COVARIANCE-ROBUST AND CONFORMAL BANDS FOR LONGITUDINAL FUNCTIONAL TREATMENT EFFECTS

Anonymous authors
Paper under double-blind review

ABSTRACT

Repeatedly randomized longitudinal studies support two distinct forms of uniform inference: a confidence band for the population treatment-effect surface and a prediction band for a new subject's latent conditional-mean treatment-effect curve. For the population target, we combine working-independence series regression, a subject-block CR3 covariance estimator, and subject-level Gaussian multiplier calibration. For the latent target, pooled functional principal component analysis learns a joint eigensystem and raw personal history predicts the scores. We incorporate this predictor into subject-level conformal calibration using recovery sets for the unobserved calibration curves. A truncation adjustment retains the full latent target. Coverage is marginal over an exchangeable future subject under the stated recovery-event conditions. Gaussian model-based bands provide a separate reference with known-parameter and fitted asymptotic guarantees. The experiments retain the population SCB analysis and evaluate the raw-history predictor under the original model. Separate end-to-end experiments validate the implemented conformal pipeline under Gaussian innovations with both Gaussian and heavy-tailed random effects. Its high observed coverage comes with substantial band-width inflation from recovery uncertainty; the model-based reference bands can undercover. In a same-realization comparison, residual-studentized exact-F recovery gives mean paired relative width reductions of 13.4–16.2% while observed simultaneous coverage remains between $.9980$ and 1.0000. For a separate rank-four covariate-dependent working predictor, we report fixed-grid GCV-label coverage empirically and distinguish it from the finite-sample exact-label oracle guarantee.

1 INTRODUCTION

Repeatedly randomized longitudinal studies contain information at two levels. Across subjects, they identify how the mean treatment effect varies with observed covariates. A population confidence band quantifies uncertainty about this mean effect. For a new subject, the observed history can also inform a persistent departure from the population effect. An individual prediction band covers that subject's latent conditional-mean effect curve, excluding the additional noise in a future outcome. Both bands seek simultaneous coverage over their evaluation domains. They differ in the target being covered: the population mean effect or a new subject's latent effect curve. Section 2.1 defines the common model and these two targets.

Table 1: The two inferential targets use different sampling units, data, and coverage statements.

Target	Sampling unit	Information used	Guarantee
Population mean effect	Subject	Repeated randomized outcomes	Asymptotic simultaneous confidence over the covariate domain
New subject's latent effect curve	Calibration subject	History and disjoint randomized recovery visits	Marginal simultaneous prediction over the curve domain

Both problems concern calibration with longitudinal data, but the source of difficulty differs. For the population target, repeated visits make visit-level standard errors invalid. For the latent target, the

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.

curve needed for calibration is not observed. The methods below address these two failures without treating pointwise uncertainty as simultaneous curve uncertainty.

For the population target, treating visits as independent understates uncertainty, while estimating and inverting a functional covariance operator can be unstable. We separate estimation from uncertainty quantification. For the reported population construction, treatment-specific means are fitted in an additive dictionary under working independence; one score is formed for each subject, and a subject-block CR3 sandwich estimates the covariance. A Gaussian multiplier maximum then calibrates simultaneous coverage. The argument uses high-dimensional Gaussian approximation and comparison (Chernozhukov et al., 2013); it does not require a correct within-subject covariance model.

For personal prediction, functional principal component analysis separates learning population eigenfunctions from predicting a new person's scores (Yao et al., 2005). We estimate a shared joint covariance for both arms, then use the person's raw history to predict the scores. Mean adjustment with a shared covariance is one of the alternatives studied by Jiang and Wang (2010). The prediction formulas also relate to Ding et al. (2022); the original pooled branch does not adopt their covariate-dependent covariance. This preserves the original model while allowing histories shorter than the random-effect dimension.

The latent target creates a separate calibration problem: calibration curves are not observed. We therefore distinguish the FPCA predictor from the subject-level rule that calibrates its prediction error. A recovery set for each calibration curve produces a feasible upper score. Ranking these scores and accounting for the future recovery event yields a conformal band around the FPCA predictor. The reported center may be truncated, but the band retains the omitted predicted contribution and the full latent target. This construction does not require recovery outcomes from the future person.

The two inferential layers have separate assumptions and evidence. The SCB remains covariance-robust under its subject-score conditions. The protected personal conformal guarantee requires exchangeability and valid recovery sets; it is marginal over the future subject and history. The direct branch instead has an exact-label oracle guarantee. The component experiments evaluate the raw-history predictor and Gaussian model-based bands under the original model, revealing improved prediction with history and finite-sample undercoverage. The separate end-to-end experiment in Section 4.3 evaluates recovery-set conformal calibration under Gaussian innovations and documents its conservative coverage and substantial width cost. A same-seed residual-studentized exact-F branch reduces mean width relative to the externally bounded-noise branch, although both remain conservative in the reported cells.

We also study a direct covariate-dependent FPCA branch inspired by Ding et al. (2022). It extends the working covariance to record-varying covariates and two arms, learns a joint weighted FPCA, and predicts from two history records. Its exact-label oracle has a finite-grid conformal rank guarantee. The implemented GCV point labels do not inherit that guarantee, so we report their coverage as simulation evidence and identify the recovery-gap condition separately. This branch remains distinct from the population SCB and the original guarded personal band.

2 Targets and methods

2.1 Model and inferential targets

For subject $i = 1, \ldots, n$ and visit $t = 1, \ldots, T_i$, observe $(Y_{it}, A_{it}, X_{it}, Z_{it})$. Here Y_{it} is the outcome, $A_{it} \in \{0, 1\}$ is the assigned treatment, $X_{it} \in \mathcal{X} \subset \mathbb{R}^d$ is a vector of observed covariates, and $Z_{it} \in \mathcal{Z}$ indexes within-subject variation, such as time. Write $x = (x_1, \ldots, x_d)^\top$ and z for evaluation values of these covariates. To describe the targets, consider the model

$$Y_{it} = \mu_{A_{it}}(X_{it}) + g_{i,A_{it}}(Z_{it}) + \varepsilon_{it}. \tag{1}$$

The function $\mu_a(x)$ is the model's across-subject mean under treatment a. We impose no additive restriction across the coordinates of x at this model level. The estimation and coverage results below state their own dictionary, representation, and approximation conditions. The process $g_{ia}(z)$ is subject i's persistent deviation from this mean under treatment a, and ε_{it} is the visit-level innovation. The processes g_{ia} have mean zero across subjects and may be dependent across treatment arms. Subjects are independent. The procedures impose different conditions on within-subject dependence, stated in

Section 3 and the corresponding appendices. Marginal centering of g_{ia} does not imply zero residual mean conditional on observed visits and covariates. A causal interpretation also requires consistency, no interference, and sequential randomization; informative histories can still change the target of unweighted population regression.

The population target is the mean treatment-effect surface

$$\tau(x) = \mu_1(x) - \mu_0(x), \qquad x \in \mathcal{X}.$$

For a subject, write $u = (x, z)$ and fix a prediction domain $\mathcal{U} \subseteq \mathcal{X} \times \mathcal{Z}$. The individual target is

$$\tau_i(u) = \tau(x) + g_{i1}(z) - g_{i0}(z). \tag{2}$$

This latent conditional-mean contrast excludes the innovation in a future outcome. A population confidence band covers τ simultaneously over $\mathcal{X}$. For a new subject with observed record history $\mathcal{H}_i$, with length $h_i = |\mathcal{H}_i|$, an individual prediction band aims to satisfy

$$\mathbb{P}\{\tau_i(u) \in \mathcal{C}_i(u) \text{ for every } u \in \mathcal{U}\} \geq 1 - \gamma.$$

Here $\gamma \in (0, 1)$ is the target error probability, and coverage is marginal over the new subject and that subject's history. The Gaussian construction below restricts this claim to a fixed grid. Conformal calibration is the method used to construct individual prediction bands; simultaneous coverage is a property sought by both the confidence and prediction bands.

2.2 Population confidence bands

For the population construction developed here, choose the full-rank centered additive dictionary $B_K(x) \in \mathbb{R}^{p_K}$ described in Appendix A.1, and write $\mu_a(x) = B_K(x)^\top \theta_{aK} + r_{aK}(x)$. This is a dictionary projection with an explicit remainder, not an additive restriction on the general model in (1). Stack the two arm-specific coefficients in β_K, set

$$c_K(x) = \{B_K(x)^\top, -B_K(x)^\top\}^\top, \qquad \tau_K(x) = c_K(x)^\top \beta_K.$$

For each visit, define

$$d_{it} = \{A_{it} B_K(X_{it})^\top, (1 - A_{it}) B_K(X_{it})^\top\}^\top.$$

Let the rows of the subject-level design matrix D_i be $d_{it}^\top$, and $Y_i = (Y_{i1}, \ldots, Y_{iT_i})^\top$. Thus D_i is $T_i \times 2p_K$ and β_K is $2p_K$-dimensional. Precisely, β_K solves $\mathbb{E}\{D_i^\top (Y_i - D_i \beta_K)\} = 0$, so τ_K is an unweighted visit-distribution projection contrast. Interpreting the band as inference for τ requires the residual score condition and negligible projection bias specified in Appendix A. Randomization and marginal mean-zero random effects alone do not imply those conditions. With $G_n = \sum_i D_i^\top D_i$, the working-independence estimator is

$$\widehat{\beta} = G_n^{-1} \sum_i D_i^\top Y_i, \qquad \widehat{\tau}(x) = c_K(x)^\top \widehat{\beta}.$$

The word "independence" describes only this working precision.

For the HeartSteps sensitivity analysis, M0 is the complete contrast dictionary; M1–M3 impose prespecified day or decision-slot restrictions. We cross these dictionaries with working independence and a frozen random-intercept precision. Because a restriction or working precision can change the randomization-averaged projection under misspecification, we report projection and full-surface coverage separately. Appendix D.2 gives the dictionaries, precisions, and target calculations.

Define the three subject-level quantities

$$\widehat{e}_i = Y_i - D_i \widehat{\beta}, \qquad P_i = D_i G_n^{-1} D_i^\top, \qquad \widetilde{g}_i = D_i^\top (I_{T_i} - P_i)^{-1} \widehat{e}_i.$$

The subject-block CR3 covariance and pointwise standard error are

$$\widehat{A}_n = n^{-1} G_n, \qquad \widehat{\Omega}_n = n^{-1} \sum_i \widetilde{g}_i \widetilde{g}_i^\top,$$

$$\widehat{V}_\beta = n^{-1} \widehat{A}_n^{-1} \widehat{\Omega}_n \widehat{A}_n^{-1}, \qquad \widehat{s}^2(x) = c_K(x)^\top \widehat{V}_\beta c_K(x).$$

For independent $\xi_i \sim N(0, 1)$, conditional on the data, compute

$$\mathbb{T}_n^*(x) = \frac{n^{-1/2} \sum_i \xi_i c_K(x)^\top \widehat{A}_n^{-1} \widetilde{g}_i}{\widehat{\sigma}_K(x)}, \qquad \widehat{\sigma}_K(x) = \sqrt{n}\, \widehat{s}(x).$$

If $c_{1-\alpha}^*$ is the conditional $(1 - \alpha)$-quantile of the maximum of $|\mathbb{T}_n^*(x)|$ on a prespecified grid, the reported band is

$$\mathcal{C}_{1-\alpha}^{\text{pop}}(x) = [\widehat{\tau}(x) \pm c_{1-\alpha}^* \widehat{s}(x)]. \tag{3}$$

2.3 PERSONAL FPCA AND CONFORMAL PREDICTION

The fitting people, calibration people and future person are disjoint. We learn a shared eigensystem from the fitting sample and predict each person's scores directly from raw history. Recovery information protects the latent calibration labels; it is not the input to the future person's point predictor.

Pooled fitting and history prediction. Choose finite, linearly independent numerical dictionaries $a(x)$ for the arm means and $b(z)$ for the persistent deviations. They may be selected in advance or learned on an auxiliary sample independent of the two fitting groups, calibration subjects, and future subject; in the latter case we condition on that selection. The exact finite-dimensional model-based results assume

$$\mu_a(x) = a(x)^\top \eta_a, \qquad g_{ia}(z) = b(z)^\top \gamma_{ia}, \qquad \Gamma = \mathrm{Cov}\{(\gamma_{i0}^\top, \gamma_{i1}^\top)^\top\},$$

where b has p coordinates and $D = 2p$. Under a fixed reference measure ν on $\mathcal{Z}$, a chosen linearly independent dictionary $b_0(z) \in \mathbb{R}^p$ with $H = \int b_0(z) b_0(z)^\top d\nu(z) > 0$ is orthonormalized as $b = H^{-1/2} b_0$. To preserve $g_{ia} = b_0^\top \gamma_{ia,0} = b^\top \gamma_{ia}$, set $\gamma_{ia} = H^{1/2} \gamma_{ia,0}$. For the two-arm stack, let $T = \mathrm{diag}(H^{1/2}, H^{1/2}) \in \mathbb{R}^{2p \times 2p}$ and $\Gamma = T \Gamma_0 T^\top$ before taking the joint eigendecomposition. The mean dictionary a need not be additive or orthonormal. These choices remain subject to the representation, Gram-matrix, moment, and variance assumptions in Appendix B.1. The reported experiments use their prespecified fixed dictionaries. The dictionary supplies numerical coordinates; its fitted eigenvectors define the learned FPCs. Estimate the common mean on one group of fitting people and the joint covariance on an independent group by pooled residual second moments, with positive covariance and noise-variance clipping. Appendix B.1 gives the estimators and their assumptions. No individual regression is needed for the raw-history point predictor; recovery is used only to protect conformal calibration labels.

Write $\widehat{\Gamma} = \widehat{U} \widehat{\Lambda} \widehat{U}^\top$ and $\widehat{\phi}_{ak}(z) = b(z)^\top \widehat{U}_{ak}$, retaining both arms in the joint decomposition. For history $\mathcal{H}_i$, let F_i and B_i be the arm-specific mean and random-effect design matrices, and put $\widehat{\Phi}_i = B_i \widehat{U}$. For the fitted estimator and numerical study, use $\widehat{\Omega}_i = \widehat{\sigma}^2 I$ with the scalar noise-variance fit. The score predictor and error covariance are

$$\widehat{m}_i = \widehat{\Lambda} \widehat{\Phi}_i^\top \widehat{\mathcal{V}}_i^{-1} (Y_i - F_i \widehat{\eta}), \qquad \widehat{\mathcal{V}}_i = \widehat{\Phi}_i \widehat{\Lambda} \widehat{\Phi}_i^\top + \widehat{\Omega}_i, \tag{4}$$

$$\widehat{\mathcal{S}}_i = \widehat{\Lambda} - \widehat{\Lambda} \widehat{\Phi}_i^\top \widehat{\mathcal{V}}_i^{-1} \widehat{\Phi}_i \widehat{\Lambda}. \tag{5}$$

With no history these reduce to 0 and $\widehat{\Lambda}$. These are linear prediction quantities; they are conditional posterior moments only under the stated Gaussian model.

Let $\widehat{d}_k(z) = \widehat{\phi}_{1k}(z) - \widehat{\phi}_{0k}(z)$ and let P_M retain the first M scores. Set $\widehat{\tau}_{\mathrm{fit}}(x) = a(x)^\top (\widehat{\eta}_1 - \widehat{\eta}_0)$, using only the personal fitting sample, separately from the population SCB fit. The personal predictor, omitted predicted contribution and scale are

$$\widehat{c}_{i,M}(u) = \widehat{\tau}_{\mathrm{fit}}(x) + \widehat{d}(z)^\top P_M \widehat{m}_i, \tag{6}$$

$$\widehat{\delta}_{i,M}(u) = \widehat{d}(z)^\top (I - P_M) \widehat{m}_i, \qquad \widehat{s}_i(u) = \max\left\{ \sqrt{\widehat{d}(z)^\top \widehat{\mathcal{S}}_i \widehat{d}(z)}, s_{\min} \right\}. \tag{7}$$

The scale floor $s_{\min} > 0$ and all fitting rules are frozen before calibration. The full covariance is retained even when the displayed center is truncated. In the experiments, M is the smallest rank explaining 95% of fitted variance.

Subject-level calibration. Let $v(u)$ be a linearly independent basis spanning the mean-contrast and random-effect functions, so $\tau_i(u) = v(u)^\top \zeta_i$. It represents the full target, including discarded FPC directions. If a or b changes, the recovery dictionary and uncertainty set must cover the corresponding full effect span. An approximate span requires the recovery-design and prediction-domain error envelopes in Theorem 9; the exact-dictionary statement does not cover it without these corrections. In Appendix C, the generic recovery notation $b, \theta_i, \widehat{\theta}_i$ means $v, \zeta_i, \widetilde{\zeta}_i$ here; its coefficient dimension is $\dim(v)$, not the random-effect coordinate count p. The appendix's legal center and scale are instantiated as $\widehat{c}_{i,D}$ and $\widehat{s}_i$. For calibration person i, a disjoint randomized recovery block supplies a confidence ellipsoid

$$\mathcal{E}_i = \{\zeta : \|\zeta - \widetilde{\zeta}_i\|_{V_i} \le \beta_i\}.$$

Here V_i is recovery information, distinct from the history covariance $\widehat{\mathcal{V}}_i$. Appendix C specifies sufficient recovery conditions. Proposition 7 gives the Gaussian-innovation construction used in the end-to-end experiment. Let $\widehat{c}_{i,D} = \widehat{c}_{i,M} + \widehat{\delta}_{i,M}$. Use one feasible score per person:

$$R_i^+ = \sup_{u \in \mathcal{U}} \frac{|v(u)^\top \widetilde{\zeta}_i - \widehat{c}_{i,D}(u)| + \beta_i \|v(u)\|_{V_i^{-1}}}{\widehat{s}_i(u)}. \tag{8}$$

For m calibration people and a prespecified rank error γ_{rk}, set $k = \lceil (m+1)(1-\gamma_{\mathrm{rk}}) \rceil$ and take q^+ to be the kth order statistic of these scores, with $q^+ = \infty$ if $k > m$. Report the conformal band around the truncated personal predictor:

$$\mathcal{C}_{m+1}^{\mathrm{FPCA}}(u) = \left[\widehat{c}_{m+1,M}(u) \ \pm \ \left\{ |\widehat{\delta}_{m+1,M}(u)| + q^+ \widehat{s}_{m+1}(u) \right\} \right]. \tag{9}$$

The future person's recovery outcomes are never observed. An appended future recovery block appears only in the proof. The same protected score and cutoff also give the full-center band $[\widehat{c}_{i,D}(u) \pm q^+ \widehat{s}_i(u)]$ with the same coverage lower bound. The displayed truncated-center band contains it pointwise because $\widehat{c}_{i,D} = \widehat{c}_{i,M} + \widehat{\delta}_{i,M}$. The guard supports a truncated reported center while retaining the full latent target; it is not a width-reduction or computational-saving device. Suprema computed only on a grid guarantee that grid; continuum coverage needs exact maximization or a valid discretization bound.

The preceding construction is the original protected personal branch. The following direct branch uses a distinct working predictor and calibration score and does not alter the recovery-set band above.

2.4 Covariate-dependent FPCA for direct personal calibration

We add a personal-prediction branch whose working covariance can vary with the record-level covariate and time. It leaves the population SCB and the guarded recovery-set procedure unchanged. The target remains the full latent effect $\tau_i(u)$ on $\mathcal{G} = \{-1, -.5, 0, .5, 1\} \times \{0, .05, \ldots, 1\}$; the rank-four model below is a prediction rule, not a redefinition of that target. Throughout this branch $U_{it} = (X_{it}, Z_{it})$.

Let $b_z(z) \in \mathbb{R}^7$ and $b_x(x) \in \mathbb{R}^4$ be ordinary cubic B-spline vectors on $z \in [0,1]$ and $x \in [-1,1]$. Set $h(u) = b_x(x) \otimes b_z(z) \in \mathbb{R}^{28}$ and $q_a(u) = ((1-a)h(u)^\top, a h(u)^\top)^\top \in \mathbb{R}^{56}$. With $\theta_\mu^{\mathrm{CD}} \in \mathbb{R}^{56}$ and $C^{\mathrm{CD}} = ((C_0^{\mathrm{CD}})^\top, (C_1^{\mathrm{CD}})^\top)^\top \in \mathbb{R}^{56 \times 4}$, where $C_a^{\mathrm{CD}} \in \mathbb{R}^{28 \times 4}$, define $\ell_a(u) = C_a^{\mathrm{CD}\top} h(u)$. The working model is

$$Y_{it} \approx q_{A_{it}}(U_{it})^\top \theta_\mu^{\mathrm{CD}} + \ell_{A_{it}}(U_{it})^\top \omega_i + e_{it}. \tag{10}$$

It imposes $E_{\mathsf{W}}(\omega_i \mid \mathcal{D}_i) = 0$, $\mathrm{Cov}_{\mathsf{W}}(\omega_i \mid \mathcal{D}_i) = I_4$, $\mathrm{Var}_{\mathsf{W}}(e_i \mid \text{design}) = \sigma_{\mathrm{CD}}^2 I$, and $\mathrm{Cov}_{\mathsf{W}}(\omega_i, e_i \mid \text{design}) = 0$. Thus

$$K_{ab}^{\mathrm{CD}}(u,v) = \ell_a(u)^\top \ell_b(v). \tag{11}$$

The shared factor ω_i defines cross-arm and cross-covariate covariance jointly. This adapts the functional prediction ideas of Ding et al. (2022) to record-varying x, two treatment arms, a joint discrete FPCA, and a coefficient-difference penalty; it is not a reproduction of their estimator. In that comparison, Ding's time t corresponds to our z, and Ding's covariate z corresponds to our record-level covariate x.

Independent mean and covariance groups contain 960 people each. A fixed 768/192 split selects $\lambda \in \{.03, .3, 3\}$ before refitting on all 960. Let $D_z \in \mathbb{R}^{5 \times 7}$ and $D_x \in \mathbb{R}^{2 \times 4}$ be the second-difference matrices. With $P_0 = I_4 \otimes D_z^\top D_z + D_x^\top D_x \otimes I_7$ and $P = \mathrm{diag}(P_0, P_0)$, the mean fit is penalized least squares. For person i, let $Q_i \in \mathbb{R}^{8 \times 56}$ stack rows $q_{A_{it}}(U_{it})^\top$, and set $r_i = Y_i - Q_i \widehat{\theta}_\mu^{\mathrm{CD}} \in \mathbb{R}^8$. The covariance criterion is

$$\mathcal{L}_\lambda(C, \rho) = \sum_i \{ \log\det V_i + r_i^\top V_i^{-1} r_i \} + \lambda \, \mathrm{tr}(C^\top P C), \quad V_i = Q_i C C^\top Q_i^\top + (10^{-5} + e^\rho) I_8. \tag{12}$$

Mean validation uses RSS and covariance validation uses unpenalized quasi-NLL. Appendix H gives the two-pass optimizer, acceptance checks, and penalty scaling. These diagnostics do not establish a global optimum or an estimation rate.

For joint FPCA, stack both arms' grid loadings in $L \in \mathbb{R}^{210\times 4}$, set $W = I_{210}/210$, and diagonalize

$$L^\top W L = V\Lambda V^\top, \qquad \widehat{\phi}_k(a,u_g) = \widehat{\ell}_a(u_g)^\top v_k/\sqrt{\lambda_k}. \tag{13}$$

Eigenvalues below 10^{-10} are numerical zeros. The retained rank M is the smallest joint rank reaching 95% FVE; ties identify only an eigenspace. The standardized factors and FPCA scores differ by $\xi_i^{\mathrm{CD}} = \Lambda^{1/2}V^\top\omega_i$.

Only history records 1–2 are used for a new person. Here Q_i^H extracts the two corresponding rows of the person's design matrix Q_i. If $L_i^H = Q_i^H\widehat{C}$ and $r_i^H = y_i^H - Q_i^H\widehat{\theta}_\mu$, then the working law, conditional on the fixed design $\mathcal{D}_i$ and frozen fitted parameters, is $r_i^H = L_i^H\omega_i + e_i^H$, where $E_{\mathsf{W}}(e_i^H \mid \mathcal{D}_i) = 0$, $\mathrm{Cov}_{\mathsf{W}}(e_i^H \mid \mathcal{D}_i) = \sigma^2_{\mathrm{CD}}I_2$, $\mathrm{Cov}_{\mathsf{W}}(\omega_i, e_i^H \mid \mathcal{D}_i) = 0$, and $\sigma^2_{\mathrm{CD}} > 0$. Substituting fitted values defines the predictor; this does not assert that the actual DGP obeys the working law. Then

$$\widehat{m}_{\omega,i} = L_i^{H\top}(L_i^HL_i^{H\top} + \widehat{\sigma}^2_{\mathrm{CD}}I_2)^{-1}r_i^H, \quad \widehat{S}_{\omega,i} = I_4 - L_i^{H\top}(L_i^HL_i^{H\top} + \widehat{\sigma}^2_{\mathrm{CD}}I_2)^{-1}L_i^H. \tag{14}$$

Here $\widehat{\sigma}^2_{\mathrm{CD}}$ is the fitted counterpart of the working variance σ^2_{CD}. Writing $\widehat{d}(u) = \widehat{\ell}_1(u) - \widehat{\ell}_0(u)$ and $\Pi_M = V_MV_M^\top$, define $\widehat{\mu}^{\mathrm{CD}}_\Delta(u) = \{q_1(u) - q_0(u)\}^\top\widehat{\theta}^{\mathrm{CD}}_\mu$. The frozen center and scale are

$$\widehat{c}^{\mathrm{CD}}_{i,M}(u) = \widehat{\mu}^{\mathrm{CD}}_\Delta(u) + \widehat{d}(u)^\top\Pi_M\widehat{m}_{\omega,i}, \quad \widehat{s}_i^{\mathrm{CD}}(u) = \max\{.03, [\widehat{d}(u)^\top\widehat{S}_{\omega,i}\widehat{d}(u)]_+^{1/2}\}. \tag{15}$$

The scale retains all four working factors and adds no future measurement noise. Under the working moment model these are linear-prediction quantities; only an additional joint conditional Gaussian assumption gives a posterior interpretation.

Recovery records 7–518 produce a point label from separate arm-specific GCV fits, using a cubic spline in z plus orthonormal linear and quadratic terms in x, without an $x \times z$ interaction. Appendix H specifies the candidate dimensions and failure rules. With $m = 399$, $\gamma_{\mathrm{CD}} = .05$, and $k = 380$, define

$$R_i^{\mathrm{CD}} = \max_{u\in\mathcal{G}} \frac{|\widetilde{\tau}_i^{\mathrm{GCV}}(u) - \widehat{c}^{\mathrm{CD}}_{i,M}(u)|}{\widehat{s}_i^{\mathrm{CD}}(u)}, \qquad q_m^{\mathrm{CD}} = R^{\mathrm{CD}}_{(380)}, \tag{16}$$

$$\mathcal{C}^{\mathrm{CD}}_{m+1}(u) = [\widehat{c}^{\mathrm{CD}}_{m+1,M}(u) \pm q_m^{\mathrm{CD}}\widehat{s}^{\mathrm{CD}}_{m+1}(u)]. \tag{17}$$

The GCV-label band is assessed empirically. Replacing the calibration labels by the exact latent effects gives the oracle branch covered by Theorem 3.

3 Coverage guarantees

The population theorem retains its independent-subject score and approximation conditions. The personal conformal theorem uses the frozen FPCA predictor, exchangeable subject records and valid recovery sets. Model-based Gaussian bands have a separate statement in Appendix B.1. Complete conditions and proofs are given in the corresponding appendices.

The following continuum result concerns the centered additive spline construction under its stated primitive conditions; it is not a theorem for every choice of dictionary in the general mean model.

Theorem 1 (Population simultaneous confidence band)**.** *Under Assumptions 1–8, with the fixed-degree spline and the Appendix A grid rates in* (31) *and* (32)*, for any fixed $\alpha \in (0,1)$, the band in* (3) *satisfies*

$$\mathbb{P}\{\tau(x) \in \mathcal{C}^{\mathrm{pop}}_{1-\alpha}(x) \text{ for every } x \in \mathcal{X}\} \longrightarrow 1-\alpha.$$

The proof first linearizes the working-independence estimator, establishes uniform consistency of the CR3 covariance, compares the non-Gaussian maximum with its Gaussian counterpart, and absorbs studentization, grid, and sieve errors by anti-concentration. The direct regression target is τ_K; coverage for τ uses the explicit approximation condition relative to this projection. For fixed dictionaries and finite grids, Corollary 1 states a separate result. Neither statement establishes a causal interpretation from marginal centering alone. The result does not require the working covariance model to be correct.

Theorem 2 (Personal FPCA conformal band)**.** *Conditional on the fitting sample, assume the augmented calibration and future subject records are exchangeable, with the same prespecified history, recovery and score rules. Assume the exact effect dictionary above and measurable suprema over $\mathcal{U}$ (or maxima if $\mathcal{U}$ is the evaluation grid). Assume recovery sets with*

$$\mathbb{P}\{\zeta_{m+1} \notin \mathcal{E}_{m+1} \mid \mathcal{D}_{\mathrm{fit}}\} \le \eta_{\mathrm{rec}}.$$

Freeze the fitted FPCA rule and positive scales before calibration, and choose $\gamma_{\mathrm{rk}} + \eta_{\mathrm{rec}} \le \gamma < 1$. Then

$$\mathbb{P}\{\tau_{m+1}(u) \in \mathcal{C}^{\mathrm{FPCA}}_{m+1}(u)\ \forall u \in \mathcal{U} \mid \mathcal{D}_{\mathrm{fit}}\} \ge 1 - \gamma_{\mathrm{rk}} - \eta_{\mathrm{rec}}.$$

Proof. The common feasible-score rule makes $R^+_1, \dots, R^+_{m+1}$ exchangeable on the probability space with the appended future recovery block. The corrected rank gives $\mathbb{P}(R^+_{m+1} \le q^+ \mid \mathcal{D}_{\mathrm{fit}}) \ge 1 - \gamma_{\mathrm{rk}}$, including ties and the infinite-quantile case. On $\{\zeta_{m+1} \in \mathcal{E}_{m+1}\}$, the ellipsoid support bound makes R^+_{m+1} at least the future standardized error about the full center $\widehat{c}_{m+1,D}$. A union bound therefore covers the full-center band $[\widehat{c}_{m+1,D}(u) \pm q^+ \widehat{s}_{m+1}(u)]$ with probability at least $1 - \gamma_{\mathrm{rk}} - \eta_{\mathrm{rec}}$. Finally, (9) contains that full-center band pointwise because $\widehat{c}_{i,D} = \widehat{c}_{i,M} + \widehat{\delta}_{i,M}$. □

Only the future recovery-set failure enters this bound; the calibration sets form feasible scores without a simultaneous set-coverage requirement. The recovery-set guarantee is a substantive condition, not a consequence of fitting FPCA. Appendix C gives sufficient randomization and envelope conditions and their failure budgets. An approximate effect dictionary additionally needs both recovery-regression and prediction-domain approximation envelopes. The conformal argument accepts a fitted linear predictor without a Gaussian score law; Gaussianity is used for the separate model-based reference bands in Appendix B.1. Those bands have known-parameter guarantees and, in an exact fixed finite model, fitted asymptotic validity. Their simulated Gaussian critical values are not subject-level conformal ranks.

The protected experiment instantiates the theorem above with rank error .04 and recovery error .01 (rank 384 when $m = 399$). The direct CD branch below instead uses rank error .05 (rank 380) and has an exact-label guarantee only.

3.1 Guarantees and limits for the direct CD branch

The statements here concern the finite grid $\mathcal{G}$ and the direct score in (16). They do not modify the population SCB theorem or the guarded recovery-set theorem.

Proposition 1 (PSD kernel and weighted FPCA)**.** *For any finite collection (a_j, u_j), the matrix with entries $K^{\mathrm{CD}}_{a_j a_k}(u_j, u_k)$ is positive semidefinite. For every retained positive eigenvalue, the functions in* (13) *are orthonormal under the discrete two-arm inner product induced by W. The kernel and every full-factor prediction expressed through $L\omega$ are invariant to an orthogonal rotation of the four factor coordinates. Individual eigenfunctions are not identified within a tied eigenspace. If a tie crosses the FVE cutoff, the implementation uses its fixed measurable eigensolver ordering; the resulting truncated center need not be rotation invariant.*

Proposition 2 (Working linear predictor)**.** *Conditional on the fixed design and frozen fitted parameters, under the working factor moments in* (10) *and the exact centered history-residual law stated before* (14)*,* (14) *is the best linear predictor of ω_i and its linear prediction-error covariance. The Woodbury identity gives the equivalent precision form $(I_4 + L_i^{H\top} L_i^H / \sigma^2_{\mathrm{CD}})^{-1} L_i^{H\top} r_i^H / \sigma^2_{\mathrm{CD}}$, and its error covariance is $(I_4 + L_i^{H\top} L_i^H / \sigma^2_{\mathrm{CD}})^{-1}$. If, additionally, (ω_i, e_i) is jointly Gaussian conditional on design, these quantities are the conditional mean and covariance. No Gaussian posterior claim is made otherwise.*

Let the oracle score $S^{o,\mathrm{CD}}_i$ replace $\widetilde{\tau}^{\mathrm{GCV}}_i$ in (16) by the full latent τ_i.

Theorem 3 (Exact-label finite-grid rank guarantee)**.** *Conditional on the independent fitting data and a successful fit, suppose the m calibration subjects and one future subject are exchangeable and use the same frozen, measurable history-extraction, center, scale, grid, and exact-label score rules. Assume every center is finite and every scale is finite and strictly positive. Let $k = \lceil (m+1)(1 - \gamma_{\mathrm{CD}}) \rceil$ and take the kth of the m oracle scores, with an infinite cutoff if $k > m$. Then*

$$\Pr\{\tau_{m+1}(u) \in [\widehat{c}^{\mathrm{CD}}_{m+1,M}(u) \pm q^{o,\mathrm{CD}}_m \widehat{s}^{\mathrm{CD}}_{m+1}(u)]\ \forall u \in \mathcal{G} \mid \mathcal{D}_{\mathrm{fit}}\} \ge 1 - \gamma_{\mathrm{CD}}.$$

If the fit fails, define the reported band as the whole real line at every grid point; if $k > m$, use the same convention. The probability is marginal over the future subject and history. It is not conditional-on-history coverage, a continuum result, or simultaneous coverage of 32 future people.

Proposition 3 (Recovered-label gap)**.** *Define* $\Delta_i = \max_{u\in\mathcal{G}} |\widetilde{\tau}_i^{\mathrm{GCV}}(u) - \tau_i(u)|/\widehat{s}_i^{\mathrm{CD}}(u)$. *Then*

$$|R_i^{\mathrm{CD}} - S_i^{\mathrm{o,CD}}| \le \Delta_i.$$

Under the assumptions of Theorem 3, on a successful fit let q_S and q_R denote the kth calibration order statistics of the oracle and recovered scores, respectively. Set $\Delta_i = +\infty$ if calibration label i fails, and otherwise use the displayed finite-label definition. For a fixed finite $\epsilon \ge 0$, let $E_\epsilon = \{\max_{i\le m} \Delta_i \le \epsilon\}$ and suppose $\Pr(E_\epsilon^{\mathrm{c}} \mid \mathcal{D}_{\mathrm{fit}}) \le \eta$. For $k \le m$, on E_ϵ we have $q_S \le q_R + \epsilon$. The band with cutoff $q_R + \epsilon$ therefore has conditional coverage at least $1 - \gamma_{\mathrm{CD}} - \eta$. Failed calibration labels are assigned $R_i = +\infty$; the absolute-difference claim is restricted to finite successful labels. If $k > m$ or the fit fails, the reported band is the whole real line at every grid point. This one-ϵ inflation is not implemented, and no such event has been proved for the reported GCV labels.

Proposition 4 (Counterexample: unbiased point labels need not give latent coverage)**.** *At a one-point grid, set $\tau_i = 1$, $c_i = 0$, and $s_i = 1$. Independently across people, let $\widetilde{\tau}_i = 0$ with probability .99 and 100 with probability .01. Then $E\widetilde{\tau}_i = 1 = \tau_i$; the label error is -1 or 99 and has mean zero. Nevertheless, $R_i = |\widetilde{\tau}_i|$ is 0 or 100 while the oracle score is $S_i = 1$. For $m = 399$ and $k = 380$, if $N \sim \mathrm{Binomial}(399, .01)$ counts labels equal to 100, then $q_R = 100$ exactly when $N \ge 20$ and otherwise $q_R = 0$. Hence the reported band $[-q_R, q_R]$ covers the target exactly when $N \ge 20$, and $\Pr(N \ge 20) \le E(N)/20 = 3.99/20 = .1995$ by Markov's inequality. The number .1995 is an upper bound, not the exact binomial tail. Thus unbiased point labels do not imply latent conformal coverage.*

Remark 1. *Separately, cross-arm covariance is unidentified when each person is observed in only one arm; a finite grid does not imply continuum coverage; and an FVE threshold needs a fixed measurable rule when an eigenvalue tie crosses its cutoff.*

4 Experiments

The population experiment evaluates simultaneous coverage of τ; the personal experiments evaluate the full latent curve. The complete experimental designs, tables, and replication-level MCSE definitions are provided in Appendices D–G, H, and J.

4.1 Population band

With 60 dependent visits per subject, the subject-block CR3 band approaches its known-covariance benchmark: simultaneous coverage is .923, .932, and .962 at $n = 100, 200, 400$. In the nonparametric design at $n = 200$, coverage is .840, .952, .932, .868 for basis sizes $K = 6, 8, 10, 12$, showing bias for a small basis and deteriorating studentization for a large one. A fitted functional precision shortens the band but undercovers; working independence and a diagonal precision are nearly identical.

4.2 Personal FPCA prediction

The model-based experiment has four learned joint FPC directions, independent mean and covariance fits of size $n = 240$ or 960, and histories of $0, 2$, or 6 visits. At $n = 960$ under Gaussian effects, the truncation-adjusted band has coverage .9414, .9466, .9395 and widths $4.0311, 3.1655, 2.1982$, respectively. At six visits, naive and oracle truncation cover only .7760 and .8190, supporting retention of omitted directions.

At two visits, adjusted coverage rises from .9180 (MCSE .0067) to .9466 (.0056) as n rises from 240 to 960 under Gaussian scores; the corresponding elliptical-t_5 values are .8913 (.0080) and .9323 (.0064). The theorem is asymptotic, and the non-Gaussian study is a robustness diagnostic. These model-based bands have no conformal calibration sample and are distinct from the end-to-end procedure below.

4.3 END-TO-END CONFORMAL PREDICTION

The end-to-end study adds $m = 199$ or 399 calibration people and $r = 32$ or 128 recovery visits. It uses the exact four-dimensional recovery dictionary, an error budget of .04 for conformal rank and .01 for recovery, and either the bounded-noise radius of Proposition 7 or the exact-F radius of Proposition 8. Appendix F documents all 240 design–method cells and the disjoint development stage.

At $n = 960$, $m = 399$, $r = 128$, and two history visits, guarded coverage is 1.0000 for Gaussian effects and .9993 (MCSE .0007) for elliptical-t_5 effects, with widths 6.932 and 7.178. The infeasible latent-label benchmark has coverage .9499 and .9544 and widths 3.337 and 3.547; with $r = 32$, guarded widths increase to 18.440 and 18.702. Exact-F recovery reduces paired width by 13.4–16.2%, but coverage remains .9980–1.0000. Appendix Tables 15 and 14 give the complete comparison. Protected bands are therefore conservative, and more recovery visits matter more than the choice between the two recovery rules.

4.4 HEARTSTEPS POPULATION ANALYSIS

The public HeartSteps V1 analysis has 5,777 decision points from 37 participants (Klasnja et al., 2019). On the 15-point grid, 5,000 multiplier draws give a critical value of 2.805 and average total width .987; estimates range from $-.021$ to .319, and every interval contains zero. These are descriptive projections under the observed visit distribution and working dictionary, not established causal excursion effects.

M0 with working independence remains primary. M2 and M3 shorten its band by about 19%, but in the paired template simulation their full-surface coverage falls to .737–.742 when the omitted slot interaction is present, despite about .917 projection coverage. Appendix Tables 9 and 10 report all sensitivity fits.

In a strictly nested semi-synthetic experiment with $N \in \{37, 100, 200\}$, holdout coverage is .938, .938, .940 while width decreases from .91933 to .55482 and .39237; the full-replication coverage values are .936, .926, .945. Increasing independent subjects improves precision in this frozen mechanism, but does not prove coverage equivalence or recruitment gains under new covariate support. Appendix Tables 6– 8 give complete results; Appendix Figure 1 gives the leave-one-participant-out diagnostic.

Direct CD-FPCA confirmation. Across 12 cells, GCV-label simultaneous coverage is .9495–.9683 for the rank-four CD predictor. A structure-matched rank-eight negative control is narrower with slightly higher coverage; the DGP has no genuinely x-varying covariance. Appendix J gives the complete results.

5 DISCUSSION

The population band covers a mean treatment-effect surface, whereas the personalized band predicts a future subject's latent conditional-mean effect curve; a factual-outcome band answers neither target.

The personal procedure combines pooled FPCA, raw-history score prediction, and conformal calibration. Longer histories improve precision, but protected bands remain conservative; exact-F recovery shortens them, while additional recovery visits matter more. Both recovery rules require a complete exogenous design and conditionally spherical Gaussian innovations.

The HeartSteps band is a descriptive projection based on 37 participants. In the nested $N = 37/100/200$ experiment, widths track standard errors while holdout coverage is .938, .938, .940; the frozen templates do not prove coverage equivalence or recruitment gains. Restricted dictionaries lose full-surface coverage under the omitted interaction and remain sensitivities.

The CD-FPCA confirmation has no width advantage over the structure-matched negative control, and its DGP lacks genuinely x-varying covariance. Honest interaction recovery and continuous-domain direct bands remain open.

AI USE STATEMENT

Generative AI tools were used to reorganize the LaTeX source, edit prose, and check notation and proof arguments, and draft code for the raw-history FPCA simulation. Simulated observations are generated by the documented random-number program, not language-model sampling; the original HeartSteps observations were not altered. The authors are responsible for verifying the resulting text, mathematical arguments, and source files before submission and for the final content of the paper.

ETHICS STATEMENT

The HeartSteps analysis uses a public, de-identified data release and is reported only as an illustration of population-level inference. The proposed personalized bands should not be used to make clinical decisions without a prospectively specified recovery protocol and an assessment of the assumptions required by the selected band.

REPRODUCIBILITY STATEMENT

The supplementary material documents experimental designs, tuning, replication counts and numerical checks. The accompanying source package provides code, configurations, required inputs and manual run instructions; its README maps studies to paper results and states historical reproduction limits. Full historical simulation archives are not included. Assumptions and proofs appear in Appendices A–C and I.

REFERENCES

Abbasi-Yadkori, Y., Pál, D., and Szepesvári, C. (2011). Improved algorithms for linear stochastic bandits. *Advances in Neural Information Processing Systems*, 24, 2312–2320.

Alaa, A. M., Ahmad, Z., and van der Laan, M. (2023). Conformal meta-learners for predictive inference of individual treatment effects. *Advances in Neural Information Processing Systems*, 36.

Anderson, T. W. (1955). The integral of a symmetric unimodal function over a symmetric convex set and some probability inequalities. *Proceedings of the American Mathematical Society*, 6, 170–176.

Batra, D., Mercuri, S., and Khraishi, R. (2023). Conformal predictions for longitudinal data. *arXiv preprint arXiv:2310.02863*.

Chernozhukov, V., Chetverikov, D., and Kato, K. (2013). Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *The Annals of Statistics*, 41, 2786–2819.

de Boor, C. (2001). *A Practical Guide to Splines*, revised edition. Springer, New York.

Degras, D. A. (2011). Simultaneous confidence bands for nonparametric regression with functional data. *Statistica Sinica*, 21, 1735–1765.

Diquigiovanni, J., Fontana, M., and Vantini, S. (2025). The importance of being a band: finite-sample exact distribution-free prediction sets for functional data. *Statistica Sinica*, 35, 853–871.

Jiang, C.-R. and Wang, J.-L. (2010). Covariate adjusted functional principal components analysis for longitudinal data. *The Annals of Statistics*, 38, 1194–1226.

Klasnja, P., Smith, S., Seewald, N. J., Lee, A., Hall, K., Luers, B., Hekler, E. B., and Murphy, S. A. (2019). Efficacy of contextually tailored suggestions for physical activity: A micro-randomized optimization trial of HeartSteps. *Annals of Behavioral Medicine*, 53, 573–582.

Lee, S., Okui, R., and Whang, Y.-J. (2017). Doubly robust uniform confidence band for the conditional average treatment effect function. *Journal of Applied Econometrics*, 32, 1207–1225.

Lei, J., G'Sell, M., Rinaldo, A., Tibshirani, R. J., and Wasserman, L. (2018). Distribution-free predictive inference for regression. *Journal of the American Statistical Association*, 113, 1094–1111.

Lei, L. and Candès, E. J. (2021). Conformal inference of counterfactuals and individual treatment effects. *Journal of the Royal Statistical Society: Series B*, 83, 911–938.

Liebl, D. and Reimherr, M. (2023). Fast and fair simultaneous confidence bands for functional parameters. *Journal of the Royal Statistical Society: Series B*, 85, 842–868.

Ma, S., Yang, L., and Carroll, R. J. (2012). A simultaneous confidence band for sparse longitudinal regression. *Statistica Sinica*, 22, 95–122.

Semenova, V. and Chernozhukov, V. (2021). Debiased machine learning of conditional average treatment effects and other causal functions. *The Econometrics Journal*, 24, 264–289.

Vovk, V., Gammerman, A., and Shafer, G. (2005). *Algorithmic Learning in a Random World*. Springer.

Yao, F., Müller, H.-G., and Wang, J.-L. (2005). Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association*, 100, 577–590.

Zheng, S., Yang, L., and Härdle, W. K. (2014). A smooth simultaneous confidence corridor for the mean of sparse functional data. *Journal of the American Statistical Association*, 109, 661–673.

Bose, S. and Dempsey, W. (2025). Prediction intervals for individual treatment effects in a multiple decision point framework using conformal inference. *arXiv preprint arXiv:2512.08828*.

Javanmardi, A., Sale, Y., Hofman, P., and Hüllermeier, E. (2023). Conformal prediction with partially labeled data. *Proceedings of Machine Learning Research*, 204, 251–266.

Zheng, M., Zhou, W., and Zhu, S. (2026). Beyond prediction: Conformal inference for latent distributional parameters. *arXiv preprint arXiv:2608.03607*.

Ding, F., He, S., Jones, D. E., and Huang, J. Z. (2022). Functional PCA with covariate dependent mean and covariance structure. *Technometrics*, 64(3), 335–345. doi:10.1080/00401706.2021.2008502. The supplied arXiv:2001.11425v2 was revised on 19 August 2023.

A Population-band details

Using the model in Section 2.1, this appendix supplies the series-projection details, technical assumptions, auxiliary results, and proofs for Theorem 1.

A.1 Series approximation and covariance representation

We use the model and inferential targets in Section 2.1, equations (1) and (2). For the population confidence band, treatment assignment must satisfy the estimating equation in Assumption 1. The definitions below specify the series projection and the quantities used in the proofs.

Functional covariance representation. The processes g_{ia} have mean zero. For each arm, the spectral decomposition of its covariance kernel gives

$$g_{ia}(z) = \sum_{m \geq 1} \xi^a_{im} \varphi^a_m(z), \qquad \mathbb{E}(\xi^a_{im}) = 0, \qquad \mathrm{Cov}(\xi^a_{im}, \xi^a_{i\ell}) = \lambda_{am} \mathbf{1}(m = \ell). \tag{18}$$

Cross-arm score covariances may be nonzero; in particular, $\mathrm{Cov}(\xi^1_{im}, \xi^0_{im}) = \lambda^{10}_m$ encodes dependence between the two potential functional deviations. A finite joint representation of the two arm-specific processes will supply the working predictor for the latent subject-specific contrast. The population SCB result itself does not use this working covariance model. The population SCB validity theory below does not require estimating the eigenfunctions, eigenvalues, or cross-action covariance. It only uses independence across subjects and moment conditions on the resulting subject score.

Centered additive sieve This subsection specifies the additive dictionary used for the population construction and its spline primitive conditions. General model means may have nonzero projection remainders; coverage of the original mean contrast requires the approximation conditions stated below.

Let $b_{\ell K}(x_\ell) \in \mathbb{R}^K$ be a fixed, linearly independent centered basis. For splines, start with $K+1$ raw basis functions collected in $q_{\ell K}$, which sum to one. Under a fixed reference measure with density bounded above and away from zero, let $m_{\ell K} = \int q_{\ell K}$ and $C_{\ell K} = \int (q_{\ell K} - m_{\ell K})(q_{\ell K} - m_{\ell K})^\top$. Write its positive spectral decomposition as $C_{\ell K} = U_{\ell K}\Lambda_{\ell K}U_{\ell K}^\top$, where $U_{\ell K}$ has K orthonormal columns, and set

$$b_{\ell K}(x_\ell) = \Lambda_{\ell K}^{-1/2} U_{\ell K}^\top \{q_{\ell K}(x_\ell) - m_{\ell K}\}.$$

This removes the constant direction before whitening; its reference Gram matrix is I_K. Centering all $K+1$ raw functions and retaining them would leave a linear dependency. The reference measure and knots are fixed in advance; a basis trained on independent fitting data may be conditioned upon. Define

$$B_K(x) = \{1, b_{1K}(x_1)^\top, \ldots, b_{dK}(x_d)^\top\}^\top, \qquad p_K = 1 + dK.$$

For $a \in \{0,1\}$, write

$$\mu_a(x) = B_K(x)^\top \theta_{aK} + r_{aK}(x), \qquad \beta_K = (\theta_{1K}^\top, \theta_{0K}^\top)^\top. \tag{19}$$

Throughout, the coefficients in (19) are chosen so that their stacked vector is the population projection in Assumption 1; the remainders are defined relative to that projection. Thus the sieve target below and the estimating-equation target are the same object. This definition alone does not make r_{aK} small. For interpretation as the model mean, impose the additional score condition

$$\mathbb{E}\left[\sum_t d_{it}\{g_{i,A_{it}}(Z_{it}) + \varepsilon_{it}\}\right] = 0.$$

For example, zero conditional mean of the bracketed residual given the complete visit/treatment design is sufficient. Marginal centering of $g_{ia}(z)$ is not sufficient when visits, covariates, or assignment depend on subject history. Without this score condition and the explicit approximation bound below, inference concerns τ_K, the unweighted visit-distribution projection contrast, not a causal mean surface. Set

$$c_K(x) = \{B_K(x)^\top, -B_K(x)^\top\}^\top, \quad \tau_K(x) = c_K(x)^\top \beta_K, \quad b_K = \sup_{x \in \mathcal{X}} |\tau(x) - \tau_K(x)|.$$

For a visit, define the $2p_K$-vector

$$d_{it} = \{A_{it}B_K(X_{it})^\top, (1 - A_{it})B_K(X_{it})^\top\}^\top.$$

Let D_i be the $T_i \times 2p_K$ matrix with rows $d_{it}^\top$, and let $Y_i = (Y_{i1}, \ldots, Y_{iT_i})^\top$.

A.2 Covariance-robust estimation and calibration

Let $G_n = \sum_{i=1}^n D_i^\top D_i$. The main estimator is

$$\widehat{\beta} = G_n^{-1} \sum_{i=1}^n D_i^\top Y_i, \qquad \widehat{\tau}(x) = c_K(x)^\top \widehat{\beta}. \tag{20}$$

The word "independence" refers only to the working precision in (20); the true vector $u_i = Y_i - D_i\beta_K$ may have an arbitrary covariance matrix Σ_i.

Let

$$\widehat{e}_i = Y_i - D_i\widehat{\beta}, \qquad P_i = D_i G_n^{-1} D_i^\top, \qquad \widetilde{e}_i = (I_{T_i} - P_i)^{-1}\widehat{e}_i,$$

and define the adjusted cluster score $\widetilde{g}_i = D_i^\top \widetilde{e}_i$. Thus the correction is a subject-block CR3 correction, not an observationwise diagonal correction. Set

$$\widehat{A}_n = n^{-1}G_n, \qquad \widehat{\Omega}_n = n^{-1}\sum_{i=1}^n \widetilde{g}_i\widetilde{g}_i^\top, \qquad \widehat{V}_\beta = n^{-1}\widehat{A}_n^{-1}\widehat{\Omega}_n\widehat{A}_n^{-1}. \tag{21}$$

Thus $\widehat{V}_\beta = n^{-1}\widehat{V}_K$, where $\widehat{V}_K = \widehat{A}_n^{-1}\widehat{\Omega}_n\widehat{A}_n^{-1}$ is the unscaled sandwich matrix used in the asymptotic arguments below. The pointwise standard error is

$$\widehat{s}^2(x) = c_K(x)^\top \widehat{V}_\beta c_K(x) = n^{-1}\widehat{\sigma}_K^2(x). \tag{22}$$

Equivalently, $\widehat{s}^2(x) = c_K(x)^\top G_n^{-1}(\sum_i \widetilde{g}_i\widetilde{g}_i^\top)G_n^{-1}c_K(x)$, which is the normalization used in the simulation code.

Conditionally on the data, generate independent $\xi_i \sim N(0,1)$ and set

$$\mathbb{T}_n^*(x) = \frac{n^{-1/2}\sum_{i=1}^n \xi_i c_K(x)^\top \widehat{A}_n^{-1}\widetilde{g}_i}{\widehat{\sigma}_K(x)}. \tag{23}$$

For a grid $\mathcal{X}_n = \{x_1, \dots, x_{M_n}\}$, let

$$T_n^* = \max_{x\in\mathcal{X}_n} |\mathbb{T}_n^*(x)|, \qquad c_{1-\alpha}^* = q_{1-\alpha}(T_n^* \mid \mathcal{D}_n).$$

Here $\mathcal{D}_n = \{(Y_i, D_i) : i = 1, \dots, n\}$ is the collection of observed subject clusters, and $\mathbb{P}^*(\cdot) = \mathbb{P}(\cdot \mid \mathcal{D}_n)$ denotes conditional probability given these clusters. This sample is distinct from the fitting sample $\mathcal{D}_{\text{fit}}$ used by the latent-effect procedure. The proposed band is

$$\mathcal{C}_{1-\alpha}(x) = \left[\widehat{\tau}(x) - c_{1-\alpha}^*\widehat{s}(x), \widehat{\tau}(x) + c_{1-\alpha}^*\widehat{s}(x)\right]. \tag{24}$$

A.3 Fixed-terminal-time theory

All sequences below may depend on n. Write $A_K = \mathbb{E}(D_i^\top D_i)$, $g_i = D_i^\top u_i$, $\Omega_K = \mathbb{E}(g_i g_i^\top)$, and

$$\psi_i(x) = c_K(x)^\top A_K^{-1} g_i, \qquad \sigma_K^2(x) = \mathbb{E}\{\psi_i(x)^2\}. \tag{25}$$

Assumption 1 (Independent clusters and projection target)**.** *For every n, the clusters (Y_i, D_i) are row-wise i.i.d.; $\max_i T_i \le \bar{T} < \infty$; β_K is the unique solution of $\mathbb{E}\{D_i^\top(Y_i - D_i\beta)\} = 0$; and $\mathbb{E}g_i = 0$. The estimators are defined on $\{\lambda_{\min}(\widehat{A}_n) > c_A/2,\ h_n < 1/2\}$, with an arbitrary measurable convention on its complement. The subsequent assumptions imply that this event has probability tending to one.*

Assumption 2 (Gram matrix, envelope, and variance)**.** *For constants $0 < c_A < C_A < \infty$, $c_A \le \lambda_{\min}(A_K) \le \lambda_{\max}(A_K) \le C_A$. Let $\zeta_K = \sup_{x\in\mathcal{X}} \|B_K(x)\|_2$. Also, $\underline{\sigma}_K = \inf_{x\in\mathcal{X}} \sigma_K(x) > 0$. The deterministic envelopes ζ_K and $\underline{\sigma}_K$ may vary with n. Write $\omega_K = \|\Omega_K\|_{\text{op}}$.*

Assumption 3 (First-stage rates)**.** *For deterministic $a_n \to 0$ and $q > 4$,*

$$\|\widehat{A}_n - A_K\|_{\text{op}} = O_p(a_n), \quad \left\|n^{-1/2}\sum_{i=1}^n g_i\right\|_2 = O_p(\sqrt{p_K}), \quad \sup_n \mathbb{E}\left(\frac{\|g_i\|_2}{\sqrt{p_K}}\right)^q < \infty,$$

and $\zeta_K a_n\sqrt{p_K}/\underline{\sigma}_K = o(1)$. For bounded cluster sizes and bounded basis functions, the first display follows from standard matrix concentration under the corresponding moment envelope; we state the rate explicitly because it is the quantity used below.

Assumption 4 (High-level CR3 meat concentration)**.** *With $h_n = \max_i \|P_i\|_{\text{op}}$, $h_n = o_p(1)$. Moreover,*

$$\left\|n^{-1}\sum_{i=1}^n(\widetilde{g}_i\widetilde{g}_i^\top - g_i g_i^\top)\right\|_{\text{op}} = o_p(v_n), \quad \left\|n^{-1}\sum_{i=1}^n g_i g_i^\top - \Omega_K\right\|_{\text{op}} = o_p(v_n),$$

where $v_n > 0$ is deterministic and

$$\frac{\zeta_K^2\{v_n + a_n(1+\omega_K)\}}{\underline{\sigma}_K^2} = o(1).$$

Lemma 1 (Primitive sufficient rates for the CR3 meat)**.** *Suppose, in addition to Assumptions 1–3, that for deterministic $\mathcal{D}_K, U_n \ge 1$,*

$$\max_i \|D_i\|_{\text{op}} \le \mathcal{D}_K, \qquad \max_i \|u_i\|_2 \le U_n$$

almost surely within each row of the triangular array. Let

$$\bar{h}_n = \frac{2\mathcal{D}_K^2}{nc_A}, \qquad r_{g,n} = \mathcal{D}_K U_n \bar{h}_n + \mathcal{D}_K^2\sqrt{p_K/n},$$

and assume $\bar{h}_n = o(1)$. *Then*

$$\left\| n^{-1} \sum_i (\widetilde{g}_i \widetilde{g}_i^\top - g_i g_i^\top) \right\|_{\mathrm{op}} = O_p\{r_{g,n} \mathcal{D}_K U_n + r_{g,n}^2\},$$

$$\left\| n^{-1} \sum_i g_i g_i^\top - \Omega_K \right\|_{\mathrm{op}} = O_p \left\{ \mathcal{D}_K^2 U_n^2 \left(\sqrt{\frac{\log(2p_K)}{n}} + \frac{\log(2p_K)}{n} \right) \right\}.$$

Thus Assumption 4 follows whenever both displayed deterministic rates are $o(v_n)$ *and* v_n *satisfies its final compatibility condition.*

Proof. By Assumptions 2–3, the event $\{\lambda_{\min}(\widehat{A}_n) > c_A/2\}$ has probability tending to one. On this event, $\|G_n^{-1}\|_{\mathrm{op}} \le 2/(nc_A)$, hence $h_n \le \bar{h}_n$. The exact decomposition

$$\widetilde{g}_i - g_i = D_i^\top \{(I - P_i)^{-1} - I\} u_i - D_i^\top (I - P_i)^{-1} D_i (\widehat{\beta} - \beta_K)$$

and $\|(I - P_i)^{-1}\|_{\mathrm{op}} \le (1 - \bar{h}_n)^{-1}$, together with $\|\widehat{\beta} - \beta_K\|_2 = O_p(\sqrt{p_K/n})$, give $\max_i \|\widetilde{g}_i - g_i\|_2 = O_p(r_{g,n})$; the omitted $(1 - \bar{h}_n)^{-1}$ is bounded. Since $\max_i \|g_i\|_2 \le \mathcal{D}_K U_n$, putting $\Delta_i = \widetilde{g}_i - g_i$ gives

$$\left\| n^{-1} \sum_i (\widetilde{g}_i \widetilde{g}_i^\top - g_i g_i^\top) \right\|_{\mathrm{op}} \le 2 \max_i \|g_i\|_2 \max_i \|\Delta_i\|_2 + \max_i \|\Delta_i\|_2^2.$$

This proves the first bound. The second follows from matrix Bernstein applied to the row-wise independent, centered self-adjoint matrices $g_i g_i^\top - \mathbb{E}(g_i g_i^\top)$, whose operator norms are bounded by a constant multiple of $\mathcal{D}_K^2 U_n^2$. This proves the lemma. □

Lemma 2 (Sub-Gaussian transfer)**.** *Suppose the design envelope in Lemma 1 holds almost surely and, for constants* $K_u < \infty$ *and* $\eta > 0$,

$$\mathbb{P}(\|u_i\|_2 > t) \le 2e^{-t^2/K_u^2}, \qquad U_n = K_u \sqrt{(2 + \eta) \log n}.$$

If

$$\mathcal{D}_K^2 \mathbb{E}\{\|u_i\|_2^2 \mathbf{1}(\|u_i\|_2 > U_n)\} = o(v_n),$$

then the two conclusions of Lemma 1 and its implication for Assumption 4 remain valid with this U_n*, provided both deterministic rates in that lemma are* $o(v_n)$ *and the final compatibility condition in Assumption 4 holds.*

Proof. Let $u_i^{(L)} = u_i \mathbf{1}(\|u_i\|_2 \le U_n)$ and $g_i^{(L)} = D_i^\top u_i^{(L)}$. These truncated clusters remain row-wise independent and satisfy the deterministic envelope required by matrix Bernstein. A union bound gives

$$\mathbb{P}\left(\max_{i \le n} \|u_i\|_2 > U_n \right) \le 2n^{-(1+\eta)} = o(1).$$

On the complementary event, the original errors themselves obey the bounded envelope, so the deterministic CR3 score-replacement argument applies to the original estimator and scores. Moreover,

$$\left\| \mathbb{E}(g_i^{(L)} g_i^{(L)\top}) - \Omega_K \right\|_{\mathrm{op}} \le \mathcal{D}_K^2 \mathbb{E}\{\|u_i\|_2^2 \mathbf{1}(\|u_i\|_2 > U_n)\} = o(v_n).$$

On the no-truncation event, the deterministic CR3 score-replacement argument in Lemma 1 applies directly, using the original first-stage rate in Assumption 3. For the true meat, apply matrix Bernstein to the independent truncated scores and use

$$\left\| n^{-1} \sum_i g_i g_i^\top - \Omega_K \right\|_{\mathrm{op}} \le \left\| n^{-1} \sum_i g_i^{(L)} g_i^{(L)\top} - \mathbb{E}(g_i^{(L)} g_i^{(L)\top}) \right\|_{\mathrm{op}} + \left\| \mathbb{E}(g_i^{(L)} g_i^{(L)\top}) - \Omega_K \right\|_{\mathrm{op}}.$$

Transfer both conclusions on the no-truncation event; its complement has probability $o(1)$, completing the proof. □

Theorem 4 (Reduction to CR3 meat concentration)**.** *Under Assumptions 1–4,*

$$\sup_{x\in\mathcal{X}_n} \frac{\left|\sqrt{n}\{\widehat{\tau}(x)-\tau_K(x)\}-n^{-1/2}\sum_{i=1}^n \psi_i(x)\right|}{\sigma_K(x)} = o_p(1), \tag{26}$$

$$\sup_{x\in\mathcal{X}_n} \left|\frac{\widehat{\sigma}_K(x)}{\sigma_K(x)}-1\right| = o_p(1). \tag{27}$$

Proof. The normal equations and the definition of β_K give the exact identity

$$\sqrt{n}(\widehat{\beta}-\beta_K) = \widehat{A}_n^{-1} n^{-1/2}\sum_{i=1}^n g_i.$$

Because the eigenvalues of A_K are bounded away from zero, the resolvent identity yields $\|\widehat{A}_n^{-1} - A_K^{-1}\|_{\mathrm{op}} = O_p(a_n)$. Consequently, uniformly in x, the absolute linearization remainder is at most

$$\sqrt{2}\,\zeta_K\,\|\widehat{A}_n^{-1}-A_K^{-1}\|_{\mathrm{op}}\left\|n^{-1/2}\sum_i g_i\right\|_2 = O_p(\zeta_K a_n\sqrt{p_K}).$$

Division by $\underline{\sigma}_K$ proves (26).

Let $V_K = A_K^{-1}\Omega_K A_K^{-1}$ and $\widehat{V}_K = \widehat{A}_n^{-1}\widehat{\Omega}_n\widehat{A}_n^{-1}$, so that $\sigma_K^2(x) = c_K(x)^\top V_K c_K(x)$ and $\widehat{\sigma}_K^2(x) = c_K(x)^\top \widehat{V}_K c_K(x)$. The resolvent identity, Assumption 4, and boundedness of A_K^{-1} imply $\|\widehat{V}_K - V_K\|_{\mathrm{op}} = o_p(v_n) + O_p\{a_n(1+\omega_K)\}$. Therefore

$$\sup_x |\widehat{\sigma}_K^2(x)-\sigma_K^2(x)| \le 2\zeta_K^2\|\widehat{V}_K - V_K\|_{\mathrm{op}} = o_p(\underline{\sigma}_K^2).$$

On the event where the last supremum is smaller than $\underline{\sigma}_K^2/2$, which has probability tending to one, $\widehat{\sigma}_K^2(x) > 0$ uniformly. Finally, $|\sqrt{a}/\sqrt{b}-1| \le |a/b-1|$ for positive a, b, which proves (27). No entropy condition on M_n enters this deterministic envelope reduction; grid complexity first enters Theorem 5. □

For Gaussianization, define $X_{ij} = \psi_i(x_j)/\sigma_K(x_j)$, $j \le M_n$, and

$$S_n = \max_{j\le M_n}\left|n^{-1/2}\sum_{i=1}^n X_{ij}\right|.$$

The absolute maximum is the maximum of the $2M_n$ coordinates $(X_{ij}, -X_{ij})$.

Assumption 5 (High-dimensional Gaussian approximation)**.** *For some $B_n \ge 1$ and constants $c_0, C_0 > 0$,*

$$\max_{j\le M_n}\mathbb{E}\exp(|X_{ij}|/B_n) \le 2, \qquad \max_j \mathbb{E}|X_{ij}|^{2+k} \le B_n^k \quad (k=1,2),$$

and

$$\frac{B_n^2\log^7(2M_n n)}{n} \le C_0 n^{-c_0}.$$

The standardized coordinate variances equal one.

Let $S_n^0 = \max_{j\le M_n}|G_{0n}(x_j)|$, where the centered Gaussian vector $\{G_{0n}(x_j) : j \le M_n\}$ has covariance $\mathbb{E}(X_{ij}X_{ik})$.

Theorem 5 (CCK reduction to the oracle Gaussian field)**.** *Under Assumptions 1 and 5,*

$$\rho_{\mathrm{GA},n} := \sup_{z\in\mathbb{R}}|\mathbb{P}(S_n\le z)-\mathbb{P}(S_n^0\le z)| \longrightarrow 0. \tag{28}$$

Moreover, for a deterministic $\ell_n \asymp C_n\sqrt{\log(2M_n)}$, with $C_n\to\infty$ arbitrarily slowly, $\mathbb{P}(S_n > \ell_n) = o(1)$.

Proof. Apply Corollary 2.1, condition (E.1), of Chernozhukov et al. (2013) to the row-wise independent, centered $2M_n$-vector

$$n^{-1/2}\sum_i (X_{i1},\dots,X_{iM_n},-X_{i1},\dots,-X_{iM_n}).$$

Independence and centering follow from Assumption 1; the coordinate variances are one. Assumption 5 supplies the third and fourth moments, exponential bound, and polynomial growth margin required by (E.1). Thus all conditions of that corollary hold, so the resulting Kolmogorov error is $o(1)$, proving (28). A Gaussian union bound gives $S_n^0 = O_p\{\sqrt{\log(2M_n)}\}$; the second claim then follows from (28). No feasible-covariance comparison or separate Gaussian anti-concentration step is used here. □

Let $r_{L,n}$ denote the left-hand side of (26), let

$$r_{S,n} = \sup_{x\in\mathcal{X}_n}\left|\frac{\widehat{\sigma}_K(x)}{\sigma_K(x)}-1\right|, \qquad D_n = \max_{x\in\mathcal{X}_n}\left|\frac{\sqrt{n}\{\widehat{\tau}(x)-\tau_K(x)\}}{\widehat{\sigma}_K(x)}\right|,$$

and define $\delta_{\mathrm{lin},n} = 2r_{L,n} + 2\ell_n r_{S,n}$.

Lemma 3 (Deterministic studentization perturbation)**.** *On* $\{r_{S,n} < 1/2, S_n \le \ell_n\}$,

$$|D_n - S_n| \le \delta_{\mathrm{lin},n}. \tag{29}$$

Proof. Write $Z_j^0 = n^{-1/2}\sum_i X_{ij}$, let $|R_j| \le r_{L,n}$ be the linearization remainder, and put $s_j = \widehat{\sigma}_K(x_j)/\sigma_K(x_j) - 1$. Then

$$\max_j\left|\frac{Z_j^0+R_j}{1+s_j} - Z_j^0\right| \le 2r_{L,n} + 2r_{S,n}\max_j|Z_j^0|,$$

which is (29) on the stated event. □

Assumption 6 (Feasible covariance consistency)**.** *Set*

$$V_K = A_K^{-1}\Omega_K A_K^{-1}, \qquad \widehat{V}_K = \widehat{A}_n^{-1}\widehat{\Omega}_n\widehat{A}_n^{-1}.$$

Then

$$d_{V,n} := \frac{\zeta_K^2\|\widehat{V}_K - V_K\|_{\mathrm{op}}}{\underline{\sigma}_K^2} = o_p(1). \tag{30}$$

Lemma 4 (Uniform covariance-kernel consistency)**.** *Under Assumption 6, uniformly over $x, x' \in \mathcal{X}$, the standard deviations and correlations generated by $\widehat{V}_K$ converge to those generated by V_K. A sufficient primitive compatibility rate is*

$$\frac{\zeta_K^2\{v_n + a_n(1+\omega_K)\}}{\underline{\sigma}_K^2} = o(1).$$

Proof. The quadratic-form bound

$$\sup_{x,x'\in\mathcal{X}}|c_K(x)^\top(\widehat{V}_K - V_K)c_K(x')| \le 2\zeta_K^2\|\widehat{V}_K - V_K\|_{\mathrm{op}}$$

and $\inf_x \sigma_K(x) = \underline{\sigma}_K$ give the uniform relative standard-deviation and normalized covariance conclusions. The final display follows by substituting the operator-norm rate obtained in the proof of Theorem 4. This is only C^0 covariance consistency; the derivative geometry needed below is stated separately. □

A.4 PRIMITIVE CUBIC-SPLINE RATES AND MULTIPLIER VALIDITY

The multiplier theorem is proved directly on the evaluation grid. It uses CCK Gaussian approximation for the non-Gaussian scores, CCK Gaussian comparison for the feasible covariance, and Gaussian anti-concentration to absorb linearization, studentization, discretization, and sieve bias. It does not require a smooth loading manifold or a tube formula.

Assumption 7 (Fixed-degree B-spline and CR3 primitives)**.** *Let $d = 2$, $\mathcal{X} = [0,1]^2$, $p_K \asymp K$, and let the spline degree $q \geq 3$ be fixed. In each additive coordinate use degree-q B-splines with simple quasi-uniform knots of mesh order K^{-1} (de Boor, 2001), using the rank-K centered and whitened construction in Section A.1. The cubic case $q = 3$ is one example; the numerical applications below use finite dictionaries instead. The design density is bounded above and away from zero, treatment has uniform overlap, and the population design and score covariances satisfy*

$$c\boldsymbol{I} \preceq A_K \preceq C\boldsymbol{I}, \qquad c\boldsymbol{I} \preceq V_K \preceq C\boldsymbol{I}, \qquad \|\Omega_K\|_{\mathrm{op}} \leq C.$$

The cluster size is bounded and, for a fixed C_u, $\mathbb{P}(\|u_i\|_2 > t) \leq 2\exp(-t^2/C_u^2)$. The design also satisfies $\|\mathbb{E}[(D_i^\top D_i - A_K)^2]\|_{\mathrm{op}} \leq CK$.

Choose a fixed $A > 2$, put $U_n = C_u\sqrt{A\log n}$, choose $L_n \to \infty$ arbitrarily slowly, and define

$$\begin{aligned}
a_{K,n} &= \sqrt{\frac{K\log(2K)}{n}} + \frac{K\log(2K)}{n}, \\
q_{H,n} &= K^2\sqrt{\frac{\log n}{n}} + \frac{K^3}{n} + \frac{K^2\log n}{n}, \\
q_{\Omega,n} &= K\log n\left\{\sqrt{\frac{\log(2K)}{n}} + \frac{\log(2K)}{n}\right\}, \\
R_{V,n} &= a_{K,n} + L_n(q_{H,n} + q_{\Omega,n}).
\end{aligned}$$

Take A sufficiently large that

$$K\,\mathbb{E}\{\|u_i\|_2^2\mathbf{1}(\|u_i\|_2 > U_n)\} = o\{L_n(q_{H,n} + q_{\Omega,n})\}.$$

For $p_n = 2M_n$, assume

$$\frac{K^3\log^7(p_n n)}{n} \leq C_0 n^{-c_0}, \qquad a_{K,n}\sqrt{K\log p_n} \to 0, \qquad R_{V,n}\log^2 p_n \to 0. \tag{31}$$

All constants may depend on the fixed degree q, but not on n or K.

Assumption 8 (Grid and undersmoothing primitives)**.** *Let $\mathcal{X}_n$ be a rectangular grid with spatial mesh Δ_n and $M_n \asymp \Delta_n^{-2}$. For this additive-spline sufficient regime, assume the arm means admit additive components with a stable spline projection satisfying $b_K \leq CK^{-s}$, where $3/2 < s \leq q+1$, and*

$$K^{3/2}\Delta_n\sqrt{\log p_n} \to 0, \qquad \sqrt{n}\,K^{-(s+1/2)}\sqrt{\log p_n} \to 0. \tag{32}$$

For later use, take the principal square roots $V_K = L_{0n}L_{0n}^\top$ and $\widehat{V}_K = \widehat{L}_n\widehat{L}_n^\top$, and define

$$\theta_{0n}(x) = \frac{L_{0n}^\top c_K(x)}{\sigma_K(x)}, \qquad \widehat{\theta}_n(x) = \frac{\widehat{L}_n^\top c_K(x)}{\widehat{\sigma}_K(x)}.$$

Let $G_{0n}(x) = \theta_{0n}(x)^\top Z_0$ and $\widehat{G}_n(x) = \widehat{\theta}_n(x)^\top Z$, where Z_0, Z are standard Gaussian vectors of the required dimension. Also set

$$J_{0n}(x) = \{\partial_{x_1}\theta_{0n}(x), \partial_{x_2}\theta_{0n}(x)\}, \qquad \kappa_{0,n} = \int_{\mathcal{X}} \det\{J_{0n}(x)^\top J_{0n}(x)\}^{1/2}\,dx.$$

Also set

$$Z_n(x) = \frac{\sqrt{n}\{\widehat{\tau}(x) - \tau_K(x)\}}{\widehat{\sigma}_K(x)},$$

and, for a nearest grid point $\pi_n(x)$, define

$$\begin{aligned}
d_{\mathcal{X},n} &= \sup_{x\in\mathcal{X}} |Z_n(x) - Z_n\{\pi_n(x)\}|, \\
d_{0,n} &= \sup_{x\in\mathcal{X}} |G_{0n}(x) - G_{0n}\{\pi_n(x)\}|, \\
d_{*,n} &= \sup_{x\in\mathcal{X}} |\widehat{G}_n(x) - \widehat{G}_n\{\pi_n(x)\}|.
\end{aligned}$$

Lemma 5 (Consequences of the B-spline primitives)**.** *Under Assumptions 1, 7, and 8,*

$$\zeta_K \asymp \underline{\sigma}_K \asymp \sqrt{K}, \qquad \max_i \|D_i\|_{\mathrm{op}} = O(\sqrt{K}),$$
$$\|\widehat{A}_n - A_K\|_{\mathrm{op}} = O_p(a_{K,n}), \qquad \|\widehat{V}_K - V_K\|_{\mathrm{op}} = O_p(R_{V,n}),$$

and Assumption 4 holds with $v_n = L_n(q_{H,n} + q_{\Omega,n})$. In addition,

$$r_{L,n} = O_p(a_{K,n}\sqrt{K}), \qquad r_{S,n} = O_p(R_{V,n}), \qquad d_{V,n} = O_p(R_{V,n}).$$

Consequently, Assumptions 2–6 follow from these primitives with $B_n = CK^{3/2}$, for a sufficiently large fixed constant C, and the displayed choices of a_n, v_n. If $\pi_n(x)$ is a nearest grid point, then

$$\sup_{x\in\mathcal{X}} |Z_n(x) - Z_n\{\pi_n(x)\}| = O_p(K^{3/2}\Delta_n),$$
$$\sup_{x\in\mathcal{X}} |G_{0n}(x) - G_{0n}\{\pi_n(x)\}| = O_p(K^{3/2}\Delta_n),$$

with the analogous conditional order for the feasible Gaussian multiplier field in probability.

For geometry, choose the principal square roots of V_K and $\widehat{V}_K$. Then, for $r = 0, 1, 2$,

$$\|\widehat{\theta}_n - \theta_{0n}\|_{C^r(\mathcal{X})} = O_p(K^r R_{V,n}).$$

Thus $K^2 R_{V,n} \to 0$ is sufficient for global C^2 loading convergence. In the primary cubic case, a spline with simple interior knots is only C^2: its third derivative jumps at the knots. The retained multiplier SCB needs only the displayed grid and covariance rates, not a global third-derivative condition on a loading manifold. Higher degree splines may provide additional smoothness, but it is not invoked here.

Proof. The raw spline Gram eigenvalues are of order K^{-1}. On the orthogonal complement of the constant direction, subtracting the reference mean preserves this order: the variance is the minimum squared distance to a constant, and the raw Gram norm is equivalent to K^{-1} times the coefficient norm. Thus whitening has operator norm $O(\sqrt{K})$. Raw splines have bounded overlap and derivative norms $O(K^r)$, so the rank-corrected basis has derivative envelopes

$$\sup_x \|\partial^r B_K(x)\|_2 = O(K^{r+1/2}), \qquad r = 0, 1, 2,$$

and the same piecewise bound for $r = 3$. Matrix Bernstein, using the stated design variance bound, gives the rate $a_{K,n}$. Substituting $\mathcal{D}_K = O(\sqrt{K})$, $p_K \asymp K$, and $U_n = O(\sqrt{\log n})$ in Lemmas 1 and 2 gives $q_{H,n}$ and $q_{\Omega,n}$. The resolvent calculation in Theorem 4 then gives the covariance, linearization, and studentization rates. Moreover, $|X_{ij}| \le C\sqrt{K}\|u_i\|_2$, so $\mathbb{E}|X_{ij}|^3 = O(K^{3/2})$, $\mathbb{E}|X_{ij}|^4 = O(K^2)$, and the exponential norm is $O(\sqrt{K})$. Choosing $B_n = CK^{3/2}$ therefore verifies both normalized moments and the exponential bound in Assumption 5. This conservative choice does not assume localization of the inverse-Gram score loadings.

Uniform positive definiteness and $R_{V,n} = o(1)$ give

$$\sup_{x\in\mathcal{X}} \|\nabla\{c_K(x)/\widehat{\sigma}_K(x)\}\|_2 = O_p(K).$$

Together with $\|\sqrt{n}(\widehat{\beta} - \beta_K)\|_2 = O_p(\sqrt{K})$, this yields the first grid modulus. The Gaussian coefficient norm is $O_p(\sqrt{K})$, conditionally as well, and the normalized loading is $O(K)$-Lipschitz, which yields the other two. The principal matrix square root is locally Lipschitz under the same spectral bounds. Repeated product and quotient rules give the stated $K^r R_{V,n}$ loading rates. The knot-multiplicity smoothness rule gives the last conclusion. □

Theorem 6 (Gaussian multiplier SCB validity)**.** *Under Assumptions 1–8, for any fixed $\alpha \in (0, 1)$,*

$$\sup_{z\in\mathbb{R}} |\mathbb{P}(D_n \le z) - \mathbb{P}^*(T_n^* \le z)| = o_p(1). \tag{33}$$

If $c_{1-\alpha}^*$ is the exact conditional $(1-\alpha)$-quantile of T_n^*, then the band (24) *satisfies*

$$\mathbb{P}\{\tau(x) \in \mathcal{C}_{1-\alpha}(x),\ \forall x \in \mathcal{X}\} \longrightarrow 1 - \alpha.$$

Proof. Write $p_n = 2M_n$ to represent the absolute maximum as a coordinatewise maximum. The sub-Gaussian cluster condition and spline envelopes imply the CCK moment and exponential bounds with $B_n = CK^{3/2}$, so the first condition in (31) and Theorem 5 give the oracle Gaussian approximation.

Conditionally on the data, T_n^* is a standardized Gaussian maximum. By Lemma 5, the largest entrywise difference between its correlation matrix and the oracle correlation matrix is $O_p(R_{V,n})$. Lemma 3.1 of Chernozhukov et al. (2013) and $R_{V,n}\log^2 p_n \to 0$ therefore give a conditional Kolmogorov distance $o_p(1)$ between the two Gaussian maxima.

It remains to replace the linearized statistic by D_n. Lemmas 3 and 5 give

$$\delta_{\mathrm{lin},n} = O_p\{a_{K,n}\sqrt{K} + C_n R_{V,n}\sqrt{\log p_n}\},$$

where $C_n \to \infty$ may be chosen slowly. Conditions (31) allow this choice so that

$$\delta_{\mathrm{lin},n}\sqrt{\log p_n} = o_p(1).$$

Use the Gaussian anti-concentration bound in Lemma 2.1 of Chernozhukov et al. (2013). It absorbs this perturbation. Combining the three comparisons proves (33).

Finally, Lemma 5, (32), and the same anti-concentration bound show that replacing the grid supremum by the supremum over $\mathcal{X}$, and τ_K by τ, preserves (33) with D_n replaced by the continuous target statistic

$$\widetilde{D}_n = \sup_{x\in\mathcal{X}} \frac{\sqrt{n}|\widehat{\tau}(x) - \tau(x)|}{\widehat{\sigma}_K(x)}.$$

For completeness, let $q_n(a)$ be the deterministic a-quantile of the oracle Gaussian maximum. For every fixed $0 < \epsilon < \min(\alpha, 1-\alpha)$, the two Gaussian comparisons imply, on an event with probability tending to one,

$$q_n(1-\alpha-\epsilon) \le c^*_{1-\alpha} \le q_n(1-\alpha+\epsilon).$$

On this event, deterministic event inclusions bracket $\{\widetilde{D}_n \le c^*_{1-\alpha}\}$ between the corresponding two events at $q_n(1-\alpha-\epsilon)$ and $q_n(1-\alpha+\epsilon)$. The oracle Gaussian maximum has a continuous distribution, and its approximation error is uniform in the threshold. Taking probabilities, then letting $\epsilon \downarrow 0$, proves the coverage statement. □

A nonempty rate regime. For $3/2 < s \le q+1$, take $K \asymp n^{\kappa}$ with $1/(2s+1) < \kappa < 1/4$, a grid mesh $\Delta_n \asymp n^{-d_0}$ with $d_0 > 3\kappa/2$, and $L_n = \log\log(n+3)$. Then $K^3\log^7(p_n n)/n$ has a polynomial margin, $R_{V,n}\log^2 p_n \to 0$, and both grid and bias conditions hold. The restriction $s > 3/2$ is needed for this sufficient regime, rather than an assertion that every positive smoothness permits these rates.

Corollary 1 (Fixed-dictionary, finite-grid confidence band)**.** *Let the coefficient dimension and evaluation grid be fixed. Suppose the clusters are i.i.d.,* $\mathbb{E}g_i = 0$*,* A_K *is nonsingular,* $\mathbb{E}(\|D_i\|_{\mathrm{op}}^4 + \|g_i\|_2^2) < \infty$*, the block leverage maximum converges to zero, the CR3 covariance is consistent, and all contrast variances are positive. Then the Gaussian multiplier band has asymptotic simultaneous coverage* $1-\alpha$ *for* τ_K *on that grid. Coverage for* τ *additionally requires* $\max_x |\tau(x) - \tau_K(x)| = o(n^{-1/2})$.

Proof. The exact normal equation, the law of large numbers for $\widehat{A}_n$, and the multivariate central limit theorem give a Gaussian limit for $\sqrt{n}(\widehat{\beta} - \beta_K)$. Covariance consistency and positive contrast variances give the same conditional limit for the studentized multiplier vector. Its absolute maximum has a continuous cdf, so quantile convergence gives the claim. The stated bias bound permits replacement of τ_K by τ by Slutsky's theorem. □

This corollary, rather than the increasing-spline theorem, is the relevant asymptotic reference for the radial-basis simulation and the linear/categorical HeartSteps fit. Consistent empirical-quantile grid locations can also be used when the fixed basis is continuously differentiable near their deterministic limits, with positive limiting variances: the loading vectors then converge and the coefficient CLT applies at those random locations. The target remains the projection evaluated at the reported locations. Neither result supplies a small-cluster guarantee for 37 participants. In computation, the multiplier quantile uses R's type-7 interpolation: if $h = 1 + (B-1)(1-\alpha)$, interpolate the order statistics at $\lfloor h \rfloor$ and $\lceil h \rceil$. The asymptotic statements use the exact conditional quantile, or require $B \to \infty$ to remove Monte Carlo quantile error.

B PERSONAL FPCA PREDICTION AND MODEL-BASED BANDS

B.1 PERSONAL FPCA PREDICTION WITHIN THE CONFORMAL LAYER

The population SCB targets τ, whereas this section predicts τ_i in (2) under the same model (1). We first construct the personal predictor from raw history, then connect it to subject-level conformal calibration in Section 2.3. The Gaussian bands below describe model-based prediction uncertainty; their critical values are not empirical conformal ranks. Covariate adjustment of the mean with a shared covariance is consistent with the modeling distinction studied by Jiang and Wang (2010). The original pooled branch in this section introduces no covariate-dependent covariance extension. The separate direct working-model extension appears in Section 2.4; none of the results below are transferred to it.

Personal target and fitting representation The personal band targets an individual random curve, not just the population mean effect. It does not include a fresh measurement error for a future observed response.

Write $G_{ab}(z, z') = \mathrm{Cov}\{g_{ia}(z), g_{ib}(z')\}$. This joint covariance is shared across people. In particular, it has no additional dependence on X. The mean and persistent-process dictionaries may come from different finite, linearly independent families. They are prespecified, or selected on an independent auxiliary sample and then conditioned upon. The persistent dictionary is orthonormal under a fixed reference measure; the mean dictionary need not be additive or orthonormal. For the exact results below, assume

$$\mu_a(x) = a(x)^\mathsf{T}\eta_a, \qquad g_{ia}(z) = b(z)^\mathsf{T}\gamma_{ia}, \qquad \gamma_i = (\gamma_{i0}^\mathsf{T}, \gamma_{i1}^\mathsf{T})^\mathsf{T}, \quad \mathrm{Cov}(\gamma_i) = \Gamma > 0. \tag{34}$$

Here b contains p orthonormal numerical basis functions and $D = 2p$. These functions provide coordinates for estimation; the learned eigenfunctions below are the FPCs. The finite representation is an explicit restriction for this theory and simulation, not a theorem for an arbitrary infinite-rank process. The simulated polynomial design is one instance and does not establish identification for other dictionaries.

For a record history $\mathcal{H}_i$ of $h_i = |\mathcal{H}_i|$ visits, let F_i and B_i be the arm-specific mean and random-effect design matrices. Then

$$Y_i = F_i\eta + B_i\gamma_i + \varepsilon_i, \qquad \Omega_i = \mathrm{Var}(\varepsilon_i).$$

Throughout, conditional on the design, random effects and noise have zero means, covariances Γ and Ω_i, and zero cross-covariance. These moment restrictions also apply to the fitting people. All conditional Gaussian statements assume that, conditional on the observed design $\mathcal{D}_i$, the score law remains $N(0, \Gamma)$ and the independent noise law is $N(0, \Omega_i)$. The numerical study uses exogenous designs and $\Omega_i = \sigma^2 I$. Sequential randomization alone does not establish this conditional law when later designs depend on earlier outcomes. A separate likelihood analysis would be needed in that setting.

Pooled fitting and learned joint FPCA Split the independent training people into two groups, each containing n people. Estimate the common mean in the first group:

$$\widehat{\eta} = \left(\sum_i F_i^\mathsf{T} F_i\right)^{-1} \sum_i F_i^\mathsf{T} Y_i. \tag{35}$$

This is a pooled mean fit. No personal least-squares estimate of γ_i is required.

For people in the second group, set $r_i = Y_i - F_i\widehat{\eta}$. For this fitting estimator and its consistency theorem, specialize to homoskedastic independent noise, $\Omega_i = \sigma^2 I$. General known positive noise matrices can be used in the prediction formulas, but are not estimated by the scalar-noise fit below. With $\vartheta = (\mathrm{vech}\,\Gamma, \sigma^2)$, define the known linear operator

$$O_i\vartheta = \mathrm{vech}(B_i\Gamma B_i^\mathsf{T} + \sigma^2 I).$$

Estimate the covariance by pooled second moments:

$$\widetilde{\vartheta} = \left(\sum_i O_i^\mathsf{T} O_i\right)^{-1} \sum_i O_i^\mathsf{T}\,\mathrm{vech}(r_i r_i^\mathsf{T}). \tag{36}$$

Project the resulting covariance to a positive spectral interval and the noise variance to a positive interval. The consistency statement assumes that the true parameters lie strictly inside these intervals.

Diagonalize $\widehat{\Gamma} = \widehat{U}\widehat{\Lambda}\widehat{U}^{\mathsf{T}}$. For population notation, $\Gamma = U\Lambda U^{\mathsf{T}}$, with ordered eigenvalues λ_k, and

$$\phi_{ak}(z) = b(z)^{\mathsf{T}} U_{ak}, \qquad \xi_i = U^{\mathsf{T}}\gamma_i, \qquad g_{ia}(z) = \sum_{k=1}^{D} \xi_{ik}\phi_{ak}(z).$$

Here U_{ak} is the arm-a block of column k. The eigenfunctions are joint eigenfunctions on the two-arm function space, so the cross-arm covariance is retained. Identification of the full Γ requires pooled information about both arms within people, or additional structure. A design observing only one arm per person cannot generally identify the cross-arm block.

Personal prediction from raw history Let $\Phi_i = B_i U$, $\mathcal{V}_i = \Phi_i \Lambda \Phi_i^{\mathsf{T}} + \Omega_i$, and

$$m_i = \Lambda \Phi_i^{\mathsf{T}} \mathcal{V}_i^{-1}(Y_i - F_i\eta), \tag{37}$$

$$\mathcal{S}_i = \Lambda - \Lambda \Phi_i^{\mathsf{T}} \mathcal{V}_i^{-1} \Phi_i \Lambda. \tag{38}$$

These are the best linear score predictor and its error covariance under second moments. Under the conditional Gaussian model,

$$\xi_i \mid (Y_i, \mathcal{D}_i) \sim N(m_i, \mathcal{S}_i).$$

Outside the Gaussian model, $\mathcal{S}_i$ describes the linear prediction error covariance averaged over histories, conditional on design; it need not be a posterior covariance for a realized history.

Define the contrast functions and truncation projector by

$$d_k(z) = \phi_{1k}(z) - \phi_{0k}(z), \qquad d(z) = (d_1(z), \ldots, d_D(z))^{\mathsf{T}},$$

where P_M is the diagonal projector retaining the first M coordinates. The fitted personal predictor is

$$\widehat{c}_{i,M}(x, z) = a(x)^{\mathsf{T}}(\widehat{\eta}_1 - \widehat{\eta}_0) + \widehat{d}(z)^{\mathsf{T}} P_M \widehat{m}_i. \tag{39}$$

The full fitted covariance is used in computing $\widehat{m}_i$ and $\widehat{\mathcal{S}}_i$ before truncating the displayed predictor. The experiments choose the smallest M explaining at least 95% of fitted total variance.

When $h_i = 0$, $m_i = 0$ and $\mathcal{S}_i = \Lambda$: the predictor is the population mean and the band retains between-person uncertainty. When $h_i < D$, prediction remains well-defined because $\Omega_i > 0$. Individual design invertibility is unnecessary. The posterior score calculation follows the functional prediction principle in Ding et al. (Ding et al., 2022), Section 3.2. Their paper motivates this prediction step; it does not establish the simultaneous coverage claims made here.

Simultaneous bands and their guarantees

B.1.1 Known parameters: full-score Gaussian band

Write $u = (x, z)$ and take a fixed compact prediction domain $\mathcal{U}$. With all D scores, let

$$c_i(u) = \tau(x) + d(z)^{\mathsf{T}} m_i, \qquad s_i^2(z) = d(z)^{\mathsf{T}} \mathcal{S}_i d(z).$$

A simple critical value is $q_\chi = \sqrt{\chi^2_{D,1-\alpha}}$.

Proposition 5 (Gaussian projection band)**.** *Under the conditional Gaussian model with known parameters,*

$$\mathbb{P}\{\tau_i(u) \in [c_i(u) - q_\chi s_i(z), c_i(u) + q_\chi s_i(z)] \text{ for all } u \in \mathcal{U} \mid Y_i, \mathcal{D}_i\} \geq 1 - \alpha.$$

The inequality reflects projection of a D-dimensional ellipsoid onto a family of contrasts. It can be conservative.

A sharper critical value is the $(1-\alpha)$ quantile of

$$T_i = \sup_{u \in \mathcal{U}} \frac{|d(z)^{\mathsf{T}} \mathcal{S}_i^{1/2} Z|}{s_i(z)}, \qquad Z \sim N(0, I_D). \tag{40}$$

Assume $s_i(z) > 0$ throughout the domain. Replacing the supremum with a maximum over a finite grid guarantees coverage on that grid only.

For simulation calibration on a grid, generate B independent copies of T_i and take order statistic $k = \lceil (B+1)(1-\alpha) \rceil$, provided $k \leq B$. Under the known conditional Gaussian law, the future standardized error and these simulated copies are independent draws from the same distribution conditional on the history. No calibration subjects enter this model-based calculation. With a continuous distribution, coverage averaged over this Monte Carlo randomness equals $k/(B+1)$. Here $B = 499$, $\alpha = .05$, and $k = 475$, giving .95. This is a model-based rank argument, not a distribution-free conformal guarantee.

B.1.2 Truncation adjustment

Here q_i denotes either the Gaussian-maximum critical value for person i or the common projection value q_χ, with the domain matched to that calibration. The omitted predicted contribution is

$$\delta_{i,M}(z) = d(z)^\mathsf{T}(I - P_M)m_i.$$

The error around the truncated center is

$$\tau_i(u) - c_{i,M}(u) = \delta_{i,M}(z) + d(z)^\mathsf{T}(\xi_i - m_i).$$

Consequently, the band

$$c_{i,M}(u) \ \pm \ \{|\delta_{i,M}(z)| + q_i s_i(z)\} \tag{41}$$

contains the corresponding full-center band pointwise. It inherits its simultaneous coverage guarantee for every M, including a data-selected M. Thus the guard permits a truncated displayed center for the full latent target; it is not intended to reduce width or computation. The practical version substitutes fitted quantities everywhere. This adjustment preserves the full posterior uncertainty as well as the omitted posterior mean; deleting discarded directions from the variance is generally invalid.

B.1.3 Estimated parameters

We state the scope of the asymptotic guarantee explicitly: If dictionaries are selected on an independent auxiliary sample, all statements below are conditional on that selection. Dimensions remain fixed, and the representation, Gram, moment, boundedness, and variance conditions must hold uniformly on events whose probability tends to one.

(A1) The representation (34) is exact, with fixed dimensions. Training people are independent and identically distributed within each fitting group, the groups have the same population model, and the future person is independent of both fitting groups.

(A2) Histories have bounded length; design functions are bounded on the fitting and prediction domains; fourth moments exist. The two population matrices $\mathbb{E}(F_i^\mathsf{T} F_i)$ and $\mathbb{E}(O_i^\mathsf{T} O_i)$ are positive definite.

(A3) The true covariance and noise variance are strictly positive and interior to the clipping intervals.

(A4) For the coverage conclusion, the stated conditional Gaussian model holds. On the prediction domain the limiting $s_i(z)$ is bounded away from zero. The number of prediction grid points and Monte Carlo draws is fixed.

Theorem 7 (Fitted finite-dimensional bands)**.** *Under (A1)–(A3), $\widehat{\eta} - \eta = O_p(n^{-1/2})$ and $\widehat{\Gamma} - \Gamma = O_p(n^{-1/2})$, with the same rate for the noise variance. Under (A4), the fitted full-score Gaussian grid band calibrated with $B = 499$, $k = 475$ has*

$$\liminf_{n\to\infty} \mathbb{P}\{\tau_i(u) \text{ is covered for every grid point } u\} \geq .95.$$

The fitted truncation-adjusted band (41) *has the same lower bound. The fitted χ^2 projection band has asymptotic simultaneous coverage at least .95 on the compact domain. Probabilities average over training, the future person's history and random effect, and any calibration randomness.*

This is not a finite-sample exactness claim for fitted bands. It is not coverage conditional on a fixed personal random effect. No eigenvalue gap is needed for full-score coverage or the containment argument, since they can be expressed in fixed coordinates. Consistency of separately labelled eigenfunctions requires the usual separation of the corresponding eigenspaces.

B.1.4 NON-GAUSSIAN AND INFINITE-RANK LIMITS

With only finite second moments, the best linear predictor remains meaningful. A full-score Chebyshev radius $\sqrt{D/\alpha}$ gives a known-parameter guarantee conditional on design and averaged over history, using the full linear error covariance. It does not justify Gaussian conditional-history calibration. The elliptical-t_5 experiment is therefore a robustness check, not an experiment covered by the Gaussian theorem.

For an infinite-rank process, one must also control the omitted random tail and its contribution to the history likelihood. A 95% variance-explained threshold alone does not provide a uniform tail envelope. Such an extension is outside the claims of this construction.

B.2 PROOFS AND TECHNICAL DETAILS

Score prediction and Gaussian projection For $r_i = Y_i - F_i\eta$, $\mathrm{Cov}(\xi_i, r_i) = \Lambda\Phi_i^\mathsf{T}$ and $\mathrm{Var}(r_i) = \mathcal{V}_i$. Minimizing the mean squared error of Ar_i gives $A = \Lambda\Phi_i^\mathsf{T}\mathcal{V}_i^{-1}$. Substitution gives (38). Under Gaussianity, the residual $\xi_i - m_i$ is jointly Gaussian and uncorrelated with r_i, hence independent of r_i; this gives the stated posterior law.

For Proposition 5, write $\xi_i - m_i = \mathcal{S}_i^{1/2}Z$ conditionally on history. Cauchy–Schwarz gives $|d(z)^\mathsf{T}\mathcal{S}_i^{1/2}Z| \le s_i(z)\,\|Z\|$ simultaneously in z. The event $\|Z\|^2 \le \chi^2_{D,1-\alpha}$ has probability $1-\alpha$. For Monte Carlo calibration on the finite grid, the rank of the independent future maximum among its B calibration copies and itself is uniform on $1,\dots,B+1$; thus the order-statistic rule gives $k/(B+1)$ coverage. Any zero-probability ties can be ignored under positive Gaussian variances. Finally $c_i = c_{i,M} + \delta_{i,M}$ and the triangle inequality prove containment in (41) without assuming M is deterministic.

Pooled identification in the simulated design The mean design has a nonsingular population Gram matrix because both arms have positive probability and the polynomial regressors have full support. For covariance identification, suppose $B_i\Delta B_i^\mathsf{T} + cI = 0$ almost surely for a symmetric Δ. For distinct visits in any arm pair (a,b), this gives $b(z)^\mathsf{T}\Delta_{ab}b(z') = 0$ for independently continuously distributed z, z'. The polynomial identity forces $\Delta_{ab} = 0$ for all arm pairs. The diagonal entries then give $c = 0$. Thus the pooled moment operator is injective, and its finite-dimensional population Gram matrix is positive definite. This argument uses within-person pairs; it is not an assertion of full rank for each personal history.

Rates for the pooled estimates Write $e_i = B_i\gamma_i + \varepsilon_i$. The mean normal equation is

$$\widehat{\eta} - \eta = \left(n^{-1}\sum_i F_i^\mathsf{T}F_i\right)^{-1} n^{-1}\sum_i F_i^\mathsf{T}e_i.$$

The inverse is bounded in probability by (A2). The centered average is $O_p(n^{-1/2})$ by independence and finite variance.

For the independent covariance group, put $\Delta_\eta = \widehat{\eta} - \eta$ and expand

$$r_i r_i^\mathsf{T} = e_i e_i^\mathsf{T} - e_i\Delta_\eta^\mathsf{T}F_i^\mathsf{T} - F_i\Delta_\eta e_i^\mathsf{T} + F_i\Delta_\eta\Delta_\eta^\mathsf{T}F_i^\mathsf{T}.$$

After multiplication by $O_i^\mathsf{T}\,\mathrm{vech}(\cdot)$ and averaging, each linear term is $O_p(n^{-1})$: its centered coefficient average is $O_p(n^{-1/2})$ and is independent of $\Delta_\eta = O_p(n^{-1/2})$. The quadratic term is also $O_p(n^{-1})$. The leading centered second-moment average is $O_p(n^{-1/2})$ by the fourth-moment assumption. Inverting the convergent positive definite moment Gram matrix proves the rate for $\widetilde{\vartheta}$. Interior clipping is inactive with probability tending to one and therefore preserves the rate.

Transfer to fitted bands Work in the fixed γ coordinates, avoiding eigenvector labels. The true conditional mean and covariance are

$$a_i = \Gamma B_i^\mathsf{T}(B_i\Gamma B_i^\mathsf{T} + \Omega_i)^{-1}(Y_i - F_i\eta), \qquad C_i = \Gamma - \Gamma B_i^\mathsf{T}(B_i\Gamma B_i^\mathsf{T} + \Omega_i)^{-1}B_i\Gamma.$$

Their fitted versions converge in probability to (a_i, C_i). The lower noise bound, bounded histories and designs, and positive Γ keep C_i nonsingular. Gaussian densities depend continuously on their

mean and positive definite covariance; on compact parameter neighborhoods they admit an integrable Gaussian envelope. Integration therefore gives total variation convergence of the fitted and true conditional laws. Restriction to events of arbitrarily high probability handles the random conditional means. Since total variation is bounded by one, its expectation also tends to zero.

For the target curve, there is also a uniformly $o_p(1)$ population-mean estimation error. Divide this error by the lower bound for s_i. For a grid of L points, each true standardized residual is marginally $N(0,1)$. For any threshold t independent of the target conditional residual and any $\epsilon > 0$, a union bound gives

$$\mathbb{P}\{\max_{\ell \le L} |Z_\ell| \in [t-\epsilon, t+\epsilon]\} \le \frac{4L\epsilon}{\sqrt{2\pi}},$$

regardless of their correlations. This controls the effect of the vanishing mean shift. Apply total variation transfer conditionally on training and history, integrating over the independent fitted calibration draws. Under the fitted Gaussian law the rank argument gives $475/500$ coverage. The transfer and mean-shift bounds give the asserted asymptotic lower bound under the true law.

For the compact-domain χ^2 band, Cauchy–Schwarz bounds the normalized error by the norm of a D-dimensional standard Gaussian vector plus an $o_p(1)$ term. Continuity of its radial distribution gives the result. The fitted truncation-adjusted band contains the fitted full band for each realized training sample and history; thus its conclusion follows directly.

A second-moment alternative Conditional on design, $\xi_i - m_i$ has mean zero and covariance $\mathcal{S}_i$ even without Gaussianity. Hence $\mathbb{E}\{(\xi_i - m_i)^\mathsf{T} \mathcal{S}_i^{-1} (\xi_i - m_i) \mid \mathcal{D}_i\} = D$. Markov's inequality and Cauchy–Schwarz yield simultaneous coverage at least $1-\alpha$ using radius $\sqrt{D/\alpha}$. This averages over history and assumes known parameters.

If one instead studies the unconditional linear error of a truncated predictor, write $J_i = \Lambda \Phi_i^\mathsf{T} \mathcal{V}_i^{-1} \Phi_i \Lambda = \Lambda - \mathcal{S}_i$. Its covariance is

$$W_{i,M} = \Lambda - P_M J_i - J_i P_M + P_M J_i P_M = \mathcal{S}_i + (I - P_M) J_i (I - P_M).$$

This distinction confirms why a retained-score variance alone omits relevant uncertainty. It does not turn the non-Gaussian linear error into a conditional Gaussian posterior.

C Conformal recovery-set guarantees

These details supply the recovery events used by the current FPCA conformal band. The former efficient OLS/noisy-score procedure is not part of this construction.

C.1 Conformal prediction of latent effect curves

For a new exchangeable subject, the prediction target is the latent effect curve defined by the conditional mean

$$\tau_i(u) = \mu_1(x) - \mu_0(x) + g_{i1}(z) - g_{i0}(z), \qquad u = (x,z) \in \mathcal{U}. \tag{42}$$

It excludes the innovation contrast and is neither a factual outcome nor a realized visit-level individual treatment effect. The fitting sample $\mathcal{D}_{\mathrm{fit}}$, calibration subjects $i = 1, \dots, m$, and future subject $m+1$ are disjoint. Conditional on $\mathcal{D}_{\mathrm{fit}}$, every global regression, covariance, dictionary, block rule, center, scale, and tuning parameter is frozen.

The robust results allow either a compact continuum or a fixed grid. In the continuum case, $\mathcal{U}$ is compact, all paths entering a supremum are sample-continuous, and the positive scale is bounded away from zero. The supremum is then measurable because it equals the supremum over a fixed countable dense set. Numerical maximization on a finite grid guarantees that grid only unless a valid discretization correction is supplied.

Randomized recovery labels and the legal filtration Each calibration subject is prospectively divided into a legal history block $\mathcal{H}_i$ and a disjoint recovery block. The reported center $\widehat{m}_i(u)$ uses only $\mathcal{H}_i$ and $\mathcal{D}_{\mathrm{fit}}$. At recovery visit r, treatment is randomized with known predictable propensity

$e_{ir} \in [e_{\min}, 1 - e_{\min}]$. For frozen arm regressions $\widehat{m}_{a,ir}$, define the augmented inverse-propensity weighted label

$$Q_{ir} = \widehat{m}_{1,ir} - \widehat{m}_{0,ir} + \frac{A_{ir}}{e_{ir}}(Y_{ir} - \widehat{m}_{1,ir}) - \frac{1 - A_{ir}}{1 - e_{ir}}(Y_{ir} - \widehat{m}_{0,ir}). \tag{43}$$

This is a noisy recovery label; it is not an observed latent treatment effect.

Assumption 9 (External randomization and legal recovery filtration)**.** *Let $R_{\max}$ be a deterministic upper bound on the recovery window. The sigma-field $\mathscr{G}^{\mathrm{aug}}_{i,r-1}$, available just before treatment assignment at visit r, contains $\mathcal{D}_{\mathrm{fit}}$, the latent coefficient and persistent subject effects fixed before recovery, the legal history, previous recovery queries, assignments, outcomes and missingness indicators, the current preassignment query U_{ir}, and auxiliary randomization already generated. It excludes the current assignment, the current potential innovations, and every future assignment, innovation and recovery outcome. The inclusion indicator M_{ir} is predictable. Given this sigma-field, A_{ir} is Bernoulli with success probability e_{ir} and is independent of the current potential innovations. For both arms, $\mathbb{E}\{\varepsilon_{ir}(a) \mid \mathscr{G}^{\mathrm{aug}}_{i,r-1}\} = 0$, $a \in \{0, 1\}$. This centering condition is part of identification; no sub-Gaussian envelope is required for Proposition 6.*

Because U_{ir} and M_{ir} are preassignment measurable, $x_{ir} = M_{ir} b(U_{ir})$ is predictable. A random recovery length is covered by $M_{ir} = 1\{r \le N_i\}$ when $\{N_i \ge r\} \in \mathscr{G}^{\mathrm{aug}}_{i,r-1}$. Missingness revealed only after observing the current outcome is not covered.

Proposition 6 (Identification at a recovery visit)**.** *Under Assumption 9, including its two-arm conditional innovation centering, for any frozen preassignment-measurable arm regressions,*

$$\mathbb{E}(Q_{ir} \mid \mathscr{G}^{\mathrm{aug}}_{i,r-1}) = \tau_i(U_{ir}).$$

Proof. Write $f_{ia,r} = \mu_a(X_{ir}) + g_{ia}(Z_{ir})$. Conditional on the legal preassignment sigma-field, innovation centering and randomization give

$$\mathbb{E}\left\{\frac{A}{e}(Y - \widehat{m}_1)\right\} = f_1 - \widehat{m}_1, \qquad \mathbb{E}\left\{\frac{1 - A}{1 - e}(Y - \widehat{m}_0)\right\} = f_0 - \widehat{m}_0.$$

Substitution in (43) yields $f_1 - f_0 = \tau_i(U_{ir})$. □

A primitive recovery-noise bound Set $h_{ia,r} = f_{ia,r} - \widehat{m}_{a,ir}$. For an included visit let Q_{ir} be the AIPW label in (43). Define

$$Q^{\mathrm{obs}}_{ir} = \begin{cases} Q_{ir}, & M_{ir} = 1, \\ 0, & M_{ir} = 0, \end{cases} \qquad \nu_{ir} = \begin{cases} Q_{ir} - \tau_i(U_{ir}), & M_{ir} = 1, \\ 0, & M_{ir} = 0. \end{cases}$$

The zero extensions require no outcome or pseudo-label at an omitted visit.

Assumption 10 (Primitive arm and innovation envelopes)**.** *Conditional on $\mathscr{G}^{\mathrm{aug}}_{i,r-1}$, the potential innovation $\varepsilon_{ir}(a)$ is centered and $\sigma_{a,ir}$-sub-Gaussian. On a subject-level event $E_{H,i}$, $|h_{ia,r}| \le H_{a,ir}$ for every included recovery visit, where the bounds and the residuals to which they apply are $\mathscr{G}^{\mathrm{aug}}_{i,r-1}$-measurable. Conditional on $\mathcal{D}_{\mathrm{fit}}$, $\mathbb{P}(E^c_{H,i}) \le \eta_H$.*

The event $E_{H,i}$ refers to the complete recovery window and therefore must not be used as a conditioning event in a martingale argument. To localize the process, define the first envelope-violation time

$$\kappa_i = \inf\left\{r \le R_{\max} : M_{ir} = 1 \text{ and } \max_{a \in \{0,1\}} \{|h_{ia,r}| - H_{a,ir}\} > 0\right\},$$

with $\inf \varnothing = R_{\max} + 1$, and define the stopped pseudo-noise

$$\bar{\nu}_{ir} = \begin{cases} \{Q_{ir} - \tau_i(U_{ir})\}\mathbf{1}\{r < \kappa_i\}, & M_{ir} = 1, \\ 0, & M_{ir} = 0. \end{cases}$$

Predictability of M_{ir}, $h_{ia,r}$, and $H_{a,ir}$ makes $M_{ir}\mathbf{1}\{r < \kappa_i\}$ measurable with respect to $\mathscr{G}^{\mathrm{aug}}_{i,r-1}$. On $E_{H,i}$, no violation occurs and $\bar{\nu}_{ir} = \nu_{ir}$ for every included visit.

Lemma 6 (Primitive conditional MGF)**.** *Under Assumptions 9–10, the stopped increment satisfies, for every real t,*

$$\mathbb{E}(\bar{\nu}_{ir} \mid \mathscr{G}^{\mathrm{aug}}_{i,r-1}) = 0, \qquad \mathbb{E}\{e^{t\bar{\nu}_{ir}} \mid \mathscr{G}^{\mathrm{aug}}_{i,r-1}\} \le \exp\left\{\frac{t^2 R_{ir}^2}{2} M_{ir}\mathbf{1}\{r < \kappa_i\}\right\},$$

where

$$R_{ir}^2 = \max\left\{\frac{\sigma_{1,ir}^2}{e_{ir}^2}, \frac{\sigma_{0,ir}^2}{(1-e_{ir})^2}\right\} + \frac{1}{4}\left\{\frac{H_{1,ir}}{e_{ir}} + \frac{H_{0,ir}}{1-e_{ir}}\right\}^2. \tag{44}$$

If $\sigma_{a,ir} \le \sigma_\varepsilon$ and $H_{a,ir} \le H_a$, overlap gives

$$R_{ir} \le e_{\min}^{-1}\left\{\sigma_\varepsilon^2 + \frac{(H_1+H_0)^2}{4}\right\}^{1/2} =: R_*.$$

Any larger fixed proxy is also valid. The constants are prespecified or fixed from independent fitting information; inactive branches may use any finite nonnegative extension because the exponent is multiplied by the activity indicator.

Proof. Fix a visit and condition on $\mathscr{G}^{\mathrm{aug}}_{i,r-1}$. On the predictable event $\{M_{ir} = 1, r < \kappa_i\}$, all quantities below except the current assignment and current innovations are then fixed. Write $e = e_{ir}$. Direct substitution in the AIPW label gives

$$\nu_{ir} = \frac{A_{ir} - e}{1-e}\left\{\frac{1-e}{e}h_{i1,r} + h_{i0,r}\right\} + \frac{A_{ir}}{e}\varepsilon_{ir}(1) - \frac{1-A_{ir}}{1-e}\varepsilon_{ir}(0). \tag{45}$$

Let $w = ((1-e)/e)h_{i1,r} + h_{i0,r}$. If $A = 1$, the right side of (45) is $w + \varepsilon(1)/e$. If $A = 0$, it is $-ew/(1-e) - \varepsilon(0)/(1-e)$. The conditional innovation MGF bounds therefore give

$$\mathbb{E}(e^{t\nu} \mid \mathscr{G}) \le e^{t^2\sigma_*^2/2}\{ee^{tw} + (1-e)e^{-tew/(1-e)}\}, \quad \sigma_*^2 = \max\{\sigma_1^2/e^2, \sigma_0^2/(1-e)^2\}.$$

The expression in braces is the MGF of a two-point variable that equals w with probability e and $-ew/(1-e)$ with probability $1-e$. Its conditional mean is zero and its range length is $|w|/(1-e)$. Hoeffding's lemma gives an additional variance proxy $\{|w|/(1-e)\}^2/4$. On $\{M_{ir} = 1, r < \kappa_i\}$,

$$\frac{|w|}{1-e} \le \frac{H_1}{e} + \frac{H_0}{1-e}$$

so the displayed value of R_{ir} is valid. If $M_{ir} = 0$ or $r \ge \kappa_i$, the stopped increment is zero and its conditional MGF is one. Combining the two predictable cases proves the MGF inequality. Taking the conditional expectation of (45) also proves the conditional mean-zero statement. This proof never conditions on the future-dependent event $E_{H,i}$. □

An envelope only for the treatment-effect coefficient is not enough: common arm modes can be arbitrarily large while their difference is zero. A valid route is to assume the arm residual bounds directly, or to derive them from bounded arm loadings and coefficients together with a frozen nuisance envelope. This is why η_H appears in every robust coverage budget.

Exact-dictionary recovery and ghost conformalization Assume temporarily that

$$\tau_i(u) = b(u)^\top \theta_i, \qquad \mathbb{P}(\|\theta_i\|_2 > S \mid \mathcal{D}_{\mathrm{fit}}) \le \eta_\theta. \tag{46}$$

For $\lambda > 0$, define

$$\begin{aligned} V_i &= \lambda I_p + \sum_{r=1}^{R_{\max}} x_{ir}x_{ir}^\top, \\ \widehat{\theta}_i &= V_i^{-1}\sum_{r=1}^{R_{\max}} x_{ir}Q_{ir}^{\mathrm{obs}}, \qquad (47) \\ \beta_i &= R_*\sqrt{2\log\left\{\frac{\det(V_i)^{1/2}}{\det(\lambda I_p)^{1/2}\eta_\nu}\right\}} + \sqrt{\lambda}S, \\ \mathcal{E}_i &= \{\theta : \|\theta - \widehat{\theta}_i\|_{V_i} \le \beta_i\}. \qquad (48) \end{aligned}$$

Lemma 7 (Time-uniform ridge recovery coverage)**.** *Under the preceding assumptions,*

$$\mathbb{P}(\theta_i \in \mathcal{E}_i \mid \mathcal{D}_{\mathrm{fit}}) \geq 1 - \eta_\nu - \eta_\theta - \eta_H.$$

The statement holds at every bounded predictable recovery stopping rule and under predictable missingness.

Proof. Apply the self-normalized vector-martingale inequality to the predictable vectors x_{ir} and the stopped increments $\bar{\nu}_{ir}$. The MGF bound in Lemma 6 verifies the required conditional sub-Gaussian condition without conditioning on $E_{H,i}$. Consequently, with conditional probability at least $1 - \eta_\nu$, simultaneously for all $r \leq R_{\max}$,

$$\left\| \sum_{s \leq r} x_{is} \bar{\nu}_{is} \right\|_{V_{ir}^{-1}} \leq R_* \sqrt{2 \log \left\{ \frac{\det(V_{ir})^{1/2}}{\det(\lambda I_p)^{1/2} \eta_\nu} \right\}}$$

where $V_{ir} = \lambda I_p + \sum_{s \leq r} x_{is} x_{is}^\top$. Predictable omission sets $x_{ir} = 0$, so it also sets the corresponding vector increment to zero. A bounded predictable stopping rule may be substituted for r because the inequality holds simultaneously over the entire bounded window.

Let $\mathcal{A}_i$ denote this time-uniform martingale event and let $\mathcal{B}_i = \{\|\theta_i\|_2 \leq S\}$. On $\mathcal{A}_i \cap \mathcal{B}_i \cap E_{H,i}$, the stopped and original pseudo-noise sequences coincide at every included visit. The ridge normal equation is therefore

$$V_i(\widehat{\theta}_i - \theta_i) = \sum_r x_{ir} \nu_{ir} - \lambda \theta_i = \sum_r x_{ir} \bar{\nu}_{ir} - \lambda \theta_i.$$

Multiplication by $V_i^{-1/2}$, followed by the triangle inequality, gives

$$\|\widehat{\theta}_i - \theta_i\|_{V_i} \leq \left\| \sum_r x_{ir} \nu_{ir} \right\|_{V_i^{-1}} + \|\lambda \theta_i\|_{V_i^{-1}}.$$

Since $V_i \succeq \lambda I_p$, we have $V_i^{-1} \preceq \lambda^{-1} I_p$, and hence

$$\|\lambda \theta_i\|_{V_i^{-1}} \leq \sqrt{\lambda} \|\theta_i\|_2 \leq \sqrt{\lambda} S.$$

The two terms are exactly those used in β_i, so $\theta_i \in \mathcal{E}_i$ on the intersection event. Finally,

$$\mathbb{P}(\mathcal{A}_i^c \cup \mathcal{B}_i^c \cup E_{H,i}^c \mid \mathcal{D}_{\mathrm{fit}}) \leq \eta_\nu + \eta_\theta + \eta_H$$

by the union bound. This proves the stated conditional coverage and also explains why η_H must be charged even though the martingale itself is analyzed through stopping. □

For any legal center $\widehat{m}_i$ and positive legal scale s_i, set

$$S_i^o = \sup_{u \in \mathcal{U}} \frac{|b(u)^\top \theta_i - \widehat{m}_i(u)|}{s_i(u)}, \tag{49}$$

$$R_i^+ = \sup_{u \in \mathcal{U}} \frac{|b(u)^\top \widehat{\theta}_i - \widehat{m}_i(u)| + \beta_i \|b(u)\|_{V_i^{-1}}}{s_i(u)}. \tag{50}$$

Let $\eta_{\mathrm{rec}} = \eta_\nu + \eta_\theta + \eta_H$, choose $\gamma_{\mathrm{rk}} > 0$, and put $k = \lceil (m+1)(1 - \gamma_{\mathrm{rk}}) \rceil$. The corrected critical value q^+ is the kth order statistic of the m calibration scores when $k \leq m$, and is $+\infty$ otherwise.

For a desired total miscoverage level α, the user must choose the rank and recovery budgets before examining the calibration scores so that

$$\gamma_{\mathrm{rk}} + \eta_\nu + \eta_\theta + \eta_H \leq \alpha. \tag{51}$$

The finite-sample rank is nontrivial only when $k \leq m$, equivalently $\gamma_{\mathrm{rk}} \geq 1/(m+1)$. If this condition fails, the stated $+\infty$ convention remains valid but produces an uninformative band.

Robust ghost algorithm. The following objects have distinct roles and are fixed by a common rule for all calibration and future subjects.

1. Fit the arm-specific nuisance regressions $\widehat{m}_{a,ir}$ using $\mathcal{D}_{\mathrm{fit}}$, and freeze them before any calibration or future recovery record is processed. These regressions enter only the AIPW label Q_{ir}.
2. From the subject's legal history, construct the prediction center $\widehat{m}_i(u)$ and the positive scale $s_i(u)$. Neither object may use that subject's recovery outcomes.
3. For each calibration subject, use the disjoint recovery block to form $x_{ir} = M_{ir}b(U_{ir})$ and the included-only values Q_{ir}^{obs}, compute $\widehat{\theta}_i$, V_i, and β_i, and then evaluate the single subject-level score R_i^+.
4. Compute the corrected order statistic q^+ from the m calibration scores. For the future subject, report $\widehat{m}_{m+1}(u) \pm q^+ s_{m+1}(u)$. The reported future band uses no future recovery outcome.

The proof appends a recovery block for the future subject on an enlarged probability space. This ghost block is generated by the same prespecified kernel as the calibration recovery blocks, but it is never observed or used by the reported procedure.

Theorem 8 (Finite-sample exact-dictionary ghost band)**.** *Conditional on $\mathcal{D}_{\mathrm{fit}}$, suppose the $m+1$ augmented subject records are exchangeable and the same measurable history, recovery, missingness, center, scale, and score rules are applied to all records. If $\gamma_{\mathrm{tot}} = \gamma_{\mathrm{rk}} + \eta_{\mathrm{rec}} < 1$, then*

$$C_{m+1}(u) = [\widehat{m}_{m+1}(u) - q^+ s_{m+1}(u), \widehat{m}_{m+1}(u) + q^+ s_{m+1}(u)] \tag{52}$$

satisfies

$$\mathbb{P}\{\tau_{m+1}(u) \in C_{m+1}(u)\ \forall u \in \mathcal{U} \mid \mathcal{D}_{\mathrm{fit}}\} \geq 1 - \gamma_{\mathrm{tot}}.$$

Only the future recovery event is charged.

Proof. Condition on $\mathcal{D}_{\mathrm{fit}}$. Append to the future record an independent recovery block drawn from the same prespecified conditional kernel used for every calibration subject. By assumption, the augmented subject records are exchangeable, and every center, scale, recovery set, and score is obtained by applying the same measurable map to its own augmented record. Hence $R_1^+, \ldots, R_m^+, R_{m+1}^+$ are exchangeable.

Let $k = \lceil (m+1)(1-\gamma_{\mathrm{rk}}) \rceil$. The usual exchangeable-rank argument, with conservative treatment of ties, gives

$$\mathbb{P}(R_{m+1}^+ \leq q^+ \mid \mathcal{D}_{\mathrm{fit}}) \geq 1 - \gamma_{\mathrm{rk}}.$$

This statement remains true under the infinite critical-value convention when $k > m$.

Now let $E_{m+1} = \{\theta_{m+1} \in \mathcal{E}_{m+1}\}$. On this event, for every prediction location u, Cauchy–Schwarz in the V_{m+1} norm gives

$$\begin{aligned} |b(u)^\top \theta_{m+1} - \widehat{m}_{m+1}(u)| &\leq |b(u)^\top \widehat{\theta}_{m+1} - \widehat{m}_{m+1}(u)| \\ &\quad + |b(u)^\top (\theta_{m+1} - \widehat{\theta}_{m+1})| \\ &\leq |b(u)^\top \widehat{\theta}_{m+1} - \widehat{m}_{m+1}(u)| + \beta_{m+1} \|b(u)\|_{V_{m+1}^{-1}}. \end{aligned}$$

After division by $s_{m+1}(u) > 0$ and taking the supremum over u, this is precisely $S_{m+1}^o \leq R_{m+1}^+$. Therefore

$$\{R_{m+1}^+ \leq q^+\} \cap E_{m+1} \subseteq \{S_{m+1}^o \leq q^+\}.$$

Lemma 7 gives $\mathbb{P}(E_{m+1}^c \mid \mathcal{D}_{\mathrm{fit}}) \leq \eta_{\mathrm{rec}}$. No independence between E_{m+1} and the rank event is needed: the union bound implies

$$\mathbb{P}(S_{m+1}^o \leq q^+ \mid \mathcal{D}_{\mathrm{fit}}) \geq 1 - \gamma_{\mathrm{rk}} - \eta_{\mathrm{rec}}.$$

Finally, $S_{m+1}^o \leq q^+$ is exactly the simultaneous band-coverage event and contains no appended recovery outcome. Integrating out the ghost block therefore leaves the probability of the reported event unchanged. Notice that the proof uses only the future recovery event; it never intersects the calibration ellipsoid events. □

Lemma 8 (Fixed-design robust OLS ellipsoid)**.** *Suppose the complete recovery design $X_i \in \mathbb{R}^{T_R \times p}$ is fixed at the start of recovery, or exogenous and conditioned upon, and is full rank. Let $\mathscr{F}_{r-1}$ contain the preceding recovery errors and assume, for every real t and each r, $\mathbb{E}\{\exp(t\nu_r) \mid X_i, \mathcal{D}_{\mathrm{fit}}, \mathscr{F}_{r-1}\} \le \exp(t^2 R_*^2/2)$. Then the OLS estimator obeys*

$$\mathbb{P}\left\{\|\widehat{\theta}_i^{\mathrm{OLS}} - \theta_i\|_{X_i^\top X_i} \le 2R_*\sqrt{2\{p\log 5 + \log(1/\eta_\nu)\}} \mid X_i, \mathcal{D}_{\mathrm{fit}}\right\} \ge 1 - \eta_\nu.$$

Proof. Write $G = X^\top X$, which is positive definite by the full-rank assumption, and put $W = G^{-1/2}X^\top \nu$. For a fixed unit vector $v \in \mathbb{R}^p$,

$$v^\top W = \sum_{r=1}^{T_R} c_r \nu_r, \qquad c = XG^{-1/2}v.$$

The coefficient vector has squared norm

$$\sum_r c_r^2 = v^\top G^{-1/2} X^\top X G^{-1/2} v = 1.$$

Conditional on the fixed design, iterating the displayed conditional MGF bound and using $\sum_r c_r^2 = 1$ therefore gives

$$\mathbb{P}(v^\top W > t \mid X, \mathcal{D}_{\mathrm{fit}}) \le \exp\{-t^2/(2R_*^2)\}.$$

Let $\mathcal{N}$ be a one-half net of the unit sphere with $|\mathcal{N}| \le 5^p$. If $W \ne 0$, choose $v \in \mathcal{N}$ within distance one-half of $W/\|W\|_2$. Then $v^\top W \ge \|W\|_2/2$, and hence

$$\|W\|_2 \le 2\max_{v\in\mathcal{N}} v^\top W.$$

A union bound over the net shows that, with probability at least $1 - \eta_\nu$,

$$\|W\|_2 \le 2R_*\sqrt{2\{p\log 5 + \log(1/\eta_\nu)\}}.$$

Finally,

$$\|\widehat{\theta}^{\mathrm{OLS}} - \theta\|_G = \|G^{1/2}G^{-1}X^\top\nu\|_2 = \|W\|_2,$$

which proves the result. □

This OLS lemma is not claimed for arbitrary adaptive predictable designs: the inverse full-design matrix then depends on future design rows and its net coefficients need not be predictable.

Corollary 2 (Fixed-design OLS ghost band)**.** *Under the exact dictionary, suppose the calibration and ghost recovery matrices are generated by the same fixed or exogenous full-rank rule. Define*

$$\beta_i^{\mathrm{OLS}} = 2R_*\sqrt{2\{p\log 5 + \log(1/\eta_\nu)\}}$$

and

$$R_i^{+,\mathrm{OLS}} = \sup_{u\in\mathcal{U}} \frac{|b(u)^\top \widehat{\theta}_i^{\mathrm{OLS}} - \widehat{m}_i(u)| + \beta_i^{\mathrm{OLS}}\|b(u)\|_{(X_i^\top X_i)^{-1}}}{s_i(u)}.$$

If the OLS ellipsoid fails with probability at most η_{OLS}, the corrected $(1-\gamma_{\mathrm{rk}})$ quantile of the calibration $R_i^{+,\mathrm{OLS}}$ scores gives simultaneous coverage at least $1 - \gamma_{\mathrm{rk}} - \eta_{\mathrm{OLS}}$. When R_* is obtained from the stopped primitive-envelope construction above, $\eta_{\mathrm{OLS}} = \eta_\nu + \eta_H$.

Proof. On the OLS ellipsoid event, Cauchy–Schwarz in the $X_i^\top X_i$ norm gives, for every u,

$$|b(u)^\top(\widehat{\theta}_i^{\mathrm{OLS}} - \theta_i)| \le \beta_i^{\mathrm{OLS}}\|b(u)\|_{(X_i^\top X_i)^{-1}}.$$

Thus the OLS feasible score dominates the latent oracle score for the future subject. The proof of Theorem 8 now applies word for word with R_i^+ replaced by $R_i^{+,\mathrm{OLS}}$: exchangeability pays the rank error, and the union bound pays only the future OLS-ellipsoid failure. If the scalar MGF proxy is valid only before the envelope-violation time, the stopped-process argument adds η_H to that failure probability. □

Approximate dictionaries Now allow

$$\tau_i(u) = b(u)^\top \theta_i + r_i(u). \tag{53}$$

Let r_i^R collect $M_{ir} r_i(U_{ir})$. Suppose identically constructed honest envelopes obey

$$\mathbb{P}(\|r_i^R\|_2 > B_{R,i} \mid \mathcal{D}_{\text{fit}}) \le \eta_R, \qquad \mathbb{P}\left\{\sup_u \frac{|r_i(u)|}{s_i(u)} > a_{p,i} \mid \mathcal{D}_{\text{fit}}\right\} \le \eta_U.$$

These bounds must be fixed by the fitting sample, the legal subject record, or prespecified scientific knowledge. The theorem does not estimate them from the unobserved future latent curve. If they are available only through the known data-generating distribution, the resulting procedure is an oracle-envelope procedure. A legal prediction-domain envelope also implies the recovery bound

$$B_{R,i} = a_{p,i} \left\{ \sum_r M_{ir} s_i(U_{ir})^2 \right\}^{1/2}.$$

Lemma 9 (Recovery approximation bias)**.** *For $V_i = \lambda I + X_i^\top X_i$,*

$$\|X_i^\top r_i^R\|_{V_i^{-1}} \le \|r_i^R\|_2.$$

Proof. The squared norm on the left is

$$\|X_i^\top r_i^R\|_{V_i^{-1}}^2 = (r_i^R)^\top X_i V_i^{-1} X_i^\top r_i^R.$$

Take a singular-value decomposition $X_i = U_i D_i W_i^\top$. If d_{ij} denotes an eigenvalue of $X_i^\top X_i$, then the corresponding nonzero eigenvalue of $X_i V_i^{-1} X_i^\top$ is

$$\frac{d_{ij}}{\lambda + d_{ij}},$$

which lies in $[0, 1]$. Every remaining eigenvalue is zero. Hence $0 \preceq X_i V_i^{-1} X_i^\top \preceq I$, and therefore

$$(r_i^R)^\top X_i V_i^{-1} X_i^\top r_i^R \le (r_i^R)^\top r_i^R.$$

Taking square roots proves the claim. □

Inflate the radius to

$$\beta_i^{\text{app}} = R_* \sqrt{2 \log \left\{ \frac{\det(V_i)^{1/2}}{\det(\lambda I_p)^{1/2} \eta_\nu} \right\}} + \sqrt{\lambda} S + B_{R,i}$$

and define

$$R_i^{+,\text{app}} = \sup_u \frac{|b(u)^\top \widehat{\theta}_i - \widehat{m}_i(u)| + \beta_i^{\text{app}} \|b(u)\|_{V_i^{-1}} + a_{p,i} s_i(u)}{s_i(u)}.$$

Theorem 9 (Finite-sample approximate-dictionary ghost band)**.** *Under Theorem 8 and the two approximation envelopes, the corrected $(1 - \gamma_{\text{rk}})$ quantile of the calibration $R_i^{+,\text{app}}$ scores yields simultaneous coverage at least*

$$1 - \gamma_{\text{rk}} - \eta_\nu - \eta_\theta - \eta_H - \eta_R - \eta_U.$$

If one joint approximation event supplies both envelopes, its failure probability is charged only once. To target total miscoverage α, the six separately constructed budgets must satisfy

$$\gamma_{\text{rk}} + \eta_\nu + \eta_\theta + \eta_H + \eta_R + \eta_U \le \alpha,$$

with the obvious removal of one term when a single joint event controls both approximation errors.

Proof. On the approximate model, $Q_i = X_i \theta_i + r_i^R + \nu_i$ over the included recovery visits. The ridge normal equation therefore becomes

$$V_i(\widehat{\theta}_i - \theta_i) = X_i^\top \nu_i + X_i^\top r_i^R - \lambda \theta_i.$$

Consider the intersection of the stopped martingale event, the coefficient event $\{\|\theta_i\|_2 \le S\}$, the arm-envelope event, and the recovery approximation event $\{\|r_i^R\|_2 \le B_{R,i}\}$. On this intersection, the triangle inequality in the V_i^{-1} norm, Lemma 9, and $\|\lambda\theta_i\|_{V_i^{-1}} \le \sqrt{\lambda}S$ give

$$\begin{aligned}\|\widehat{\theta}_i - \theta_i\|_{V_i} &\le \|X_i^\top \nu_i\|_{V_i^{-1}} + \|X_i^\top r_i^R\|_{V_i^{-1}} + \|\lambda\theta_i\|_{V_i^{-1}} \\ &\le R_* \sqrt{2\log\left\{\frac{\det(V_i)^{1/2}}{\det(\lambda I_p)^{1/2}\eta_\nu}\right\}} + B_{R,i} + \sqrt{\lambda}S = \beta_i^{\mathrm{app}}.\end{aligned}$$

On the prediction-domain event $\sup_u |r_i(u)|/s_i(u) \le a_{p,i}$, Cauchy–Schwarz then gives, for every u,

$$\begin{aligned}|\tau_i(u) - \widehat{m}_i(u)| &= |b(u)^\top\theta_i + r_i(u) - \widehat{m}_i(u)| \\ &\le |b(u)^\top\widehat{\theta}_i - \widehat{m}_i(u)| + |b(u)^\top(\theta_i - \widehat{\theta}_i)| + |r_i(u)| \\ &\le |b(u)^\top\widehat{\theta}_i - \widehat{m}_i(u)| + \beta_i^{\mathrm{app}}\|b(u)\|_{V_i^{-1}} + a_{p,i}s_i(u).\end{aligned}$$

After standardization and maximization, the future approximate feasible score dominates its latent oracle score on this intersection. Identical construction of $B_{R,i}$ and $a_{p,i}$ preserves exchangeability of the augmented feasible scores. The corrected-rank argument pays γ_{rk}. A union bound for the future subject pays $\eta_\nu + \eta_\theta + \eta_H + \eta_R + \eta_U$. No calibration recovery or approximation event is intersected. If one joint approximation event implies both required bounds, the same event appears only once in this union bound. □

Proxy-to-oracle quantile equivalence For row m, let $(R_{i,m}^+, S_{i,m}^o)$, $i = 1, \dots, m$, be iid coupled pairs with cdfs F_m^+ and F_m^o. Let F be a proper continuous cdf, $a \in (0,1)$, and $q = F^{-1}(a)$.

Theorem 10 (Triangular-array proxy quantiles)**.** *Suppose (i) for every $\epsilon > 0$, $\mathbb{P}(|R_{i,m}^+ - S_{i,m}^o| > \epsilon) \to 0$; (ii) $\|F_m^o - F\|_\infty \to 0$; (iii) F is strictly locally increasing at q; and (iv) the two scores are finite almost surely. With $k_m = \lceil (m+1)a \rceil$, both corrected empirical order statistics converge in probability to q, and their difference is $o_p(1)$.*

Proof. For every $\epsilon > 0$, writing $\pi_m(\epsilon) = \mathbb{P}(|R^+ - S^o| > \epsilon)$,

$$F_m^o(t-\epsilon) - \pi_m(\epsilon) \le F_m^+(t) \le F_m^o(t+\epsilon) + \pi_m(\epsilon).$$

The left inequality follows because $\{S^o \le t - \epsilon, |R^+ - S^o| \le \epsilon\} \subseteq \{R^+ \le t\}$. The right inequality follows from $\{R^+ \le t, |R^+ - S^o| \le \epsilon\} \subseteq \{S^o \le t + \epsilon\}$.

Convergence in probability of the score difference permits a deterministic sequence $\epsilon_m \downarrow 0$ such that $\pi_m(\epsilon_m) \to 0$. A proper continuous cdf on the real line is uniformly continuous. Hence

$$\sup_t |F(t+\epsilon_m) - F(t)| + \sup_t |F(t-\epsilon_m) - F(t)| \longrightarrow 0.$$

Combining this fact with $\|F_m^o - F\|_\infty \to 0$ in the preceding cdf sandwich gives $\|F_m^+ - F\|_\infty \to 0$.

For either iid row, the Dvoretzky–Kiefer–Wolfowitz inequality gives

$$\mathbb{P}(\|\widehat{F}_m - F_m\|_\infty > t) \le 2e^{-2mt^2}.$$

Thus both empirical cdfs converge uniformly in probability to F. Moreover $k_m/m = a + O(m^{-1})$, and $k_m \le m$ eventually because $a < 1$.

Fix a sufficiently small $\delta > 0$. Strict local increase at q and the definition $q = F^{-1}(a)$ imply

$$F(q-\delta) < a < F(q+\delta).$$

The two strict gaps remain positive after replacing F by either empirical cdf and a by k_m/m, with probability tending to one. On that event, the corresponding k_mth order statistic lies in $(q-\delta, q+\delta]$. Since δ is arbitrary, both empirical order statistics converge in probability to q. Their difference is bounded by the sum of their absolute deviations from q, and is therefore $o_p(1)$. □

Corollary 3 (Primitive exact-dictionary score coupling)**.** *Consider a sequence of exact-dictionary recovery problems. Let $B_m = \sup_u \|b_m(u)\|_2$, let $s_{\min,m} = \inf_u s_{i,m}(u) > 0$, and suppose that a generic subject's recovery Gram matrix satisfies*

$$\lambda_{\min}(X_i^\top X_i) \geq cR_m$$

with probability tending to one for some $c > 0$. Suppose also that $\eta_{\nu,m} + \eta_{\theta,m} + \eta_{H,m} \to 0$ and

$$\frac{B_m}{s_{\min,m}\sqrt{\lambda_m + cR_m}} \left[R_{*,m} \sqrt{p_m \log\left\{1 + \frac{R_m B_m^2}{\lambda_m p_m}\right\} + 2\log(1/\eta_{\nu,m})} + \sqrt{\lambda_m} S_m \right] \longrightarrow 0. \quad (54)$$

Then $R^+_{i,m} - S^o_{i,m} = o_p(1)$. Consequently, if the oracle-score laws also satisfy conditions (ii)–(iv) of Theorem 10, the feasible and oracle corrected empirical quantiles differ by $o_p(1)$.

Proof. On the recovery ellipsoid event, for every u,

$$|b_m(u)^\top(\widehat{\theta}_i - \theta_i)| \leq \beta_{i,m} \|b_m(u)\|_{V_{i,m}^{-1}}.$$

Applying this inequality in both directions to the absolute prediction errors yields

$$0 \leq R^+_{i,m} - S^o_{i,m} \leq 2\beta_{i,m} \sup_u \frac{\|b_m(u)\|_{V_{i,m}^{-1}}}{s_{i,m}(u)}. \quad (55)$$

On the stable-design event,

$$\sup_u \|b_m(u)\|_{V_{i,m}^{-1}} \leq \frac{B_m}{\sqrt{\lambda_m + cR_m}}.$$

It remains to bound $\beta_{i,m}$. Since $\operatorname{tr}(X_i^\top X_i) \leq R_m B_m^2$, concavity of the logarithm gives

$$\log \frac{\det(V_{i,m})}{\det(\lambda_m I_{p_m})} \leq p_m \log\left\{1 + \frac{R_m B_m^2}{\lambda_m p_m}\right\}.$$

Substitution in the definition of $\beta_{i,m}$, followed by (54), makes the right side of (55) tend to zero. The recovery ellipsoid and stable-design events have probability tending to one, so the score coupling holds in probability. The quantile conclusion is then an immediate application of Theorem 10. □

For the approximate dictionary, the same argument applies if, in addition,

$$\frac{B_m B_{R,m}}{s_{\min,m}\sqrt{\lambda_m + cR_m}} + a_{p,m} \longrightarrow 0, \qquad \eta_{R,m} + \eta_{U,m} \longrightarrow 0.$$

The first term controls approximation on the recovery design. The second controls approximation on the prediction domain. Neither may be omitted.

C.2 Gaussian-innovation recovery for the end-to-end experiment

The experiment in Section 4.3 instantiates the generic recovery-set condition of Theorem 2 as follows. It uses the original model with independent Gaussian visit innovations. The random-effect coefficients may have either of the two distributions specified in the experiment; their Gaussianity is not required for this recovery result. This provides an alternative to the AIPW martingale construction in Section C.1.

Let

$$v(x,z) = \{1, \sqrt{3}x, \sqrt{5}(3x^2-1)/2, \sqrt{3}(2z-1)\}^\top.$$

This is the experiment's instance of the effect dictionary v in Section 2.3, with the same coefficient ζ_i. The specified mean and random functions have this exact finite representation. Each subject's arm-specific conditional mean is exactly $v(x,z)^\top \beta_{ia}$, where the intercept combines the population and random intercepts. Stack $\beta_i = (\beta_{i0}^\top, \beta_{i1}^\top)^\top$, and write $C = (-I_4, I_4)$, so that $\zeta_i = C\beta_i$ and $\tau_i(x,z) = v(x,z)^\top \zeta_i$. The r recovery rows are $W_{ir} = \{(1 - A_{ir})v(U_{ir})^\top, A_{ir} v(U_{ir})^\top\}$. Stacking these rows gives $W_i \in \mathbb{R}^{r \times 8}$. Histories and recovery blocks are prospectively fixed and disjoint.

Conditional on the persistent coefficients, fitting sample, history and exogenous recovery design W_i, assume

$$Y_i^{\mathrm{rec}} = W_i\beta_i + \varepsilon_i^{\mathrm{rec}}, \qquad \varepsilon_i^{\mathrm{rec}} \sim N(0, \sigma_\varepsilon^2 I_r), \qquad 0 < \sigma_\varepsilon \le \bar{\sigma}.$$

The externally specified upper bound $\bar{\sigma}$ is fixed before calibration; it is not the fitted noise estimate used in the FPCA predictor.

If W_i has column rank eight, define

$$\widetilde{\zeta}_i = C(W_i^\top W_i)^{-1} W_i^\top Y_i^{\mathrm{rec}}, \qquad K_i = C(W_i^\top W_i)^{-1} C^\top, \tag{56}$$

$$b_{\mathrm{rec}} = \bar{\sigma}\sqrt{\chi^2_{4,1-\eta_{\mathrm{rec}}}}, \qquad \mathcal{E}_i = \{\zeta : (\zeta - \widetilde{\zeta}_i)^\top K_i^{-1}(\zeta - \widetilde{\zeta}_i) \le b_{\mathrm{rec}}^2\}. \tag{57}$$

Here $\chi^2_{d,p}$ denotes the p quantile of a chi-square variable with d degrees of freedom. The recovery precision is $V_i = K_i^{-1}$ in Section 2.3; it differs from the within-history covariance $\widehat{\mathcal{V}}_i$ in the FPC score predictor. If the recovery design is rank deficient, take $\mathcal{E}_i = \mathbb{R}^4$ and $R_i^+ = +\infty$. This rule retains the subject and preserves the common exchangeable score rule.

Proposition 7 (Recovery coverage under Gaussian innovations)**.** *Under the preceding conditions,* $\Pr(\zeta_i \notin \mathcal{E}_i \mid \mathcal{D}_{\mathrm{fit}}) \le \eta_{\mathrm{rec}}$. *Consequently, the band in* (9)*, using the score*

$$R_i^+ = \max_{u \in \mathcal{U}_G} \frac{|v(u)^\top \widetilde{\zeta}_i - \widehat{c}_{i,D}(u)| + b_{\mathrm{rec}}\sqrt{v(u)^\top K_i v(u)}}{\widehat{s}_i(u)}$$

for a full-rank recovery design and $+\infty$ *otherwise, has simultaneous grid coverage at least* $1 - \gamma_{\mathrm{rk}} - \eta_{\mathrm{rec}}$ *under the exchangeability and frozen-rule conditions of Theorem 2.*

Proof. When W_i has full column rank, K_i is positive definite because C has row rank four. Conditional on the variables in the stated innovation model, $\widetilde{\zeta}_i - \zeta_i$ is normal with covariance $\sigma_\varepsilon^2 K_i$. Whiten this four-dimensional vector to obtain

$$\frac{(\widetilde{\zeta}_i - \zeta_i)^\top K_i^{-1}(\widetilde{\zeta}_i - \zeta_i)}{\sigma_\varepsilon^2} \sim \chi^2_4.$$

Since $\bar{\sigma} \ge \sigma_\varepsilon$, the conditional failure probability of (57) is at most η_{rec}. It is zero on the rank-deficient branch. Integrating over the histories, designs and persistent coefficients gives the displayed conditional-on-fit bound. On the recovery event, Cauchy–Schwarz bounds $|v^\top(\zeta_i - \widetilde{\zeta}_i)|$ by $b_{\mathrm{rec}}\sqrt{v^\top K_i v}$, simultaneously over the grid. The common score rule is therefore a feasible upper score. Theorem 2 supplies the rank argument and the truncation containment step. □

The individual regression in (56) constructs only the calibration confidence set. The future person's point prediction continues to use raw-history FPC score prediction; future recovery outcomes are unnecessary. In the experiment $\bar{\sigma} = .5$ and the innovation standard deviation is .35, with $\eta_{\mathrm{rec}} = .01$ and $\gamma_{\mathrm{rk}} = .04$. Thus the recovery bound is conservative and its effect on band width is part of the reported comparison. The result does not cover an arbitrary non-Gaussian innovation law or an invalid noise upper bound.

A residual-studentized recovery set. The reported same-realization comparison evaluates this branch alongside the externally bounded-noise rule above. Their assumptions are not nested: both use a complete exogenous design and conditionally spherical Gaussian innovations; this branch removes the external noise upper bound but requires $R > 8$ and a within-person residual scale. Conditional on the persistent coefficients, fitting sample, history, and complete exogenous recovery design, suppose

$$Y_i^{\mathrm{rec}} = W_i\beta_i + \varepsilon_i^{\mathrm{rec}}, \qquad \varepsilon_i^{\mathrm{rec}} \sim N(0, \sigma_i^2 I_R), \quad 0 < \sigma_i^2 < \infty,$$

where $R > 8$ and $\mathrm{rank}(W_i) = 8$. Let $\mathscr{F}_i^F = \sigma(\mathcal{D}_{\mathrm{fit}}, \text{persistent coefficients}, \text{history}, W_i, R)$, where W_i is the complete exogenous recovery design. All conditional probabilities in this paragraph and proposition are given $\mathscr{F}_i^F$. Let $P_i^{\mathrm{rec}} = W_i(W_i^\top W_i)^{-1}W_i^\top$ and

$$s_{\varepsilon,i}^2 = \frac{\|(I_R - P_i^{\mathrm{rec}})Y_i^{\mathrm{rec}}\|_2^2}{R - 8}, \qquad b_{F,i} = s_{\varepsilon,i}\sqrt{4F_{4,R-8,1-\eta_{\mathrm{rec}}}},$$

where $s_{\varepsilon,i}$ is the nonnegative square root and $F_{d_1,d_2,p}$ is the p quantile of an F_{d_1,d_2} distribution. Define

$$\mathcal{E}_i^F = \{\zeta : (\zeta - \widetilde{\zeta}_i)^\top K_i^{-1}(\zeta - \widetilde{\zeta}_i) \le b_{F,i}^2\}. \tag{58}$$

Proposition 8 (Exact residual-studentized recovery)**.** *On the full-rank $R > 8$ branch,* $\Pr(\zeta_i \in \mathcal{E}_i^F \mid \mathscr{F}_i^F) = 1 - \eta_{\mathrm{rec}}$. *If $R \le 8$ or W_i is rank deficient, set $\mathcal{E}_i^F = \mathbb{R}^4$ and the feasible score to infinity. Otherwise use*

$$R_i^{+,F} = \max_{u \in \mathcal{U}_G} \frac{|v(u)^\top \widetilde{\zeta}_i - \widehat{c}_{i,D}(u)| + b_{F,i}\sqrt{v(u)^\top K_i v(u)}}{\widehat{s}_i(u)}.$$

Under the exchangeability and frozen-rule conditions of Theorem 2, the protected grid band has coverage at least $1 - \gamma_{\mathrm{rk}} - \eta_{\mathrm{rec}}$.

Proof. Since C has row rank four and W_i has column rank eight, $K_i = C(W_i^\top W_i)^{-1}C^\top > 0$. Conditional on $\mathscr{F}_i^F$, $\widetilde{\zeta}_i - \zeta_i = C(W_i^\top W_i)^{-1}W_i^\top \varepsilon_i^{\mathrm{rec}} \sim N(0, \sigma_i^2 K_i)$, so its standardized quadratic form is χ_4^2. Also $I_R - P_i^{\mathrm{rec}}$ is an orthogonal projection of rank $R-8$, hence $(R-8)s_{\varepsilon,i}^2/\sigma_i^2 \sim \chi_{R-8}^2$. Their cross-covariance is $\sigma_i^2 C(W_i^\top W_i)^{-1}W_i^\top(I_R - P_i^{\mathrm{rec}}) = 0$; joint Gaussianity makes the quadratic forms independent. Therefore

$$\frac{(\widetilde{\zeta}_i - \zeta_i)^\top K_i^{-1}(\widetilde{\zeta}_i - \zeta_i)/4}{s_{\varepsilon,i}^2} \sim F_{4,R-8},$$

which proves the exact ellipsoid probability, also after integrating the conditioning variables. Ellipsoid support and Cauchy–Schwarz give the upper score. The original ghost theorem supplies the rank step and future recovery failure budget; the invalid branch is the whole space. □

This pivot requires a complete exogenous design and conditionally spherical joint Gaussian innovations. It does not follow from marginal or martingale Gaussian errors under an adaptive design. Penalized or GCV residuals, or a global noise estimate, do not preserve this exact F claim. Its random scale and finite-denominator-degrees quantile can make the set wider than the bounded-noise set. In the archived paired simulation this rule reduced mean width, but that empirical comparison does not weaken the complete-design and spherical-Gaussian conditions above. The claim does not cover $R = 8$ or arbitrary missingness.

D Experimental details

D.1 Simulation studies

The simulations keep the two inferential targets separate. The population experiment evaluates the subject-block CR3 Gaussian-multiplier SCB for τ on a two-dimensional grid. The personalized experiment records the latent subject-specific treatment-effect curve itself, separates history from recovery visits, and evaluates simultaneous latent-curve coverage. Pointwise or factual-trajectory coverage is not used as a substitute for either target.

Population SCB The population experiment uses independent subjects, $T = 60$ visits, visit-level treatment $A_{it} \sim \mathrm{Bernoulli}(0.5)$, and mutually independent $X_{1it}, X_{2it}, Z_{it} \sim \mathrm{Unif}(0,1)$. Outcomes follow

$$Y_{it} = \mu_{A_{it}}(X_{1it}, X_{2it}) + \sum_{k=1}^{2} \xi_{ik}^{A_{it}} \phi_k^{A_{it}}(Z_{it}) + \varepsilon_{it}, \qquad \varepsilon_{it} \sim N(0, 0.5^2).$$

The loading functions are

$$\phi^1(z) = \sqrt{2}\{\sin(\pi z/2), \cos(\pi z/2)\}^\top, \qquad \phi^0(z) = \sqrt{2}\{\sin(3\pi z/2), \cos(3\pi z/2)\}^\top,$$

and the two score pairs are independent across k, with

$$\mathrm{Cov}\{(\xi_{i1}^0, \xi_{i1}^1)^\top\} = \begin{pmatrix} 1/5 & 1/18 \\ 1/18 & 1/9 \end{pmatrix}, \qquad \mathrm{Cov}\{(\xi_{i2}^0, \xi_{i2}^1)^\top\} = \begin{pmatrix} 1/25 & 1/48 \\ 1/48 & 1/36 \end{pmatrix}.$$

This construction preserves dependence across repeated visits and across actions within a participant.

For the exact-span experiment, each additive component uses K Gaussian radial functions $b_j(x) = \exp\{-5(x - c_j)^2\}$, with equally spaced centers from zero to one. The action-specific coefficient

vectors are fixed before simulation. For $j = 1, \ldots, K$, their two additive blocks are

$$\theta_{0j}^{(1)} = 0.70\cos\{j\pi/(K+1)\} + 0.25, \qquad \theta_{0j}^{(2)} = -0.45\sin\{j\pi/(K+1)\} + 0.15,$$
$$\theta_{1j}^{(1)} = 0.80\sin\{j\pi/(K+1)\} + 0.55, \qquad \theta_{1j}^{(2)} = 0.55\cos\{j\pi/(K+1)\} - 0.10.$$

Every method uses working-independence least squares, one CR3-adjusted score per participant, and a Gaussian multiplier maximum on a prespecified 15-by-15 grid. The oracle benchmark uses the known conditional covariance but the same fitted mean structure and grid.

Table 2 reports the exact-span sample-size ladder at $K = 8$. The Rademacher multiplier is included only as a calibration sensitivity. The critical-value ratio is the feasible Gaussian multiplier critical value divided by the oracle value.

Table 2: Population SCB experiment: exact-basis multiplier results. Entries are simultaneous coverage of the population treatment-effect surface; the final column is the mean feasible-Gaussian/oracle critical-value ratio.

Subjects	Replications	Gaussian CR3	Rademacher CR3	Oracle	Critical ratio
25	250	0.808	0.704	0.940	0.988
50	250	0.908	0.876	0.972	0.996
100	1000	0.923	0.907	0.939	0.998
200	1000	0.932	0.927	0.937	0.999
400	1000	0.962	0.958	0.964	0.999

The critical-value ratio is already close to one at small n, while CR3 coverage approaches the oracle coverage as the number of independent subjects grows. The remaining small-sample gap is therefore driven mainly by studentization rather than multiplier critical-value calibration.

The nonparametric approximation experiment replaces the exact-span means by

$$\mu_1(x) = 1 + 4\sin(2\pi x_1) - 1.5\cos(2\pi x_2), \qquad \mu_0(x) = 2 - \sin(2\pi x_1) + 1.5\cos(2\pi x_2).$$

Table 3 varies the radial-basis dimension. The small dictionary has visible approximation bias. At $K = 8$, the bias is small relative to the band half-width and coverage is close to 0.95; larger dictionaries do not improve coverage at $n = 200$ because their increased estimation cost offsets the small reduction in approximation error.

Table 3: Gaussian-CR3 multiplier sensitivity to basis size for the nonparametric target ($n = 200$, $T = 60$, 250 replications). Here SE/SD is the grid-average of replication-averaged pointwise standard errors divided by the grid-average of pointwise empirical standard deviations across successful replications.

K	Coverage	Supremum bias / half-width	SE/SD	Mean width
6	0.840	0.442	1.003	0.352
8	0.952	0.048	0.997	0.428
10	0.932	0.057	0.973	0.448
12	0.868	0.064	0.910	0.450

Table 4 changes only the working precision in the estimating equation; every feasible row still uses participant-level CR3 scores and the Gaussian multiplier SCB. Working independence and diagonal marginal precision are essentially identical. Estimated functional precision shortens the band but undercovers and has one numerical failure at $n = 100$. Oracle precision demonstrates the possible efficiency gain when the covariance is known, but it is not a feasible procedure.

The source package includes the historical diagnostic code, but the original seed and invocation records for the early oracle ladder are incomplete; exact regeneration of those historical entries is not certified. The diagnostic ladder in Table 5 isolates why earlier SCB implementations failed. The ideal known-covariance Gaussian reference is well calibrated. Severe basis misspecification and unstable plug-in functional covariance estimation, rather than the Gaussian multiplier critical value itself, produce the catastrophic undercoverage. These diagnostics support working independence with the empirical CR3 sandwich as the primary feasible population procedure.

Table 4: Working-precision sensitivity in the exact-span DGP (1,000 replications per row, $T = 60$, $K = 8$). Here SE/SD is the grid-average of replication-averaged pointwise standard errors divided by the grid-average of pointwise empirical standard deviations across successful replications. Coverage, width, RMSE, and SE/SD summarize successful replicates; failure rate uses all scheduled replications.

n	Working precision	Coverage	Mean width	RMSE	SE/SD	Failure rate
100	Independence	0.922	0.607	0.092	1.004	0
100	Diagonal marginal	0.923	0.607	0.092	1.004	0
100	Estimated functional	0.906	0.520	0.081	0.963	0.001
100	Oracle precision	0.927	0.470	0.071	1.004	0
200	Independence	0.932	0.429	0.065	1.006	0
200	Diagonal marginal	0.934	0.429	0.065	1.006	0
200	Estimated functional	0.920	0.339	0.052	0.992	0
200	Oracle precision	0.946	0.331	0.050	1.007	0

Table 5: Completed non-policy SCB diagnostic ladder. These are diagnostic comparisons, not recommended feasible procedures.

Diagnostic	Settings or comparison	Simultaneous coverage
Ideal Gaussian reference	Exact basis, known covariance	0.960
Fixed-basis approximation	$K = 4$ / $K = 6$ / $K = 8$	0.000 / 0.920 / 0.950
Plug-in covariance	Raw / PSD plus fixed eigenvalue floor	0.000 / 0.192

HeartSteps-calibrated SCB stress test. We also retain the randomized arm of the semi-synthetic experiment based on the HeartSteps V1 visit times, availability patterns, pre-decision activity, and decision slots. Treatment is generated from the known 0.6 probability of receiving either type of activity suggestion. Potential outcomes on the $\log(1 + \text{steps}_{30})$ scale combine prespecified additive action means, a participant random intercept, and an autoregressive innovation. In 200 replications with 1,000 Gaussian multiplier draws, the 95% simultaneous coverage is 0.905 (Monte Carlo standard error 0.021), pointwise coverage is approximately 0.99, and mean total width is 0.919. With only 37 independent participants, this is a small-sample stress test rather than confirmation of nominal coverage.

The paired extension uses 384 replications per DGP and preserves the 37-person visit templates and 15-point grid. Within a replication, all DGPs and all eight dictionary–precision methods share treatment assignments, random intercepts, innovations, and Gaussian multipliers. D00 satisfies the day and quadratic-slot restrictions, whereas D11 restores both omitted components and equals the full calibrated contrast. The simulation evaluates both the common full surface and each dictionary-and-precision-specific randomization-averaged projection. Its coverage is finite-grid empirical evidence, not a finite-sample 95% guarantee or a continuous-domain claim.

HeartSteps strictly nested $N = 37/100/200$ semi-synthetic sample-size experiment. We separately examine the effect of increasing the number of independently generated participant clusters while holding the fitted HeartSteps mechanism fixed. The complete M0 baseline and treatment-effect models are fitted to the 37 observed participants and then frozen. Conditional on these fits and the finite participant-template library, outcomes are generated as

$$Y_{it}^* = \widehat{\mu}_0(X_{it}) + A_{it}^* \widehat{\tau}_{\text{M0}}(X_{it}) + b_i^* + e_{it}^*,$$

where $A_{it}^* \sim \text{Bernoulli}(.6)$, $b_i^* \sim \mathcal{N}(0, \widehat{\sigma}_b^2)$, and $e_{it}^* = \widehat{\rho} e_{i,t-1}^* + \eta_{it}^*$, with independent Gaussian innovations. The frozen calibration gives $\widehat{\sigma}_b = .3731$, innovation standard deviation 2.4202, and $\widehat{\rho} = -.0120$.

Each $N = 100$ sample retains the preceding construction: the 37 original participant-level visit and covariate templates plus 63 templates sampled with replacement from the same library. The $N = 200$ sample appends 100 further resampled templates. All participants receive newly generated treatments, random intercepts, innovations, and outcomes and have distinct identifiers. Separate prespecified random-number streams generate the added 100 participants and multiplier rows, so the first 37 and

100 participants are exact prefixes of the $N = 200$ sample. In particular, every $N = 37$ and $N = 100$ replication reproduces the preceding experiment.

All three sample sizes use M0 with working independence, participant-block CR3, and 1,000 participant-level Gaussian multiplier draws. The evaluation grid is frozen from the cleaned 37-participant template: the 10th, 50th, and 90th percentiles of standardized recent steps crossed with five decision slots, at standardized day zero. Under this correctly specified frozen mechanism, the full treatment-effect surface and its M0 population projection coincide on the grid; this equality need not hold under treatment-effect misspecification. Table 6 summarizes 1,000 strictly paired replications.

Table 6: HeartSteps strictly nested $N = 37/100/200$ semi-synthetic experiment. Coverage is simultaneous over the fixed 15-point grid; pointwise coverage is averaged over its 15 locations. Parentheses contain Monte Carlo standard errors across 1,000 paired replications.

N	Sim. cov. (MCSE)	Pointwise cov. (MCSE)	Width (MCSE)	Critical	Mean SE	Failures
37	.936 (.00774)	.99327 (.00088)	.91718 (.00164)	2.7924	.16422	0/1000
100	.926 (.00828)	.99253 (.00092)	.55457 (.00065)	2.7965	.09915	0/1000
200	.945 (.00721)	.99447 (.00078)	.39193 (.00034)	2.7993	.07001	0/1000

Table 7 separates the full-replication comparison from the common holdout used below. Across all 1,000 replications, the mean paired width ratios are .60588, .70729, and .42843 for $100/37$, $200/100$, and $200/37$. The corresponding coverage differences are $-.010$, .019, and .009, with MCSEs .01058, .00931, and .01054. Thus, the full-replication estimate at $N = 200$ is closer to .95 than the estimate at $N = 100$, but the comparison does not establish a monotone or general coverage improvement.

Table 7: Paired standard-band comparisons. The holdout uses the same replications 501–1,000 later used to evaluate the frozen DGP-oracle cutoffs. Parentheses contain paired Monte Carlo standard errors.

Scope	Comparison	Mean paired width ratio	Coverage difference	Mean paired SE ratio
All 1,000	100/37	.60588 (.00092)	−.010 (.01058)	.60483
All 1,000	200/100	.70729 (.00070)	.019 (.00931)	.70643
All 1,000	200/37	.42843 (.00072)	.009 (.01054)	.42726
Holdout 500	100/37	.60474 (.00129)	.000 (.01498)	.60342
Holdout 500	200/100	.70776 (.00100)	.002 (.01218)	.70695
Holdout 500	200/37	.42792 (.00102)	.002 (.01564)	.42660

To diagnose whether the width reduction is mainly an artifact of empirical coverage differences, we add an infeasible split-sample DGP-oracle analysis. For sample size N and replication r, define

$$T_{N,r} = \max_{u \in \mathcal{G}_{15}} \frac{|\widehat{\tau}_{N,r}(u) - \tau(u)|}{\widehat{\mathrm{se}}_{N,r}(u)}.$$

For each N, q_N is the 476th order statistic of $T_{N,r}$ from replications 1–500, where $476 = \lceil 501 \times .95 \rceil$. We freeze these three cutoffs before evaluating replications 501–1,000 and form the oracle band $\widehat{\tau}_N(u) \pm q_N \widehat{\mathrm{se}}_N(u)$.

The holdout standard-band coverages are .938, .938, and .940, while widths continue to fall. The DGP-oracle widths also fall substantially, and their decomposition shows larger changes in mean SE than in the cutoff. This supports precision improvement as the main explanation for the narrower bands in this frozen experiment. The oracle holdout coverages .964, .976, and .940 are not strictly matched, however, so this is not an equal-coverage efficiency comparison. It uses the known simulation target and is infeasible in applications. Its reported evaluation MCSEs condition on the frozen q_N and do not include calibration-cutoff uncertainty.

These are conditional semi-synthetic, finite-grid results. They average over template resampling, treatments, random effects, and multiplier draws while conditioning on the original fitted mechanism, standardization, grid, and finite template library. They do not account for uncertainty in the fitted data-generating mechanism, give a finite-sample 95% guarantee, establish simultaneous coverage over a continuous domain, or imply that recruiting 163 new participants from the target population would reproduce the same width or coverage changes. Actual recruitment could add covariate support and heterogeneity absent from the resampled template library.

Table 8: Split-sample DGP-oracle diagnostic. Panel A evaluates the standard and oracle bands on the common 500-replication holdout. Panel B gives the exact mean-width decomposition $\overline{W}^o_B/\overline{W}^o_A = (q_B/q_A)(\overline{\mathrm{SE}}_B/\overline{\mathrm{SE}}_A)$, where the two parenthesized factors are multiplied. Oracle MCSEs condition on the frozen cutoffs and exclude cutoff-estimation uncertainty.

Panel A: common holdout replications 501–1,000

N	Standard cov.	Standard width	q_N	Oracle cov. (MCSE)	Oracle width (MCSE)	Mean SE
37	.938	.91933	2.8993	.964 (.00834)	.95460 (.00226)	.16462
100	.938	.55482	3.0099	.976 (.00685)	.59694 (.00086)	.09916
200	.940	.39237	2.7876	.940 (.01063)	.39064 (.00041)	.07007

Panel B: oracle width-ratio decomposition

Comparison	Oracle width ratio of means	q ratio	Mean-SE ratio
100/37	.62533	1.03814	.60235
200/100	.65441	.92616	.70659
200/37	.40922	.96148	.42562

D.2 HEARTSTEPS ANALYSIS

We apply the population layer to the public HeartSteps V1 micro-randomized trial (Klasnja et al., 2019). Restricting to available decision times with an observed 30-minute step outcome and consistent randomization and delivery indicators gives 5,777 decision points from 37 participants. The outcome is $\log(1 + \text{steps})$. Active and anti-sedentary suggestions are combined as $A_{it} = 1$, no suggestion is coded as $A_{it} = 0$, and the trial's known probability of any suggestion is 0.6; the observed message rate is 0.592. The participant, rather than the decision point, is the sampling unit for the CR3 score.

We interpret the fitted contrasts as descriptive projections under the observed visit distribution and chosen working dictionary. A causal excursion-effect interpretation would require additional target and estimating-equation justification under longitudinal feedback; randomization alone does not establish that interpretation here.

For the population analysis, the working mean has separate coefficients by action for an intercept, standardized pre-decision steps, standardized study day, and four decision-slot indicators. We use all participants, participant-block CR3 scores, and 5,000 Gaussian multiplier draws. The SCB is evaluated at the product of three pre-decision-activity quantiles and five daily decision slots, with study day fixed at its standardized mean. The critical value is 2.805 and the average total width is 0.987. None of the 15 grid-point intervals excludes zero. This is inference for the historical randomized design, not evaluation of a newly learned adaptive policy.

For the leave-one-participant-out check, we remove each participant in turn, refit the working mean and CR3 band, and retain the full-sample grid. The largest absolute change in a point estimate is 0.100, attained when participant 30 is omitted. The median of the 37 participant-level maximum changes is 0.053. Across refits, the average total width ranges from 0.972 to 1.015, and no simultaneous interval excludes zero. These refits assess sensitivity to a single participant; they do not provide a new coverage statement or compensate for having only 37 independent clusters.

An additional prespecified sensitivity crosses four related treatment-contrast dictionaries with working independence and a fixed-calibration random- intercept precision. M0 is the seven-coordinate analysis above; M1 removes treatment-by-day moderation; M2 replaces four slot indicators by centered orthonormal linear and quadratic trends; M3 applies both restrictions. All retain the complete baseline mean, participant-block CR3, and the same 15-point grid. M0 with working independence remains primary. The restricted fits describe different observed-visit projections, and the same-outcome RI weights are exploratory plug-ins whose estimation contribution is not included in the reported sandwich.

The contrast dimensions of M0–M3 are $7, 6, 5$, and 4. For the RI branch, $V_i = \sigma^2 I_{T_i} + \tau^2 \mathbf{1}\mathbf{1}^\top$ is frozen from the HeartSteps template before outcomes are generated. Subject blocks are premultiplied by $V_i^{-1/2}$, after which we apply the same participant-block CR3 construction and common subject-level Gaussian multipliers. Each target is the randomization-averaged projection for its fixed visit template, dictionary, and precision. Under independent Bernoulli treatment with $p = .6$, diagonal terms use $E(A_{it}^2) = p$ and off-diagonal terms use $E(A_{it}A_{is}) = p^2$.

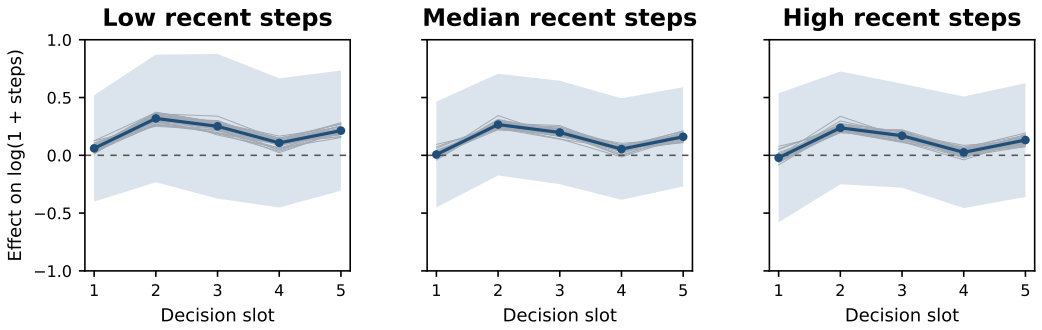


Figure 1: HeartSteps fitted population projection contrasts. The dark curve is the full-sample estimate and the shaded region is its nominal 95% simultaneous band. Each gray curve is the point estimate from a refit that omits one participant and uses the same 15 grid points; adjacent decision slots are joined only to aid reading. All panels use the common vertical range $[-1, 1]$.

Table 9: HeartSteps descriptive projection sensitivity on the fixed 15-point grid. RI denotes exploratory same-outcome plug-in random-intercept GLS followed by participant-block CR3; its bands omit the influence-function contribution from estimating the working weights. "Range" is the fitted contrast range; the last column counts grid intervals excluding zero. M0–WI remains the primary analysis.

Spec.	Precision	q	Critical	Width	Range	Excludes zero
M0	WI	7	2.805	0.987	$[-0.021, 0.319]$	0
M0	RI	7	2.813	0.977	$[0.001, 0.290]$	0
M1	WI	6	2.813	0.989	$[-0.014, 0.311]$	0
M1	RI	6	2.811	0.975	$[0.008, 0.281]$	0
M2	WI	5	2.746	0.799	$[0.040, 0.242]$	0
M2	RI	5	2.735	0.789	$[0.056, 0.213]$	0
M3	WI	4	2.749	0.801	$[0.046, 0.235]$	0
M3	RI	4	2.742	0.791	$[0.061, 0.206]$	0

The paired template simulation uses 384 replications for each DGP, with all four DGPs sharing actions, participant random intercepts, innovations, and multiplier draws within a replication. D00 satisfies both restrictions; D11 adds treatment-by-day moderation and the slot component orthogonal to the constant, linear, and quadratic slot basis, reproducing the calibrated full contrast. Table 10 reports working independence. Under D00, M2 and M3 reduce width from .916 to about .719 while their simultaneous coverage is .917. Under D11, their projection coverage remains .917, but coverage of the full surface falls to .742 and .737.

Table 10: HeartSteps-template simulation with working independence. Entries are means over 384 independent replications; parentheses are replication-level MCSEs. Coverage is simultaneous on the 15-point grid. No model is selected from these results.

DGP	Spec.	Full-surface coverage	Projection coverage	Width
D00	M0	.901 (.015)	.901 (.015)	.916 (.003)
D00	M1	.898 (.015)	.898 (.015)	.916 (.003)
D00	M2	.917 (.014)	.917 (.014)	.719 (.003)
D00	M3	.917 (.014)	.917 (.014)	.718 (.003)
D11	M0	.901 (.015)	.901 (.015)	.916 (.003)
D11	M1	.901 (.015)	.896 (.016)	.917 (.003)
D11	M2	.742 (.022)	.917 (.014)	.719 (.003)
D11	M3	.737 (.022)	.917 (.014)	.719 (.003)

D.3 RAW-HISTORY FPCA PREDICTION EXPERIMENTS

These experiments evaluate the fitted personal predictor and its model-based bands. They do not use a conformal calibration sample. The resulting model-based coverage is not evidence of coverage for the conformal procedure in Section 2.3, which is evaluated in the separate end-to-end study in Section 4.3.

Prespecified design Use $b(z) = (1, \sqrt{3}(2z-1))^\mathsf{T}$, so $D = 4$, and

$$a(x) = \left(1, \sqrt{3}x, \frac{\sqrt{5}}{2}(3x^2-1)\right)^\mathsf{T}, \quad \eta_0 = (.5, .3, .1)^\mathsf{T}, \quad \eta_1 = (.9, .5, -.05)^\mathsf{T}.$$

At each visit, $X \sim \mathrm{Unif}[-1,1]$, $Z \sim \mathrm{Unif}[0,1]$, and $A \sim \mathrm{Bernoulli}(.5)$ independently. Noise is $N(0, .35^2)$. Set $\Gamma = Q\,\mathrm{diag}(.55, .30, .12, .03)Q^\mathsf{T}$. Starting from I_4, right-multiply planar rotations for pairs $(1,2),(1,3),(1,4),(2,3),(2,4),(3,4)$, in that order. For pair (j,k), the angle is $.15 + .1((j+k) \bmod 4)$, and the off-diagonal rotation entries are $R_{jk} = -\sin(\text{angle})$, $R_{kj} = \sin(\text{angle})$.

Random effects are either Gaussian or $\gamma = \sqrt{3/V}\,\Gamma^{1/2} Z_0$ with $V \sim \chi^2_5$ and $Z_0 \sim N(0, I_4)$ independent. The latter retains covariance Γ; noise remains Gaussian. Each training person has eight visits. There are $n = 240$ or 960 people *in each fitting group*, for total training sizes 480 and 1920. The smaller training samples are prefixes of the larger samples. Covariance eigenvalues are clipped to $[10^{-5}, 5]$, and noise variance to $[.001, 2]$.

There are 384 independent replications per distribution, four future people per replication, and personal histories $H = 0, 2, 6$ nested within a six-visit history. The prediction grid combines five x values $(-1, -.5, 0, .5, 1)$ and 21 equally spaced z values, yielding 105 points. Coverage means that the entire latent curve is covered on this grid. Monte Carlo standard errors (MCSEs) use the 384 independent replication averages, accounting for the four people within each replication. Widths are averaged over people and grid points. RMSE is the square root of the aggregated squared prediction error. Oracle, fitted-240, and fitted-960 regimes are all retained.

Table 11: Original model, truncation-adjusted grid band. Parentheses give MCSEs. The nominal coverage is .95; n is the size of each fitting group.

Law	Regime	H	Coverage (MCSE)	Width (MCSE)	RMSE
Gaussian	Oracle	0	0.9518 (0.0052)	4.0281 (0.0036)	0.8519
Gaussian	Oracle	2	0.9609 (0.0048)	3.1730 (0.0141)	0.6752
Gaussian	Oracle	6	0.9596 (0.0049)	2.2173 (0.0117)	0.4720
Gaussian	$n = 240$	0	0.9303 (0.0066)	4.0338 (0.0148)	0.8571
Gaussian	$n = 240$	2	0.9180 (0.0067)	3.1397 (0.0175)	0.6896
Gaussian	$n = 240$	6	0.8971 (0.0081)	2.1592 (0.0131)	0.4911
Gaussian	$n = 960$	0	0.9414 (0.0061)	4.0311 (0.0080)	0.8524
Gaussian	$n = 960$	2	0.9466 (0.0056)	3.1655 (0.0154)	0.6797
Gaussian	$n = 960$	6	0.9395 (0.0059)	2.1982 (0.0120)	0.4792
Elliptical t_5	Oracle	0	0.9375 (0.0062)	4.0280 (0.0035)	0.8877
Elliptical t_5	Oracle	2	0.9473 (0.0057)	3.1630 (0.0142)	0.6969
Elliptical t_5	Oracle	6	0.9596 (0.0050)	2.1906 (0.0119)	0.4472
Elliptical t_5	$n = 240$	0	0.9089 (0.0072)	4.0002 (0.0266)	0.8890
Elliptical t_5	$n = 240$	2	0.8913 (0.0080)	3.0675 (0.0222)	0.7150
Elliptical t_5	$n = 240$	6	0.8529 (0.0110)	2.0398 (0.0165)	0.4904
Elliptical t_5	$n = 960$	0	0.9238 (0.0066)	4.0137 (0.0123)	0.8878
Elliptical t_5	$n = 960$	2	0.9323 (0.0064)	3.1450 (0.0162)	0.7010
Elliptical t_5	$n = 960$	6	0.9453 (0.0058)	2.1588 (0.0129)	0.4534

Primary result: truncation-adjusted Gaussian grid band Table 11 reports all 18 primary cells. For Gaussian effects with $n = 960$, coverage is .9414, .9466, and .9395 for $H = 0, 2, 6$, while width decreases from 4.0311 to 3.1655 to 2.1982. The corresponding RMSEs decrease from .8524 to .6797 to .4792. Thus personal histories improve precision, while the fitted procedure still shows some undercoverage.

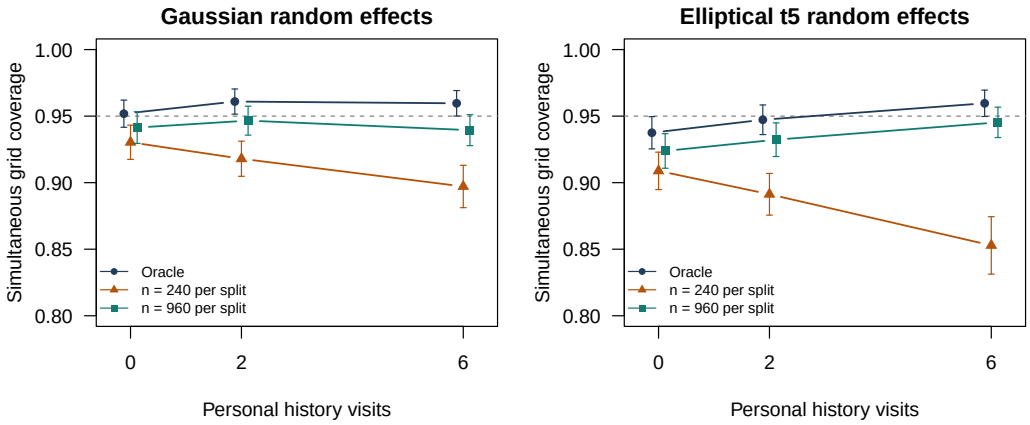


Figure 2: Primary band coverage on the prediction grid. Bars show ± 1.96 MCSE; the dashed line is .95. These are simulation uncertainty bars, not additional prediction bands.

At $n = 240$ and $H = 2$, Gaussian coverage is .9180 (MCSE .0067). The elliptical-t_5 coverage in that cell is .8913 (MCSE .0080), increasing to .9323 (MCSE .0064) at $n = 960$. These cells are retained as evidence of finite-sample limitations and sensitivity to a non-Gaussian score law.

Table 12: All comparison methods for Gaussian effects and $n = 960$. Parentheses give MCSEs.

H	Method	Coverage (MCSE)	Width (MCSE)
2	Full	0.9408 (0.0059)	3.0947 (0.0144)
2	Truncation-adjusted	0.9466 (0.0056)	3.1655 (0.0154)
2	Naive truncation	0.8861 (0.0077)	2.9067 (0.0145)
2	χ^2 projection	0.9909 (0.0024)	3.9939 (0.0185)
2	Pointwise normal	0.8522 (0.0087)	2.5413 (0.0118)
2	Chebyshev	1.0000 (0.0000)	11.5974 (0.0538)
6	Full	0.9212 (0.0069)	2.0533 (0.0113)
6	Truncation-adjusted	0.9395 (0.0059)	2.1982 (0.0120)
6	Naive truncation	0.7760 (0.0104)	1.8372 (0.0114)
6	χ^2 projection	0.9870 (0.0028)	2.6321 (0.0144)
6	Pointwise normal	0.8079 (0.0102)	1.6749 (0.0092)
6	Chebyshev	1.0000 (0.0000)	7.6432 (0.0418)

Why simultaneous calibration and truncation treatment matter At $H = 2$, naive truncation covers .8861, compared with .9466 for the adjusted band. The paired coverage increase is .0605 (MCSE .0059). At $H = 6$, naive coverage is .7760, compared with .9395 after adjustment; the paired increase is .1634 (MCSE .0092). This failure is not solely an estimation effect: at $H = 6$ the oracle naive band covers .8190, whereas the oracle adjusted band covers .9596.

Pointwise normal intervals have simultaneous coverage only .8522 at $n = 960, H = 2$. The χ^2 projection band covers .9909 with width 3.9939; the Chebyshev band covers 1.0000 with width 11.5974. The latter methods illustrate the cost of broader bounds, rather than evidence that the primary Gaussian fitted band is exact. The full method list and every one of the 108 design–method cells appear in Appendix E. In the Gaussian experiments, mean relative covariance error decreases from .1789 at $n = 240$ to .0910 at $n = 960$. The method nevertheless needs further estimation-uncertainty analysis before a finite-sample fitted coverage claim is appropriate.

E Complete original-model results

Full, adjusted, and naive denote the Gaussian grid bands with full scores, truncation adjustment, and unadjusted truncation, respectively. The electronic table also supplies width MCSEs and all remaining recorded metrics.

Table 13: All 108 original-model cells; coverage is simultaneous on the 105-point grid.

Law	Regime	H	Method	Coverage (MCSE)	Width	RMSE
Gaussian	Oracle	0	Full	0.9518 (0.0052)	4.0281	0.8519
Gaussian	Oracle	0	Truncation-adjusted	0.9518 (0.0052)	4.0281	0.8519
Gaussian	Oracle	0	Naive truncation	0.9349 (0.0064)	3.8839	0.8519
Gaussian	Oracle	0	χ^2 projection	0.9948 (0.0018)	5.2512	0.8519
Gaussian	Oracle	0	Pointwise normal	0.8796 (0.0083)	3.3414	0.8519
Gaussian	Oracle	0	Chebyshev	1.0000 (0.0000)	15.2483	0.8519
Gaussian	Oracle	2	Full	0.9564 (0.0051)	3.0997	0.6737
Gaussian	Oracle	2	Truncation-adjusted	0.9609 (0.0048)	3.1730	0.6752
Gaussian	Oracle	2	Naive truncation	0.9154 (0.0067)	2.9081	0.6752
Gaussian	Oracle	2	χ^2 projection	0.9928 (0.0021)	4.0034	0.6737
Gaussian	Oracle	2	Pointwise normal	0.8815 (0.0080)	2.5474	0.6737
Gaussian	Oracle	2	Chebyshev	1.0000 (0.0000)	11.6251	0.6737
Gaussian	Oracle	6	Full	0.9466 (0.0059)	2.0689	0.4594
Gaussian	Oracle	6	Truncation-adjusted	0.9596 (0.0049)	2.2173	0.4720
Gaussian	Oracle	6	Naive truncation	0.8190 (0.0097)	1.8480	0.4720
Gaussian	Oracle	6	χ^2 projection	0.9896 (0.0026)	2.6474	0.4594
Gaussian	Oracle	6	Pointwise normal	0.8600 (0.0090)	1.6845	0.4594
Gaussian	Oracle	6	Chebyshev	1.0000 (0.0000)	7.6873	0.4594
Gaussian	$n = 240$	0	Full	0.9303 (0.0066)	4.0338	0.8571
Gaussian	$n = 240$	0	Truncation-adjusted	0.9303 (0.0066)	4.0338	0.8571
Gaussian	$n = 240$	0	Naive truncation	0.9076 (0.0074)	3.8993	0.8571
Gaussian	$n = 240$	0	χ^2 projection	0.9915 (0.0023)	5.2620	0.8571
Gaussian	$n = 240$	0	Pointwise normal	0.8424 (0.0093)	3.3482	0.8571
Gaussian	$n = 240$	0	Chebyshev	1.0000 (0.0000)	15.2797	0.8571
Gaussian	$n = 240$	2	Full	0.9121 (0.0069)	3.0724	0.6884
Gaussian	$n = 240$	2	Truncation-adjusted	0.9180 (0.0067)	3.1397	0.6896
Gaussian	$n = 240$	2	Naive truncation	0.8509 (0.0093)	2.8937	0.6896
Gaussian	$n = 240$	2	χ^2 projection	0.9818 (0.0033)	3.9688	0.6884
Gaussian	$n = 240$	2	Pointwise normal	0.8112 (0.0097)	2.5254	0.6884
Gaussian	$n = 240$	2	Chebyshev	1.0000 (0.0000)	11.5246	0.6884
Gaussian	$n = 240$	6	Full	0.8724 (0.0090)	2.0195	0.4791
Gaussian	$n = 240$	6	Truncation-adjusted	0.8971 (0.0081)	2.1592	0.4911
Gaussian	$n = 240$	6	Naive truncation	0.7240 (0.0121)	1.8184	0.4911
Gaussian	$n = 240$	6	χ^2 projection	0.9688 (0.0044)	2.5899	0.4791
Gaussian	$n = 240$	6	Pointwise normal	0.7415 (0.0117)	1.6480	0.4791
Gaussian	$n = 240$	6	Chebyshev	1.0000 (0.0000)	7.5206	0.4791
Gaussian	$n = 960$	0	Full	0.9414 (0.0061)	4.0311	0.8524
Gaussian	$n = 960$	0	Truncation-adjusted	0.9414 (0.0061)	4.0311	0.8524
Gaussian	$n = 960$	0	Naive truncation	0.9232 (0.0068)	3.8920	0.8524
Gaussian	$n = 960$	0	χ^2 projection	0.9922 (0.0022)	5.2544	0.8524
Gaussian	$n = 960$	0	Pointwise normal	0.8535 (0.0088)	3.3434	0.8524
Gaussian	$n = 960$	0	Chebyshev	1.0000 (0.0000)	15.2575	0.8524
Gaussian	$n = 960$	2	Full	0.9408 (0.0059)	3.0947	0.6778
Gaussian	$n = 960$	2	Truncation-adjusted	0.9466 (0.0056)	3.1655	0.6797
Gaussian	$n = 960$	2	Naive truncation	0.8861 (0.0077)	2.9067	0.6797
Gaussian	$n = 960$	2	χ^2 projection	0.9909 (0.0024)	3.9939	0.6778
Gaussian	$n = 960$	2	Pointwise normal	0.8522 (0.0087)	2.5413	0.6778
Gaussian	$n = 960$	2	Chebyshev	1.0000 (0.0000)	11.5974	0.6778
Gaussian	$n = 960$	6	Full	0.9212 (0.0069)	2.0533	0.4660
Gaussian	$n = 960$	6	Truncation-adjusted	0.9395 (0.0059)	2.1982	0.4792
Gaussian	$n = 960$	6	Naive truncation	0.7760 (0.0104)	1.8372	0.4792
Gaussian	$n = 960$	6	χ^2 projection	0.9870 (0.0028)	2.6321	0.4660
Gaussian	$n = 960$	6	Pointwise normal	0.8079 (0.0102)	1.6749	0.4660
Gaussian	$n = 960$	6	Chebyshev	1.0000 (0.0000)	7.6432	0.4660
Elliptical t_5	Oracle	0	Full	0.9375 (0.0062)	4.0280	0.8877
Elliptical t_5	Oracle	0	Truncation-adjusted	0.9375 (0.0062)	4.0280	0.8877
Elliptical t_5	Oracle	0	Naive truncation	0.9277 (0.0064)	3.8850	0.8877
Elliptical t_5	Oracle	0	χ^2 projection	0.9818 (0.0033)	5.2512	0.8877
Elliptical t_5	Oracle	0	Pointwise normal	0.8815 (0.0081)	3.3414	0.8877
Elliptical t_5	Oracle	0	Chebyshev	0.9987 (0.0009)	15.2483	0.8877
Elliptical t_5	Oracle	2	Full	0.9440 (0.0059)	3.0933	0.6943
Elliptical t_5	Oracle	2	Truncation-adjusted	0.9473 (0.0057)	3.1630	0.6969
Elliptical t_5	Oracle	2	Naive truncation	0.9023 (0.0071)	2.9049	0.6969
Elliptical t_5	Oracle	2	χ^2 projection	0.9798 (0.0036)	3.9966	0.6943
Elliptical t_5	Oracle	2	Pointwise normal	0.8789 (0.0085)	2.5431	0.6943
Elliptical t_5	Oracle	2	Chebyshev	0.9993 (0.0007)	11.6052	0.6943
Elliptical t_5	Oracle	6	Full	0.9492 (0.0055)	2.0446	0.4374
Elliptical t_5	Oracle	6	Truncation-adjusted	0.9596 (0.0050)	2.1906	0.4472
Elliptical t_5	Oracle	6	Naive truncation	0.8574 (0.0088)	1.8271	0.4472
Elliptical t_5	Oracle	6	χ^2 projection	0.9870 (0.0028)	2.6186	0.4374
Elliptical t_5	Oracle	6	Pointwise normal	0.8848 (0.0081)	1.6662	0.4374
Elliptical t_5	Oracle	6	Chebyshev	1.0000 (0.0000)	7.6037	0.4374
Elliptical t_5	$n = 240$	0	Full	0.9089 (0.0072)	4.0002	0.8890
Elliptical t_5	$n = 240$	0	Truncation-adjusted	0.9089 (0.0072)	4.0002	0.8890
Elliptical t_5	$n = 240$	0	Naive truncation	0.8874 (0.0080)	3.8864	0.8890

Law	Regime	H	Method	Coverage (MCSE)	Width	RMSE
Elliptical t_5	$n = 240$	0	χ^2 projection	0.9655 (0.0045)	5.2375	0.8890
Elliptical t_5	$n = 240$	0	Pointwise normal	0.8385 (0.0097)	3.3327	0.8890
Elliptical t_5	$n = 240$	0	Chebyshev	0.9987 (0.0009)	15.2085	0.8890
Elliptical t_5	$n = 240$	2	Full	0.8828 (0.0083)	3.0093	0.7143
Elliptical t_5	$n = 240$	2	Truncation-adjusted	0.8913 (0.0080)	3.0675	0.7150
Elliptical t_5	$n = 240$	2	Naive truncation	0.8275 (0.0104)	2.8519	0.7150
Elliptical t_5	$n = 240$	2	χ^2 projection	0.9486 (0.0058)	3.8975	0.7143
Elliptical t_5	$n = 240$	2	Pointwise normal	0.7956 (0.0104)	2.4800	0.7143
Elliptical t_5	$n = 240$	2	Chebyshev	0.9987 (0.0009)	11.3174	0.7143
Elliptical t_5	$n = 240$	6	Full	0.8268 (0.0120)	1.9109	0.4914
Elliptical t_5	$n = 240$	6	Truncation-adjusted	0.8529 (0.0110)	2.0398	0.4904
Elliptical t_5	$n = 240$	6	Naive truncation	0.7305 (0.0135)	1.7445	0.4904
Elliptical t_5	$n = 240$	6	χ^2 projection	0.9173 (0.0097)	2.4593	0.4914
Elliptical t_5	$n = 240$	6	Pointwise normal	0.7103 (0.0137)	1.5649	0.4914
Elliptical t_5	$n = 240$	6	Chebyshev	0.9837 (0.0056)	7.1412	0.4914
Elliptical t_5	$n = 960$	0	Full	0.9238 (0.0066)	4.0137	0.8878
Elliptical t_5	$n = 960$	0	Truncation-adjusted	0.9238 (0.0066)	4.0137	0.8878
Elliptical t_5	$n = 960$	0	Naive truncation	0.9128 (0.0072)	3.8754	0.8878
Elliptical t_5	$n = 960$	0	χ^2 projection	0.9772 (0.0038)	5.2411	0.8878
Elliptical t_5	$n = 960$	0	Pointwise normal	0.8678 (0.0085)	3.3350	0.8878
Elliptical t_5	$n = 960$	0	Chebyshev	0.9987 (0.0009)	15.2191	0.8878
Elliptical t_5	$n = 960$	2	Full	0.9284 (0.0065)	3.0772	0.6993
Elliptical t_5	$n = 960$	2	Truncation-adjusted	0.9323 (0.0064)	3.1450	0.7010
Elliptical t_5	$n = 960$	2	Naive truncation	0.8750 (0.0082)	2.8941	0.7010
Elliptical t_5	$n = 960$	2	χ^2 projection	0.9792 (0.0035)	3.9744	0.6993
Elliptical t_5	$n = 960$	2	Pointwise normal	0.8542 (0.0092)	2.5289	0.6993
Elliptical t_5	$n = 960$	2	Chebyshev	0.9993 (0.0007)	11.5408	0.6993
Elliptical t_5	$n = 960$	6	Full	0.9251 (0.0067)	2.0190	0.4453
Elliptical t_5	$n = 960$	6	Truncation-adjusted	0.9453 (0.0058)	2.1588	0.4534
Elliptical t_5	$n = 960$	6	Naive truncation	0.8151 (0.0103)	1.8137	0.4534
Elliptical t_5	$n = 960$	6	χ^2 projection	0.9831 (0.0037)	2.5897	0.4453
Elliptical t_5	$n = 960$	6	Pointwise normal	0.8288 (0.0098)	1.6479	0.4453
Elliptical t_5	$n = 960$	6	Chebyshev	1.0000 (0.0000)	7.5201	0.4453

F End-to-end conformal experiment

This experiment uses the original shared-covariance data-generating model in Appendix D, with the two-coordinate random basis per arm. The population coefficients, rotated joint covariance with eigenvalues $(.55, .30, .12, .03)$, exogenous uniform covariates, Bernoulli treatment assignments, and innovation standard deviation .35 are unchanged. The two latent laws are Gaussian and standardized elliptical t_5 with the same covariance. The latter uses one common chi-square scale for all four coefficients of a subject; innovations remain Gaussian.

F.1 Samples and fitted FPCA predictor

Independent mean-fitting and covariance-fitting groups each contain $n = 240$ or 960 people, with eight visits per person. The mean uses pooled regression; the covariance and noise use pooled residual second-moment equations on the other group. Eigenvalues are clipped to $[10^{-5}, 5]$ and the noise variance to $[.001, 2]$. The eigensystem is learned from this covariance. The retained rank is the smallest explaining at least 95% of its variance. The predictor in (4) uses histories of $h = 0, 2, 6$ visits and the full score covariance, with scale floor .03. The two sample sizes share prefixes within a replication. Neither calibration nor future people enter either fitting group.

There are $m = 199$ or 399 calibration people and four future people. Each person's persistent coefficients remain fixed across all visits. Histories are prefixes of visits $1, \ldots, 6$; the disjoint recovery block starts at visit seven and contains $r = 32$ or 128 visits. Both calibration sizes and recovery lengths use paired prefixes. Unused observations are not selected by their outcomes. All future people are excluded from the calibration cutoffs, including when $m = 199$ and the larger calibration pool is available. Future recovery blocks are generated only to audit the augmented proof events; they do not enter the reported prediction.

F.2 Recovery, methods and comparisons

The original recovery set is (57), with the external bound $\bar{\sigma} = .5$ and failure budget $\eta_{\mathrm{rec}} = .01$. The paired postprocessor also uses the residual-studentized exact-F set (58). Under the same complete

exogenous design and conditionally spherical Gaussian recovery model, it replaces the external noise upper bound by a within-person residual scale and requires $R > 8$; neither recovery assumption set contains the other. Proposition 8 states its exact coverage result. The feasible score is computed over all 105 grid points. The rank error is $\gamma_{\mathrm{rk}} = .04$; the required order is $k = \lceil (m+1)(1-.04) \rceil$, namely 192 or 384. A singular recovery design receives the full coefficient space and an infinite score. No such subject is omitted.

The five methods use identical fitting and history data:

1. *Full*: the full FPCA center with half-width $q^{+}\widehat{s}$.
2. *Guard*: the primary truncated center with half-width $|\widehat{\delta}_M| + q^{+}\widehat{s}$.
3. *No guard*: the truncated center with half-width $q^{+}\widehat{s}$, omitting the mean correction. No general guarantee is assigned to this diagnostic.
4. *Unprotected*: a guarded truncated band calibrated using recovery point estimates without their confidence radii, at rank error .05. Its latent-curve coverage is not guaranteed by the stated recovery argument.
5. *Oracle*: the same guarded truncated predictor calibrated at rank error .05 using the actual latent calibration curves. These labels are unavailable in practice. They are confined to this explicitly infeasible benchmark and evaluation.

Table 15 shows that the protected methods are much wider than the latent-label benchmark. The unprotected recovery method is also conservative in these cells, but its observed coverage does not supply the missing recovery-set guarantee. The full and no-guard bands have identical widths; their centers differ. Figure 3 displays the history comparisons. Table 16 reports every prespecified fitting-size, calibration-size and recovery-length combination at $h = 2$ for the primary method. No additional setting was selected after observing coverage.

Archived same-realization exact-F comparison. This is a deterministic, paired postprocessing analysis of all 768 final raw archives from job 4867181, rather than a fresh independent confirmatory sample. It reuses the same FPCA fits, observations, 384 seeds per latent law, and four future people per replication. The postprocessor covers $h \in \{0, 2, 6\}$; the primary tables report $n = 960$, $m = 399$, $h = 2$, and $R \in \{32, 128\}$. It replaces only the recovery radius by $s_{\varepsilon,i}\{4F_{4,R-8,.99}\}^{1/2}$ and keeps $\gamma_{\mathrm{rk}} = .04$. Job 4875745 completed on 48 CPUs and produced all 92,160 expected rows. For every seed, it reconstructed the archived bounded scores, cutoffs, Full bands, and Guard bands before forming the paired comparison. MCSEs use the 384 independent replication averages, after first averaging the four future people within a replication.

F.3 Replication, uncertainty and verification

A disjoint development stage has 12 independent replications per latent law. The final stage has 384 per law, 768 in total, each with four future people. Every replication evaluates two fitting sizes, three history lengths, two recovery lengths, two calibration sizes and five methods: 480 future–method records. The final output contains 368,640 such records, 92,160 replication–method means, 1,536 fitting diagnostic records, and 240 summary cells. No unbounded future band occurred in this run.

For coverage, width and the other averages, the Monte Carlo standard error is the sample standard deviation of the 384 independent replication means divided by $\sqrt{384}$. This accounts for dependence among the four future predictions sharing a fitting and calibration sample. Paired comparisons likewise use replication-level differences. The RMSE is the square root of the mean squared error over future curves and grid points. The grid has five equally weighted x values $(-1, -.5, 0, .5, 1)$ and 21 equally spaced z values on $[0, 1]$. These are finite-grid coverage measurements.

The preflight checks the complete effect dictionary, no-history prediction, learned-FPCA and coordinate-form equivalence, full covariance, truncation containment, noiseless recovery, singular-design handling, corrected ranks, ties, seed separation and absence of latent-truth or future-recovery leakage into feasible predictions. For every final seed, a separate audit reconstructs the primary $n = 960, h = 2, r = 128, m = 399$ path from raw data using coordinate prediction and QR recovery. All cells receive record-count, key, seed and numerical-value checks. All failures are retained; coverage itself is not a stopping or tuning criterion.

Table 14: Paired bounded-noise and exact-F recovery results at $n = 960$, $m = 399$, and $h = 2$. The final column is exact-F minus bounded mean width; parentheses give its replication-level MCSE. The latent-oracle Guard widths are 3.337 for Gaussian effects and 3.547 for elliptical-t_5 effects, with coverage $.9499$ and $.9544$, respectively.

Law	R	Center	Bounded $\bar{\sigma}$		Exact F		Δ width (MCSE)
			Cov.	Width	Cov.	Width	
G	32	Full	1.0000	18.364	1.0000	15.389	-2.975 (.033)
G	32	Guard	1.0000	18.440	1.0000	15.465	-2.975 (.033)
G	128	Full	1.0000	6.856	0.9993	5.880	-0.976 (.007)
G	128	Guard	1.0000	6.932	0.9993	5.956	-0.976 (.007)
t_5	32	Full	1.0000	18.633	1.0000	15.692	-2.941 (.053)
t_5	32	Guard	1.0000	18.702	1.0000	15.761	-2.941 (.053)
t_5	128	Full	0.9993	7.109	0.9980	6.143	-0.966 (.017)
t_5	128	Guard	0.9993	7.178	0.9980	6.212	-0.966 (.017)

Table 15: End-to-end comparison at $n = 960$, $m = 399$, $h = 2$, and $r \in \{32, 128\}$. Coverage and width are followed by their replication-level MCSEs. The latent-label oracle is infeasible.

Effect law	r	Method	Coverage (MCSE)	Width (MCSE)
Gaussian	32	Full conformal	1.0000 (0.0000)	18.364 (0.103)
Gaussian	32	Guarded conformal	1.0000 (0.0000)	18.440 (0.104)
Gaussian	32	No truncation guard	1.0000 (0.0000)	18.364 (0.103)
Gaussian	32	Unprotected recovery	0.9993 (0.0007)	4.993 (0.025)
Gaussian	32	Latent-label oracle	0.9499 (0.0056)	3.337 (0.017)
Gaussian	128	Full conformal	1.0000 (0.0000)	6.856 (0.033)
Gaussian	128	Guarded conformal	1.0000 (0.0000)	6.932 (0.034)
Gaussian	128	No truncation guard	1.0000 (0.0000)	6.856 (0.033)
Gaussian	128	Unprotected recovery	0.9753 (0.0038)	3.704 (0.018)
Gaussian	128	Latent-label oracle	0.9499 (0.0056)	3.337 (0.017)
Elliptical t_5	32	Full conformal	1.0000 (0.0000)	18.633 (0.225)
Elliptical t_5	32	Guarded conformal	1.0000 (0.0000)	18.702 (0.226)
Elliptical t_5	32	No truncation guard	1.0000 (0.0000)	18.633 (0.225)
Elliptical t_5	32	Unprotected recovery	0.9941 (0.0019)	5.213 (0.044)
Elliptical t_5	32	Latent-label oracle	0.9544 (0.0053)	3.547 (0.031)
Elliptical t_5	128	Full conformal	0.9993 (0.0007)	7.109 (0.090)
Elliptical t_5	128	Guarded conformal	0.9993 (0.0007)	7.178 (0.091)
Elliptical t_5	128	No truncation guard	0.9993 (0.0007)	7.109 (0.090)
Elliptical t_5	128	Unprotected recovery	0.9720 (0.0043)	3.903 (0.032)
Elliptical t_5	128	Latent-label oracle	0.9544 (0.0053)	3.547 (0.031)

Table 16: Guarded conformal bands at $h = 2$ across all prespecified fitting sizes, calibration sizes and recovery lengths. n counts people in each of the two fitting groups.

Effect law	n	m	r	Coverage (MCSE)	Width (MCSE)
Gaussian	240	199	32	1.0000 (0.0000)	18.800 (0.141)
Gaussian	240	199	128	1.0000 (0.0000)	7.157 (0.044)
Gaussian	240	399	32	1.0000 (0.0000)	18.613 (0.122)
Gaussian	240	399	128	1.0000 (0.0000)	7.141 (0.043)
Gaussian	960	199	32	1.0000 (0.0000)	18.572 (0.124)
Gaussian	960	199	128	1.0000 (0.0000)	6.959 (0.036)
Gaussian	960	399	32	1.0000 (0.0000)	18.440 (0.104)
Gaussian	960	399	128	1.0000 (0.0000)	6.932 (0.034)
Elliptical t_5	240	199	32	1.0000 (0.0000)	19.711 (0.338)
Elliptical t_5	240	199	128	0.9980 (0.0011)	7.699 (0.140)
Elliptical t_5	240	399	32	1.0000 (0.0000)	19.481 (0.318)
Elliptical t_5	240	399	128	0.9980 (0.0011)	7.658 (0.138)
Elliptical t_5	960	199	32	1.0000 (0.0000)	18.895 (0.234)
Elliptical t_5	960	199	128	0.9993 (0.0007)	7.209 (0.090)
Elliptical t_5	960	399	32	1.0000 (0.0000)	18.702 (0.226)
Elliptical t_5	960	399	128	0.9993 (0.0007)	7.178 (0.091)

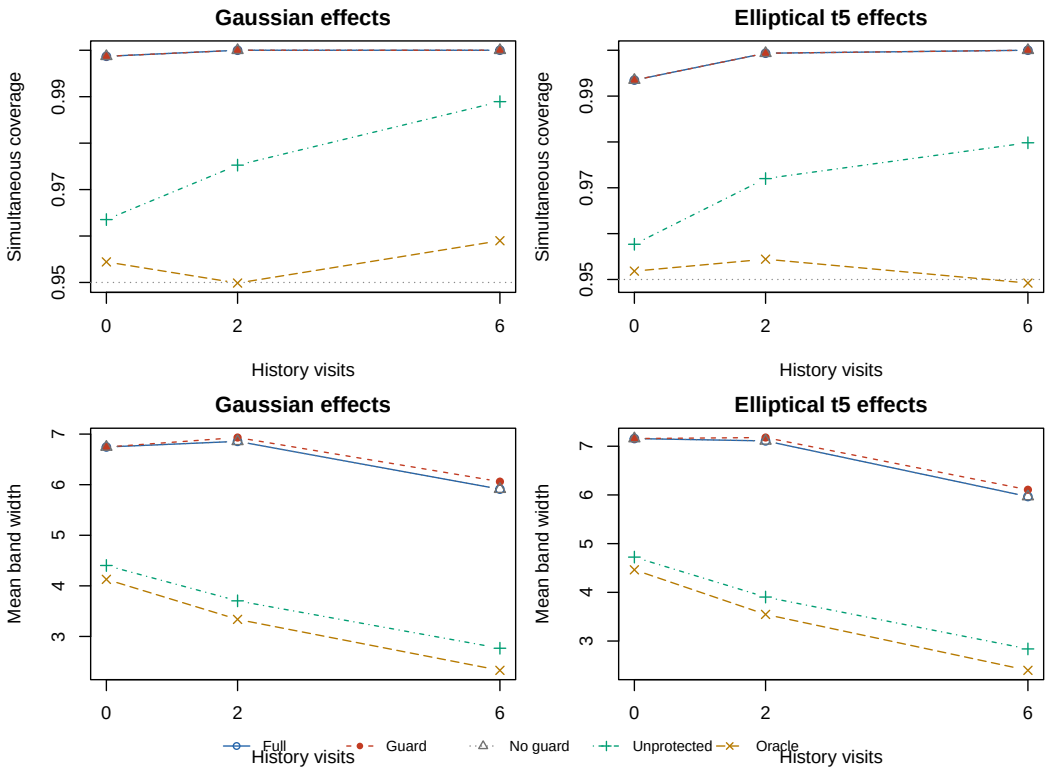


Figure 3: End-to-end simultaneous grid coverage and mean width across history lengths, at $n = 960$ per fitting group, $m = 399$ and $r = 128$. The horizontal reference is .95. Protected calibration gives high coverage with substantial width inflation.

The development and final seed maps are $971515000 + 100000\ell + j$ and $371515000 + 100000\ell + j$, respectively, where $\ell = 1, 2$ indexes the Gaussian and elliptical-t_5 laws and j indexes replications. The source package provides generation, aggregation, and audit scripts with study-specific run instructions. The exact-F analysis uses the replication-level outputs generated by the end-to-end module. Full historical simulation archives are not included.

G COMPLETE END-TO-END CONFORMAL RESULTS

All 240 prespecified cells are reported below. G denotes Gaussian random effects and T denotes elliptical t_5 random effects. n is the size of each fitting group; h is history length, r recovery length, and m calibration size. Full and Guard have the stated recovery-set guarantee; No guard and Unprotected are diagnostics; Oracle uses unavailable latent calibration labels. Coverage is simultaneous on the 105-point grid. Parentheses give coverage MCSEs from independent replication averages. Width is mean total band width. The full machine-readable results include MCSEs for width and the other reported averages.

Table 17: All end-to-end experiment cells, with 384 independent replications and four future people per replication in each latent law.

Law	n	h	r	m	Method	Coverage (MCSE)	Width	RMSE
G	240	0	32	199	Full	1.0000 (0.0000)	14.996	0.864
G	240	0	32	199	Guard	1.0000 (0.0000)	14.996	0.864
G	240	0	32	199	No guard	1.0000 (0.0000)	14.996	0.864
G	240	0	32	199	Unprotected	0.9850 (0.0032)	5.318	0.864
G	240	0	32	199	Oracle	0.9538 (0.0054)	4.293	0.864
G	240	0	32	399	Full	1.0000 (0.0000)	14.823	0.864
G	240	0	32	399	Guard	1.0000 (0.0000)	14.823	0.864
G	240	0	32	399	No guard	1.0000 (0.0000)	14.823	0.864
G	240	0	32	399	Unprotected	0.9883 (0.0027)	5.298	0.864
G	240	0	32	399	Oracle	0.9525 (0.0053)	4.277	0.864
G	240	0	128	199	Full	0.9987 (0.0009)	6.918	0.864
G	240	0	128	199	Guard	0.9987 (0.0009)	6.918	0.864
G	240	0	128	199	No guard	0.9987 (0.0009)	6.918	0.864
G	240	0	128	199	Unprotected	0.9583 (0.0051)	4.521	0.864
G	240	0	128	199	Oracle	0.9538 (0.0054)	4.293	0.864
G	240	0	128	399	Full	0.9987 (0.0009)	6.896	0.864
G	240	0	128	399	Guard	0.9987 (0.0009)	6.896	0.864
G	240	0	128	399	No guard	0.9987 (0.0009)	6.896	0.864
G	240	0	128	399	Unprotected	0.9596 (0.0049)	4.507	0.864
G	240	0	128	399	Oracle	0.9525 (0.0053)	4.277	0.864
G	240	2	32	199	Full	1.0000 (0.0000)	18.729	0.690
G	240	2	32	199	Guard	1.0000 (0.0000)	18.800	0.694
G	240	2	32	199	No guard	1.0000 (0.0000)	18.729	0.694
G	240	2	32	199	Unprotected	0.9993 (0.0007)	5.159	0.694
G	240	2	32	199	Oracle	0.9518 (0.0054)	3.545	0.694
G	240	2	32	399	Full	1.0000 (0.0000)	18.542	0.690
G	240	2	32	399	Guard	1.0000 (0.0000)	18.613	0.694
G	240	2	32	399	No guard	1.0000 (0.0000)	18.542	0.694
G	240	2	32	399	Unprotected	0.9987 (0.0009)	5.126	0.694
G	240	2	32	399	Oracle	0.9518 (0.0054)	3.523	0.694
G	240	2	128	199	Full	1.0000 (0.0000)	7.086	0.690
G	240	2	128	199	Guard	1.0000 (0.0000)	7.157	0.694
G	240	2	128	199	No guard	1.0000 (0.0000)	7.086	0.694
G	240	2	128	199	Unprotected	0.9720 (0.0042)	3.865	0.694
G	240	2	128	199	Oracle	0.9518 (0.0054)	3.545	0.694
G	240	2	128	399	Full	1.0000 (0.0000)	7.070	0.690
G	240	2	128	399	Guard	1.0000 (0.0000)	7.141	0.694
G	240	2	128	399	No guard	1.0000 (0.0000)	7.070	0.694
G	240	2	128	399	Unprotected	0.9707 (0.0043)	3.844	0.694
G	240	2	128	399	Oracle	0.9518 (0.0054)	3.523	0.694
G	240	6	32	199	Full	1.0000 (0.0000)	18.499	0.471
G	240	6	32	199	Guard	1.0000 (0.0000)	18.642	0.485
G	240	6	32	199	No guard	1.0000 (0.0000)	18.499	0.485
G	240	6	32	199	Unprotected	1.0000 (0.0000)	4.535	0.485
G	240	6	32	199	Oracle	0.9570 (0.0053)	2.515	0.485
G	240	6	32	399	Full	1.0000 (0.0000)	18.266	0.471
G	240	6	32	399	Guard	1.0000 (0.0000)	18.409	0.485
G	240	6	32	399	No guard	1.0000 (0.0000)	18.266	0.485
G	240	6	32	399	Unprotected	1.0000 (0.0000)	4.511	0.485
G	240	6	32	399	Oracle	0.9622 (0.0050)	2.510	0.485
G	240	6	128	199	Full	1.0000 (0.0000)	6.084	0.471
G	240	6	128	199	Guard	1.0000 (0.0000)	6.226	0.485
G	240	6	128	199	No guard	1.0000 (0.0000)	6.084	0.485
G	240	6	128	199	Unprotected	0.9837 (0.0033)	2.908	0.485
G	240	6	128	199	Oracle	0.9570 (0.0053)	2.515	0.485
G	240	6	128	399	Full	1.0000 (0.0000)	6.061	0.471
G	240	6	128	399	Guard	1.0000 (0.0000)	6.204	0.485
G	240	6	128	399	No guard	1.0000 (0.0000)	6.061	0.485
G	240	6	128	399	Unprotected	0.9857 (0.0030)	2.905	0.485
G	240	6	128	399	Oracle	0.9622 (0.0050)	2.510	0.485
G	960	0	32	199	Full	1.0000 (0.0000)	14.848	0.862
G	960	0	32	199	Guard	1.0000 (0.0000)	14.848	0.862

Law	n	h	r	m	Method	Coverage (MCSE)	Width	RMSE
G	960	0	32	199	No guard	1.0000 (0.0000)	14.848	0.862
G	960	0	32	199	Unprotected	0.9876 (0.0028)	5.216	0.862
G	960	0	32	199	Oracle	0.9538 (0.0055)	4.145	0.862
G	960	0	32	399	Full	1.0000 (0.0000)	14.649	0.862
G	960	0	32	399	Guard	1.0000 (0.0000)	14.649	0.862
G	960	0	32	399	No guard	1.0000 (0.0000)	14.649	0.862
G	960	0	32	399	Unprotected	0.9876 (0.0028)	5.202	0.862
G	960	0	32	399	Oracle	0.9544 (0.0053)	4.129	0.862
G	960	0	128	199	Full	0.9987 (0.0009)	6.759	0.862
G	960	0	128	199	Guard	0.9987 (0.0009)	6.759	0.862
G	960	0	128	199	No guard	0.9987 (0.0009)	6.759	0.862
G	960	0	128	199	Unprotected	0.9648 (0.0047)	4.419	0.862
G	960	0	128	199	Oracle	0.9538 (0.0055)	4.145	0.862
G	960	0	128	399	Full	0.9987 (0.0009)	6.745	0.862
G	960	0	128	399	Guard	0.9987 (0.0009)	6.745	0.862
G	960	0	128	399	No guard	0.9987 (0.0009)	6.745	0.862
G	960	0	128	399	Unprotected	0.9635 (0.0048)	4.403	0.862
G	960	0	128	399	Oracle	0.9544 (0.0053)	4.129	0.862
G	960	2	32	199	Full	1.0000 (0.0000)	18.496	0.683
G	960	2	32	199	Guard	1.0000 (0.0000)	18.572	0.687
G	960	2	32	199	No guard	1.0000 (0.0000)	18.496	0.687
G	960	2	32	199	Unprotected	0.9980 (0.0011)	5.033	0.687
G	960	2	32	199	Oracle	0.9486 (0.0059)	3.346	0.687
G	960	2	32	399	Full	1.0000 (0.0000)	18.364	0.683
G	960	2	32	399	Guard	1.0000 (0.0000)	18.440	0.687
G	960	2	32	399	No guard	1.0000 (0.0000)	18.364	0.687
G	960	2	32	399	Unprotected	0.9993 (0.0007)	4.993	0.687
G	960	2	32	399	Oracle	0.9499 (0.0056)	3.337	0.687
G	960	2	128	199	Full	1.0000 (0.0000)	6.883	0.683
G	960	2	128	199	Guard	1.0000 (0.0000)	6.959	0.687
G	960	2	128	199	No guard	1.0000 (0.0000)	6.883	0.687
G	960	2	128	199	Unprotected	0.9772 (0.0038)	3.729	0.687
G	960	2	128	199	Oracle	0.9486 (0.0059)	3.346	0.687
G	960	2	128	399	Full	1.0000 (0.0000)	6.856	0.683
G	960	2	128	399	Guard	1.0000 (0.0000)	6.932	0.687
G	960	2	128	399	No guard	1.0000 (0.0000)	6.856	0.687
G	960	2	128	399	Unprotected	0.9753 (0.0038)	3.704	0.687
G	960	2	128	399	Oracle	0.9499 (0.0056)	3.337	0.687
G	960	6	32	199	Full	1.0000 (0.0000)	18.399	0.458
G	960	6	32	199	Guard	1.0000 (0.0000)	18.548	0.472
G	960	6	32	199	No guard	1.0000 (0.0000)	18.399	0.472
G	960	6	32	199	Unprotected	1.0000 (0.0000)	4.452	0.472
G	960	6	32	199	Oracle	0.9596 (0.0050)	2.339	0.472
G	960	6	32	399	Full	1.0000 (0.0000)	18.158	0.458
G	960	6	32	399	Guard	1.0000 (0.0000)	18.307	0.472
G	960	6	32	399	No guard	1.0000 (0.0000)	18.158	0.472
G	960	6	32	399	Unprotected	1.0000 (0.0000)	4.423	0.472
G	960	6	32	399	Oracle	0.9590 (0.0052)	2.330	0.472
G	960	6	128	199	Full	1.0000 (0.0000)	5.929	0.458
G	960	6	128	199	Guard	1.0000 (0.0000)	6.078	0.472
G	960	6	128	199	No guard	1.0000 (0.0000)	5.929	0.472
G	960	6	128	199	Unprotected	0.9876 (0.0028)	2.776	0.472
G	960	6	128	199	Oracle	0.9596 (0.0050)	2.339	0.472
G	960	6	128	399	Full	1.0000 (0.0000)	5.911	0.458
G	960	6	128	399	Guard	1.0000 (0.0000)	6.060	0.472
G	960	6	128	399	No guard	1.0000 (0.0000)	5.911	0.472
G	960	6	128	399	Unprotected	0.9889 (0.0026)	2.767	0.472
G	960	6	128	399	Oracle	0.9590 (0.0052)	2.330	0.472
T	240	0	32	199	Full	1.0000 (0.0000)	15.539	0.856
T	240	0	32	199	Guard	1.0000 (0.0000)	15.539	0.856
T	240	0	32	199	No guard	1.0000 (0.0000)	15.539	0.856
T	240	0	32	199	Unprotected	0.9766 (0.0037)	5.728	0.856
T	240	0	32	199	Oracle	0.9512 (0.0055)	4.715	0.856
T	240	0	32	399	Full	1.0000 (0.0000)	15.290	0.856
T	240	0	32	399	Guard	1.0000 (0.0000)	15.290	0.856
T	240	0	32	399	No guard	1.0000 (0.0000)	15.290	0.856
T	240	0	32	399	Unprotected	0.9785 (0.0036)	5.664	0.856
T	240	0	32	399	Oracle	0.9512 (0.0055)	4.648	0.856
T	240	0	128	199	Full	0.9928 (0.0021)	7.463	0.856
T	240	0	128	199	Guard	0.9928 (0.0021)	7.463	0.856
T	240	0	128	199	No guard	0.9928 (0.0021)	7.463	0.856
T	240	0	128	199	Unprotected	0.9596 (0.0049)	4.921	0.856
T	240	0	128	199	Oracle	0.9512 (0.0055)	4.715	0.856
T	240	0	128	399	Full	0.9928 (0.0021)	7.394	0.856
T	240	0	128	399	Guard	0.9928 (0.0021)	7.394	0.856
T	240	0	128	399	No guard	0.9928 (0.0021)	7.394	0.856
T	240	0	128	399	Unprotected	0.9570 (0.0052)	4.863	0.856
T	240	0	128	399	Oracle	0.9512 (0.0055)	4.648	0.856
T	240	2	32	199	Full	1.0000 (0.0000)	19.654	0.690

Law	n	h	r	m	Method	Coverage (MCSE)	Width	RMSE
T	240	2	32	199	Guard	1.0000 (0.0000)	19.711	0.694
T	240	2	32	199	No guard	1.0000 (0.0000)	19.654	0.694
T	240	2	32	199	Unprotected	0.9909 (0.0026)	5.514	0.694
T	240	2	32	199	Oracle	0.9577 (0.0053)	3.901	0.694
T	240	2	32	399	Full	1.0000 (0.0000)	19.424	0.690
T	240	2	32	399	Guard	1.0000 (0.0000)	19.481	0.694
T	240	2	32	399	No guard	1.0000 (0.0000)	19.424	0.694
T	240	2	32	399	Unprotected	0.9909 (0.0026)	5.482	0.694
T	240	2	32	399	Oracle	0.9577 (0.0053)	3.839	0.694
T	240	2	128	199	Full	0.9980 (0.0011)	7.642	0.690
T	240	2	128	199	Guard	0.9980 (0.0011)	7.699	0.694
T	240	2	128	199	No guard	0.9980 (0.0011)	7.642	0.694
T	240	2	128	199	Unprotected	0.9688 (0.0047)	4.216	0.694
T	240	2	128	199	Oracle	0.9577 (0.0053)	3.901	0.694
T	240	2	128	399	Full	0.9980 (0.0011)	7.601	0.690
T	240	2	128	399	Guard	0.9980 (0.0011)	7.658	0.694
T	240	2	128	399	No guard	0.9980 (0.0011)	7.601	0.694
T	240	2	128	399	Unprotected	0.9701 (0.0043)	4.160	0.694
T	240	2	128	399	Oracle	0.9577 (0.0053)	3.839	0.694
T	240	6	32	199	Full	1.0000 (0.0000)	18.658	0.479
T	240	6	32	199	Guard	1.0000 (0.0000)	18.780	0.494
T	240	6	32	199	No guard	1.0000 (0.0000)	18.658	0.494
T	240	6	32	199	Unprotected	0.9987 (0.0009)	4.638	0.494
T	240	6	32	199	Oracle	0.9551 (0.0052)	2.645	0.494
T	240	6	32	399	Full	1.0000 (0.0000)	18.447	0.479
T	240	6	32	399	Guard	1.0000 (0.0000)	18.569	0.494
T	240	6	32	399	No guard	1.0000 (0.0000)	18.447	0.494
T	240	6	32	399	Unprotected	0.9987 (0.0009)	4.594	0.494
T	240	6	32	399	Oracle	0.9525 (0.0053)	2.623	0.494
T	240	6	128	199	Full	1.0000 (0.0000)	6.216	0.479
T	240	6	128	199	Guard	1.0000 (0.0000)	6.338	0.494
T	240	6	128	199	No guard	1.0000 (0.0000)	6.216	0.494
T	240	6	128	199	Unprotected	0.9779 (0.0037)	3.029	0.494
T	240	6	128	199	Oracle	0.9551 (0.0052)	2.645	0.494
T	240	6	128	399	Full	1.0000 (0.0000)	6.190	0.479
T	240	6	128	399	Guard	1.0000 (0.0000)	6.312	0.494
T	240	6	128	399	No guard	1.0000 (0.0000)	6.190	0.494
T	240	6	128	399	Unprotected	0.9766 (0.0038)	3.013	0.494
T	240	6	128	399	Oracle	0.9525 (0.0053)	2.623	0.494
T	960	0	32	199	Full	1.0000 (0.0000)	15.190	0.854
T	960	0	32	199	Guard	1.0000 (0.0000)	15.190	0.854
T	960	0	32	199	No guard	1.0000 (0.0000)	15.190	0.854
T	960	0	32	199	Unprotected	0.9785 (0.0036)	5.555	0.854
T	960	0	32	199	Oracle	0.9538 (0.0051)	4.537	0.854
T	960	0	32	399	Full	1.0000 (0.0000)	14.952	0.854
T	960	0	32	399	Guard	1.0000 (0.0000)	14.952	0.854
T	960	0	32	399	No guard	1.0000 (0.0000)	14.952	0.854
T	960	0	32	399	Unprotected	0.9779 (0.0036)	5.512	0.854
T	960	0	32	399	Oracle	0.9518 (0.0053)	4.463	0.854
T	960	0	128	199	Full	0.9922 (0.0022)	7.238	0.854
T	960	0	128	199	Guard	0.9922 (0.0022)	7.238	0.854
T	960	0	128	199	No guard	0.9922 (0.0022)	7.238	0.854
T	960	0	128	199	Unprotected	0.9635 (0.0045)	4.783	0.854
T	960	0	128	199	Oracle	0.9538 (0.0051)	4.537	0.854
T	960	0	128	399	Full	0.9935 (0.0020)	7.157	0.854
T	960	0	128	399	Guard	0.9935 (0.0020)	7.157	0.854
T	960	0	128	399	No guard	0.9935 (0.0020)	7.157	0.854
T	960	0	128	399	Unprotected	0.9577 (0.0050)	4.724	0.854
T	960	0	128	399	Oracle	0.9518 (0.0053)	4.463	0.854
T	960	2	32	199	Full	1.0000 (0.0000)	18.826	0.675
T	960	2	32	199	Guard	1.0000 (0.0000)	18.895	0.680
T	960	2	32	199	No guard	1.0000 (0.0000)	18.826	0.680
T	960	2	32	199	Unprotected	0.9935 (0.0020)	5.249	0.680
T	960	2	32	199	Oracle	0.9531 (0.0057)	3.593	0.680
T	960	2	32	399	Full	1.0000 (0.0000)	18.633	0.675
T	960	2	32	399	Guard	1.0000 (0.0000)	18.702	0.680
T	960	2	32	399	No guard	1.0000 (0.0000)	18.633	0.680
T	960	2	32	399	Unprotected	0.9941 (0.0019)	5.213	0.680
T	960	2	32	399	Oracle	0.9544 (0.0053)	3.547	0.680
T	960	2	128	199	Full	0.9993 (0.0007)	7.140	0.675
T	960	2	128	199	Guard	0.9993 (0.0007)	7.209	0.680
T	960	2	128	199	No guard	0.9987 (0.0009)	7.140	0.680
T	960	2	128	199	Unprotected	0.9707 (0.0045)	3.949	0.680
T	960	2	128	199	Oracle	0.9531 (0.0057)	3.593	0.680
T	960	2	128	399	Full	0.9993 (0.0007)	7.109	0.675
T	960	2	128	399	Guard	0.9993 (0.0007)	7.178	0.680
T	960	2	128	399	No guard	0.9993 (0.0007)	7.109	0.680
T	960	2	128	399	Unprotected	0.9720 (0.0043)	3.903	0.680
T	960	2	128	399	Oracle	0.9544 (0.0053)	3.547	0.680

Law	n	h	r	m	Method	Coverage (MCSE)	Width	RMSE
T	960	6	32	199	Full	1.0000 (0.0000)	18.493	0.453
T	960	6	32	199	Guard	1.0000 (0.0000)	18.632	0.468
T	960	6	32	199	No guard	1.0000 (0.0000)	18.493	0.468
T	960	6	32	199	Unprotected	1.0000 (0.0000)	4.507	0.468
T	960	6	32	199	Oracle	0.9512 (0.0055)	2.424	0.468
T	960	6	32	399	Full	1.0000 (0.0000)	18.288	0.453
T	960	6	32	399	Guard	1.0000 (0.0000)	18.428	0.468
T	960	6	32	399	No guard	1.0000 (0.0000)	18.288	0.468
T	960	6	32	399	Unprotected	1.0000 (0.0000)	4.472	0.468
T	960	6	32	399	Oracle	0.9492 (0.0056)	2.398	0.468
T	960	6	128	199	Full	1.0000 (0.0000)	5.985	0.453
T	960	6	128	199	Guard	1.0000 (0.0000)	6.124	0.468
T	960	6	128	199	No guard	1.0000 (0.0000)	5.985	0.468
T	960	6	128	199	Unprotected	0.9818 (0.0034)	2.859	0.468
T	960	6	128	199	Oracle	0.9512 (0.0055)	2.424	0.468
T	960	6	128	399	Full	1.0000 (0.0000)	5.967	0.453
T	960	6	128	399	Guard	1.0000 (0.0000)	6.106	0.468
T	960	6	128	399	No guard	1.0000 (0.0000)	5.967	0.468
T	960	6	128	399	Unprotected	0.9798 (0.0035)	2.838	0.468
T	960	6	128	399	Oracle	0.9492 (0.0056)	2.398	0.468

H CD-FPCA IMPLEMENTATION DETAILS

The time basis has dimension seven on fixed support $z \in [0, 1]$, with three type-8 quartile knots computed from the full mean-fitting sample; the covariate basis has dimension four and fixed support $[-1, 1]$. Both use ordinary, not pre-orthogonalized, cubic B-splines. Here $D_z \in \mathbb{R}^{5\times 7}$ and $D_x \in \mathbb{R}^{2\times 4}$ are second-difference matrices. The penalty in (12) is a second-difference coefficient penalty rather than the integrated derivative penalty in Ding et al. (2022), and its quasi-likelihood is not divided by the number of people. Both directions share one λ.

For a fixed fitting sample size n, the summed criterion with penalty λ_{sum} is equivalent in its exact minimizers to the mean criterion with $\lambda_{\text{mean}} = \lambda_{\text{sum}}/n$. Thus a numerical penalty grid is meaningful only with its stated scaling. The validation fit with 768 people and refit with 960 people use the same frozen sum-scale convention. Frequent selection of the largest candidate motivates a wider-grid sensitivity analysis, but does not by itself identify the source of band-width differences or establish under-regularization.

Mean validation uses RSS and covariance validation uses unpenalized quasi-NLL. The covariance implementation runs two prespecified L-BFGS-B passes. The first uses `maxit`=180 and `factr`=10^7; the second starts from the first and uses `maxit`=300, `factr`=10^4, and `pgtol`=10^{-5}. A solution is accepted only when it is finite, objective-nonincreasing, has code zero and gradient infinity norm at most .02, and triggers no Cholesky-failure sentinel.

Each simulated person has six generated history records, but the frozen predictor uses only records 1–2. Records 3–6 are ignored, and records 7–518 form the disjoint recovery block. Calibration and test subjects never enter basis construction, penalty selection, or fitting. Failed calibration labels have infinite scores; a failed fitted predictor gives an infinite band rather than being silently counted as a success. The true-oracle diagnostic changes only the label used in the score.

The weighted FPCA uses both arms and all 105 grid points. Because the original spline columns are not orthonormal, diagonalizing $C^\top C$ would be wrong; the loading Gram matrix $L^\top W L$ in (13) supplies the required metric. The working rank is four. The smooth data-generating law has eight nonzero latent coordinates, and every coverage evaluation uses the full eight-coordinate latent effect on the 105-point grid.

The arm-specific recovery regression is deliberately narrower than the CD working mean: it is additive in time and covariate and has no interaction. This matches the present simulation design but is not an honest recovery set for a general interacting latent surface. Additive recovery avoids this particular interaction misspecification in the present DGP, but does not itself establish finite-sample latent coverage. The future recovery block is generated only to evaluate labels and scores; it is absent from the future center, scale, cutoff, and reported band.

For arm a, let n_{ia} be the person's number of recovery records. The candidate spline dimensions are the distinct legal values $K_z = \max\{4, \lceil c n_{ia}^{1/5} \rceil + 3\}$ for $c \in \{1, 1.5, 2, 3\}$. With $p_{K_z} = K_z + 2$,

GCV is $n_{ia}\mathrm{RSS}/(n_{ia} - p_{K_z})^2$. A candidate is legal only if $n_{ia} > p_{K_z}$, its QR fit is full rank and finite, and the denominator is positive; absence of a legal candidate is a label failure.

Optimization records both passes, candidate failures, final status, initial and final objective values, and gradient norms. The accepted numerical fits support evaluation of this frozen implementation only. They do not prove global optimality, consistency, or a rate for the CD estimator. For every orthogonal $O \in \mathbb{R}^{4\times4}$, $\mathcal{L}_\lambda(CO, \rho) = \mathcal{L}_\lambda(C, \rho)$ because both $CC^\top$ and the trace penalty are rotation invariant. At a full-column-rank exact stationary point this gives a six-dimensional rotation orbit and flat tangent directions; it does not assert six exact zero directions at a rank-deficient or approximate stationary point. The factor C is nonunique even when its kernel and oracle rank are well defined. The implementation imposes no canonical gauge. Gauge fixing could be studied for numerical stability, but the present evidence does not attribute the two-pass optimization or width differences to this symmetry.

I PROOFS FOR THE DIRECT CD BRANCH

Proof of Proposition 1. For coefficients c_j, $\sum_{j,k} c_j c_k K^{\mathrm{CD}}_{a_j a_k}(u_j, u_k) = \| \sum_j c_j \ell_{a_j}(u_j)\|_2^2 \geq 0$. For positive λ_k, λ_l, the discrete inner product of the two functions in (13) is $v_k^\top L^\top W L v_l/(\sqrt{\lambda_k \lambda_l}) = \mathbf{1}\{k = l\}$. If L is replaced by LO, O orthogonal, then $LL^\top$ and all fitted values $L\omega$ are unchanged after the inverse rotation of ω. Repeated eigenvalues permit rotations within their eigenspace, proving the stated qualification. □

Proof of Proposition 2. For this proof, suppress hats and write $\sigma^2 = \sigma^2_{\mathrm{CD}}$. The covariance of (ω_i, r_i^H) is $\begin{pmatrix} I_4 & L_i^{H\top} \\ L_i^H & L_i^H L_i^{H\top} + \sigma^2 I \end{pmatrix}$. Because $\sigma^2 > 0$, $V_H = L_i^H L_i^{H\top} + \sigma^2 I_2$ is positive definite. For A minimizing $E_{\mathsf{W}}(\|\omega_i - A r_i^H\|^2 \mid \mathcal{D}_i)$, the normal equation is $AV_H = L_i^{H\top}$. It therefore gives $m_\omega = L_i^{H\top} V_H^{-1} r_i^H$ and $S_\omega = I_4 - L_i^{H\top} V_H^{-1} L_i^H$. Woodbury yields $S_\omega = (I_4 + L_i^{H\top} L_i^H/\sigma^2)^{-1}$ and the stated precision expression. For a joint conditionally Gaussian vector under the same fixed design and frozen-parameter conditioning, the same block formula is its conditional distribution; second moments alone only establish the linear-prediction statement. □

Proof of Theorem 3. On a successful fit, after conditioning on the independent fitting data, the finite, strictly positive-scale score map is the same measurable function of each subject's history and latent effect. Exchangeability therefore makes the $m + 1$ oracle scores exchangeable. With random tie breaking, the future rank is uniform on $1, \ldots, m + 1$; rank at most k implies that its score is no larger than the kth calibration score. Thus coverage is at least $k/(m + 1) \geq 1 - \gamma_{\mathrm{CD}}$; without random tie breaking, weak comparison is conservative. If $k > m$, or if the fit fails, the band is the whole real line and coverage is one. Mixing the success and failure branches preserves the lower bound. On the success branch the score event is exactly simultaneous inclusion at every point of $\mathcal{G}$. □

Proof of Proposition 3. For every grid point, the reverse triangle inequality gives $\big||\widetilde{\tau} - c| - |\tau - c|\big|/s \leq |\widetilde{\tau} - \tau|/s$. Taking maxima on the two left-hand score arrays changes their maxima by no more than the maximum right-hand side. On E_ϵ, every calibration oracle score obeys $S_i \leq R_i + \epsilon$, so the corresponding order statistics satisfy $q_S \leq q_R + \epsilon$. The oracle rank probability is taken conditional only on $\mathcal{D}_{\mathrm{fit}}$, not on E_ϵ. Deterministically, $\{S_{m+1} \leq q_S\} \cap E_\epsilon \subseteq \{S_{m+1} \leq q_R + \epsilon\}$. A union bound therefore gives probability at least $1 - \gamma_{\mathrm{CD}} - \eta$. The $k > m$ and fit-failure branches use the whole real line. □

The noisy-label example shows why no proof of the oracle theorem can simply replace $S_i^{o,\mathrm{CD}}$ by R_i^{CD}. The empirical GCV results below are therefore evidence for the implemented recovery rule, not a latent-label finite-sample theorem.

J CD-FPCA CONFIRMATION EXPERIMENT

The independent confirmation crosses linear and smooth signals, Gaussian and elliptical-t_5 random coefficients, and Gaussian, Rademacher, and skewed two-point innovations. Each of the 12 cells has 128 replications, independent mean and covariance fitting groups of 960 people, 399 calibration people, and 32 test people. The old rank-eight predictor and new rank-four CD predictor use identical generated data and seeds. The estimand is always the full latent effect, including all eight smooth-law coordinates. MCSEs cluster the 32 test people within replication.

The frozen DGP has $X \sim \mathrm{Unif}[-1, 1]$, $Z \sim \mathrm{Unif}[0, 1]$, and $A \sim \mathrm{Bernoulli}(.5)$ independently. Both arm means are quadratic in X. Each arm uses the first four shifted orthonormal Legendre coordinates in Z. The covariance is X-invariant. Its rotated four-dimensional principal block has eigenvalues $(.55, .30, .12, .03)$; the linear law has rank four, while the smooth law adds variances $.12^2$ and $.08^2$ in the third and fourth coordinates of each arm, respectively, and has rank eight. Gaussian coefficients are standard normal before the covariance square root; the elliptical-t_5 draw is multiplied by $\sqrt{3/\chi^2_5}$ to retain unit covariance. Innovation standard deviation is .35. Rademacher noise is symmetric, and skewed two-point noise has upper-mass probability .2 with values scaled to mean zero and variance $.35^2$. These specifications are frozen in `studies/cd_fpca/model8_core.R` (functions `Pz`, `Mx`, `truth8`, and `drawrec`). Thus this experiment contains no truly X-varying covariance and does not directly test the setting that motivates the CD extension.

Table 18: Direct CD-FPCA with GCV point labels. Coverage is simultaneous on the 105-point grid; parentheses are MCSEs across 128 replication means.

Signal	Coefficient law	Innovation	Coverage	Mean width
Linear	t_5	Gaussian	.9612 (.0029)	4.841 (.057)
Smooth	t_5	Gaussian	.9495 (.0040)	5.554 (.110)
Linear	Gaussian	Gaussian	.9631 (.0033)	4.461 (.045)
Smooth	Gaussian	Gaussian	.9604 (.0028)	4.920 (.063)
Linear	t_5	Rademacher	.9592 (.0038)	4.706 (.061)
Smooth	t_5	Rademacher	.9551 (.0032)	5.396 (.072)
Linear	Gaussian	Rademacher	.9683 (.0028)	4.357 (.045)
Smooth	Gaussian	Rademacher	.9629 (.0032)	4.842 (.042)
Linear	t_5	Skewed	.9609 (.0032)	4.786 (.054)
Smooth	t_5	Skewed	.9585 (.0035)	5.470 (.072)
Linear	Gaussian	Skewed	.9661 (.0028)	4.436 (.040)
Smooth	Gaussian	Skewed	.9563 (.0034)	4.868 (.049)

Across these cells, the paired new-minus-old GCV-label coverage differences range from $-.0093$ to $-.0005$, while paired width differences range from 0.739 to 1.149: the new bands are wider in every cell. Here `old_fit8` is the comparator rerun on the same data with the same GCV labels, direct truncated-center score, and rank-380 cutoff. It is not the earlier protected rank-384 experiment. Pointwise coverage is higher for the new predictor in every cell even though simultaneous coverage is lower, so the miss geometry differs. The archived summaries contain no separate cutoff and scale decomposition, and we do not attribute the width increase to either. More specifically, `old_fit8` is a structure-matched parametric negative control: it uses the DGP-correct quadratic-in-x arm mean class, per-arm four-coordinate z dictionary, and x-invariant pooled covariance form, while estimating all parameters from training data. The present DGP has no truly x-varying covariance, so this confirmation does not evaluate a CD-FPCA advantage under that alternative.

Oracle coverage ranges from .9448 to .9531. At Monte Carlo resolution it is consistent with the nominal exact-label rank guarantee. This does not turn the GCV point-label experiment into a finite-sample latent-coverage result.

All 1,536 final fits were accepted and all labels succeeded; no infinite bands occurred. Of 4,608 penalty candidates, 4,602 passed the frozen diagnostics and six were rejected. The upper endpoint $\lambda = 3$ was selected in 1,524 of 1,536 fits, and accepted final gradient infinity norms ranged from 1.19×10^{-4} to .0127. The quota failure left 1,375 full raw files; the 161 missing same-seed tasks were rerun into an audited compact bundle. Independent reconstruction verified all 1,536 task keys and all 196,608 result rows. This storage repair changed neither seeds nor scientific settings. The

Table 19: Cellwise GCV comparison: new minus old coverage and width (paired MCSEs in parentheses).

Signal	Law	Innovation	Coverage difference	Width difference
Linear	t_5	Gaussian	-.0029 (.0024)	.832 (.067)
Smooth	t_5	Gaussian	-.0029 (.0028)	1.149 (.109)
Linear	Gaussian	Gaussian	-.0051 (.0030)	.797 (.044)
Smooth	Gaussian	Gaussian	-.0032 (.0031)	.904 (.064)
Linear	t_5	Rademacher	-.0022 (.0024)	.787 (.062)
Smooth	t_5	Rademacher	-.0063 (.0025)	.987 (.066)
Linear	Gaussian	Rademacher	-.0037 (.0029)	.739 (.040)
Smooth	Gaussian	Rademacher	-.0061 (.0029)	.863 (.039)
Linear	t_5	Skewed	-.0032 (.0023)	.884 (.056)
Smooth	t_5	Skewed	-.0005 (.0029)	.998 (.074)
Linear	Gaussian	Skewed	-.0059 (.0023)	.785 (.039)
Smooth	Gaussian	Skewed	-.0093 (.0030)	.873 (.045)

Table 20: Exact-label oracle CD-FPCA results; parentheses are MCSEs.

Signal	Law	Innovation	Coverage	Mean width
Linear	t_5	Gaussian	.9531 (.0032)	4.627 (.055)
Smooth	t_5	Gaussian	.9451 (.0040)	5.373 (.107)
Linear	Gaussian	Gaussian	.9512 (.0036)	4.211 (.044)
Smooth	Gaussian	Gaussian	.9507 (.0031)	4.715 (.064)
Linear	t_5	Rademacher	.9509 (.0040)	4.468 (.061)
Smooth	t_5	Rademacher	.9514 (.0034)	5.213 (.067)
Linear	Gaussian	Rademacher	.9531 (.0034)	4.096 (.044)
Smooth	Gaussian	Rademacher	.9521 (.0037)	4.652 (.041)
Linear	t_5	Skewed	.9495 (.0041)	4.542 (.054)
Smooth	t_5	Skewed	.9517 (.0035)	5.271 (.067)
Linear	Gaussian	Skewed	.9492 (.0033)	4.178 (.040)
Smooth	Gaussian	Skewed	.9448 (.0040)	4.663 (.045)

source release provides a clean-run workflow in `studies/cd_fpca`: it generates 1,536 full raw files before aggregation. The historical mixed archive described above is not bundled.

The experiment supports the prespecified operational criterion: every GCV cell estimate is at most .97, and the only estimate below .95 is .9495 with MCSE .0040. It does not prove a population coverage upper bound, universal 95% recovered-label coverage, conditional-on-history validity, continuum coverage, or a width improvement.