Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Within group policy detail and predictive recovery in regional innovation assessment

Haizhou Mao

mhz@zjitc.edu.cn

Zhejiang Industry and Trade Vocational College

Article

Keywords: policy measurement, technology transfer, incremental validity, compositional data, nested validation, signal recovery

Posted Date: November 27th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11716389/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Detailed policy coding preserves distinctions that broad indicators suppress, but its predictive value depends on whether those distinctions can be learned. This study examines published counts of eleven technology transfer instruments across 31 Chinese provincial regions. A nested comparison retains the same coarse group features and adds within-group logratio coordinates with a separate penalty. Prior outcomes, research capacity, economic scale and recorded tool quantity are held common. Provincial and macroregional holdouts evaluate technology contract amounts and invention grants in 2020. Under the primary four-group partition, adding detail raises contract mean squared error by 48.0% in provincial holdouts and 41.3% in macroregional holdouts. For grants, error rises by 8.3% in provincial holdouts but falls by 11.6% in macroregional holdouts. Sensitivities examine zero replacement, quantity treatment and the source study's three-category partition. Fixed-design simulations inject three known within-group directions after removing their linear coarse-feature associations. With a stipulated conditional signal fraction of 0.25 and normal noise, the seven-coordinate procedure achieves a five-percent validation mean squared error reduction in 30.0–45.8% of repetitions; a matched coarse model supplied with the correct direction succeeds in 76.6–80.4%. The empirical comparisons therefore concern conditional predictive utility, while the oracle gap diagnoses a substantial recovery cost. Neither negative validation results nor positive block-transfer results identify policy effectiveness. The study provides an inspectable case audit that separates retained information, learned information and the prediction task.

MSC 2020: 62J07; 62P20. JEL: C52; C53; O38.

1 Introduction

Science policy assessment often turns policy texts into regional indicators. Coding instruments separately can preserve differences in formal provision, whereas broad groups make comparisons easier to estimate and communicate. Neither choice establishes that the resulting indicators add usable information about regional outcomes. Policy mix research distinguishes instruments, strategies and processes, and emphasizes the difficulty of operational mapping (Flanagan et al., 2011; Rogge & Reichardt, 2016; Meissner & Kergroach, 2021; Howoldt, 2024). The present study addresses a narrower question: when broad features are already retained, what predictive increment is supplied by differences among tools within the same group?

A comparison of unrelated indicators cannot answer this question cleanly. Replacing entropy and aggregate ratios with a full logratio composition changes the functions offered to the estimator as well as their dimension. Applying one penalty to every policy feature also couples the retention of broad information to the estimation of additional detail. A detailed model can then perform poorly because it abandons a useful coarse relationship or because its shared shrinkage is poorly suited to learning additional directions. Its failure would be difficult to interpret as evidence about information lost through aggregation.

We therefore keep the coarse features unchanged and append a distinct within-group block. The coarse-only model is an exact boundary of the augmented model: setting the fine-block penalty to infinity returns every candidate coarse model. Training folds choose the two penalties, and outer folds assess the resulting procedure. This defines an identifiable comparison of two nested linear function classes. It does not compare all possible functions of the data or identify a causal effect of coding resolution.

The application is deliberately small and historical. A published matrix records eleven selected technology transfer instruments in policies valid on 1 January 2020 across mainland China's 31 provincial regions (Liu et al., 2024). We link these counts to official 2020 contract amounts and invention grants, conditioning on 2019 outcomes and regional capacity. Provincial and macroregional holdouts address different use cases. We also inject known signals into the fixed covariate design and compare a seven-coordinate procedure with a calibration model that knows the correct single signal direction. This separates the availability of a representable signal from the practical ability to learn it.

The contribution is a case-based evaluation design and its substantive application, rather than a new ridge estimator or a general verdict on policy detail. The evidence shows heterogeneous task dependence: detail worsens contract predictions, while grants benefit under macroregional but not provincial holdouts. Simulations show how recovery costs can coexist with information that is known to be present. These are conditional predictive findings. They neither estimate the effectiveness of technology transfer policies nor establish that more detailed coding is generally preferable or dispensable.

2 Data and measurement

2.1 Published coding and official outcomes

The policy matrix is transcribed from Liu et al. (2024), Table 1, printed page 92. That study examined 107 comprehensive provincial policies valid on 1 January 2020. Its eleven retained instruments contain 1,564 coded occurrences. We reuse the published counts; we do not claim access to a newly assembled text corpus or independent inter-rater reliability. Figure 1 of the source, printed page 91, supplies provincial document counts ranging from one to seven and summing to 107. Counts of occurrences and documents measure recording volume, not implementation intensity.

The published instrument order is procurement; funding support; personnel training; information support; infrastructure; tax preferences; financial support; rights ownership; income rewards; assessment and supervision; and intellectual property protection. Funding support and financial support remain separate source categories. The source originally considered fourteen categories and omitted three before publishing the selected matrix. Accordingly, retaining all eleven tools does not mean observing the full policy system.

Technology market contract amounts are from China Statistical Yearbook 2025, Table 20–26 (National Bureau of Statistics of China, 2025). Annual invention grants are from the 2020 statistical annual report of the China National Intellectual Property Administration. Both 2019 and 2020 values are used. Research

expenditure comes from the 2019 official bulletin (National Bureau of Statistics of China et al., 2020); gross domestic product uses the common 2019 provincial table in Chengkou Statistical Yearbook 2020 (Chengkou County Statistics Bureau, 2021). Research and development intensity is recomputed using the same gross domestic product vintage. All outcome, scale and intensity variables enter as natural logarithms. No missing-value imputation is applied.

Contract amounts concern registered technology contracts, rather than realized industrial revenue (National Bureau of Statistics of China, 2022). Grants concern knowledge outputs and can reflect earlier applications and examination lags. The assembled source tables do not establish a common contract-level link from the provincial policies to either aggregate. In particular, the regional attribution of contracts cannot be equated with local use or beneficiary location without further metadata. We therefore treat the published regional outcome as the prediction target and make no claims about local implementation or transaction mechanisms.

Table 1
Primary data and boundaries

Measurement	Coverage	Definition	Source
Policy matrix	31 × 11 selected tools	1,564 occurrences; stock valid 1 January 2020	Liu et al. 2024, Table 1
Policy documents	31 document counts	107 total; range 1–7	Liu et al. 2024, Fig. 1
Contract amounts	2019 and 2020	Registered contract amounts, 100m yuan	NBS 2025, Table 20–26
Invention grants	2019 and 2020	Annual counts, not cumulative patents	CNIPA 2020, Table <link rid="tb2">2</link>–2
Capacity and scale	2019	R&D spending and common GDP vintage	Official bulletin; Chengkou yearbook

All procedures have 31 provincial policy units. No text recoding, implementation panel or causal treatment date is reconstructed.

2.2 Group definitions and retained distinctions

The primary partition retains the four analytical groups used in this reanalysis: K contains training and infrastructure; F contains funding support, tax preferences and income rewards; D contains procurement; S contains information support, financial support, rights, assessment and supervision, and intellectual property protection. These labels are operational names for a study-defined partition. They are not attributed to Liu et al.'s original classification, and F is not an exhaustive financing category. The partition is held fixed to audit a specific aggregation rather than selected for predictive performance.

To address its substantive arbitrariness, an additional sensitivity uses the classification explicitly described by Liu et al. (2024), printed pages 91–92. Among the retained tools, supply comprises funding

support, training, information and infrastructure; demand comprises procurement; environment comprises tax preferences, financial support, rights, rewards, assessment and intellectual property protection. It yields two coarse and eight within-group coordinates. Both partitions are reported. The source grouping itself omits the three categories absent from the retained matrix, so this sensitivity is not an independent reconstruction of complete demand-side policy.

For either partition, add 0.5 to each original tool count before forming group sums. Three isometric logratio coordinates of the four sums define the primary coarse block C. Separate isometric logratio transformations within K, F and S define 1, 2 and 4 coordinates respectively; procurement, a singleton, supplies none. Their concatenation defines the seven-column fine block W. The fixed Helmert bases and category order are supplied in code. Logratio coordinates respect the relative nature of compositional information (Aitchison, 1982; Egozcue et al., 2003).

Together, group shares and within-group proportions determine the smoothed eleven-tool composition. However, the resulting additive linear predictor is a specified function class, not every function of that composition. Coarse coordinates formed from arithmetic group sums are not a linear subspace of an arbitrary full-tool logratio regression. Our nesting concerns the explicitly retained coarse columns and appended fine columns. It avoids conflating feature replacement with feature addition, without claiming universal equivalence among compositional geometries.

3 Analytical design

3.1 Common benchmark and separate penalties

The primary unpenalized benchmark X contains an intercept, log 2019 gross domestic product, log 2019 research intensity, the log 2019 same-domain outcome, and I = log(1 + N), where N is the raw eleven-tool total. The last term holds recorded tool quantity common while comparing allocation. This changes the estimand from the value of all policy records to the incremental value of within-group allocation conditional on recorded volume. A quantity benchmark uses X alone; the coarse model adds C; the augmented model adds the identical C and W.

$$\min_{\beta,\alpha,\delta} \frac{1}{m}\|y - X\beta - C\alpha - W\delta\|^2 + \lambda_C\|\alpha\|^2 + \lambda_W\|\delta\|^2$$

1

Equation 1 minimizes average training squared loss plus separate coarse and fine squared-norm penalties. Benchmark coefficients remain unpenalized. Infinity removes the corresponding block exactly. The coarse-only candidate family is therefore the fine-infinity face of the augmented family. Ridge supplies stable estimates with correlated columns (Hoerl & Kennard, 1970); it does not remove measurement error or endogeneity. A larger candidate family can still perform worse because training selection is uncertain.

Each block is centered using training observations and divided by the square root of its average training variance. A common scale within each block preserves its isotropic penalty under orthogonal rotations. Scaling and benchmark residualization are recomputed inside every training fold. Coarse and fine penalties each range over ten candidates: powers of ten from 10^{-4} through 10^{4} and infinity, giving 100 pairs. Inner mean squared error chooses the pair; ties within 10^{-12} favor stronger fine shrinkage first and then stronger coarse shrinkage. The fine-infinity coarse candidates use exactly the same preprocessing and scaling as the coarse-only model.

3.2 Validation tasks and increments

Outer leave-one-province-out evaluation trains on thirty regions and predicts the excluded region. The macroregional evaluation excludes each official East, Central, West and Northeast group, containing 10, 6, 12 and 3 provinces (National Bureau of Statistics of China, 2011). Its training samples contain 21, 25, 19 and 28 regions. Both outer designs tune using inner provincial holdouts within the available training data. The macroregional exercise evaluates transfer of a provincially tuned procedure; it is not optimized by inner geographic blocking. These administrative-economic regions are not spatial buffers, and provinces can remain neighbors across a boundary (Roberts et al., 2017).

Nested evaluation separates parameter selection from the outer outcomes (Cawley & Talbot, 2010). Mean squared error is pooled over the same 31 outer predictions. The primary increment is augmented minus coarse mean squared error; root mean squared error is also reported in log units. Province-specific losses and selected penalties remain inspectable. Overlapping training samples and four unequal blocks do not support treating fold losses as independent observations. We provide descriptive comparisons without independent-fold tests or precise geographic generalization intervals (Bates et al., 2024).

Pseudocounts of 0.25 and 1 vary zero replacement. Another sensitivity moves quantity from X into the coarse penalized block, preserving it identically in the coarse and augmented models. The source-category partition tests whether the retained four-group aggregation drives the comparison. These checks are not used to choose a preferred headline specification.

The earlier non-nested representations and frozen predictions through 2024 for contracts and 2025 for grants are retained as historical diagnostics in Online Resource 2. They are not treated as additional primary validations. Their levels are frozen at 2020; later comparisons largely reuse the same prediction differences and do not increase the number of policy units. We separately tabulate policy increments under capacity-only and prior-outcome benchmarks, rather than treating improved baseline prediction as proof that an indicator's value changes.

3.3 Signal recovery with known-direction references

The simulation follows an explicit objective, data generation, estimators and performance specification (Morris et al., 2019). It retains the actual 31-region covariates. The mean under the null is X times the full-sample benchmark coefficient. Three prespecified directions use the first within-group coordinate in K, F and S. Each is residualized against the full fixed X and C and normalized to root mean square one. This

makes its observed-design linear association with the retained coarse features zero. It is a stipulated design contrast, not an intervention or an estimated policy mechanism.

For conditional signal fraction q, the added mean is sigma times sqrt[q/(1 − q)] times the standardized direction. Sigma is the benchmark residual standard deviation with 26 residual degrees of freedom. The shared null has q = 0; each direction also has q = 0.25 and 0.50. Noise is either independent normal with variance sigma squared or independent Rademacher signs multiplied by normalized leverage-adjusted residual magnitudes (HC2 adjustment). The latter is a heteroskedastic sensitivity, not a claim that this residual distribution is correct. Neither noise regime includes geographic dependence.

Each scenario has 500 repetitions, with seed 20261026 and common random numbers across directions and signal strengths within a noise regime. We compare the augmented seven-coordinate procedure with two known-direction references. The matched-coarse oracle retains the same coarse block and two-penalty tuning grid, but receives the correct single residual direction instead of seven fine features. The ideal oracle uses X and that direction only, omitting the stipulated zero coarse coefficients. Under the null the K direction supplies an irrelevant oracle candidate. Both references use knowledge of the generating mechanism unavailable in empirical analysis. The reference changes block dimension and shrinkage geometry as well as direction knowledge; its gap diagnoses the stated procedures rather than an isolated effect of knowing a direction. The full-design residualization defines the synthetic truth; it does not use simulated outcomes to construct estimator inputs.

Every fitted procedure is evaluated by outer provincial prediction. Success means at least a 5% mean squared error reduction against the coarse procedure. The same threshold and squared-loss scale apply to validation error and expected independent-outcome error. The latter averages squared deviation from the known mean plus the known noise variance at the fixed covariates. It is a synthetic conditional risk, not an empirical confidence interval. Monte Carlo standard errors describe finite simulation repetition uncertainty. Following White et al. (2024), implementation checks include exact nesting and cached prediction operators compared with direct fits.

This redesign is retrospective and motivated by review. Its initial plan preceded inspection of the new results. A matched oracle and source-based partition were added during exploratory development after successful initial execution and source verification, before numerical result inspection; both additions are identified in the accompanying plan. Neither the analysis nor its revisions are preregistered. Simulation frequencies are not probabilities that an empirical policy signal exists.

4 Results

4.1 Retaining the same coarse model changes the interpretation

Table 2 gives the primary comparison. For contracts, the augmented model has RMSE 0.4068 rather than 0.3345 under provincial holdouts, corresponding to a 48.0% increase in mean squared error. Under

macroregional holdouts the values are 0.4055 and 0.3412, a 41.3% increase. These are failures of the separately tuned augmentation procedure to improve on unchanged coarse features. They cannot be attributed to replacing those coarse functions with a different full-tool model.

Grants show a different pattern. Provincial RMSE rises from 0.1278 to 0.1330, or 8.3% in mean squared error. Macroregional RMSE falls from 0.1580 to 0.1486, an 11.6% mean squared error reduction. The two validation schemes therefore do not support the same ranking. The block-transfer improvement is a positive result and is not discarded to maintain a universal negative conclusion. It also does not establish significance, mechanism or a population-wide advantage from four unequal held-out regions.

Table 2
Primary nested comparison

Outcome	Holdout	Quantity baseline	Coarse	Coarse plus fine	MSE change
Contracts	Province	0.3300	0.3345	0.4068	+ 48.0%
Contracts	Macroregion	0.3412	0.3412	0.4055	+ 41.3%
Grants	Province	0.1278	0.1278	0.1330	+ 8.3%
Grants	Macroregion	0.1580	0.1580	0.1486	-11.6%

RMSE values are in natural-log outcome units. MSE change compares augmented to coarse, not to the quantity baseline. Lower is better. Both models retain identical benchmark and coarse columns.

Case losses qualify the pooled percentages. For provincial contracts, thirty predictions equal the coarse predictions within numerical tolerance; the entire added squared loss of 1.6628 occurs in Hainan. Under macroregional contract transfer, most loss is also in Hainan, within the held-out East block. For grants, thirteen provincial cases improve but their gains do not offset other losses. Macroregional grant gains occur in the West block, while Northeast worsens and East and Central predictions equal the coarse model. Xinjiang supplies 52.4% of the net pooled grant gain.

Exploratory deletions were prompted by these observed losses and are not prespecified validation. Removing Hainan and completely repeating all nested fits leaves contract RMSE higher for the augmented model: 0.3320 versus 0.3043 provincially and 0.3734 versus 0.3506 macroregionally. Removing Xinjiang and refitting preserves the directional grant pattern: 0.1261 versus 0.1209 provincially, but 0.1344 versus 0.1392 macroregionally. Thus neither the negative contract comparison nor the positive blocked grant comparison is solely removed by excluding its most influential case. These selected diagnostics do not establish general case robustness.

4.2 Sensitivity to grouping and quantity

Under the retained four-group partition, both alternative pseudocounts preserve the contract deterioration in both holdout designs. The grant comparison also remains directionally mixed: provincial mean squared error increases, whereas macroregional error decreases. Moving quantity into the shared coarse penalty

gives the same directional pattern. Thus the grant macroregional result is not eliminated by the three specified zero and quantity sensitivities, although its magnitude changes.

The source study's supply–demand–environment partition yields the same directional pattern. Contract RMSE rises from 0.3346 to 0.3977 provincially and from 0.3412 to 0.4049 macroregionally. Grant RMSE rises from 0.1279 to 0.1385 provincially but falls from 0.1582 to 0.1490 macroregionally, an 11.3% MSE reduction. Agreement between these two partitions supports the observed task dependence within this matrix; it does not validate their theoretical boundaries or generalize to other coding systems.

The capacity-only comparisons distinguish benchmark prediction from incremental policy information. Prior outcomes substantially improve baseline prediction: provincial RMSE falls from 0.7843 to 0.3309 for contracts and from 0.3732 to 0.1196 for grants. But most earlier policy procedures fail to improve even the capacity-only baseline. For example, full-tool contract prediction is identical to that baseline under both holdouts, while provincial grant error rises. One notable exception is the capacity-only macroregional grant result, which improves from 0.3648 to 0.3560 and loses its increment after prior grants enter. The evidence for indicator-value reversal is therefore limited; the large baseline improvement is not itself a major new finding about policy measurement. Online Resource 1 reports matched increments under both benchmark families.

4.3 Learning fine directions is harder than representing them

Table 3 reports normal-noise recovery at q = 0.25. The augmented model obtains a 5% validation mean squared error reduction in 33.0%, 40.8% and 36.4% of contract repetitions for K, F and S directions. The corresponding grant frequencies are 30.0%, 45.8% and 36.2%. With the same retained coarse features and tuning grid, the matched known-direction reference succeeds in 76.6–79.8% of contract repetitions and 78.6–80.4% of grant repetitions. Monte Carlo standard errors are approximately 1.8–2.2 percentage points. The gap is much larger than the Monte Carlo uncertainty.

Table 3
Known-signal recovery at q equals 0.25

Outcome	Direction	Seven coordinates	Matched oracle	Ideal oracle
Contracts	K	33.0 (2.1)	76.6 (1.9)	80.2 (1.8)
Contracts	F	40.8 (2.2)	79.8 (1.8)	85.0 (1.6)
Contracts	S	36.4 (2.2)	79.4 (1.8)	86.2 (1.5)
Grants	K	30.0 (2.0)	78.6 (1.8)	83.8 (1.6)
Grants	F	45.8 (2.2)	80.4 (1.8)	85.8 (1.6)
Grants	S	36.2 (2.1)	79.6 (1.8)	86.4 (1.5)

Percent of 500 normal-noise repetitions with at least 5% validation MSE gain over the coarse procedure. Parentheses give Monte Carlo standard errors in percentage points. Oracles are simulation references, not available empirical predictors.

The ideal reference that omits unnecessary coarse coefficients succeeds still more often, in 80.2–86.2% of contract and 83.8–86.4% of grant repetitions. The matched reference is the more directly comparable diagnostic because both it and the augmented procedure retain the same coarse family. Its advantage shows that learning seven unknown directions can be costly even when the fixed sample supports recovery of the correct direction supplied in advance.

Expected independent-outcome error gives a related, distinct assessment using the same 5% squared-loss threshold. At q = 0.25 under normal noise, the augmented procedure succeeds in 27.0–47.0% of contract and 21.8–53.4% of grant repetitions. The matched reference succeeds in 92.8–97.4% and 95.4–96.2%, respectively. Validation success and expected-risk success need not agree in an individual repetition. The complete null, stronger-signal and heteroskedastic scenarios, including unfavorable results for either reference, are supplied rather than selected for presentation.

Under the normal-noise null, the ideal reference obtains the specified validation gain in 31.8% of contract and 31.4% of grant repetitions, whereas expected-risk gain occurs in 17.2% and 17.8%, respectively. The matched-coarse reference obtains expected-risk gain in 0.0% and 0.2%, respectively. The ideal reference can benefit from omitting unnecessary coarse terms even under the null. A favorable validation threshold is consequently not an error-controlled significance test.

These results do not identify the reason for an empirical null. They establish a concrete distinction between representability and recoverability for this design and the stated estimators. The positive macroregional grant result remains an empirical task-specific increment; the oracle gap is synthetic calibration. Neither should be substituted for the other.

5 Discussion

5.1 Conditional information and the prediction task

This comparison addresses an ambiguity of replacing feature sets: it leaves coarse information available while adding within-group allocation. The findings support task-specific conclusions. Contracts deteriorate across both prediction tasks. Grants improve under macroregional transfer but deteriorate when an individual province is excluded. The ranking of procedures depends on which outcome and transfer setting the assessment is intended to serve.

The practical implication is conditional. An assessment should retain a transparent coarse representation as a candidate, test the additional block separately, and report the task where an increment occurs. A favorable blocked result should not justify claiming general superiority, and a failed provincial comparison should not justify abandoning all detailed coding. Policy mix maps serve description and institutional

analysis as well as prediction (Meissner & Kergroach, 2021; Howoldt, 2024). Predictive adequacy is only one validity question.

Our nested construction also narrows the statistical interpretation. It identifies the performance difference between two stated selection procedures and retains a common coarse function class. It does not estimate an intrinsic value of information independent of estimation, establish that the grouping is theoretically optimal, or discover interactions between instruments. The separate penalties are conventional regularization components; the contribution lies in how they make a conditional information question inspectable in this case.

5.2 What the oracle gap establishes

The simulations now include a counterfactual that the empirical data cannot supply: an estimator told the correct direction. When that reference succeeds much more often than the seven-coordinate model, a negative empirical comparison cannot be explained solely by asserting that thirty-one regions are incapable of supporting any useful signal. In the stipulated scenarios, this same design supports useful recovery when direction uncertainty is removed. Learning the direction and controlling nuisance coefficients are material parts of the difficulty.

Conversely, an oracle advantage does not show that such a direction exists in the historical policy matrix. The generating signals are three linear contrasts after fixed-design residualization, not measured mechanisms. The seven-coordinate procedure may have little to recover in reality, or its source coding may omit relevant implementation. Independent policies, measurement validation or linked outcomes would be required to distinguish these explanations. Stating this boundary is more informative than converting a simulated recovery rate into the power of an observational policy test.

5.3 Boundaries and further evidence

The study has only 31 provincial policy units and one primary outcome year. More simulation repetitions reduce Monte Carlo error; later outcomes do not add policy units. Regional dependence, unobserved policy histories and outcome-specific reporting remain unresolved. The four-group labels are an operational partition, while the source categories supply a theoretically recognizable alternative; neither partition is validated as a measure of actual delivery. Training-fold selection protects outer evaluation but cannot make a 100-pair tuning search stable in every small sample.

Source limitations are material. The 2020 official Ningxia contract value is 16.8054 rather than the source paper's 16.8053 in 100 million yuan; we use the official value. In later descriptive extensions, the 2022 yearbook national total exceeds the sum of regional and overseas entries by 9,301 units of 10,000 yuan, approximately 0.00195%. A separate initial Torch release differs from many yearbook regional entries. These issues are documented and no regional value is adjusted to force reconciliation. They concern the archived later-year diagnostics and do not alter the primary 2020 computation. The regional attribution question nevertheless limits the substantive interpretation of all contract aggregates.

The policy snapshot does not reveal enactment dates, amendments, expenditure, uptake or compliance. Prior outcomes can embody earlier policy effects; conditioning on them is appropriate for the stated incremental prediction question, not for estimating total policy effects (Shmueli, 2010). Annual grants can reflect prior applications rather than contemporaneous policy responses. These limitations should guide new data collection rather than invite stronger claims from the same counts.

A decisive extension would independently validate the source coding, recover dated policy records, and link instruments to compatible implementation and contract-level outcomes. An independent regional setting would test whether the observed task dependence and recovery gap recur. Until then, the contribution is an openly reproducible case audit with a better-defined comparison and calibrated limits, rather than a general theory of policy information resolution.

6 Conclusion

Preserving the same coarse predictors makes the incremental question explicit. In this 31-region application, within-group policy detail worsens contract prediction under both holdout designs; grant prediction improves under macroregional but not provincial holdouts. A known-direction reference recovers injected signals much more often than the model that must learn seven fine coordinates. The combination supports task-specific conclusions about usable information and a separate diagnosis of recovery costs. It does not establish policy effectiveness, absent latent information, or a universal benefit from detailed coding.

Declarations

Funding: Funding details are supplied in the cover letter and withheld here for anonymous review.

Competing interests: The author declares no competing interests.

Ethics approval and consent to participate: Not applicable; this secondary analysis uses published provincial aggregates and includes no recruited participants or identifiable individual records.

Data Availability: All tabular inputs used in this study, including the published provincial P1–P11 tool counts, document counts, official statistical values and derived analysis variables, are supplied in Online Resource 2 (ESM_2_Data_and_Code.zip). The complete analysis table is data/province_data_extended.csv; data/primary_inputs.csv isolates the primary inputs. Variable definitions, units, coding and transformations are documented in data/variable_dictionary.csv, data/source_registry.csv and DATA_DOCUMENTATION.txt. The package includes executable processing, analysis, simulation and plotting code, complete numerical outputs, retrospective plans and reproduction instructions. Online Resource 1 (ESM_1_Supplementary_Methods.pdf) provides supplementary methods and tables. The study reuses published aggregate counts; it did not collect the underlying 107 policy documents or the original study's document-level coding records, which are not held as part of this reanalysis. Original third-party publications are cited with source links rather than redistributed. There are

no access restrictions on the supplied analysis inputs and code for peer review. Historical non-nested analyses are separately identified in legacy_v3.zip and must not be combined with the revised primary results.

Generative AI assistance: ChatGPT (OpenAI) assisted literature discovery, analysis-code preparation and checking, restructuring, English-language polishing, and preparation and verification of the revision materials. The author is responsible for verifying and approving the final work.

References

1. Aitchison J (1982) The statistical analysis of compositional data. J Roy Stat Soc: Ser B (Methodol) 44(2):139–160. https://doi.org/10.1111/j.2517-6161.1982.tb01195.x
2. Bates S, Hastie T, Tibshirani R (2024) Cross-validation: What does it estimate and how well does it do it? J Am Stat Assoc 119(546):1434–1445. https://doi.org/10.1080/01621459.2023.2197686
3. Cawley GC, Talbot NLC (2010) On over-fitting in model selection and subsequent selection bias in performance evaluation. J Mach Learn Res 11:2079–2107. https://www.jmlr.org/papers/v11/cawley10a.html
4. Chengkou County Statistics Bureau (2021) Chengkou statistical yearbook 2020, Table 14.2: Major national economic indicators by province, 2019. https://www.cqck.gov.cn/zwgk_270/fdzdgknr/tjxx_55116/sjfb/tjnj/202111/t20211104_9926553.html
5. China National Intellectual Property Administration (2020) Statistical annual report 2020, Table 2–2: Domestic invention patent grants by region. https://www.cnipa.gov.cn/tjxx/jianbao/year2020/b/b2.html
6. Egozcue JJ, Pawlowsky-Glahn V, Mateu-Figueras G, Barceló-Vidal C (2003) Isometric logratio transformations for compositional data analysis. Math Geol 35(3):279–300. https://doi.org/10.1023/A:1023818214614
7. Flanagan K, Uyarra E, Laranja M (2011) Reconceptualising the policy mix for innovation. Res Policy 40(5):702–713. https://doi.org/10.1016/j.respol.2011.02.005
8. Hoerl AE, Kennard RW (1970) Ridge regression: Biased estimation for nonorthogonal problems. Technometrics 12(1):55–67. https://doi.org/10.1080/00401706.1970.10488634
9. Howoldt D (2024) Characterising innovation policy mixes in innovation systems. Res Policy 53(2):104902. https://doi.org/10.1016/j.respol.2023.104902
10. Liu Z, He Z, Jia J (2024) The influence of policy tools on the transformation of scientific and technological achievements: fsQCA configuration effect analysis based on 31 data sets. J Xihua Univ (Philosophy Social Sciences) 43(2):88–102. https://doi.org/10.12189/j.issn.1672-8505.2024.02.010(In Chinese)
11. Meissner D, Kergroach S (2021) Innovation policy mix: Mapping and measurement. J Technol Transf 46:197–222. https://doi.org/10.1007/s10961-019-09767-4

12. Morris TP, White IR, Crowther MJ (2019) Using simulation studies to evaluate statistical methods. Stat Med 38(11):2074–2102. https://doi.org/10.1002/sim.8086
13. National Bureau of Statistics of China, Ministry of Science and Technology of China, & Ministry of Finance of China (2020) Statistical bulletin on national science and technology expenditure, 2019. https://m.mof.gov.cn/czxw/202008/t20200831_3578026.htm
14. National Bureau of Statistics of China (2011) Classification of the eastern, central, western and northeastern regions. https://www.stats.gov.cn/zt_18555/zthd/sjtjr/dejtjkfr/tjkp/202302/t20230216_1909741.htm
15. National Bureau of Statistics of China (2022) Statistical survey system for the technology market. https://www.stats.gov.cn/fw/bmdcxmsp/bmzd/202302/t20230215_1907187.html
16. National Bureau of Statistics of China (2025) China statistical yearbook 2025, Table 20–26: Technology market transactions by region. https://www.stats.gov.cn/sj/ndsj/2025/html/C20-26.jpg
17. Roberts DR, Bahn V, Ciuti S, Boyce MS, Elith J, Guillera-Arroita G, Hauenstein S, Lahoz-Monfort JJ, Schröder B, Thuiller W, Warton DI, Wintle BA, Hartig F, Dormann CF (2017) Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. Ecography 40(8):913–929. https://doi.org/10.1111/ecog.02881
18. Rogge KS, Reichardt K (2016) Policy mixes for sustainability transitions: An extended concept and framework for analysis. Res Policy 45(8):1620–1635. https://doi.org/10.1016/j.respol.2016.04.004
19. Shmueli G (2010) To explain or to predict? Stat Sci 25(3):289–310. https://doi.org/10.1214/10-STS330
20. White IR, Pham TM, Quartagno M, Morris TP (2024) How to check a simulation study. Int J Epidemiol 53(1):dyad134. https://doi.org/10.1093/ije/dyad134

Figures

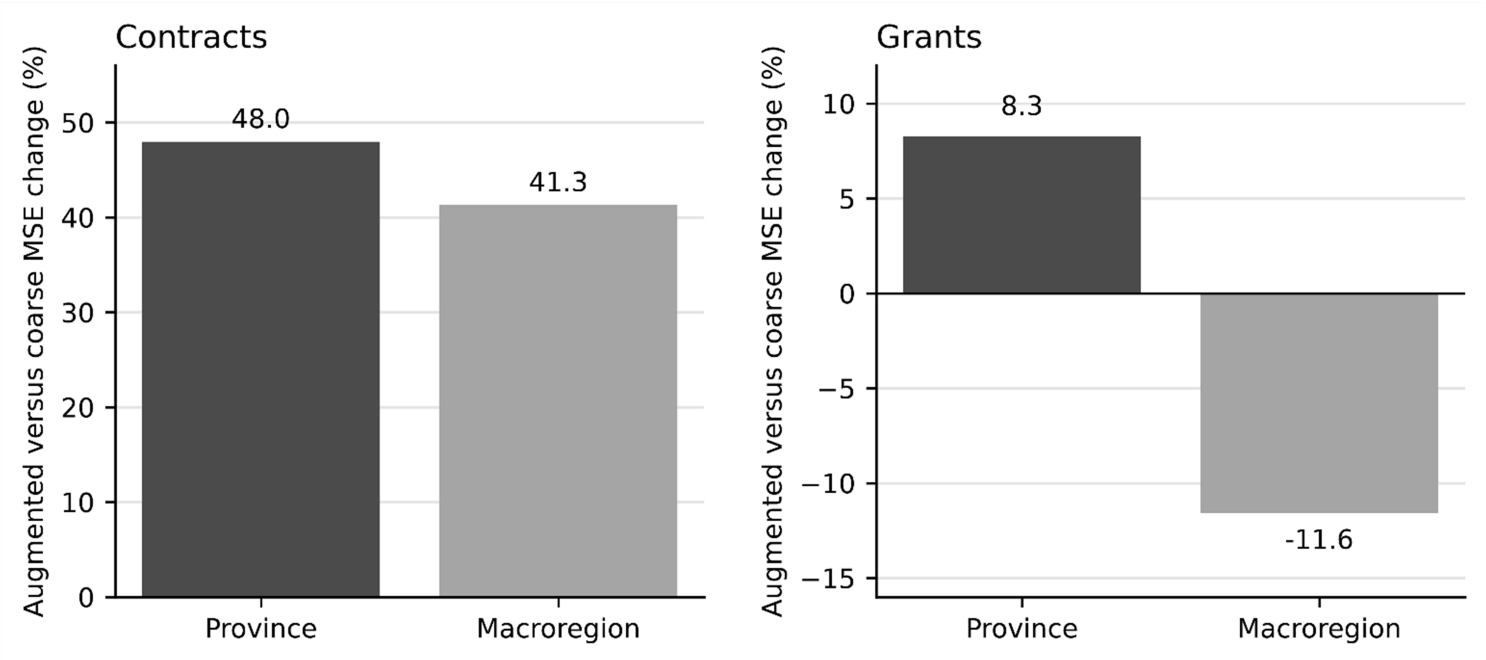

Figure 1

Conditional increment of within-group detail under the primary partition. Positive values indicate higher held-out mean squared error. The coarse columns are identical in both models; penalties are selected within outer training data. The four macroregional folds are not independent replications.

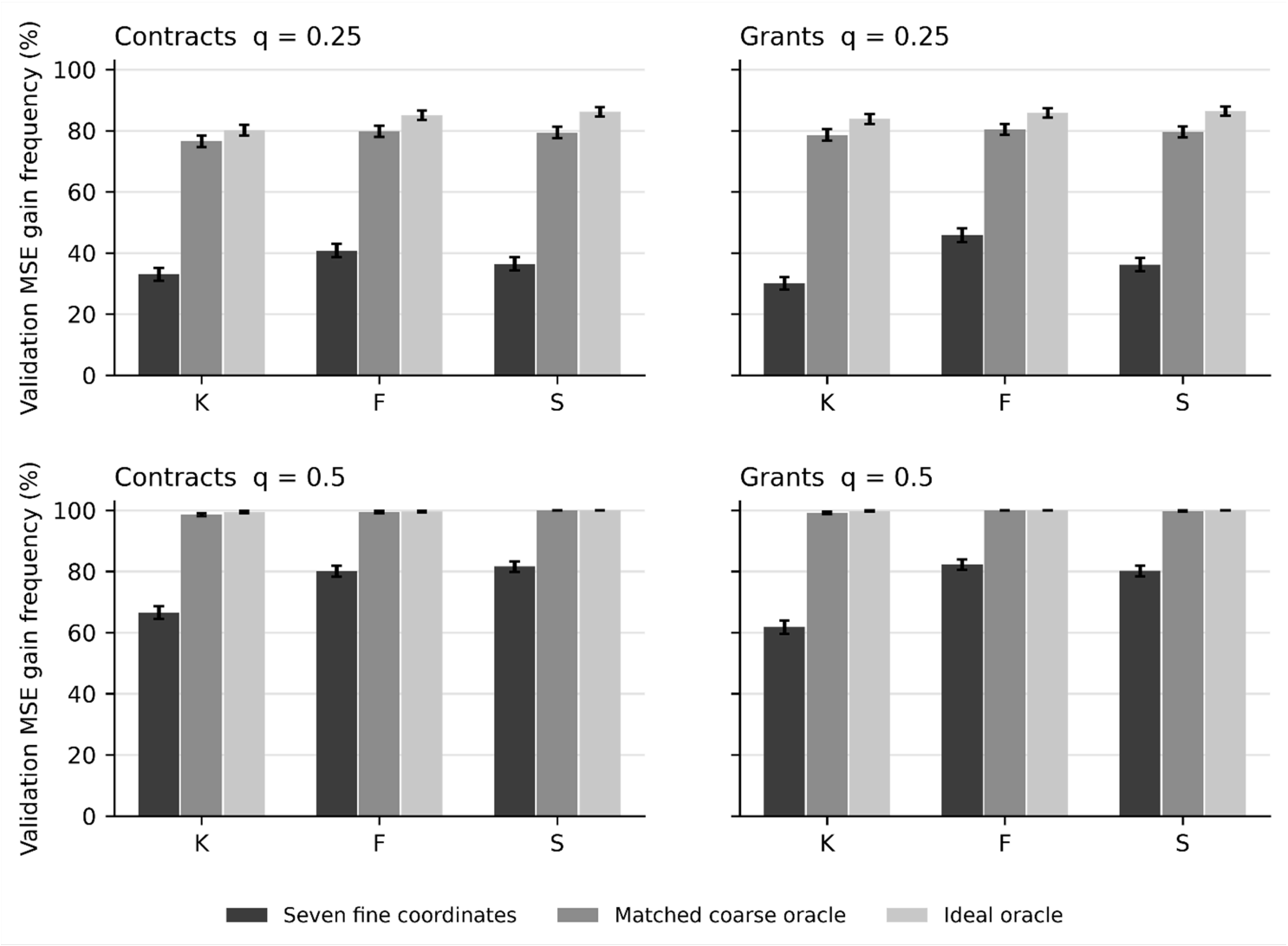

Figure 2

Recovery of known directions under normal noise. Success is at least 5% validation mean squared error reduction against the coarse procedure. Each scenario has 500 repetitions; error bars are one Monte Carlo standard error. Both oracle references have unavailable generating-model knowledge. Artificial outcomes are not empirical policy evidence.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- ESM1SupplementaryMethods.pdf

- ESM2DataandCode.zip