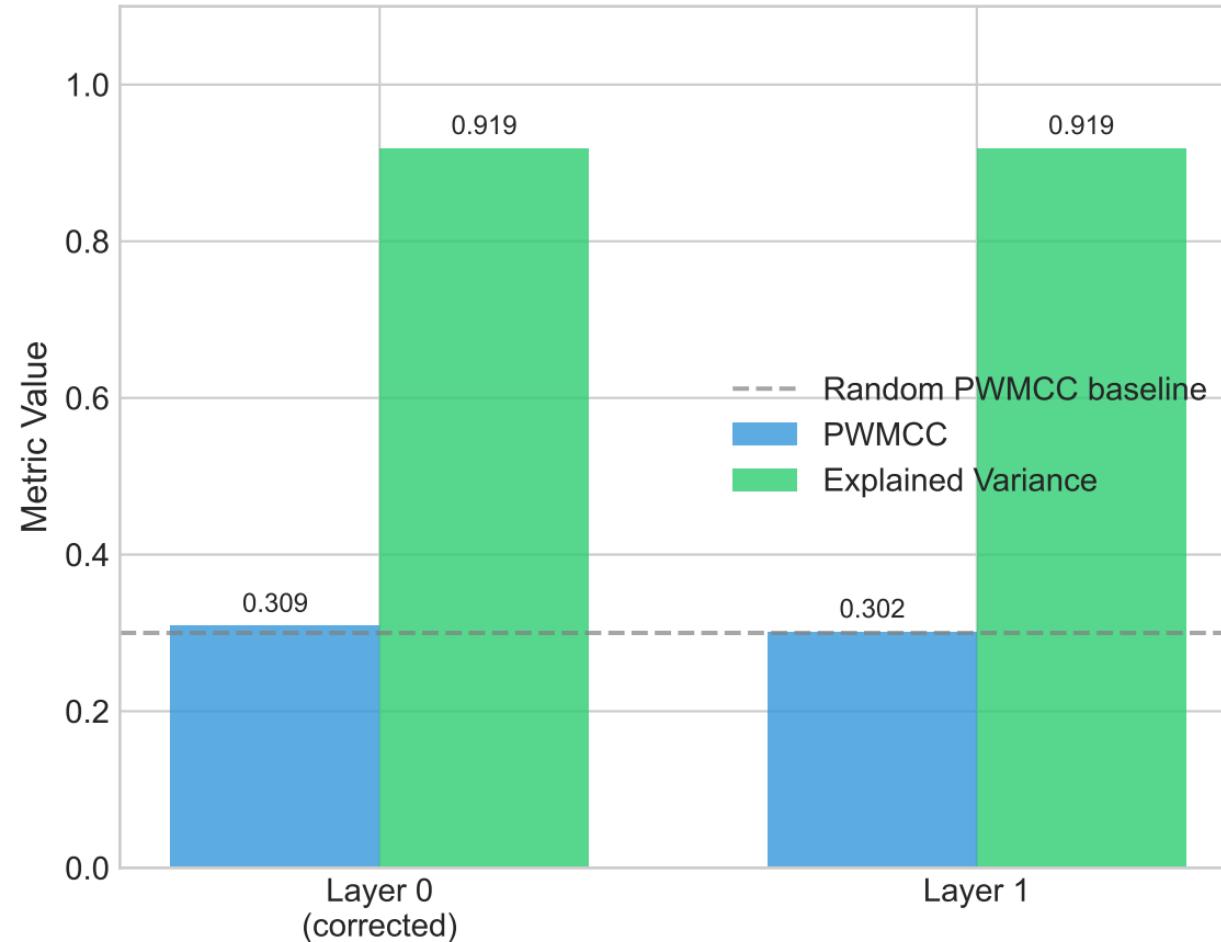


Layer-Dependent Stability



Interpretation

INTERPRETATION:

Layer 0 ($\text{PWMCC} = 0.309$, corrected):

- Same as random baseline
- Original 0.047 was measurement bug
- Decoder-based PWMCC is correct method

Layer 1 ($\text{PWMCC} = 0.302$):

- Same as random baseline
- No consistent feature learning
- Excellent reconstruction

CONCLUSION:

ALL layers show random-baseline stability.
SAE instability is layer-independent.