Irrelevant Donors: Causes and Consequences of Spurious Pre-Treatment Fits in Causal Panel Methods

Frederico Guilherme Nogueira (frederico.nogueira@gmail.com)

=> Core Geometric Assumption

Assumption 2.1 (Approximate Irrelevance with Structured or Unstructured Contamination)

Let the pre-treatment outcome matrix be partitioned as $X = [X_{\text{rel}}, X_{\text{irrel}}]$, where relevant donors exhibit alignment bounded away from zero: $a_i \geq c > 0$ with positive probability. Irrelevant donors satisfy two key conditions that together drive diagnostic degradation:

(i) **Vanishing Alignment (Economic Irrelevance)**. The projection of X_{irrel} onto the factor space of the treated unit vanishes asymptotically, ensuring that irrelevant donors provide negligible signal about the counterfactual. This condition holds even for donors with partial factor overlap if orthogonal components dominate the variance in their loadings, as their contribution to the treated unit's structure becomes negligible relative to noise.

(ii) **Admissible Geometric Complexity (Saturation Mechanism)**. The effective rank of X_{irrel} may increase with T_0, yielding the contamination ratio

$$\kappa \equiv \frac{\text{rank}\,(X_{\text{irrel}})}{T_0},$$

which controls the attenuation in projection-based diagnostics (formalized in Proposition 3.1). This setup encompasses:

- **Unstructured irrelevance**: Fragmented donors span high-dimensional subspaces, with $\text{rank}\,(X_{\text{irrel}}) \approx N_{\text{irrel}}$.
- **Structured irrelevance**: Donors load onto a bounded set of orthogonal factors, resulting in $\text{rank}\,(X_{\text{irrel}}) \ll N_{\text{irrel}}$.

The number of relevant donors satisfies $N_{\text{rel}} = o(T_0)$, and idiosyncratic errors follow standard moment conditions. Quantitatively, the aggregate alignment of irrelevant donors satisfies $\sum_{i \in \text{irrel}} a_i^2 = o(T_0)$, ensuring their collective signal contribution is asymptotically negligible relative to idiosyncratic variation (e.g., $a_i = O(T_0^{-\gamma})$ with $\gamma > 1/2$ suffices when κ is bounded).

=> Core Statistical Assumptions

We impose the following assumptions to facilitate the high-dimensional asymptotic analysis in Sections 3–5.

Assumption 2.2 (Strict Latent Exogeneity). For the pre-treatment period $t \leq T_0$, the idiosyncratic errors are strictly exogenous with respect to the factors and treatment assignment:

$$\mathbb{E}[\epsilon_{it} \mid \mathcal{D}, \Lambda, F] = 0$$

and ϵ_{it} is independent across units and time.

Remark: Strict exogeneity is required because projection estimators utilize the entire pre-treatment block simultaneously. Feedback from ϵ_{it} to future factors would invalidate the projection geometry.

Assumption 2.3 (Low-Rank Structure). The number of latent factors r is fixed and finite, with $r \ll \min(N, T_0)$.

Assumption 2.4 (Bounded Moments). The factors f_t and loadings λ_i are uniformly bounded. The errors ϵ_{it} are sub-Gaussian with variance proxy σ_ϵ^2.

Assumption 2.5 (Factor Persistence). The post-treatment factor path $\{f_t\}_{t>T_0}$ is generated by the same stochastic process as the pre-treatment path, or satisfies stationarity conditions sufficient to bound the shift in the factor covariance matrix Σ_F.

Proposition 3.1 (Unified Asymptotic Decomposition Unstructured and Structured Irrelevance).

Let $r = y_0 - P_X y_0$ be the projection residual, where $P_X y_0 = X\widehat{w}$ and $\widehat{w} = (X'X)^\dagger X' y_0$ minimize $\| y_0 - Xw \|_2^2$. Here $P_X = X(X'X)^\dagger X'$ denotes the orthogonal projection onto span(X), with † the Moore–Penrose inverse.

Under Assumptions 2.1–2.5 and Definition 3.2, decompose the donor pool as $N = N_{\text{rel}} + N_{\text{irrel}}$. As $T_0 \to \infty$ with $rank(X_{\text{irrel}})/T_0 \to \kappa \geq 0$ holding N_{rel} fixed or growing sublinearly ($N_{\text{rel}}/T_0 \to 0$), the normalized squared residual satisfies:

$$\frac{1}{T_0} \| r \|^2 \to_p \underbrace{\mathcal{M}(\delta, \mathcal{U}) \cdot \delta^2}_{\substack{\text{Attenuated Relevance} \\ \text{Violation}}} + \underbrace{(1-\kappa) \cdot \sigma_\epsilon^2}_{\text{Attenuated Noise}}$$

Where δ^2 is the squared structural relevance violation (Definition 3.1), σ_ϵ^2 is the idiosyncratic variance, $\mathcal{M}(\delta, \mathcal{U}) \in [0,1]$ is the signal masking factor (depends on contamination structure), and κ is the contamination ratio. The noise attenuation $(1-\kappa)$ is universal across all cases, while the violation attenuation $\mathcal{M}(\delta, \mathcal{U})$ exhibits case-specific behavior:

(i) The Precise Formula (Finite Sample Decomposition): The effective rank decomposes into signal capture and geometric saturation:

$$\frac{k_{\text{eff}}}{T_0} \approx \underbrace{\underbrace{\frac{r}{T_0}}_{\text{Classical Term } (\to 0)}}_{\substack{\text{Signal + Standard} \\ \text{Overfitting}}} + \underbrace{\underbrace{\frac{\min(rank(X_{\text{irrel}}), T_0 - r)}{T_0}}_{\text{Structural Term } (\to\kappa)}}_{\text{Spurious Fit}}$$

(Note: Relevant donors contribute the factor dimension r. Irrelevant donors fill the remaining noise space $(T_0 - r)$ up the rank of their outcome matrix X_{irrel}.)

It is crucial to distinguish classical overfitting from the spurious fit mechanism: **Standard Overfitting:** Occurs when relevant donors (N_{rel}) fit the idiosyncratic noise of the treated unit. This is a variance issue (r/T_0 term), remediable by increasing T_0 or regularization; **Spurious Fit (κ-Mechanism):** Occurs when irrelevant donors fit the noise or the structural mismatch δ. This is a bias issue. The projection loads on dimensions orthogonal to the true factors F, masking the non-existence of the counterfactual. This form of overfitting persists asymptotically if $\kappa > 0$ and requires donor purification for dense geometries.

Noise attenuation is governed by the rank fraction: $(1-\kappa) = 1 - \frac{rank(X_{\text{irrel}})}{T_0}$.

Relevance violation attenuation depends on contamination structure:

- **Case A (Unstructured/Dense Irrelevance):** When irrelevant donors are mutually uncorrelated with rank(X_{irrel}) $\approx N_{\text{irrel}}$, the irrelevant subspace $\mathcal{U}$ is Haar-distributed. By concentration of measure on the Grassmannian: $\mathcal{M} \to (1-\kappa)$ (due to isotropy). Thus, signal and noise are attenuated identically: $\frac{1}{T_0} \parallel r \parallel^2 \to_p (\delta^2 + \sigma_\epsilon^2)(1-\kappa)$.
- **Case B (Structured/Low-Rank Irrelevance):** When irrelevant donors share orthogonal factors with rank(X_{irrel}) $\ll N_{\text{irrel}}$, the irrelevant subspace $\mathcal{U}$ is deterministic. Signal attenuation depends on the principal angle θ between δ and $\mathcal{U}$: $\mathcal{M} = \parallel (I - P_{\mathcal{U}})u_\delta \parallel^2$, This yields:
  - **Best case (for detection):** $\delta \perp \mathcal{U} \Rightarrow \mathcal{M} = 1$ (no relevance violation masking, full diagnostic power)
  - **Worst case (for detection):** $\delta \in \mathcal{U} \Rightarrow \mathcal{M} = 0$ (complete relevance violation masking)
  - **Baseline case (isotropic violation under unstructured irrelevance):** If the violation direction u_δ is uniformly distributed over the noise space N and the irrelevant subspace U is Haar-distributed (as in the dense irrelevance regime), then $\mathbb{E}[\mathcal{M}] = 1 - rank(X_{\text{irrel}})/(T_0 - r)$.

In many economic applications, irrelevant donors may be structured (e.g., agricultural states sharing common regional shocks orthogonal to the treated unit). In this case, the effective saturation κ is determined by the rank of the irrelevant block ($\text{rank}(X_{\text{irrel}}) \ll N_{\text{irrel}}$), preserving more projection-based diagnostic signal.

(ii) The Asymptotic Limit and Structural Saturation: As $T_0 \to \infty$ with N_{rel} fixed, the Classical Term is asymptotically negligible ($r/T_0 \to 0$), standard overfitting becomes negligible, but the Structural Term persists, defining κ, the contamination ratio or the asymptotic structural saturation parameter:

$$\kappa = \lim_{T_0 \to \infty} \frac{\min(rank(X_{\text{irrel}}), T_0 - r)}{T_0}$$

The value of κ depends on the correlation structure of the irrelevant donors:

- **Dense irrelevance:** rank(X_{irrel}) $\approx N_{\text{irrel}} \Rightarrow \kappa = \min(N_{\text{irrel}}/T_0, 1)$
- **Structured irrelevance:** rank(X_{irrel}) $= s \Rightarrow \kappa = \min(s/T_0, 1)$

(iii) Regime Classification: The behavior of the residual depends on the saturation level κ:

(a) Well-Specified Regime ($\kappa = 0$):

$$k_{\text{eff}} \to r \Longrightarrow \frac{1}{T_0} \parallel r \parallel^2 \to_p (\delta^2 + \sigma_\epsilon^2)(1{-}0)$$

Interpretation: The residual accurately reflects structural bias plus noise. Projection-based pre-treatment fit is a valid diagnostic for relevance violations.

(b) Partial Spurious Fit ($0 < \kappa < 1$):

- **Noise:** Always attenuated by factor $(1{-}\kappa)$
- **Relevance violation:**
  - Case A Unstructured: Attenuated by $(1{-}\kappa)$, masking is proportional

- Case B Structured: Attenuated by $\mathcal{M}(\delta, \mathcal{U}) = \| (I - P_{\mathcal{U}})u_\delta \|^2 \in [0,1]$, masking varies with alignment

Interpretation: Irrelevant donors partially saturate the noise subspace. Detection power degrades, but extent depends on whether contamination is fragmented (Case A) or clustered (Case B).

(c) Spurious Fit Trap ($\kappa \to 1$):

$$k_{\text{eff}} \approx T_0 \Longrightarrow \frac{1}{T_0} \| r \|^2 \to_p 0$$

Interpretation: As the effective rank approaches the full outcome dimension, corresponding to maximal-rank contamination in which irrelevant donors span the orthogonal complement, the donor space saturates the outcome space. Both relevance violation and noise are perfectly interpolated, causing pre-treatment RMSE to converge to zero regardless of the magnitude of the relevance violation δ. Consequently, fit-based diagnostics become asymptotically uninformative.

Remark (Convergence and Practical Applicability)

(a) Convergence as $\boldsymbol{\kappa} \to \mathbf{1}^-$: Proposition 3.1 establishes that for any fixed contamination level $\kappa < 1$, the residual norm converges to the stated limit as $T_0 \to \infty$. However, we do not claim uniform convergence over all $\kappa \in [0,1)$ simultaneously. As κ approaches unity, the rate of convergence slows and the approximation quality degrades. This degradation is not a phase transition but a continuous process: diagnostic power erodes smoothly as irrelevant donors increasingly saturate the noise subspace. The limiting behavior as $\kappa \to 1^-$ is precisely characterized in Theorem 4.2, which shows that the detection boundary diverges as $(1-\kappa)^{-1/4}$, making violations progressively harder to detect even though the residuals remain asymptotically well-defined.

(b) Finite-Sample Considerations: The regime classifications in part (iii) rely on asymptotics where the noise subspace dimension $T_0 - r \to \infty$. In empirical applications with short pre-treatment periods, common in comparative case studies (e.g., Abadie et al. 2010 uses $T_0 = 19$), the noise subspace is limited, and high-dimensional random matrix approximations (e.g., Marchenko-Pastur concentration) may exhibit slower convergence. Nonetheless, the finite-sample detection boundary in Theorem 4.2(i-b), which includes an explicit $\log T_0$ correction, applies exactly for any $T_0 \geq r$ via sub-Gaussian concentration inequalities. For practical calibration, researchers should interpret regime boundaries as qualitative guides rather than sharp thresholds when T_0 is small.

(c) Practical Interpretation: For applied work, the key takeaway is that $\kappa < 1$ is a necessary condition for diagnostic validity, not a sufficient one. Even when $\kappa = 0.7$ (seemingly far from saturation), 70% of relevance violations are masked, severely degrading detection power (Theorem 4.2). Conservative practice enforces $\kappa \leq 0.5$ via ex-ante donor curation, balancing diagnostic sensitivity against sample size. This threshold balances two considerations: (i) preserving at least 50% of diagnostic relevance violation, and (ii) maintaining sufficient sample size for stable weight estimation. Stricter thresholds ($\kappa \leq 0.3$) may be warranted when T_0 is small or violations are expected to be modest.

Proof.

Objective:
We wish to find the limit of the Normalized Squared Residual R:

$$R = \frac{1}{T_0} \parallel r \parallel_2^2 \text{ as } T_0 \to \infty$$

Step 1: Model Setup and Estimator Definition

First, we define the mathematical objects and the estimator.

1. Estimator: The projection estimator $\widehat{w}$ minimizes the distance between the treated unit y_0 and the donors X. The solution is the orthogonal projection.

$$\hat{y}_0 = P_X y_0$$

where $P_X = X(X^\top X)^\dagger X^\top$ is the unique orthogonal projection matrix onto span(X).

2. Residual: The residual is the difference between the observed outcome and the projection.

$$r = y_0 - \hat{y}_0 = (I - P_X)y_0$$

3. Data Structure:

- $y_0 = F\lambda_0 + \epsilon_0$ (Assumption 2.1).
- $X = [X_{\text{rel}}, X_{\text{irrel}}]$.
- Relevant donors span the factors: span$(X_{\text{rel}}) \approx$ span(F).
- Irrelevant donors are orthogonal to factors: span$(X_{\text{irrel}}) \perp$ span(F).

Step 2: Geometric Decomposition of the Treated Unit

We must split the treated unit vector y_0 into a **Signal Component** (spanned by factors) and a **Noise/Violation Component** (orthogonal to factors).

Definition (True Factor Projectors):
Let F be the matrix of true factors.

- $P_F = F(F^\top F)^{-1}F^\top$: Projector onto the Signal Space $\mathcal{S}$.
- $Q = I - P_F$: Projector onto the Noise Space $\mathcal{N}$ (Orthogonal Complement).

Decomposition:
Apply the identity matrix $I = P_F + Q$ to y_0:

$$\begin{aligned} y_0 &= (P_F + Q)(F\lambda_0 + \epsilon_0) \\ &= P_F(F\lambda_0) + Q(F\lambda_0) + P_F(\epsilon_0) + Q(\epsilon_0) \end{aligned}$$

Analysis of Terms:

1. $P_F(F\lambda_0)$: Since $F\lambda_0$ is already in the span of F, projecting it doesn't change it.

$$= F\lambda_0$$

2. $Q(F\lambda_0)$ **(The Structural Violation):** This is the part of the treated unit's signal that is *orthogonal* to the donor factors.

- *Why* $\sqrt{T_0}$? The Euclidean norm squared of a persistent relevance violation grows linearly with time (T_0). To define a stable population parameter δ^2 (Mean Squared Error), we must normalize the vector.
- We define δ_T such that $\sqrt{T_0}\delta_T = Q(F\lambda_0)$.
- Assuming the violation component satisfies a LLN so that $\| Q(F\lambda_0) \|^2/T_0 \to \delta^2$.
- Thus, $\| \sqrt{T_0}\delta_T \|^2/T_0 = \| \delta_T \|^2 \to \delta^2$.

3. $P_F(\epsilon_0)$: This is noise projected onto a low-dimensional space ($r \ll T_0$). By the Law of Large Numbers, its energy is negligible ($O_p(r/T_0) \to 0$). We ignore it for the leading order derivation.
4. $Q(\epsilon_0)$: This is the noise remaining in the high-dimensional space.

Resulting Equation for y_0:

$$y_0 = \underbrace{F\lambda_0}_{\in \mathcal{S}} + \underbrace{\sqrt{T_0}\delta_T}_{\in \mathcal{N}} + \underbrace{Q\epsilon_0}_{\in \mathcal{N}}$$

Step 3: Geometric Decomposition of the Donor Span

Now we determine what the estimator P_X actually projects onto. We use **Random Matrix Theory** to approximate the subspaces.

1. Relevant Donors (X_{rel}):
Since they load on the factors, their span approximates the Signal Space $\mathcal{S}$.

- **Theorem Used: Davis-Kahan $\sin\Theta$ Theorem (1970)**. This bounds the angle between the perturbed subspace (observed donors) and the true subspace (factors). For high Signal-to-Noise Ratio, the angle $\to 0$.
- *Result:* $\text{span}(X_{\text{rel}}) \approx \text{span}(F) = \mathcal{S}$, under the eigenvalue separation implied by Assumption 2.1.

2. Irrelevant Donors (X_{irrel}):
These lie in the Noise Space $\mathcal{N}$. Let $\mathcal{U} = \text{span}(QX_{\text{irrel}})$ be the specific subspace they span. We need the **dimension** (rank) of $\mathcal{U}$ to know how much noise they capture.

- **Theorem Used: Marchenko-Pastur Law (1967)**.
  Under standard nondegeneracy and moment conditions ensuring MP behavior for the sample covariance $\frac{1}{T_0}X_{\text{irrel}}^\top X_{\text{irrel}}$ (e.g., independent rows, or weakly dependent with suitable mixing, with uniformly bounded fourth moments), the irrelevant block is full rank so that whenever $N_{\text{irrel}}/T_0 \le \kappa^- < 1$, the sample covariance is full rank with probability tending to one:

$$\lambda_{\min} \to (1 - \sqrt{\frac{N_{\text{irrel}}}{T_0}})^2 > 0$$

- *Result:* Assume the population covariance of X_{irrel} has eigenvalues bounded away from zero. The dimension of $\mathcal{U}$ is exactly $k_{\text{eff}} = \text{rank}(X_{\text{irrel}}) \approx \min(N_{\text{irrel}}, T_0 - r)$.

3. Total Projector (P_X):
We invoke High-Dimensional Geometry.

- **Theorem Used: Vershynin (2018, Thm 4.6.1)**. Two high-dimensional random vectors from orthogonal distributions are approximately orthogonal.
- Since $S \perp N$, standard random subspace incoherence results imply that the principal angles between $\mathrm{span}(X_{\mathrm{rel}})$ and $\mathrm{span}(X_{\mathrm{irrel}})$ converge to $\pi/2$, so that $\| P_{X_{\mathrm{rel}}} P_{X_{\mathrm{irrel}}} \| = o_p(1)$,
- and cross-projection terms are asymptotically negligible.
- Under Assumption 2.1 and sub-Gaussianity of X_{irrel}, standard results on random subspace incoherence imply $\| P_X - (P_F + P_U) \| = o_p(1)$ in operator norm.
- Consequently, $P_{X_{\mathrm{rel}}} \to_p P_F$ and $P_X \to_p P_F + P_U$, with convergence in operator norm (or Frobenius norm).

where P_F projects onto Signal, and $P_{\mathcal{U}}$ projects onto the Irrelevant Subspace inside Noise.

Step 4: The Residual Expansion

We substitute the decomposition of P_X (Step 3) and y_0 (Step 2) into the residual definition.

$$r = (I - P_X)y_0 \approx (I - (P_F + P_{\mathcal{U}}))y_0$$

Algebraic Expansion:

$$r \approx (I - P_F - P_{\mathcal{U}})(\underbrace{F\lambda_0}_{\text{Signal}} + \underbrace{\sqrt{T_0}\delta_T + Q\epsilon_0}_{\text{Violation/Noise}})$$

Annihilation:

1. **Signal Term:** $(I - P_F - P_{\mathcal{U}})F\lambda_0$.
   Since $F\lambda_0 \in \mathcal{S}$, $P_F(F\lambda_0) = F\lambda_0$.

$$= F\lambda_0 - F\lambda_0 - 0 = 0$$

(The relevant donors define the signal space and successfully explain the signal).

2. **Violation/Noise Terms:**
   These vectors live in $\mathcal{N}$, so $P_F(\dots) = 0$.
   However, $P_{\mathcal{U}}$ acts on them.
   The operator becomes:

$$(I - 0 - P_{\mathcal{U}}) = (I - P_{\mathcal{U}})$$

Final Residual Vector:

$$r \approx (I - P_{\mathcal{U}})\left(\sqrt{T_0}\delta_T + Q\epsilon_0\right)$$

Interpretation: The residual is composed of the Structural Violation and the Noise, **minus** whatever part the Irrelevant Donors ($P_{\mathcal{U}}$) managed to capture (spuriously).

Step 5: The Norm Calculation (The Trace Trick)

We compute the normalized squared norm $R = \frac{1}{T_0} \| r \|_2^2$.

$$R = \frac{1}{T_0} \| (I - P_{\mathcal{U}})\sqrt{T_0}\delta_T + (I - P_{\mathcal{U}})Q\epsilon_0 \|^2$$

Expand the square $\| A+B \|^2 = \| A \|^2 + \| B \|^2 + 2A^\top B$:

$$R = \underbrace{\frac{1}{T_0} \| (I - P_{\mathcal{U}})\sqrt{T_0}\delta_T \|^2}_{\text{Relevance Violation Term}} + \underbrace{\frac{1}{T_0} \| (I - P_{\mathcal{U}})Q\epsilon_0 \|^2}_{\text{Noise Term}} + \underbrace{\text{Cross Term}}_{\to 0}$$

Cross term: $\frac{2}{T_0}(\sqrt{T_0}\delta_T)'(I - P_{\mathcal{U}})Q\epsilon_0$. Since ϵ_0 is independent of δ (Assumption 2.2), this has mean zero and variance $\text{Var}(\delta_T' \epsilon_0/\sqrt{T_0}) = \| \delta_T \|^2$ and $\sigma_\epsilon^2/T_0 = O(1/T_0) \to 0$.

(The cross term is $o_p(1)$ by Cauchy–Schwarz and concentration of sub-Gaussian quadratic forms, since $\| (I - P_U)\delta_T \| = O(1)$ and $\| (I - P_U)Q\epsilon_0 \| = O_p(\sqrt{T_0})$.)

5.1 Analyzing the Noise Term (Universal Attenuation)

We calculate the expected value of the Noise Term:

$$\mathbb{E}[N_{term}] = \frac{1}{T_0}\mathbb{E}[\epsilon_0^\top Q(I - P_{\mathcal{U}})^\top (I - P_{\mathcal{U}})Q\epsilon_0]$$

Algebraic Simplification:

1. $P_{\mathcal{U}}$ is a subspace of Q. Thus $Q(I - P_{\mathcal{U}}) = Q - P_{\mathcal{U}}$.
2. Let $M = Q - P_{\mathcal{U}}$. This is a projector, so $M^\top M = M$.
3. The term becomes $\epsilon_0^\top M \epsilon_0$.

The Trace Trick:
For $\epsilon_0 \sim (0, \sigma^2 I)$, $\mathbb{E}[\epsilon^\top M \epsilon] = \sigma^2 \text{tr}(M)$.

$$\text{tr}(M) = \text{tr}(Q) - \text{tr}(P_{\mathcal{U}})$$

Using Dimensions (From Marchenko-Pastur in Step 3):

- $\text{tr}(Q) = \dim(\mathcal{N}) = T_0 - r$.
- $\text{tr}(P_{\mathcal{U}}) = \dim(\mathcal{U}) = k_{\text{eff}}$.

Result:

$$\mathbb{E}[N_{term}] = \frac{\sigma_\epsilon^2}{T_0}(T_0 - r - k_{\text{eff}}) = \sigma_\epsilon^2\left(1 - \frac{r}{T_0} - \frac{k_{\text{eff}}}{T_0}\right)$$

As $T_0 \to \infty$, this converges to $(1-\kappa)\sigma_\epsilon^2$.

Theorem Used: **Hanson-Wright Inequality (Rudelson & Vershynin, 2013)** ensures the random quadratic form concentrates tightly around this expected value.

5.2 Analyzing the Relevance Violation Term (Case Specific)

$$V_{term} = \| (I - P_{\mathcal{U}})\delta_T \|^2$$

Let $u_{\delta_T} = \delta_T / \| \delta_T \|$ be the unit vector of the violation.

$$= \| \delta_T \|^2 \cdot \| (I - P_{\mathcal{U}}) u_{\delta_T} \|^2$$

Case A: Dense Irrelevance (Random Subspace)

- X_{irrel} is random noise.
- **Theorem Used: Rotational Invariance of Gaussian Matrices**. The subspace $\mathcal{U}$ is uniformly distributed (Haar Measure) on the Grassmannian manifold.
- Projecting a fixed vector onto a random subspace of size κ captures exactly fraction κ of the energy in expectation.

$$\mathbb{E}[\| P_{\mathcal{U}} u_{\delta_T} \|^2] = \frac{\dim(\mathcal{U})}{\dim(\mathcal{N})} = \kappa$$

- Therefore, the remaining energy is $(1-\kappa)$.
- *Result:* $(1-\kappa)\delta^2$.

Case B: Structured Irrelevance (Fixed Subspace)

- X_{irrel} shares fixed factors. $\mathcal{U}$ is a deterministic subspace.
- We cannot use the Haar measure average.
- We define the geometric masking factor $\mathcal{M}(\delta_T, \mathcal{U}) = \| (I - P_{\mathcal{U}}) u_{\delta_T} \|^2$. Equivalently, if θ is the principal angle between δ_T and $\mathcal{U}$, then $\mathcal{M} = \sin^2(\theta)$ (Pythagorean theorem in subspace decomposition).
- *Result:* $\mathcal{M}(\delta, \mathcal{U})\delta^2$.

Step 6: Final Result

Combining Step 5.1 (Noise) and Step 5.2 (Relevance Violation):

$$\frac{1}{T_0} \| r \|^2 \to_p \underbrace{\mathcal{M}(\delta, \mathcal{U}) \cdot \delta^2}_{\text{Attenuated Relevance Violation}} + \underbrace{(1-\kappa) \cdot \sigma_\epsilon^2}_{\text{Attenuated Noise}}$$

This completes the derivation.