer weakens a binary East–West account. For 396 fixed Kreise over 2012–2021, 87.2% of the variation in current-price output-per-resident growth lies within Länder. The East median exceeds the West median, but the empirical distributional overlap is 0.627. These are descriptive quantities from one span using current-price workplace output and standard hours. They are neither real-productivity estimates nor evidence that East status is a treatment.

Table 18 Capital-stock and signed actual-hours accounting

Target	Estimate	Pointwise p	Max-t p_F
Unweighted, unconditional initial level	-1.421	0.039	0.106
Population weighted, conditional initial level	1.942	0.002	0.025
Unweighted aging, signed hours	2.390	0.026	0.053
Population weighted aging, signed hours	2.805	0.012	0.018
Population weighted aging, capital per hour	3.763	0.019	0.058

Source: Authenticated frozen Spec 045A direct-test and familywise rows joined by exact estimand identifier.

Sample: The numerator is a gross fixed-asset stock chain index; the denominator uses actual hours. A positive signed-hours term means that slower hours growth raises capital per hour by identity.

The later capital system preserves the same 32 Land-window rows as the convergence design and separates the gross fixed-asset stock index from actual hours. The unweighted unconditional initial-level coefficient for capital per hour clears its pointwise test but not its three-outcome familywise test. A population-weighted conditional coefficient clears both. For the aging transition, however, no capital-stock numerator row clears its familywise test. The only familywise result in the displayed weighted decomposition is the signed hours term, where lower hours growth raises capital per hour by identity. The ratio therefore supplies no evidence of an investment or automation mechanism.

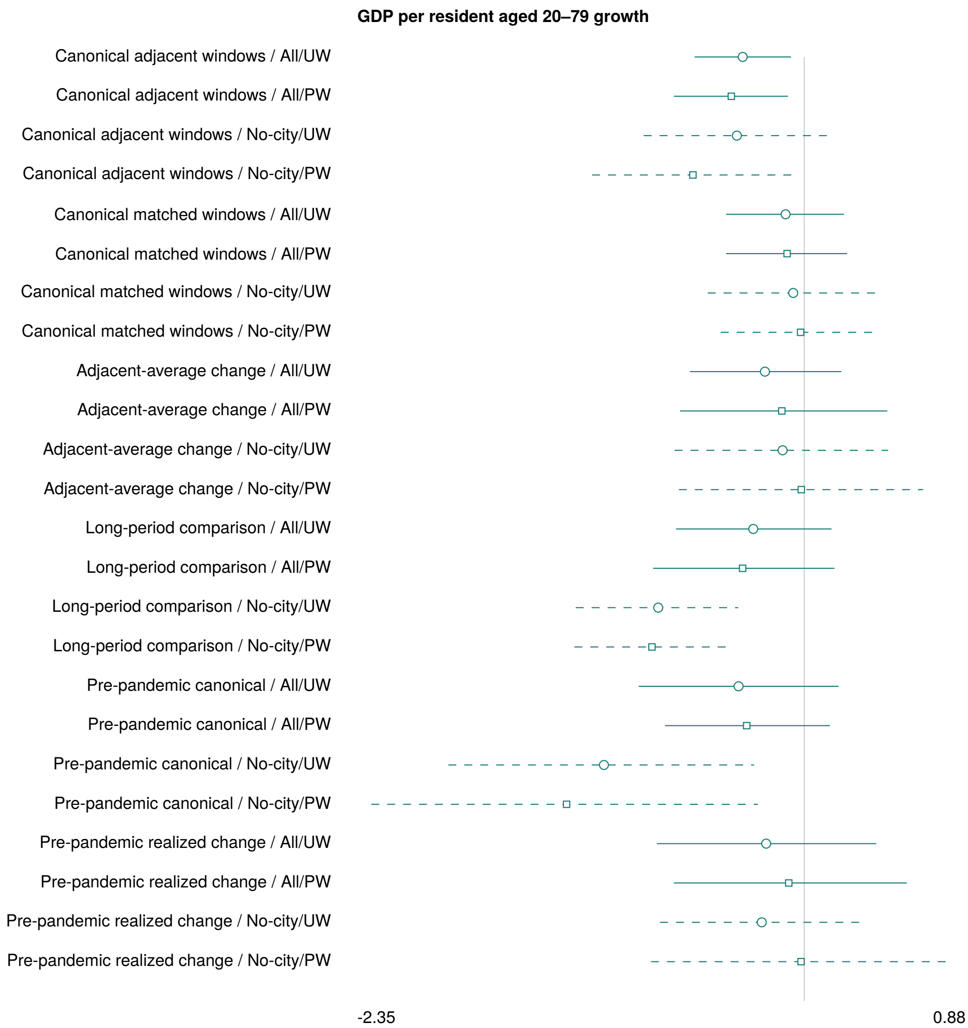

Figure 6 Länder specification map: GDP per resident aged 20–79 growth

Source: Spec 005 frozen target estimates and WCR11 intervals.

Sample: All 24 registered panel runs. Circles are unweighted; squares are population-weighted; dashed intervals exclude city states.

Table 19 Länder estimates: resident ages 20–79 denominator

Design	Unit	All/UW	All/PW	No-city/UW	No-city/PW
Canonical adjacent windows	pp	-0.317 [-0.565, -0.068]	-0.376 [-0.672, -0.084]	-0.347 [-0.827, 0.141]	-0.573 [-1.093, -0.062]
Canonical matched windows	pp	-0.096 [-0.402, 0.204]	-0.088 [-0.403, 0.221]	-0.057 [-0.498, 0.380]	-0.019 [-0.430, 0.391]
Adjacent-average change	pp	-0.202 [-0.588, 0.190]	-0.116 [-0.640, 0.428]	-0.112 [-0.668, 0.432]	-0.015 [-0.645, 0.612]
Annual Land and year fixed effects	EUR	-974 [-2364, 407]	-1221 [-2767, 329]	-431 [-1740, 876]	-792 [-3195, 1569]
Long-period comparison	pp	-0.263 [-0.661, 0.140]	-0.318 [-0.779, 0.155]	-0.752 [-1.177, -0.340]	-0.785 [-1.184, -0.396]
Pre-pandemic canonical	pp	-0.338 [-0.853, 0.176]	-0.296 [-0.717, 0.132]	-1.032 [-1.833, -0.259]	-1.224 [-2.230, -0.239]
Pre-pandemic realized change	pp	-0.196 [-0.759, 0.371]	-0.080 [-0.672, 0.528]	-0.219 [-0.743, 0.315]	-0.017 [-0.790, 0.760]

Source: Spec 005 frozen target estimates; WCR11 intervals.
Sample: All registered sample and weight variants. Values scale a 10% higher old-to-young ratio.

Table 20 Länder estimates: employed-person denominator

Design	Unit	All/UW	All/PW	No-city/UW	No-city/PW
Canonical adjacent windows	pp	-0.367 [-0.616, -0.116]	-0.425 [-0.717, -0.135]	-0.402 [-0.820, 0.012]	-0.585 [-1.020, -0.148]
Canonical matched windows	pp	-0.042 [-0.129, 0.044]	-0.060 [-0.147, 0.029]	-0.145 [-0.343, 0.047]	-0.178 [-0.417, 0.059]
Adjacent-average change	pp	0.041 [-0.063, 0.146]	0.121 [-0.065, 0.309]	0.006 [-0.156, 0.161]	0.080 [-0.197, 0.354]
Annual Land and year fixed effects	EUR	534 [-539, 1629]	459 [-1016, 1932]	1091 [-816, 3006]	977 [-2943, 4857]
Long-period comparison	pp	-0.250 [-0.499, -0.001]	-0.292 [-0.559, -0.021]	-0.504 [-0.902, -0.104]	-0.532 [-0.796, -0.274]
Pre-pandemic canonical	pp	-0.051 [-0.213, 0.111]	-0.147 [-0.432, 0.139]	-0.526 [-1.074, 0.000]	-0.684 [-1.319, -0.054]
Pre-pandemic realized change	pp	0.116 [-0.022, 0.257]	0.155 [-0.005, 0.318]	0.000 [-0.081, 0.082]	0.094 [-0.144, 0.332]

Source: Spec 005 frozen target estimates; WCR11 intervals.
Sample: All registered sample and weight variants. Values scale a 10% higher old-to-young ratio.

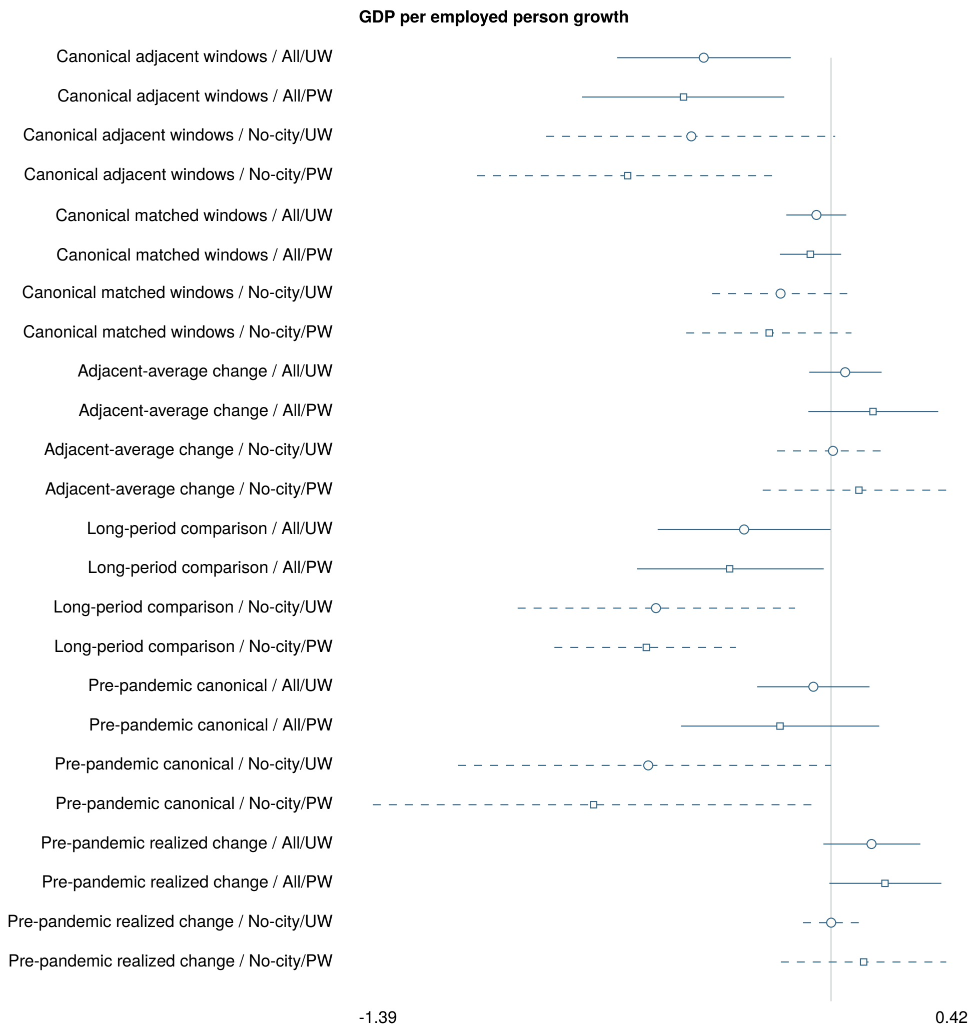

Figure 7 Länder specification map: GDP per employed person growth

Source: Spec 005 frozen target estimates and WCR11 intervals.

Sample: All 24 registered panel runs. Circles are unweighted; squares are population-weighted; dashed intervals exclude city states.

Taken together, the composition benchmark survives weights, city-state exclusion, and every one-Land deletion, but its point estimate attenuates in the regime and geography checks; the pooled slope tests do not reject equality. Predetermined cohort pressure does not reproduce the full coefficient precisely. Annual evidence separates a negative between-Land association from positive output-per-hour dynamics, and corrected direct tests resolve selected within–between differences. Land-specific trends remove the apparent within-Land precision. The two-window model confirms unconditional catch-up while leaving aging and the GDP-per-resident East restrictions outside their familywise tests. Kreis heterogeneity and capital accounting further narrow the interpretation. The next section asks where the surviving accounting associations appear and what they do not establish.

SECTION 6

How Regions Adjust: What the Accounts Show

The accounting evidence asks where the association with adjacent-average realized aging appears in the available workplace accounts. It does not observe the technology, migration, participation, or organizational response that may produce those accounts. The first layer retains six outcome-specific regressions. The second applies one common-X design and shared covariance to seven exact-identity components. The third uses resident employment to split the labor-market margin annually and then places the same five-part identity inside the two-window convergence design.

6.1 Outcome-Specific Accounting Regressions

Figure 8 reports the original six targets on the same $\log(1.1)$ scale as the growth results: percentage points of annual growth associated with a 10% higher old-to-young ratio. Each regression conditions on its own pre-window outcome and the initial manufacturing share.

6.2 Output per Hour

Output per actual hour remains close to zero. The pooled/full estimate is -0.019 percentage points with a 95% WCR11 interval of $[-0.188, 0.148]$. The early 2001–2011 window estimate is -0.010 with interval $[-0.127, 0.105]$. Both intervals span zero. Thus, the negative period-composition benchmark does not reappear as evidence that faster adjacent-average aging reduced output per hour in these accounting equations.

This comparison is informative but not decisive. The composition and accounting treatments differ, and these equations use outcome-specific baseline controls. They therefore remain separate from the common-X system below.

6.3 Capital Deepening Is a Ratio

Capital deepening is positive in both point estimates, but only the pooled/full-window interval excludes zero. Across 2001–2011 and 2011–2021, the estimate is 0.498 percentage points with interval $[0.141, 0.870]$ and $p = 0.014$. In the early 2001–2011 window, the estimate is 0.407 with $[-0.160, 0.977]$ and $p = 0.139$. The intervals are wide, overlap materially, and do not constitute a covariance-aware difference test.

The early-window estimate is also highly sensitive to Hamburg. On the unscaled log-point coefficient, deleting Hamburg moves the estimate from 4.268 to 14.698, and Hamburg's Cook distance is 4.30. More fundamentally, the capital object is a gross-fixed-asset stock chain index, not capital services. Equation (2.8) adds a second boundary: $K/\mathcal{H}$ can rise because K rises, $\mathcal{H}$ falls, or both. The common-X system estimates those terms separately, but the result remains a conditional capital-deepening association rather than evidence that aging produced investment or induced automation.

6.4 The Residual Does Not Name the Mechanism

The growth-accounting residual does not preserve its sign. The pooled/full estimate is −0.105 with interval [−0.321, 0.114], whereas the early 2001–2011 estimate is 0.169 with [−0.282, 0.624]. Both intervals span zero. The residual may contain utilization, labor quality, allocation, markups, sector composition, and measurement error. Calling it total factor productivity would make the label more ambitious, not the evidence more informative.

6.5 The Common-X Decomposition

The common-X system applies the same observations, regressors, effects, weights, and bootstrap draws to all seven components. The system changes the question. It does not ask whether separately controlled equations happen to have different signs; it asks where one shared conditional association is located while enforcing the accounting identities.

For a 10% larger adjacent-average old-to-young transition, the primary unweighted association is 0.289 annual log percentage points for output per resident aged 20–79 and 0.388 for output per hour. Hours per workplace-employed person contribute −0.064 points and workplace employment per resident contributes −0.035. Under the single-step component-family test, only output per hour excludes zero: its simultaneous interval is [0.010, 0.766] and $p = 0.045$. Capital per hour is 0.490 points, but its single-step interval [−0.153, 1.134] spans zero and its single-step p-value is 0.133.

The arithmetic is exact rather than rhetorical. Output per resident equals output per hour plus hours per employed person plus workplace employment per resident at both the observation and coefficient levels. Capital per hour equals capital per resident minus hours per resident. Reconciliation errors are below 10^{-10}. The decomposition locates a conditional association; it does not identify why the capital stock or hours changed.

The same discipline changes the economic reading of output per hour. The 0.388 point ratio coefficient and the −0.464 point actual-hours coefficient imply a raw workplace-output coefficient of -0.076 points. In the point estimates, output is approximately flat while hours decline. The output-per-hour association is therefore consistent with labor-input compression, selection, or restructuring; it is not evidence that aging raised output or induced a productivity-enhancing technology response.

6.6 The Capital Numerator and Hours Denominator

The primitive point estimates narrow the ratio's interpretation. Capital-stock index growth is 0.027 points with pointwise interval [−0.333, 0.386] and $p = 0.860$. Actual-hours growth is −0.464 points with interval [−0.759, −0.169] and $p = 0.014$. Their signed contribution contrast is −0.437 points and is not resolved under either family procedure ($p = 0.110$ pointwise). Thus, in the point estimates, the positive capital-per-hour association is accounted for by an actual-hours decline, not measured capital-stock growth. The unresolved direct contrast prevents the word "mainly" from carrying an inferential claim. The ratio is not evidence of investment, capital services, or automation.

6.7 Direct Tests, Power, and Sensitivity

The corrected contrast family excludes four tests that merely renamed an accounting component. Under the single-step correction, output per hour minus the workplace-employment-per-resident margin is resolved at 0.423 points, with simultaneous interval [0.099, 0.748] and $p = 0.016$. Output per hour minus hours per worker has stepdown $p = 0.042$ but single-step $p = 0.058$ and a simultaneous interval that crosses zero. The procedures answer different multiplicity questions and are not interchangeable. Simulation-calibrated MDES rows retain the same 16-cluster family and make imprecision visible without asserting equality.

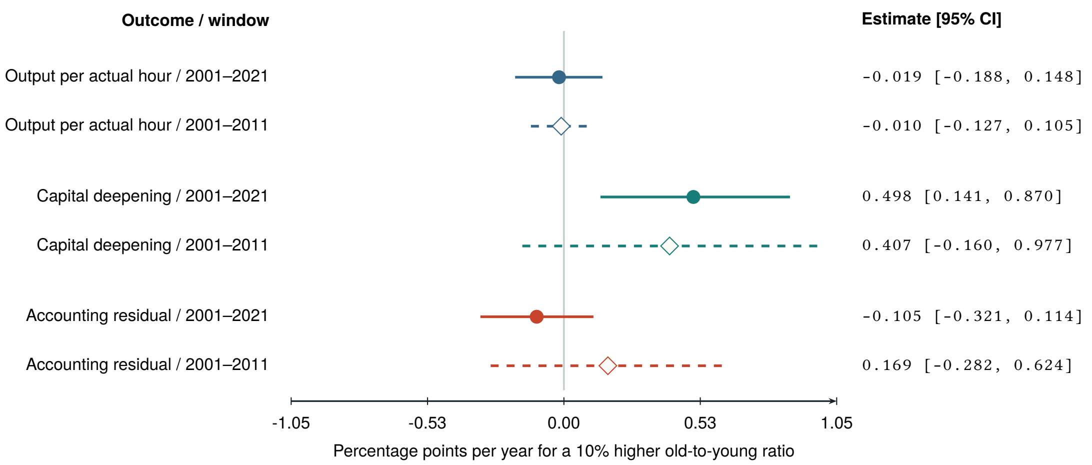

Figure 8 Conditional accounting associations with regional aging. Points are OLS estimates scaled by ln(1.10); horizontal lines are equal-tailed 95% WCR11 intervals. Filled circles with solid intervals use both ten-year windows. Open diamonds with dashed intervals use 2001–2011 only.

Source: Frozen evidence-recovery release *(see replication package)*; content digest *(see replication package)*.

Sample: Six accounting-only targets for 16 Länder. The pooled/full variant has 32 Land-window observations; the early 2001–2011 variant has 16. Estimates condition on baseline outcome, initial manufacturing share, and period effects where applicable. These are accounting associations, not direct technology-adoption, innovation, automation, efficiency, mediation, or causal estimates.

Table 21 Corrected Common-X accounting associations with regional aging

Outcome	Estimate	Pointwise CI	Pointwise p	Simultaneous CI	Single-step p	RW stepdown p
GDP / resident ages 20–79	0.289	[0.007, 0.572]	0.045	[-0.125, 0.704]	0.172	0.161
GDP / actual hour	0.388	[0.136, 0.640]	0.016	[0.010, 0.766]	0.045	0.045
Actual hours / workplace-employed person	-0.064	[-0.158, 0.031]	0.133	[-0.181, 0.054]	0.337	0.290
Workplace employment / resident ages 20–79	-0.035	[-0.142, 0.072]	0.519	[-0.263, 0.193]	0.991	0.519
Capital-stock index / actual hour	0.490	[0.152, 0.829]	0.011	[-0.153, 1.134]	0.133	0.128
Capital-stock index / resident ages 20–79	0.392	[0.036, 0.747]	0.028	[-0.299, 1.082]	0.302	0.290
Actual hours / resident ages 20–79	-0.099	[-0.238, 0.041]	0.157	[-0.276, 0.079]	0.316	0.290

Source: Frozen correctness release *(see replication package)*.
Sample: Primary unweighted common-X system: 32 Land-window observations and 16 Land clusters. Estimates and intervals are annual log percentage points for a 10% larger adjacent-average old-to-young transition. Pointwise WCR-t and Single-step max-t p-values invert their displayed intervals; RW stepdown p-values are a separate multiplicity result.

Table 22 Output-per-hour association in primitive terms

Term	Effect	Interpretation
Output per actual hour	0.388	Reported ratio
Actual hours	-0.464	Primitive denominator growth
Implied raw workplace output	-0.076	$(Q/H) + H = Q$

Source: Common-sample output and actual-hours accounting system.
Sample: Annual percentage points for a 10% larger adjacent-average aging transition. The implied raw-output coefficient is an exact same-system identity, not a separately selected regression.

Table 23 Primitive capital-stock and actual-hours associations

Target	Estimate	Pointwise 95% CI	Pointwise p
Capital-stock index growth	0.027	[-0.333, 0.386]	0.860
Actual-hours growth	-0.464	[-0.759, -0.169]	0.014
Capital growth plus hours growth	-0.437	[-0.982, 0.108]	0.110

Source: Frozen correctness release *(see replication package)*.
Sample: Annual log percentage points for a 10% larger aging transition. The capital object is a gross-fixed-asset stock chain index. These rows are accounting associations, not investment or automation estimates.

Table 24 Direct-null tests of substantive accounting contrasts

Contrast	Estimate	Pointwise CI	Pointwise p	Simultaneous CI	Single-step p	RW stepdown p
Output/hour minus hours/worker	0.452	[0.125, 0.779]	0.019	[-0.015, 0.919]	0.058	0.042
Output/hour minus workplace employment/resident	0.423	[0.191, 0.655]	0.006	[0.099, 0.748]	0.016	0.016
Hours/worker minus workplace employment/resident	-0.029	[-0.167, 0.110]	0.695	[-0.331, 0.274]	1.000	0.781
Capital growth plus hours growth	-0.437	[-0.982, 0.108]	0.110	[-1.240, 0.365]	0.440	0.320
Capital/hour minus output/hour	0.102	[-0.535, 0.740]	0.588	[-0.700, 0.904]	0.998	0.781

Source: Frozen correctness release *(see replication package)*.
Sample: Each contrast is estimated as a transformed outcome under the same 32-observation system and bootstrapped under its own null. Units are annual log percentage points for a 10% larger aging transition. Accounting relabelings and cross-sample difference tests are excluded.

Table 25 Common-X coefficient concentration by sample and window

Variant	Q/N	Q/H	K/H	K/N	K	H
All Lander	0.289	0.388	0.490	0.392	0.027	-0.464
No city states	0.460	0.486	0.672	0.646	0.084	-0.588
Eastern group deleted	0.027	0.162	-0.181	-0.316	-0.460	-0.279
Western Lander only	0.016	0.109	-0.191	-0.284	-0.428	-0.237
2001–2011	0.269	0.515	0.386	0.140	-0.106	-0.492
2011–2021	0.306	0.326	0.521	0.502	0.080	-0.441

Source: Frozen correctness release *(see replication package)*.
Sample: Annual log percentage points for a 10% larger aging transition. These are descriptive concentration diagnostics. Different samples are not linked by corrected difference-test inference. K is the capital-stock index and H is aggregate actual hours.

The sensitivity coefficients reveal geographic concentration. Capital per hour is 0.490 points in the full sample but −0.181 after deleting the five eastern non-city Länder and −0.191 in the western-Länder sample. Output per resident falls from 0.289 to 0.027 and 0.016, respectively. These literal coefficients are descriptive concentration diagnostics, not corrected cross-sample difference tests. The positive full-sample capital and output transition associations are concentrated in eastern Germany and are consistent with convergence, restructuring, or other East-specific forces. Stable-vintage census evidence remains unavailable rather than approximated; the annual variant is now executed separately in the annual dynamics design.

The original bridge left the final E^W/N^R term unsplit. A later extension admits one compact ETR resident-employment revision, and the annual system verifies the annual identity

$$Q^W/N^R = (Q^W/\mathcal{H}^W)(\mathcal{H}^W/E^W)(E^W/E^R)(E^R/N^R)$$

for every supported Land-year and first difference. In that system, output per hour is positive and familywise precise within Länder and in annual changes. Resident-employment intensity is not familywise precise in the within, between, or first-difference equation. The negative between-Land workplace-concentration coefficient is precise, but an aggregate E^W/E^R ratio does not reveal bilateral flows or individual commuting behavior.

The convergence system further shows why the old East-concentration diagnostic was not a mechanism result. Neither aging coefficient nor any GDP-per-resident East restriction rejects its five-outcome adjusted null. The familywise East timing results instead concern hours per worker and workplace concentration and rest on only five East clusters. This is a more exact location of the accounting pattern, not evidence that aging, convergence, migration, or commuting caused it.

The capital extension then places the capital numerator and signed hours denominator inside the same four convergence registrations. The population-weighted conditional initial-level coefficient for capital per hour is precise under its three-member familywise correction. For the aging-transition target, however, the capital-stock numerator never clears that correction, and the corresponding total capital-per-hour row is also unresolved. Only the weighted signed-hours term does. The later evidence therefore strengthens the decomposition while weakening the mechanism claim: what survives is an hours margin under one registration, not an investment or automation response.

Direct technology adoption, patent innovation, digital infrastructure, and automation exposure also remain unestimated because comparable historical panels cannot be constructed from the available sources. Accounting can show where a change appears. It does not identify causal mediation, and it does not turn a residual into a technology measure.

SECTION 7

Threats, Power, and Decisive Tests

7.1 Sorting and Post-Unification Convergence

The main composition regressor is contemporaneous with the outcome window. Young adults can move toward productive regions, older residents can remain in declining regions, and workplace output can grow through incoming commuters. The 1991–2001 window adds a distinct historical transition: reunification, capital reallocation, and East–West convergence are not ordinary decade effects. Initial output, child dependency, period effects, and manufacturing shares absorb some observable structure, but they do not make regional age composition exogenous.

The regime and geography estimates show that this concern is empirical rather than ceremonial. Removing the 1990s reduces the headline effect from −0.317 to −0.096 percentage points. Deleting the eastern group or adding East-by-window intercepts produces estimates near −0.09, while the western-only estimate is slightly positive. The registered pooled interaction model provides the missing formal comparison. Neither the post-2000 slope contrast nor the eastern–western contrast rejects, and the joint exhaustive WCR11 test gives $p = 0.220$. The large coefficient is descriptively concentrated in the 1990s and East–West geography, but statistical slope instability is not established.

For that reason, the identification ladder moves toward baseline composition and timing-specific cohort-transition exposure. The cohort design executes one eligible 2001–2011 reduced form using leave-focal-Land-out composite transition factors and reports concentration, balance, placebo, relevance, deletion, influence, and power diagnostics. The narrow exposure support and imprecise first stage make the limitation visible. Even a strong first stage would not establish an exclusion restriction by itself; the completed result remains a reduced form.

7.2 Sixteen Clusters and Many Correlated Questions

Sixteen Länder provide institutional comparability but limited independent regional variation. The frozen analysis therefore uses WCR11 inference, leave-one-Land refits, and observation-level influence diagnostics. The appendix reports compact named extrema for every run; the complete products contain 812 deletion estimates and 5,104 influence records. They show where a small sample is sensitive. They do not authorize deleting an inconvenient Land or promoting another model.

Paper 0.9.3 removes Monte Carlo uncertainty from the central sensitivity tests by enumerating the complete Rademacher distribution. For 16 clusters this means all 65,536 sign patterns; each one-Land refit uses all 32,768 patterns. The full benchmark gives exhaustive WCR11 $p = 0.010$, but its CRV3 jackknife p-value is 0.070. The post-2000 and geography variants are imprecise under both. The disagreement is itself informative: with 16 clusters, the inference method is part of the uncertainty rather than a footnote to it.

Exhaustive enumeration removes Monte Carlo error because every Rademacher sign pattern enters the reference distribution. It does not make WCR11 generally exact finite-sample inference. The paper therefore reports the procedure as "exhaustive WCR11 enumeration" and presents its p-values beside CRV3 rather than treating either method as an oracle.

The same limitation applies to cross-model interpretation. A negative estimate in one equation and a positive estimate in another do not establish heterogeneity. Nor does one interval excluding zero while another includes it. The corrected annual routine therefore tests each within–between contrast with one direct restriction and one direct-null studentization. The aligned draws then form a five-outcome single-step family. Output per resident, output per hour, and workplace relative to resident employment reject equality after that adjustment; the hours-per-worker and resident-employment-intensity differences do not. Pointwise, familywise, and CRV3 decisions remain visibly distinct. Output per hour also survives the pre-Covid and census-transition exclusions, but it does not survive Land-specific trends under exhaustive WCR11 enumeration. Every East-bearing restriction depends on only five eastern non-city clusters. The direct tests narrow the conclusion rather than decorate the sign pattern.

7.3 Census, Price, and Location Concepts

The population data cross the 2011 census and later Census 2022 revisions. The Länder outcomes use chain-linked real workplace GDP, whereas the Kreis layer uses current-price workplace output. Resident employment is now observed at published precision, but it is not age matched to the 20–79 denominator. Workplace concentration is an aggregate stock ratio, not a bilateral commuting flow. These boundaries are retained in the variable and source contracts. A stable-vintage 2011 sensitivity, bilateral commuting, and functional labor markets remain blocked extensions rather than unreported assumptions. The 2022 census change lies outside the registered 2001–2021 windows.

The revision-local boundary tests remove the annual change ending in 2011 and separately end the annual sample in 2019. Neither operation reverses the positive first-difference coefficients for output per resident or output per hour. This reduces concern about one census transition and the pandemic years, but it cannot establish vintage invariance. The stable-Census design remains blocked because the required pre-2010 exact-age cells do not exist under a compatible basis.

Official BKG boundaries now permit a Länder-level spatial sensitivity. The three window-specific residual Moran statistics are insignificant, and the 300 km and 500 km panel-Conley adjustments do not widen benchmark uncertainty. With 16 spatial units, these calculations cannot identify spillovers or rescue the design from sorting. They answer the covariance question that can be asked at the benchmark geography.

7.4 What Would Change the Interpretation

Four additions would materially change the interpretation. Repeated eligible cohort windows could show whether the imprecise 2001–2011 result is specific to one narrow exposure support. Compatible pre-2010 Census-2011-basis ages could test measurement stability. Bilateral commuting could turn workplace concentration into governed origin–destination flows. Direct technology, investment-flow, or organizational evidence could test claims that no accounting ratio can identify. A governed Kreis geography could extend the new Länder covariance sensitivity to smaller spatial units, but that is a future research project rather than a condition for the present submission. The current systems answer narrower questions; they do not supply these missing objects.

SECTION 8

Conclusion: Composition Is Not Adjustment

The full German benchmark is negative and economically substantial, but it is not a portable estimate of adjustment to aging. A 10% higher period-average old-to-young ratio is associated with 0.317 percentage points lower annual growth in workplace output per resident aged 20–79. Exhaustive WCR11 gives $p = 0.010$; CRV3 gives $p = 0.070$. Once the 1990s are removed, the point estimate attenuates to −0.096. Western-only and eastern-group-deleted comparisons are also much smaller. The pooled interaction tests do not reject period or geographic slope equality, so those patterns remain strong descriptive concentration rather than established structural breaks.

The annual results provide the strongest estimand evidence. Corrected direct tests reject within–between equality after five-outcome adjustment for output per resident, output per hour, and workplace relative to resident employment. The remaining two differences are unresolved. Primitive coefficients also show why the ratio matters: the positive output-per-hour point estimate combines falling hours with approximately flat raw output. That pattern is labor-input compression in the point estimates, not evidence of productivity-enhancing adaptation. Land-specific trends make both central annual within coefficients imprecise.

The convergence, Kreis, and spatial systems sharpen that result without settling it. Lower initial output is associated with faster subsequent growth in the unconditional model. Conditional convergence is pointwise but not familywise precise, and neither aging coefficient survives its familywise test. At the Kreis level, East and West growth distributions overlap substantially and most variation lies within Länder. The governed Länder Conley sensitivity does not widen uncertainty, and residual Moran tests are insignificant. Geography describes the pattern; it does not supply a treatment.

The strongest conclusion is therefore a distinction with observable consequences. Composition describes a regional equilibrium state. Transition describes how that state changes. Convergence describes dependence on initial conditions. Adjustment describes the margins through which the accounts move. The evidence identifies conditional composition, reduced-form, accounting, and convergence associations. Regional sorting, migration, commuting, shared shocks, and post-unification convergence remain alternative explanations. It does not establish a causal effect of aging on growth, technology, investment, commuting, or adaptation.

The full German benchmark resembles Hayashi's cross-country estimate, but the two designs remain comparable benchmarks rather than identical estimands. German regional evidence supports an external-validity warning: the sign of an aggregate coefficient can survive while its economic interpretation changes. Composition is not adjustment. The accounts make that distinction measurable; they do not make it causal.

Declaration of generative AI and AI-assisted technologies in the manuscript preparation process

During the preparation of this work, the author used GitHub Copilot and Copilot Chat and OpenAI ChatGPT to assist with code development, reproducibility checks, literature organization, manuscript structure, and language editing. The author independently verified the source data, references, code, numerical results, and interpretation, reviewed and edited all AI-assisted output, and takes full responsibility for the content of the publication.

References

Acemoglu, D., Autor, D., Beirne, K., and Scott, A. (2026). Baby Busts and Growth Booms: Demographic Change and the Macroeconomy. Technical Report 35401, National Bureau of Economic Research.

Acemoglu, D. and Restrepo, P. (2017). Secular Stagnation? The Effect of Aging on Economic Growth in the Age of Automation. *American Economic Review*, 107(5):174–179.

Acemoglu, D. and Restrepo, P. (2022). Demographics and Automation. *Review of Economic Studies*, 89(1):1–44.

Adão, R., Kolesár, M., and Morales, E. (2019). Shift-Share Designs: Theory and Inference. *Quarterly Journal of Economics*, 134(4):1949–2010.

André, C., Gal, P., and Schief, M. (2026). Employment and Productivity in an Ageing Society: Japan in International Perspective. *The Journal of the Economics of Ageing*, 33:100624.

Bode, E., Dohse, D., and Stolzenburg, U. (2023). Aging and Regional Productivity Growth in Germany. *Review of Regional Research*, 43:409–432.

Bönte, W., Falck, O., and Heblich, S. (2009). The Impact of Regional Age Structure on Entrepreneurship. *Economic Geography*, 85(3):269–287.

Börsch-Supan, A., Hunkler, C., and Weiss, M. (2021). Big Data at Work: Age and Labor Productivity in the Service Sector. *The Journal of the Economics of Ageing*, 19:100319.

Borusyak, K., Hull, P., and Jaravel, X. (2022). Quasi-Experimental Shift-Share Research Designs. *Review of Economic Studies*, 89(1):181–213.

Brunow, S. and Hirte, G. (2006). Age Structure and Regional Economic Growth. *Jahrbuch für Regionalwissenschaft*, 26(1):3–23.

Brunow, S. and Hirte, G. (2009). The Age Pattern of Human Capital and Regional Productivity: A Spatial Econometric Study on German Regions. *Papers in Regional Science*, 88(4):799–823.

Canay, I. A., Santos, A., and Shaikh, A. M. (2021). The Wild Bootstrap with a “Small” Number of “Large” Clusters. *Review of Economics and Statistics*, 103(2):346–363.

Conley, T. G. (1999). GMM Estimation with Cross Sectional Dependence. *Journal of Econometrics*, 92(1):1–45.

Daniele, F., Honiden, T., and Lembcke, A. C. (2019). Aging and Productivity Growth in OECD Regions: Combatting the Economic Impact of Aging through Productivity Growth? Technical Report 2019/08, OECD.

Dauth, W., Findeisen, S., Suedekum, J., and Woessner, N. (2021). The Adjustment of Labor Markets to Robots. *Journal of the European Economic Association*, 19(6):3104–3153.

Feyrer, J. (2007). Demographics and Productivity. *Review of Economics and Statistics*, 89(1):100–109.

Goldsmith-Pinkham, P., Sorkin, I., and Swift, H. (2020). Bartik Instruments: What, When, Why, and How. *American Economic Review*, 110(8):2586–2624.

Gregory, T. and Patuelli, R. (2015). Demographic Ageing and the Polarization of Regions—An Exploratory Space–Time Analysis. *Environment and Planning A*, 47(5):1192–1210.

Hayashi, F. (2026). The Root Cause of Japan's Reversal of Fortune: Implications for China, the U.S., and Other Mature Economies. Technical Report 4553699, Social Science Research Network. First draft September 2023; August 2026 revision.

Hoshi, T. (2026). Demographic Challenges and Economic Stagnation in Japan. *The Japanese Economic Review*, 77:411–433.

Hoshi, T., Kodama, N., and Li, H. (2025). The Effect of Aging on Entrepreneurship and Aggregate Productivity. *The Journal of the Economics of Ageing*, 32:100605.

Kameda, K., Lu, Z., Fukui, M., and Chikahiro, M. (2026). Population Ageing and the Productivity of Public Capital: A Panel Cointegration Analysis of Japanese Prefectural Data. *The Journal of the Economics of Ageing*, 34:100636.

Kim, H. and Lee, B. S. (2023). Aging Workforce, Wages, and Productivity: Do Older Workers Drag Productivity Down in Korea? *The Journal of the Economics of Ageing*, 24:100444.

Lindh, T. and Malmberg, B. (1999). Age Structure Effects and Growth in the OECD, 1950–1990. *Journal of Population Economics*, 12(3):431–449.

MacKinnon, J. G., Nielsen, M. Ø., and Webb, M. D. (2023). Cluster-Robust Inference: A Guide to Empirical Practice. *Journal of Econometrics*, 232(2):272–299.

Maestas, N., Mullen, K. J., and Powell, D. (2023). The Effect of Population Aging on Economic Growth, the Labor Force, and Productivity. *American Economic Journal: Macroeconomics*, 15(2):306–332.

Mitze, T. and Reinkowski, J. (2011). Testing the Neoclassical Migration Model: Overall and Age-Group Specific Results for German Regions. *Journal for Labour Market Research*, 43(4):277–297.

Monte, F., Redding, S. J., and Rossi-Hansberg, E. (2018). Commuting, Migration, and Local Employment Elasticities. *American Economic Review*, 108(12):3855–3890.

Romano, J. P. and Wolf, M. (2005). Stepwise Multiple Testing as Formalized Data Snooping. *Econometrica*, 73(4):1237–1282.