systems

MDPI

Article

Structural Similarity and the Persistence of DEA Benchmarks in Regional Innovation Systems

Leyla Gamidullaeva [1,*], Sergey Vasin [2] and Natalia Roslyakova [3]

[1] Department of Management and Public Administration, Penza State University, Penza 440026, Russia
[2] Research Institute of Fundamental and Applied Research, Penza State University, Penza 440026, Russia
[3] Laboratory of Integrated Arctic Research, Department of Integrated Scientific Research of the Karelian Scientific Centre of the Russian Academy of Sciences, 11 Pushkinskaya St., Petrozavodsk 185035, Russia
* Correspondence: la_iem@pnzgu.ru

Highlights

Please indicate how your work links to systems science via your contributions to systems practice, theory, and/or methodology.

- The candidate population defines the benchmarking model's comparison boundary; this is distinct from the boundary of the regional innovation system itself.
- Separating structural similarity, initial selection and two later outcomes makes the reassessment of regional benchmarks explicit.

What are the main findings and/or the implications of the main findings?

- For the Russian regions studied, benchmarks selected through similarity in economic structure showed no clear persistence advantage over those obtained from random reference sets. An advantage in innovative output relative to expenditure could persist even when the original benchmark could no longer contribute to the same recipient's optimal DEA projection.
- Regional authorities can use benchmark-persistence analysis to reassess comparison regions and select cases for closer examination of innovation performance and its underlying conditions. The findings support periodic reassessment of regional benchmarks; they do not establish that transferring a benchmark region's policies would improve outcomes.

Abstract

Regional innovation policy requires well-founded choices of regions for comparative learning. This study examines whether similarity in economic structure helps identify persistent innovation benchmarks across 85 Russian territories in 2019–2021. Employment shares describe economic structure, while data envelopment analysis (DEA) relates innovation expenditure to innovative output in the following year. Structurally selected benchmarks are compared with alternative approaches, including random reference sets. We distinguish ratio retention (a sustained advantage in innovative output relative to expenditure) from projection participation (the possibility of contributing to the same recipient's later optimal DEA projection). The latter criterion is exploratory. In the main specification, ratio retention occurs in 78.3% of initially selected peer–recipient pairs, while both conditions hold in 55.1% (joint retention). Structural selection shows no clear persistence advantage over random reference sets. Sustained relative performance therefore does not ensure continued DEA reference status. For regional authorities, the framework supports the periodic reassessment of comparison regions and the selection of cases for closer investigation of innovation performance. It does not establish the effects of policy transfer. Broad sector categories, nominal monetary indicators and a single subsequent evaluation window limit generalization.

Check for updates

Academic Editors: Mitsuru Kodama and Ying Fan

Received: 10 June 2026
Revised: 23 October 2026
Accepted: 2 November 2026
Published: 5 November 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

Keywords: regional innovation systems; data envelopment analysis; structural similarity; reference sets; benchmark persistence; source reconciliation; industrial structure

1. Introduction

Regional innovation systems bring together firms, research organizations, public institutions and the capabilities accumulated through their interactions. Their performance depends on the organization of these relationships and on the environment in which they operate [1,2]. A region's innovation expenditure and innovative output are observable aggregates of this system, but neither describes its internal coordination. Comparing these aggregates therefore requires a decision about which other regions provide meaningful reference points.

Structural similarity offers one basis for defining a regional comparison set. Comparable employment structures can reduce obvious differences in the activities through which innovation is produced and commercialized, but they do not imply institutional, cognitive or organizational proximity. This distinction is consistent with the proximity literature and with research on complementary interregional linkages [3,4]. Li et al. [5], for example, distinguish structure-based from linkage-based technological proximity and obtain different relationships with interregional innovation spillovers. Zhao and Wang [6] likewise show that geographical, organizational and cognitive proximity need not operate in the same way in regional innovation performance. Sectoral resemblance can therefore justify a comparison without establishing that the corresponding region is a durable performance benchmark.

The choice of the benchmark is also a substantive issue in DEA. Ruiz and Sirvent [7] formulate benchmark selection jointly with target setting; Borrás et al. [8] construct targets on the part of the frontier spanned by preselected peer candidates; and Chatzistamoulou et al. [9] examine peer influence and the learning role of reference units. Outside DEA, Olariaga-Santa-Cruz et al. [10] select reference regions for innovation and competitiveness benchmarking by combining superior outcomes with similarity in the regional resource base. These studies make peer suitability an explicit analytical problem. Regional applications raise a further question: whether a peer chosen under an ex ante similarity rule retains its observed advantage in a later window, and whether the result reflects the similarity criterion rather than simply a smaller comparison set.

We examine this question using employment-based structural selection. The baseline DEA relates innovation-activity expenditure in 2019 to the volume of innovative goods, works and services in 2020. The selected peer is evaluated in the 2020-expenditure/2021-output window in two ways: whether its output–expenditure ratio remains higher than the recipient's and whether it can still contribute to the recipient's optimal DEA projection. Original candidates are held fixed for this second check. Both assessments are retrospective and descriptive; neither estimates the effect of expenditure or a regional intervention.

The contribution is empirical and methodological in the design of benchmark evaluation. For the observed sample, employment-based restriction produces nearer peers and reduces reliance on a few frontier regions. However, comparison with random candidate sets depends on the criterion: structural peers retain their ratio advantage less often than the control medians, while joint retention places them slightly above those medians. Their rates remain within the respective replication ranges. The study therefore documents a distinction between the two forms of persistence and a limit to treating structural resemblance as evidence of a durable DEA reference. The evaluation design separates candidate construction, external-peer availability and subsequent outcomes under fixed candidates,

https://doi.org/10.3390/systems14111416

with controls at three neighborhood sizes. It develops neither a new DEA estimator nor a new peer-selection algorithm or a general theory of regional innovation systems.

The knowledge added is a follow-up of the same originally selected regional peers and evidence that the interpretation of persistence changes when projection participation is required. Persistence matters whenever an initial comparison organizes later learning or monitoring: a region can retain a ratio advantage while no longer contributing to the recipient's frontier projection. This question is not specific to Russian statistics. Neither outcome alone establishes a transferable practice. A reversal can reflect recipient catch-up, and projection participation is conditional on the model and candidate population.

From a systems perspective, the contribution concerns the specification and reassessment of a comparison model. Its finite candidate population is not a delineation of all actors and relationships in a regional innovation system. We examine how this modeling choice changes the references available to a region and whether the original references retain their performance advantage and projection participation when observations change. The results support reporting these outcomes separately, together with coverage, instead of using a single retention percentage as evidence of general benchmark quality.

2. Regional Systems and the Choice of Benchmarks

Regional innovation studies use different inputs, outputs and time structures because they represent different stages of innovation. Zemtsov and Kotsemir [11] evaluate Russian regions using resource and patenting indicators over a long period, while Firsova and Chernyshova [12] combine preliminary clustering with an output-oriented DEA Malmquist analysis of Russian regional innovation development. Shin and Kim [13] more recently apply Window DEA to the dynamic efficiency of 16 Korean regional innovation systems and relate the resulting estimates to regional determinants. Regional DEA, temporal comparison and structurally differentiated regional data are therefore established elements of the literature. The issue examined here is narrower: the subsequent behavior of an explicitly selected regional peer.

A second strand of work represents internal regional relationships more directly. Broekel [14] examines collaboration intensity and innovation efficiency in German regions using a conditional efficiency approach. Avilés-Sacoto et al. [15] model two production stages and outputs shared by firms within a Mexican regional innovation system, while Barra and Ruggiero [16] investigate knowledge and institutional spillovers with Italian regional data and double-bootstrap DEA. These studies are important for interpretation because a scalar regional output cannot stand in for collaboration, shared resources or institutional quality. The expenditure–output frontier used here is intentionally narrower: its projection weights are coefficients of comparison, not observed interorganizational ties.

Recent network DEA also makes the production stages explicit. Stergiou et al. [17] connect knowledge generation, productive performance and environmental outcomes across 199 European regions and report a trade-off between knowledge-generation efficiency and productive performance. Such models help locate differences within an innovation process. Network DEA represents linked production stages; it does not by itself observe interregional collaboration. Applying it here would require intermediate outputs and consistent resource allocation across stages, which the two monetary aggregates do not supply.

Methodologically, the review by de Oliveira et al. [18] places combinations of DEA, clustering and multicriteria methods within an established field. Ruiz and Sirvent [7] and Borrás et al. [8] address peer suitability and target construction; Chatzistamoulou et al. [9] examine peer influence and learning. Guevel et al. [19] incorporate expert judgments through weight restrictions when constructing improvement targets. Olariaga-Santa-Cruz et al. [10] identify reference regions from performance and resource similarity without

estimating a DEA frontier. These approaches address the suitability of a comparison or improvement plan. Our evaluation follows the original selected peer into a subsequent window, distinguishing ratio advantage from continued projection participation, and compares structural restriction with random restrictions of similar size and scale. The published eight-group classification [20] is retained as a comparator, without assuming homogeneous technology within each group.

Structural resemblance and persistence need not move together. Similar employment compositions may reduce differences in the broad activities through which innovation is produced and commercialized, yet they do not identify the institutional or knowledge relationships that connect organizations across regions. This is consistent with recent proximity research: Li et al. [5] distinguish structure-based and linkage-based technological proximity, while Zhao and Wang [6] report different roles for geographical, organizational and cognitive proximity. A structurally close region may also face the same demand or supply shock and may begin with only a small performance advantage. Conversely, a more distant benchmark may retain a larger advantage but offer less transferable organizational experience. The empirical analysis can distinguish these observed patterns; the available aggregate data cannot identify the mechanism that produces them.

Ecosystem research extends the interpretation beyond resemblance between regional attributes. Zhao et al. [21], studying the Shenzhen–Hong Kong cross-border region, consider relational proximity together with environmental integration, complementary advantages and common goals. This supports considering both similarity and complementarity when defining a meaningful comparison. It does not establish that adding these dimensions will improve the persistence of DEA benchmarks; that proposition requires a separate temporal evaluation.

DEA under constant and variable returns to scale provides established ways of constructing an empirical production boundary [22,23]. Restricting observations changes that boundary even if the underlying measurements remain fixed. We distinguish the candidate set, the optimal projection, a selected external peer and two subsequent properties of that peer: ratio retention and projection participation, defined in Section 3.4. An initial frontier score cannot establish either later property.

The study addresses three questions. First, how does structural selection compare with random candidate sets on ratio retention and projection participation? Second, how do these comparisons depend on neighborhood size, and what changes when frontier screening is replaced by a nearest-higher-ratio rule? Third, do alternative specifications, a separate cooperation association and a nominal temporal decomposition support stronger interpretations? The last analyses are supplementary and do not identify regional policy effects.

The random-set exercise assesses procedural sensitivity on the observed dataset. It is not a DEA bootstrap or a substitute for the statistical framework required for frontier inference [24]. All research questions and sensitivity checks are retrospective and exploratory. Earlier results were already known when the revised analytical protocol was recorded, so the later observation window is not presented as an untouched confirmatory test set.

3. Materials and Methods

3.1. Observations and Source Reconciliation

The main analysis contains 85 non-overlapping territories in the reporting frame of the Russian statistical sources for 2019–2021. Each territory contributes one baseline observation and one later expenditure–output pair. Federal district and national aggregates are excluded. Arkhangelsk Oblast excludes Nenets Autonomous Okrug, and Tyumen Oblast excludes Khanty-Mansi and Yamalo-Nenets Autonomous Okrugs; the okrugs enter separately. The

source frame includes Crimea and Sevastopol. Retaining source identifiers is a statistical documentation choice and makes no determination about territorial status.

Innovation variables are reconciled against the 2021, 2022 and 2023 editions of Indicators of Innovation in the Russian Federation [25–27]. The verification covers 595 cells: two expenditure series, three output series and two cooperation series for 85 territories. Of these, 568 archival values agree within the precision of the published tables, 22 previously absent values are recovered, one discrepancy is resolved in favor of the published value, and four cooperation percentages remain unsuitable for the required territorial definition. The source register records the original value, published entry, territorial transformation, accepted value, table and page for every cell.

The 22 recovered entries comprise 13 output values and nine cooperation entries. A dash in these cooperation tables is coded as zero because the source legend defines it as an absence of the phenomenon. It is not treated as an unreported observation. The archival 2019 cooperation share for Yaroslavl Oblast is 20.3%, whereas the source reports 20.2%; we adopt 20.2% without assigning an unverified cause to the discrepancy. All five monetary series are positive and complete for the 85 territories after reconciliation.

Territorial adjustment uses subtraction only for absolute monetary values. Published inclusive oblast totals are reduced by the separately reported autonomous okrug totals. For example, the reconstructed 2021 output for Arkhangelsk Oblast excluding Nenets Autonomous Okrug is 31,172.4 − 14.1 = 31,158.3 million rubles. Rounding uncertainty can reach 0.1 million rubles for a difference in two published one-decimal figures. The five reconciled territorial sums reproduce the national published monetary totals within 0.1–0.3 million rubles. Percentages cannot be reconstructed by subtraction. The two cooperation series for the narrow Arkhangelsk and Tyumen territories are therefore excluded, leaving 83 observations in each supplementary regression.

We preserve additional archival decimal places when compatible with published precision, without treating the extra decimals as independently verified. A sensitivity uses source-published or source-derived rounded monetary values.

Employment is verified against the 2020 edition of Regions of Russia [28]: total employment in Table 3.3 (pp. 118–119) and the 2019 activity breakdown in Table 3.5 (pp. 142–145). All 1190 archived entries—85 totals and 1105 activity values—match the publication exactly. The register identifies each source cell and the exclusive Arkhangelsk and Tyumen territories. The source also supplies 85 water-supply, sewerage, waste-management and remediation values omitted from the 13-column archive. These complete a 14-category sensitivity profile. Total-minus-category-sum differences are at most 0.3 thousand persons and are compatible with published rounding. All archived analytical values remain unchanged; later revised totals are not substituted into this release.

3.2. Variables and the Structural Comparison Rule

Innovation expenditure and innovative output are both regional totals in million nominal rubles. Expenditure refers to innovation activities under the statistical definition introduced in 2019; output is the reported volume of innovative goods, works and services. We do not combine expenditure with a separate overlapping R&D expense measure. GRP per capita, wages and unemployment do not enter the revised frontier. This avoids mixing monetary totals with per capita welfare outcomes or assigning a desirable direction to unemployment. It also narrows the interpretation: the resulting frontier concerns the relation between two innovation aggregates, not a comprehensive regional efficiency. Table 1 summarizes the variables and their analytical roles.

Table 1. Variables and analytical roles.

Variable	Observation Years	Role and Unit
Innovation expenditure X	2019 and 2020	DEA input and denominator; million nominal rubles
Innovative output Y	2019–2021	DEA output, later outcome and regression control; million nominal rubles
Total employment E	2019	Share denominator, scale matching and regression control; thousand persons
Sectoral employment	2019	11 main-profile shares; 13 archived columns and 14 source categories for sensitivity
Joint R&D participation	2019 and 2020	Supplementary association; percent of organizations engaged in innovation activities
Published group	Fixed archival membership	Eight-group comparator; no claim of homogeneous technology

Source: innovation series [25–27]; verified employment from Rosstat [28], Supplementary File S1, Supplementary File S2; original group membership [20]. All 85 territories enter DEA; 83 enter the cooperation regressions.

The baseline pair is $X_{i,2019}$ and $Y_{i,2020}$. The later pair is $X_{i,2020}$ and $Y_{i,2021}$. A one-year lag allows an expenditure observation to precede the associated output observation, but does not establish the gestation period of innovation. Sales can embody investments made over several years; expenditure can finance activities that will not generate sales within the window. The same-year sensitivity therefore tests a timing choice rather than selecting a causally correct production lag.

The main specification retains the 11-activity profile used in the preceding analysis, so that primary verification does not redefine the baseline after its results are known. Each activity value is divided by total regional employment in 2019. The activities are agriculture, forestry and fishing; mining; manufacturing; construction; trade and vehicle repair; transportation and storage; accommodation and food services; information and communication; real estate; education; and health and social work.

Source verification identifies archival Zan_Lok as electricity, gas, steam and air-conditioning supply, and Zan_Dr as the published aggregate of other activities [28]. Their earlier identification uncertainty is therefore resolved. Both are included in the 13-column sensitivity; the 14-category sensitivity additionally restores the separately published water and waste-service column. The main profile's exclusions are a specification choice, not evidence that these activities are irrelevant. Shares are not renormalized, and omitted activities are not reconstructed from a residual.

$$d_{ij} = \frac{1}{2}\sum_{s \in S} \left| \frac{E_{is}}{E_i} - \frac{E_{js}}{E_j} \right| \quad (1)$$

Here E_i is total employment, E_{is} is employment in retained activity s, and S contains the 11 activities in the main specification. Because that profile is incomplete, d_{ij} is a scaled absolute distance over selected shares, not a total-variation distance between complete compositions. It measures neither geographic separation nor an observed knowledge link.

For each recipient i, the main external candidate set contains its 15 nearest neighbors; ties are resolved by the stable source identifier. The recipient is added to the DEA reference set to ensure feasibility. Neighborhood sizes of 10 and 20 and broader 13- and 14-category profiles provide sensitivity checks. Fourteen refers to the source's published categories,

including its heterogeneous other-activity aggregate; it does not imply a fully disaggregated industry classification. Fifteen is a working neighborhood size, not an empirically established optimum.

3.3. Frontier Estimation and External-Peer Selection

For each recipient i and reference set C_i, the output-oriented variable-returns-to-scale (VRS) problem maximizes feasible output subject to the observed expenditure constraint and convexity [23]. All inputs and outputs in this step refer to the baseline pair.

$$\begin{gathered} Y_i^* = max_\lambda \sum_{j \in C_i} \lambda_j Y_j \\ \text{subject to } \sum_{j \in C_i} \lambda_j X_j \leq X_i, \sum_{j \in C_i} \lambda_j = 1, \lambda_j \geq 0 \end{gathered} \tag{2}$$

The reported score is $e_i = Y_i/Y_i^*$, bounded by zero and one. The projection is obtained in two steps. First, output is maximized. Second, output is held at Y_i^* while projected expenditure is minimized. This sequence, called lexicographic optimization, removes input slack at the maximal output. The optimal face comprises all weight combinations that attain both optima. We identify every observation that can receive positive weight in at least one such combination.

A selected external peer must belong to this set, differ from the recipient, and have a strictly higher baseline ratio Y/X. Among qualifying peers, we choose the one with the smallest structural distance, resolving ties by source identifier. A relative tolerance of 10^{-10} is used for the ratio comparison.

Considering all optimal weight combinations prevents the solver's choice among equivalent solutions from determining the reported peer. The selected region can contribute to an optimal projection; it need not reproduce that projection alone. If no external observation satisfies the rule, no peer is recorded. Such cases remain in the coverage denominator, while conditional retention rates use only recipients with an initially selected external peer. We report all candidate identifiers, all possible external optimal references, one optimal set of weights and the feasible minimum and maximum weight for each possible reference.

A constant-returns-to-scale (CRS) sensitivity removes the convexity restriction [22]. With one input and one output, the CRS benchmark is determined by the maximum candidate output–expenditure ratio. The main VRS frontier retains a scale-dependent piecewise-linear boundary. Both models are descriptive envelopes of regional aggregates; neither treats an entire regional innovation system as an observed firm-level production process.

3.4. Comparators and Definitions of Benchmark Persistence

The structural model is compared with a pooled frontier and with the eight published regional groups [20]. Their original sizes are 1, 4, 4, 1, 12, 21, 40 and 2. We compute their descriptive implications for all 85 territories, including the absence of an external benchmark for a singleton. Small-group scores are not used as validated local technologies, and repeated region-year records are not added to increase group size. The previously assigned substantive regional types do not enter the revised analysis.

Two random-set controls assess whether structural similarity adds information beyond restricting the number and scale of candidate regions. At each k = 10, 15 and 20, the size-only control selects k external regions uniformly without replacement for every recipient. The size-and-scale control draws the same number from each employment-size quartile as in that recipient's structural neighborhood. Quartiles use 2019 total employment, with cut points of 311.1, 508.6 and 1052.2 thousand persons. Matching preserves this coarse scale composition, not every feature of the structural neighborhood.

Each of the six main-profile controls has 999 replications, giving 5994 in total. Reproducible random streams use seed components 20260924, k and a matching indicator. The

recipient is included in every DEA set, and the same optimal-face and external-peer rules apply. Each draw's candidate set is held fixed for its later projection check.

A separate sensitivity repeats both controls 999 times for the full 14-category profile at k = 15, using that profile's distances and neighborhood quartile composition. These 1998 additional replications give 7992 across both profiles; the two profiles are reported separately.

A further comparator omits DEA. It selects the nearest external region with a higher baseline output–expenditure ratio among the same 15 structural candidates. This separates the empirical role of frontier screening from the role of proximity alone. Because this simpler rule can return a peer when the DEA rule does not, we report both its full eligible set and the recipients shared by the two rules.

Benchmark persistence is the general term for the later properties of an originally selected peer. We assess two properties—ratio retention and projection participation—and call their joint occurrence joint retention.

Ratio retention means that the original peer p(i) retains a strictly higher later output–expenditure ratio than recipient i. Equation (3) defines the conditional rate. Coverage is the fraction of all 85 territories receiving a qualifying baseline external peer. We also report retained pairs relative to all 85 territories, so that a selective rule is not judged only within its eligible subset.

$$b_i = 1[\frac{Y_{p(i),2021}}{X_{p(i),2020}} > \frac{Y_{i,2021}}{X_{i,2020}}],\ Retention = \frac{\sum_{i \in P} b_i}{\mid P \mid} \quad (3)$$

P is the fixed set of recipients with a qualifying baseline external peer; coverage = |P|/85. Each recipient in P defines one peer–recipient pair, and the same peer region may serve several recipients. All conditional retention rates use this original set P. The ratio comparison uses the same relative tolerance as at selection.

Projection participation means that the original peer can still contribute to the same recipient's later optimal DEA projection. We keep the baseline candidate identifiers, including the recipient, and re-estimate the VRS projection with X2020 and Y2021. Output is maximized first and input is then minimized at that output. The original peer satisfies the criterion if it can receive positive weight in at least one solution on the resulting optimal face. It need not become the newly selected nearest peer; membership of a pooled frontier alone is insufficient.

Joint retention requires both ratio retention and projection participation. All three conditional rates use the same baseline set P. A later loss of projection participation is an outcome, not a reason to remove a pair from the denominator.

For verification, a third linear program maximizes the original peer's weight while holding the later optimal output and minimum input fixed; a weight above 10^{-7} is treated as positive. This independently checks optimal-face membership for all 425 recipient–method cases across pooled, published-group and structural k = 10, 15 and 20 models, plus 18 random-control spot checks. Regional results and maximum weights are in Supplementary File S1 (ProjectionPersistence). The original k = 15 random peer assignments and ratio summaries are reproduced exactly. These are computational checks on the same data, not an independent empirical replication.

Initial advantages differ across methods. We divide the positive log ratio gaps among all ordered regional pairs into three equally populated strata, using baseline information only. The peer/recipient ratio cut points are 1.9251 and 4.5732. Standardized ratio retention is the unweighted mean of the three within-stratum rates. This coarse adjustment reduces differences in initial-gap distributions; it does not eliminate differences within strata or identify a treatment effect. To compare methods on the same recipients, each random replication is also evaluated on the recipients for which both rules select a baseline external peer. We

calculate structural-minus-control differences for ratio retention, projection participation and joint retention on that common set.

Replication ranges for the random controls are the empirical 2.5th and 97.5th percentiles across reference-set replications, with the observed data held fixed. They describe variation in the selection procedure. They are not confidence intervals for a regional population and are not used to derive population *p*-values. Overlapping neighborhoods and repeated selection of the same peer region also preclude treating peer–recipient pairs as independent trials.

3.5. Supplementary Cooperation Association

To examine a measured aspect of regional interaction separately from the frontier, we estimate two ordinary least-squares regressions for the 83 territorially comparable observations. The cooperation indicator is the percentage of organizations engaged in innovation activities that participated in joint R&D projects. It records participation, not the strength, quality, location or direction of collaboration. The dependent variable is log innovative output in the next year. Controls are log expenditure, log current output, log total employment in 2019, and the 2019 mining and manufacturing employment shares.

$$\begin{aligned} log\ Y_{i,t+1} = \alpha_t + \beta_{1t} log X_{it} + \beta_{2t} log Y_{it} + \beta_{3t} log E_{i,2019} \\ + \beta_{4t} Mining_{i,2019} + \beta_{5t} Manufacturing_{i,2019} + \gamma_t \frac{Coop_{it}}{10} + \varepsilon_{it} \end{aligned} \quad (4)$$

The coefficient on Coop/10 corresponds to a ten-percentage-point difference. We report HC3 standard errors [29] and two-sided 95% confidence intervals using the residual degrees of freedom. These intervals are not robust to spatial dependence. Leave-one-out prediction errors are computed with and without cooperation, and a sensitivity excludes the source-coded zero cooperation observations. These are regressions on observed output, not second-stage regressions on estimated DEA scores. We do not interpret their coefficients as spillovers or policy effects.

3.6. Nominal Malmquist Diagnostic

We supplement the benchmark-persistence analysis with a pooled CRS Malmquist decomposition [30] of the two lagged expenditure–output pairs. This directly separates relative movement toward the monetary frontier from the movement of the frontier itself, while retaining the same 85 observations. Let $r_{i0} = Y_{i,2020}/X_{i,2019}$, $r_{i1} = Y_{i,2021}/X_{i,2020}$, and $h_t = \max_i r_{it}$. Under one-input, one-output CRS, the output distance is $D_t(x,y) = y/(h_t x)$. The usual cross-distance index reduces to the ratio change below.

$$M_i = \frac{r_{i1}}{r_{i0}},\ EC_i = \frac{r_{i1}/h_1}{r_{i0}/h_0},\ FS = \frac{h_1}{h_0},\ M_i = EC_i \times FS \quad (5)$$

EC is relative efficiency change and FS is the nominal frontier shift. Their product equals M for every territory. The use of lagged pairs is explicit: this is not a conventional contemporaneous production panel. No regional input or output deflators are applied. We therefore call FS a nominal frontier shift and do not interpret it as technical progress. In this simple CRS specification the decomposition adds an explanation of the moving reference boundary, not information beyond the observed ratios. The calculation is an exploratory diagnostic and is not used to optimize the structural neighborhood or peer-selection rule.

4. Results

4.1. Benchmark Availability and Two Persistence Outcomes

Structural DEA at k = 15 selects an external peer for 69 of 85 territories, giving 81.2% coverage. Ratio retention occurs in 54 of these 69 peer–recipient pairs (78.3%), equivalent to

63.5% of all territories. Pooled DEA and the published groups have higher ratio-retention rates, but their coverage differs. Table 2 reports counts and denominators for each rule. Projection participation is assessed separately below.

Table 2. Coverage and ratio retention of selected external benchmarks.

Rule	Retained/ Selected	Coverage %	Ratio Retention %	Retained/ All 85%	Standardized %	Median Distance
Structural k = 15	54/69	81.2	78.3	63.5	73.8	0.0638
Pooled DEA	70/80	94.1	87.5	82.4	90.2	0.1084
Published groups	56/66	77.6	84.8	65.9	86.0	0.1013
Random same size	54/64	75.3	84.4	63.5	81.8	0.1016
Random size and scale	56/66	77.6	83.8	65.9	81.9	0.0965
Nearest higher ratio	51/79	92.9	64.6	60.0	69.3	0.0522

Source: developed by the authors.

Selected and retained counts refer to peer–recipient pairs; a distinct peer region may serve more than one recipient. Random rows report separate medians over 999 replications at k = 15, so a ratio of median counts need not equal the median rate. Conditional and standardized rates concern ratio retention; coverage and retained/all use all 85 territories. Standardization equally weights three initial-advantage strata. Distance is the 11-activity measure in Equation (1). Replication ranges appear in the text and Figure 1.

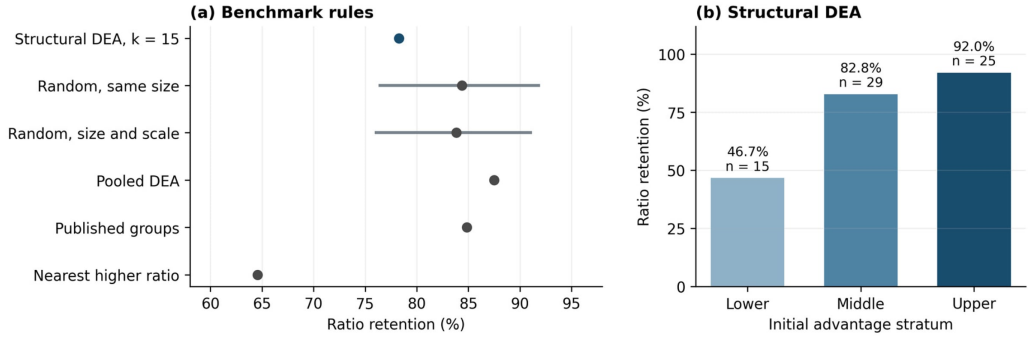


Figure 1. Ratio retention under alternative rules (**a**) and within the structural model by initial-advantage stratum (**b**). Horizontal ranges show the 2.5th–97.5th percentiles across 999 random-set replications, not confidence intervals. The structural strata contain 15, 29 and 25 recipients; ratio cut points are 1.9251 and 4.5732. This figure does not measure projection participation. Source: developed by the authors.

The structural method's median selected-peer distance is 0.0638, compared with 0.1084 for pooled DEA. This is consistent with how candidates are constructed; it is not an independent validation of contextual comparability. Structural DEA uses 20 distinct peers, with no peer serving more than eight recipients. Pooled DEA concentrates selection on four regions, with one serving 33 recipients. The published groups use 12 distinct peers, with a maximum of 21 selections. Structural restriction spreads reliance across more observations, while its conditional ratio-retention rate is lower than that of the pooled frontier.

At k = 15, median ratio-retention rates are 84.4% for the size-only random control and 83.8% for the size-and-scale control. Their central 95% replication ranges are 76.5–91.8% and 76.1–91.0%. The structural rate of 78.3% is below both medians but within both ranges, providing no clear ratio-retention advantage for structural selection. These ranges describe procedural variation in fixed data and do not establish a population disadvantage.

The mean structural VRS score is 0.4834, compared with 0.2607 for pooled DEA and 0.4947 for the published groups. Higher restricted scores are expected when comparator observations are removed from the feasible envelope. The score comparison therefore cannot establish that the regions became more efficient or that structural grouping improved the underlying innovation process.

Projection participation changes the assessment (Table 3). In 39 of the 69 initial structural peer–recipient pairs, the original peer can still contribute to the later optimal projection; 38 pairs also meet ratio retention. Thus, 16 pairs meet ratio retention only, one meets projection participation only, and 14 meet neither criterion. Joint retention is 55.1%, or 44.7% of all 85 territories. The size-only and size-and-scale random medians are 53.2% and 52.2%, with central 95% replication ranges of 42.6–63.9% and 40.6–64.2%. Structural joint retention is slightly above both medians but within both ranges. Table 3 reports the corresponding outcomes for the pooled and published-group frontiers.

Table 3. Ratio retention, projection participation and joint retention at k = 15.

Rule	Selected Pairs	Ratio Retention %	Projection Participation %	Joint Retention %	Joint/ All 85%
Structural k = 15	69	78.3	56.5	55.1	44.7
Pooled DEA	80	87.5	71.3	70.0	65.9
Published groups	66	84.8	33.3	33.3	25.9
Random same size	64	84.4	55.7	53.2	40.0
Random size and scale	66	83.8	54.0	52.2	41.2

Source: authors' calculations. Counts refer to peer–recipient pairs, and all conditional rates use the original baseline-selected recipients. Candidate identifiers are held fixed for the later projection. Random rows report separate medians over 999 replications at k = 15, not the outcomes of a representative draw. Joint retention requires both criteria; joint/all uses all 85 territories.

4.2. Initial Advantage and the Role of Frontier Screening

Within structural DEA at k = 15, ratio retention increases with initial advantage: it occurs in seven of 15 pairs in the lower stratum, 24 of 29 in the middle and 23 of 25 in the upper. The corresponding rates are 46.7%, 82.8% and 92.0% (Figure 1b). Equal weighting reduces structural ratio retention to 73.8%; random-control medians are 81.8% and 81.9%. Structural ratio retention remains below both medians after this coarse adjustment. These rates do not describe projection participation.

Recipient populations differ across procedures. At k = 15, when comparisons are restricted to recipients with a selected baseline peer under both rules, median structural-minus-control differences remain negative for ratio retention and positive for joint retention. All central 95% replication ranges include zero. These common-recipient comparisons reduce differences in coverage, but do not constitute randomized treatment comparisons or population inference. The method-specific differences and ranges are reported in Supplementary File S1 (MethodComparison).

The nearest-higher-ratio rule covers more territories and selects closer peers than structural DEA (Table 2). On their 69 common recipients, ratio retention occurs in 54 DEA-selected pairs and 45 pairs selected by the simpler rule (78.3% versus 65.2%). However, the median initial advantage is also larger under DEA: 3.74 versus 1.94. The difference cannot therefore be attributed to an isolated causal contribution of DEA. Equal weighting of initial-advantage strata narrows the retention gap (Table 2).

4.3. Sensitivity to Neighborhood and Model Choices

Larger structural neighborhoods increase coverage in this sample, but neither ratio retention nor joint retention changes monotonically (Tables 4 and 5). These results provide no basis for treating 15 candidates as an optimal neighborhood size.

Table 4. Sensitivity of structural benchmark selection.

Specification	Selected Pairs	Ratio Retained	Ratio Retention %	Mean Score
Structural k = 10	64	52	81.3	0.5443
Structural k = 15	69	54	78.3	0.4834
Structural k = 20	73	60	82.2	0.4506
All 13 archived columns, k = 15	69	54	78.3	0.4864
All 14 source categories, k = 10	63	50	79.4	0.5584
All 14 source categories, k = 15	69	54	78.3	0.4882
All 14 source categories, k = 20	72	60	83.3	0.4471
CRS k = 15	79	62	78.5	0.2343
Same-year baseline	64	51	79.7	0.5068
Published monetary precision	69	54	78.3	0.4834

Source: developed by the authors.

Table 5. Neighborhood-specific structural and random-control comparisons.

k	Rule	Selected Pairs	Coverage %	Ratio Retention % [Replication Range]	Joint Retention % [Replication Range]
10	Structural	64	75.3	81.3	57.8
10	Random same size	57	67.1	84.2 [74.1–92.0]	56.1 [44.2–67.9]
10	Random size and scale	60	70.6	83.6 [75.0–91.4]	55.0 [42.1–67.8]
15	Structural	69	81.2	78.3	55.1
15	Random same size	64	75.3	84.4 [76.5–91.8]	53.2 [42.6–63.9]
15	Random size and scale	66	77.6	83.8 [76.1–91.0]	52.2 [40.6–64.2]
20	Structural	73	85.9	82.2	56.2
20	Random same size	68	80.0	84.5 [76.7–92.3]	52.2 [40.9–62.7]
20	Random size and scale	70	82.4	83.8 [76.7–90.4]	50.0 [39.1–61.6]

Source: authors' calculations. Each random row summarizes its own 999 replications. Brackets show central 95% replication ranges, not confidence intervals. Selected pair counts, coverage and rates in random rows are separate medians. Conditional rates use baseline-selected recipients; joint retention requires both ratio retention and projection participation. Observations and baseline structural profiles are unchanged.

All rows use 85 territories; selected counts refer to peer–recipient pairs. Except where indicated, the model is VRS, k = 15, with the 11-activity profile and the 2019-expenditure/2020-

output baseline. The same-year row uses 2020 expenditure and output; later evaluation remains 2021 output relative to 2020 expenditure. Scores across specifications do not represent changes in regional performance. Projection participation is evaluated for the baseline VRS candidates at k = 10, 15 and 20 and the broader activity profiles, not for every specification in this table.

Matching each neighborhood size to its own random controls preserves the ordering by criterion (Table 5). With 10, 15 and 20 candidates, structural ratio retention is below both control medians, while joint retention is above them. Every structural rate lies within the corresponding central 95% replication range. For both controls at all three neighborhood sizes, common-recipient difference ranges for ratio retention, projection participation and joint retention include zero (Supplementary File S1, MethodComparison). These comparisons do not identify a generally superior neighborhood or method.

Using all 13 archived columns or all 14 source categories at k = 15 gives the same aggregate ratio-retention count, 54 of 69, as the 11-activity baseline. However, each broader profile changes the selected-peer assignment or the availability for 19 of the 85 recipients. Joint retention is 39/69 (56.5%) with 13 columns and 37/69 (53.6%) with all 14 categories, compared with 38/69 (55.1%) in the baseline. Equal aggregate ratio retention therefore does not establish identical benchmark assignments or projection participation.

The full 14-category profile has its own 999-draw controls at k = 15. Its joint-retention rate of 53.6% is slightly above the control medians of 53.0% and 50.8%, but remains within both central 95% replication ranges. All six common-recipient difference ranges include zero. The ordering by criterion therefore persists without establishing superiority. Full-profile results at k = 10 and 20 are descriptive sensitivities; no additional random controls are fitted for those specifications. Detailed assignments, retention rates and control ranges are in Supplementary File S1 (SectorSensitivity).

Ratio retention is 62/79 under CRS and 51/64 when 2020 expenditure is matched to 2020 output (Table 4). The uniformly rounded monetary specification leaves every selected peer unchanged and changes the mean score only from 0.483400 to 0.483407. Published monetary precision is therefore not consequential for peer selection in this dataset.

Source completion has a larger effect than monetary rounding. The previous complete-case population contained 75 territories and yielded ratio retention in 49 of 61 selected pairs. That result is reproduced exactly when the reconciled dataset is restricted to those territories. Adding the ten previously excluded territories changes the selected peer for ten of the original recipients and changes 17 of their scores. With fixed-size local neighborhoods, an additional territory can replace an earlier neighbor; the resulting score change need not have the same direction for every recipient. Complete source coverage is consequently part of the reference-set design, not merely a reporting issue.

4.4. Cooperation and Later Innovative Output

The cooperation coefficient changes sign between the two windows, and both HC3 95% confidence intervals include zero (Table 6). The data therefore provide no stable positive conditional association under this specification. This does not establish that cooperation is unimportant: an aggregate participation percentage is an incomplete measure of the relationships emphasized in regional systems research.

Dependent variable: log innovative output. Each model has 83 territories; brackets show HC3 95% confidence intervals with 76 residual degrees of freedom. Sector shares are fractions, and cooperation is measured per ten percentage points. R^2 = 0.9314 and 0.9208. Intervals are not spatially robust. Full standard errors and sensitivity estimates are in Supplementary File S1 (Regressions sheet).

Table 6. Conditional associations with next-year innovative output.

Regressor	2019 → 2020	2020 → 2021
Intercept	−0.3953 [−1.8016, 1.0110]	0.7698 [−0.7486, 2.2882]
Log innovation expenditure	0.1585 [−0.1055, 0.4225]	0.1164 [−0.1457, 0.3785]
Log current innovative output	0.6486 [0.3566, 0.9406]	0.8056 [0.5775, 1.0337]
Log employment in 2019	0.2873 [−0.0630, 0.6377]	0.0665 [−0.3085, 0.4415]
Mining employment share	1.5269 [−1.6140, 4.6677]	−4.4314 [−7.4955, −1.3673]
Manufacturing employment share	4.2117 [0.0184, 8.4050]	−0.0234 [−4.2360, 4.1893]
Joint R&D participation per 10 pp	0.0197 [−0.2591, 0.2984]	−0.1235 [−0.3712, 0.1243]

Source: developed by the authors.

Including cooperation yields leave-one-out log-output RMSEs of 0.6906 and 0.7056, compared with 0.6777 and 0.7043 when it is omitted. The cooperation variable does not improve this prediction criterion in either window. Removing source-coded zero cooperation observations leaves 78 and 79 territories and produces coefficients of −0.0482 and −0.0983, again with intervals containing zero. These findings do not support deriving network-strengthening recommendations from this indicator alone. They also do not estimate the effects of particular types of partners or the benefits of interregional knowledge flows.

4.5. What the Nominal Temporal Decomposition Shows

The pooled CRS nominal Malmquist index has a geometric mean of 0.9414 (Table 7). Seventy-five of the 85 territories move closer to the period-specific CRS frontier, but only 40 improve their own lagged output–expenditure ratio. Relative catch-up and improvement in a region's own ratio therefore give different accounts of change.

Table 7. Pooled CRS decomposition for the two nominal lagged pairs.

Component	Geometric Mean	Median	Territories Above 1
Nominal index M	0.9414	0.9396	40
Relative efficiency change EC	2.7411	2.7359	75
Nominal frontier shift FS	0.3434	0.3434	0

Source: developed by the authors.

N = 85. First pair: X2019 and Y2020; second pair: X2020 and Y2021. M = EC × FS for each territory. FS is common under pooled one-input, one-output CRS. These are nominal lagged-ratio diagnostics, not estimates of real productivity or technical change. All regional components are in Supplementary File S1 (Malmquist sheet).

This apparent divergence is explained by the reference boundary. The maximum baseline ratio is 126.2963 in Ingushetia, whereas the maximum later ratio is 43.3746 in Murmansk Oblast. Ingushetia's small 2019 expenditure denominator contributes to its extreme baseline ratio. As the maximum observed ratio falls, many territories improve their

relative position without improving their own ratio. The decomposition therefore cautions against labeling relative catch-up as system-wide innovation progress. It neither separates price changes from quantities nor establishes technical regression; both would require a more complete production and price measurement design.

4.6. A Traceable Regional Example

Penza Oblast illustrates the distinction between selecting a peer, constructing a DEA projection and assessing later benchmark persistence. Under the main 11-activity specification with 15 external candidates, Bryansk and Omsk Oblasts are possible external references. The proximity rule selects Bryansk Oblast. Table 8 reports the baseline comparison and subsequent outcomes; Figure 2 places this pair among the 69 original structural comparisons.

Table 8. Penza Oblast under main 11-activity structural DEA with 15 external candidates.

Quantity	Result
Baseline expenditure/output, million rubles	5177.11/30,627.10
Projected output, million rubles	50,508.17
VRS score	0.6064
Possible external optimal references	Bryansk Oblast; Omsk Oblast
Selected peer/structural distance	Bryansk Oblast/0.0531
Peer-to-recipient initial/later ratio	2.4419/3.7297
Projected-output decrease if selected peer is removed	22.1%
Later projection participation/maximum feasible weight	Yes/0.3241
Later projection participation with full 14-category profile	No/maximum weight 0

Source: developed by the authors.

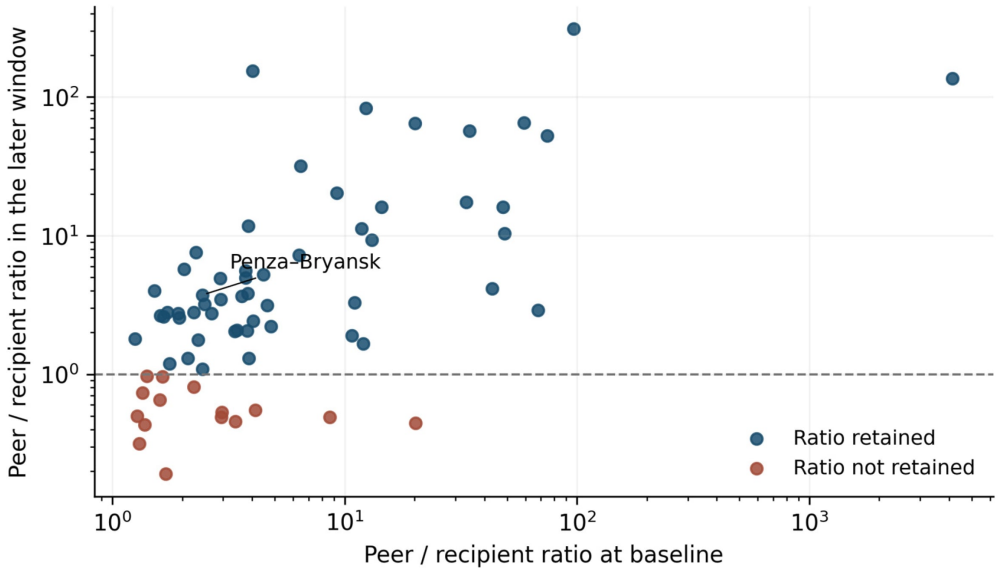


Figure 2. Initial and later peer-to-recipient output–expenditure ratios for the 69 original structural peer–recipient pairs. Both axes are logarithmic. Points above the horizontal line satisfy ratio retention; those below it do not. Penza–Bryansk is identified as a traceable example. Each point represents a selected comparison, not an independent policy experiment. Source: developed by the authors.

The peer is selected from all qualifying regions that can contribute to an optimal projection, as described in Section 3.3. The projection is a model reference, not a forecast or

intervention effect. Original candidate membership and baseline weights are in Supplementary File S1 (Benchmarks and ProjectionWeights); the later fixed-candidate check is in ProjectionPersistence.

At k = 15, Penza–Bryansk meets both ratio retention and projection participation under the main 11-activity profile. Removing Bryansk from the baseline candidate set reduces projected output by 22.1%, a separate measure of its influence on that projection. The 13- and 14-category profiles also select Bryansk initially, but give it zero feasible weight in the later optimal projection. Ratio retention therefore persists across these profiles, while projection participation depends on activity coverage and the resulting candidate set.

A substantive comparison would need to establish which innovative products generate the difference, how firms are classified in the statistical survey, whether sales reflect earlier investments, and whether comparable organizations face different commercialization constraints. The model identifies a focused comparative inquiry; it does not predict a policy-induced output gain or justify an expenditure cut.

5. Discussion

5.1. Contribution and Comparison with Previous Research

The empirical contribution is evidence on whether the same originally selected regions remain useful model references over time. Structural selection produces nearer peers and distributes reliance across more regions, but it does not demonstrate a clear persistence advantage over the random controls. The 16 pairs with ratio retention alone in the main specification illustrate why sustained performance advantage must be distinguished from projection participation. The evaluation design makes both outcomes explicit under fixed candidates, matched controls and reported coverage at three neighborhood sizes.

This finding is compatible with evidence connecting regional efficiency to capabilities beyond sector composition. Zemtsov and Kotsemir [11] associate Russian regional innovation efficiency with accumulated knowledge, established universities and entrepreneurial activity. Shin and Kim [13] find declining dynamic efficiency across Korean regional innovation systems and higher efficiency where commercialization infrastructure is more developed. These results help explain why regions with similar employment shares can differ in innovation performance. They do not test whether a structurally selected peer retains its advantage over a particular recipient, so they neither confirm nor contradict our retention comparison directly.

Temporal DEA findings likewise depend on what is measured. Firsova and Chernyshova [12] report average productivity growth in their cluster-based study of Russian regions, whereas Shin and Kim [13] report declining dynamic efficiency in Korea. Our nominal lagged-pair decomposition cannot be used to choose between these findings: its indicators, period and frontier differ, and its frontier shift does not identify technical change. Stergiou et al. [17] further show why efficiency at one stage need not coincide with performance at another. Together, these comparisons support evaluating the particular outcome attached to a benchmark instead of inferring durable superiority from an aggregate efficiency score.

The comparison with the nearest-higher-ratio rule separates candidate construction from frontier screening. Among the 69 common recipients, DEA-selected peers retain their advantage more often, but they also start from larger output–expenditure advantages. The difference is consequently an observed property of the full selection procedure, not an isolated effect of DEA. Relative to approaches that incorporate suitable or preselected peers into target construction [7,8], this result suggests a useful distinction between deciding who may enter the reference set and deciding which observations become active frontier

references. The two stages can alter benchmark quality in different ways and should be reported separately.

Within benchmark-selection research, the results qualify the presumption that a suitable initial comparison necessarily remains a useful DEA reference. Ruiz and Sirvent [7], Borrás et al. [8] and Guevel et al. [19] formulate suitability through peer choice, target alignment or expert restrictions; they do not provide the same structural-versus-random follow-up of the original peer. Chatzistamoulou et al. [9] analyze peer influence through removal, addressing dependence on the reference set rather than projection participation when observations change. These are related questions, and the cited results are not treated as direct replications.

Out-of-sample evaluation itself is established in DEA: Ouenniche and Tone [31] develop an evaluation framework for classification, illustrated by bankruptcy prediction. Our retrospective regional peer check addresses a different outcome. It is presented neither as the first subsequent evaluation in DEA nor as an independent confirmatory test.

5.2. Why Structural Similarity May Be Insufficient

Randomization applies to candidate sets. Every set then undergoes the same DEA optimization and external-peer selection rule as a structural set. The comparison asks whether employment similarity adds persistence beyond this common screening procedure and the restriction on candidate numbers. Random sets can include structurally distant regions with large initial performance advantages, which may survive greater relative deterioration. The initial-gap comparison in Section 4.2 is consistent with this interpretation. These margins help interpret ratio retention; they do not determine projection participation.

Projection participation depends on the original peer, the recipient's later expenditure and the other candidate regions. Even with a higher output–expenditure ratio, the original peer may be displaced from the optimal projection as the recipient's position or the competing references change. This possibility gives the distinction practical meaning: a sustained performance advantage need not preserve the same model reference. The observed differences do not identify a causal mechanism favoring structural similarity.

Sectoral employment shares describe the allocation of labor across broad activities. They do not measure the quality of research organizations, firms' commercialization capabilities or the knowledge available through external partners. Li et al. [5] find different spillover associations for structure-based and linkage-based technological proximity, with a weaker role for the former. Zhao and Wang [6] distinguish the roles of geographical, organizational and cognitive proximity, including mediation through absorptive capacity. These findings are consistent with the possibility that resemblance in activities leaves important differences in learning and commercialization unresolved. They are explanatory precedents, not evidence that those mechanisms caused the retention pattern in this sample.

Additional dimensions could be tested using independently measured information. Patent or publication profiles can describe technological and cognitive relatedness. Research and commercialization infrastructure, together with governance arrangements, can inform institutional and organizational comparisons. Joint projects, co-patenting and partner composition can describe network relations. Complementary capabilities may matter even when profiles are dissimilar, as the interregional and ecosystem literature suggests [4,21].

Adding every available proximity measure would not automatically improve benchmarking. A stronger test would specify a small set of candidate dimensions before later outcomes are observed, then assess their incremental contribution against the same controls, with coverage and initial gaps reported.

The observation window permits alternative explanations. A uniform proportional shock to both regional ratios preserves their ordering. Changes in relative responses, in-

novation portfolios, expenditure timing, reporting or prices can reverse it. Even with the ordering unchanged, movements of other candidates or of the recipient's expenditure position can displace the original peer from the optimal projection. The pandemic-period aggregates cannot distinguish these mechanisms. Moreover, a reversal can represent recipient catch-up. Neither persistence outcome measures the welfare value of a gap or the success of regional policy.

The cooperation regressions cannot distinguish between these explanations. The participation share does not show a stable conditional association with later output, but it records neither the partners involved nor their capabilities or the content of their collaboration. Its result cannot rule out effects of particular institutional arrangements or knowledge networks. Earlier ecosystem work emphasizes intermediary arrangements connecting actors, resources and knowledge flows [32]. The implication for further research is to measure those relations directly and test whether they improve later benchmark evaluation after controlling for initial gaps. Their relevance cannot be inferred from the absence of a stable coefficient on an aggregate participation percentage.

5.3. Practical Use of Benchmark Persistence Analysis

Policymakers can use the framework to prioritize comparisons for investigation and decide when a benchmark shortlist needs reassessment. An analytical team first records the territorial frame, variables, candidate rule, selected peers, initial gaps and projection sensitivity. When comparable later observations become available, it checks both the original peer's ratio advantage and its possible projection participation using the original candidates, before selecting new peers. This keeps reassessment separate from replacement of a benchmark. The two outcomes guide interviews, data verification and organizational case studies; differentiated regional policy [33] and stakeholder feedback [34] provide the context for this use.

The two criteria support four follow-up decisions. A peer meeting joint retention remains a candidate for investigation as both a performance and a DEA reference. A peer meeting ratio retention alone may remain an informative performance comparison, but does not meet the projection-participation criterion. Projection participation without ratio retention calls for examining the role of regional scale and the VRS frontier before describing the peer as a superior performance case. If neither criterion is met, the team checks catch-up, shocks and measurement before refreshing the comparison.

The absence of a baseline external peer indicates a gap in the available reference set, not an absence of improvement opportunities. Strong sensitivity to peer removal calls for verifying that observation and considering alternatives. These outcomes guide analytical work; they do not prescribe an intervention.

This use complements conventional regional scoreboards and target-oriented DEA. The European Commission's Regional Innovation Scoreboard [35], for example, compares regional innovation profiles and performance groups. Our procedure attaches a later outcome to a specific peer–recipient assignment, makes the absence of an external benchmark visible and evaluates the selection rule against explicit alternatives. It can therefore help decide whether a previous comparison remains informative, while a scoreboard describes a region's broader relative position. This contribution is this additional evaluation, not a replacement for multidimensional regional assessment.

For Penza Oblast, the Bryansk comparison directs attention to product composition, the firms contributing most to reported innovative sales, expenditure timing and commercialization conditions. If the advantage is concentrated in a few firms or reflects sales generated by earlier investment, the regional comparison offers little basis for policy transfer. If comparable activities operate under different organizational or commercialization

arrangements, those arrangements can be examined separately. The present data do not distinguish between these explanations. At the project level, earlier work on innovation portfolios shows that life-cycle stage and project characteristics are relevant to portfolio decisions [36].

At k = 15, the Penza–Bryansk pair meets ratio retention under all three activity profiles, but projection participation only under the main 11-activity specification. This dependence on activity coverage needs examination before the comparison is treated as a durable DEA reference. The pair remains a candidate for investigating commercialization differences; Omsk is another possible baseline reference in the main specification. Neither comparison identifies a support instrument. Budget allocation, expenditure reduction or expected policy gains require evidence about a defined intervention and a credible counterfactual.

5.4. Transferability and External Validity

The transferable component is the evaluation sequence: define candidates and selection rules, record original peers and coverage, then follow those peers into later observations with a fixed comparison set. Random candidate controls, neighborhood sensitivity and reporting of ratio retention separately from projection participation can be applied wherever comparable repeated regional data exist. Case selection and learning often occur after the observations supporting the original comparison, so the question extends beyond Russia. Transferability of the procedure does not imply transferability of the estimated rates or method ordering.

Implementation requires local adaptation. This application depends on the Russian territorial frame, innovation-survey definitions, nominal expenditure and sales, the 2019 employment archive and inclusive oblast–okrug adjustments. The 11-activity profile, k = 10–20 range, one-year lag and employment quartiles are application choices. Other settings require reconciliation of sector classifications, territorial scale, prices, survey coverage and innovation lags. Projection participation can be defined in a multi-input or multi-output model by examining the later optimal face under fixed candidates, with its orientation and slack rule specified. The present scalar ratio criterion cannot be transferred unchanged; an appropriate multidimensional performance criterion would have to be justified separately.

We would not expect an identical numerical result across national innovation systems. Where sector profiles are closely associated with stable technological capabilities and institutions, structural selection could be more informative about later performance. Where production stages and external knowledge links cut across sectors, or where relative shocks dominate a short window, employment resemblance may be less informative. These are conditional expectations derived from the literature [4–6,17,21], not findings of a cross-country test. Replication should therefore retain the evaluation logic while re-estimating the empirical comparison with local data over several observation windows; the Russian ordering of methods is an outcome to test, not an assumption to export.

5.5. Limitations and Further Validation

Primary reconciliation confirms all 1190 archived employment entries against one documented Rosstat release, but it does not establish that broad activity shares adequately represent innovation capabilities. The source-complete 14-category profile still contains an aggregate of other activities. At k = 15, broadening the 11-activity baseline changes peer assignments or availability for 19 of 85 recipients, despite unchanged aggregate ratio retention. Joint retention also changes. Benchmark identities are therefore sensitive to how structural similarity is operationalized; the main exclusions and equal activity weights have no claim to optimality. Regional comparisons require an explicit classification, a consistent statistical release and checks of individual assignments alongside aggregate rates.

The short period spans the pandemic and supplies only one later comparison window. Nominal expenditure and output reflect prices, sector composition, reporting coverage and investment timing as well as physical innovation activity. A one-year lag does not remove these influences. A common monetary rescaling within a window would leave relative peer rankings unchanged, but region-specific and sector-specific price changes can alter them. The nominal Malmquist decomposition should therefore not be read as real productivity growth.

The parsimonious frontier omits knowledge stocks, human capital, innovation quality and environmental or social outcomes. Neighborhood size and equal sector weights are explicit rather than optimal; matching uses employment quartiles and ratio-gap adjustment uses broad strata. All common-recipient difference ranges at k = 10, 15 and 20 include zero, and these procedural ranges do not quantify sampling uncertainty. The projection-participation criterion is model-dependent and does not establish a learning relationship. HC3 intervals are not spatially robust, and no reported comparison identifies a causal effect. The checks establish consistency for specified alternatives, not robustness to every defensible model.

The analysis is retrospective. The projection-participation criterion and random controls at k = 10 and 20 were specified after the earlier ratio-retention results had been examined. The source-complete profile and its controls were added after primary employment reconciliation. These are exploratory extensions, not preregistered confirmation, and there remains only one subsequent window. A stronger test would fix candidates, activity coverage and both outcome definitions before observing additional periods, then repeat the comparison across windows. Separately measured institutional and network information would allow a test of whether those dimensions add persistence beyond sector profiles.

6. Conclusions

Employment-based structural similarity provides a transparent comparison rule, but the results do not show that it identifies more persistent regional innovation benchmarks than random reference sets. In the main specification with 15 external candidates, ratio retention is 78.3% and joint retention is 55.1%; both lie within the respective random-control replication ranges. A sustained performance advantage therefore need not coincide with projection participation. The ordering by criterion persists across the tested neighborhood sizes without establishing general method superiority.

The empirical contribution is the assessment of these distinct outcomes for the same regional benchmark relationships under fixed candidate sets, with explicit reporting of coverage, initial performance gaps and matched controls. Structural similarity provides a transparent starting point for regional comparison, but continued benchmarking relevance requires reassessment. For regional authorities, the framework can inform the selection and review of cases for closer investigation of innovation performance; it does not establish the transferability of particular policies or practices.

Checks using broader activity profiles show that stable aggregate retention rates can conceal changes in individual benchmark assignments. The nominal decomposition and cooperation regressions remain supplementary diagnostics. Further research should extend the observation period and examine measured institutional, technological and network dimensions. Substantive conclusions remain conditional on the activity-profile specification, the use of nominal monetary indicators and a single subsequent evaluation window. Future research could also consider the alignment of education with employers' competency requirements [37].

Supplementary Materials: The following supporting information can be downloaded at https://www.mdpi.com/article/10.3390/systems14111416/s1. Supplementary File S1, Systems_4400220_Supplementary_Tables.xlsx, contains an Index and 14 analytical sheets: Data, Benchmarks, ProjectionWeights, NearestRatio, Malmquist, Regressions, SourceRegister, RandomReplications, ProjectionPersistence, MethodComparison, EmploymentAudit, EmploymentCells, EmploymentSource and SectorSensitivity. Supplementary File S2, Systems_4400220_Reproducibility.zip, contains executable code, source pages, provenance records, checksums and detailed exports. It includes the complete 1190-cell employment audit, the 14-category source data, 5994 main-profile random replications and 1998 additional full-profile replications.

Author Contributions: Conceptualization, L.G. and S.V.; methodology, L.G., S.V. and N.R.; investigation and writing—original draft preparation, L.G.; data curation and formal analysis, N.R.; writing—review and editing, L.G., S.V. and N.R.; visualization, L.G.; project administration, L.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Russian Science Foundation, grant No. 25-28-20328, "Models and Mechanisms for Optimizing the Structure of the Regional Economy for Sustainable Industrial Development", https://rscf.ru/project/25-28-20328/ (accessed on 1 November 2026).

Institutional Review Board Statement: Not applicable. The study analyzes aggregate regional statistics and does not involve individual participants or identifiable personal records.

Informed Consent Statement: Not applicable.

Data Availability Statement: The reader-facing analytical data and complete reported results are provided in Supplementary File S1. The executable code, protocols, source manifest and checksums are provided in Supplementary File S2. Published innovation-source tables are identified in references [25–27]. No firm-level or personal data are distributed.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Cooke, P.; Gomez Uranga, M.; Etxebarria, G. Regional innovation systems: Institutional and organisational dimensions. *Res. Policy* **1997**, *26*, 475–491. [CrossRef]
2. Martin, R.; Sunley, P. Complexity thinking and evolutionary economic geography. *J. Econ. Geogr.* **2007**, *7*, 573–601. [CrossRef]
3. Boschma, R. Proximity and Innovation: A Critical Assessment. *Reg. Stud.* **2005**, *39*, 61–74. [CrossRef]
4. Balland, P.A.; Boschma, R. Complementary interregional linkages and Smart Specialisation: An empirical study on European regions. *Reg. Stud.* **2021**, *55*, 1059–1070. [CrossRef]
5. Li, Y.; Sun, T.; Sun, Y. Linkage- and structure-based technological proximity and interregional spillovers of innovation growth. *Growth Change* **2024**, *55*, e12695. [CrossRef]
6. Zhao, S.; Wang, J. Proximity and regional innovation performance: The mediating role of absorptive capacity. *J. Sci. Technol. Policy Manag.* **2025**, *16*, 1094–1110. [CrossRef]
7. Ruiz, J.L.; Sirvent, I. Identifying suitable benchmarks in the way toward achieving targets using data envelopment analysis. *Int. Trans. Oper. Res.* **2022**, *29*, 1749–1768. [CrossRef]
8. Borrás, F.; Ruiz, J.L.; Sirvent, I. Planning improvements through data envelopment analysis (DEA) benchmarking based on a selection of peers. *Socio-Econ. Plan. Sci.* **2024**, *95*, 102020. [CrossRef]
9. Chatzistamoulou, N.; Kounetas, K.; Tsekouras, K. Knowledge flows in Data Envelopment Analysis. The role of peer effects. *Omega* **2024**, *129*, 103137. [CrossRef]
10. Olariaga-Santa-Cruz, M.; Rodríguez-Castellanos, A.; Barrutia-Güenaga, J.; Rincón-Díez, V. How to identify reference regions for comparison of innovation and competitiveness. *Reg. Stud. Reg. Sci.* **2024**, *11*, 362–383. [CrossRef]
11. Zemtsov, S.; Kotsemir, M. An assessment of regional innovation system efficiency in Russia: The application of the DEA approach. *Scientometrics* **2019**, *120*, 375–404. [CrossRef]
12. Firsova, A.; Chernyshova, G. Efficiency Analysis of Regional Innovation Development Based on DEA Malmquist Index. *Information* **2020**, *11*, 294. [CrossRef]
13. Shin, D.J.; Kim, B.H.S. Analyzing the Dynamic Efficiency and Critical Factors of the Regional Innovation System: A Case Study of Korea. *Int. Reg. Sci. Rev.* **2025**, *48*, 437–459. [CrossRef]
14. Broekel, T. Collaboration Intensity and Regional Innovation Efficiency in Germany—A Conditional Efficiency Approach. *Ind. Innov.* **2012**, *19*, 155–179. [CrossRef]

15. Avilés-Sacoto, S.V.; Cook, W.D.; Güemes-Castorena, D.; Zhu, J. Modelling Efficiency in Regional Innovation Systems: A Two-Stage Data Envelopment Analysis Problem with Shared Outputs within Groups of Decision-Making Units. *Eur. J. Oper. Res.* **2020**, *287*, 572–582. [CrossRef]
16. Barra, C.; Ruggiero, N. On the impact of knowledge and institutional spillovers on RIS efficiency. Evidence from Italian regional level. *Growth Change* **2022**, *53*, 702–752. [CrossRef]
17. Stergiou, E.; Kounetas, K.; Tsekouras, K. Innovation Efficiency, Productive Performance and Undesirable Outputs across European Regions: Are there any missing links? *J. Environ. Manag.* **2024**, *367*, 122053. [CrossRef]
18. de Oliveira, M.S.; Steffen, V.; de Francisco, A.C.; Trojan, F. Integrated data envelopment analysis, multi-criteria decision making, and cluster analysis methods: Trends and perspectives. *Decis. Anal. J.* **2023**, *8*, 100271. [CrossRef]
19. Guevel, H.P.; Ramón, N.; Aparicio, J. Benchmarking and Target Setting in Weight Restriction Context. *Mathematics* **2025**, *13*, 1175. [CrossRef]
20. Gamidullaeva, L.A.; Roslyakova, N.A. Cluster-Econometric Analysis of Russian Regions: Implications for Differentiated Economic Policy. *Econ. Reg.* **2025**, *21*, 283–300. [CrossRef]
21. Zhao, Y.; Lyu, L.; Grimes, S. An Analytical Framework for Cross-border Regional Innovation Ecosystems: The Case of Shenzhen–Hong Kong Cross-border Region. *Tijdschr. Voor Econ. Soc. Geogr.* **2024**, *115*, 402–417. [CrossRef]
22. Charnes, A.; Cooper, W.W.; Rhodes, E. Measuring the efficiency of decision making units. *Eur. J. Oper. Res.* **1978**, *2*, 429–444. [CrossRef]
23. Banker, R.D.; Charnes, A.; Cooper, W.W. Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis. *Manag. Sci.* **1984**, *30*, 1078–1092. [CrossRef]
24. Simar, L.; Wilson, P.W. Estimation and inference in two-stage, semi-parametric models of production processes. *J. Econom.* **2007**, *136*, 31–64. [CrossRef]
25. Ditkovsky, K.; Evnevich, E.; Fridlyanova, S.; Fursov, K.; Gokhberg, L.; Gracheva, G.; Kuznetsova, I.; Martynova, S.; Ratay, T.; Rosovetskaya, L.; et al. *Indicators of Innovation in the Russian Federation 2021: Data Book*; HSE University: Moscow, Russia, 2021. (In Russian) [CrossRef]
26. Ditkovskiy, K.; Fridlyanova, S.; Gokhberg, L.; Gracheva, G.; Kuznetsova, I.; Martynova, S.; Ratay, T.; Rosovetskaya, L.; Roud, V.; Vlasova, V. *Indicators of Innovation in the Russian Federation 2022: Data Book*; HSE University: Moscow, Russia, 2022. (In Russian) [CrossRef]
27. Ditkovsky, K.; Fridlyanova, S.; Gokhberg, L.; Gracheva, G.; Kuznetsova, I.; Martynova, S.; Ratay, T.; Rosovetskaya, L.; Streltsova, E.; Vlasova, V. *Indicators of Innovation in the Russian Federation 2023: Data Book*; Table 9.6; HSE University: Moscow, Russia, 2023. (In Russian) [CrossRef]
28. Federal State Statistics Service (Rosstat). *Regions of Russia. Socio-Economic Indicators. 2020*; Rosstat: Moscow, Russia, 2020; pp. 118–145.
29. MacKinnon, J.G.; White, H. Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties. *J. Econom.* **1985**, *29*, 305–325. [CrossRef]
30. Färe, R.; Grosskopf, S.; Norris, M.; Zhang, Z. Productivity Growth, Technical Progress, and Efficiency Change in Industrialized Countries. *Am. Econ. Rev.* **1994**, *84*, 66–83.
31. Ouenniche, J.; Tone, K. An out-of-sample evaluation framework for DEA with application in bankruptcy prediction. *Ann. Oper. Res.* **2017**, *254*, 235–250. [CrossRef]
32. Gamidullaeva, L. Towards Combining the Innovation Ecosystem Concept with Intermediary Approach to Regional Innovation Development. *Int. J. Econ. Bus. Adm.* **2018**, *6*, 39–53. [CrossRef]
33. Tödtling, F.; Trippl, M. One size fits all? Towards a differentiated regional innovation policy approach. *Res. Policy* **2005**, *34*, 1203–1219. [CrossRef]
34. Parygin, D.; Sadovnikova, N.; Gamidullaeva, L.; Finogeev, A.; Rashevskiy, N. Tools and Technologies for Sustainable Territorial Development in the Context of a Quadruple Innovation Helix. *Sustainability* **2022**, *14*, 9086. [CrossRef]
35. European Commission. Regional Innovation Scoreboard. 2025. Available online: https://research-and-innovation.ec.europa.eu/statistics/performance-indicators/regional-innovation-scoreboard_en (accessed on 16 October 2026).
36. Tolstykh, T.; Gamidullaeva, L.; Shmeleva, N. Approach to the Formation of an Innovation Portfolio in Industrial Ecosystems Based on the Life Cycle Concept. *J. Open Innov. Technol. Mark. Complex.* **2020**, *6*, 151. [CrossRef]
37. Deev, M.; Gamidullaeva, L.; Finogeev, A.; Finogeev, A.; Vasin, S. The Convergence Model of Education for Sustainability in the Transition to Digital Economy. *Sustainability* **2021**, *13*, 11441. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.