DIAGNOSING CROSS-VARIATE ACCESS IN TIME SERIES FORECASTING: BENEFITS, EDGE REDUNDANCY, AND ALLOCATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Multivariate time-series forecasting is often framed through channel-independent or channel-dependent architectures, but this model-level distinction does not reveal which variables and forecast objectives benefit from cross-variate access, or what connectivity preserves those benefits. We introduce a controlled diagnostic framework using binary structural masks fixed throughout training and paired retraining to distinguish access benefit, budget sufficiency, and connection allocation. With input and prediction lengths both fixed at 96, we conduct primary experiments on four benchmarks using a compact iTransformer. We find that access benefits vary across datasets, receiving variables, and forecast objectives: aggregate noninferiority after removing cross-variate access can coexist with increased error for individual variables. After separately optimizing Short (steps 1–48) and Far (steps 49–96), we find stronger Short access benefits on Solar and ECL with both compact iTransformer and SAMformer. Positive access benefits also coexist with substantial edge redundancy: on Solar and ECL with compact iTransformer, randomly removing 90% of direct cross-variate edges satisfies a prespecified 1% noninferiority criterion for joint-96 mean squared error. Exploratory Solar sparsity curves further show sharply increasing Short error under extreme sparsification, while Far error remains close to the fully connected baseline. Beyond edge count, allocation also affects performance: at fixed budgets, restoring cross-variate coverage consistently reduces local error for previously isolated receivers across three real-data settings. In controlled synthetic experiments, connections to predictive-parent proxies outperform non-parent controls matched in edge count, receiver degree, and coverage. These findings support evaluating cross-variate interaction jointly in terms of forecast objectives, connection budgets, and allocation, providing empirical guidance for selective interaction mechanisms.

1 INTRODUCTION

Multivariate time-series forecasting must decide how a prediction for one variable should use the histories of other variables. Cross-variate access can provide useful predictive information but can also expose models to spurious correlations. Existing models adopt different designs: channel-independent methods such as PatchTST process variables separately, whereas channel-dependent methods such as iTransformer explicitly model interactions among variables (Nie et al., 2023; Liu et al., 2024). Source selection and variable-subset selection further expand how information is exchanged (Zhao & Shen, 2024; Vareille et al., 2024).

Comparisons between complete models simultaneously involve differences in interaction mechanisms, model capacity, and training behavior, making it difficult to isolate the benefits of cross-variate access and the connection structures needed to retain them. Aggregate error can also conceal differences in benefit across receiving variables and forecast objectives. Under matched conditions, we investigate three questions in sequence: do access benefits vary across datasets, receivers, and forecast objectives? When access is beneficial, can sparse connections preserve predictive performance? At a fixed connection budget, how do receiver coverage and source relevance affect prediction?

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

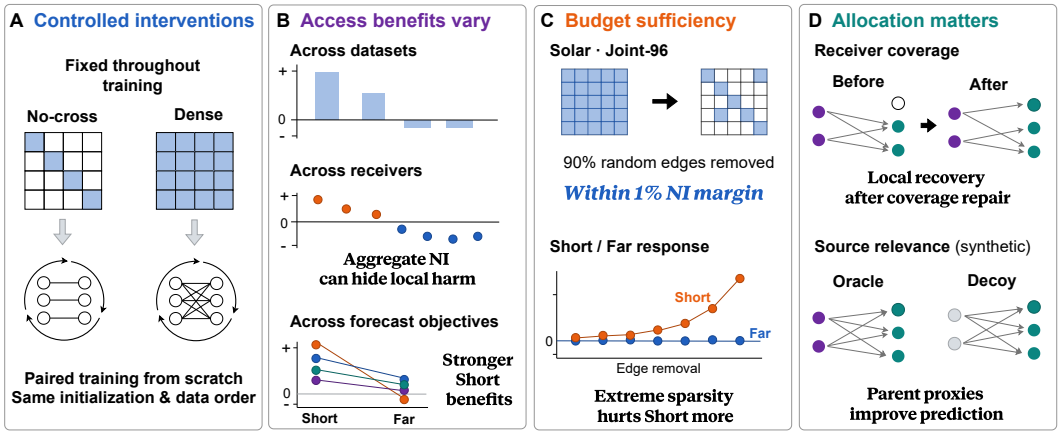


Figure 1: Overview of the diagnostic framework and main findings. (A) Fixed structural masks under paired retraining. (B) Access benefits across datasets, receivers, and forecast objectives. (C) Budget sufficiency and objective-dependent sparsity responses. (D) Receiver-coverage interventions on real data and source-relevance controls in synthetic worlds.

We introduce a controlled diagnostic framework to answer these questions. Within each paired comparison, we hold the backbone and training protocol fixed and control cross-variate access using binary structural masks determined before training and held constant throughout, while preserving self-access. Each structural condition is trained from scratch in paired experimental blocks sharing initialization and data order. We compare fully connected (Dense) and self-only (No-cross) structures, progressively sparsify connections, and reallocate edges at fixed budgets to measure access benefit, budget sufficiency, and allocation effects, respectively. We also optimize near and far forecast objectives separately to test the objective dependence of access benefits (Figure 1).

These controlled experiments yield three main contributions: (1) **Controlled diagnostic framework.** Fixed structural masks, paired retraining, and noninferiority analysis separately test the benefits of cross-variate access, the sufficiency of tested connection budgets, and allocation effects at fixed budgets. (2) **Heterogeneous access benefits and edge redundancy.** We reveal differences in access benefits across datasets, receivers, and forecast objectives. Aggregate noninferiority can conceal local damage, and positive access benefits can coexist with substantial edge redundancy. (3) **Access-allocation diagnostics.** Receiver-coverage interventions on real data and matched-source controls on synthetic data show that coverage repair reduces local error for previously isolated receivers and that predictively relevant sources retain an advantage under matched structural conditions, revealing allocation effects beyond edge count.

2 Related Work

Understanding channel dependence. Prior analyses of channel-independent and channel-dependent forecasting explain their relative performance through both model properties and dataset structure. Han et al. (2024b) characterize a capacity–robustness trade-off: cross-channel modeling offers greater representational capacity but can be less robust under distribution shift. Abdelmalak et al. (2025) show that lookback-window selection and the strength of cross-channel dependencies affect comparative performance. We vary permitted cross-variate routes within a backbone under matched training conditions, distinguishing the benefit of access from the sufficiency of an edge budget.

Partial and selective interaction. Existing selective forecasting mechanisms control communication through masks, structured aggregation, or adaptive filtering. CCM organizes interaction through channel clustering (Chen et al., 2024), while SOFTS aggregates information through a shared series core (Han et al., 2024a). TimeFilter filters spatial–temporal dependencies at the patch level (Hu et al., 2025). Lee et al. (2026) introduce dataset-driven channel masks that combine channel similarity with learnable dataset-specific parameters to modulate attention. We instead fix binary structural constraints from initialization and retrain under each constraint to measure how edge budget and allocation affect error.

Predictive sources and forecast objectives. LIFT exploits local lead–lag relationships to select leading indicators (Zhao & Shen, 2024), while ChronoEpilogi identifies multiple minimal predictive variable subsets for a target (Vareille et al., 2024). These methods seek informative sources; our controlled synthetic contrasts test whether predictive source relevance affects error after matching edge count, receiver degree, and coverage. CATS uses future-position-dependent queries to organize forecasting (Kim et al., 2024). Whereas such queries change the prediction mechanism, we compare dependence on cross-variate access under separately trained near (Short) and far (Far) objectives at a fixed total prediction length. This comparison measures objective-conditioned demand; the masks remain shared across output positions.

3 Experimental Framework

Each experimental condition consists of a binary structural mask fixed before training and a forecasting model trained from scratch under that mask. Within each paired structural comparison, the backbone, forecast objective, and training protocol are held fixed while the permitted cross-variate connections vary. The resulting contrasts compare predictive performance after adaptation to each access structure. For N variables, the rows of $\mathbf{M} \in \{0,1\}^{N\times N}$ index receivers (queries) and the columns index sources (keys).

3.1 Persistent structural interventions

Before the softmax of every layer and attention head, disallowed query–key logits are set to $-\infty$; self-loops are always retained ($M_{ii} = 1$). Dense ($\mathbf{M} = \mathbf{1}$) permits all cross-variate connections; No-cross ($\mathbf{M} = \mathbf{I}$) permits each variable to access only its own history. Masks in the main experiments are shared across layers, heads, and output positions; specific sparsity and allocation structures are introduced with the corresponding experiments.

We operationalize cross-variate access demand as the paired change in test MSE from Dense to No-cross after retraining each condition under its fixed mask. This contrast is specific to the dataset, forecast objective, backbone, and training protocol.

The primary backbone is a compact iTransformer-style model (Liu et al., 2024); we use SAMformer (Ilbert et al., 2024) for prespecified secondary replications on Solar and ECL. Both use variate-token attention. RevIN (Kim et al., 2022) computes statistics separately for each input instance and variable, and all other non-attention modules operate per variable; attention is the only module that mixes observations from different variables. No-cross therefore blocks sample-dependent cross-variate information flow, while model parameters remain shared across variables. Architecture, normalization, and channel-embedding details are given in Appendix A.1.

All main experiments on Solar, ECL, ETTm1, and Exchange use $L = H = 96$, with preprocessing based on training-set statistics only. Sampling frequencies, splits, and physical spans are listed in Appendix A, Table A1; cross-dataset comparisons use the same step-indexed intervals, which correspond to different physical lead times.

3.2 Paired training

Each paired block shares the seed, initialization, and data order; conditions differ only in the structural mask. Training uses validation-MSE early stopping, and the best validation checkpoint is used for test evaluation. Confirmatory analyses must pass prespecified convergence and pairing-integrity checks; execution settings, retention rules, and the audit protocol are given in Appendices A.2–A.4.

In addition to joint-96, we train separately for Short (steps 1–48) and Far (49–96). The model still outputs the full 96 steps, but the training loss, validation selection, test evaluation, and interpretation are restricted to the target band; post hoc segment evaluations of joint-96 models are marked separately.

3.3 Statistical framework

For each paired block b, the log-MSE ratio of condition A relative to comparator B is

$$\delta_b(A, B) = \log \frac{\mathrm{MSE}_b(A)}{\mathrm{MSE}_b(B)}. \tag{1}$$

Dense is the default comparator; experiments that use a different matched structural control say so explicitly. Across blocks we report $100[\exp(\bar{\delta}) - 1]$, where positive values indicate that A has higher MSE. Two-sided 95% Student-t intervals are computed on the paired log-ratios and back-transformed to the percentage scale.

The noninferiority tolerance is a 1% margin frozen before confirmatory runs; a condition is judged noninferior on the current objective only when the one-sided 95% upper bound on the relative MSE change does not exceed 1%. Tests use unrounded values. Sequential tests, multiple-comparison corrections, and small-sample exact tests are given in Appendices A.5–A.6.

4 Cross-Variate Access Benefits Vary across Datasets, Receivers, and Forecast Objectives

4.1 Access benefits differ across datasets

We first test whether cross-variate access improves overall MSE on forecasting tasks across different datasets. For the compact iTransformer joint-96 task, we compare 20 paired Dense–No-cross training blocks for each of four datasets and summarize the geometric mean paired MSE change of No-cross relative to Dense as described in Section 3.3.

Cross-variate access provides clear aggregate benefits on Solar and ECL, whereas ETTm1 and Exchange show no positive access benefit. Removing access increases MSE by $+12.72\%$ on Solar and $+8.63\%$ on ECL. By comparison, No-cross changes MSE relative to Dense by -1.31% on ETTm1 and -1.13% on Exchange, with one-sided 95% upper bounds of -0.87% and -0.51%, respectively; both satisfy the prespecified 1% overall-MSE noninferiority criterion. Full aggregate effects, confidence intervals, and noninferiority decisions are reported in Appendix B.1; model performance references and dataset-specific convergence diagnostics are given in Appendices A.1 and A.3.

Thus, under the current backbone and training protocol, the aggregate benefit of cross-variate access varies across datasets. Aggregate error, however, combines performance across all receivers; determining whether this dataset-level conclusion also applies to individual variables requires further decomposition.

4.2 Aggregate effects conceal receiver-level heterogeneity

Aggregate access benefits are not distributed uniformly across receivers. We conduct a post hoc per-variable decomposition of the existing 20 paired joint-96 Dense–No-cross blocks for each dataset, summarize the paired log-MSE ratios for each variable, and characterize the effect distribution using back-transformed relative changes (Table 1).

Access benefits extend to almost all variables on Solar, whereas aggregate noninferiority on ETTm1 conceals local damage: 3 of 7 variables incur more than 1% higher error under No-cross, including a $+10.40\%$ increase for LULL (variable counts in Table 1; full LULL estimate and calculation in Appendix B.1). Aggregate MSE pools receivers before forming the ratio, so overall noninferiority

Table 1: Receiver counts by geometric mean paired MSE change of No-cross relative to Dense (20 blocks per dataset). The $\pm 1\%$ thresholds are descriptive, not significance or noninferiority decisions.

Dataset	Variables	Damage $> 1\%$	Within $\pm 1\%$	Improvement $> 1\%$
Solar	137	136	0	1
ECL	321	263	29	29
ETTm1	7	3	0	4
Exchange	8	1	5	2

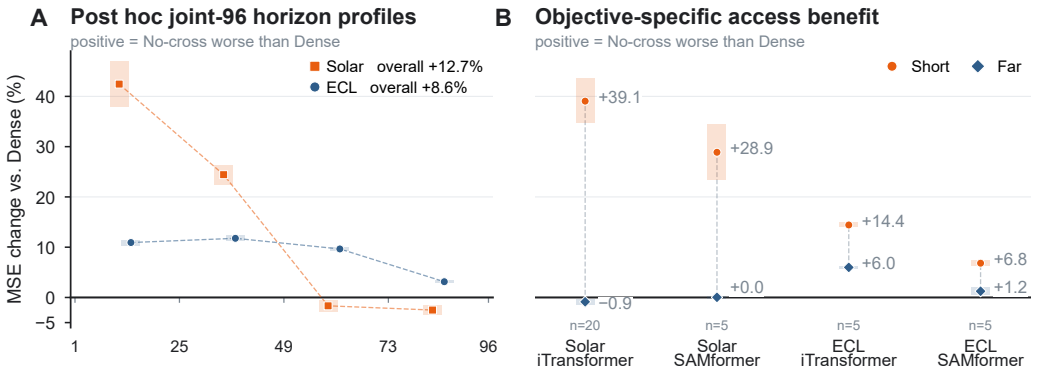


Figure 2: Dataset- and objective-dependent No-cross effects relative to Dense ($L = H = 96$). (A) Post hoc 24-step segment effects from compact iTransformer joint-96 models on Solar and ECL ($n = 20$ paired blocks each). (B) Short/Far effects after objective-specific training, validation selection, and evaluation ($n = 20$ for Solar/compact iTransformer; $n = 5$ per secondary setting). Points and intervals show geometric mean paired MSE changes and two-sided 95% CIs; positive values indicate higher No-cross error.

can conceal local damage. This also occurs under sparse access: Random-90% satisfies the joint-96 criterion on Solar and ECL (Section 5.1), although 2/137 and 35/321 receivers, respectively, have more than 1% higher MSE than paired Dense (descriptive categories; Appendix C.4).

Beyond receivers, overall joint-96 error also combines different forecast positions. We therefore examine whether the benefits of cross-variate access vary with the forecast objective.

4.3 Access benefits vary across forecast objectives

Post hoc segment evaluations of joint-96 models show that aggregate effects conceal differences across forecast intervals (Figure 2A). On Solar, the No-cross MSE change falls from $+42.45\%$ at H1–24 to -2.51% at H73–96, with negative effects in both latter intervals. On ECL, all four segment effects remain positive, but damage falls from $+11.77\%$ at H25–48 to $+3.13\%$ at H73–96. Thus, the aggregate joint-96 effect combines segment effects that differ in magnitude and even direction; full segment estimates are given in Appendix B.2.

Because these segment results come from jointly trained models, differences across intervals may reflect interactions in the joint loss. We therefore use the Solar joint-96 results to define Short (steps 1–48) and Far (49–96), freezing this division before objective-specific retraining. We compare access benefits after separately optimizing the two objectives under the protocol in Section 3.2.

After separate optimization, access benefits remain substantially stronger for Short than for Far (Figure 2B). In the primary 20-block Solar/compact iTransformer comparison, No-cross increases Short MSE by $+39.05\%$, whereas Far MSE changes by -0.86%. The Short–Far interaction, summarized from differences in paired log-MSE ratios, is $+40.27\%$ (95% CI $[+36.12\%, +44.54\%]$; positive in all 20 blocks; definition in Appendix A.7). The persistence of this difference after separate optimization shows that competition between intervals in the joint loss cannot alone explain the near–far contrast.

Three prespecified secondary analyses—ECL/compact iTransformer, Solar/SAMformer, and ECL/SAMformer—replicate this direction, with positive paired interactions in all five blocks of each setting (Figure 2B; full statistics in Appendix B.3). All four tested settings support stronger access benefits on Short than on Far with a 96-step input window, but the sign reversal of the Far effect occurs only on Solar/compact iTransformer.

The benefits of cross-variate access therefore need to be characterized jointly with the dataset, receiver, and forecast objective. For settings with clear benefits, the connection budgets needed to retain those benefits remain to be identified. Section 5 progressively reduces cross-variate connections to test whether the evaluated sparse budgets preserve Dense performance.

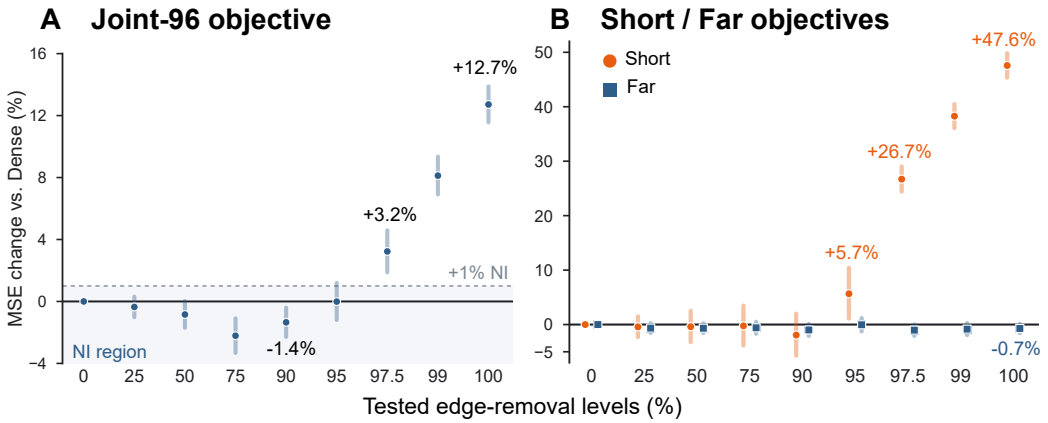


Figure 3: Objective-dependent sensitivity to sparsification on Solar/compact iTransformer ($L = H = 96$). (A) Joint-96 ($n = 20$ paired blocks). (B) Exploratory Short (1–48) and Far (49–96) curves ($n = 5$, seeds 2–6), using the same nested masks across objectives and reusing matching Dense/No-cross models from Figure 2B. Models are trained, selected, and evaluated separately by objective. Points show geometric mean paired MSE changes relative to each objective's Dense baseline; bars show pointwise two-sided 95% CIs. In (A), the dashed line marks the 1% noninferiority margin; ordered tests use one-sided 95% upper bounds (Appendix C.1). Removal levels are equally spaced categories; vertical scales differ.

5 Positive Access Benefit Can Coexist with Substantial Edge Redundancy

5.1 Joint-96 performance remains stable under substantial sparsification

To test whether positive access benefits require dense connectivity, we progressively remove direct cross-variate attention edges on Solar/compact iTransformer and retrain under nested random masks across 20 paired blocks.

After randomly removing 90% of cross-variate edges, overall joint-96 MSE changes by -1.35% relative to Dense (one-sided 95% upper bound -0.56%), satisfying the prespecified 1% noninferiority criterion (Figure 3A). At 95% removal, performance narrowly meets noninferiority in the ordered testing procedure but fails the Holm sensitivity adjustment. Removing 97.5% of edges, leaving only 466 connections, produces significant degradation, with MSE increasing by $+3.23\%$. Testing procedures are given in Appendices A.5–A.6, and full budget results and testing status in Appendix C.1.

At 90% removal, a supplementary ECL/compact iTransformer replication ($n = 10$) yields a -1.10% MSE change relative to Dense (one-sided 95% upper bound -0.565%), meeting the same 1% criterion, while No-cross increases MSE by $+8.57\%$ in these blocks (Appendix C.3). Thus, positive access benefits coexist with substantial direct-edge redundancy on both datasets.

To distinguish direct-edge removal from loss of cross-variate reachability, we audit the fixed masks used in these 90%-removal runs. Every receiver retains at least one direct cross-variate source, and the masks permit paths through one or two attention layers for an average of 76.9% of ordered cross-variate source–receiver pairs on Solar and 96.4% on ECL (Appendix C.4). The noninferior MSE is thus observed in masks with far fewer direct edges but broad remaining structural access.

On Solar, overall joint-96 MSE can still conceal differences across forecast intervals. A post hoc decomposition of trained checkpoints shows that Random-95% already incurs appreciable near-term damage, which is diluted in the aggregate metric by slight improvements farther ahead (Appendix B.2). This motivates a further test: do Short and Far still exhibit different budget–performance relationships after separate optimization?

5.2 Short forecasts are more sensitive to extreme sparsification

We conduct an exploratory budget scan across five paired blocks on Solar/compact iTransformer, with separate training, validation selection, and evaluation for Short and Far. Each sparse condition is retrained from scratch; Dense and No-cross endpoints reuse the models for the corresponding blocks in Section 4.3, characterizing the budget–performance relationships of both objectives within the same paired experiments. Full curves and paired interactions are given in Appendix C.2.

The two curves diverge under extreme sparsification (Figure 3B). Point estimates remain close to Dense at 25–90% removal, but Short's one-sided 95% upper bounds exceed 1% at all four tested levels; this exploratory scan therefore does not establish Short noninferiority at the 1% margin. At 90% removal, the Short estimate is -1.93% with an upper bound of $+1.07\%$. Short error then increases by $+5.66\%$ and $+26.71\%$ at 95% and 97.5% removal, while Far point estimates remain close to Dense. At the No-cross endpoint, Short and Far changes are $+47.58\%$ and -0.72%, respectively.

The sparsity curves characterize budget–performance relationships under the tested random allocations, but the same budget can support different connection structures. The next section fixes the edge count at a joint-96 budget that already exhibits aggregate degradation and tests whether reallocating connections can improve prediction.

6 Beyond Edge Count: Receiver Coverage and Predictive Source Relevance

6.1 Receiver-side allocation affects damage at a fixed budget

On Solar/iTransformer joint-96, randomly removing 97.5% of connections causes aggregate degradation. We therefore fix the remaining budget at 466 edges and compare five structures across 20 paired blocks—Random, Coverage-1, Coverage-2, Query-balanced, and Bi-balanced (full definitions in Appendix D.1)—to test whether changing connection allocation improves performance. Coverage-1 minimally rewires Random: it preserves total edge count and the source-side out-degree sequence, redirecting only surplus edges to receivers whose original cross-variate in-degree is zero so that every receiver obtains at least one cross-variate edge. The other conditions progressively raise or balance receiver degree; Bi-balanced additionally balances source out-degree.

Allocation clearly affects overall error. On joint-96, Random incurs $+3.23\%$ MSE damage relative to Dense; Coverage-1 and Coverage-2 reduce this damage to $+2.67\%$ and $+2.59\%$, and Query-balanced further reduces it to $+2.05\%$ (-1.15% relative to Random, Holm-adjusted $p < 0.001$; Figure 4A). No additional benefit of Bi-balanced over Query-balanced is detected (Appendix D.1). Thus, receiver-side allocation changes aggregate performance at this fixed budget, but balanced allocation remains insufficient to recover Dense performance.

6.2 Zero cross-variate coverage is a reproducible local vulnerability

Aggregate improvement does not reveal which receivers benefit. Coverage-1 performs only the minimal rewiring needed to eliminate zero cross-variate coverage, allowing us to test whether receivers that originally lacked incoming cross-variate edges recover locally. We conduct a post hoc test-set re-evaluation of existing Solar Dense, Random-97.5%, and Coverage-1 checkpoints, dividing variables into original-zero and original-nonzero receivers according to the Random mask before rewiring.

Minimal rewiring reduces error for original-zero receivers: on joint-96, Coverage-1 changes their MSE by -7.26% relative to Random-97.5%, with improvements in all 20 families. Original-nonzero receivers show no clear change, and overall MSE changes by only -0.54% (Figure 4C). In post hoc segment evaluations, both the isolation damage and the recovery from minimal rewiring among original-zero receivers are larger on Short (Figure 4B).

Local recovery replicates in two prespecified Short-only settings: original-zero receiver MSE falls by 18.98% on Solar/SAMformer and 5.47% on ECL/iTransformer, improving in all five families in each setting, while aggregate changes remain close to zero (Figure 4C; Table D3).

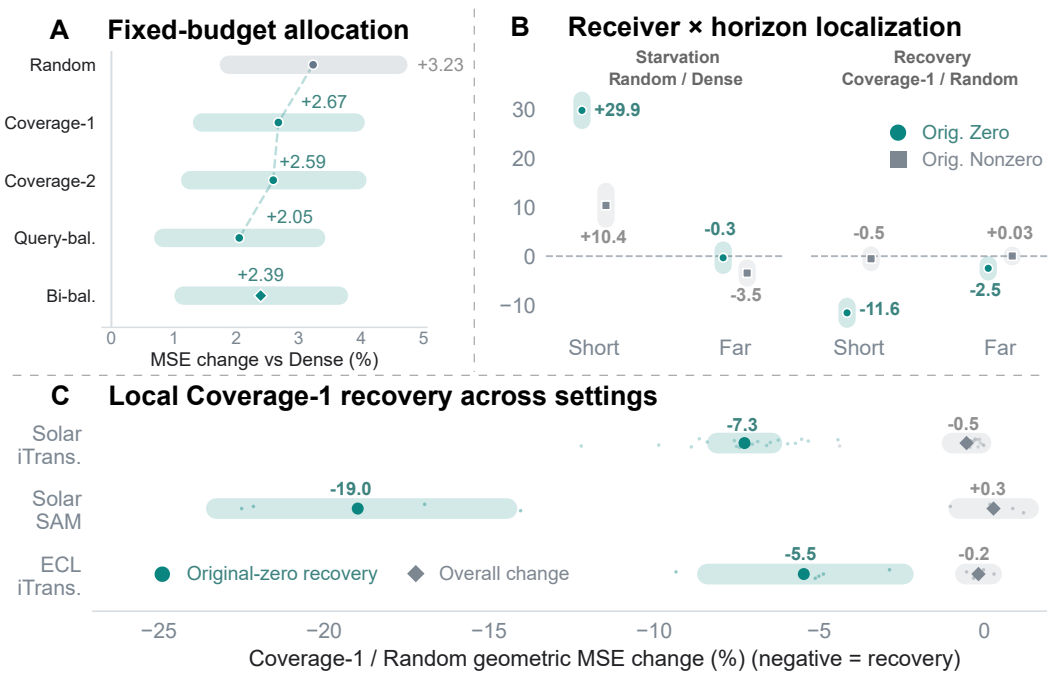


Figure 4: Receiver-side allocation and local recovery. (A) Solar/iTransformer joint-96 effects relative to Dense at 466 cross-variate edges ($n = 20$ paired blocks). (B) Post hoc Short/Far decomposition of these blocks: Random relative to Dense (left), Coverage-1 relative to Random (right); receiver groups are defined by pre-rewiring Random coverage. (C) Coverage-1 relative to Random for original-zero receivers and overall: Solar/iTransformer uses joint-96 ($n = 20$ families); Solar/SAMformer and ECL/iTransformer use Short ($n = 5$ each). Large markers and intervals show geometric mean paired MSE changes and two-sided 95% CIs; small dots in (C) represent individual families. Negative values indicate improvement.

On Solar/SAMformer, original-nonzero receiver MSE increases by $+1.26\%$ relative to Random, illustrating a trade-off between receiver groups at the fixed budget.

A complete lack of incoming cross-variate edges is therefore a reproducible local vulnerability, but does not account for all sparsification damage: original-nonzero receivers on Solar also incur Short damage relative to Dense (Figure 4B), and coverage repair does not restore overall Dense performance (Appendices D.2–D.3). Coverage prevents variables from being isolated, but receivers may still lack valuable predictive information from their directly connected sources. Whether connections to predictively relevant source variables provide further benefits remains a separate question.

6.3 Predictive source relevance contributes beyond matched budget, degree, and coverage

Real data lack directly verifiable predictive-parent labels, preventing attribution of residual damage to particular sources. We therefore use a controlled synthetic source-relevance probe to test whether source identity affects prediction after matching edge count, receiver degree, and coverage. Each synthetic receiver is driven by its own history and lagged contributions from two parent factors, observed through noisy proxy sources (full generative parameters in Appendix D.4).

Because the predictable contributions of the parent factors decay with forecast horizon, we use the H1–12 near-term objective selected during calibration and held fixed in independent test worlds to examine the predictive value of source relevance. The model retains $L = H = 96$; training and validation use all-channel MSE over the first 12 steps, while the primary test metric is receiver MSE over the same interval (Appendix D.4). In five independent test worlds, Oracle connects only to proxy groups of the true predictive-parent factors, whereas Decoy connects only to non-parent proxy groups, in addition to preserved self-access. At every tested degree $k \in \{2, 4, 8, 16\}$, the two conditions exactly match edge budget, receiver degree, and full receiver coverage (Figure 5A).

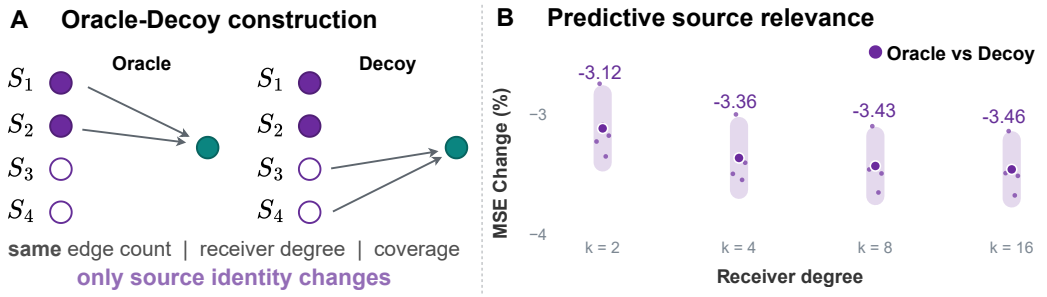


Figure 5: Predictive source relevance in synthetic worlds. (A) Oracle and Decoy connect to parent-factor and non-parent proxy groups, respectively, matching edge count, receiver degree, and coverage. (B) Oracle relative to Decoy in H1–12 receiver MSE across five test worlds. Large markers and intervals show geometric mean paired changes and pointwise two-sided 95% CIs at each degree; small dots represent individual worlds. Negative values favor Oracle.

Oracle outperforms Decoy at all four tested degrees. We first average the four Oracle–Decoy log-MSE ratios within each world, then summarize with the world as the independent statistical unit. Oracle reduces receiver MSE by 3.34% relative to Decoy, with improvement in all five worlds (one-sided exact $p = 0.03125$; Figure 5B). Thus, in this controlled DGP, source relevance provides predictive gains after matching budget, receiver degree, and coverage.

The real-data experiments show that receiver coverage affects local error, while the synthetic contrast shows that source relevance can further affect performance beyond structural budget and coverage. Edge count alone is therefore insufficient to characterize the utility of cross-variate access.

7 Conclusion and Discussion

Persistent structural interventions and paired retraining distinguish access benefit, tested-budget sufficiency, and allocation. Access benefits vary across datasets, receivers, and forecast objectives. Positive benefits can coexist with substantial direct-edge redundancy, and aggregate noninferiority can conceal local damage. At fixed budgets, coverage repair reduces local error on real data, while predictively relevant sources retain an advantage under matched edge count, receiver degree, and coverage in synthetic worlds.

The Short–Far contrast reflects demand under a fixed input window. Forecast position also changes the visibility of same-phase seasonal references, so the contrast cannot isolate the effect of forecast distance (Appendices E.1–E.3). Prespecified interventions on input-window length could test whether demand changes track the visibility boundary of seasonal references.

Stronger Short than Far access benefits after separate training, local recovery after coverage repair, and joint-96 noninferiority at 90% removal replicate across settings. Full budget curves, Far improvements from No-cross, and Query-balanced gains come from Solar/compact iTransformer. In this setting, Short–Far sparsity divergence is exploratory (five paired blocks), and stronger Short local coverage-repair gains are post hoc. Direct source-relevance evidence comes from synthetic worlds with known parent factors.

The tested backbones use variate-token attention and masks shared across output positions. These experiments identify objective-conditioned demand without implementing horizon-specific routing; masking dense logits provides no sparse-kernel speedup.

For selective interaction design, these findings suggest calibrating connection budgets on validation data for the target forecast objective and evaluating receiver-level error alongside aggregate performance. At a fixed budget, allocation strategies should account for receiver coverage and predictive source relevance. These criteria provide an empirical basis for deciding how much cross-variate access to retain and which receivers and sources to connect.

AI USE STATEMENT

We used generative AI tools to discover and organize related work; develop research questions and experimental protocols; assist with experiment implementation and result analysis; and draft sections, translate and polish text, and format the LaTeX manuscript. The authors reviewed AI-assisted references, code, analyses, and text and take responsibility for all results, claims, citations, and content in the paper.

ETHICS STATEMENT

This work uses publicly available multivariate forecasting benchmarks and controlled synthetic time series. It does not involve human subjects, private data, or deployed decision systems.

REPRODUCIBILITY STATEMENT

Persistent masks, paired-block seeds, convergence gates, and estimand definitions are specified in Section 3 and in Appendix A. Dataset splits and physical spans are listed in Table A1 in Appendix A. Evidence grades (confirmatory, prespecified secondary, post hoc, exploratory) are stated with the corresponding results. Implementation details, statistical formulas, and additional diagnostics referenced in the main text are provided in the appendix.

REFERENCES

Ibram Abdelmalak, Kiran Madhusudhanan, Jungmin Choi, Christian Klötergens, Vijaya Krishna Yalavarthi, Maximilian Stubbemann, and Lars Schmidt-Thieme. Channel dependence, limited lookback windows, and the simplicity of datasets: How biased is time series forecasting? *arXiv preprint arXiv:2502.09683*, 2025.

Jialin Chen, Jan Eric Lenssen, Aosong Feng, Weihua Hu, Matthias Fey, Leandros Tassiulas, Jure Leskovec, and Rex Ying. From similarity to superiority: Channel clustering for time series forecasting. In *Advances in Neural Information Processing Systems*, 2024.

Pierre Foret, Ariel Kleiner, Hossein Mobahi, and Behnam Neyshabur. Sharpness-aware minimization for efficiently improving generalization. In *International Conference on Learning Representations*, 2021.

Lu Han, Xu-Yang Chen, Han-Jia Ye, and De-Chuan Zhan. SOFTS: Efficient multivariate time series forecasting with series-core fusion. In *Advances in Neural Information Processing Systems*, 2024a.

Lu Han, Han-Jia Ye, and De-Chuan Zhan. The capacity and robustness trade-off: Revisiting the channel independent strategy for multivariate time series forecasting. *IEEE Transactions on Knowledge and Data Engineering*, 36(11):7129–7142, 2024b. doi: 10.1109/TKDE.2024.3400008.

Yifan Hu, Guibin Zhang, Peiyuan Liu, Disen Lan, Naiqi Li, Dawei Cheng, Tao Dai, Shu-Tao Xia, and Shirui Pan. TimeFilter: Patch-specific spatial-temporal graph filtration for time series forecasting. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 24893–24911. PMLR, 2025.

Romain Ilbert, Ambroise Odonnat, Vasilii Feofanov, Aladin Virmaux, Guillaume Paolo, Themis Palpanas, and Ievgen Redko. SAMformer: Unlocking the potential of transformers in time series forecasting with sharpness-aware minimization and channel-wise attention. In *Proceedings of the 41st International Conference on Machine Learning*, 2024.

Dongbin Kim, Jinseong Park, Jaewook Lee, and Hoki Kim. Are self-attentions effective for time series forecasting? In *Advances in Neural Information Processing Systems*, 2024.

Taesung Kim, Jinhee Kim, Yunwon Tae, Cheonbok Park, Jang-Ho Choi, and Jaegul Choo. Reversible instance normalization for accurate time-series forecasting against distribution shift. In *International Conference on Learning Representations*, 2022.

Seunghan Lee, Taeyoung Park, and Kibok Lee. Dataset-driven channel masks in transformers for multivariate time series. In *ICASSP 2026 – IEEE International Conference on Acoustics, Speech and Signal Processing*, 2026.

Yong Liu, Tengge Hu, Haoran Zhang, Haixu Wu, Shiyu Wang, Lintao Ma, and Mingsheng Long. iTransformer: Inverted transformers are effective for time series forecasting. In *International Conference on Learning Representations*, 2024.

Yuqi Nie, Nam H. Nguyen, Phanwadee Sinthong, and Jayant Kalagnanam. A time series is worth 64 words: Long-term forecasting with transformers. In *International Conference on Learning Representations*, 2023.

Etienne Vareille, Michele Linardi, Ioannis Tsamardinos, and Vassilis Christophides. ChronoEpilogi: Scalable time-series variable selection with multiple solutions. In *Advances in Neural Information Processing Systems*, 2024.

Lifan Zhao and Yanyan Shen. Rethinking channel dependence for multivariate time series forecasting: Learning from leading indicators. In *International Conference on Learning Representations*, 2024.

The appendices follow the order of main-text Sections 3–7.

A Experimental Framework and Statistical Protocols

This appendix supports Section 3 with dataset settings, model configurations, training and integrity protocols, and statistical estimands. Experiment-specific configuration files freeze all numeric thresholds; the rules below govern the retention or exclusion of confirmatory batches.

Table A1: Sampling frequencies, splits, and physical spans for $L = H = 96$ (Liu et al., 2024, Appendix A.1). Splits are chronological; ETTm1 uses the official split.

Dataset	Variates	Sampling	Split	Input span ($L = 96$)	Short (1–48)	Far (49–96)
Solar	137	10 min	70/10/20	16 h	10 min–8 h	8 h 10 min–16 h
ECL	321	1 h	70/10/20	4 d	1–48 h	49–96 h
ETTm1	7	15 min	12/4/4 mo.	24 h	15 min–12 h	12 h 15 min–24 h
Exchange	8	1 d	70/10/20	96 d	1–48 d	49–96 d

A.1 Model Architectures and Configurations

A.1.1 Compact iTransformer-Style Backbone

The primary backbone is a compact iTransformer-style encoder (Liu et al., 2024) modified to include learned channel-identity embeddings:

- **Encoder layers**: 2
- **Hidden dimension** (d_{model}): 128
- **Attention heads**: 8
- **Feed-forward dimension** (d_{ff}): 256
- **Dropout rate**: 0.1
- **Normalization**: Reversible Instance Normalization (RevIN; Kim et al., 2022) applied per-variate without affine parameters.
- **Channel-identity embedding**: A learned d_{model}-dimensional vector added to each variate token. This embedding allows the model to distinguish channel identities but carries no contemporaneous cross-variate observation data.
- **Attention mechanism**: Variate-token attention where tokens represent individual variates across the entire lookback window $L = 96$.
- **Masking implementation**: For a structural mask $\mathbf{M} \in \{0, 1\}^{N \times N}$, cross-variate logits where $M_{ij} = 0$ are set to $-\infty$ via PyTorch `masked_fill(-inf)` before the softmax operator. Diagonal elements are always preserved ($M_{ii} = 1$), ensuring every receiver maintains self-attention to its own historical token. Structural masks are shared across all layers and attention heads in main experiments.

A.1.2 SAMFormer Backbone

Secondary replications on Solar and ECL use a local PyTorch adaptation of SAMformer (Ilbert et al., 2024; official repository):

- **Architecture**: RevIN, single channel-wise attention layer, linear forecast projection head.
- **Hidden dimension**: 16 in both Solar and ECL experiments.
- **Optimization**: Sharpness-Aware Minimization (SAM; Foret et al., 2021) with base optimizer Adam, constant learning rate 10^{-3}, weight decay 10^{-5}, and SAM perturbation radius $\rho_{\text{SAM}} = 0.5$.
- **Batch size**: 256 for Solar band-specific and Coverage-1 experiments; 32 for ECL band-specific experiments.

- **Early stopping**: Patience 50 and `min_delta` $= 0$; the epoch caps and Solar convergence extensions are listed in Table A3.
- **Masking implementation**: Identical logits-level `masked_fill(-inf)` applied to the channel-attention module before softmax.

A.1.3 Optimization Hyperparameters

The compact iTransformer runs use the following optimization settings; SAMformer uses the settings specified above:

- **Optimizer**: Adam ($\beta_1 = 0.9, \beta_2 = 0.999, \epsilon = 10^{-8}$)
- **Learning rate**: $\eta = 10^{-4}$
- **Weight decay**: $\lambda = 10^{-5}$
- **Batch size**: 32
- **Gradient clipping**: Maximum L_2 gradient norm 1.0
- **Loss function**: Mean Squared Error (MSE) computed over the targeted forecast positions ($H = 96$ in joint experiments; $h \in [1, 48]$ in Short; $h \in [49, 96]$ in Far).

A.1.4 Compact-Backbone Performance Reference

The compact Dense values average 20 paired blocks (seeds 2–21). The original iTransformer results come from Liu et al. (2024, Table 10) and provide a published performance reference. All cross-variate access and sparsity effects in this paper use the compact model's paired Dense runs as their comparator.

Table A2: Joint-96 test MSE ($L = H = 96$).

Dataset	Compact	Original
Solar	0.1992	0.203
ECL	0.1411	0.148
ETTm1	0.3298	0.334
Exchange	0.0887	0.086

A.2 Deterministic Execution

Runs configure `CUBLAS_WORKSPACE_CONFIG=:4096:8` and enable PyTorch deterministic algorithms where supported.

A.3 Convergence Gates and Retention Rules

Reported batches use the setting-specific convergence configurations below; confirmatory batches must satisfy their applicable prespecified gates.

Table A3: Training budgets and early-stopping configurations by setting.

Setting / Backbone	Safety Cap (Epochs)	Patience	Early-Stopping Rule
Solar / compact iTransformer (joint, nested, allocation, band-specific)	50	5	`min_delta` $= 10^{-4}$
ECL / compact iTransformer (joint-96 and band-specific)	150	5	`min_delta` $= 10^{-4}$
ETTm1, Exchange / compact iTransformer	10	3	`min_delta` $= 10^{-4}$
Solar / SAMformer (band-specific)	400	50	`min_delta` $= 0$
Solar / SAMformer (Coverage-1)	400	50	`min_delta` $= 0$
ECL / SAMformer (band-specific)	400	50	`min_delta` $= 0$
Synthetic factor–proxy V5 (Oracle/Decoy)	60	8	`min_delta` $= 10^{-4}$

The Solar Short/No-cross runs at seeds 0 and 2 use the 600-epoch convergence-extension results; the other SAMformer endpoints use the 400-epoch configurations.

A.3.1 PRESPECIFIED GATE AND RETENTION CRITERIA

For the 20-block Solar nested-sparsity and band-specific batches:

- **Window criterion**: At most 3 of 20 runs may select a best epoch located within the final 2-epoch safety window.
- **Safety-cap criterion**: At most 9 of 20 runs may hit the safety cap without triggering early stopping.

For the ETTm1 joint-96 endpoint, cap attainment is evaluated together with best-checkpoint concentration: 13/20 Dense runs reached the 10-epoch cap, while 2/20 selected the final epoch as the best validation checkpoint; the batch was retained.

The reported ECL joint-96 endpoint uses the 150-epoch configuration on seeds 2–21; all 40 Dense and No-cross runs early-stopped before the cap. The supplementary Random-90% runs on seeds 2–11 also early-stop before the cap (epochs 51–76; Appendix C.3).

A.4 INTEGRITY CHECKS AND SHA-256 AUDIT

Before any paired block is entered into confirmatory analysis, an automated integrity audit verifies the following invariants:

1. **Initialization hash**: SHA-256 of initial parameter tensors matches across paired conditions.
2. **Data-order hash**: SHA-256 digest of sample index sequences in data loaders matches across conditions.
3. **Dataset split fingerprint**: Verification that train/val/test split boundaries match benchmark definitions.
4. **Mask hash**: SHA-256 of the binary mask tensor is verified at instantiation and re-verified upon evaluation.
5. **Checkpoint reload hash**: Verification that reloaded model weights match the saved validation checkpoint.

Blocks failing any integrity check are automatically excluded from analysis.

A.5 STATISTICAL TESTING AND ERROR AGGREGATION

A.5.1 PAIRED LOG-MSE RATIO

For each paired block $b \in \{1, \ldots, n\}$, the performance of candidate condition A relative to comparator B (where $B =$ Dense by default) is defined as:

$$\delta_b(A, B) = \log \frac{\mathrm{MSE}_b(A)}{\mathrm{MSE}_b(B)}.$$

Across n paired blocks, the geometric mean percentage change is computed as:

$$\Delta\mathrm{MSE}(A, B) = 100 \cdot \left[\exp\left(\frac{1}{n} \sum_{b=1}^{n} \delta_b(A, B) \right) - 1 \right] = 100 \cdot [\exp(\bar{\delta}) - 1].$$

A positive percentage indicates higher test error (performance degradation) under condition A relative to B.

A.5.2 CONFIDENCE INTERVALS

Two-sided 95% confidence intervals are constructed in the log-ratio space using the Student-t distribution:

$$\left[\bar{\delta} - t_{0.975,n-1}\frac{s}{\sqrt{n}},\ \bar{\delta} + t_{0.975,n-1}\frac{s}{\sqrt{n}}\right],$$

where s is the sample standard deviation of $\{\delta_b\}_{b=1}^{n}$. The interval endpoints are then transformed back to the percentage scale via $100[\exp(\cdot) - 1]$.

A.5.3 Multiple Testing and Exact Permutation Tests

- **Ordered Sequential Testing**: For the nested sparsity sequence ($25\% \to 50\% \to 75\% \to 90\% \to 95\% \to 97.5\% \to 99\% \to 100\%$), non-inferiority is evaluated in a fixed prespecified sequence at $\alpha = 0.05$. Formal non-inferiority claims terminate at the first failed level; subsequent levels are reported as descriptive.
- **Holm-Bonferroni Correction**: Prespecified non-nested comparison families apply Holm step-down adjustments; descriptive contrasts with reported p-values use nominal values.
- **Exact Sign-Flip Permutation Test**: In 5-pair replication experiments, when all 5 pairs show directional concordance ($5/5$ positive or $5/5$ negative), the exact one-sided permutation p-value is:

$$p = \left(\frac{1}{2}\right)^5 = \frac{1}{32} = 0.03125.$$

A.6 One-Percent Non-Inferiority Margin

The non-inferiority margin $\epsilon = 1.0\%$ is a study-specific operational threshold frozen prior to confirmatory execution. A candidate condition A is considered non-inferior to comparator B if and only if the upper bound of the one-sided 95% confidence interval for relative MSE change satisfies:

$$\Delta\text{MSE}_{\text{upper}} = 100 \cdot \left[\exp\left(\bar{\delta} + t_{0.95,n-1} \cdot \frac{s}{\sqrt{n}}\right) - 1\right] \leq +1.0\%.$$

On the Solar Dense anchor (MSE ≈ 0.199), a 1% relative threshold corresponds to an absolute MSE increment of approximately $+0.00199 \approx 0.002$.

A.7 Band-Interaction Estimand

To compare cross-variate demand between Short ($h \in [1, 48]$) and Far ($h \in [49, 96]$) forecast objectives, we define the paired band-interaction estimand I_b for each paired block b:

$$I_b = \delta_b^{\text{Short}} - \delta_b^{\text{Far}}, \qquad \text{where} \quad \delta_b^{\text{band}} = \log \frac{\text{MSE}_b^{\text{band}}(\text{No-cross})}{\text{MSE}_b^{\text{band}}(\text{Dense})}.$$

$\text{MSE}_b^{\text{Short}}$ and $\text{MSE}_b^{\text{Far}}$ come from models trained and selected independently for the respective band objective, rather than from out-of-band predictions of a single joint model. The summary interaction is:

$$\Delta\text{Interaction} = 100 \cdot [\exp(\bar{I}) - 1] = 100 \cdot \left[\frac{1 + \Delta\text{MSE}_{\text{Short}}/100}{1 + \Delta\text{MSE}_{\text{Far}}/100} - 1\right].$$

This ratio-of-ratios formulation preserves algebraic closure.

B Cross-Variate Access-Benefit Results

Following Section 4, this appendix reports dataset-level access effects, post hoc horizon decompositions, and separately trained Short/Far comparisons. The segment table also supplies the Random-95% diagnostic used in Section 5.1.

B.1 JOINT-96 DATASET ENDPOINTS

Table B1 supplies the full uncertainty estimates for the dataset-level results in Section 4.1.

Table B1: Joint-96 No-cross versus Dense on the compact iTransformer.

Dataset	Blocks	Change [95% CI] (%)	One-sided 95% upper (%)	1% NI criterion
Solar	20	+12.716 [+11.562, +13.882]	+13.678	Not met
ECL	20	+8.633 [+8.282, +8.985]	+8.924	Not met
ETTm1	20	-1.305 [-1.829, -0.778]	-0.870	Met
Exchange	20	-1.129 [-1.877, -0.374]	-0.506	Met

All changes and bounds are percentages. Brackets are two-sided 95% confidence intervals; the separate upper-bound column is the one-sided 95% bound used for the 1% noninferiority criterion. Decisions use unrounded values. All settings use 20 paired blocks (seeds 2–21); ECL uses a 150-epoch convergence extension following earlier epoch-cap diagnostics. "Not met" indicates failure to establish noninferiority under the stated criterion.

B.1.1 RECEIVER-LEVEL EXAMPLE: ETTM1 LULL

The LULL example in Section 4.2 uses the same 20 joint-96 Dense–No-cross blocks (seeds 2–21), with MSE evaluated on this receiver alone. For each seed b, we compute $\delta_{b,\mathrm{LULL}} = \log[\mathrm{MSE}_{b,\mathrm{LULL}}(\text{No-cross})/\,\mathrm{MSE}_{b,\mathrm{LULL}}(\text{Dense})]$ and apply the geometric summary and Student-t interval in Appendix A.5. The resulting No-cross change is **+10.404% [95% CI: +9.520%, +11.295%]**, with higher error in **20/20 blocks**. This rounds to the +10.40% reported in the main text. Across all seven ETTm1 receivers, three geometric effects exceed +1% and four are below -1%, reproducing the descriptive counts in Table 1.

This receiver analysis is post hoc; the interval is pointwise and unadjusted across receivers.

B.2 POST HOC JOINT-96 SEGMENT EFFECTS

The 24-step estimates below are post hoc evaluations of models trained on the joint-96 objective. They motivate band-specific retraining in Section 4.3; they do not identify objective-conditioned demand. The Solar Random-95% column also supports the segment diagnostic in Section 5.1.

Table B2: Post hoc 24-step segment effects relative to Dense for joint-96 models (20 paired blocks).

Segment	Solar No-cross	ECL No-cross	Solar Random-95%
1-24	+42.446 [+38.027, +47.006]	+10.925 [+10.367, +11.486]	+3.377 [-0.559, +7.468]
25-48	+24.426 [+22.499, +26.384]	+11.766 [+11.189, +12.346]	+3.259 [+1.102, +5.461]
49-72	-1.662 [-2.812, -0.498]	+9.651 [+9.272, +10.030]	-3.607 [-4.731, -2.470]
73-96	-2.506 [-3.418, -1.585]	+3.134 [+2.918, +3.350]	-1.818 [-2.769, -0.858]
Overall	+12.716 [+11.562, +13.882]	+8.633 [+8.282, +8.985]	-0.007 [-1.199, +1.199]

All values are percentages, with two-sided 95% CIs in brackets. These segment evaluations do not involve band-specific retraining. Solar Random-95% has positive first-half point estimates and negative second-half estimates; its H1–24 interval crosses zero. The joint-96 result averages MSE before taking the relative change and is not the arithmetic mean of segment percentage changes.

B.3 BAND-SPECIFIC SHORT/FAR TRANSFER ESTIMATES

Table B3 reports the complete band-specific No-cross versus Dense estimates and paired interactions summarized in Section 4.3. The interaction is defined in Appendix A.7; five-block extensions use an exact one-sided sign-flip test.

Table B3: Separately trained Short/Far No-cross effects and their within-block interaction; all effects and two-sided 95% CIs are percentages.

Setting	Blocks	Short [95% CI]	Far [95% CI]	Interaction [95% CI]	Positive interactions	Exact one-sided p
Solar / iTrans.	20	+39.055 [+34.745, +43.503]	-0.865 [-1.391, -0.335]	+40.268 [+36.119, +44.544]	20/20	—
ECL / iTrans.	5	+14.417 [+14.007, +14.829]	+5.986 [+5.699, +6.274]	+7.955 [+7.467, +8.446]	5/5	0.03125
Solar / SAM.	5	+28.881 [+23.479, +34.519]	+0.035 [-0.170, +0.241]	+28.835 [+23.497, +34.405]	5/5	0.03125
ECL / SAM.	5	+6.818 [+6.268, +7.371]	+1.238 [+0.519, +1.963]	+5.511 [+5.126, +5.898]	5/5	0.03125

The interaction is formed by subtracting paired log ratios before aggregation (Appendix A.7), rather than subtracting the displayed percentages. The exact one-sided sign-flip p-value is 0.03125 in each five-block extension. Solar/iTransformer uses seeds 2–21; the other settings use seeds 0–4. For Solar, Far was designated the primary objective and Short the secondary objective before retraining. In ECL/iTransformer, seeds 0 and 1 had also been used for earlier epoch-cap diagnostics.

C Connection Budgets and Edge Redundancy

This appendix supports Section 5 with the Solar budget curves and allocation contrast, followed by the ECL replication and post hoc diagnostics at 90% edge removal.

C.1 Joint-96 Random-Sparsity Curve

Table C1 provides the joint-96 random-budget results for Section 5.1. Structural masks retain all self edges. The full cross-variate budget is 18,632 edges; reported removal percentages label tested conditions, while the retained-edge column gives the actual integer budget.

Table C1: Solar/compact iTransformer joint-96 nested-random curve (20 paired seed/mask-family blocks, seeds 2–21).

Removed (%)	Edges	Change [95% CI] (%)	One-sided upper (%)	Nominal NI p	Ordered result	Holm p
0	18632	0.000 [0.000, 0.000]	—	—	Reference	—
25	13974	-0.355 [-1.016, +0.311]	+0.194	0.00022	Pass	0.00109
50	9316	-0.843 [-1.704, +0.026]	-0.125	0.00015	Pass	0.00089
75	4658	-2.213 [-3.324, -1.089]	-1.285	< 0.0001	Pass	< 0.0001
90	1863	-1.347 [-2.290, -0.395]	-0.561	< 0.0001	Pass	0.00021
95	932	-0.007 [-1.199, +1.199]	+0.989	0.04824	Pass	0.19298
97.5	466	+3.228 [+1.881, +4.592]	+4.354	0.99874	First failure	1.00000
99	186	+8.122 [+6.903, +9.356]	+9.140	1.00000	Not tested	1.00000
100	0	+12.716 [+11.562, +13.882]	+13.678	1.00000	Not tested	1.00000

All effects and bounds are percentages. Fixed-sequence NI tests use the frozen order 25%, 50%, 75%, 90%, 95%, 97.5%, 99%, 100%, stopping at the first failure (97.5%). Rows at 99% and 100% are descriptive after that stop; nominal p-values and pointwise intervals are still shown for completeness. Holm p-values adjust all eight non-Dense comparisons as a sensitivity analysis. The 95% removal condition passes the ordered test (unrounded upper bound 0.988702%) but not the Holm sensitivity test (p = 0.192978). A zero change at Dense is an identity, not an estimated noninferiority result.

C.2 Objective-Specific Random-Sparsity Curves

The following two tables support the exploratory budget scan in Section 5.2, reporting separately trained Short/Far curves and their within-block interactions. All structural masks retain self edges.

Table C2: Exploratory separately trained Short (H1–48) and Far (H49–96) curves on the same five seeds (2–6).

Removed (%)	Edges	Short [95% CI]	Short upper	Far [95% CI]	Far upper
0	18632	0.000 [0.000, 0.000]	—	0.000 [0.000, 0.000]	—
25	13974	-0.428 [-2.309, +1.489]	+1.041	-0.654 [-1.546, +0.246]	+0.036
50	9316	-0.399 [-3.239, +2.525]	+1.838	-0.677 [-1.556, +0.211]	+0.004
75	4658	-0.245 [-3.833, +3.477]	+2.601	-0.594 [-1.680, +0.505]	+0.248
90	1863	-1.930 [-5.700, +1.992]	+1.068	-0.988 [-2.081, +0.118]	-0.140
95	932	+5.659 [+1.117, +10.405]	+9.284	-0.037 [-1.246, +1.187]	+0.901
97.5	466	+26.707 [+24.421, +29.035]	+28.491	-1.029 [-2.072, +0.026]	-0.220
99	186	+38.270 [+36.083, +40.491]	+39.972	-0.841 [-1.942, +0.272]	+0.013
100	0	+47.575 [+45.357, +49.827]	+49.301	-0.718 [-1.570, +0.141]	-0.059

Values and confidence limits are percentages. Each objective is compared with its own paired Dense baseline. The 70 intermediate runs use the same nested masks across objectives; Dense and No-cross endpoints reuse the matching seeds from the 20-block experiment. The sparsity grid was informed by the earlier seed-42 pilot, which is excluded. Brackets are pointwise two-sided 95% CIs; upper bounds are one-sided 95%. These are exploratory summaries, without a confirmatory fixed-sequence or multiplicity-adjusted NI claim. Short at 90% removal has upper bound 1.067620%, above the 1% margin. The No-cross effects differ from Table B3 because only seeds 2–6 are used here.

Table C3: Exploratory paired Short–Far interactions for Table C2.

Removed (%)	Interaction [95% CI] (%)	Positive blocks
0	0.000 [0.000, 0.000]	—
25	+0.228 [-1.993, +2.499]	4/5
50	+0.280 [-2.839, +3.499]	4/5
75	+0.351 [-3.891, +4.781]	4/5
90	-0.951 [-4.838, +3.095]	1/5
95	+5.698 [+0.086, +11.625]	5/5
97.5	+28.024 [+25.718, +30.373]	5/5
99	+39.443 [+37.084, +41.842]	5/5
100	+48.642 [+45.706, +51.638]	5/5

For each seed and removal level, the Far log-MSE ratio is subtracted from the Short log-MSE ratio, followed by the same geometric summary and pointwise two-sided 95% t interval. Values are percentages, not percentage-point differences of the displayed effects. Positive values indicate greater relative damage on Short. Intervals are descriptive and unadjusted across tested levels; the grid does not identify a continuous sparsity threshold.

C.3 ECL Joint-96 Replication at 90% Edge Removal

This supplementary replication evaluates ECL/compact iTransformer joint-96 ($L = H = 96$) using 10 paired blocks, seeds 2–11; the seed-21 pilot is excluded. Each independently generated Random-90% mask retains 10,272 of the 102,720 cross-variate edges, with all self edges preserved. Comparisons use matching Dense and No-cross conditions on these seeds. This ten-seed subset overlaps the 20-block endpoint analysis in Table B1.

Table C4: ECL joint-96 test MSE by seed. SD is the sample standard deviation across the ten blocks.

Seed	Dense MSE	No-cross MSE	Random-90% MSE
2	0.140340	0.154788	0.141058
3	0.141953	0.154860	0.139241
4	0.142479	0.153004	0.139327
5	0.142073	0.152902	0.141750
6	0.141136	0.152662	0.139436
7	0.142010	0.153399	0.139120
8	0.139527	0.153464	0.139419
9	0.141959	0.152840	0.139519
10	0.140576	0.153380	0.138950
11	0.141190	0.152992	0.139900
Mean	0.141324	0.153429	0.139772
SD	0.000941	0.000781	0.000911

Using paired log ratios and the Student-t bounds in Appendix A.5, No-cross increases MSE by $+8.566\%$ relative to Dense (one-sided 95% lower bound $+7.984\%$), with higher error in all 10 blocks (one-sided exact sign-flip $p = 1/1024 \approx 0.00098$). Random-90% changes MSE by -1.098% (one-sided 95% upper bound -0.565%), satisfying the prespecified $+1\%$ noninferiority margin; its error is lower than Dense in 9/10 blocks. The No-cross effect differs from Table B1 because this analysis uses seeds 2–11 rather than 2–21. This replication tests a single budget point; the full sparsity curves remain Solar analyses.

All sparse runs select their best validation checkpoint at epochs 46–71 and early-stop at epochs 51–76, below the 150-epoch cap. Reported pairing checks match the dataset, split, data order, and initialization across conditions. Initialization and configuration were additionally checked by independent reconstruction.

C.4 Post Hoc Local Effects and Structural Reachability at 90% Removal

We decompose the existing joint-96 Random-90% and paired Dense runs on Solar (seeds 2–21, $n = 20$) and ECL (seeds 2–11, $n = 10$). For each block and forecast interval, we average horizon MSE before forming the paired log-MSE ratio. Table C5 reports geometric mean changes and pointwise two-sided 95% t intervals. These interval effects come from jointly trained models, not separately optimized Short/Far models.

Table C5: Post hoc Random-90% MSE changes relative to paired Dense by forecast interval (%).

Interval	Solar ($n = 20$)	ECL ($n = 10$)
H1–96	-1.347 [-2.290, -0.395]	-1.098 [-1.753, -0.440]
H1–48	$+0.282$ [-1.674, $+2.276$]	-1.409 [-2.349, -0.461]
H49–96	-2.692 [-3.774, -1.598]	-0.847 [-1.331, -0.360]
H1–24	$+0.540$ [-2.453, $+3.625$]	-1.606 [-2.578, -0.624]
H25–48	$+0.167$ [-1.429, $+1.789$]	-1.262 [-2.209, -0.305]
H49–72	-2.913 [-4.023, -1.790]	-1.240 [-1.915, -0.561]
H73–96	-2.462 [-3.662, -1.247]	-0.448 [-0.777, -0.119]

At 90% removal, aggregate improvement is not uniform across receiving variables (Table C6). Each receiver is classified by its geometric mean paired MSE change across joint-96 blocks; receiver MSE averages all 96 forecast steps. The $\pm 1\%$ thresholds are descriptive categories, not variable-wise significance or noninferiority decisions.

Table C6: Receiver counts by paired joint-96 Random-90% MSE change relative to Dense.

Dataset	Receivers	$> +1\%$	Within $\pm 1\%$	$< -1\%$
Solar	137	2	48	87
ECL	321	35	78	208

We also audit the saved structural masks themselves (Table C7). In-degree counts direct cross-variate sources per receiver, excluding self-access. The two-layer percentage counts ordered source–receiver pairs reachable through one or two allowed attention edges, again excluding self-pairs. Every receiver retains at least one direct cross-variate source at 90% removal, while zero-in-degree receivers emerge in Solar at 97.5% removal. Reachability describes permitted paths, not measured information use.

Table C7: Direct receiver coverage and two-layer structural reachability of saved random masks.

Setting	n	Edges	In-degree mean/min	Zero (%)	Two-hop (%)
Solar, 90%	20	1,863	13.60 / 4	0.00	76.90
ECL, 90%	10	10,272	32.00 / 14	0.00	96.38
Solar, 95%	20	932	6.80 / 1	0.00	32.28
Solar, 97.5%	20	466	3.40 / 0	3.91	10.44

The Solar and ECL mask tensors and metadata match the hashes recorded for their respective training runs. ECL sparse horizon and receiver MSEs were obtained by evaluation-only replay of saved checkpoints; aggregate MSE agrees with the training exports within 1.6×10^{-9}.

D Receiver Coverage and Predictive Source Relevance

Following Section 6, this appendix defines the fixed-budget allocation conditions, reports receiver-level recovery, and specifies the synthetic source-relevance experiment.

D.1 Fixed-Budget Allocation Conditions

At a fixed budget of 466 cross-variate edges (97.5% sparsity), five structural topologies were evaluated across 20 paired blocks:

1. **Random**: Edges sampled uniformly at random without replacement across off-diagonal entries.
2. **Coverage-1 (Min-query-degree 1)**: Minimal rewiring of Random masks ensuring every receiver has in-degree ≥ 1 while preserving total edge count (466) and source degree sequence.
3. **Coverage-2 (Min-query-degree 2)**: Rewiring ensuring every receiver has in-degree ≥ 2.
4. **Query-balanced**: Equalizing receiver in-degrees ($k_i \in \{3, 4\}$, $\bar{k} = 3.40$) while leaving source degrees unconstrained.
5. **Bi-balanced**: Simultaneously balancing receiver in-degrees ($k_i \in \{3, 4\}$) and source out-degrees ($k_j \in \{3, 4\}$) via bipartite matching.

Table D1: Overall Solar joint-96 effects for the five 466-edge allocation conditions.

Allocation Condition	Mean MSE	Geometric ΔMSE vs Dense	Two-sided 95% CI	Contrast vs Random
Random	0.205595	+3.228%	[+1.881%, +4.592%]	Reference
Coverage-1	0.204485	+2.672%	[+1.450%, +3.909%]	-0.539% (nominal $p = 0.016$)
Coverage-2	0.204323	+2.590%	[+1.260%, +3.936%]	-0.618% (nominal $p = 0.009$)
Query-balanced	0.203239	+2.046%	[+0.830%, +3.277%]	-1.145% (Holm $p = 0.000123$)
Bi-balanced	0.203921	+2.389%	[+1.153%, +3.641%]	-0.812% (Holm $p = 0.000386$)

Brackets report two-sided 95% confidence intervals, and the final column compares each condition with its paired Random mask. Coverage-1 and Coverage-2 have descriptive nominal p-values; Query-balanced and Bi-balanced have Holm-adjusted p-values from the prespecified comparison family.

In the direct paired contrast, Bi-balanced has $+0.336\%$ MSE relative to Query-balanced (95% CI $[-0.20\%, +0.88\%]$; Holm-adjusted $p = 0.90$), with no detected additional gain from balancing source out-degree.

D.2 Solar Receiver-by-Horizon Post Hoc Decomposition

To assess whether damage from severe edge sparsity is uniformly distributed or concentrated among isolated variables, we performed a checkpoint-only re-evaluation on the 20 existing Dense, Random-97.5% (466 edges), and Coverage-1 checkpoints without retraining.

Receivers were partitioned using their degree in the paired Random mask before rewiring:

- **Original Zero Receivers**: Receivers allocated zero incoming cross-variate edges under Random-97.5%.
- **Original Nonzero Receivers**: Receivers allocated ≥ 1 cross-variate edges under Random-97.5%.

Table D2: Receiver-stratified damage and matched-budget recovery in Solar/compact iTransformer-style.

Contrast	Receiver Subgroup	Short Band ($h \in [1, 48]$) ΔMSE (95% CI)	Far Band ($h \in [49, 96]$) ΔMSE (95% CI)
Random-97.5% vs Dense	All Receivers	+11.172% [+8.419%, +13.996%]	-3.352% [-4.370%, -2.323%]
	Original Nonzero	+10.402% [+7.609%, +13.267%]	-3.461% [-4.484%, -2.428%]
	Original Zero	**+29.882% [+27.831%, +31.965%]**	-0.327% [-1.874%, +1.244%]
Coverage-1 vs Random-97.5%	All Receivers	-1.043% [-1.786%, -0.296%]	-0.058% [-0.219%, +0.104%]
	Original Nonzero	-0.530% [-1.290%, +0.235%]	+0.032% [-0.128%, +0.193%]
	Original Zero	**-11.618% [-12.921%, -10.296%]**	-2.493% [-3.227%, -1.753%]

All values are geometric mean paired MSE changes across 20 families; the receiver-by-horizon analyses are post hoc.

D.3 Cross-Setting Coverage-1 Local Recovery

Coverage-1 rewiring was evaluated across three settings to assess whether the local vulnerability of zero-coverage receivers transfers across architectures and datasets.

Table D3: Local recovery of isolated receivers under matched-budget Coverage-1 rewiring.

Setting	Backbone	Training / Eval Endpoint	n	Original Zero ΔMSE (95% CI)	Original Nonzero ΔMSE (95% CI)	Overall ΔMSE (95% CI)
Solar	compact iTransformer	Joint-96 aggregate	20	**-7.265% [-8.089, -6.433]**	-0.242% [-0.690, +0.209]	-0.539% [-0.977, -0.098]
Solar	SAMformer	Short-only (H1–48)	5	**-18.981% [-23.265, -14.458]**	+1.262% [+0.076, +2.462]	+0.279% [-0.774, +1.343]
ECL	compact iTransformer	Short-only (H1–48)	5	**-5.468% [-8.390, -2.452]**	+0.049% [-0.301, +0.400]	-0.177% [-0.574, +0.222]

Changes and CI limits are percentages; n counts paired mask families. Negative values indicate error reduction relative to paired Random masks. Original-zero receivers improved in 20/20 Solar/iTransformer families, 5/5 Solar/SAMformer families ($p = 0.03125$), and 5/5 ECL held-out mask families ($p = 0.03125$). Solar/iTransformer is post hoc; Solar/SAMformer is prespecified with its paired Random baseline reused; ECL uses prespecified held-out mask families.

All entries in Table D3 use the same aggregation. Within each family and receiver group, we first average MSE across receivers, take the paired log ratio of Coverage-1 to Random, and then average the family-level log ratios across families. The intervals are two-sided 95% t intervals on the family-level log ratios, transformed to percentage changes. For the overall row, the group contains all receivers and therefore reduces to the paired overall-MSE ratio. Receiver groups are defined from the pre-rewiring Random mask.

D.4 Synthetic Factor–Proxy Worlds and Source-Relevance Protocol

To isolate the predictive value of cross-variate edges independently of network degree, we construct synthetic factor–proxy dynamical systems with ground-truth parent graphs.

D.4.1 Latent Factor Dynamics

Each world instantiates $K = 8$ independent latent AR(1) processes:

$$f_k(t) = \rho f_k(t-1) + \eta_k(t), \qquad \eta_k(t) \sim \mathcal{N}(0, 1-\rho^2), \quad k = 1, \dots, 8.$$

D.4.2 Proxy Sources

Each factor generates $P = 8$ noisy proxy sources (total $S = 64$ source channels):

$$s_{k,p}(t) = f_k(t) + \xi_{k,p}(t), \qquad \xi_{k,p}(t) \sim \mathcal{N}(0, \sigma^2_{\text{proxy}}).$$

D.4.3 Receiver Dynamics

Each receiver $i \in \{1, \dots, 64\}$ is driven by its own autoregressive history and a signed lag-one mixture of two fixed parent factors $k_1(i), k_2(i) \in \{1, \dots, 8\}$:

$$r_i(t) = \alpha r_i(t-1) + \gamma \cdot \left[w_{i,1} f_{k_1(i)}(t-1) + w_{i,2} f_{k_2(i)}(t-1) \right] + \epsilon_i(t),$$

where $w_{i,1}, w_{i,2} \in \{-1, +1\}$ are fixed parent signs, and $\epsilon_i(t) \sim \mathcal{N}(0, \sigma^2_{\text{rec}})$.

Table D4: Generative parameters and evaluation settings for the synthetic factor–proxy worlds.

Generative Parameter	Symbol	Value
Number of latent factors	K	8
Proxy sources per factor	P	8 (64 total sources)

Continued on next page

Table D4 (continued)

Generative Parameter	Symbol	Value
Number of receivers	N_{rec}	64
True parent factors per receiver	P_{rec}	2
Factor AR coefficient	ρ	0.45
Receiver self-AR coefficient	α	0.40
Cross-signal coupling scale	γ	0.65
Proxy noise standard deviation	σ_{proxy}	0.35
Receiver noise standard deviation	σ_{rec}	0.55
Burn-in time steps	T_{burn}	512
Retained time steps	T_{total}	6,000
Context and forecast horizon	$L = H$	96
Primary evaluation endpoint	—	Receiver MSE on H1–12
Evaluated degree grid	k	$\{2, 4, 8, 16\}$
Oracle/Decoy held-out worlds	—	Worlds 8–12

For the Oracle/Decoy runs, the model retains a 96-step output head, but training minimizes MSE over forecast steps H1–12 across all 128 channels (64 proxy sources and 64 receivers). Early stopping and best-checkpoint selection use validation MSE over the same steps and channels. The primary test endpoint in Figure 5 is H1–12 MSE restricted to the 64 receiver channels; all-channel test MSE is recorded separately.

D.4.4 Oracle versus Decoy Protocol

At each matched degree $k \in \{2, 4, 8, 16\}$:

- **Oracle**: Connects each receiver i strictly to proxy sources belonging to its two true parent factors $(k_1(i), k_2(i))$.
- **Decoy**: Connects each receiver i to the same number of proxy sources belonging to two non-parent factors $(k_3(i), k_4(i)$ where $k_3, k_4 \notin \{k_1, k_2\})$.

D.4.5 Oracle–Decoy Results and Aggregation Across Degrees

For world $w \in \{8, 9, 10, 11, 12\}$ and receiver degree $k \in \{2, 4, 8, 16\}$, let $d_{w,k}$ be the log ratio of Oracle to Decoy H1–12 receiver MSE. The degree-specific effects in Figure 5B summarize the five paired worlds at each k. For the aggregate contrast reported in Section 6.3, we first average the four log ratios within each world, then summarize the five world-level effects:

$$\bar{d}_w = \frac{1}{4} \sum_{k \in \{2,4,8,16\}} d_{w,k}, \qquad \Delta_{\text{degree-averaged}} = 100 \left[\exp\left(\frac{1}{5} \sum_{w=8}^{12} \bar{d}_w \right) - 1 \right].$$

Table D5: Oracle–Decoy effects on the 64 receiver channels at H1–12.

Receiver degree	Worlds	Oracle / Decoy change [95% CI] (%)	Worlds favoring Oracle	Exact one-sided p
2	5	-3.116 [-3.401, -2.831]	5/5	—
4	5	-3.363 [-3.630, -3.096]	5/5	—
8	5	-3.430 [-3.682, -3.178]	5/5	—
16	5	-3.459 [-3.702, -3.215]	5/5	—
Within-world average across degrees	5	**-3.342 [-3.602, -3.082]**	**5/5**	**0.03125**

Effects and two-sided 95% confidence intervals use the geometric summary of paired log-MSE ratios. Degree-specific intervals are pointwise and unadjusted across degrees. For the final row, the interval is computed

from the five within-world degree averages with four degrees of freedom. The exact one-sided sign-flip test enumerates all $2^5 = 32$ sign assignments to these five world-level effects. The independent unit is the world; the four degrees within each world are repeated conditions. Negative changes favor Oracle.

E Supplementary Diagnostics for the Discussion

This appendix supports the discussion in Section 7 with seasonal reference geometry, the Solar and ECL horizon diagnostics, and supplementary within-series predictability baselines. The empirical diagnostics are post hoc.

E.1 Mathematical Formulation of Seasonal Reference Availability

Let P_s denote the fundamental seasonal period of a time series, L be the input lookback window length, and $h \geq 1$ be the forward forecast horizon step. The historical input window for a forecast origin t spans $[t - L + 1, t]$, and the target position is $t + h$.

A same-phase historical observation at lag mP_s ($m \in \mathbb{N}^+$) is located at timestamp $t + h - mP_s$. This seasonal reference point lies within the available input window if and only if:

$$t - L + 1 \leq t + h - mP_s \leq t.$$

Subtracting t and adding mP_s across the inequality yields the exact horizon visibility condition:

$$mP_s - L + 1 \leq h \leq mP_s.$$

For the most immediate same-phase seasonal reference ($m = 1$):

- **Entry Horizon**: The reference point first becomes available inside the input window at:
$$h_{\text{entry}} = P_s - L + 1.$$
- **Exit Horizon**: The reference point remains available until $h = P_s$, beyond which $t + h - P_s > t$ falls outside the historical window (requiring $m = 2$ references, which enter at $h = 2P_s - L + 1$).

E.1.1 Receptive-Field Geometry Implication

When the lookback length L is fixed, advancing the forecast horizon h simultaneously increases temporal lead time and alters the availability of within-series seasonal anchors. Horizon profiles can therefore reflect both lead-time information decay and seasonal-reference visibility geometry.

E.2 Solar Daily-Transition Diagnostic

Solar is sampled at 10-minute intervals. Under the daily-period hypothesis ($P_s = 144$ steps) and $L = 96$, the previous-day same-phase reference enters at $h = 144 - 96 + 1 = 49$.

A post hoc per-horizon reanalysis of the Solar Joint-96 profile focuses on the transition around $h = 49$: H37–48 remains damaging, H49–60 is near zero, and H61–84 favors No-cross.

E.3 ECL Weekly-Boundary Diagnostic

ECL is sampled at 1-hour intervals. Under the weekly-period hypothesis ($P_s = 168$ steps) and input length $L = 96$, the immediately preceding same-phase weekly observation enters the input window at:

$$h_{\text{entry}} = 168 - 96 + 1 = 73.$$

Table E1 reports the per-horizon geometric MSE degradation of No-cross relative to Dense for both the Joint-96 model and the separately trained Far-objective model around this theoretical boundary.

Table E1: Per-horizon relative MSE change around the theoretical weekly entry boundary $h = 73$ in ECL.

Horizon (h)	Joint-96 geometric ΔMSE	Far-objective geometric ΔMSE
68	+7.601%	+7.086%
69	+6.833%	+6.336%
70	+6.031%	+5.675%
71	+5.086%	+4.587%
72	+3.986%	+3.615%
73 (Boundary)	**+3.099%**	**+2.954%**
74	+3.252%	+2.941%
75	+3.331%	+2.831%
76	+3.298%	+2.776%

E.3.1 Slope Diagnostic Across Seeds

To assess whether the flattening around $h = 73$ is consistent across runs, we fit ordinary least squares lines to pointwise log-MSE ratios separately for each seed over pre-boundary ($h = 68$–72) and post-boundary ($h = 73$–76) windows.

Table E2: Seed-level slope diagnostic for ECL Far-objective retrained checkpoints (units: log-ratio percentage points per horizon step).

Seed	Pre-boundary slope ($h = 68$–72)	Post-boundary slope ($h = 73$–76)	$\lvert\text{post}\rvert < \lvert\text{pre}\rvert$
0	-0.933	-0.187	Yes
1	-0.863	+0.094	Yes
2	-0.821	-0.025	Yes
3	-0.732	-0.130	Yes
4	-0.777	-0.065	Yes

All five seeds flatten in slope magnitude. The inspection windows were selected after viewing the aggregate profile, and the Far model jointly optimizes horizons 49–96.

E.4 Within-Series Predictability Baselines

Table E3 reports the Solar Far ($h \in [49, 96]$) MSE values on the same test evaluation slices. The post hoc zero-training baselines show that No-cross remains far below persistence and the input-window mean, so Far still has learnable within-variate signal. The comparison does not identify why Dense does not benefit from additional cross-variate access.

Table E3: Within-series baselines and trained-model endpoints for Solar Far on matched test slices.

Forecast	Definition	Test MSE
Persistence	$\hat{y}_{t+h} = y_t$	2.003
Input-window mean	$\hat{y}_{t+h} = L^{-1} \sum_{i=0}^{L-1} y_{t-i}$	0.687
Dense	Band-specific trained model	0.215
No-cross	Band-specific trained model	0.213