

Article

An AHP-MAP Framework for Transparent and Evidence-Calibrated Evaluation of Technology-Transfer Performance in New R&D Institutions

Ting Li ^{1,2}, Zhiwei Zhao ¹, Han Bao ¹, Ziqiang Zhang ³, Jizhe Zhang ^{3,*}  and Zhongyu Ma ³

¹ Gansu Provincial Computing Center, Lanzhou 730000, China

² Institute of Chinese Ethnic Information Technology, Northwest Minzu University, Lanzhou 730030, China

³ College of Artificial Intelligence and Computer Science, Northwest Normal University, Lanzhou 730070, China

* Correspondence: 202511160298@nwnu.edu.cn

Highlights

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

- This study models technology-transfer performance as a hierarchical system integrating expert knowledge with institutional evidence.
- The AHP–MAP framework enables auditable cross-level attribution and robust organizational evaluation under uncertainty.

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

- Across 36 new R&D institutions, MAP calibration preserves AHP consistency while improving discrimination, explanatory performance, and ranking stability.
- The framework supports transparent policy evaluation and institutional management by balancing structured expert priors with observed evidence.

Abstract

Evaluating technology-transfer performance in new R&D institutions requires an evaluation system that remains auditable while allowing empirical calibration with observed outcomes. This study proposes an AHP-MAP framework that combines the analytic hierarchy process with maximum a posteriori weight calibration. The primary and local AHP weights define the modes of Dirichlet priors, while a two-part occurrence–magnitude model links composite scores to zero-inflated and long-tailed monetized technology-transfer outputs. A common prior-strength parameter is selected from a prespecified candidate set through stratified four-fold cross-validation. Using 23 indicators from 36 institutions, the proposed method improves post-selection out-of-fold performance relative to the fixed AHP baseline. The Spearman correlation increases from 0.655 to 0.712, the area under the receiver operating characteristic curve increases from 0.742 to 0.805, the root mean square error for positive-output magnitude decreases from 1.481 to 1.453, and the mean joint negative log-likelihood decreases from 2.006 to 1.974. Compared with the purely data-driven model under the same formulation with $\kappa = 0$, AHP-MAP remains closer to the original AHP structure and avoids excessive weight concentration. Its additive structure also enables exact decomposition into dimension-level and indicator-level contributions. Multiple-initialization, transformation, bootstrap, jackknife, and Monte Carlo analyses provide complementary evidence on weight and ranking stability across alternative analytical



Academic Editors: Jun Li, Sadia Riaz, Sawsan Malik and Shijie Zhang

Received: 8 July 2026

Revised: 1 August 2026

Accepted: 4 August 2026

Published: 5 August 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 \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

settings and sample sizes. The framework provides a transparent and auditable basis for performance assessment, weight adjustment, and institutional diagnosis.

Keywords: new R&D institutions; technology transfer; analytic hierarchy process; maximum a posteriori estimation; interpretable statistical calibration; multi-criteria evaluation; ranking stability

1. Introduction

Technology transfer is a key mechanism through which research organizations convert scientific resources into industrial value, economic outcomes, and regional innovation capacity. Recent studies on university technology transfer indicate that this field is moving from descriptive analyses of institutional arrangements and output statistics toward more systematic explanations of governance mechanisms, organizational capabilities, and performance outcomes [1]. For innovation-driven and publicly funded research organizations, technology-transfer efficiency depends not only on the quantity of research outputs, but also on how institutional resources are organized, governed, and converted into market-oriented value [2]. New R&D institutions are expected to bridge universities, enterprises, governments, and markets. Their performance evaluation therefore cannot rely on a single output indicator, but must jointly reflect talent resources, operational capacity, R&D activities, and technology-transfer outcomes.

Technology-transfer performance evaluation faces two methodological challenges. The first challenge concerns structural auditability. The indicator system is usually hierarchical and policy-oriented, so the evaluation results must remain traceable, comparable, and auditable. Expert-based methods such as the analytic hierarchy process (AHP) are useful for preserving such structural semantics, but the resulting weights may be affected by subjective judgment and may not fully reflect observed institutional performance. The second challenge concerns output heterogeneity. Monetized technology-transfer outputs often exhibit zero inflation and long-tailed distributions. Many institutions may have no monetized transfer output in a given period, whereas a small number of institutions may generate very high transfer income. This distributional feature makes it difficult for a purely linear and expert-weighted scoring system to capture the empirical relationship between indicator performance and actual transfer output.

Recent technology-transfer research has increasingly shifted from simple input–output statistics toward mechanism-oriented explanation. At the organizational level, the ambidexterity of university technology transfer offices has been found to influence transfer performance, suggesting that commercialization outcomes depend on the configuration of organizational capabilities rather than on isolated output indicators [3]. At the collaboration level, social capital in university–industry collaboration shapes cross-industry knowledge sharing, indicating that relational resources and knowledge-sharing pathways are important mechanisms behind technology-transfer performance [4]. At the institutional governance level, contractual provisions concerning intellectual property, disclosure, and confidentiality affect the boundary conditions of knowledge diffusion in publicly funded research [5]. Evidence from NASA technology licensing further shows that licensing policies may generate commercialization effects and subsequent innovation spillovers [6]. These studies suggest that performance evaluation should not only rank institutions, but also provide traceable evidence concerning the indicators associated with performance differences.

Multi-criteria decision analysis (MCDA) provides a structured basis for aggregating heterogeneous indicators and has developed from classical weighting and ranking methods

toward robust ordinal regression and preference-robust decision support [7]. This development is particularly relevant to technology-transfer evaluation because policy-oriented rankings should retain their hierarchical meanings while remaining stable under changes in preferences, parameters, and samples. Recent research on explainable multi-criteria decision making further emphasizes transparency throughout data preparation, weight determination, score aggregation, and decision support [8]. However, existing approaches rarely combine expert-derived hierarchical weights, zero-inflated and long-tailed transfer outputs, prior-constrained statistical calibration, and ranking-stability analysis within a unified evaluation process. Fixed expert weights cannot incorporate observed performance evidence directly, whereas unrestricted data-driven weighting may produce excessive weight concentration and weaken consistency with the original hierarchical evaluation system.

To address this research gap, this study proposes an AHP-MAP framework for technology-transfer performance evaluation in new R&D institutions. The primary and local AHP weights define the modes of hierarchical Dirichlet priors and are calibrated using observed monetized technology-transfer outputs. A two-part occurrence–magnitude mechanism separately models whether monetized output occurs and the magnitude of positive output. A common prior-strength parameter controls the balance between the original expert structure and the observed evidence and is selected from a prespecified candidate set through stratified four-fold cross-validation. The calibrated score retains the additive hierarchical structure of AHP, which permits exact decomposition into primary-dimension and secondary-indicator contributions.

The study is guided by the following research questions.

- **RQ1.** How can expert-derived AHP weights and observed performance evidence be integrated while preserving the hierarchical semantics and auditability of the existing evaluation system?
- **RQ2.** Does prior-constrained MAP calibration improve the empirical relationship between composite scores and monetized technology-transfer outputs relative to the fixed AHP baseline?
- **RQ3.** Do the calibrated weights and institutional rankings remain structurally consistent and stable across alternative prior strengths, initial values, indicator transformations, resampling procedures, institutional omissions, and prespecified sample-size conditions?

The main contributions of this study are threefold. First, it develops a prior-constrained hierarchical calibration method that treats expert-derived AHP weights as structural prior information and estimates primary and local weights under simplex constraints. This formulation incorporates observed performance evidence without replacing the organization and auditability of the original indicator system. Second, it introduces an occurrence–magnitude mechanism for zero-inflated and long-tailed monetized outputs and establishes a cross-validation procedure for selecting the common prior-strength parameter. Third, it provides an empirical evaluation using data from 36 new R&D institutions. The model comparison includes fixed AHP, AHP-MAP, and a purely data-driven specification that uses the same occurrence–magnitude likelihood, hierarchical simplex constraints, fold partition, and optimization procedure as AHP-MAP, with the prior strength fixed at $\kappa = 0$. The evaluation further examines multiple-initialization convergence, indicator-transformation sensitivity, weight calibration, institutional ranking, contribution decomposition, bootstrap and jackknife stability, descriptive institutional-type differences, and Monte Carlo sample-size experiments under prespecified simulation settings.

The remainder of this paper is organized as follows. Section 2 reviews the related literature and identifies the research gap. Section 3 formulates the technology-transfer performance evaluation problem and develops the research hypotheses. Section 4 presents

the AHP-MAP methodology, including data transformation, AHP prior construction, occurrence–magnitude modeling, prior-strength selection, MAP optimization, and contribution decomposition. Section 5 reports the empirical design, model comparison, weight calibration, institutional ranking, and robustness analyses. Section 6 discusses the theoretical and managerial implications, implementation process, and methodological boundaries. Section 7 concludes the study and outlines future research directions.

2. Literature Review

2.1. Technology-Transfer Performance Evaluation

Technology-transfer performance evaluation is closely related to innovation management, university–industry collaboration, intellectual property governance, and regional innovation policy. Traditional evaluation systems commonly use observable indicators such as patents, technology contracts, technology-transfer income, licensing revenue, incubated enterprises, and R&D investment. These indicators provide a basis for cross-institutional comparison, but they do not fully represent the organizational process through which scientific resources are transformed into market-oriented outcomes. Research on university commercialization shows that outsourcing commercialization activities to specialized technology transfer offices can alter commercialization performance by introducing professional capabilities and external organizational resources [9]. Systematic evidence on university–industry R&D collaboration also indicates that performance measurement should cover different collaboration outputs and knowledge-transfer channels rather than rely exclusively on final commercialization indicators [10]. Living Labs further demonstrate that platform-based and participatory mechanisms influence knowledge recombination, application development, and technology-transfer patterns [11].

The organizational role of technology transfer offices has also expanded. These offices may operate as coordinators, brokers, and active intermediaries between academic organizations and market actors [12]. Evidence from National Institutes of Health proof-of-concept centers suggests that intermediary support can reduce early-stage uncertainty and facilitate the movement of scientific discoveries toward commercial applications [13]. Commercialization outcomes from university–industry collaboration are also shaped by combinations of organizational conditions rather than by isolated factors [14]. Technology-transfer performance evaluation should therefore account for the capabilities and organizational mechanisms that support the generation and realization of transfer outcomes.

Intellectual property governance constitutes another important component of technology-transfer evaluation. Academic patent commercialization involves invention disclosure, ownership arrangements, licensing decisions, market demand, and institutional incentives [15]. Recent measurement research emphasizes that invention commercialization should be identified through market-oriented use and implementation rather than solely through patent counts or invention disclosures [16]. Intellectual property arrangements may also produce heterogeneous social and institutional consequences, which means that stronger protection does not automatically correspond to higher public value [17]. University patent litigation may further increase uncertainty and affect subsequent technology-transfer activities [18]. These findings support an evaluation structure that combines commercialization outcomes with organizational, scientific, and institutional capabilities.

These issues are especially relevant to new R&D institutions, which usually combine public missions, market-oriented operating mechanisms, and flexible organizational structures. Empirical research on publicly supported technology transfer offices shows that organizational capacity and public support jointly influence technology-transfer performance [19]. Evidence from the biopharmaceutical industry also indicates that university–industry interaction contributes to firm-level innovation performance [20]. The value

created through technology transfer therefore depends on both internal institutional capabilities and external industrial relationships. An evaluation system for new R&D institutions should accordingly integrate talent resources, organizational development, R&D activities, and technology-transfer realization within a coherent hierarchical structure.

2.2. AHP and Robust Multi-Criteria Evaluation

Multi-criteria decision analysis provides a structured methodological foundation for evaluating systems involving multiple objectives and heterogeneous indicators. Policy-oriented evaluations often require the aggregation of indicators with different units, scales, and managerial meanings. The analytic hierarchy process is widely used in such settings because it decomposes a complex evaluation problem into hierarchical levels and derives weights from pairwise expert comparisons. Subjective weighting methods remain important in multi-criteria decision making because they make expert preferences explicit and auditable [21]. Preference-elicitation approaches based on deck-of-cards ordinal regression similarly demonstrate the importance of transparent preference construction in multi-criteria models [22]. Multiple-criteria sorting methods further extend structured evaluation from simple ranking to classification and ordered assignment [23].

AHP-based evaluation nevertheless depends on the consistency and representativeness of expert judgments. Once the elicited weights are fixed, the evaluation system has limited capacity to respond to observed differences between the original preference structure and institutional performance evidence. Robust ordinal regression addresses this issue from the perspective of preference uncertainty by identifying conclusions that remain valid across multiple preference models compatible with the available information [24]. Interactive preference learning combines ordinal regression with sequential learning and supports multi-criteria choices under evolving preference information [25]. These developments show that indicator weights and institutional rankings can be examined as uncertain and updateable components rather than as immutable outputs of a single elicitation procedure.

Ranking robustness is particularly important when evaluation results influence resource allocation, policy support, and institutional incentives. Distributionally robust optimization has been used to reduce ranking sensitivity under uncertain preference distributions in group decision making [26]. The integration of outranking methods and Bayesian networks provides another approach to representing uncertainty in multi-criteria decision support [27]. Hierarchical preference-elicitation procedures can also combine heterogeneous expert judgments and different weighting approaches within a unified multi-attribute utility structure [28]. These studies provide methodological support for preserving expert knowledge while allowing uncertainty and new information to influence evaluation outcomes.

Sensitivity analysis offers a direct means of examining whether scores and rankings remain stable under changes in weights and model parameters. Robustness coefficients can quantify the stability of multi-criteria results and support comparisons among alternative rankings [29]. Simultaneous modification of multiple weights or preference parameters may produce ranking changes that cannot be detected through one-factor-at-a-time analysis [30]. Recent innovation-performance research also demonstrates the importance of alternative measurements, robustness checks, and supplementary analyses when empirical conclusions depend on model specification [31].

These developments provide effective approaches for examining preference uncertainty and ranking stability, but they offer limited guidance on calibrating an existing hierarchical expert-weight system against observed institutional outcomes. When the original AHP structure must remain the auditable reference, the primary and local AHP weight vectors can define the modes of Dirichlet priors under positive prior strength. A common

prior-strength parameter can then regulate the permissible departure from the original hierarchy. This formulation makes the balance between expert judgment and observed evidence explicit and allows the degree of calibration to be selected and examined systematically.

2.3. Interpretable Evaluation and Prior-Constrained Calibration

Interpretability is a central requirement in management evaluation and decision support because users must be able to trace how input information produces scores, rankings, and resource-allocation recommendations. Research on explainable artificial intelligence in operational research emphasizes performance, attribution, and responsibility as complementary requirements for analytical decision systems [32]. The XAI 2.0 research agenda similarly highlights the usability, verifiability, and decision relevance of model explanations [33]. Explainable decision-support research further shows that transparency and user understanding can improve the practical use of model-generated recommendations [34].

Existing interpretability methods follow different analytical routes. Post-hoc methods based on Shapley values explain predictions generated by complex statistical or machine-learning models. Research on Shapley values has clarified how different attribution formulations correspond to different sensitivity assumptions [35]. Practical SHAP applications demonstrate how individual predictions can be decomposed into feature contributions [36]. Explainable deep-learning studies further combine predictive models, feature attribution, and supplementary statistical analysis to support policy interpretation [37]. These methods are particularly useful when the underlying predictor contains nonlinear interactions or internal structures that cannot be inspected directly.

The present study adopts a different route to interpretability. It does not rely on a black-box predictor that requires post-hoc attribution. Its interpretability follows from the additive hierarchical score structure, through which each institutional score can be decomposed exactly into primary-dimension and secondary-indicator contributions. Changes introduced by calibration can also be traced through primary, local, and global weight drift. This structural form of explanation is consistent with explainable multi-criteria decision making, which emphasizes transparency across data preparation, preference modeling, evaluation, and decision support [8].

Prior-constrained calibration extends this transparent score structure by incorporating observed outcome evidence without discarding the original expert hierarchy. Maximum a posteriori estimation combines an outcome-based likelihood with Dirichlet priors defined on the primary and local weight simplexes. Under the parameterization adopted in this study, the AHP weight vectors define the prior modes when the prior-strength parameter is positive. Larger prior strength produces stronger shrinkage toward the expert structure, whereas smaller prior strength permits greater adjustment according to the observed outcomes. Setting the prior strength to zero removes the informative AHP component and yields the corresponding purely data-driven specification under the same likelihood and hierarchical constraints.

The distribution of monetized technology-transfer outputs also requires a model that distinguishes different stages of transfer realization. Some institutions report no monetized output, while the positive observations exhibit a strongly right-skewed distribution. A two-part occurrence–magnitude specification separates the probability that monetized output occurs from the magnitude observed after occurrence. This separation reflects the substantive distinction between establishing a realizable technology-transfer pathway and increasing the monetary scale of an existing pathway. It also permits model performance to be evaluated through occurrence discrimination, positive-output magnitude fitting, joint predictive loss, and rank association rather than through a single undifferentiated criterion.

2.4. Research Gap and Analytical Positioning

The reviewed studies establish important foundations for technology-transfer measurement, hierarchical multi-criteria evaluation, preference-robust decision support, and transparent analytical interpretation. Technology-transfer research identifies the organizational, collaborative, and institutional conditions associated with commercialization outcomes. AHP and related MCDA methods provide auditable hierarchical weighting structures, while robust MCDA research offers methods for examining uncertainty and ranking stability. Explainable modeling research further clarifies how analytical outputs can be made traceable to their inputs.

Four gaps remain at the intersection of these research streams. Technology-transfer evaluation commonly relies on fixed indicator weights or direct output statistics, while limited attention has been given to treating expert-derived primary and local weights as prior information that can be updated by observed evidence. Probabilistic and robust MCDA studies mainly address uncertain preferences, decision dependencies, or alternative ranking structures and rarely combine hierarchical AHP priors with a two-part model for zero-inflated and right-skewed technology-transfer outputs. Existing studies also provide limited guidance for selecting prior strength from empirical data, which may lead to either insufficient calibration or excessive departure from the original evaluation structure. Institutional evaluations further tend to report a single set of weights and rankings without jointly examining parameter sensitivity, initialization dependence, transformation sensitivity, resampling stability, influential observations, and performance under different sample-size conditions.

The present study addresses these gaps through a hierarchical AHP-MAP framework. The primary and local AHP weights define the modes of Dirichlet priors under positive prior strength, while a common prior-strength parameter regulates the degree of weight adjustment and is selected from a prespecified candidate set through stratified cross-validation. Monetized technology-transfer output is represented through separate occurrence and magnitude components. The resulting additive scores retain exact dimension-level and indicator-level decomposition. The empirical analysis evaluates predictive performance and structural consistency together with multiple-initialization convergence, indicator-transformation sensitivity, bootstrap and jackknife stability, descriptive institutional-type differences, and Monte Carlo sample-size experiments under prespecified simulation settings.

3. Problem Formulation

This section formulates technology-transfer performance evaluation in new R&D institutions as a prior-constrained and evidence-calibrated multi-criteria evaluation problem. The formulation retains the hierarchical structure of the original AHP evaluation system while incorporating observed monetized technology-transfer outputs into weight calibration. Figure 1 illustrates the relationships among institutional indicators, hierarchical weights, output evidence, and calibrated evaluation results.

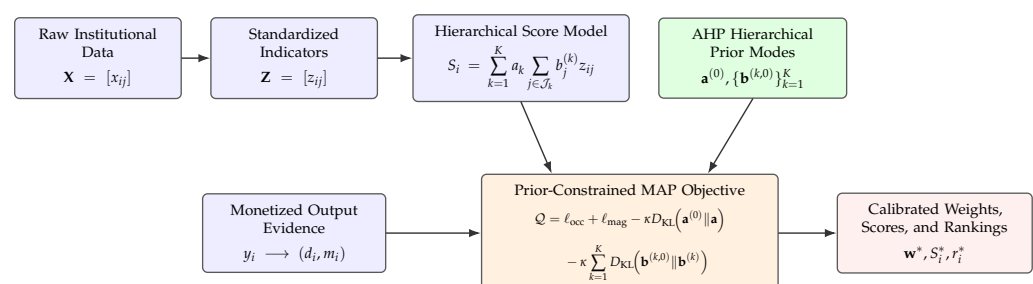


Figure 1. Mathematical structure of prior-constrained evidence calibration.

3.1. Evaluation Setting and Hierarchical Score Model

Let $\mathcal{I} = \{1, 2, \dots, N\}$ denote the set of evaluated new R&D institutions, and let $\mathcal{J} = \{1, 2, \dots, M\}$ denote the set of evaluation indicators. The raw observation of institution $i \in \mathcal{I}$ for indicator $j \in \mathcal{J}$ is denoted by x_{ij} , and the resulting data matrix is $\mathbf{X} = [x_{ij}] \in \mathbb{R}_+^{N \times M}$.

The indicator system contains K primary dimensions indexed by $\mathcal{K} = \{1, 2, \dots, K\}$. In this study, the four primary dimensions represent talent reserve, operation and development, R&D activities, and technology transfer. Each dimension k contains a secondary-indicator subset $\mathcal{J}_k$. These subsets form a non-overlapping partition of the complete indicator set, so that $\mathcal{J} = \bigcup_{k=1}^K \mathcal{J}_k$ and $\mathcal{J}_k \cap \mathcal{J}_l = \emptyset$ for $k \neq l$.

All indicators are benefit-type indicators. Because they differ substantially in scale and distributional shape, a logarithmic transformation and min–max normalization are applied before score aggregation.

$$\tilde{x}_{ij} = \log(1 + x_{ij}), \quad (1a)$$

$$z_{ij} = \frac{\tilde{x}_{ij} - \min_{m \in \mathcal{I}} \tilde{x}_{mj}}{\max_{m \in \mathcal{I}} \tilde{x}_{mj} - \min_{m \in \mathcal{I}} \tilde{x}_{mj} + \varepsilon_z}, \quad (1b)$$

where $\varepsilon_z = 10^{-12}$ prevents numerical instability when an indicator has little or no cross-institutional variation. The standardized matrix is denoted by $\mathbf{Z} = [z_{ij}] \in [0, 1]^{N \times M}$.

Let $\mathbf{a} = (a_1, \dots, a_K)^\top$ denote the primary-dimension weight vector, and let $\mathbf{b}^{(k)} = (b_j^{(k)})_{j \in \mathcal{J}_k}$ denote the local weight vector within dimension k . These vectors satisfy

$$a_k \geq 0, \quad \sum_{k=1}^K a_k = 1, \quad b_j^{(k)} \geq 0, \quad \sum_{j \in \mathcal{J}_k} b_j^{(k)} = 1.$$

The dimension-level score, overall institutional score, and global indicator weight are defined jointly as

$$\begin{aligned} S_{ik} &= \sum_{j \in \mathcal{J}_k} b_j^{(k)} z_{ij}, \\ S_i &= \sum_{k=1}^K a_k S_{ik} = \sum_{k=1}^K a_k \sum_{j \in \mathcal{J}_k} b_j^{(k)} z_{ij} = \sum_{j=1}^M w_j z_{ij}, \\ w_j &= a_k b_j^{(k)}, \quad j \in \mathcal{J}_k. \end{aligned} \quad (2)$$

The hierarchical formulation preserves the meanings of the primary dimensions and the local relationships among their secondary indicators. It also permits each overall score to be traced through dimension-level and indicator-level contributions.

3.2. Monetized Output Evidence and Occurrence–Magnitude Representation

Let $y_i \geq 0$ denote the monetized technology-transfer output of institution i . The output evidence is used to calibrate the hierarchical weights, while all 23 indicators continue to determine the resulting institutional score.

Monetized output includes zero observations and a strongly right-skewed distribution among positive observations. The outcome is therefore represented through separate occurrence and magnitude components.

$$d_i = \mathbb{I}(y_i > 0), \quad d_i \sim \text{Bernoulli}(\pi_i), \quad \pi_i = \sigma(\beta_0 + \beta_1 S_i), \quad (3a)$$

$$m_i = \log(1 + y_i), \quad m_i \mid d_i = 1 \sim \mathcal{N}(\mu_i, \sigma_y^2), \quad \mu_i = \gamma_0 + \gamma_1 S_i. \quad (3b)$$

Here, $\sigma(u) = 1/[1 + \exp(-u)]$ is the logistic function. The mechanism parameters are collected in $\Theta = \{\beta_0, \beta_1, \gamma_0, \gamma_1, \sigma_y\}$.

The occurrence and magnitude log-likelihoods are

$$\ell_{\text{occ}} = \sum_{i=1}^N [d_i \log \pi_i + (1 - d_i) \log(1 - \pi_i)], \quad (4a)$$

$$\ell_{\text{mag}} = -\frac{1}{2} \sum_{i:d_i=1} \left[\log(2\pi\sigma_y^2) + \frac{(m_i - \mu_i)^2}{\sigma_y^2} \right]. \quad (4b)$$

The combined output-evidence term is $\ell(\Theta, \mathbf{a}, \{\mathbf{b}^{(k)}\}) = \ell_{\text{occ}} + \ell_{\text{mag}}$. The occurrence component represents whether an institution realizes monetized technology transfer, while the magnitude component represents the monetary scale achieved after occurrence.

The institutional indicators and monetized outputs in the current data matrix correspond to the same reporting period. The occurrence–magnitude mechanism therefore calibrates the hierarchical weights using contemporaneous institutional evidence and supports analysis of same-period score–output relationships. Multi-period observations can further extend this framework by examining the temporal pathways linking earlier R&D inputs to subsequent technology-transfer income.

3.3. Prior-Constrained Calibration Problem

The existing AHP evaluation system supplies the initial hierarchical weight structure. Let $\mathbf{a}^{(0)} = (a_1^{(0)}, \dots, a_K^{(0)})^\top$ denote the AHP-derived primary weights, and let $\mathbf{b}^{(k,0)} = (b_j^{(k,0)})_{j \in \mathcal{J}_k}$ denote the AHP-derived local weights within dimension k .

For two probability vectors $\mathbf{p}$ and $\mathbf{q}$ defined on the same simplex, their Kullback–Leibler divergence is

$$D_{\text{KL}}(\mathbf{p} \parallel \mathbf{q}) = \sum_h p_h \log \frac{p_h}{q_h}.$$

A common prior-strength parameter κ is used for both the primary and local weight vectors. The corresponding Dirichlet priors are parameterized as

$$\begin{aligned} \mathbf{a} \mid \kappa &\sim \text{Dirichlet}(\mathbf{1} + \kappa \mathbf{a}^{(0)}), \\ \mathbf{b}^{(k)} \mid \kappa &\sim \text{Dirichlet}(\mathbf{1} + \kappa \mathbf{b}^{(k,0)}), \quad k = 1, \dots, K. \end{aligned} \quad (5)$$

For $\kappa > 0$, the primary and local AHP weight vectors are the modes of their corresponding Dirichlet priors. After terms independent of the calibrated weights are removed, the joint log-prior is equivalent to a KL-divergence penalty. The resulting objective can be written compactly as

$$\begin{aligned} \mathcal{R}(\mathbf{a}, \{\mathbf{b}^{(k)}\}) &= D_{\text{KL}}(\mathbf{a}^{(0)} \parallel \mathbf{a}) + \sum_{k=1}^K D_{\text{KL}}(\mathbf{b}^{(k,0)} \parallel \mathbf{b}^{(k)}), \\ \log p(\mathbf{a}, \{\mathbf{b}^{(k)}\} \mid \kappa) &= -\kappa \mathcal{R}(\mathbf{a}, \{\mathbf{b}^{(k)}\}) + C, \\ \mathcal{Q} &= \ell_{\text{occ}} + \ell_{\text{mag}} - \kappa \mathcal{R}(\mathbf{a}, \{\mathbf{b}^{(k)}\}), \end{aligned} \quad (6)$$

where C is independent of the calibrated weights.

The calibration problem maximizes Equation (6) over the hierarchical weights and the occurrence–magnitude parameters under the primary and local simplex constraints. Larger values of κ impose stronger shrinkage toward the AHP structure, whereas smaller positive values allow greater adjustment according to the observed output evidence.

When $\kappa = 0$, the informative AHP component is removed and the model becomes the corresponding purely data-driven specification. This specification retains the same occurrence–magnitude likelihood, hierarchical score structure, simplex constraints, data partition, and optimization procedure as AHP-MAP. The comparison therefore isolates the effect of the AHP-derived prior rather than differences in model architecture or estimation conditions.

The calibrated solution is defined as

$$\left(\mathbf{a}^*, \{\mathbf{b}^{(k)*}\}_{k=1}^K, \Theta^*\right) = \arg \max_{\mathbf{a}, \{\mathbf{b}^{(k)}\}, \Theta} \mathcal{Q}. \quad (7)$$

The resulting global weights, institutional scores, and ranks are

$$\begin{aligned} w_j^* &= a_k^* b_j^{(k)*}, \quad j \in \mathcal{J}_k, \\ S_i^* &= \sum_{k=1}^K a_k^* \sum_{j \in \mathcal{J}_k} b_j^{(k)*} z_{ij} = \sum_{j=1}^M w_j^* z_{ij}, \\ r_i^* &= 1 + \sum_{\substack{\ell \in \mathcal{I} \\ \ell \neq i}} \mathbb{I}(S_\ell^* > S_i^*). \end{aligned} \quad (8)$$

A smaller value of r_i^* represents a higher institutional position.

3.4. Model Complexity and Research Hypotheses

The simplex constraints leave $K - 1$ free primary-weight parameters and $\sum_{k=1}^K (|\mathcal{J}_k| - 1) = M - K$ free local-weight parameters. The nominal model complexity is therefore

$$\begin{aligned} d_{\text{weight}} &= (K - 1) + \sum_{k=1}^K (|\mathcal{J}_k| - 1) = M - 1, \\ d_{\text{total}} &= d_{\text{weight}} + 5 = M + 4, \\ \frac{N}{d_{\text{total}}} &= \frac{36}{27} \approx 1.33. \end{aligned} \quad (9)$$

For $K = 4$ and $M = 23$, the hierarchical weight system contains twenty-two nominal free parameters. The occurrence–magnitude mechanism contributes five additional parameters. The complete model therefore contains twenty-seven nominal free parameters for 36 observed institutions.

The Dirichlet prior does not reduce this nominal parameter count. It restricts effective weight flexibility by penalizing departures from the AHP structure and discouraging unsupported movement toward simplex boundaries. The empirical analysis consequently evaluates predictive performance together with weight drift, weight concentration, multiple-initialization convergence, resampling stability, institutional influence, and sensitivity to sample size.

H1. *Relative to the fixed AHP baseline, prior-constrained MAP calibration improves the empirical consistency between institutional composite scores and monetized technology-transfer outputs. The improvement is evaluated through occurrence discrimination, positive-output magnitude fitting, joint predictive loss, and rank association.*

H2. *Relative to the purely data-driven specification, prior-constrained MAP calibration maintains greater consistency with the original AHP hierarchy and produces lower weight drift and less excessive weight concentration.*

H3. *The calibrated weights and institutional rankings exhibit limited sensitivity across alternative prior strengths, initial values, indicator transformations, resampling procedures, institutional omissions, and prespecified simulation settings with different sample sizes.*

The formulated problem requires a calibration procedure that jointly estimates the hierarchical weights and the occurrence–magnitude parameters under simplex and prior constraints. Section 4 presents the AHP-MAP procedure used to obtain the calibrated solution and evaluate the three hypotheses.

4. Methodology

This section presents the AHP-MAP procedure used to solve the prior-constrained calibration problem formulated in Section 3. The procedure combines AHP-based prior construction, occurrence–magnitude MAP calibration, constrained alternating optimization, and hierarchical contribution decomposition.

4.1. Overview of the AHP-MAP Framework

Figure 2 illustrates the computational workflow of the AHP-MAP procedure. Standardized institutional indicators and monetized technology-transfer outputs constitute the empirical inputs. The AHP-derived primary and local weights provide hierarchical prior information. The calibration procedure alternately updates the occurrence–magnitude parameters and the hierarchical weights. The resulting outputs include calibrated weights, institutional scores and rankings, and dimension-level and indicator-level contributions.

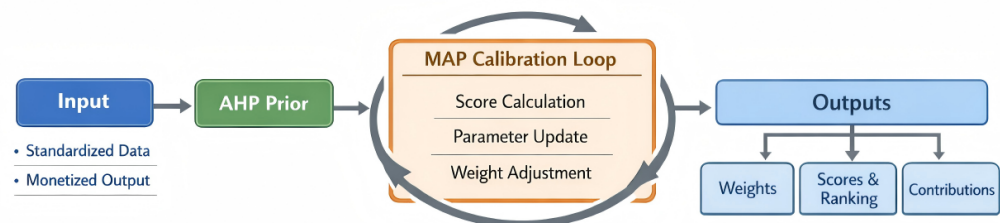


Figure 2. Computational workflow of the AHP-MAP framework.

The hierarchical score structure is maintained throughout the calibration process. The primary weights determine the relative importance of the four evaluation dimensions, while the local weights determine the relative importance of the secondary indicators within each dimension. Their products form the global indicator weights used to calculate institutional scores.

The framework differs from fixed AHP because the expert-derived weights are treated as updateable prior information rather than immutable final weights. It also differs from unrestricted data-driven weighting because the AHP hierarchy continues to regulate the calibrated solution through an explicit prior-strength parameter. The data-driven comparison model retains the same occurrence–magnitude likelihood, hierarchical score structure, simplex constraints, fold partition, and numerical optimizer, while setting $\kappa = 0$.

4.2. AHP Expert Prior Construction

The expert-prior weights are constructed using the analytic hierarchy process. At each level of the hierarchical indicator system, experts compare the relative importance of the corresponding criteria through pairwise comparison matrices. For a group containing q criteria, the judgment matrix and its normalized principal eigenvector are obtained from

$$\begin{aligned}
 \mathbf{A} &= [a_{uv}]_{q \times q}, & a_{uv} > 0, & & a_{uv} = \frac{1}{a_{vu}}, & & a_{uu} = 1, \\
 \mathbf{A}\mathbf{p} &= \lambda_{\max}\mathbf{p}, \\
 p_u^{\text{norm}} &= \frac{p_u}{\sum_{v=1}^q p_v}, & u &= 1, \dots, q.
 \end{aligned} \tag{10}$$

The element a_{uv} represents the relative importance of criterion u over criterion v . The normalized principal eigenvector provides the local priority weights. The consistency of each judgment matrix is evaluated through

$$CI = \frac{\lambda_{\max} - q}{q - 1}, \quad CR = \frac{CI}{RI}, \quad CR < 0.10, \quad (11)$$

where RI denotes the random consistency index corresponding to the order of the judgment matrix.

This procedure yields the primary AHP weight vector $\mathbf{a}^{(0)} = (a_1^{(0)}, \dots, a_K^{(0)})^\top$ and the local AHP weight vectors $\mathbf{b}^{(k,0)} = (b_j^{(k,0)})_{j \in \mathcal{J}_k}$. The corresponding global AHP weight of secondary indicator $j \in \mathcal{J}_k$ is $w_j^{(0)} = a_k^{(0)} b_j^{(k,0)}$.

The primary and local AHP vectors define the modes of the Dirichlet priors used in MAP calibration. They preserve the hierarchical meanings of the original evaluation system without requiring the calibrated weights to remain identical to the expert-derived values.

4.3. MAP Calibration with Dirichlet Shrinkage

The calibrated primary and local weights remain on their corresponding simplexes,

$$a_k \geq 0, \quad \sum_{k=1}^K a_k = 1, \quad b_j^{(k)} \geq 0, \quad \sum_{j \in \mathcal{J}_k} b_j^{(k)} = 1.$$

A common prior-strength parameter κ is applied to the primary and local weight vectors. The prior distributions and the resulting log-posterior objective are expressed jointly as

$$\mathbf{a} \mid \kappa \sim \text{Dirichlet}(\mathbf{1} + \kappa \mathbf{a}^{(0)}), \quad (12a)$$

$$\mathbf{b}^{(k)} \mid \kappa \sim \text{Dirichlet}(\mathbf{1} + \kappa \mathbf{b}^{(k,0)}), \quad k = 1, \dots, K, \quad (12b)$$

$$\begin{aligned} \mathcal{L}_{\text{MAP}} = & \ell_{\text{occ}} + \ell_{\text{mag}} + \kappa \sum_{k=1}^K a_k^{(0)} \log a_k \\ & + \kappa \sum_{k=1}^K \sum_{j \in \mathcal{J}_k} b_j^{(k,0)} \log b_j^{(k)} + C. \end{aligned} \quad (12c)$$

Here, C is independent of the calibrated parameters. Under positive prior strength, the modes of the primary and local Dirichlet distributions correspond exactly to the associated AHP weight vectors. The common value of κ establishes a consistent shrinkage scale across the hierarchical levels and avoids introducing separate tuning parameters for every dimension.

Larger values of κ produce stronger shrinkage toward the AHP hierarchy, while smaller positive values permit greater adjustment according to the observed output evidence. When $\kappa = 0$, both Dirichlet distributions reduce to uniform distributions over their corresponding simplexes, and the AHP-dependent terms disappear from the posterior objective. The resulting specification is therefore data-driven under the same model structure and estimation conditions.

Maximizing Equation (12c) under the hierarchical simplex constraints is equivalent to maximizing the KL-regularized objective in Equation (6). The occurrence and magnitude likelihoods remain those defined in Equations (4a) and (4b).

4.4. Selection of the Prior Strength

The prior strength is selected from the prespecified candidate set $\mathcal{K}_\kappa = \{5, 10, 20, 50, 100, 200\}$. The occurrence indicator d_i is used as the stratification variable. The observations are divided into four folds so that the four zero-output institutions are allocated across the validation folds.

The same fold partition is retained for all candidate values of κ , ensuring that differences in validation performance arise from prior strength rather than from different data partitions. Each candidate is estimated on three folds and evaluated on the remaining fold. Prior-strength selection uses only the resulting validation predictions.

Let $\mathcal{T}_f$ and $\mathcal{V}_f$ denote the calibration and validation sets of fold f . The out-of-fold occurrence and magnitude predictions for institution $i \in \mathcal{V}_f$ are

$$\hat{\pi}_{i,\kappa}^{(-f)} = \sigma \left(\hat{\beta}_{0,\kappa}^{(-f)} + \hat{\beta}_{1,\kappa}^{(-f)} \hat{S}_{i,\kappa}^{(-f)} \right), \quad (13a)$$

$$\hat{\mu}_{i,\kappa}^{(-f)} = \hat{\gamma}_{0,\kappa}^{(-f)} + \hat{\gamma}_{1,\kappa}^{(-f)} \hat{S}_{i,\kappa}^{(-f)}. \quad (13b)$$

The superscript $(-f)$ indicates estimation without validation fold f . The mean joint negative log-likelihood for the validation fold is

$$\begin{aligned} \text{JNLL}_{\kappa,f} = & -\frac{1}{|\mathcal{V}_f|} \left\{ \sum_{i \in \mathcal{V}_f} \left[d_i \log \hat{\pi}_{i,\kappa}^{(-f)} + (1 - d_i) \log \left(1 - \hat{\pi}_{i,\kappa}^{(-f)} \right) \right] \right. \\ & \left. + \sum_{\substack{i \in \mathcal{V}_f \\ d_i=1}} \log \phi \left(m_i; \hat{\mu}_{i,\kappa}^{(-f)}, \hat{\sigma}_{y,\kappa}^{(-f)} \right) \right\}, \end{aligned} \quad (14)$$

where $\phi(m; \mu, \sigma)$ denotes the Gaussian probability density function with mean μ and standard deviation σ .

The mean loss, fold-based standard error, and principal selection rule are

$$\overline{\text{JNLL}}_\kappa = \frac{1}{4} \sum_{f=1}^4 \text{JNLL}_{\kappa,f}, \quad (15a)$$

$$SE_\kappa = \frac{s_\kappa}{\sqrt{4}}, \quad (15b)$$

$$\kappa^* = \arg \min_{\kappa \in \mathcal{K}_\kappa} \overline{\text{JNLL}}_\kappa, \quad (15c)$$

where s_κ is the sample standard deviation of the four fold-level losses.

The one-standard-error region is reported as a sensitivity diagnostic, while the minimum mean out-of-fold joint negative log-likelihood remains the prespecified selection criterion. The selected prior strength is subsequently used for all main-model estimation and comparison procedures.

4.5. Alternating Constrained Optimization

The posterior objective contains two parameter blocks consisting of the occurrence–magnitude parameters and the hierarchical weights. These blocks are updated alternately. At outer iteration t , the current hierarchical weights produce

$$S_i^{(t)} = \sum_{k=1}^K a_k^{(t)} \sum_{j \in \mathcal{J}_k} b_j^{(k,t)} z_{ij}. \quad (16)$$

Given the current scores, the occurrence and magnitude parameters are updated through

$$(\beta_0^{(t+1)}, \beta_1^{(t+1)}) = \arg \max_{\beta_0, \beta_1} \ell_{\text{occ}}, \quad (17a)$$

$$(\gamma_0^{(t+1)}, \gamma_1^{(t+1)}, \sigma_y^{(t+1)}) = \arg \max_{\gamma_0, \gamma_1, \sigma_y} \ell_{\text{mag}}, \quad \sigma_y \geq 0.25. \quad (17b)$$

The occurrence block is estimated through logistic regression. The magnitude block is estimated on the positive-output subset through Gaussian regression.

After updating the mechanism parameters, the primary and local weights are optimized jointly through sequential least squares programming. This update minimizes the negative log-posterior while enforcing all primary and local simplex constraints.

$$\begin{aligned} (\mathbf{a}^{(t+1)}, \{\mathbf{b}^{(k,t+1)}\}_{k=1}^K) &= \arg \min_{\mathbf{a}, \{\mathbf{b}^{(k)}\}} -\mathcal{L}_{\text{MAP}}(\mathbf{a}, \{\mathbf{b}^{(k)}\}, \Theta^{(t+1)}) \\ \text{subject to } a_k &\geq 0, \quad \sum_{k=1}^K a_k = 1, \\ b_j^{(k)} &\geq 0, \quad \sum_{j \in \mathcal{J}_k} b_j^{(k)} = 1. \end{aligned} \quad (18)$$

Let $w_j^{(t)} = a_k^{(t)} b_j^{(k,t)}$ for $j \in \mathcal{J}_k$. Convergence is assessed jointly through the relative posterior-objective change and the maximum global-weight change,

$$\begin{aligned} \Delta_{\mathcal{L}}^{(t)} &= \frac{|\mathcal{L}_{\text{MAP}}^{(t)} - \mathcal{L}_{\text{MAP}}^{(t-1)}|}{|\mathcal{L}_{\text{MAP}}^{(t-1)}| + \varepsilon_c}, \\ \Delta_w^{(t)} &= \max_{j \in \mathcal{J}} |w_j^{(t)} - w_j^{(t-1)}|, \\ \Delta_{\mathcal{L}}^{(t)} &< \delta_{\mathcal{L}} \quad \text{and} \quad \Delta_w^{(t)} < \delta_w. \end{aligned} \quad (19)$$

The numerical constant is $\varepsilon_c = 10^{-12}$, and both convergence thresholds are set to $\delta_{\mathcal{L}} = \delta_w = 10^{-7}$. The maximum number of outer alternating iterations is 30. The SLSQP function tolerance is 10^{-9} , and the maximum number of inner SLSQP iterations is 120.

4.6. Convergence and Initialization Assessment

For fixed hierarchical weights, the institutional scores are fixed, and the occurrence and magnitude parameters are estimated through low-dimensional regression subproblems. The logistic occurrence log-likelihood is concave in (β_0, β_1) . Conditional on the positive-output observations, the Gaussian magnitude block has a closed-form regression structure subject to the lower bound imposed on σ_y .

The complete posterior objective is not jointly concave in all hierarchical parameters because the primary and local weights enter each global weight through the product $w_j = a_k b_j^{(k)}$. The alternating procedure therefore seeks a numerically stable stationary solution under the simplex constraints rather than claiming a guaranteed global optimum.

Initialization sensitivity is assessed using 22 starting points,

$$\mathcal{H} = \{h_{\text{AHP}}, h_{\text{uniform}}, h_{\text{near},1}, \dots, h_{\text{near},10}, h_{\text{diffuse},1}, \dots, h_{\text{diffuse},10}\}.$$

The AHP initialization uses $\mathbf{a}^{(0,h_{\text{AHP}})} = \mathbf{a}^{(0)}$ and $\mathbf{b}^{(k,0,h_{\text{AHP}})} = \mathbf{b}^{(k,0)}$. The uniform initialization assigns $a_k^{(0,h_{\text{uniform}})} = 1/K$ and $b_j^{(k,0,h_{\text{uniform}})} = 1/|\mathcal{J}_k|$. The AHP-neighbor starts are generated through small random perturbations of the AHP weights followed by simplex normalization. The diffuse starts are sampled from symmetric Dirichlet distributions.

For successful initialization h , let $\mathcal{L}_h^{\text{final}}$ denote the converged posterior objective. The retained run is

$$h^* = \arg \max_{h \in \mathcal{H}_{\text{success}}} \mathcal{L}_h^{\text{final}}, \quad (20)$$

where $\mathcal{H}_{\text{success}}$ contains the optimization runs that satisfy the numerical success criteria.

The converged runs are compared using their posterior objectives, global-weight differences, institutional-ranking correlations, and Top-10 overlap values. Agreement across structurally different starting points provides numerical evidence that the reported solution is not driven by a single initialization, although it does not constitute a proof of global optimality. The complete procedure for prior-strength selection, alternating calibration, initialization assessment, and final-solution determination is summarized in Algorithm 1.

Algorithm 1: Cross-validated prior-strength selection and AHP-MAP calibration

Input: Standardized indicator matrix $\mathbf{Z}$; outputs $\{y_i\}_{i=1}^N$; indicator partition $\{\mathcal{I}_k\}_{k=1}^K$; AHP weights $\mathbf{a}^{(0)}$ and $\{\mathbf{b}^{(k,0)}\}_{k=1}^K$; candidate set $\mathcal{K}_\kappa$; four stratified folds; $T_{\max}$; $\delta_{\mathcal{L}}$; δ_w .

Output: Selected κ^* ; calibrated weights; mechanism parameters; institutional scores; rankings; contribution decompositions.

- 1 Construct $d_i = \mathbb{I}(y_i > 0)$ and $m_i = \log(1 + y_i)$ for observations satisfying $d_i = 1$;
- 2 **foreach** $\kappa \in \mathcal{K}_\kappa$ **do**
- 3 **for** $f \leftarrow 1$ **to** 4 **do**
- 4 Form calibration set $\mathcal{T}_f$ and validation set $\mathcal{V}_f$;
- 5 Initialize the hierarchical weights using the AHP prior;
- 6 Fit AHP-MAP on $\mathcal{T}_f$ with prior strength κ ;
- 7 Obtain $\hat{\tau}_{i,\kappa}^{(-f)}$ and $\hat{\mu}_{i,\kappa}^{(-f)}$ for $i \in \mathcal{V}_f$;
- 8 Compute JNLL $_{\kappa,f}$ using Equation (14);
- 9 **end**
- 10 Compute $\overline{\text{JNLL}}_\kappa$ using Equation (15a);
- 11 **end**
- 12 Select κ^* using Equation (15c);
- 13 Construct the 22-member initialization set $\mathcal{H}$;
- 14 **foreach** $h \in \mathcal{H}$ **do**
- 15 Initialize $\mathbf{a}^{(0,h)}$, $\{\mathbf{b}^{(k,0,h)}\}_{k=1}^K$, and $\Theta^{(0,h)}$;
- 16 **for** $t \leftarrow 1$ **to** $T_{\max}$ **do**
- 17 Compute institutional scores using Equation (16);
- 18 Update the occurrence parameters using Equation (17a);
- 19 Update the magnitude parameters using Equation (17b);
- 20 Update the hierarchical weights using Equation (18);
- 21 Compute $\Delta_{\mathcal{L}}^{(t,h)}$ and $\Delta_w^{(t,h)}$ using Equation (19);
- 22 **if** $\Delta_{\mathcal{L}}^{(t,h)} < \delta_{\mathcal{L}}$ **and** $\Delta_w^{(t,h)} < \delta_w$ **then**
- 23 **break**
- 24 **end**
- 25 **end**
- 26 Store the optimization status, posterior objective, weights, scores, and rankings;
- 27 **end**
- 28 Select the successful run with the largest posterior objective using Equation (20);
- 29 Set the retained parameters as $\mathbf{a}^*$, $\{\mathbf{b}^{(k)*}\}_{k=1}^K$, and Θ^* ;
- 30 Compute the final scores, rankings, and contribution decompositions;

4.7. Evaluation Outputs and Contribution Decomposition

The calibrated global weights, institutional scores, and rankings follow Equation (8). The institution-specific contributions, normalized dimension shares, and global-weight changes are defined by

$$C_{ij}^* = w_j^* z_{ij}, \quad j = 1, \dots, M, \quad (21a)$$

$$C_{ik}^* = a_k^* \sum_{j \in \mathcal{J}_k} b_j^{(k)*} z_{ij}, \quad k = 1, \dots, K, \quad (21b)$$

$$S_i^* = \sum_{j=1}^M C_{ij}^* = \sum_{k=1}^K C_{ik}^*, \quad (21c)$$

$$\rho_{ik}^* = \frac{C_{ik}^*}{S_i^* + \varepsilon_z}, \quad (21d)$$

$$\Delta w_j = w_j^* - w_j^{(0)}. \quad (21e)$$

A positive Δw_j indicates that calibration increases the global weight of indicator j , whereas a negative value indicates a decrease relative to the original AHP system.

Equation (21c) verifies the exact arithmetic reconstruction of every institutional score. This property demonstrates that the calibrated additive evaluation system is fully traceable at both the dimension and indicator levels. The resulting contribution values provide a transparent account of how each institutional score is formed and support auditable comparisons of the relative roles of different dimensions and indicators in the final evaluation.

4.8. Implementation Settings

The numerical settings used for model estimation, convergence assessment, and robustness analysis are consolidated in Table 1.

Table 1. Numerical configuration for AHP-MAP estimation and robustness analyses.

Setting	Value	Purpose
Numerical stability constants	$\varepsilon_z = 10^{-12}, \varepsilon_c = 10^{-12}$	Prevention of zero denominators in normalization and relative objective-change calculations
Convergence thresholds	$\delta_{\mathcal{L}} = 10^{-7}, \delta_w = 10^{-7}$	Termination based on posterior-objective and global-weight changes
SLSQP configuration	Tolerance 10^{-9} ; maximum 120 iterations	Constrained optimization of the primary and local weights
Alternating-update configuration	Maximum 30 iterations; minimum $\sigma_y = 0.25$	Outer block updates and numerical stabilization of the magnitude model
Prior-strength configuration	Candidates $\{5, 10, 20, 50, 100, 200\}$; selected value 50	Cross-validated control of shrinkage toward the AHP hierarchy
Prior-strength selection	Stratified four-fold validation; minimum mean JNLL	Selection based on post-selection out-of-fold predictive loss
Initialization configuration	22 starting points; random seed 42	Assessment of dependence on AHP, uniform, neighboring, and diffuse initial values
Bootstrap analysis	500 replications	Evaluation of weight and institutional-ranking stability
Monte Carlo analysis	200 replications per sample size	Evaluation of weight and ranking recovery under prespecified sample-size conditions

5. Empirical Study

This section evaluates the AHP-MAP framework using data from new R&D institutions. The empirical analysis proceeds from data characteristics and AHP baseline construction to transformation selection, prior-strength determination, numerical convergence, out-of-fold model comparison, weight calibration, institutional ranking, resampling stability, sample-size sensitivity, and institutional-type analysis.

5.1. Data, Baselines, and Evaluation Metrics

5.1.1. Sample and Variables

The empirical sample consists of 36 new R&D institutions. The evaluation system contains 23 secondary indicators grouped into four primary dimensions, including talent reserve, operation and development, R&D activities, and technology transfer. Complete observations are available for all institutions and indicators, forming an empirical matrix containing 828 institutional-indicator observations. Monetized technology-transfer output is also available for all 36 institutions. Among them, 32 institutions report positive monetized output and four report zero output.

All secondary indicators are benefit-type indicators. The raw indicators differ substantially in unit, scale, zero frequency, and distributional shape. The logarithmic transformation and min–max normalization defined in Equations (1a) and (1b) are therefore applied before score aggregation and MAP calibration.

Table 2 reports the descriptive statistics. The institutional data show substantial variation across talent resources, operating performance, R&D activity, and technology-transfer realization. The mean absolute skewness across the 23 raw indicators is 4.115. Several technology-transfer indicators contain a substantial proportion of observed zeros, including patent transfer and licensing, technology equity participation enterprises, incubated sci-tech enterprises, and valuation of incubated enterprises. These characteristics support the use of logarithmic indicator transformation and the occurrence–magnitude representation of monetized output.

The indicators and monetized outputs correspond to the same reporting period. The empirical analysis therefore examines contemporaneous score–output relationships and does not interpret the estimated associations as time-lagged or causal effects.

Table 2. Descriptive statistics for the 23 institutional indicators.

Dimension	Indicator	Mean	Std.	Min	Median	Max	Zero Ratio	Skewness
Talent Reserve	R&D Personnel	45.11	46.02	2.00	30.50	224.00	0.000	2.837
Talent Reserve	Senior R&D Titles	15.75	25.41	1.00	9.50	153.00	0.000	4.751
Talent Reserve	High-Level Talent Introduction	4.83	10.13	0.00	1.00	57.00	0.306	4.263
Operation and Development	Total Income	29,960.45	72,730.90	61.86	5016.90	358,357.00	0.000	3.535
Operation and Development	Technology Transfer Income	10,047.33	32,764.89	0.00	640.00	195,386.37	0.389	5.473
Operation and Development	Technology Contract Turnover	5921.88	15,411.16	0.00	699.50	88,460.00	0.250	4.727
Operation and Development	Technology Service Income	2480.58	5766.97	0.00	356.71	30,015.00	0.250	3.646
R&D Activities	R&D Expenditure	4017.79	9203.10	78.00	1363.36	54,237.00	0.000	4.943
R&D Activities	New Innovation Platforms	1.92	2.12	0.00	1.50	8.00	0.278	1.571
R&D Activities	New Investment Amount	1048.14	2706.79	0.00	7.50	12,800.00	0.500	3.276
R&D Activities	New Instruments and Equipment	2038.02	8573.24	0.00	115.69	50,917.00	0.139	5.628
R&D Activities	Total Number of Projects	39.25	132.09	2.00	9.00	794.00	0.000	5.645

Table 2. Cont.

Dimension	Indicator	Mean	Std.	Min	Median	Max	Zero Ratio	Skewness
R&D Activities	Provincial-Level or Above Projects	9.19	20.00	0.00	3.50	103.00	0.028	4.071
R&D Activities	Total Project Funding	3007.13	5204.65	3.00	637.97	23,248.00	0.000	2.535
R&D Activities	Fiscal Support Funding	995.10	2328.06	0.00	275.00	12,306.10	0.028	4.025
R&D Activities	Patent Outputs	26.00	61.95	0.00	9.50	373.00	0.056	5.316
R&D Activities	Scientific Papers	10.78	23.39	0.00	5.00	139.00	0.250	4.977
Technology Transfer	Patent Transfer and Licensing	227.11	1031.53	0.00	0.00	5774.00	0.889	4.976
Technology Transfer	Technology Equity Participation Enterprises	0.44	1.21	0.00	0.00	5.00	0.806	3.296
Technology Transfer	Owned and Implemented Invention Patents	2.22	3.68	0.00	1.00	19.00	0.306	3.278
Technology Transfer	Incubated Sci-Tech Enterprises	1.33	2.84	0.00	0.00	10.00	0.611	2.490
Technology Transfer	Valuation of Incubated Enterprises	4890.97	21,186.77	0.00	0.00	126,500.00	0.667	5.722
Technology Transfer	Served Enterprises in the Past Two Years	173.31	505.47	0.00	14.00	2190.00	0.139	3.676

5.1.2. AHP Baseline and Consistency

Table 3 reports the primary AHP weights and the consistency ratios of the corresponding judgment matrices. Technology transfer receives the highest primary weight, followed by R&D activities, operation and development, and talent reserve. This hierarchy emphasizes realized technology-transfer outcomes while retaining the supporting roles of institutional resources and R&D capacity.

All judgment matrices satisfy the consistency requirement $CR < 0.10$, indicating acceptable consistency in the expert judgments used to construct the hierarchical AHP prior.

Table 3. Primary AHP weights and judgment-matrix consistency.

Judgment Matrix	Primary Weight	Consistency Ratio	Consistency Result
Primary dimensions	–	0.0097	Pass
Talent reserve	0.0954	0.0133	Pass
Operation and development	0.1601	0.0000	Pass
R&D activities	0.2772	0.0055	Pass
Technology transfer	0.4673	0.0173	Pass

5.1.3. Model Parameterization

The hierarchical weight structure contains three primary-weight degrees of freedom and nineteen local-weight degrees of freedom. The occurrence–magnitude mechanism introduces five additional parameters. Table 4 summarizes the sample composition and nominal parameterization.

The complete model contains 27 nominal free parameters for 36 institutions, corresponding to an institution-to-parameter ratio of 1.33. The Dirichlet prior does not reduce the nominal parameter count. It regularizes the hierarchical weights toward the AHP prior modes and limits unsupported movement toward the simplex boundaries.

Table 4. Sample composition and nominal model parameterization.

Category	Measure	Value
Sample composition	Number of institutions	36
	Institutions with positive monetized output	32
	Institutions with zero monetized output	4
	Primary dimensions	4
	Secondary indicators	23
Nominal model parameterization	Primary-weight degrees of freedom	3
	Local-weight degrees of freedom	19
	Total weight degrees of freedom	22
	Occurrence–magnitude parameters	5
	Total nominal free parameters	27
	Institution-to-parameter ratio	1.33

5.1.4. Baseline Models and Evaluation Metrics

Three evaluation models are compared. The fixed AHP baseline directly uses the expert-derived hierarchical weights. The AHP-MAP model calibrates the primary and local weights using the cross-validation-selected prior strength $\kappa = 50$.

The purely data-driven model is defined as the same occurrence–magnitude model with $\kappa = 0$. It uses the same hierarchical score formulation, simplex constraints, fold partition, multiple-initialization procedure, and SLSQP optimizer as AHP-MAP. The comparison therefore isolates the contribution of the informative AHP prior.

The evaluation metrics cover four complementary dimensions. Pearson, Spearman, and Kendall coefficients measure the association between institutional scores and monetized output. The occurrence component is evaluated using the area under the receiver operating characteristic curve (AUC), accuracy, balanced accuracy, Brier score, and log loss. The positive-output magnitude component is evaluated using root mean square error (RMSE), mean absolute error (MAE), and the coefficient of determination R^2 . Joint probabilistic performance is evaluated using the mean joint negative log-likelihood (JNLL).

Structural preservation is evaluated using rank correlation with the fixed AHP result, Top-10 overlap, score correlation, primary-weight L_1 drift, global-weight L_1 drift, and the effective number of weights. The L_1 drift measures the sum of absolute weight changes relative to the AHP baseline. The effective number of weights is calculated as

$$N_{\text{eff}} = \frac{1}{\sum_{j=1}^M w_j^2}. \quad (22)$$

A larger N_{eff} indicates that global importance is distributed across a broader set of indicators. The interquartile range used in the resampling analysis is the difference between the 75th and 25th percentiles of the bootstrap ranks.

5.2. Transformation and Variance Diagnostics

Three transformation procedures are compared, including logarithmic transformation followed by min–max normalization, Yeo–Johnson transformation followed by min–max

normalization, and rank-percentile transformation. Table 5 reports their distributional, fitting, variance, and structural-consistency characteristics.

The mean absolute skewness of the raw indicators is 4.115. The log-based transformation reduces this value to 0.836 and produces the lowest full-sample joint NLL, the highest score-output Spearman coefficient, the highest occurrence AUC, and the lowest magnitude RMSE among the three examined procedures. The Breusch–Pagan p -value is 0.336 under this specification.

The Yeo–Johnson transformation produces the lowest mean absolute skewness, while the rank-percentile transformation provides a fully ordinal representation. The log-based transformation provides the strongest combined performance across model fit and consistency with the principal ranking and weight structure. It is therefore retained in the main analysis.

Table 5. Indicator-transformation, model-fit, variance, and structural-consistency diagnostics.

Diagnostic Category	Metric	Log1p + Min–Max	Yeo–Johnson + Min–Max	Rank Percentile
Distributional diagnostic	Mean absolute skewness	0.836	0.308	0.457
Model-fit diagnostic	Mean joint NLL	1.644	1.741	1.756
Score association	Spearman coefficient	0.768	0.706	0.705
Occurrence performance	AUC	0.930	0.914	0.797
Magnitude performance	RMSE	1.233	1.327	1.243
Variance diagnostic	Breusch–Pagan p -value	0.336	0.219	0.237
Structural consistency	Ranking correlation	1.000	0.964	0.899
Structural consistency	Global-weight L_1 difference	0.000	0.075	0.119

Note. Mean absolute skewness is calculated across the 23 transformed indicators. Ranking correlation and global-weight L_1 difference are measured relative to the log1p–min–max specification. NLL denotes negative log-likelihood, AUC denotes area under the receiver operating characteristic curve, and RMSE denotes root mean square error.

5.3. Prior-Strength Selection and Convergence

The six candidate prior strengths are evaluated using the stratified four-fold procedure specified in Section 4.4. The same fold partition is used for every candidate so that differences in validation performance are attributable to prior strength rather than to different data partitions.

Table 6 reports the selection results. The mean out-of-fold JNLL reaches its minimum value of 1.974 at $\kappa = 50$. The same specification produces an out-of-fold Spearman coefficient of 0.712, an occurrence AUC of 0.805, and a magnitude RMSE of 1.453. The values $\kappa = 50$, 100, and 200 lie within one standard error of the minimum. The prespecified minimum-loss rule selects $\kappa = 50$ for the principal analysis.

Figure 3 presents the mean out-of-fold joint loss and the corresponding fold-based standard errors.

Table 6. Cross-validated performance and selection diagnostics across candidate prior strengths.

Metric	$\kappa = 5$	$\kappa = 10$	$\kappa = 20$	$\kappa = 50$	$\kappa = 100$	$\kappa = 200$
Mean OOF JNLL	3.909	3.164	3.094	1.974	1.984	1.995
Fold-based SE	0.425	0.535	0.497	0.158	0.153	0.153
OOF Spearman	0.727	0.722	0.707	0.712	0.667	0.671
Occurrence AUC	0.734	0.859	0.789	0.805	0.773	0.750
Magnitude RMSE	1.519	1.514	1.476	1.453	1.463	1.471
Successful folds	4/4	4/4	4/4	4/4	4/4	4/4
Within one SE	No	No	No	Yes	Yes	Yes

Note. OOF denotes out-of-fold, JNLL denotes joint negative log-likelihood, SE denotes standard error, AUC denotes area under the receiver operating characteristic curve, and RMSE denotes root mean square error. Bold values identify the minimum mean OOF JNLL, the minimum magnitude RMSE, and the minimum fold-based SE. The principal model uses $\kappa = 50$, selected according to the minimum mean OOF JNLL.

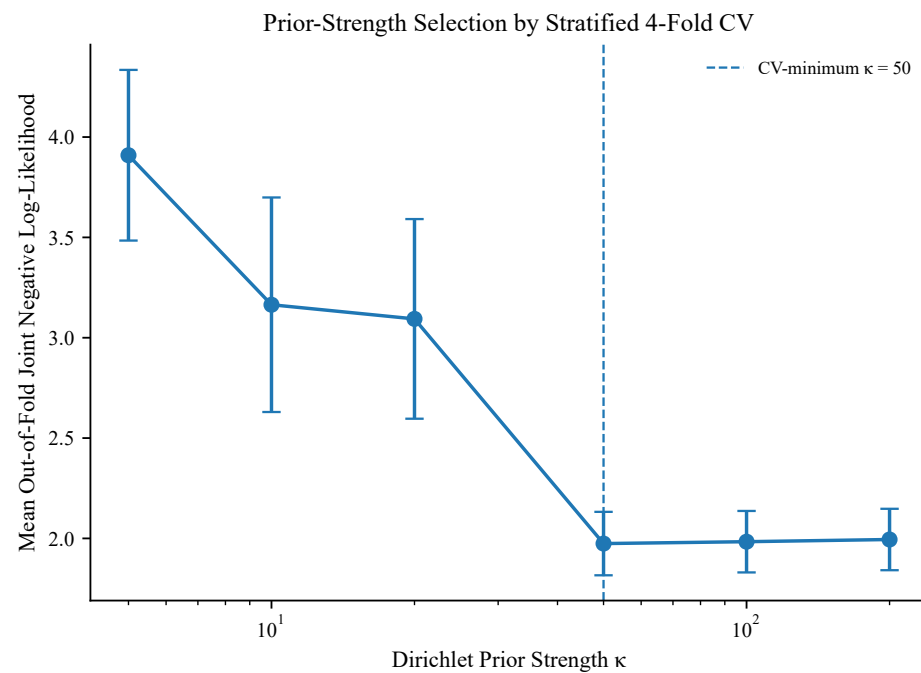


Figure 3. Cross-validated joint loss across candidate prior strengths.

Numerical convergence is assessed using 22 starting points, including the AHP prior, uniform weights, ten AHP-neighbor initializations, and ten diffuse random initializations. Figure 4 presents representative objective trajectories.

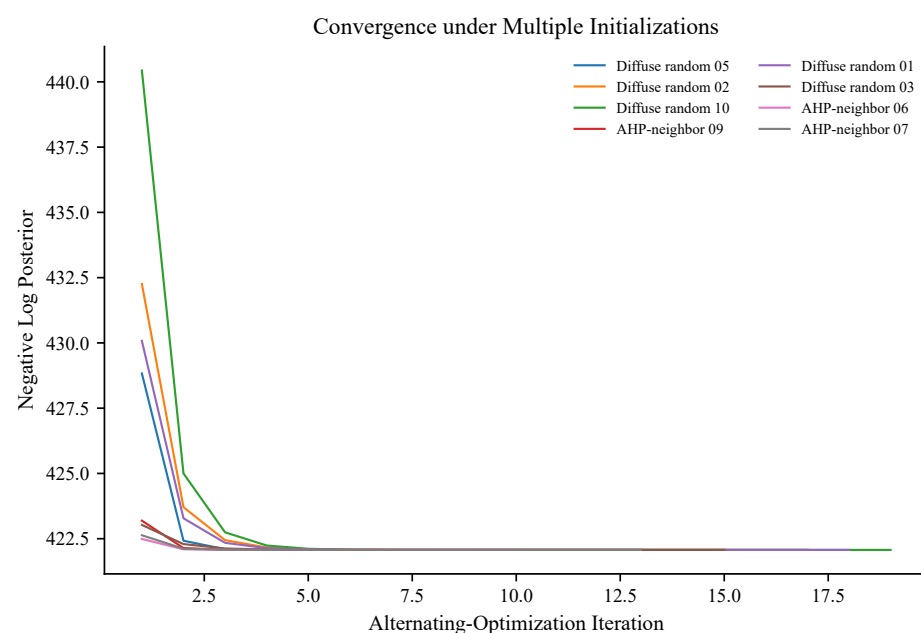


Figure 4. Negative log-posterior convergence across representative initializations.

The 21 runs satisfying the optimizer success criterion converge within 10–19 alternating iterations. Their final solutions are nearly identical. The largest objective difference from the selected solution is 5.63×10^{-6} , the largest global-weight L_1 difference is 2.58×10^{-5} , and the largest single-weight difference is 5.89×10^{-6} . Every successful terminal solution produces a ranking correlation of 1.000 and a Top-10 overlap of 1.000 with the selected result. Table 7 summarizes the convergence diagnostics across the multiple initializations.

Table 7. Convergence diagnostics across multiple initializations.

Measure	Result
Total starting points	22
Runs satisfying optimizer success criterion	21
Alternating-iteration range	10–19
Maximum objective difference from selected solution	5.63×10^{-6}
Maximum global-weight L_1 difference	2.58×10^{-5}
Maximum absolute single-weight difference	5.89×10^{-6}
Minimum rank correlation with selected solution	1.000
Minimum Top-10 overlap with selected solution	1.000

The close agreement among successful runs demonstrates strong numerical consistency across different initializations. The nearly identical posterior objectives, calibrated weights, institutional rankings, and Top-10 sets confirm that the reported solution is highly stable with respect to the initialization strategy under the examined optimization settings.

5.4. Out-of-Fold Model Comparison

The three models are evaluated using a common stratified four-fold out-of-fold protocol. After selecting $\kappa = 50$, the corresponding AHP-MAP out-of-fold predictions are used for model comparison. The same fold structure is retained throughout prior-strength selection and comparative evaluation, ensuring consistent data partitions and directly comparable performance estimates across the three models. This design provides a unified post-selection out-of-fold assessment under a common validation framework.

Table 8 reports the results. Relative to the fixed AHP baseline, AHP-MAP increases the Pearson coefficient from 0.579 to 0.661, the Spearman coefficient from 0.655 to 0.712, and the Kendall coefficient from 0.485 to 0.552. The occurrence AUC increases from 0.742 to 0.805, the Brier score decreases from 0.089 to 0.072, and the occurrence log loss decreases from 0.332 to 0.292.

For positive-output magnitude, AHP-MAP decreases RMSE from 1.481 to 1.453 and MAE from 1.138 to 1.089, while increasing R^2 from 0.246 to 0.275. The mean joint NLL decreases from 2.006 to 1.974.

The purely data-driven model achieves the highest Pearson coefficient and occurrence accuracy. AHP-MAP achieves the highest Spearman and Kendall coefficients, the highest occurrence AUC, the lowest Brier score, the lowest occurrence log loss, the strongest magnitude fitting, and the lowest joint NLL among the three models.

Although the purely data-driven model attains higher threshold-based accuracy, AHP-MAP achieves a lower occurrence log loss and a substantially lower joint NLL. These results demonstrate that AHP-MAP provides better-calibrated occurrence probabilities and stronger joint probabilistic performance in the current sample.

AHP-MAP maintains a rank correlation of 0.962 with the fixed AHP result and an effective weight number of 14.137. The purely data-driven model has a rank correlation of 0.816 and an effective weight number of 7.923. These results indicate that AHP-MAP improves empirical score–output consistency while retaining broader indicator participation and closer agreement with the original hierarchical evaluation structure.

Table 8. Post-selection out-of-fold performance and structural preservation.

Evaluation Aspect	Metric	Fixed AHP	AHP-MAP	Pure Data-Driven
Score association	Pearson coefficient	0.579	0.661	0.694
Score association	Spearman coefficient	0.655	0.712	0.636
Score association	Kendall coefficient	0.485	0.552	0.501
Occurrence performance	AUC	0.742	0.805	0.668
Occurrence performance	Accuracy	0.861	0.889	0.917
Occurrence performance	Balanced accuracy	0.484	0.609	0.734
Occurrence performance	Brier score	0.089	0.072	0.090
Occurrence performance	Log loss	0.332	0.292	1.943
Magnitude performance	RMSE	1.481	1.453	1.588
Magnitude performance	MAE	1.138	1.089	1.165
Magnitude performance	R^2	0.246	0.275	0.134
Joint performance	Mean JNLL	2.006	1.974	4.913
Structural preservation	Rank correlation	1.000	0.962	0.816
Structural preservation	Top-10 overlap	1.000	0.700	0.500
Structural preservation	Effective number of weights	13.531	14.137	7.923
Structural preservation	Global-weight L_1 drift	0.000	0.258	1.125
Structural preservation	Primary-weight L_1 drift	0.000	0.208	0.653

Note. AUC denotes area under the receiver operating characteristic curve, RMSE denotes root mean square error, MAE denotes mean absolute error, and JNLL denotes joint negative log-likelihood. Lower values indicate better performance for the Brier score, log loss, RMSE, MAE, JNLL, and weight-drift measures. Structural-preservation measures are evaluated relative to the fixed AHP result.

The model-comparison results support H1 through improved score–output association and occurrence–magnitude performance relative to fixed AHP. They also support H2 because AHP-MAP produces substantially lower weight drift, higher effective weight participation, and closer ranking agreement with fixed AHP than the purely data-driven specification.

5.5. Weight Calibration Results

The calibrated primary weights are 0.0660 for talent reserve, 0.2641 for operation and development, 0.2236 for R&D activities, and 0.4463 for technology transfer. Operation and development receives the largest positive adjustment, while technology transfer remains the dimension with the highest overall weight.

Figure 5 presents the principal directions of global-weight calibration. The horizontal axis reports the difference $\Delta w_j = w_j^{\text{MAP}} - w_j^{\text{AHP}}$.

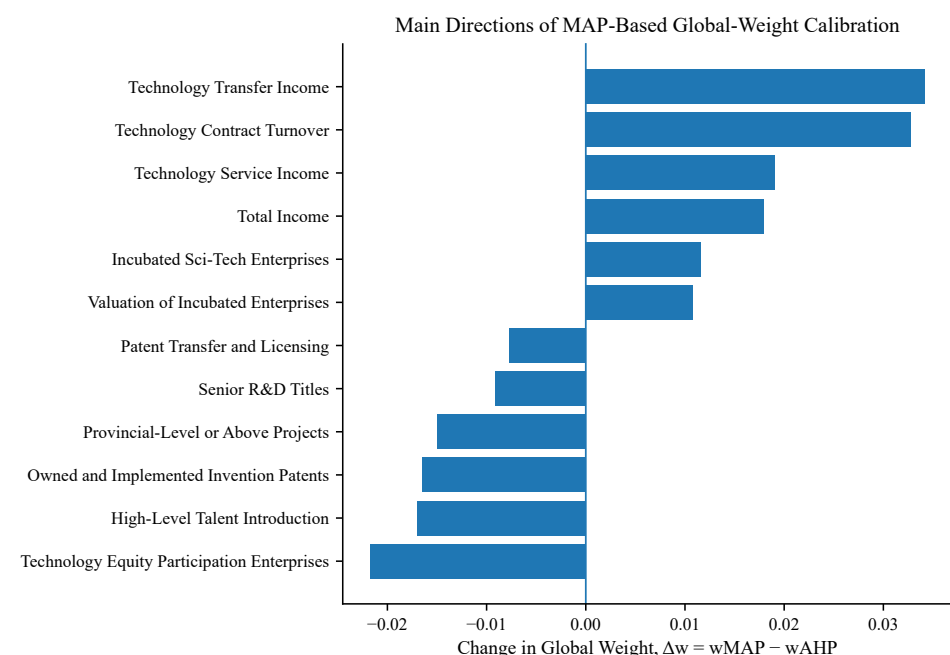
**Figure 5.** Indicator-level global-weight adjustments relative to fixed AHP.

Table 9 reports the principal indicator-level changes. Technology transfer income, technology contract turnover, technology service income, and total income receive the largest positive adjustments. Incubated sci-tech enterprises and valuation of incubated enterprises also receive positive adjustments. The remaining indicators continue to contribute to the hierarchical score, with their calibrated weights reflecting both the original AHP structure and the observed institutional evidence.

Bootstrap direction-stability values exceed 0.98 for the four principal operation-and-development indicators, high-level talent introduction, and provincial-level or above projects. The corresponding calibration directions are reproduced in most bootstrap samples.

Table 9. Principal indicator-level global-weight adjustments and bootstrap diagnostics.

Dimension	Indicator	AHP Weight	MAP Weight	Δw_j	Bootstrap 2.5%	Bootstrap 97.5%	Direction Stability
Operation and Development	Technology Transfer Income	0.0534	0.0875	0.0342	0.0588	0.1418	0.998
Operation and Development	Technology Contract Turnover	0.0534	0.0862	0.0328	0.0579	0.1356	0.988
Operation and Development	Technology Service Income	0.0267	0.0458	0.0191	0.0298	0.0796	0.994
Operation and Development	Total Income	0.0267	0.0446	0.0179	0.0302	0.0845	0.998
Technology Transfer	Technology Equity Participation Enterprises	0.1669	0.1452	−0.0217	0.1039	0.1717	0.922
Talent Reserve	High-Level Talent Introduction	0.0533	0.0363	−0.0170	0.0232	0.0508	0.986
Technology Transfer	Owned and Implemented Invention Patents	0.1041	0.0876	−0.0165	0.0602	0.1123	0.882
R&D Activities	Provincial-Level or Above Projects	0.0612	0.0462	−0.0150	0.0334	0.0565	0.998
Technology Transfer	Incubated Sci-Tech Enterprises	0.0443	0.0559	0.0116	0.0324	0.0808	0.714
Technology Transfer	Valuation of Incubated Enterprises	0.0286	0.0394	0.0108	0.0220	0.0581	0.876

Note. $\Delta w_j = w_j^{\text{MAP}} - w_j^{\text{AHP}}$. Direction stability denotes the proportion of successful bootstrap replications in which the direction of the weight adjustment agrees with the complete-sample AHP-MAP result.

The observed weight changes describe how the current-period output evidence modifies the hierarchical evaluation weights. They do not establish causal effects of individual indicators on technology-transfer performance.

5.6. Institutional Scores and Rankings

Table 10 reports the ten highest-ranked institutions under AHP-MAP. Inst. 04 obtains the highest calibrated score of 0.5261, followed by Inst. 22, Inst. 21, Inst. 06, and Inst. 20. The first seven institutions under AHP-MAP are also included among the leading institutions under fixed AHP, indicating continuity between the expert-derived and calibrated evaluation results.

Inst. 16 moves from 11th to 8th, Inst. 12 moves from 14th to 9th, and Inst. 18 moves from 13th to 10th. These adjustments reflect differences introduced after incorporating occurrence–magnitude evidence into the hierarchical weights.

Table 10. Top 10 institutional rankings under AHP-MAP and comparison with fixed AHP.

Institution	MAP Score	MAP Rank	Monetized Output	Log-Transformed Output	AHP Score	AHP Rank	Rank Change
Inst. 04	0.5261	1	28,225.07	10.2480	0.5201	2	−1
Inst. 22	0.5166	2	24,648.00	10.1125	0.5277	1	1
Inst. 21	0.4928	3	31,081.15	10.3444	0.4545	4	−1
Inst. 06	0.4846	4	64,422.73	11.0732	0.4485	5	−1
Inst. 20	0.4704	5	121,529.00	11.7079	0.4663	3	2
Inst. 08	0.4627	6	13,035.81	9.4755	0.4466	6	0
Inst. 14	0.4105	7	21,681.00	9.9842	0.3874	7	0
Inst. 16	0.3849	8	16,545.25	9.7139	0.3407	11	−3
Inst. 12	0.3785	9	201,984.67	12.2160	0.3310	14	−5
Inst. 18	0.3753	10	17,821.17	9.7882	0.3338	13	−3

Note. Log-transformed output denotes $\log(1 + \text{Output})$. Rank change is calculated as the AHP-MAP rank minus the fixed-AHP rank. A negative value therefore indicates an upward movement under AHP-MAP.

The bootstrap intervals in Figure 6 provide complementary stability information for all 36 institutional rankings.

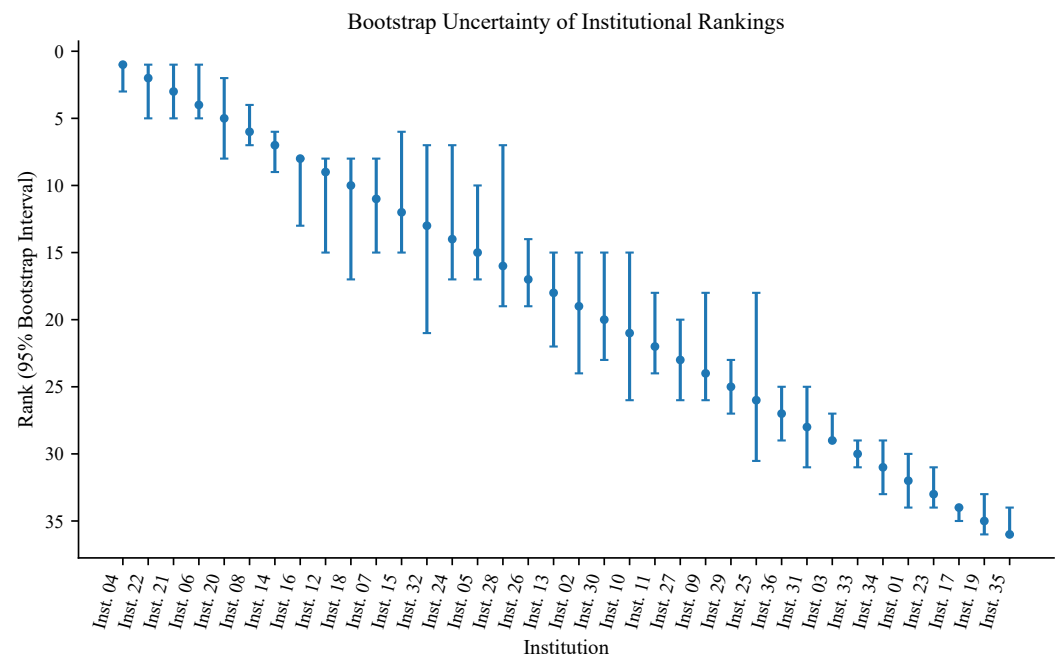


Figure 6. Bootstrap rank intervals for the 36 institutions.

5.7. Bootstrap, Jackknife, and Contribution Stability

Figure 6 presents the 95% bootstrap intervals of the institutional rankings. The leading institutions display relatively concentrated ranking intervals and high Top-10 inclusion probabilities. Inst. 04, Inst. 06, Inst. 20, Inst. 21, and Inst. 22 are included in the Top 10 in every bootstrap replication.

All 500 bootstrap replications complete successfully. The mean rank correlation with the base AHP-MAP ranking is 0.982, and the mean Top-10 overlap is 0.855. The mean rank IQR is 2.083, and the mean rank standard deviation is 1.612. The mean weight-order correlation is 0.968, while the mean overlap among the five highest-weight indicators is 0.955.

All 36 leave-one-institution-out jackknife runs also complete successfully. The minimum rank correlation with the complete-sample ranking is 0.993. The maximum global-weight L_1 difference is 0.137, the maximum single-weight difference is 0.0226, and the largest institutional rank movement is five positions.

The hierarchical contribution decomposition reconstructs every institutional score exactly. Both the maximum and mean absolute differences between the final score and the sum of its indicator contributions are zero. Table 11 consolidates the bootstrap, jackknife, and additive-decomposition diagnostics.

These results support H3 by demonstrating strong stability of the calibrated weights and institutional rankings under bootstrap resampling and single-institution omission. The high rank correlations and Top-10 overlap values confirm the consistency of the institutional ordering, while the weight-order correlation and Top-5 indicator overlap demonstrate the stability of the calibrated importance structure. The zero reconstruction differences further verify the exact arithmetic fidelity and full traceability of the additive contribution decomposition.

Table 11. Bootstrap, jackknife, and additive-decomposition diagnostics.

Analysis	Measure	Value
Bootstrap	Requested replications	500
Bootstrap	Successful replications	500
Bootstrap	Mean rank correlation with base	0.982
Bootstrap	Mean Top-10 overlap	0.855
Bootstrap	Mean rank IQR	2.083
Bootstrap	Mean rank standard deviation	1.612
Bootstrap	Mean weight-order correlation	0.968
Bootstrap	Mean Top-5 indicator overlap	0.955
Jackknife	Successful leave-one-institution-out runs	36/36
Jackknife	Maximum global-weight L_1 difference	0.137
Jackknife	Maximum single-weight difference	0.0226
Jackknife	Minimum rank correlation with base	0.993
Jackknife	Maximum institutional rank movement	5
Contribution decomposition	Maximum absolute reconstruction difference	0.000
Contribution decomposition	Mean absolute reconstruction difference	0.000

5.8. Monte Carlo Sample-Size Analysis

The Monte Carlo analysis evaluates sample sizes of $N = 24, 36, 72$, and 144 . Each model and sample-size combination is repeated 200 times using random seed 42. Fixed AHP retains the original expert-derived weights, AHP-MAP uses $\kappa = 50$, and the purely data-driven model uses $\kappa = 0$.

All estimated models use the same occurrence–magnitude likelihood and hierarchical simplex constraints. The maximum number of alternating iterations is 30, the SLSQP function tolerance is 10^{-9} , the maximum number of inner SLSQP iterations is 120, and the lower bound for the magnitude standard deviation is $\sigma_y = 0.25$. Weight recovery is measured using the global-weight L_1 difference from the simulation reference weights. Ranking recovery is measured using the Spearman correlation and Top-10 overlap with the simulation reference ranking.

Figure 7 presents the recovery results. AHP-MAP achieves the smallest mean global-weight L_1 difference at every examined sample size. Its mean difference decreases from 0.219 at $N = 24$ to 0.210 at $N = 36$, 0.179 at $N = 72$, and 0.152 at $N = 144$. The mean Spearman correlation with the simulation reference ranking increases from 0.967 to 0.971, 0.978, and 0.984 across the corresponding sample sizes. The mean Top-10 overlap increases from 0.764 at $N = 24$ to 0.838 at $N = 144$.

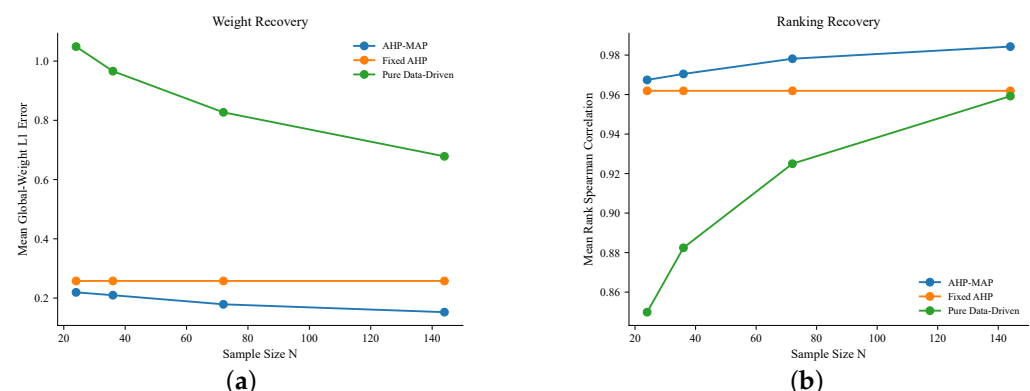


Figure 7. Monte Carlo weight and ranking recovery across sample sizes. (a) Mean global-weight L_1 difference across sample sizes; (b) mean Spearman rank correlation and Top-10 overlap across sample sizes.

Table 12 summarizes the comparative recovery results. AHP-MAP combines the smallest mean weight differences with the highest mean ranking correlations across the four prespecified sample sizes.

Table 12. Monte Carlo recovery metrics across sample sizes.

Sample Size	Model	Mean Weight L_1 Difference	Mean Rank Spearman	Mean Top-10 Overlap
24	AHP-MAP	0.219	0.967	0.764
24	Fixed AHP	0.258	0.962	0.700
24	Pure Data-Driven	1.049	0.850	0.712
36	AHP-MAP	0.210	0.971	0.774
36	Fixed AHP	0.258	0.962	0.700
36	Pure Data-Driven	0.966	0.882	0.749
72	AHP-MAP	0.179	0.978	0.811
72	Fixed AHP	0.258	0.962	0.700
72	Pure Data-Driven	0.827	0.925	0.779
144	AHP-MAP	0.152	0.984	0.838
144	Fixed AHP	0.258	0.962	0.700
144	Pure Data-Driven	0.679	0.959	0.826

The results indicate improved weight and ranking recovery for AHP-MAP as the simulated sample size increases under the examined data-generation and estimation settings. They should not be interpreted as a general sample-complexity guarantee outside these prespecified simulation conditions.

5.9. Institution-Type Analysis

The 36 institutions comprise 29 enterprise-led institutions, four government-led institutions, and three university-led institutions. Table 13 reports descriptive results for the three organizational forms.

The mean AHP-MAP score is 0.342 for enterprise-led institutions, 0.174 for government-led institutions, and 0.244 for university-led institutions. The within-type Spearman correlation between AHP-MAP scores and monetized outputs is 0.843 for enterprise-led institutions, 0.800 for government-led institutions, and 0.500 for university-led institutions.

Leave-one-type-out estimation produces high agreement with the complete-sample ranking. Excluding government-led institutions yields a rank correlation of 0.998, while excluding university-led institutions yields a rank correlation of 0.993. The common cross-validated prior strength $\kappa = 50$ is retained across the institutional types to maintain a consistent evaluation basis under the observed subgroup sizes.

Table 13. Institution-type profiles and within-type score–output associations.

Institution Type	N	Mean Monetized Output	Zero-Output Ratio	Mean AHP Score	Mean MAP Score	Within-Type Spearman
Enterprise-led	29	23447.59	0.103	0.327	0.342	0.843
Government-led	4	1222.76	0.250	0.170	0.174	0.800
University-led	3	54524.08	0.000	0.227	0.244	0.500

The subgroup coefficients provide descriptive evidence of positive score–output associations within each observed institutional type. The government-led and university-led groups contain four and three institutions, respectively, so the results are most appropriately used to characterize the observed subgroup patterns. These findings provide an initial empirical basis for future cross-type validation using larger and more balanced institutional samples.

5.10. Summary of Empirical Findings

The empirical results provide convergent support for H1–H3. The log-based transformation offers the strongest combined fit among the examined transformations, and stratified four-fold selection identifies $\kappa = 50$ within a stable moderate-to-strong prior region. Relative to fixed AHP, AHP-MAP improves score–output association, occurrence discrimination, positive-output magnitude fitting, and joint probabilistic performance. Relative to the same model with $\kappa = 0$, it produces lower weight drift, broader indicator participation, and closer ranking agreement with the original AHP system. Multiple-initialization, bootstrap, jackknife, exact contribution reconstruction, and prespecified Monte Carlo analyses provide complementary evidence on numerical, resampling, and sample-size stability. The institutional-type results remain descriptive because the government-led and university-led subgroups are small.

6. Discussion

6.1. Theoretical Implications

The first theoretical implication concerns the relationship between expert knowledge and empirical evidence in hierarchical performance evaluation. The AHP-MAP framework does not replace the expert-designed indicator system with an unrestricted data-driven model. Instead, the primary and local AHP weights define the modes of Dirichlet priors under positive prior strength, while observed institutional outcomes provide evidence for calibrated adjustment under the same hierarchical simplex constraints. The cross-validated prior strength makes the balance between structural continuity and empirical adaptation an explicit and data-selectable component of the evaluation model.

The second implication arises from the occurrence–magnitude representation of monetized technology-transfer output. The realization of positive output and the scale achieved after realization correspond to related but distinct institutional processes. Separating these two components allows the evaluation model to reflect the coexistence of zero observations and strongly skewed positive outputs without reducing technology-transfer performance to a single linear relationship. This structure also provides a clearer analytical connection between institutional scores and different stages of market realization.

The third implication concerns transparent interpretation. The additive hierarchical score enables exact reconstruction through primary-dimension and secondary-indicator contributions. This structural transparency allows each institutional score to be traced directly to its underlying dimensions and indicators. The contribution values reveal how the calibrated evaluation system forms each score, while weight drift quantifies how observed evidence adjusts the original hierarchy. Together, these measures provide an auditable representation of evaluation composition and support clear comparison of the relative roles of dimensions and indicators in institutional performance assessment.

The framework further connects empirical calibration with uncertainty assessment. Multiple-initialization comparison examines numerical dependence on starting values, transformation diagnostics examine sensitivity to data preparation, and bootstrap and jackknife procedures examine sampling variation and institutional influence. Monte Carlo analysis provides an additional assessment of weight and ranking recovery under the

prespecified simulation settings. Together, these components extend hierarchical evaluation from a single deterministic ranking toward an evidence-calibrated process with explicit stability information.

6.2. Managerial and Policy Implications

Evaluation agencies can retain the established AHP hierarchy while using AHP-MAP to update the relative weights with institutional evidence. The fixed AHP score and calibrated score should be reported together with global-weight changes, contribution decompositions, and ranking-stability information. This parallel reporting structure preserves policy continuity and allows evaluators to identify which results arise from the original expert system and which arise from evidence-based calibration.

For managers of new R&D institutions, the increased weights of technology-transfer income, technology contract turnover, technology service income, and total income indicate that commercialization pathways, contract formation, and market-oriented service capacity are closely associated with monetized output in the current sample. Dimension-level and indicator-level contributions can help determine whether an institution's score is mainly supported by talent reserve, operation and development, R&D activities, or direct technology-transfer realization. Institutions with strong research resources but limited monetized output can use the contribution profiles to identify gaps in contract execution, technology services, incubation, and market connection.

Talent cultivation, foundational research, high-level projects, and knowledge accumulation remain important components of long-term institutional capacity. Their calibrated weight changes only reflect their relationship with monetized output during the current reporting period. Evaluation practice should therefore retain these dimensions in the hierarchical system and examine substantial weight changes through expert review. A dual evaluation structure that reports current commercialization performance together with long-term capability development can reduce the risk that short-term monetary indicators dominate strategic resource allocation.

Ranking results should also be interpreted together with stability information. Bootstrap intervals, rank standard deviations, and Top-10 inclusion probabilities distinguish stable leading institutions from institutions whose positions depend on small score differences. Resource allocation and performance classification can therefore use score ranges and contribution profiles together with point rankings rather than treating every rank difference as substantively meaningful.

6.3. Implementation Protocol for Evaluation Agencies

The AHP-MAP procedure can be implemented through seven sequential steps.

1. Fix the evaluation period, indicator definitions, statistical units, hierarchical structure, and expert judgment matrices.
2. Apply the selected indicator transformation and calculate the fixed AHP weights, institutional scores, and rankings.
3. Divide the institutions according to output occurrence and use stratified cross-validation to compare the prespecified candidate values of κ .
4. Estimate the final AHP-MAP model using the selected κ and multiple initializations, retaining the successful solution with the largest posterior objective.
5. Conduct bootstrap and jackknife analyses to obtain weight intervals, ranking intervals, and institutional influence measures.
6. Report the fixed AHP and AHP-MAP weights, scores, rankings, contribution decompositions, and stability results in parallel.

7. Recalibrate the model when a new evaluation cycle provides an updated institutional dataset under consistent indicator definitions and statistical standards.

6.4. Methodological Boundaries

The empirical analysis contains 36 institutions and 27 nominal free parameters. Dirichlet shrinkage, cross-validation, multiple initializations, bootstrap analysis, jackknife analysis, and Monte Carlo experiments provide complementary checks on numerical and sampling stability. These procedures reduce the risk of unsupported weight adjustment but do not substitute for external validation using larger and independently collected institutional samples.

The current dataset is cross-sectional. R&D investment, talent recruitment, patent formation, contract execution, enterprise incubation, and monetized realization may occur in different periods. The estimated relationships therefore describe contemporaneous institutional associations and do not establish time-lagged or causal effects. Panel data covering multiple evaluation cycles are required to examine the temporal pathways from institutional inputs and capabilities to subsequent technology-transfer outcomes.

The sample is drawn from new R&D institutions in Gansu Province, and enterprise-led institutions account for most observations. The institutional-type results consequently provide descriptive subgroup evidence rather than precise estimates of differences among organizational forms. Cross-regional samples with more balanced institutional types are needed before the calibrated relationships can be generalized to broader populations. The current out-of-fold comparison also uses the same four-fold structure for prior-strength selection and model assessment. Larger samples can support nested cross-validation with separate parameter-selection and performance-evaluation layers.

Monetized technology-transfer output captures an important form of market realization but does not represent the complete public value of new R&D institutions. Knowledge spillovers, talent development, public services, industrial coordination, and long-term social impact may not be reflected fully in monetary output during a single reporting period. Future evaluation systems should therefore combine monetized output with complementary indicators of contracts, licensing, enterprise collaboration, knowledge diffusion, and public value.

7. Conclusions and Future Directions

This study developed an AHP-MAP framework for technology-transfer performance evaluation in new R&D institutions. The framework combines an expert-designed AHP hierarchy with Dirichlet prior regularization, an occurrence–magnitude output mechanism, cross-validated prior-strength selection, and constrained alternating optimization. The resulting institutional scores preserve the original hierarchical semantics and allow exact decomposition into primary-dimension and secondary-indicator contributions.

Within the current sample and evaluation setting, AHP-MAP strengthens the empirical relationship between composite scores and monetized technology-transfer outputs relative to fixed AHP. Compared with the same model under $\kappa = 0$, it preserves closer alignment with the original AHP structure, produces lower weight drift, and maintains broader participation across the indicator set. The multiple-initialization, transformation, bootstrap, jackknife, and prespecified Monte Carlo analyses jointly demonstrate the numerical consistency and sampling stability of the calibrated weights and institutional rankings. These findings support AHP-MAP as a transparent evidence-calibration framework that complements expert evaluation and provides an auditable basis for institutional assessment.

Future research can extend the framework through institution–year panel data that capture the temporal pathways among R&D investment, talent accumulation, patent for-

mation, contract execution, incubation, and transfer realization. The calibration outcomes can also be enriched by incorporating contract activity, licensing, enterprise collaboration, knowledge spillovers, and public-value indicators. Full Bayesian estimation can further provide posterior intervals for weights, scores, and rankings. Larger and more diverse datasets, together with independent or nested validation designs, can strengthen external validity and support broader cross-regional and policy-oriented applications.

Author Contributions: Conceptualization, T.L.; methodology, T.L. and H.B.; software, T.L., Z.Z. (Zhiwei Zhao) and J.Z.; validation, Z.Z. (Zhiwei Zhao) and J.Z.; formal analysis, Z.Z. (Zhiwei Zhao); investigation, T.L., Z.Z. (Ziqiang Zhang) and Z.M.; resources, Z.Z. (Ziqiang Zhang); data curation, H.B.; writing—original draft preparation, T.L.; writing—review and editing, Z.Z. (Zhiwei Zhao) and H.B.; visualization, H.B.; supervision, Z.M. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Key R&D Foundation of Gansu Province under Grant No. 23YFGA0079; the “FeiTian Scholar” Supporting Program of Gansu Province under Grant No. 2024FTQN-020; the Natural Science Foundation of Gansu Province under Grant No. 25JRRA093; and the Youth Doctoral Support Program of Gansu Universities under Grant No. 2025QB-014.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data supporting the findings of this study are available from the corresponding author upon reasonable request.

Acknowledgments: The authors would like to thank Sile Hu for her valuable support in providing and organizing the data used in this study. The authors also thank the editors and reviewers for their valuable comments and suggestions.

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

Abbreviations

The following abbreviations are used in this manuscript.

AHP	Analytic hierarchy process
MAP	Maximum a posteriori estimation
AHP-MAP	Analytic hierarchy process–maximum a posteriori framework
XAI	Explainable artificial intelligence
MCDA	Multi-criteria decision analysis
R&D	Research and development
AUC	Area under the receiver operating characteristic curve
RMSE	Root mean square error
MAE	Mean absolute error
JNLL	Joint negative log-likelihood
OOF	Out-of-fold
IQR	Interquartile range
SE	Standard error
KL	Kullback–Leibler
CI	Consistency index
RI	Random consistency index
CR	Consistency ratio
SLSQP	Sequential least squares programming

References

1. Cunningham, J.A.; Menter, M.; Starke, F. The evolution of university technology transfer research: A text mining approach. *J. Technol. Transf.* **2025**, *50*, 1231–1268. <https://doi.org/10.1007/s10961-024-10133-2>.

2. Zarea, F.; Douglas, E.J.; Obschonka, M.; Davidsson, P.; Audretsch, D.B.; Hutmacher, D.W. When the marketplace comes to the research laboratory: Technology transfer efficiency of innovation-driven publicly funded research centers. *J. Technol. Transf.* **2025**, *50*, 2556–2586. <https://doi.org/10.1007/s10961-025-10188-9>.
3. Fernández-Alles, M.; Rodríguez-González, T.; Villanueva-Flores, M.; Díaz-Fernández, M. Impact of the ambidexterity of University Technology Transfer Offices on their performance: The Spanish context. *Technol. Forecast. Soc. Chang.* **2025**, *218*, 124224. <https://doi.org/10.1016/j.techfore.2025.124224>.
4. Sa'a, E.A.; Asplund, F. Unpacking social capital in University–Industry Collaborations: Pathways to cross-industry knowledge sharing. *Technovation* **2025**, *140*, 103160. <https://doi.org/10.1016/j.technovation.2024.103160>.
5. Lie, H.T.; Egelie, K.J.; Grimpe, C.; Sørheim, R. The fine print of collaboration: How contractual provisions govern IP and disclosure in publicly funded research. *Res. Policy* **2025**, *54*, 105336. <https://doi.org/10.1016/j.respol.2025.105336>.
6. Giebel, M.; Rösner, A. Reaching for the society: The commercialization effects of NASA technology licensing. *Res. Policy* **2025**, *54*, 105337. <https://doi.org/10.1016/j.respol.2025.105337>.
7. Greco, S.; Słowiński, R.; Wallenius, J. Fifty years of multiple criteria decision analysis: From classical methods to robust ordinal regression. *Eur. J. Oper. Res.* **2025**, *323*, 351–377. <https://doi.org/10.1016/j.ejor.2024.07.038>.
8. Shi, C.; Yao, Y. Explainable multi-criteria decision-making: A three-way decision perspective. *Int. J. Approx. Reason.* **2025**, *187*, 109528. <https://doi.org/10.1016/j.ijar.2025.109528>.
9. Nugent, A.; Chan, H.F. Outsourcing university research commercialization to a sophisticated technology transfer office: Evidence from Australian universities. *Technovation* **2023**, *125*, 102762. <https://doi.org/10.1016/j.technovation.2023.102762>.
10. Cohen, M.; Fernandes, G.; Godinho, P. Measuring the impacts of university-industry R&D collaborations: A systematic literature review. *J. Technol. Transf.* **2025**, *50*, 345–374. <https://doi.org/10.1007/s10961-024-10114-5>.
11. Abi Saad, E.; Agogué, M. Living Labs in science-industry collaborations: Roles, design, and application patterns. *Technovation* **2024**, *135*, 103066. <https://doi.org/10.1016/j.technovation.2024.103066>.
12. Chen, Z.; Little, V.J.; Thuan, N.H. The evolving role of technology transfer offices in the entrepreneurial university: Go-betweens or playmakers? *J. Technol. Transf.* **2025**, *50*, 1060–1079. <https://doi.org/10.1007/s10961-024-10123-4>.
13. Bailey, A.G.; Reingold, B.M.; Johnson, J.D.; O'Connor, A.C. Paths towards commercialization: Evidence from NIH proof of concept centers. *J. Technol. Transf.* **2025**, *50*, 2509–2531. <https://doi.org/10.1007/s10961-025-10187-w>.
14. Tolín, G.; Piccaluga, A. Commercializing technology from university-industry collaborations: A configurational perspective on organizational factors. *J. Bus. Res.* **2025**, *188*, 115105. <https://doi.org/10.1016/j.jbusres.2024.115105>.
15. Miranda, F.J.; Chamorro-Mera, A.; Rubio, S. The commercialization of academic patents: A critical literature review and a research agenda. *Technol. Innov.* **2025**, *24*, 149–160. <https://doi.org/10.1080/19498241.2024.2420705>.
16. Bottai, C.; de Rassenfosse, G.; Raiteri, E. A new approach to measuring invention commercialization: An application to the SBIR program. *Res. Policy* **2025**, *54*, 105302. <https://doi.org/10.1016/j.respol.2025.105302>.
17. Castaldi, C.; Giuliani, E.; Kyle, M.; Nuvolari, A. Are intellectual property rights working for society? *Res. Policy* **2024**, *53*, 104936. <https://doi.org/10.1016/j.respol.2023.104936>.
18. Ascione, G.S.; Ciucci, L.; Detotto, C.; Sterzi, V. University patent litigation in the United States: Do we have a problem? *Res. Policy* **2024**, *53*, 104909. <https://doi.org/10.1016/j.respol.2023.104909>.
19. Belgin, Ö.; Apaydın Avşar, B. Performance of publicly supported technology transfer offices in Türkiye: A DEA–MPI–GM(1,N) approach. *J. Technol. Transf.* **2026**. <https://doi.org/10.1007/s10961-026-10335-w>.
20. Giunta, A.; Pierucci, E.; Vurchio, D. University–industry interaction and firms' innovation performance: new evidence from the biopharmaceutical industry. *J. Technol. Transf.* **2025**. <https://doi.org/10.1007/s10961-025-10275-x>.
21. Kizielewicz, B.; Tomczyk, T.; Gandor, M.; Sałabun, W. Subjective weight determination methods in multi-criteria decision-making: A systematic review. *Procedia Comput. Sci.* **2024**, *246*, 5396–5407. <https://doi.org/10.1016/j.procs.2024.09.673>.
22. Barbati, M.; Greco, S.; Lami, I.M. The Deck-of-cards-based Ordinal Regression method and its application for the development of an ecovillage. *Eur. J. Oper. Res.* **2024**, *319*, 845–861. <https://doi.org/10.1016/j.ejor.2024.07.010>.
23. Belahcène, K.; Mousseau, V.; Ouerdane, W.; Pirlot, M.; Sobrie, O. A guided tour of multiple criteria sorting models and methods. *Ann. Oper. Res.* **2024**, *343*, 785–845. <https://doi.org/10.1007/s10479-024-06278-w>.
24. Gilbert, H.; Ouaguenouni, M.; Öztürk, M.; Spanjaard, O. Robust ordinal regression for subsets comparisons with interactions. *Eur. J. Oper. Res.* **2025**, *320*, 146–159. <https://doi.org/10.1016/j.ejor.2024.07.021>.
25. Grillo, M.; Kotłowski, W.; Kadziński, M. Ordinal regression meets online learning: Interactive preference learning for multiple criteria choice and ranking with provable guarantees. *Eur. J. Oper. Res.* **2025**, *327*, 1003–1022. <https://doi.org/10.1016/j.ejor.2025.05.045>.
26. Shao, L.; Liu, J.; Fu, C.; Zhu, N.; Chen, H. Alternative ranking in trust network group decision-making: A distributionally robust optimization method. *Eur. J. Oper. Res.* **2025**, *327*, 986–1002. <https://doi.org/10.1016/j.ejor.2025.05.052>.
27. Cebesoy, M.; Tuncer Şakar, C.; Yet, B. Multicriteria decision support under uncertainty: Combining outranking methods with Bayesian networks. *Ann. Oper. Res.* **2025**, *355*, 2971–2998. <https://doi.org/10.1007/s10479-024-06064-8>.

28. Gómez-Jiménez, A.; Jiménez-Martín, A.; Chergui, Z.; Marín, S. A hierarchical elicitation process based on the combination of different weighting methods within multi-attribute utility theory in a group decision-making context. *Ann. Oper. Res.* **2026**. <https://doi.org/10.1007/s10479-026-07146-5>.
29. Paradowski, B.; Wątróbski, J.; Sałabun, W. Novel coefficients for improved robustness in multi-criteria decision analysis. *Artif. Intell. Rev.* **2025**, *58*, 298. <https://doi.org/10.1007/s10462-025-11307-6>.
30. Więckowski, J.; Sałabun, W. A new sensitivity analysis method for decision-making with multiple parameters modification. *Inf. Sci.* **2024**, *678*, 120902. <https://doi.org/10.1016/j.ins.2024.120902>.
31. Qiao, Y.; Wang, S. Firms' Structural Positions in Patent Citation Networks and Innovation Performance: Evidence from a Large-Scale Chinese Dataset. *Systems* **2026**, *14*, 351. <https://doi.org/10.3390/systems14040351>.
32. De Bock, K.W.; Coussement, K.; De Caigny, A.; Słowiński, R.; Baesens, B.; Boute, R.N.; Choi, T.M.; Delen, D.; Kraus, M.; Lessmann, S.; et al. Explainable AI for operational research: A defining framework, methods, applications, and a research agenda. *Eur. J. Oper. Res.* **2024**, *317*, 249–272. <https://doi.org/10.1016/j.ejor.2023.09.026>.
33. Longo, L.; Brcic, M.; Cabitza, F.; Choi, J.; Confalonieri, R.; Del Ser, J.; Guidotti, R.; Hayashi, Y.; Herrera, F.; Holzinger, A.; et al. Explainable Artificial Intelligence (XAI) 2.0: A manifesto of open challenges and interdisciplinary research directions. *Inf. Fusion* **2024**, *106*, 102301. <https://doi.org/10.1016/j.inffus.2024.102301>.
34. Coussement, K.; Abedin, M.Z.; Kraus, M.; Maldonado, S.; Topuz, K. Explainable AI for enhanced decision-making. *Decis. Support Syst.* **2024**, *184*, 114276. <https://doi.org/10.1016/j.dss.2024.114276>.
35. Borgonovo, E.; Plischke, E.; Rabitti, G. The many Shapley values for explainable artificial intelligence: A sensitivity analysis perspective. *Eur. J. Oper. Res.* **2024**, *318*, 911–926. <https://doi.org/10.1016/j.ejor.2024.06.023>.
36. Ponce-Bobadilla, A.V.; Schmitt, V.; Maier, C.S.; Mensing, S.; Stodtmann, S. Practical guide to SHAP analysis: Explaining supervised machine learning model predictions in drug development. *Clin. Transl. Sci.* **2024**, *17*, e70056. <https://doi.org/10.1111/cts.70056>.
37. Fan, Z.; Fan, P.; Wang, Y. Analysis of Influencing Factors and Prediction of Provincial Energy Poverty in China Based on Explainable Deep Learning. *Systems* **2026**, *14*, 319. <https://doi.org/10.3390/systems14030319>.

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.