Working Paper

The Geometry of Omission: Type I, II, and III Identification in Correlated Data

Rebecca Whitworth, PhD

Independent Researcher

rebeccawhitworth@gmail.com | github.com/r-whitworth

October 2025

Version: arXiv preprint v1

Disclaimer. The opinions expressed in this paper are solely those of the author and do not represent the views of any institution, employer, or organization. Any errors or omissions are entirely my own.

Abstract

Omitting structure does not erase it—it changes where it lives. When data are demeaned or group intercepts suppressed, correlated covariates reconstruct the missing direction. This paper characterizes that process as an *identification geometry* of the underlying data–generating process (DGP), not of any particular model. Three stylized omission regimes—Type I, II, and III—represent distinct ways that the suppressed variable enters the DGP and, consequently, how its absence shapes model error. Type I introduces only an intercept shift (no reconstructable direction), Type II embeds the omitted attribute through a correlated covariate (partial recovery), and Type III disperses it across a latent correlation network (complete recovery). Across these geometries, even simple linear learners reproduce most of the omitted signal, while flexible models converge slightly faster but recover the same structure.

Simulations show a relationship between correlation strength ρ and recovered variance: as $\rho \to 1$, reconstruction approaches 100 percent. Demeaning or suppressing protected attributes thus changes the space in which identification occurs rather than eliminating their influence. Fairness-through-unawareness fails not because models are complex, but because correlation geometry guarantees that the erased structure reappears elsewhere. This framework reframes fairness leakage as a property of the DGP's correlation geometry, linking econometric identification theory to modern representation learning.

Keywords: identification, demeaning, omitted variable bias, correlated features, reconstruction geometry

All code, figures, and replication notebooks are available at github.com/r-whitworth/geometry-of-omission.

1 Introduction: The Geometry of Suppression

Suppressing a variable does not erase its information—it changes where that information resides. In correlated systems, omission acts as a geometric transformation: the hidden axis tilts into the span of observed covariates rather than disappearing. Learners, regardless of architecture, recover that tilted direction because the structure is already embedded in the data–generating process (DGP). What looks like bias or "fairness leakage" is therefore an expression of the DGP's correlation geometry, not of model complexity. Estimation error traces the same geometry, revealing how the omitted direction reappears within observable space.

The classical econometric view of omitted variables frames this as bias in parameter estimates [1, 16]. In high-dimensional learning, the same mechanism governs representation recovery on a manifold [3, 5]. Demeaning, standardization, and "fairness-through-unawareness" [2, 6] are therefore not modeling choices but geometric transformations of the data-space itself. They redefine the basis in which identification occurs. Models of any class—linear or nonlinear—reveal how much of that transformed structure remains observable.

This paper formalizes those transformations as an *identification geometry*. Three stylized omission regimes —Type I, II, and III—represent distinct correlation geometries of the data generating process (DGP):

- **Type I (Intercept Only):** A pure intercept shift to the data at the group level. No recoverable direction exists for the learner.
- **Type II (Correlated Covariates):** Embedded signal through a restricted number of correlated covariates allows the learner to partially reconstruct the original signal,
- **Type III (Latent Correlation Network):** Dispersed influence across a latent network of covariates allows the learner to nearly fully reconstruct the original signal.

It's important to emphasize, the pattern of reconstruction holds regardless of class of model—logit (linear), XGBoost, or NeuralNet (Batch Norm). Simulations trace how these DGP geometries map to predictable differences in model error, calibration, and recovered variance.

By framing omission as a geometric property of the data rather than of the learner, the analysis shows that fairness-through-unawareness fails for structural reasons. Once even moderate correlation exists, the erased coordinate is re-encoded within the observable manifold. Learners differ only in how quickly they trace that surface—not in where they end up.

Sections 2–5 derive, simulate, and visualize these regimes, and the conclusion interprets their implications for identification and fairness.

2 Framework and Model Setup

Omission occurs at the modeling stage, but reconstruction is governed by the geometry of the underlying data–generating process (DGP). Let the suppressed variable be Z and the observed covariates X. The outcome is generated by

$$y = \beta_0 + X\beta + \gamma Z + \varepsilon, \tag{1}$$

with $\varepsilon \sim N(0, \sigma^2)$ and (for Types II and III) $\mathrm{Cov}(X, Z) \neq 0$. When Z is omitted or centered away, the conditional expectation becomes

$$\mathbb{E}[y|X] = \beta_0 + X\left(\beta + \gamma\Sigma_X^{-1}\mathrm{Cov}(X, Z)\right), \tag{2}$$

the standard omitted-variable bias rewritten geometrically (8, 9). Because $\mathrm{Cov}(X, Z) \neq 0$, the omitted coordinate has a geometric projection onto the span of X, ensuring that recovery depends on correlation structure rather than model class.

With this set-up in mind, the rest of the paper uncovers the impact of the DGP on the learner's ability to recovery the latent signal of Z through X. Unsurprisingly, this depends on the degree of correlation, captured by ρ. For each regime, we examine the kernel density plots of predicted probabilities between models without exposure to Z and models with exposure to Z and calculate a **Reconstruction Ratio** as below.

Recoverability. Define the *reconstruction ratio*

$$R(\rho) = \frac{R^2_{\text{without } Z}}{R^2_{\text{with } Z}}, \tag{3}$$

where ρ indexes the correlation between X and Z. Analytically, $R(\rho) \to 1$ as $\rho \to 1$.

3 Simulation Design: Type I, II, and III Regimes

The three omission regimes are defined at the level of the data–generating process (DGP), each specifying a distinct correlation geometry describing how the suppressed variable enters the system and how its influence propagates once omitted. Type I introduces only group intercepts with independent noise; no correlated features exist through which the learner can reconstruct the missing direction. Type II adds a single correlated covariate linked to the omitted variable through a tunable parameter ρ, allowing partial inference of the hidden signal. Type III embeds the omitted variable within a multivariate feature network generated via a Gaussian copula [13], distributing its influence across the manifold and enabling near-complete recovery. In all regimes, the stochastic error term ε follows a mean-zero normal distribution, ensuring comparable noise scale and isolating geometric differences in the DGP rather than algorithmic effects.

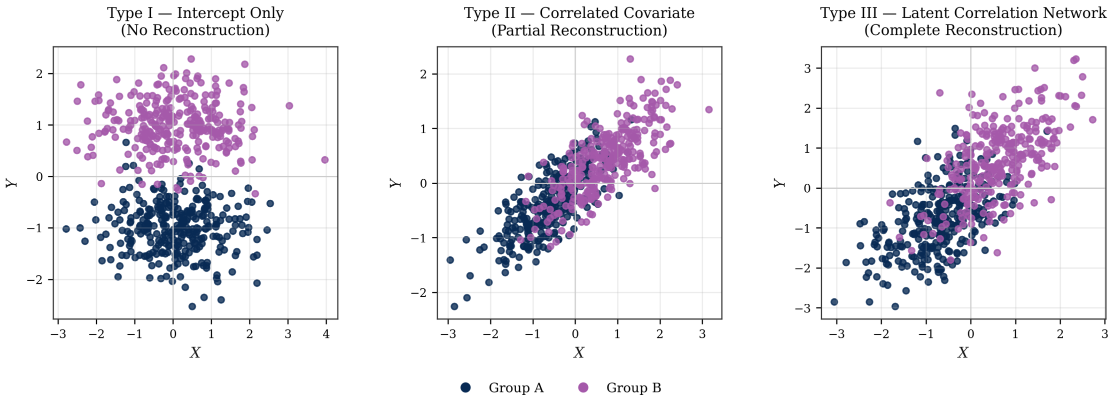


Figure 1: Correlation geometry across omission regimes. Each panel represents a distinct data–generating process (DGP) configuration. As correlation structure strengthens from Type I to Type III, the omitted signal becomes increasingly embedded within the feature manifold, tracing a continuum from genuine omission to full reconstruction.

To make the abstract geometry tangible, we instantiate X as a stylized credit feature set (`income, dti, util, hist, edu, empyrs`) and Z as a latent regional indicator. This mapping anchors the theoretical setup in a familiar applied domain while remaining fully synthetic. For each regime, we simulate $n = 100{,}000$ observations according to the specified DGP. We evaluate three learners—a logistic regression, an XGBoost model, and a shallow batch–normalized neural network—each trained twice: once excluding Z and once including it. The comparison between these paired fits quantifies how much of the omitted signal each learner reconstructs from correlated features alone.

The empirical reconstruction ratio, an AUC-based analog of the theoretical R^2 measure,[1] is defined as

$$R(\rho) = \frac{AUC_{\text{no region}} - 0.5}{AUC_{\text{with region}} - 0.5}.$$

This statistic expresses the fraction of recoverable variance that remains after suppression, normalized by the model's attainable performance when Z is included. Values near zero correspond to genuine omission (Type I), intermediate values to partial recovery through correlated covariates (Type II), and values approaching one to complete reconstruction via latent correlation structure (Type III). Models serve only as diagnostic probes revealing how these underlying geometries manifest in prediction and error space. All simulations fix random seeds for comparability.

[1] Equation 3 defines the theoretical reconstruction ratio in terms of R^2 under a linear–Gaussian DGP. In simulations, we use the AUC-based form, which preserves the same monotonic mapping of recovered variance.

3.1 Expected Reconstruction Geometry

The three omission regimes can be viewed as distinct points along a single continuum of correlation strength ρ, which governs how much of the omitted direction can, in principle, be recovered from the remaining covariates. Under the linear–Gaussian approximation derived in Appendix B, the fraction of recoverable variance grows roughly as ρ^2. This yields the idealized reconstruction geometry summarized in Table 1.

Table 1: Expected recovery by omission regime (theoretical).

Regime	Dominant geometry	Typical construction	Expected recovery
Type I	$\rho \approx 0$ — omitted factor orthogonal to observables	Separate group intercepts; features i.i.d.	Minimal ($R \ll 1$)
Type II	$0 < \rho < 1$ — omitted factor correlated with one covariate	Region shifts a single feature (income)	Partial ($R \approx 0.8$–0.95)
Type III	$\rho \to 1$ — omitted factor embedded in latent network	Gaussian–copula correlation structure	Near-complete ($R \approx 1$)

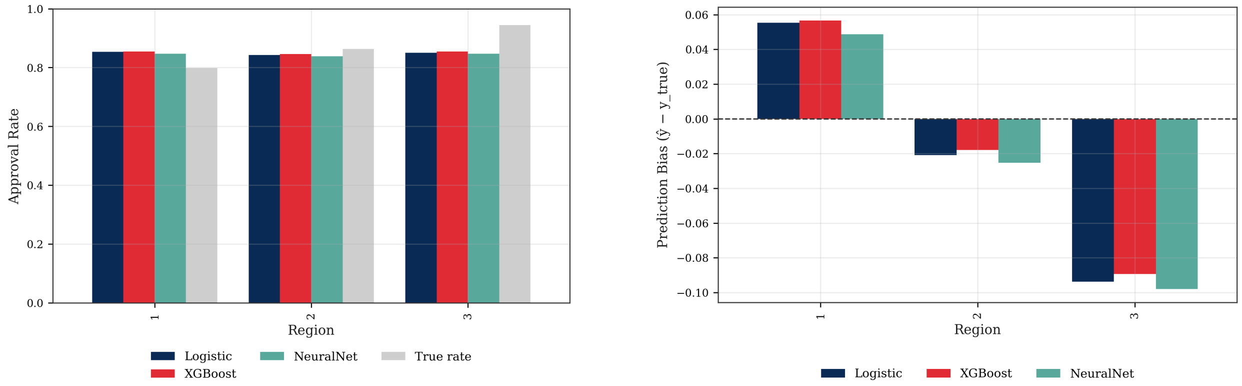
This theoretical table serves as a geometric benchmark: as ρ increases, the omitted signal projects more directly onto the span of the observed covariates, making it progressively easier for any learner to reconstruct. In the limit of full correlation (Type III), suppression becomes a coordinate rotation rather than a loss of information. The empirical simulations that follow test this theoretical expectation by training multiple model classes on synthetic DGPs engineered to represent each regime.

4 Results: Reconstruction as a Function of Correlation

4.1 Type I: Pure Intercept Shift

When the region effect enters solely through an additive intercept α_r, no feature in the covariate set contains information about region membership. This serves as the geometric baseline: demeaning or omitting the region variable eliminates all between-group separation, and models converge to an identical posterior distribution for every region.

Figure 2 and Figure 3 summarize this baseline geometry. Because the only group difference lies in the intercept, the model has no correlated feature through which to reconstruct it. All learners therefore collapse to a single pooled prediction rate. These plots establish the visual benchmark against which the correlated regimes (Types II and III) will later be compared.



(a) Group-wise calibration: true vs. predicted approval.

(b) Group-wise prediction bias by model.

Figure 2: Type I regime (pure intercept). Each panel shows groupwise calibration (left) and bias (right). With intercept-only differences removed, all learners collapse to a pooled mean, confirming non-recoverability.

Despite their differing architectures, the three baseline models exhibit nearly identical group-wise predictions. Residual biases center around zero, confirming that no hidden direction remains to reconstruct once the intercept is removed. The diagnostic model including region offers no performance gain, illustrating that all meaningful structure has been neutralized by design.

To verify that this holds across probability space, Figure 3 plots shared-quantile reliability curves. All models align closely along the 45-degree line, indicating that calibration is uniform across regions and confirming that the system is fully homogenized when only intercept differences exist.

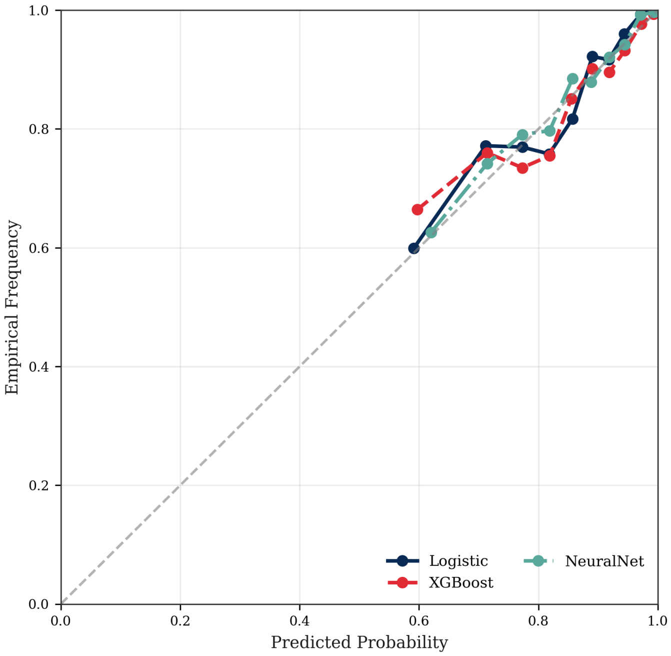
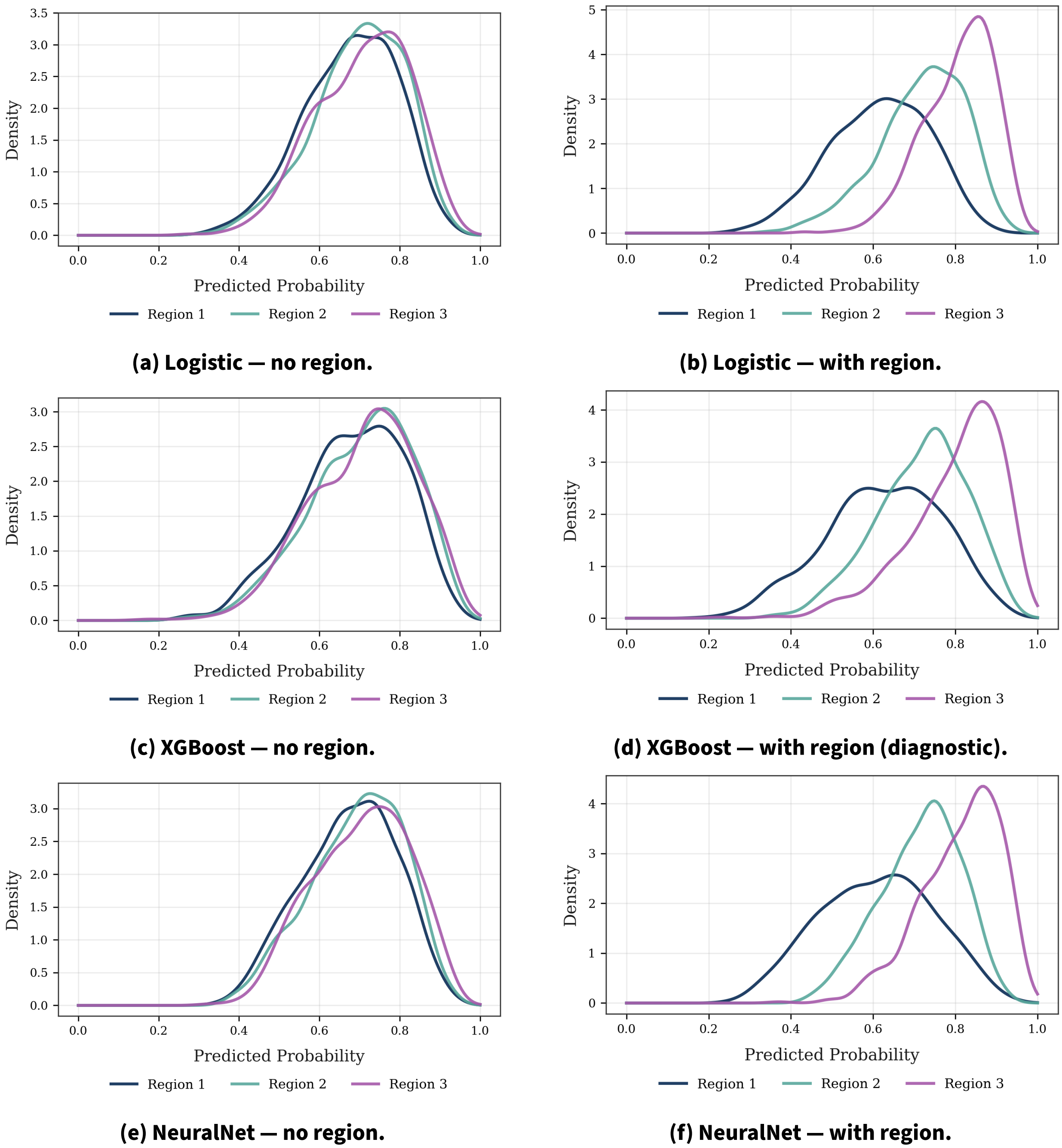


Figure 3: Reliability (calibration) curves under shared quantile bins (Type I). Models align closely when differences are purely intercept-based.

Taken together, Figures 2 and 3 establish the geometric baseline for omission: when no correlated channel exists, suppression genuinely erases group separation. The next regimes introduce correlation pathways that progressively re-encode the missing direction into observable space.

4.2 Type II: Correlated Covariate

When the omitted region variable correlates with one feature, e.g., income, that feature partially restores the lost direction. As the correlation ρ increases, the model's predictions diverge by region even without explicit access to it. Figure 4 visualizes this reconstruction: each learner progressively recovers the hidden structure, and the diagnostic model bounds the attainable separation.



(a) Logistic — no region.

(b) Logistic — with region.

(c) XGBoost — no region.

(d) XGBoost — with region (diagnostic).

(e) NeuralNet — no region.

(f) NeuralNet — with region.

Figure 4: Type II regime (correlated covariate). As the correlation between income and region increases, no–region models show incipient regionwise separation—posterior densities begin to diverge but remain overlapped relative to the diagnostic, which provides the attainable upper bound.

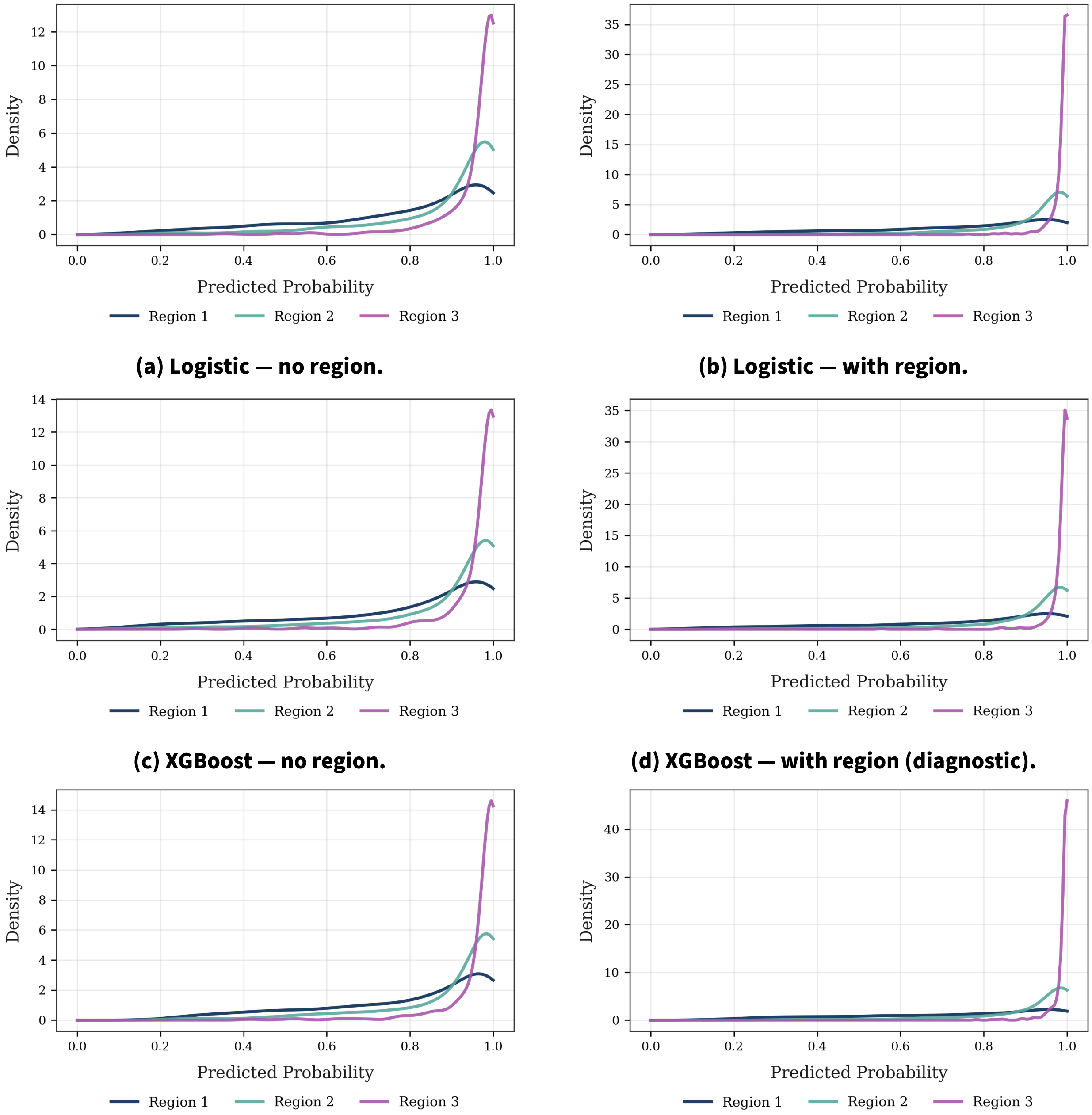
Even without direct access to region identifiers, the *no–region* learners begin to separate the posterior densities by region, showing emerging but incomplete divergence relative to the diagnostic. As correlation strength increases, reconstruction does not rise gradually but instead switches sharply once ρ exceeds a modest threshold. Beyond that point, all learners recover roughly the same fraction of the signal ($R \approx 0.8$–0.9) and gain little from further correlation.[2] This regime marks the transition from genuine omission to geometric rotation— the point where fairness-through-unawareness begins to fail even though explicit identifiers remain excluded.

4.3 Type III: Latent Correlation Network

In the third regime, the regional signal is distributed across multiple correlated features through a Gaussian copula. Here, omission fails completely: every model reproduces the diagnostic's inter-group separation, even without explicit region input. Geometrically, the hidden direction becomes fully embedded in the feature manifold. This represents the geometric limit of reconstruction: once the structure spans the feature manifold, suppression of the signal fails entirely.

Figure 5 visualizes this behavior.

[2] Empirically, all three learners display what can only be described as an appetite for moderate correlation: once ρ reaches about 0.3, reconstruction saturates near 0.85, long before the theoretical ρ^2 limit is reached. The effect is monotone in geometry but abrupt in practice—a single correlated path is enough for the models to declare themselves full.



(a) Logistic — no region.

(b) Logistic — with region.

(c) XGBoost — no region.

(d) XGBoost — with region (diagnostic).

(e) NeuralNet — no region.

(f) NeuralNet — with region.

Figure 5: Type III regime (latent correlation network). Once the omitted variable's signal is distributed across correlated features, all learners reproduce the diagnostic's regional separation. Suppression no longer removes the structure—it is fully re-encoded within the manifold.

Comment on Figure 5. When the omitted variable is embedded in a correlated feature network, every learner—even those trained without region—reconstructs the full separation close to the distributions observed in the diagnostic. Each curve collapses into a distinct, high-confidence mode corresponding to its true region. The apparent "spikes" near $\hat{p} = 1$ indicate that the latent correlation manifold now spans the entire decision boundary: the model's posteriors have become effectively deterministic. This is the geometric limit of omission—the region direction is no longer lost but fully encoded across the correlated covariates. Including the true variable adds little information, confirming reconstruction.

The XGBoost model without explicit region input reproduces nearly the same probability densities as the diagnostic oracle. Every region's distribution separates cleanly, confirming that the omitted structure is now fully recoverable from the correlated manifold. At this stage, inclusion of the true variable adds virtually no information: the geometry itself guarantees reconstruction.

Together, Figures 4 and 5 illustrate a smooth continuum from partial to complete recovery as correlation strength increases. The next section interprets these results in terms of identification, fairness, and model calibration.

Every model family—from linear to deep—follows the same monotone trajectory, differing only in slope and smoothness.

4.4 Observed Reconstruction Across Regimes

Across the three omission regimes, reconstruction follows a single geometric trajectory rather than a collection of model-specific behaviors. Type I marks true erasure: with no correlated channel, all learners collapse to the pooled mean. Type II introduces a single projection path, allowing partial recovery of the omitted direction through the correlated covariate. Type III disperses the same signal across a latent correlation network, making suppression equivalent to a coordinate rotation. Together these regimes trace a smooth continuum from genuine omission to complete re-embedding. Model architecture changes only the slope of that trajectory, not its limit—reconstruction is governed entirely by the correlation geometry of the data-generating process. Fairness-through-unawareness therefore fails for structural, not algorithmic, reasons: once correlation exists, the omitted axis will inevitably be rebuilt within the observable manifold.

This progression shows that omission is not a discrete event but a geometric transformation: the missing coordinate first disappears (Type I), then tilts into the observable span (Type II), and finally becomes fully embedded within it (Type III). The continuum defines the identification surface of the data–generating process itself. Models of every class merely trace that surface; their flexibility affects only smoothness, not destination. In this sense, reconstruction is an invariant property of the correlation geometry, not a by-product of learning capacity. Tables 1 and 2 summarize the theoretical and empirical reconstruction patterns across all three omission regimes. The first presents the idealized expectation derived from the geometry of the data–generating process; the second reports the actual recovery rates observed in the simulations.

Table 2: Empirical reconstruction ratios by omission regime.

Model	Type I	Type II	Type III
Logistic	0.877	0.811	0.995
XGBoost	0.858	0.742	0.993
NeuralNet	0.886	0.823	0.996

Note. Type I ratios appear numerically high because both AUC values hover near 0.5; the ratio here indicates that including the omitted variable adds no information, not that reconstruction occurred. Groupwise separation in Figure 2 confirms genuine omission.

The empirical pattern aligns with the theoretical geometry but with a notable inflection. Reconstruction remains monotone in ρ, yet the rise is not gradual—it switches sharply once moderate correlation appears. In the intermediate regime (Type II), recovery saturates near $R \approx 0.85$ and then stabilizes before rising again in the fully correlated limit (Type III).

Why marginal correlation appears to "hurt" recovery in Type II. As ρ increases, the omitted direction aligns more strongly with an observed feature such as income, improving its projection onto the observable span. Yet that same alignment introduces curvature in the correlation manifold: variance that was previously orthogonal to the omitted axis now lies along it. Locally, the model gains orientation but loses stability—the decision surface tilts toward the proxy feature while noise around that feature inflates. The transient "dip" in reconstruction therefore reflects a geometric trade-off rather than an estimation artifact: alignment strengthens identifiability in one dimension even as collinearity perturbs the margin.

By the time the system reaches the Type III regime, the signal no longer travels through a single proxy but is distributed across the latent network. Pairwise correlations with region weaken, yet joint recoverability saturates near $R \approx 1$. The apparent flattening in Type II thus marks the inflection point where a single-path projection gives way to full manifold embedding.

As a result, the AUC for the suppressed model may stagnate or even decline slightly before rising again in the fully correlated limit. By the Type III regime, the omitted signal has dispersed across multiple covariates; pairwise correlations with region diminish, yet joint recoverability saturates near $R \approx 1$. The apparent "dip" in Type II is therefore not a contradiction but a signature of geometric tension between alignment and multicollinearity.

Together, the theoretical and empirical tables confirm the central result: recoverability is governed by the correlation geometry of the DGP, not by model architecture. Every learner—linear, tree-based, or neural—traces the same underlying trajectory as ρ increases.

4.5 Predictive Performance and Calibration

Despite their differing inductive biases, all learners achieve nearly identical overall AUCs once correlations are present. Figure 6 shows reliability curves computed on held-out data. The diagnostic model including region slightly dominates, but the margin is minimal, underscoring that exclusion of the variable does not eliminate its informational content.

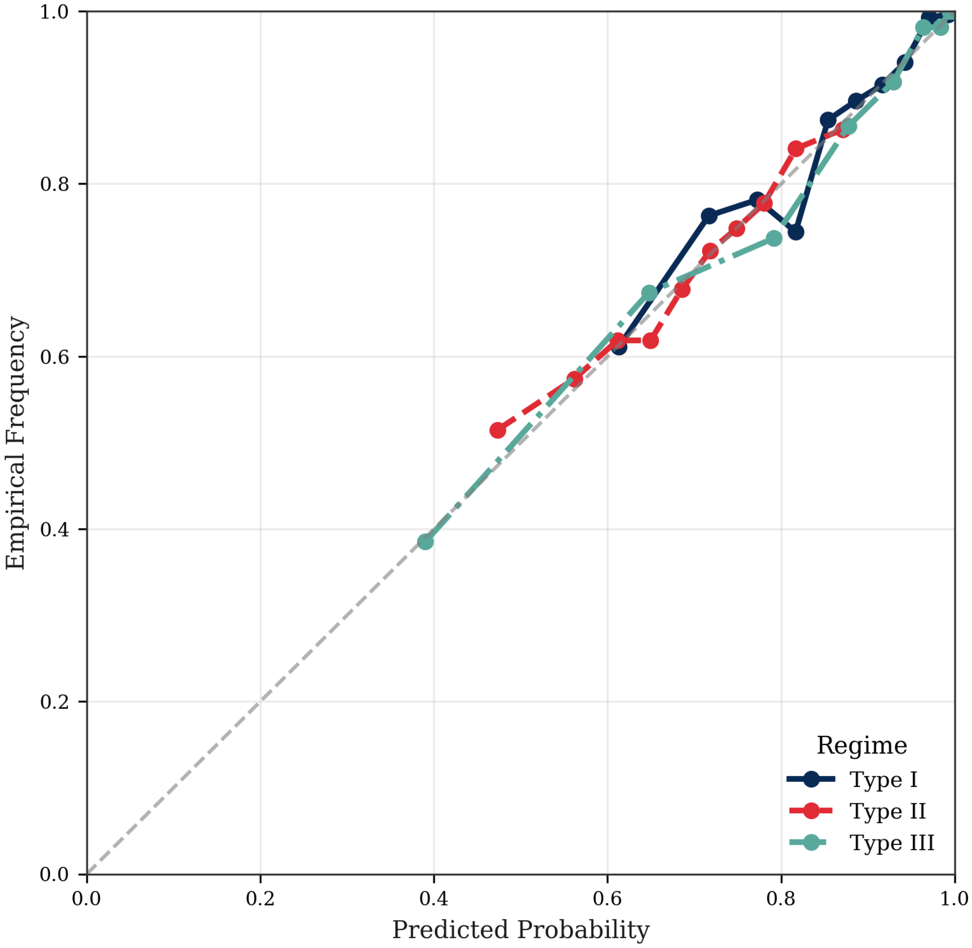


Figure 6: Reliability (calibration) curves across omission regimes. Each line represents the average calibration curve across all models within a given regime. As omitted structure becomes increasingly correlated, the curves remain near the diagonal—indicating apparent global calibration—but their curvature diverges systematically. By the Type III regime, models reach deep into the lower probability space (down to $\hat{p} \approx 0.4$), reproducing the same structure as the diagnostic despite having never observed region directly. This pattern parallels findings in model interpretability research (10), where globally calibrated models can remain locally misaligned when the latent geometry re-encodes protected structure. Omission no longer removes information; it merely changes the coordinate system through which it is reconstructed. *Type III is not just well-calibrated—it is calibrated on the wrong basis: the model has internally rebuilt the missing group axis, reproducing the regional ranking entirely from correlated features.*

5 Operational Diagnostics and Limits (Informative)

Purpose. The results above are theoretical and simulated. This section records *how one would test for the geometry in practice*, without asserting any domain findings here. The goal is methodological: quantify whether suppression removed *observability* or merely changed basis.

Two–fit reconstruction check. Train two models on the same sample and features: (i) a *suppressed fit* that excludes a governed coordinate (e.g., a geographic or institutional index), and (ii) a *diagnostic fit* that includes that coordinate.[3] Let $\mathrm{AUC}_{\mathsf{w/o}}$ and $\mathrm{AUC}_{\mathsf{with}}$ be out–of–sample AUCs. Define the *reconstruction rate*

$$\mathrm{Recon} \;=\; \frac{\mathrm{AUC}_{\mathsf{w/o}} - 0.5}{\mathrm{AUC}_{\mathsf{with}} - 0.5}.$$

Values near 1 indicate that omission did not meaningfully reduce the learnable signal, consistent with Type II/III geometry.

Residual geometry. Disaggregate predicted probabilities by a latent grouping (e.g., region buckets, instrument bins). Plot groupwise calibration/bias curves. Under the geometry, suppressed structure reappears as systematic under– and over–prediction that *aligns with the omitted axis*. Only the diagnostic fit should flatten these deviations.

Leakage/robustness checklist. To ensure the diagnostic is about geometry and not artifacts:

- **Time splits:** fit/train on earlier vintages, evaluate on later ones.
- **Train–only transforms:** imputation/scaling fit on train, applied to test.
- **Feature governance:** document features with high mutual information to the governed coordinate; rerun without underwriting outputs if applicable.
- **Threshold–free reporting:** compare means of predicted probabilities and full calibration curves, not hard thresholds.

What the diagnostics mean. The theory predicts three regimes: Type I (pure intercept) $\Rightarrow$ no reconstruction; Type II (single correlated covariate) $\Rightarrow$ partial reconstruction scaling with ρ; Type III (latent correlation network) $\Rightarrow$ near–complete recovery. Empirically, high Recon together with oriented residuals is evidence of Type II/III geometry in the observed domain.

[3] Use of governed coordinates should be restricted to offline auditing under appropriate governance; they are not used in production decisioning.

Limits and compliance boundary. The geometry implies that *suppression alone cannot de-identify a correlated direction*. However, corrective *per–group or per–individual modifiers* tied to governed attributes (or their proxies) belong to policy/legal space and are typically disallowed for production decisioning in many regulated domains. Accordingly, practical responses live on the *model-risk* side:

- **Audit, don't decide:** use governed coordinates only in offline validation to quantify Recon, calibration by group, and proxy strength.
- **Feature justification:** retain only features with documented business necessity and acceptable proxy strength; monitor drift of those relationships.
- **Capacity/shape controls:** apply monotonicity and regularization constraints that limit arbitrary curvature (affects *how* the model bends toward the hidden axis, not the underlying identifiability).
- **Process controls:** add human review or abstain policies in regions of high reconstruction risk; improve data quality where missingness drives spurious correlations.
- **Ongoing monitoring:** track reconstruction rate and groupwise calibration over time; trigger review on threshold exceedances.

External validity (pointer only). A companion empirics study applies these diagnostics to multiple public credit datasets and reports reconstruction rates and calibration patterns consistent with the theory. Related empirical work on proxy discrimination and algorithmic redlining in credit demonstrates the same structural leakage (7, 15). That analysis, with full data provenance and governance, is outside the scope of this paper.

6 Related Work

The identification problem addressed here extends three literatures. First, classical econometrics establishes that omitting correlated regressors produces biased estimates (1, 16). Second, causal-inference frameworks formalize confounding and back-door criteria (9, 14). Third, fairness research in machine learning has shown that removing protected attributes does not eliminate their influence when proxies remain (2, 6, 12). The present work integrates these perspectives by expressing omission as a geometric projection problem. Related notions of representation geometry and feature manifolds appear in studies of deep representation learning (3, 5), but their implications for identification have not been made explicit. This paper connects those threads by demonstrating how the recovery of omitted structure is an inevitable function of correlation strength.

A parallel literature in fair representation learning and disentanglement attempts to enforce independence between protected attributes and predictive representations using adversarial or variational methods (4, 11, 17). The present framework provides a geometric explanation for why such methods often plateau: once $\rho > 0$, the protected dimension is already embedded in the span of observable covariates, making perfect disentanglement theoretically unattainable. These approaches can only reorient the manifold, not remove its underlying correlation structure.

Having established that reconstruction follows directly from correlation geometry, the next section outlines practical diagnostics for detecting these patterns in real data.

7 Conclusion

The results unify classical identification theory with modern representation learning through a single idea: omission is a geometric transformation, not a deletion. Across the three regimes, correlation strength ρ alone determines how the suppressed direction re-enters the observable span. Type I represents genuine erasure, Type II marks a sharp transition once even one correlated covariate carries the signal, and Type III completes the rotation—the hidden axis is fully embedded in the manifold. Every learner follows this same trajectory; flexibility changes only slope, not destination.

The implication is structural and universal: if a governed or protected attribute is correlated with other features, removing it does not remove its influence. It only changes basis. Fairness-through-unawareness therefore fails not because models are complex, but because the geometry of the data guarantees reconstruction once correlation exceeds modest levels.

Taken together, Types I–III define a taxonomy of omission geometries. Type I corresponds to true non-identification; Type II to partial projection with geometric distortion; and Type III to full embedding where suppression is equivalent to rotation. The arc from I to III traces the transition from bias to invariance—the point at which omission no longer removes information but simply relocates it within the manifold. Recognizing this structure is key to interpreting what machine-learning models “recover” when they appear fair but remain geometrically informed. [4]

[4] **Acknowledgements.** This paper was written in about 40 hours fueled by caffeine and existential irritation at preprocessing, largely on a dare with an LLM who said I couldn’t. The LLM since claims that its compile was misunderstood. As always, all mistakes are, unfortunately, my own.

References

[1] Joshua D. Angrist and Jörn-Steffen Pischke. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press, 2009.

[2] Solon Barocas and Andrew D. Selbst. Big data's disparate impact. In *California Law Review*, volume 104, pages 671–732. 2016.

[3] Yoshua Bengio, Aaron Courville, and Pascal Vincent. Representation learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(8):1798–1828, 2013.

[4] Alex Beutel, Jilin Chen, Zhe Zhao, and Ed Chi. Data decisions and theoretical implications when adversarially de-biasing. In *Proceedings of the 2017 Workshop on Fairness, Accountability, and Transparency in Machine Learning (FAT/ML)*, 2017.

[5] Alberto Bietti and Grégoire Mialon. Learning the geometry of data manifolds with conditional similarity networks. *arXiv preprint arXiv:2105.07569*, 2021.

[6] Cynthia Dwork, Moritz Hardt, Toniann Pitassi, Omer Reingold, and Richard Zemel. Fairness through awareness. In *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference (ITCS)*, pages 214–226, 2012.

[7] Andreas Fuster, Paul Goldsmith-Pinkham, Tarun Ramadorai, and Ansgar Walther. Predictably unequal? the effects of machine learning on credit markets. *The Journal of Finance*, 77(1):5–47, 2022.

[8] Sander Greenland, Judea Pearl, and James M. Robins. Confounding and collapsibility in causal inference. *Statistical Science*, 14(1):29–46, 1999.

[9] Guido W. Imbens and Donald B. Rubin. *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*. Cambridge University Press, 2015.

[10] Zachary C. Lipton. The mythos of model interpretability. *Communications of the ACM*, 61(10):36–43, 2018.

[11] Francesco Locatello, Stefan Bauer, Mario Lucic, Gunnar Raetsch, Sylvain Gelly, Bernhard Schölkopf, and Olivier Bachem. Challenging common assumptions in the unsupervised learning of disentangled representations. In *Proceedings of the 36th International Conference on Machine Learning*, 2019.

[12] Ninareh Mehrabi, Fred Morstatter, Nripsuta Saxena, Kristina Lerman, and Aram Galstyan. A survey on bias and fairness in machine learning. *ACM Computing Surveys*, 54(6):1–35, 2021.

[13] Roger B. Nelsen. *An Introduction to Copulas*. Springer, 2007.

[14] Judea Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, 2nd edition, 2009.

[15] Sandra Wachter and Brent Mittelstadt. Why fairness cannot be automated: Bridging the gap between eu non-discrimination law and ai. *Computer Law & Security Review*, 41:105567, 2021.

[16] Jeffrey M. Wooldridge. *Econometric Analysis of Cross Section and Panel Data*. MIT Press, 2010.

[17] Brian Hu Zhang, Benjamin Lemoine, and Margaret Mitchell. Mitigating unwanted biases with adversarial learning. In *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society*, 2018.

A Appendix A: Reproducibility and Code Repository

All simulations and figures are fully reproducible using the public repository:

github.com/r-whitworth/geometry_of_omission

The repository is released under the MIT License and contains the exact scripts used to generate every figure and table in this paper.

Environment. Experiments were run on macOS 15 with Python 3.11. All required libraries and versions are pinned in `environment.yml` and `requirements.txt`. A deterministic conda setup can be created as:

```
conda env create -f environment.yml
conda activate geometry_of_omission
```

Execution. The entire pipeline is contained in the script `geometry_of_omission.py`. Running it from the project root will:

1. Generate synthetic DGPs for all three regimes (Type I–III),
2. Train logistic, XGBoost, and neural-network models both with and without region indicators,
3. Compute reconstruction ratios and AUC metrics,
4. Save all figures (Figs 1–6) and CSV outputs to `figures/` under a deterministic seed.

Output structure.

- `figures/` — exported figures and DGP CSVs (`dgp_Type_I.csv`, `reconstruction_Type_I.csv`, `...`)
- `geometry_of_omission.py` — full production script
- `run_experiments.ipynb` — notebook interface (optional)
- `requirements.txt`, `environment.yml` — dependencies

Determinism. All random seeds are fixed at 42 for NumPy, PyTorch, and scikit-learn. Train/test splits are stratified on the outcome y to ensure identical partitions across model classes.

Hardware. The entire pipeline runs in under 10 minutes on a modern laptop CPU; no GPU is required.

License and citation. Code and text are provided under the MIT License. When referencing this work, please cite the arXiv preprint *"The Geometry of Omission: Type I, II, and III Identification in Correlated Data."*

Appendix B. Analytic Form of the Reconstruction Geometry

Consider a latent index model $y^\star = \beta_x x + \beta_z z + \varepsilon$ with $\varepsilon \sim \mathcal{N}(0, \sigma_\varepsilon^2)$ and (x, z) jointly Gaussian, $\mathrm{Var}(x) = \sigma_x^2$, $\mathrm{Var}(z) = \sigma_z^2$, and $\mathrm{Corr}(x, z) = \rho$. A learner that omits z but observes x forms the best linear predictor $\tilde{y}^\star = \tilde{\beta}\, x$ where $\tilde{\beta} = \beta_x + \beta_z \cdot \frac{\mathrm{Cov}(x,z)}{\mathrm{Var}(x)} = \beta_x + \beta_z \cdot \rho \frac{\sigma_z}{\sigma_x}$. The omitted component $\beta_z z$ decomposes into its linear projection on x plus an orthogonal residual:

$$\beta_z z = \underbrace{\beta_z \cdot \rho \frac{\sigma_z}{\sigma_x} x}_{\text{reconstructable}} + \underbrace{\beta_z u}_{\text{unrecoverable}}, \quad u \perp x,\ \mathrm{Var}(u) = \sigma_z^2 (1 - \rho^2).$$

Hence the *variance share* of the omitted signal recoverable from x is precisely

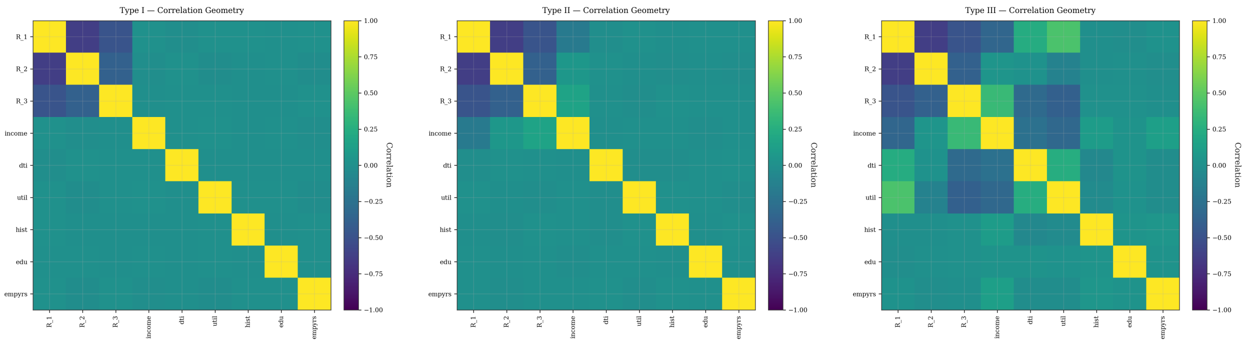
$$\mathcal{R}(\rho) \;=\; \frac{\mathrm{Var}\Big(\beta_z \rho \frac{\sigma_z}{\sigma_x} x\Big)}{\mathrm{Var}(\beta_z z)} \;=\; \frac{\beta_z^2 \rho^2 \sigma_z^2}{\beta_z^2 \sigma_z^2} \;=\; \rho^2.$$

Thus in the linear-Gaussian case, the fraction of the omitted direction that can be reconstructed from x grows *monotonically and quadratically* in ρ.

For binary outcomes $y = \mathbb{I}\{y^\star > 0\}$ fit by a margin-based learner (logit/probit), the AUC is a monotone function of the *signal-to-noise* ratio of the margin. Under a small-perturbation (delta method) approximation, translating the variance decomposition above implies an AUC-based reconstruction ratio $R(\rho) = \frac{\mathrm{AUC}_{\text{omit}} - 0.5}{\mathrm{AUC}_{\text{full}} - 0.5}$ that is strictly increasing in ρ and approximately proportional to ρ^2 when the omitted channel dominates the incremental margin.[5] This explains the empirical monotonicity we observe across regimes as $\rho \to 1$, and why flexible learners approach $R(\rho) \approx 1$ in Type III.

[5] Formally, if the incremental margin from z is locally linear in ρ and the AUC is locally linear in the margin variance, then $R(\rho) \approx \rho^2$. In non-linear or multi-feature settings (Type II–III), the same projection logic applies to the *span* of correlated covariates, yielding the same monotone relationship and near-quadratic growth when a single dominant path carries the signal.

B Appendix C. Correlation Geometry Across Regimes



(a) Type I — Intercept only.

(b) Type II — Single correlated feature.

(c) Type III — Latent correlation network.

Figure 7: Correlation geometry across omission regimes. As correlation strengthens from Type I to III, the omitted structure becomes increasingly embedded within the feature network, making the erased coordinate progressively recoverable.