

Supplement to: Estimation and inference in penalized function-on-function regression with dependent residual curves

Fabian Scheipl*

Department of Statistics, LMU Munich

Munich Center for Machine Learning (MCML)

Table of contents

S1 Simulation study	2
S1.1 Coverage per setting	2
S1.2 Lower tails, number of curves, signal and truth roughness	9
S1.3 Error structures and extended families	12
S1.4 Studentized errors	14
S1.5 Estimation	18
S1.6 Relative error, half-width and detection per setting	24
S1.7 Satterthwaite critical values beyond the main factorial	27
S1.8 A zero coefficient surface	28
S1.9 A principal component term for a low-rank covariate	29
S2 Plasmode study	30
S2.1 Coverage per dataset, truth source and residual source	31
S2.2 Estimation by truth source; relative error and detection	37
S2.3 Truth source against residual source; heartbeat errors; block flips	40
S2.4 Studentized errors	42
S3 Comparators	43
S3.1 AR(1) working model	43
S3.2 Random curve effect, GLS and curve bootstrap	44
S3.3 Functional GEE	50
S4 Approaches that did not work	54
S4.1 Intervals centred at the NCV fit	54
S4.2 mgcv's neighbourhood covariance for curve-blocked NCV	62
S4.3 Smoothing-parameter uncertainty term for REML	65

*All code and text were generated by large language models in collaboration with the author; see the statement on the use of generative AI at the end of the main text of the paper.

S5 Misspecified truths and basis size	66
S5.1 Truths outside the spline space	66
S5.2 Basis size with representable truths	69
S6 Data and code	72
References	74

This supplement accompanies the paper. S1 and S2 give the full results of the simulation and the plasmode study, S3 the comparators, S4 the approaches that did not work, S5 misspecified truths and basis size, S6 data and code. The CL2 algebra, the descriptions of both studies’ set-ups and the computing times are in Appendices A–C of the paper. Every number is computed from the study’s summary files when the document is rendered.

Coverage, its 5% quantile over the grid (lower tail), relative error, relative half-width, detection rate and interval score are defined in Section 2.4 of the paper. In addition, a *setting* is one combination of design factors of the simulation, with 200 replicate data sets, and the *wrong-sign rate* is the share of all grid points at which an interval excludes zero on the wrong side. Monte Carlo (MC) standard errors are computed over replicates; for a grid-averaged coverage, from the replicate-level grid averages, so the correlation of coverage indicators across grid points is accounted for. REML + CL2 and NCV + CL2 denote the REML or curve-blocked NCV fit with the CL2 covariance of Appendix A of the paper and normal critical values; model-based intervals use mgcv’s V_c for REML fits and V_p for NCV fits, because in the mgcv build of the study V_c of an NCV fit holds only the smoothing-parameter correction and is not a valid covariance (S6). Error processes of the main factorial: independent, rough stationary (an exponential correlation with a nugget) and smooth (from DTI residual curves). Signal levels: low, default and high ($R^2 = 0.2, 0.5, 0.8$ for Gaussian responses; $\sqrt{S} = 0.5$ (default) and 1 on the link scale for Poisson and binary responses, where “low” and “default” coincide). Bases: the default, and bases with $2\times$ and $4\times$ finer knots. Functional covariate: rich (the default spectrum) or low-rank. The plasmode truths are the fitted means of the REML fit (*REML-fitted truth*), of the NCV fit (*NCV-fitted truth*), and of a refit with every smoothing parameter at the geometric mean of the REML and NCV values (*midpoint truth*).

S1 Simulation study

The design is described in Appendix B.1 of the paper.

S1.1 Coverage per setting

Table 1: Grid-averaged coverage (MC SE) of the mean and $\beta(s, t)$ per setting and interval. REML model / NCV model: model-based intervals of the REML / NCV fit; NCV wid.: NCV + CL2 widened in quadrature by the NCV–REML difference (S4.1(b)), with the Bayesian or the frequentist (freq.) CL2; AR(1): AR(1) working model.

block	family	error	G	signal	truth	D	basis	estimand	NCV+CL2	NCV wid.	NCV wid. (freq.)	NCV model	REML+CL2	REML model	AR(1)
main factorial	Gaussian	indep.	40	low	smooth	61	default	$\beta(s, t)$	0.984 (0.002)	0.988 (0.001)	0.924 (0.004)	0.984 (0.002)	0.990 (0.001)	0.994 (0.001)	
main factorial	Gaussian	indep.	40	low	smooth	61	default	$E(Y X)$	0.946 (0.003)	0.952 (0.003)	0.910 (0.004)	0.952 (0.002)	0.956 (0.002)	0.967 (0.002)	
main factorial	Gaussian	indep.	100	low	smooth	61	default	$\beta(s, t)$	0.982 (0.001)	0.989 (0.001)	0.939 (0.003)	0.983 (0.001)	0.989 (0.001)	0.992 (0.001)	
main factorial	Gaussian	indep.	100	low	smooth	61	default	$E(Y X)$	0.950 (0.002)	0.955 (0.002)	0.922 (0.003)	0.952 (0.002)	0.961 (0.002)	0.968 (0.002)	
main factorial	Gaussian	rough	40	low	smooth	61	default	$\beta(s, t)$	0.898 (0.007)	0.978 (0.003)	0.972 (0.003)	0.648 (0.010)	0.925 (0.004)	0.635 (0.009)	
main factorial	Gaussian	rough	40	low	smooth	61	default	$E(Y X)$	0.883 (0.005)	0.943 (0.003)	0.937 (0.003)	0.546 (0.007)	0.929 (0.003)	0.617 (0.005)	
main factorial	Gaussian	rough	100	low	smooth	61	default	$\beta(s, t)$	0.924 (0.004)	0.982 (0.002)	0.976 (0.002)	0.682 (0.009)	0.941 (0.003)	0.671 (0.008)	
main factorial	Gaussian	rough	100	low	smooth	61	default	$E(Y X)$	0.920 (0.003)	0.954 (0.002)	0.948 (0.002)	0.595 (0.006)	0.942 (0.002)	0.641 (0.005)	
main factorial	Gaussian	smooth	40	low	smooth	61	default	$\beta(s, t)$	0.902 (0.007)	0.979 (0.002)	0.974 (0.003)	0.633 (0.010)	0.923 (0.004)	0.609 (0.010)	
main factorial	Gaussian	smooth	40	low	smooth	61	default	$E(Y X)$	0.895 (0.005)	0.949 (0.003)	0.944 (0.003)	0.545 (0.007)	0.930 (0.003)	0.601 (0.006)	
main factorial	Gaussian	smooth	100	low	smooth	61	default	$\beta(s, t)$	0.925 (0.004)	0.983 (0.002)	0.977 (0.002)	0.667 (0.009)	0.944 (0.003)	0.658 (0.009)	
main factorial	Gaussian	smooth	100	low	smooth	61	default	$E(Y X)$	0.921 (0.003)	0.956 (0.002)	0.951 (0.002)	0.586 (0.006)	0.944 (0.002)	0.631 (0.005)	
main factorial	Gaussian	indep.	40	default	smooth	61	default	$\beta(s, t)$	0.982 (0.002)	0.989 (0.001)	0.937 (0.003)	0.983 (0.002)	0.989 (0.001)	0.992 (0.001)	
main factorial	Gaussian	indep.	40	default	smooth	61	default	$E(Y X)$	0.948 (0.002)	0.954 (0.002)	0.921 (0.003)	0.954 (0.002)	0.958 (0.002)	0.968 (0.002)	
main factorial	Gaussian	indep.	100	default	smooth	61	default	$\beta(s, t)$	0.982 (0.001)	0.989 (0.001)	0.951 (0.003)	0.983 (0.001)	0.986 (0.001)	0.989 (0.001)	
main factorial	Gaussian	indep.	100	default	smooth	61	default	$E(Y X)$	0.950 (0.002)	0.958 (0.002)	0.933 (0.002)	0.953 (0.002)	0.961 (0.002)	0.967 (0.001)	
main factorial	Gaussian	rough	40	default	smooth	61	default	$\beta(s, t)$	0.922 (0.005)	0.982 (0.002)	0.976 (0.002)	0.682 (0.009)	0.927 (0.004)	0.638 (0.008)	
main factorial	Gaussian	rough	40	default	smooth	61	default	$E(Y X)$	0.920 (0.003)	0.956 (0.002)	0.951 (0.002)	0.603 (0.005)	0.931 (0.003)	0.624 (0.005)	
main factorial	Gaussian	rough	100	default	smooth	61	default	$\beta(s, t)$	0.934 (0.004)	0.983 (0.001)	0.977 (0.002)	0.700 (0.007)	0.940 (0.003)	0.672 (0.007)	0.972 (0.002)
main factorial	Gaussian	rough	100	default	smooth	61	default	$E(Y X)$	0.931 (0.003)	0.958 (0.002)	0.953 (0.002)	0.623 (0.005)	0.942 (0.002)	0.646 (0.005)	0.934 (0.003)
main factorial	Gaussian	smooth	40	default	smooth	61	default	$\beta(s, t)$	0.924 (0.005)	0.983 (0.002)	0.978 (0.002)	0.670 (0.009)	0.926 (0.004)	0.615 (0.009)	
main factorial	Gaussian	smooth	40	default	smooth	61	default	$E(Y X)$	0.922 (0.004)	0.959 (0.003)	0.954 (0.003)	0.591 (0.006)	0.931 (0.003)	0.606 (0.005)	
main factorial	Gaussian	smooth	100	default	smooth	61	default	$\beta(s, t)$	0.934 (0.004)	0.983 (0.002)	0.978 (0.002)	0.684 (0.008)	0.944 (0.003)	0.657 (0.008)	0.993 (0.001)
main factorial	Gaussian	smooth	100	default	smooth	61	default	$E(Y X)$	0.929 (0.003)	0.959 (0.002)	0.954 (0.002)	0.607 (0.005)	0.944 (0.002)	0.634 (0.005)	0.972 (0.002)
main factorial	Gaussian	indep.	40	high	smooth	61	default	$\beta(s, t)$	0.981 (0.001)	0.989 (0.001)	0.948 (0.003)	0.983 (0.001)	0.984 (0.001)	0.988 (0.001)	
main factorial	Gaussian	indep.	40	high	smooth	61	default	$E(Y X)$	0.948 (0.002)	0.956 (0.002)	0.930 (0.002)	0.954 (0.002)	0.957 (0.002)	0.966 (0.002)	
main factorial	Gaussian	indep.	100	high	smooth	61	default	$\beta(s, t)$	0.979 (0.001)	0.990 (0.001)	0.959 (0.002)	0.980 (0.001)	0.979 (0.001)	0.983 (0.001)	
main factorial	Gaussian	indep.	100	high	smooth	61	default	$E(Y X)$	0.950 (0.002)	0.960 (0.002)	0.941 (0.002)	0.954 (0.002)	0.958 (0.002)	0.963 (0.001)	
main factorial	Gaussian	rough	40	high	smooth	61	default	$\beta(s, t)$	0.932 (0.003)	0.984 (0.001)	0.978 (0.002)	0.695 (0.007)	0.927 (0.004)	0.637 (0.008)	
main factorial	Gaussian	rough	40	high	smooth	61	default	$E(Y X)$	0.927 (0.003)	0.959 (0.002)	0.955 (0.002)	0.626 (0.005)	0.932 (0.002)	0.628 (0.005)	
main factorial	Gaussian	rough	100	high	smooth	61	default	$\beta(s, t)$	0.936 (0.003)	0.982 (0.001)	0.976 (0.002)	0.703 (0.007)	0.941 (0.003)	0.670 (0.006)	
main factorial	Gaussian	rough	100	high	smooth	61	default	$E(Y X)$	0.934 (0.002)	0.958 (0.002)	0.954 (0.002)	0.638 (0.005)	0.942 (0.002)	0.651 (0.004)	
main factorial	Gaussian	smooth	40	high	smooth	61	default	$\beta(s, t)$	0.931 (0.004)	0.984 (0.002)	0.978 (0.002)	0.684 (0.008)	0.926 (0.004)	0.617 (0.008)	
main factorial	Gaussian	smooth	40	high	smooth	61	default	$E(Y X)$	0.928 (0.003)	0.960 (0.002)	0.956 (0.003)	0.610 (0.005)	0.932 (0.003)	0.611 (0.005)	
main factorial	Gaussian	smooth	100	high	smooth	61	default	$\beta(s, t)$	0.938 (0.003)	0.984 (0.001)	0.978 (0.002)	0.693 (0.007)	0.943 (0.003)	0.653 (0.007)	
main factorial	Gaussian	smooth	100	high	smooth	61	default	$E(Y X)$	0.933 (0.003)	0.959 (0.002)	0.955 (0.002)	0.621 (0.005)	0.944 (0.002)	0.636 (0.004)	
main factorial	Poisson	indep.	40	default	smooth	61	default	$\beta(s, t)$	0.985 (0.001)	0.990 (0.001)	0.936 (0.003)	0.986 (0.001)	0.989 (0.001)	0.993 (0.001)	
main factorial	Poisson	indep.	40	default	smooth	61	default	$E(Y X)$	0.947 (0.002)	0.953 (0.002)	0.916 (0.003)	0.954 (0.002)	0.958 (0.002)	0.968 (0.002)	
main factorial	Poisson	indep.	100	default	smooth	61	default	$\beta(s, t)$	0.982 (0.001)	0.990 (0.001)	0.949 (0.002)	0.983 (0.001)	0.986 (0.001)	0.989 (0.001)	
main factorial	Poisson	indep.	100	default	smooth	61	default	$E(Y X)$	0.948 (0.002)	0.956 (0.002)	0.929 (0.002)	0.952 (0.002)	0.960 (0.002)	0.966 (0.002)	
main factorial	Poisson	rough	40	default	smooth	61	default	$\beta(s, t)$	0.921 (0.005)	0.980 (0.002)	0.972 (0.003)	0.705 (0.009)	0.927 (0.004)	0.684 (0.008)	
main factorial	Poisson	rough	40	default	smooth	61	default	$E(Y X)$	0.910 (0.004)	0.949 (0.003)	0.943 (0.003)	0.612 (0.006)	0.930 (0.003)	0.657 (0.005)	
main factorial	Poisson	rough	100	default	smooth	61	default	$\beta(s, t)$	0.936 (0.003)	0.982 (0.002)	0.976 (0.002)	0.715 (0.007)	0.942 (0.003)	0.698 (0.007)	
main factorial	Poisson	rough	100	default	smooth	61	default	$E(Y X)$	0.929 (0.003)	0.956 (0.002)	0.951 (0.002)	0.632 (0.005)	0.943 (0.002)	0.664 (0.004)	
main factorial	Poisson	smooth	40	default	smooth	61	default	$\beta(s, t)$	0.926 (0.005)	0.983 (0.002)	0.976 (0.003)	0.691 (0.009)	0.928 (0.004)	0.658 (0.009)	
main factorial	Poisson	smooth	40	default	smooth	61	default	$E(Y X)$	0.914 (0.004)	0.954 (0.003)	0.948 (0.003)	0.603 (0.007)	0.934 (0.003)	0.642 (0.005)	
main factorial	Poisson	smooth	100	default	smooth	61	default	$\beta(s, t)$	0.935 (0.004)	0.984 (0.002)	0.978 (0.002)	0.706 (0.008)	0.945 (0.003)	0.682 (0.008)	
main factorial	Poisson	smooth	100	default	smooth	61	default	$E(Y X)$	0.927 (0.003)	0.957 (0.002)	0.952 (0.002)	0.617 (0.005)	0.945 (0.002)	0.653 (0.005)	
main factorial	Poisson	indep.	40	high	smooth	61	default	$\beta(s, t)$	0.981 (0.001)	0.988 (0.001)	0.936 (0.003)	0.983 (0.001)	0.986 (0.001)	0.990 (0.001)	
main factorial	Poisson	indep.	40	high	smooth	61	default	$E(Y X)$	0.939 (0.003)	0.950 (0.002)	0.918 (0.003)	0.948 (0.002)	0.955 (0.002)	0.966 (0.002)	
main factorial	Poisson	indep.	100	high	smooth	61	default	$\beta(s, t)$	0.976 (0.002)	0.986 (0.001)	0.948 (0.002)	0.979 (0.002)	0.979 (0.001)	0.985 (0.001)	
main factorial	Poisson	indep.	100	high	smooth	61	default	$E(Y X)$	0.941 (0.002)	0.954 (0.002)	0.930 (0.003)	0.947 (0.002)	0.955 (0.002)	0.962 (0.002)	
main factorial	Poisson	rough	40	high	smooth	61	default	$\beta(s, t)$	0.922 (0.004)	0.979 (0.002)	0.970 (0.002)	0.715 (0.008)	0.929 (0.004)	0.703 (0.007)	
main factorial	Poisson	rough	40	high	smooth	61	default	$E(Y X)$	0.902 (0.004)	0.947 (0.003)	0.940 (0.003)	0.627 (0.006)	0.930 (0.003)	0.674 (0.005)	
main factorial	Poisson	rough	100	high	smooth	61	default	$\beta(s, t)$	0.933 (0.003)	0.981 (0.001)	0.974 (0.002)	0.723 (0.006)	0.940 (0.003)	0.705 (0.007)	
main factorial	Poisson	rough	100	high	smooth	61	default	$E(Y X)$	0.922 (0.003)	0.953 (0.002)	0.947 (0.002)	0.643 (0.005)	0.940 (0.002)	0.672 (0.005)	
main factorial	Poisson	smooth	40	high	smooth	61	default	$\beta(s, t)$	0.923 (0.004)	0.979 (0.002)	0.972 (0.002)	0.699 (0.008)	0.930 (0.004)	0.683 (0.008)	
main factorial	Poisson	smooth	40	high	smooth	61	default	$E(Y X)$	0.905 (0.004)	0.949 (0.003)	0.943 (0.003)	0.610 (0.006)	0.934 (0.003)	0.656 (0.005)	
main factorial	Poisson	smooth	100	high	smooth	61	default	$\beta(s, t)$	0.929 (0.003)	0.980 (0.002)	0.973 (0.002)	0.711 (0.007)	0.944 (0.003)	0.696 (0.007)	
main factorial	Poisson	smooth	100	high	smooth	61	default	$E(Y X)$	0.923 (0.003)	0.954 (0.002)	0.949 (0.002)	0.625 (0.006)	0.945 (0.002)	0.662 (0.004)	
main factorial	binary	indep.	40	default	smooth	61	default	$\beta(s, t)$	0.973 (0.005)	0.979 (0.004)	0.893 (0.007)	0.974 (0.004)	0.973 (0.005)	0.984 (0.004)	
main factorial	binary	indep.	40	default	smooth	61	default	$E(Y X)$	0.923 (0.004)	0.935 (0.004)	0.879 (0.005)	0.927 (0.004)	0.925 (0.004)	0.952 (0.003)	

Table 1: Grid-averaged coverage (MC SE) of the mean and $\beta(s, t)$ per setting and interval. REML model / NCV model: model-based intervals of the REML / NCV fit; NCV wid.: NCV + CL2 widened in quadrature by the NCV–REML difference (S4.1(b)), with the Bayesian or the frequentist (freq.) CL2; AR(1): AR(1) working model. (*continued*)

block	family	error	G	signal	truth	D	basis	estimand	NCV+CL2	NCV wid.	NCV wid. (freq.)	NCV model	REML+CL2	REML model	AR(1)
main factorial	binary	indep.	100	default	smooth	61	default	$\beta(s, t)$	0.982 (0.002)	0.987 (0.001)	0.911 (0.005)	0.982 (0.002)	0.990 (0.001)	0.994 (0.001)	
main factorial	binary	indep.	100	default	smooth	61	default	$E(Y X)$	0.939 (0.003)	0.945 (0.003)	0.896 (0.004)	0.942 (0.003)	0.949 (0.003)	0.963 (0.002)	
main factorial	binary	rough	40	default	smooth	61	default	$\beta(s, t)$	0.790 (0.014)	0.948 (0.005)	0.935 (0.005)	0.627 (0.017)	0.937 (0.004)	0.800 (0.008)	
main factorial	binary	rough	40	default	smooth	61	default	$E(Y X)$	0.842 (0.007)	0.930 (0.003)	0.920 (0.004)	0.587 (0.008)	0.927 (0.003)	0.747 (0.006)	
main factorial	binary	rough	100	default	smooth	61	default	$\beta(s, t)$	0.901 (0.008)	0.971 (0.003)	0.958 (0.004)	0.738 (0.011)	0.951 (0.003)	0.819 (0.007)	
main factorial	binary	rough	100	default	smooth	61	default	$E(Y X)$	0.876 (0.005)	0.933 (0.003)	0.922 (0.003)	0.628 (0.007)	0.940 (0.003)	0.752 (0.005)	
main factorial	binary	smooth	40	default	smooth	61	default	$\beta(s, t)$	0.785 (0.013)	0.953 (0.005)	0.944 (0.005)	0.569 (0.016)	0.926 (0.005)	0.749 (0.010)	
main factorial	binary	smooth	40	default	smooth	61	default	$E(Y X)$	0.857 (0.006)	0.942 (0.003)	0.935 (0.003)	0.550 (0.008)	0.921 (0.004)	0.704 (0.006)	
main factorial	binary	smooth	100	default	smooth	61	default	$\beta(s, t)$	0.896 (0.008)	0.972 (0.003)	0.962 (0.003)	0.708 (0.012)	0.952 (0.003)	0.779 (0.008)	
main factorial	binary	smooth	100	default	smooth	61	default	$E(Y X)$	0.883 (0.004)	0.939 (0.003)	0.930 (0.003)	0.601 (0.007)	0.941 (0.003)	0.722 (0.005)	
main factorial	binary	indep.	40	high	smooth	61	default	$\beta(s, t)$	0.978 (0.002)	0.985 (0.002)	0.912 (0.005)	0.980 (0.002)	0.986 (0.003)	0.991 (0.003)	
main factorial	binary	indep.	40	high	smooth	61	default	$E(Y X)$	0.937 (0.003)	0.945 (0.003)	0.895 (0.004)	0.943 (0.003)	0.949 (0.003)	0.964 (0.003)	
main factorial	binary	indep.	100	high	smooth	61	default	$\beta(s, t)$	0.980 (0.002)	0.987 (0.001)	0.928 (0.004)	0.982 (0.002)	0.989 (0.001)	0.993 (0.001)	
main factorial	binary	indep.	100	high	smooth	61	default	$E(Y X)$	0.943 (0.002)	0.951 (0.002)	0.911 (0.003)	0.948 (0.002)	0.959 (0.002)	0.967 (0.002)	
main factorial	binary	rough	40	high	smooth	61	default	$\beta(s, t)$	0.905 (0.008)	0.974 (0.003)	0.959 (0.004)	0.755 (0.010)	0.940 (0.004)	0.813 (0.008)	
main factorial	binary	rough	40	high	smooth	61	default	$E(Y X)$	0.871 (0.005)	0.935 (0.003)	0.922 (0.003)	0.651 (0.007)	0.929 (0.003)	0.761 (0.005)	
main factorial	binary	rough	100	high	smooth	61	default	$\beta(s, t)$	0.930 (0.004)	0.980 (0.002)	0.968 (0.002)	0.780 (0.008)	0.951 (0.003)	0.820 (0.007)	
main factorial	binary	rough	100	high	smooth	61	default	$E(Y X)$	0.913 (0.003)	0.949 (0.002)	0.938 (0.002)	0.699 (0.005)	0.944 (0.002)	0.766 (0.004)	
main factorial	binary	smooth	40	high	smooth	61	default	$\beta(s, t)$	0.897 (0.008)	0.974 (0.003)	0.962 (0.004)	0.722 (0.011)	0.932 (0.004)	0.765 (0.009)	
main factorial	binary	smooth	40	high	smooth	61	default	$E(Y X)$	0.874 (0.005)	0.942 (0.003)	0.931 (0.003)	0.624 (0.007)	0.923 (0.003)	0.713 (0.006)	
main factorial	binary	smooth	100	high	smooth	61	default	$\beta(s, t)$	0.925 (0.005)	0.980 (0.002)	0.970 (0.002)	0.759 (0.008)	0.952 (0.003)	0.783 (0.007)	
main factorial	binary	smooth	100	high	smooth	61	default	$E(Y X)$	0.907 (0.004)	0.948 (0.003)	0.939 (0.003)	0.663 (0.006)	0.943 (0.002)	0.729 (0.005)	
misreg.	Gaussian	misreg.	40	high	smooth	61	default	$\beta(s, t)$	0.937 (0.004)	0.974 (0.003)	0.956 (0.004)	0.752 (0.008)	0.949 (0.004)	0.761 (0.007)	
misreg.	Gaussian	misreg.	40	high	smooth	61	default	$E(Y X)$	0.919 (0.004)	0.946 (0.003)	0.934 (0.004)	0.696 (0.007)	0.938 (0.003)	0.729 (0.005)	
misreg.	Gaussian	misreg.	40	high	rough	61	default	$\beta(s, t)$	0.820 (0.008)	0.928 (0.003)	0.901 (0.004)	0.595 (0.009)	0.945 (0.003)	0.739 (0.006)	
misreg.	Gaussian	misreg.	40	high	rough	61	default	$E(Y X)$	0.885 (0.004)	0.932 (0.003)	0.918 (0.003)	0.635 (0.006)	0.934 (0.003)	0.706 (0.005)	
misreg.	Gaussian	misreg.	100	high	smooth	61	default	$\beta(s, t)$	0.937 (0.004)	0.976 (0.002)	0.959 (0.003)	0.739 (0.008)	0.951 (0.003)	0.753 (0.007)	0.875 (0.005)
misreg.	Gaussian	misreg.	100	high	smooth	61	default	$E(Y X)$	0.930 (0.004)	0.952 (0.003)	0.941 (0.003)	0.706 (0.006)	0.942 (0.003)	0.730 (0.005)	0.844 (0.005)
misreg.	Gaussian	misreg.	100	high	rough	61	default	$\beta(s, t)$	0.863 (0.006)	0.934 (0.003)	0.908 (0.004)	0.628 (0.007)	0.944 (0.003)	0.729 (0.006)	
misreg.	Gaussian	misreg.	100	high	rough	61	default	$E(Y X)$	0.911 (0.003)	0.938 (0.003)	0.926 (0.003)	0.665 (0.005)	0.937 (0.003)	0.706 (0.005)	
misreg.	Poisson	misreg.	40	high	smooth	61	default	$E(Y X)$	0.792 (0.007)	0.910 (0.004)	0.900 (0.004)	0.486 (0.010)	0.902 (0.004)	0.634 (0.005)	
misreg.	Poisson	misreg.	40	high	rough	61	default	$E(Y X)$	0.773 (0.007)	0.910 (0.003)	0.902 (0.003)	0.454 (0.009)	0.899 (0.003)	0.614 (0.005)	
misreg.	Poisson	misreg.	100	high	smooth	61	default	$E(Y X)$	0.801 (0.007)	0.922 (0.003)	0.915 (0.003)	0.438 (0.009)	0.910 (0.003)	0.571 (0.004)	
misreg.	Poisson	misreg.	100	high	rough	61	default	$E(Y X)$	0.780 (0.006)	0.918 (0.003)	0.912 (0.003)	0.409 (0.007)	0.906 (0.003)	0.558 (0.004)	
misreg.	binary	misreg.	40	high	smooth	61	default	$E(Y X)$	0.919 (0.004)	0.932 (0.003)	0.884 (0.004)	0.901 (0.004)	0.944 (0.003)	0.945 (0.003)	
misreg.	binary	misreg.	40	high	rough	61	default	$E(Y X)$	0.892 (0.004)	0.908 (0.004)	0.854 (0.005)	0.871 (0.005)	0.925 (0.003)	0.924 (0.003)	
misreg.	binary	misreg.	100	high	smooth	61	default	$E(Y X)$	0.931 (0.003)	0.941 (0.003)	0.902 (0.003)	0.909 (0.004)	0.954 (0.002)	0.943 (0.003)	
misreg.	binary	misreg.	100	high	rough	61	default	$E(Y X)$	0.909 (0.003)	0.919 (0.003)	0.878 (0.004)	0.882 (0.004)	0.932 (0.003)	0.918 (0.003)	
dense grid	Gaussian	indep.	100	default	smooth	241	default	$\beta(s, t)$	0.976 (0.001)	0.988 (0.001)	0.956 (0.002)	0.978 (0.001)	0.976 (0.001)	0.980 (0.001)	
dense grid	Gaussian	indep.	100	default	smooth	241	default	$E(Y X)$	0.947 (0.002)	0.957 (0.002)	0.938 (0.002)	0.951 (0.002)	0.956 (0.002)	0.961 (0.001)	
dense grid	Gaussian	rough	100	default	smooth	241	default	$\beta(s, t)$	0.917 (0.004)	0.988 (0.001)	0.987 (0.001)	0.406 (0.006)	0.940 (0.002)	0.359 (0.004)	
dense grid	Gaussian	rough	100	default	smooth	241	default	$E(Y X)$	0.926 (0.003)	0.966 (0.002)	0.965 (0.002)	0.348 (0.004)	0.941 (0.002)	0.367 (0.003)	
dense grid	Gaussian	smooth	100	default	smooth	241	default	$\beta(s, t)$	0.916 (0.004)	0.989 (0.001)	0.988 (0.001)	0.391 (0.007)	0.944 (0.002)	0.350 (0.004)	
dense grid	Gaussian	smooth	100	default	smooth	241	default	$E(Y X)$	0.923 (0.003)	0.967 (0.002)	0.966 (0.002)	0.333 (0.004)	0.943 (0.002)	0.353 (0.003)	
dense grid	Poisson	indep.	100	default	smooth	241	default	$\beta(s, t)$	0.978 (0.001)	0.988 (0.001)	0.955 (0.002)	0.979 (0.001)	0.978 (0.001)	0.983 (0.001)	
dense grid	Poisson	indep.	100	default	smooth	241	default	$E(Y X)$	0.947 (0.002)	0.958 (0.002)	0.936 (0.002)	0.951 (0.002)	0.957 (0.002)	0.963 (0.001)	
dense grid	Poisson	rough	100	default	smooth	241	default	$\beta(s, t)$	0.915 (0.004)	0.988 (0.001)	0.986 (0.001)	0.418 (0.006)	0.938 (0.002)	0.374 (0.004)	
dense grid	Poisson	rough	100	default	smooth	241	default	$E(Y X)$	0.922 (0.003)	0.965 (0.002)	0.964 (0.002)	0.357 (0.004)	0.941 (0.002)	0.380 (0.003)	
dense grid	Poisson	smooth	100	default	smooth	241	default	$\beta(s, t)$	0.915 (0.004)	0.988 (0.001)	0.987 (0.001)	0.406 (0.007)	0.943 (0.003)	0.369 (0.004)	
dense grid	Poisson	smooth	100	default	smooth	241	default	$E(Y X)$	0.921 (0.003)	0.966 (0.002)	0.965 (0.002)	0.342 (0.004)	0.943 (0.002)	0.370 (0.003)	
dense grid	binary	indep.	100	default	smooth	241	default	$\beta(s, t)$	0.984 (0.001)	0.990 (0.001)	0.937 (0.003)	0.985 (0.001)	0.990 (0.001)	0.993 (0.001)	
dense grid	binary	indep.	100	default	smooth	241	default	$E(Y X)$	0.951 (0.002)	0.956 (0.002)	0.921 (0.003)	0.955 (0.002)	0.962 (0.002)	0.970 (0.002)	
dense grid	binary	rough	100	default	smooth	241	default	$\beta(s, t)$	0.868 (0.010)	0.986 (0.001)	0.984 (0.001)	0.464 (0.010)	0.938 (0.002)	0.450 (0.005)	
dense grid	binary	rough	100	default	smooth	241	default	$E(Y X)$	0.861 (0.006)	0.953 (0.002)	0.951 (0.002)	0.365 (0.005)	0.938 (0.002)	0.452 (0.003)	
dense grid	binary	smooth	100	default	smooth	241	default	$\beta(s, t)$	0.866 (0.009)	0.986 (0.001)	0.984 (0.001)	0.422 (0.010)	0.943 (0.002)	0.420 (0.005)	
dense grid	binary	smooth	100	default	smooth	241	default	$E(Y X)$	0.872 (0.005)	0.959 (0.002)	0.958 (0.002)	0.337 (0.005)	0.941 (0.002)	0.418 (0.003)	
rough truth	Gaussian	indep.	100	low	rough	61	default	$\beta(s, t)$	0.933 (0.004)	0.938 (0.004)	0.843 (0.007)	0.933 (0.004)	0.948 (0.003)	0.957 (0.002)	
rough truth	Gaussian	indep.	100	low	rough	61	default	$E(Y X)$	0.942 (0.002)	0.945 (0.002)	0.913 (0.003)	0.945 (0.002)	0.948 (0.002)	0.956 (0.002)	
rough truth	Gaussian	smooth	100	low	rough	61	default	$\beta(s, t)$	0.803 (0.006)	0.945 (0.002)	0.934 (0.003)	0.515 (0.006)	0.942 (0.003)	0.645 (0.007)	
rough truth	Gaussian	smooth	100	low	rough	61	default	$E(Y X)$	0.912 (0.003)	0.948 (0.002)	0.943 (0.002)	0.579 (0.005)	0.943 (0.002)	0.627 (0.005)	
rough truth	Gaussian	indep.	100	default	rough	61	default	$\beta(s, t)$	0.953 (0.003)	0.955 (0.003)	0.894 (0.004)	0.955 (0.003)	0.961 (0.002)	0.967 (0.002)	
rough truth	Gaussian	indep.	100	default	rough	61	default	$E(Y X)$	0.946 (0.002)	0.949 (0.002)	0.926 (0.002)	0.951 (0.002)	0.952 (0.002)	0.958 (0.002)	

Table 1: Grid-averaged coverage (MC SE) of the mean and $\beta(s, t)$ per setting and interval. REML model / NCV model: model-based intervals of the REML / NCV fit; NCV wid.: NCV + CL2 widened in quadrature by the NCV–REML difference (S4.1(b)), with the Bayesian or the frequentist (freq.) CL2; AR(1): AR(1) working model. (*continued*)

block	family	error	G	signal	truth	D	basis	estimand	NCV+CL2	NCV wid.	NCV wid. (freq.)	NCV model	REML+CL2	REML model	AR(1)
rough truth	Gaussian	smooth	100	default	rough	61	default	$\beta(s, t)$	0.824 (0.007)	0.935 (0.003)	0.924 (0.003)	0.528 (0.007)	0.941 (0.003)	0.642 (0.007)	
rough truth	Gaussian	smooth	100	default	rough	61	default	$E(Y X)$	0.920 (0.002)	0.948 (0.002)	0.944 (0.002)	0.597 (0.004)	0.944 (0.002)	0.631 (0.004)	
rough truth	Gaussian	indep.	100	high	rough	61	default	$\beta(s, t)$	0.954 (0.002)	0.959 (0.002)	0.916 (0.003)	0.959 (0.002)	0.963 (0.002)	0.969 (0.002)	
rough truth	Gaussian	indep.	100	high	rough	61	default	$E(Y X)$	0.946 (0.002)	0.949 (0.002)	0.933 (0.002)	0.951 (0.002)	0.952 (0.002)	0.958 (0.001)	
rough truth	Gaussian	smooth	100	high	rough	61	default	$\beta(s, t)$	0.900 (0.005)	0.950 (0.003)	0.940 (0.003)	0.587 (0.007)	0.942 (0.003)	0.649 (0.006)	
rough truth	Gaussian	smooth	100	high	rough	61	default	$E(Y X)$	0.952 (0.002)	0.950 (0.002)	0.946 (0.002)	0.614 (0.004)	0.944 (0.002)	0.635 (0.004)	
rough truth	Poisson	indep.	100	default	rough	61	default	$\beta(s, t)$	0.952 (0.003)	0.955 (0.002)	0.889 (0.004)	0.955 (0.002)	0.960 (0.002)	0.966 (0.002)	
rough truth	Poisson	indep.	100	default	rough	61	default	$E(Y X)$	0.944 (0.002)	0.947 (0.002)	0.922 (0.002)	0.948 (0.002)	0.950 (0.002)	0.957 (0.002)	
rough truth	Poisson	smooth	100	default	rough	61	default	$\beta(s, t)$	0.811 (0.007)	0.932 (0.003)	0.920 (0.003)	0.530 (0.007)	0.943 (0.003)	0.670 (0.007)	
rough truth	Poisson	smooth	100	default	rough	61	default	$E(Y X)$	0.918 (0.003)	0.947 (0.002)	0.942 (0.002)	0.607 (0.004)	0.944 (0.002)	0.652 (0.004)	
rough truth	Poisson	indep.	100	high	rough	61	default	$\beta(s, t)$	0.950 (0.002)	0.955 (0.002)	0.903 (0.004)	0.955 (0.002)	0.961 (0.002)	0.968 (0.002)	
rough truth	Poisson	indep.	100	high	rough	61	default	$E(Y X)$	0.940 (0.002)	0.946 (0.002)	0.925 (0.002)	0.945 (0.002)	0.949 (0.002)	0.957 (0.002)	
rough truth	Poisson	smooth	100	high	rough	61	default	$\beta(s, t)$	0.860 (0.007)	0.935 (0.003)	0.924 (0.004)	0.578 (0.007)	0.940 (0.003)	0.677 (0.006)	
rough truth	Poisson	smooth	100	high	rough	61	default	$E(Y X)$	0.919 (0.003)	0.945 (0.002)	0.940 (0.002)	0.617 (0.005)	0.943 (0.002)	0.661 (0.004)	
rough truth	binary	indep.	100	default	rough	61	default	$\beta(s, t)$	0.911 (0.004)	0.921 (0.004)	0.784 (0.008)	0.912 (0.004)	0.935 (0.003)	0.950 (0.002)	
rough truth	binary	indep.	100	default	rough	61	default	$E(Y X)$	0.926 (0.003)	0.934 (0.003)	0.886 (0.004)	0.929 (0.003)	0.927 (0.003)	0.944 (0.002)	
rough truth	binary	smooth	100	default	rough	61	default	$\beta(s, t)$	0.799 (0.009)	0.940 (0.004)	0.923 (0.004)	0.587 (0.010)	0.948 (0.003)	0.766 (0.008)	
rough truth	binary	smooth	100	default	rough	61	default	$E(Y X)$	0.872 (0.005)	0.933 (0.003)	0.924 (0.003)	0.596 (0.007)	0.940 (0.002)	0.718 (0.005)	
rough truth	binary	indep.	100	high	rough	61	default	$\beta(s, t)$	0.919 (0.005)	0.929 (0.005)	0.820 (0.008)	0.921 (0.005)	0.944 (0.003)	0.954 (0.003)	
rough truth	binary	indep.	100	high	rough	61	default	$E(Y X)$	0.934 (0.002)	0.938 (0.002)	0.898 (0.003)	0.939 (0.002)	0.942 (0.002)	0.952 (0.002)	
rough truth	binary	smooth	100	high	rough	61	default	$\beta(s, t)$	0.803 (0.006)	0.931 (0.003)	0.912 (0.004)	0.598 (0.006)	0.947 (0.003)	0.767 (0.007)	
rough truth	binary	smooth	100	high	rough	61	default	$E(Y X)$	0.899 (0.003)	0.941 (0.002)	0.931 (0.003)	0.655 (0.005)	0.943 (0.002)	0.729 (0.004)	
smooth effect	Gaussian	indep.	100	default	smooth	61	default	$E(Y X)$	0.956 (0.002)	0.965 (0.001)	0.933 (0.002)	0.959 (0.002)	0.961 (0.001)	0.968 (0.001)	
smooth effect	Gaussian	smooth	100	default	smooth	61	default	$E(Y X)$	0.925 (0.003)	0.957 (0.002)	0.949 (0.002)	0.628 (0.005)	0.938 (0.002)	0.668 (0.004)	
smooth effect	Gaussian	smooth	100	default	smooth	61	default	$E(Y X)$	0.925 (0.003)	0.957 (0.002)	0.950 (0.002)	0.614 (0.005)	0.937 (0.002)	0.650 (0.004)	
smooth effect	Poisson	indep.	100	default	smooth	61	default	$E(Y X)$	0.947 (0.002)	0.958 (0.002)	0.921 (0.002)	0.950 (0.002)	0.961 (0.001)	0.969 (0.001)	
smooth effect	Poisson	rough	100	default	smooth	61	default	$E(Y X)$	0.910 (0.003)	0.951 (0.002)	0.943 (0.002)	0.632 (0.005)	0.939 (0.002)	0.689 (0.004)	
smooth effect	Poisson	smooth	100	default	smooth	61	default	$E(Y X)$	0.911 (0.003)	0.952 (0.002)	0.945 (0.002)	0.620 (0.005)	0.936 (0.002)	0.668 (0.004)	
smooth effect	binary	indep.	100	default	smooth	61	default	$E(Y X)$	0.945 (0.002)	0.954 (0.002)	0.887 (0.004)	0.949 (0.003)	0.963 (0.002)	0.974 (0.002)	
smooth effect	binary	rough	100	default	smooth	61	default	$E(Y X)$	0.864 (0.006)	0.934 (0.004)	0.917 (0.004)	0.648 (0.007)	0.936 (0.002)	0.772 (0.004)	
smooth effect	binary	smooth	100	default	smooth	61	default	$E(Y X)$	0.877 (0.005)	0.945 (0.003)	0.933 (0.003)	0.629 (0.007)	0.936 (0.002)	0.743 (0.005)	
families	scaled t	indep.	100	default	smooth	61	default	$\beta(s, t)$	0.982 (0.001)	0.990 (0.001)	0.954 (0.002)	0.983 (0.001)	0.985 (0.001)	0.988 (0.001)	
families	scaled t	indep.	100	default	smooth	61	default	$E(Y X)$	0.950 (0.002)	0.959 (0.002)	0.936 (0.002)	0.953 (0.002)	0.960 (0.001)	0.965 (0.001)	
families	scaled t	smooth	100	default	smooth	61	default	$\beta(s, t)$	0.940 (0.004)	0.985 (0.001)	0.979 (0.002)	0.703 (0.008)	0.944 (0.003)	0.671 (0.007)	
families	scaled t	smooth	100	default	smooth	61	default	$E(Y X)$	0.938 (0.003)	0.962 (0.002)	0.958 (0.002)	0.636 (0.005)	0.947 (0.002)	0.644 (0.005)	
families	beta	indep.	100	default	smooth	61	default	$\beta(s, t)$	0.979 (0.002)	0.988 (0.001)	0.947 (0.003)	0.981 (0.002)	0.984 (0.001)	0.989 (0.001)	
families	beta	indep.	100	default	smooth	61	default	$E(Y X)$	0.946 (0.002)	0.956 (0.002)	0.930 (0.002)	0.953 (0.002)	0.958 (0.002)	0.966 (0.001)	
families	beta	smooth	100	default	smooth	61	default	$\beta(s, t)$	0.929 (0.004)	0.982 (0.002)	0.976 (0.002)	0.691 (0.008)	0.941 (0.003)	0.659 (0.007)	
families	beta	smooth	100	default	smooth	61	default	$E(Y X)$	0.924 (0.003)	0.956 (0.002)	0.951 (0.002)	0.623 (0.005)	0.941 (0.002)	0.635 (0.004)	
families	negative binomial	indep.	100	default	smooth	61	default	$\beta(s, t)$	0.982 (0.002)	0.990 (0.001)	0.944 (0.003)	0.983 (0.001)	0.988 (0.001)	0.991 (0.001)	
families	negative binomial	indep.	100	default	smooth	61	default	$E(Y X)$	0.950 (0.002)	0.958 (0.002)	0.930 (0.003)	0.954 (0.002)	0.962 (0.002)	0.968 (0.002)	
families	negative binomial	smooth	100	default	smooth	61	default	$\beta(s, t)$	0.932 (0.004)	0.983 (0.002)	0.977 (0.002)	0.699 (0.009)	0.943 (0.003)	0.669 (0.008)	
families	negative binomial	smooth	100	default	smooth	61	default	$E(Y X)$	0.925 (0.003)	0.956 (0.002)	0.951 (0.002)	0.622 (0.005)	0.943 (0.002)	0.641 (0.005)	
families	negative binomial	indep.	100	default	smooth	61	default	$\beta(s, t)$	0.962 (0.002)	0.982 (0.001)	0.955 (0.002)	0.929 (0.003)	0.972 (0.002)	0.945 (0.002)	
families	negative binomial	indep.	100	default	smooth	61	default	$E(Y X)$	0.934 (0.002)	0.950 (0.002)	0.932 (0.002)	0.868 (0.004)	0.952 (0.002)	0.898 (0.003)	
families	negative binomial	smooth	100	default	smooth	61	default	$\beta(s, t)$	0.913 (0.005)	0.984 (0.001)	0.981 (0.002)	0.558 (0.008)	0.938 (0.003)	0.531 (0.006)	
families	negative binomial	smooth	100	default	smooth	61	default	$E(Y X)$	0.913 (0.003)	0.957 (0.002)	0.954 (0.002)	0.486 (0.005)	0.939 (0.002)	0.529 (0.004)	
exact AR(1)	Gaussian	AR(1)	100	default	smooth	61	default	$\beta(s, t)$	0.933 (0.003)	0.984 (0.001)	0.979 (0.002)	0.684 (0.007)	0.940 (0.003)	0.651 (0.007)	0.989 (0.001)
exact AR(1)	Gaussian	AR(1)	100	default	smooth	61	default	$E(Y X)$	0.930 (0.003)	0.959 (0.002)	0.955 (0.002)	0.603 (0.005)	0.942 (0.002)	0.625 (0.004)	0.961 (0.002)
sign-changing	Gaussian	sign-chg.	100	default	smooth	61	default	$\beta(s, t)$	0.935 (0.003)	0.985 (0.001)	0.977 (0.001)	0.782 (0.005)	0.942 (0.003)	0.700 (0.006)	0.992 (0.001)
sign-changing	Gaussian	sign-chg.	100	default	smooth	61	default	$E(Y X)$	0.915 (0.003)	0.957 (0.002)	0.951 (0.002)	0.642 (0.005)	0.942 (0.002)	0.646 (0.004)	0.963 (0.002)
sign-changing	Poisson	sign-chg.	100	default	smooth	61	default	$\beta(s, t)$	0.935 (0.003)	0.986 (0.001)	0.977 (0.001)	0.791 (0.005)	0.942 (0.003)	0.728 (0.006)	
sign-changing	Poisson	sign-chg.	100	default	smooth	61	default	$E(Y X)$	0.905 (0.003)	0.951 (0.002)	0.944 (0.002)	0.653 (0.005)	0.941 (0.002)	0.670 (0.005)	
sign-changing	binary	sign-chg.	100	default	smooth	61	default	$\beta(s, t)$	0.927 (0.005)	0.977 (0.002)	0.956 (0.003)	0.841 (0.007)	0.959 (0.002)	0.885 (0.005)	
sign-changing	binary	sign-chg.	100	default	smooth	61	default	$E(Y X)$	0.882 (0.005)	0.945 (0.003)	0.930 (0.003)	0.724 (0.006)	0.943 (0.002)	0.794 (0.005)	
misreg. (AR(1))	Gaussian	misreg.	100	high	rough	61	default	$\beta(s, t)$	0.863 (0.006)	0.934 (0.003)	0.908 (0.004)	0.628 (0.007)	0.944 (0.003)	0.729 (0.006)	0.860 (0.005)
misreg. (AR(1))	Gaussian	misreg.	100	high	rough	61	default	$E(Y X)$	0.911 (0.003)	0.938 (0.003)	0.926 (0.003)	0.665 (0.005)	0.937 (0.003)	0.706 (0.005)	0.838 (0.004)
heteroskedastic	Gaussian	var(t)	40	default	smooth	61	default	$\beta(s, t)$	0.918 (0.005)	0.979 (0.003)	0.973 (0.003)	0.662 (0.009)	0.926 (0.004)	0.619 (0.009)	
heteroskedastic	Gaussian	var(t)	40	default	smooth	61	default	$E(Y X)$	0.919 (0.004)	0.957 (0.003)	0.951 (0.003)	0.593 (0.006)	0.932 (0.003)	0.614 (0.005)	
heteroskedastic	Gaussian	var(t)	100	default	smooth	61	default	$\beta(s, t)$	0.933 (0.004)	0.983 (0.002)	0.977 (0.002)	0.686 (0.008)	0.944 (0.003)	0.662 (0.007)	0.991 (0.001)
heteroskedastic	Gaussian	var(t)	100	default	smooth	61	default	$E(Y X)$	0.928 (0.003)	0.959 (0.002)	0.954 (0.002)	0.614 (0.005)	0.944 (0.002)	0.642 (0.005)	0.972 (0.002)
heteroskedastic	Gaussian	var(z)	40	default	smooth	61	default	$\beta(s, t)$	0.925 (0.004)	0.985 (0.002)	0.979 (0.002)	0.668 (0.009)	0.939 (0.003)	0.619 (0.009)	

Table 1: Grid-averaged coverage (MC SE) of the mean and $\beta(s, t)$ per setting and interval. REML model / NCV model: model-based intervals of the REML / NCV fit; NCV wid.: NCV + CL2 widened in quadrature by the NCV–REML difference (S4.1(b)), with the Bayesian or the frequentist (freq.) CL2; AR(1): AR(1) working model. (*continued*)

block	family	error	G	signal	truth	D	basis	estimand	NCV+CL2	NCV wid.	NCV wid. (freq.)	NCV model	REML+CL2	REML model	AR(1)
heteroskedastic	Gaussian	var(z)	40	default	smooth	61	default	$E(Y X)$	0.912 (0.004)	0.955 (0.003)	0.950 (0.003)	0.560 (0.007)	0.934 (0.003)	0.588 (0.006)	
heteroskedastic	Gaussian	var(z)	100	default	smooth	61	default	$\beta(s, t)$	0.935 (0.004)	0.984 (0.001)	0.978 (0.002)	0.683 (0.009)	0.949 (0.003)	0.660 (0.008)	0.992 (0.001)
heteroskedastic	Gaussian	var(z)	100	default	smooth	61	default	$E(Y X)$	0.923 (0.003)	0.955 (0.002)	0.950 (0.002)	0.572 (0.006)	0.944 (0.002)	0.606 (0.005)	0.949 (0.003)
heteroskedastic	Gaussian	var(t,z)	40	default	smooth	61	default	$\beta(s, t)$	0.924 (0.005)	0.983 (0.002)	0.977 (0.002)	0.667 (0.009)	0.940 (0.003)	0.629 (0.009)	
heteroskedastic	Gaussian	var(t,z)	40	default	smooth	61	default	$E(Y X)$	0.911 (0.004)	0.955 (0.003)	0.949 (0.003)	0.563 (0.007)	0.936 (0.003)	0.596 (0.005)	
heteroskedastic	Gaussian	var(t,z)	100	default	smooth	61	default	$\beta(s, t)$	0.934 (0.004)	0.984 (0.001)	0.978 (0.002)	0.688 (0.009)	0.950 (0.003)	0.666 (0.008)	0.990 (0.001)
heteroskedastic	Gaussian	var(t,z)	100	default	smooth	61	default	$E(Y X)$	0.922 (0.003)	0.955 (0.002)	0.950 (0.002)	0.580 (0.006)	0.944 (0.002)	0.615 (0.005)	0.951 (0.002)
low-rank covariate	Gaussian	indep.	100	default	smooth	61	default	$\beta(s, t)$	0.986 (0.001)	0.994 (0.001)	0.953 (0.003)	0.986 (0.001)	0.991 (0.001)	0.994 (0.001)	
low-rank covariate	Gaussian	indep.	100	default	smooth	61	default	$E(Y X)$	0.948 (0.002)	0.956 (0.002)	0.933 (0.002)	0.952 (0.002)	0.960 (0.002)	0.966 (0.001)	
low-rank covariate	Gaussian	smooth	100	default	smooth	61	default	$\beta(s, t)$	0.936 (0.004)	0.987 (0.001)	0.982 (0.002)	0.712 (0.009)	0.945 (0.003)	0.666 (0.010)	
low-rank covariate	Gaussian	smooth	100	default	smooth	61	default	$E(Y X)$	0.930 (0.003)	0.958 (0.002)	0.954 (0.002)	0.607 (0.005)	0.944 (0.002)	0.633 (0.005)	
low-rank covariate	Poisson	indep.	100	default	smooth	61	default	$\beta(s, t)$	0.987 (0.001)	0.994 (0.001)	0.953 (0.003)	0.987 (0.001)	0.992 (0.001)	0.994 (0.001)	
low-rank covariate	Poisson	indep.	100	default	smooth	61	default	$E(Y X)$	0.946 (0.002)	0.954 (0.002)	0.928 (0.003)	0.950 (0.002)	0.959 (0.002)	0.965 (0.002)	
low-rank covariate	Poisson	smooth	100	default	smooth	61	default	$\beta(s, t)$	0.939 (0.004)	0.989 (0.001)	0.984 (0.002)	0.735 (0.009)	0.946 (0.004)	0.693 (0.009)	
low-rank covariate	Poisson	smooth	100	default	smooth	61	default	$E(Y X)$	0.928 (0.003)	0.957 (0.002)	0.952 (0.002)	0.617 (0.006)	0.945 (0.002)	0.652 (0.005)	
low-rank covariate	binary	indep.	100	default	smooth	61	default	$\beta(s, t)$	0.987 (0.002)	0.992 (0.001)	0.915 (0.005)	0.987 (0.002)	0.993 (0.001)	0.996 (0.001)	
low-rank covariate	binary	indep.	100	default	smooth	61	default	$E(Y X)$	0.937 (0.003)	0.944 (0.003)	0.899 (0.004)	0.941 (0.003)	0.945 (0.003)	0.960 (0.002)	
low-rank covariate	binary	smooth	100	default	smooth	61	default	$\beta(s, t)$	0.904 (0.008)	0.977 (0.003)	0.966 (0.003)	0.734 (0.012)	0.954 (0.004)	0.789 (0.010)	
low-rank covariate	binary	smooth	100	default	smooth	61	default	$E(Y X)$	0.890 (0.004)	0.939 (0.003)	0.931 (0.003)	0.609 (0.007)	0.940 (0.003)	0.718 (0.005)	
basis size	Gaussian	indep.	100	default	smooth	61	2* finer knots	$\beta(s, t)$	0.994 (0.001)	0.997 (0.000)	0.959 (0.002)	0.994 (0.001)	0.997 (0.000)	0.998 (0.000)	
basis size	Gaussian	indep.	100	default	smooth	61	2* finer knots	$E(Y X)$	0.962 (0.001)	0.970 (0.001)	0.934 (0.002)	0.964 (0.001)	0.973 (0.001)	0.978 (0.001)	
basis size	Gaussian	indep.	100	default	smooth	61	4* finer knots	$\beta(s, t)$	0.999 (0.000)	0.999 (0.000)	0.964 (0.002)	0.999 (0.000)	1.000 (0.000)	1.000 (0.000)	
basis size	Gaussian	indep.	100	default	smooth	61	4* finer knots	$E(Y X)$	0.972 (0.001)	0.979 (0.001)	0.934 (0.002)	0.973 (0.001)	0.982 (0.001)	0.985 (0.001)	
basis size	Gaussian	smooth	100	default	smooth	61	2* finer knots	$\beta(s, t)$	0.952 (0.003)	0.994 (0.001)	0.990 (0.001)	0.789 (0.006)	0.945 (0.003)	0.671 (0.007)	
basis size	Gaussian	smooth	100	default	smooth	61	2* finer knots	$E(Y X)$	0.933 (0.003)	0.971 (0.001)	0.965 (0.002)	0.660 (0.005)	0.947 (0.002)	0.673 (0.004)	
basis size	Gaussian	smooth	100	default	smooth	61	4* finer knots	$\beta(s, t)$	0.967 (0.002)	0.998 (0.000)	0.995 (0.000)	0.871 (0.005)	0.954 (0.002)	0.673 (0.006)	
basis size	Gaussian	smooth	100	default	smooth	61	4* finer knots	$E(Y X)$	0.937 (0.002)	0.979 (0.001)	0.973 (0.001)	0.707 (0.005)	0.951 (0.002)	0.706 (0.003)	
basis size	Poisson	indep.	100	default	smooth	61	2* finer knots	$\beta(s, t)$	0.994 (0.001)	0.997 (0.000)	0.956 (0.002)	0.995 (0.001)	0.997 (0.000)	0.998 (0.000)	
basis size	Poisson	indep.	100	default	smooth	61	2* finer knots	$E(Y X)$	0.961 (0.002)	0.969 (0.001)	0.929 (0.002)	0.963 (0.002)	0.974 (0.001)	0.978 (0.001)	
basis size	Poisson	indep.	100	default	smooth	61	4* finer knots	$\beta(s, t)$	0.999 (0.000)	0.999 (0.000)	0.961 (0.002)	0.999 (0.000)	1.000 (0.000)	1.000 (0.000)	
basis size	Poisson	indep.	100	default	smooth	61	4* finer knots	$E(Y X)$	0.970 (0.001)	0.977 (0.001)	0.927 (0.002)	0.971 (0.001)	0.982 (0.001)	0.985 (0.001)	
basis size	Poisson	smooth	100	default	smooth	61	2* finer knots	$\beta(s, t)$	0.953 (0.003)	0.995 (0.001)	0.990 (0.001)	0.805 (0.006)	0.948 (0.003)	0.702 (0.007)	
basis size	Poisson	smooth	100	default	smooth	61	2* finer knots	$E(Y X)$	0.930 (0.003)	0.970 (0.001)	0.963 (0.002)	0.666 (0.005)	0.949 (0.002)	0.697 (0.004)	
basis size	Poisson	smooth	100	default	smooth	61	4* finer knots	$\beta(s, t)$	0.970 (0.002)	0.998 (0.000)	0.995 (0.000)	0.886 (0.005)	0.955 (0.002)	0.715 (0.006)	
basis size	Poisson	smooth	100	default	smooth	61	4* finer knots	$E(Y X)$	0.935 (0.003)	0.977 (0.001)	0.970 (0.001)	0.714 (0.005)	0.952 (0.002)	0.738 (0.003)	
basis size	binary	indep.	100	default	smooth	61	2* finer knots	$\beta(s, t)$	0.994 (0.001)	0.996 (0.001)	0.909 (0.004)	0.994 (0.001)	0.997 (0.000)	0.998 (0.000)	
basis size	binary	indep.	100	default	smooth	61	2* finer knots	$E(Y X)$	0.954 (0.002)	0.959 (0.002)	0.895 (0.004)	0.956 (0.002)	0.965 (0.002)	0.975 (0.001)	
basis size	binary	indep.	100	default	smooth	61	4* finer knots	$\beta(s, t)$	0.999 (0.000)	0.999 (0.000)	0.908 (0.004)	0.999 (0.000)	0.999 (0.000)	1.000 (0.000)	
basis size	binary	indep.	100	default	smooth	61	4* finer knots	$E(Y X)$	0.963 (0.002)	0.967 (0.002)	0.891 (0.004)	0.965 (0.002)	0.974 (0.001)	0.982 (0.001)	
basis size	binary	smooth	100	default	smooth	61	2* finer knots	$\beta(s, t)$	0.917 (0.007)	0.990 (0.001)	0.983 (0.001)	0.785 (0.011)	0.958 (0.003)	0.789 (0.007)	
basis size	binary	smooth	100	default	smooth	61	2* finer knots	$E(Y X)$	0.892 (0.004)	0.959 (0.002)	0.950 (0.002)	0.641 (0.007)	0.946 (0.002)	0.753 (0.004)	
basis size	binary	smooth	100	default	smooth	61	4* finer knots	$\beta(s, t)$	0.933 (0.007)	0.997 (0.000)	0.993 (0.001)	0.845 (0.010)	0.960 (0.002)	0.795 (0.006)	
basis size	binary	smooth	100	default	smooth	61	4* finer knots	$E(Y X)$	0.899 (0.004)	0.973 (0.002)	0.964 (0.002)	0.671 (0.007)	0.950 (0.002)	0.781 (0.003)	

Table 2: Coverage (MC SE) of $\alpha(t)$ and $\gamma(t)$ per setting; for the smooth-effect settings the first group of columns holds $f(x, t)$. Interval labels as in the previous table.

block	family	error	G	signal	truth	D	basis	$\alpha(t)$ or $f(x, t)$				$\gamma(t)$			
								REML model	REML+CL2	NCV+CL2	NCV wid.	REML model	REML+CL2	NCV+CL2	NCV wid.
main factorial	Gaussian	indep.	40	low	smooth	61	default	0.954 (0.005)	0.940 (0.005)	0.937 (0.005)	0.941 (0.005)	0.958 (0.005)	0.942 (0.006)	0.935 (0.007)	0.938 (0.007)
main factorial	Gaussian	indep.	100	low	smooth	61	default	0.955 (0.005)	0.950 (0.005)	0.941 (0.005)	0.944 (0.005)	0.968 (0.004)	0.957 (0.005)	0.948 (0.005)	0.952 (0.005)
main factorial	Gaussian	rough	40	low	smooth	61	default	0.621 (0.013)	0.933 (0.007)	0.847 (0.014)	0.912 (0.008)	0.640 (0.012)	0.937 (0.007)	0.867 (0.011)	0.915 (0.008)
main factorial	Gaussian	rough	100	low	smooth	61	default	0.653 (0.011)	0.948 (0.005)	0.924 (0.007)	0.939 (0.006)	0.656 (0.012)	0.937 (0.006)	0.904 (0.008)	0.923 (0.007)
main factorial	Gaussian	smooth	40	low	smooth	61	default	0.608 (0.014)	0.930 (0.007)	0.871 (0.013)	0.919 (0.009)	0.620 (0.014)	0.945 (0.006)	0.889 (0.010)	0.926 (0.008)
main factorial	Gaussian	smooth	100	low	smooth	61	default	0.635 (0.012)	0.948 (0.006)	0.925 (0.007)	0.941 (0.006)	0.649 (0.013)	0.948 (0.006)	0.925 (0.007)	0.942 (0.006)
main factorial	Gaussian	indep.	40	default	smooth	61	default	0.956 (0.004)	0.947 (0.004)	0.940 (0.005)	0.943 (0.005)	0.959 (0.004)	0.951 (0.005)	0.941 (0.005)	0.944 (0.005)
main factorial	Gaussian	indep.	100	default	smooth	61	default	0.956 (0.004)	0.956 (0.004)	0.945 (0.005)	0.949 (0.005)	0.960 (0.004)	0.950 (0.005)	0.946 (0.005)	0.951 (0.004)
main factorial	Gaussian	rough	40	default	smooth	61	default	0.632 (0.012)	0.940 (0.006)	0.924 (0.006)	0.936 (0.006)	0.653 (0.011)	0.938 (0.006)	0.920 (0.007)	0.940 (0.006)
main factorial	Gaussian	rough	100	default	smooth	61	default	0.659 (0.010)	0.951 (0.005)	0.938 (0.006)	0.945 (0.005)	0.663 (0.012)	0.938 (0.006)	0.924 (0.006)	0.936 (0.006)
main factorial	Gaussian	smooth	40	default	smooth	61	default	0.614 (0.013)	0.937 (0.007)	0.919 (0.008)	0.936 (0.007)	0.621 (0.012)	0.946 (0.006)	0.927 (0.007)	0.946 (0.006)
main factorial	Gaussian	smooth	100	default	smooth	61	default	0.641 (0.011)	0.950 (0.005)	0.941 (0.006)	0.949 (0.005)	0.655 (0.012)	0.948 (0.005)	0.935 (0.006)	0.946 (0.006)
main factorial	Gaussian	indep.	40	high	smooth	61	default	0.954 (0.004)	0.947 (0.004)	0.939 (0.004)	0.943 (0.004)	0.959 (0.004)	0.945 (0.005)	0.935 (0.005)	0.943 (0.005)
main factorial	Gaussian	indep.	100	high	smooth	61	default	0.956 (0.004)	0.954 (0.004)	0.950 (0.004)	0.953 (0.004)	0.955 (0.004)	0.944 (0.005)	0.938 (0.005)	0.947 (0.005)
main factorial	Gaussian	rough	40	high	smooth	61	default	0.638 (0.011)	0.941 (0.005)	0.933 (0.006)	0.941 (0.006)	0.657 (0.010)	0.940 (0.006)	0.931 (0.006)	0.943 (0.006)
main factorial	Gaussian	rough	100	high	smooth	61	default	0.665 (0.010)	0.953 (0.005)	0.946 (0.005)	0.949 (0.005)	0.666 (0.011)	0.939 (0.005)	0.928 (0.006)	0.936 (0.005)
main factorial	Gaussian	smooth	40	high	smooth	61	default	0.622 (0.012)	0.940 (0.007)	0.930 (0.007)	0.943 (0.007)	0.629 (0.012)	0.945 (0.006)	0.935 (0.006)	0.949 (0.006)
main factorial	Gaussian	smooth	100	high	smooth	61	default	0.646 (0.011)	0.949 (0.005)	0.946 (0.005)	0.954 (0.005)	0.655 (0.011)	0.947 (0.005)	0.936 (0.006)	0.946 (0.005)
main factorial	Poisson	indep.	40	default	smooth	61	default	0.960 (0.004)	0.953 (0.004)	0.940 (0.005)	0.946 (0.005)	0.955 (0.005)	0.948 (0.005)	0.939 (0.005)	0.943 (0.005)
main factorial	Poisson	indep.	100	default	smooth	61	default	0.954 (0.005)	0.951 (0.005)	0.936 (0.005)	0.943 (0.005)	0.963 (0.004)	0.955 (0.005)	0.947 (0.005)	0.951 (0.005)
main factorial	Poisson	rough	40	default	smooth	61	default	0.655 (0.011)	0.935 (0.006)	0.909 (0.007)	0.937 (0.006)	0.667 (0.012)	0.937 (0.006)	0.905 (0.009)	0.932 (0.007)
main factorial	Poisson	rough	100	default	smooth	61	default	0.668 (0.011)	0.950 (0.005)	0.934 (0.006)	0.948 (0.006)	0.677 (0.011)	0.937 (0.006)	0.923 (0.007)	0.933 (0.006)
main factorial	Poisson	smooth	40	default	smooth	61	default	0.641 (0.013)	0.933 (0.007)	0.913 (0.009)	0.941 (0.007)	0.647 (0.012)	0.943 (0.006)	0.910 (0.008)	0.937 (0.007)
main factorial	Poisson	smooth	100	default	smooth	61	default	0.665 (0.012)	0.948 (0.005)	0.927 (0.007)	0.948 (0.006)	0.662 (0.012)	0.945 (0.005)	0.931 (0.006)	0.944 (0.006)
main factorial	Poisson	indep.	40	high	smooth	61	default	0.961 (0.004)	0.950 (0.004)	0.927 (0.006)	0.946 (0.005)	0.956 (0.005)	0.941 (0.005)	0.927 (0.006)	0.936 (0.005)
main factorial	Poisson	indep.	100	high	smooth	61	default	0.950 (0.005)	0.948 (0.005)	0.926 (0.006)	0.941 (0.005)	0.954 (0.004)	0.943 (0.005)	0.937 (0.005)	0.943 (0.005)
main factorial	Poisson	rough	40	high	smooth	61	default	0.677 (0.011)	0.944 (0.006)	0.898 (0.009)	0.946 (0.006)	0.669 (0.011)	0.930 (0.006)	0.899 (0.008)	0.928 (0.006)
main factorial	Poisson	rough	100	high	smooth	61	default	0.671 (0.011)	0.947 (0.005)	0.918 (0.007)	0.947 (0.005)	0.688 (0.010)	0.938 (0.006)	0.921 (0.007)	0.932 (0.006)
main factorial	Poisson	smooth	40	high	smooth	61	default	0.653 (0.012)	0.940 (0.006)	0.891 (0.009)	0.946 (0.006)	0.658 (0.011)	0.938 (0.006)	0.902 (0.007)	0.934 (0.006)
main factorial	Poisson	smooth	100	high	smooth	61	default	0.659 (0.011)	0.949 (0.005)	0.918 (0.007)	0.949 (0.005)	0.669 (0.011)	0.943 (0.005)	0.921 (0.006)	0.937 (0.006)
main factorial	binary	indep.	40	default	smooth	61	default	0.903 (0.013)	0.859 (0.014)	0.873 (0.012)	0.889 (0.012)	0.931 (0.009)	0.890 (0.011)	0.876 (0.012)	0.896 (0.012)
main factorial	binary	indep.	100	default	smooth	61	default	0.943 (0.006)	0.924 (0.007)	0.912 (0.008)	0.920 (0.007)	0.945 (0.007)	0.923 (0.008)	0.912 (0.009)	0.917 (0.009)
main factorial	binary	rough	40	default	smooth	61	default	0.712 (0.014)	0.904 (0.008)	0.844 (0.012)	0.908 (0.009)	0.724 (0.013)	0.920 (0.009)	0.799 (0.014)	0.885 (0.010)
main factorial	binary	rough	100	default	smooth	61	default	0.749 (0.011)	0.926 (0.007)	0.836 (0.013)	0.902 (0.008)	0.740 (0.012)	0.924 (0.008)	0.826 (0.014)	0.888 (0.011)
main factorial	binary	smooth	40	default	smooth	61	default	0.668 (0.016)	0.885 (0.010)	0.874 (0.011)	0.931 (0.008)	0.691 (0.015)	0.922 (0.008)	0.838 (0.012)	0.915 (0.009)
main factorial	binary	smooth	100	default	smooth	61	default	0.702 (0.013)	0.920 (0.008)	0.856 (0.012)	0.908 (0.008)	0.730 (0.013)	0.939 (0.007)	0.854 (0.012)	0.914 (0.009)
main factorial	binary	indep.	40	high	smooth	61	default	0.956 (0.005)	0.938 (0.006)	0.920 (0.007)	0.929 (0.007)	0.942 (0.007)	0.921 (0.008)	0.910 (0.009)	0.918 (0.008)
main factorial	binary	indep.	100	high	smooth	61	default	0.958 (0.004)	0.949 (0.005)	0.933 (0.006)	0.938 (0.005)	0.948 (0.005)	0.935 (0.006)	0.922 (0.007)	0.926 (0.007)
main factorial	binary	rough	40	high	smooth	61	default	0.747 (0.011)	0.907 (0.008)	0.835 (0.015)	0.910 (0.009)	0.746 (0.012)	0.929 (0.008)	0.817 (0.014)	0.892 (0.009)
main factorial	binary	rough	100	high	smooth	61	default	0.759 (0.011)	0.937 (0.006)	0.909 (0.007)	0.934 (0.006)	0.758 (0.010)	0.941 (0.005)	0.893 (0.009)	0.917 (0.007)
main factorial	binary	smooth	40	high	smooth	61	default	0.686 (0.013)	0.903 (0.008)	0.859 (0.011)	0.929 (0.007)	0.716 (0.013)	0.919 (0.008)	0.833 (0.012)	0.902 (0.009)
main factorial	binary	smooth	100	high	smooth	61	default	0.730 (0.011)	0.929 (0.007)	0.891 (0.008)	0.932 (0.006)	0.737 (0.011)	0.940 (0.006)	0.895 (0.009)	0.926 (0.007)
misreg.	Gaussian	misreg.	40	high	smooth	61	default	0.766 (0.013)	0.943 (0.007)	0.932 (0.008)	0.940 (0.008)	0.718 (0.013)	0.929 (0.007)	0.903 (0.009)	0.927 (0.008)
misreg.	Gaussian	misreg.	40	high	rough	61	default	0.754 (0.012)	0.949 (0.006)	0.924 (0.008)	0.941 (0.007)	0.678 (0.011)	0.915 (0.007)	0.872 (0.009)	0.912 (0.007)
misreg.	Gaussian	misreg.	100	high	smooth	61	default	0.766 (0.012)	0.953 (0.007)	0.948 (0.007)	0.952 (0.007)	0.719 (0.013)	0.933 (0.007)	0.920 (0.007)	0.930 (0.007)
misreg.	Gaussian	misreg.	100	high	rough	61	default	0.732 (0.011)	0.949 (0.006)	0.943 (0.007)	0.950 (0.006)	0.692 (0.013)	0.928 (0.006)	0.923 (0.006)	0.930 (0.006)
dense grid	Gaussian	indep.	100	default	smooth	241	default	0.952 (0.004)	0.951 (0.004)	0.943 (0.004)	0.949 (0.004)	0.951 (0.004)	0.943 (0.004)	0.933 (0.004)	0.939 (0.004)
dense grid	Gaussian	rough	100	default	smooth	241	default	0.382 (0.009)	0.953 (0.005)	0.941 (0.005)	0.953 (0.004)	0.397 (0.011)	0.936 (0.006)	0.919 (0.006)	0.936 (0.006)
dense grid	Gaussian	smooth	100	default	smooth	241	default	0.373 (0.010)	0.947 (0.005)	0.938 (0.006)	0.954 (0.005)	0.369 (0.010)	0.947 (0.005)	0.930 (0.006)	0.950 (0.006)
dense grid	Poisson	indep.	100	default	smooth	241	default	0.954 (0.004)	0.951 (0.004)	0.943 (0.005)	0.951 (0.004)	0.957 (0.004)	0.948 (0.004)	0.935 (0.005)	0.942 (0.004)
dense grid	Poisson	rough	100	default	smooth	241	default	0.397 (0.010)	0.952 (0.005)	0.932 (0.006)	0.956 (0.005)	0.411 (0.010)	0.936 (0.006)	0.915 (0.007)	0.934 (0.006)
dense grid	Poisson	smooth	100	default	smooth	241	default	0.380 (0.010)	0.947 (0.005)	0.925 (0.007)	0.953 (0.005)	0.387 (0.010)	0.947 (0.005)	0.923 (0.006)	0.948 (0.005)
dense grid	binary	indep.	100	default	smooth	241	default	0.957 (0.004)	0.951 (0.005)	0.940 (0.005)	0.945 (0.005)	0.960 (0.004)	0.951 (0.005)	0.942 (0.005)	0.944 (0.005)
dense grid	binary	rough	100	default	smooth	241	default	0.447 (0.012)	0.919 (0.007)	0.831 (0.013)	0.928 (0.007)	0.474 (0.011)	0.941 (0.006)	0.813 (0.014)	0.913 (0.008)
dense grid	binary	smooth	100	default	smooth	241	default	0.410 (0.011)	0.920 (0.007)	0.843 (0.012)	0.938 (0.005)	0.445 (0.011)	0.943 (0.005)	0.843 (0.012)	0.932 (0.007)
rough truth	Gaussian	indep.	100	low	rough	61	default	0.955 (0.004)	0.952 (0.005)	0.941 (0.005)	0.944 (0.005)	0.952 (0.004)	0.942 (0.005)	0.939 (0.005)	0.943 (0.004)
rough truth	Gaussian	smooth	100	low	rough	61	default	0.634 (0.012)	0.947 (0.006)	0.925 (0.007)	0.940 (0.006)	0.642 (0.012)	0.942 (0.006)	0.928 (0.007)	0.938 (0.006)
rough truth	Gaussian	indep.	100	default	rough	61	default	0.956 (0.004)	0.956 (0.004)	0.945 (0.005)	0.950 (0.004)	0.954 (0.004)	0.945 (0.004)	0.940 (0.005)	0.944 (0.004)
rough truth	Gaussian	smooth	100	default	rough	61	default	0.642 (0.012)	0.950 (0.005)	0.937 (0.006)	0.947 (0.005)	0.652 (0.012)	0.947 (0.006)	0.937 (0.006)	0.943 (0.006)
rough truth	Gaussian	indep.	100	high	rough	61	default	0.957 (0.004)	0.955 (0.004)	0.948 (0.004)	0.952 (0.004)	0.954 (0.004)	0.945 (0.004)	0.938 (0.005)	0.941 (0.005)
rough truth	Gaussian	smooth	100	high	rough	61	default	0.648 (0.011)	0.950 (0.005)	0.945 (0.005)	0.950 (0.005)	0.657 (0.011)	0.947 (0.005)	0.944 (0.005)	0.947 (0.005)
rough truth	Poisson	indep.	100	default	rough	61	default	0.952 (0.004)	0.952 (0.005)	0.941 (0.005)	0.946 (0.005)	0.952 (0.005)	0.942 (0.005)	0.936 (0.005)	0.941 (0.005)

Table 2: Coverage (MC SE) of $\alpha(t)$ and $\gamma(t)$ per setting; for the smooth-effect settings the first group of columns holds $f(x, t)$. Interval labels as in the previous table. (continued)

block	family	error	G	signal	truth	D	basis	REML model	REML+CL2	NCV+CL2	NCV wid.	REML model	REML+CL2	NCV+CL2	NCV wid.
rough truth	Poisson	smooth	100	default	rough	61	default	0.666 (0.012)	0.950 (0.005)	0.936 (0.006)	0.944 (0.006)	0.670 (0.011)	0.945 (0.005)	0.930 (0.006)	0.938 (0.006)
rough truth	Poisson	indep.	100	high	rough	61	default	0.952 (0.005)	0.950 (0.005)	0.936 (0.006)	0.950 (0.005)	0.955 (0.004)	0.945 (0.004)	0.937 (0.005)	0.945 (0.005)
rough truth	Poisson	smooth	100	high	rough	61	default	0.668 (0.012)	0.948 (0.005)	0.939 (0.006)	0.952 (0.005)	0.678 (0.011)	0.943 (0.005)	0.924 (0.006)	0.937 (0.005)
rough truth	binary	indep.	100	default	rough	61	default	0.942 (0.006)	0.921 (0.007)	0.920 (0.008)	0.926 (0.008)	0.907 (0.007)	0.877 (0.008)	0.897 (0.008)	0.906 (0.008)
rough truth	binary	smooth	100	default	rough	61	default	0.708 (0.013)	0.920 (0.007)	0.862 (0.012)	0.914 (0.008)	0.720 (0.012)	0.931 (0.006)	0.839 (0.011)	0.904 (0.007)
rough truth	binary	indep.	100	high	rough	61	default	0.957 (0.004)	0.947 (0.005)	0.939 (0.005)	0.942 (0.005)	0.928 (0.006)	0.916 (0.006)	0.918 (0.006)	0.923 (0.006)
rough truth	binary	smooth	100	high	rough	61	default	0.729 (0.011)	0.932 (0.006)	0.892 (0.008)	0.935 (0.006)	0.745 (0.011)	0.938 (0.006)	0.902 (0.008)	0.927 (0.006)
smooth effect	Gaussian	indep.	100	default	smooth	61	default	0.976 (0.001)	0.971 (0.002)	0.975 (0.001)	0.983 (0.001)	0.962 (0.004)	0.952 (0.004)	0.946 (0.005)	0.951 (0.005)
smooth effect	Gaussian	rough	100	default	smooth	61	default	0.677 (0.005)	0.945 (0.002)	0.940 (0.003)	0.975 (0.002)	0.661 (0.012)	0.938 (0.006)	0.925 (0.007)	0.935 (0.006)
smooth effect	Gaussian	smooth	100	default	smooth	61	default	0.650 (0.005)	0.942 (0.003)	0.939 (0.003)	0.973 (0.002)	0.648 (0.012)	0.950 (0.005)	0.935 (0.007)	0.947 (0.006)
smooth effect	Poisson	indep.	100	default	smooth	61	default	0.978 (0.001)	0.973 (0.001)	0.967 (0.002)	0.976 (0.001)	0.959 (0.004)	0.949 (0.005)	0.942 (0.005)	0.946 (0.005)
smooth effect	Poisson	rough	100	default	smooth	61	default	0.697 (0.005)	0.945 (0.002)	0.920 (0.004)	0.964 (0.002)	0.673 (0.011)	0.938 (0.006)	0.920 (0.007)	0.932 (0.006)
smooth effect	Poisson	smooth	100	default	smooth	61	default	0.671 (0.005)	0.943 (0.003)	0.922 (0.004)	0.965 (0.002)	0.664 (0.012)	0.946 (0.005)	0.931 (0.006)	0.942 (0.006)
smooth effect	binary	indep.	100	default	smooth	61	default	0.988 (0.001)	0.982 (0.001)	0.969 (0.002)	0.976 (0.002)	0.952 (0.006)	0.930 (0.007)	0.922 (0.008)	0.929 (0.008)
smooth effect	binary	rough	100	default	smooth	61	default	0.791 (0.005)	0.949 (0.003)	0.885 (0.008)	0.958 (0.004)	0.729 (0.013)	0.925 (0.008)	0.834 (0.014)	0.889 (0.011)
smooth effect	binary	smooth	100	default	smooth	61	default	0.754 (0.006)	0.945 (0.003)	0.895 (0.008)	0.966 (0.003)	0.723 (0.013)	0.943 (0.006)	0.850 (0.012)	0.910 (0.009)
families	scaled t	indep.	100	default	smooth	61	default	0.955 (0.004)	0.954 (0.004)	0.947 (0.004)	0.950 (0.004)	0.956 (0.004)	0.950 (0.004)	0.946 (0.005)	0.950 (0.004)
families	scaled t	smooth	100	default	smooth	61	default	0.651 (0.012)	0.951 (0.005)	0.947 (0.005)	0.955 (0.005)	0.668 (0.012)	0.950 (0.005)	0.944 (0.006)	0.952 (0.005)
families	beta	indep.	100	default	smooth	61	default	0.958 (0.004)	0.950 (0.004)	0.940 (0.005)	0.946 (0.005)	0.959 (0.004)	0.950 (0.004)	0.939 (0.005)	0.946 (0.004)
families	beta	smooth	100	default	smooth	61	default	0.642 (0.012)	0.945 (0.005)	0.935 (0.006)	0.948 (0.005)	0.656 (0.012)	0.946 (0.005)	0.931 (0.006)	0.943 (0.006)
families	negative binomial	indep.	100	default	smooth	61	default	0.954 (0.004)	0.950 (0.005)	0.935 (0.005)	0.943 (0.005)	0.965 (0.004)	0.957 (0.005)	0.948 (0.005)	0.953 (0.005)
families	negative binomial	smooth	100	default	smooth	61	default	0.647 (0.012)	0.942 (0.006)	0.924 (0.007)	0.943 (0.006)	0.658 (0.012)	0.943 (0.005)	0.926 (0.007)	0.940 (0.006)
families	negative binomial	indep.	100	default	smooth	61	default	0.896 (0.007)	0.950 (0.005)	0.923 (0.006)	0.941 (0.005)	0.881 (0.007)	0.948 (0.005)	0.927 (0.006)	0.939 (0.005)
families	negative binomial	smooth	100	default	smooth	61	default	0.554 (0.011)	0.942 (0.006)	0.915 (0.007)	0.948 (0.006)	0.541 (0.012)	0.942 (0.005)	0.909 (0.008)	0.935 (0.006)
exact AR(1)	Gaussian	AR(1)	100	default	smooth	61	default	0.636 (0.011)	0.955 (0.005)	0.939 (0.005)	0.949 (0.005)	0.645 (0.012)	0.937 (0.006)	0.921 (0.006)	0.933 (0.006)
sign-changing	Gaussian	sign-chg.	100	default	smooth	61	default	0.633 (0.011)	0.951 (0.005)	0.905 (0.007)	0.941 (0.006)	0.645 (0.012)	0.944 (0.005)	0.902 (0.007)	0.935 (0.006)
sign-changing	Poisson	sign-chg.	100	default	smooth	61	default	0.656 (0.011)	0.947 (0.005)	0.885 (0.009)	0.937 (0.006)	0.673 (0.011)	0.939 (0.006)	0.882 (0.008)	0.918 (0.007)
sign-changing	binary	sign-chg.	100	default	smooth	61	default	0.768 (0.010)	0.934 (0.006)	0.844 (0.012)	0.918 (0.007)	0.790 (0.010)	0.937 (0.006)	0.810 (0.012)	0.904 (0.008)
misreg. (AR(1))	Gaussian	misreg.	100	high	rough	61	default	0.732 (0.011)	0.949 (0.006)	0.943 (0.007)	0.950 (0.006)	0.692 (0.013)	0.928 (0.006)	0.923 (0.006)	0.930 (0.006)
heteroskedastic	Gaussian	var(t)	40	default	smooth	61	default	0.630 (0.013)	0.939 (0.006)	0.922 (0.008)	0.937 (0.007)	0.632 (0.012)	0.944 (0.006)	0.925 (0.007)	0.945 (0.006)
heteroskedastic	Gaussian	var(t)	100	default	smooth	61	default	0.650 (0.011)	0.951 (0.005)	0.943 (0.005)	0.952 (0.005)	0.661 (0.012)	0.947 (0.005)	0.932 (0.006)	0.943 (0.006)
heteroskedastic	Gaussian	var(z)	40	default	smooth	61	default	0.618 (0.014)	0.930 (0.007)	0.913 (0.008)	0.935 (0.007)	0.490 (0.014)	0.926 (0.007)	0.898 (0.009)	0.931 (0.007)
heteroskedastic	Gaussian	var(z)	100	default	smooth	61	default	0.647 (0.012)	0.945 (0.005)	0.936 (0.006)	0.946 (0.005)	0.494 (0.013)	0.938 (0.006)	0.920 (0.007)	0.945 (0.005)
heteroskedastic	Gaussian	var(t,z)	40	default	smooth	61	default	0.629 (0.014)	0.935 (0.006)	0.917 (0.007)	0.936 (0.007)	0.502 (0.014)	0.925 (0.007)	0.893 (0.009)	0.930 (0.007)
heteroskedastic	Gaussian	var(t,z)	100	default	smooth	61	default	0.655 (0.012)	0.947 (0.005)	0.939 (0.006)	0.949 (0.005)	0.510 (0.013)	0.940 (0.005)	0.920 (0.006)	0.944 (0.006)
low-rank covariate	Gaussian	indep.	100	default	smooth	61	default	0.956 (0.004)	0.956 (0.004)	0.945 (0.005)	0.950 (0.005)	0.960 (0.004)	0.950 (0.005)	0.946 (0.005)	0.951 (0.004)
low-rank covariate	Gaussian	smooth	100	default	smooth	61	default	0.642 (0.011)	0.949 (0.005)	0.939 (0.006)	0.950 (0.005)	0.655 (0.012)	0.949 (0.005)	0.934 (0.006)	0.946 (0.006)
low-rank covariate	Poisson	indep.	100	default	smooth	61	default	0.952 (0.005)	0.952 (0.005)	0.937 (0.005)	0.942 (0.005)	0.963 (0.004)	0.955 (0.005)	0.948 (0.005)	0.950 (0.005)
low-rank covariate	Poisson	smooth	100	default	smooth	61	default	0.667 (0.012)	0.947 (0.006)	0.928 (0.007)	0.948 (0.006)	0.663 (0.012)	0.945 (0.005)	0.932 (0.006)	0.944 (0.006)
low-rank covariate	binary	indep.	100	default	smooth	61	default	0.942 (0.006)	0.923 (0.007)	0.917 (0.007)	0.921 (0.007)	0.944 (0.007)	0.922 (0.008)	0.910 (0.009)	0.917 (0.009)
low-rank covariate	binary	smooth	100	default	smooth	61	default	0.702 (0.013)	0.922 (0.007)	0.868 (0.011)	0.913 (0.008)	0.731 (0.013)	0.940 (0.007)	0.867 (0.011)	0.915 (0.008)
basis size	Gaussian	indep.	100	default	smooth	61	2× finer knots	0.968 (0.003)	0.965 (0.003)	0.955 (0.004)	0.961 (0.004)	0.971 (0.003)	0.963 (0.003)	0.952 (0.004)	0.960 (0.004)
basis size	Gaussian	indep.	100	default	smooth	61	4× finer knots	0.978 (0.002)	0.975 (0.002)	0.966 (0.003)	0.971 (0.003)	0.976 (0.002)	0.973 (0.003)	0.960 (0.003)	0.968 (0.003)
basis size	Gaussian	smooth	100	default	smooth	61	2× finer knots	0.697 (0.010)	0.950 (0.005)	0.941 (0.005)	0.956 (0.005)	0.703 (0.011)	0.950 (0.005)	0.932 (0.006)	0.949 (0.005)
basis size	Gaussian	smooth	100	default	smooth	61	4× finer knots	0.739 (0.010)	0.954 (0.004)	0.944 (0.005)	0.960 (0.004)	0.734 (0.011)	0.950 (0.005)	0.934 (0.006)	0.955 (0.005)
basis size	Poisson	indep.	100	default	smooth	61	2× finer knots	0.968 (0.003)	0.964 (0.004)	0.946 (0.005)	0.955 (0.004)	0.971 (0.003)	0.964 (0.003)	0.951 (0.004)	0.955 (0.004)
basis size	Poisson	indep.	100	default	smooth	61	4× finer knots	0.978 (0.002)	0.975 (0.003)	0.958 (0.004)	0.968 (0.003)	0.978 (0.002)	0.973 (0.003)	0.959 (0.004)	0.964 (0.003)
basis size	Poisson	smooth	100	default	smooth	61	2× finer knots	0.716 (0.010)	0.949 (0.005)	0.927 (0.007)	0.956 (0.005)	0.723 (0.010)	0.950 (0.005)	0.928 (0.006)	0.943 (0.005)
basis size	Poisson	smooth	100	default	smooth	61	4× finer knots	0.752 (0.010)	0.950 (0.005)	0.931 (0.007)	0.963 (0.005)	0.769 (0.010)	0.953 (0.005)	0.929 (0.006)	0.947 (0.005)
basis size	binary	indep.	100	default	smooth	61	2× finer knots	0.961 (0.005)	0.944 (0.006)	0.933 (0.006)	0.938 (0.006)	0.958 (0.006)	0.936 (0.007)	0.921 (0.008)	0.926 (0.008)
basis size	binary	indep.	100	default	smooth	61	4× finer knots	0.969 (0.004)	0.956 (0.005)	0.937 (0.007)	0.945 (0.006)	0.966 (0.005)	0.949 (0.006)	0.928 (0.008)	0.934 (0.007)
basis size	binary	smooth	100	default	smooth	61	2× finer knots	0.728 (0.013)	0.915 (0.007)	0.866 (0.011)	0.921 (0.008)	0.762 (0.012)	0.942 (0.006)	0.865 (0.011)	0.923 (0.009)
basis size	binary	smooth	100	default	smooth	61	4× finer knots	0.713 (0.014)	0.899 (0.008)	0.877 (0.011)	0.936 (0.006)	0.779 (0.013)	0.946 (0.006)	0.872 (0.011)	0.929 (0.008)

S1.2 Lower tails, number of curves, signal and truth roughness

In the dependent main-factorial settings, the worst setting's 5% quantile of pointwise coverage for the mean and β is 0.91 for REML + CL2 at $G = 100$ and 0.89 at $G = 40$; for model-based intervals it is as low as 0.54. With 40 instead of 100 curves REML + CL2 loses 1.1–1.8 pp of coverage (means per family and estimand; coverage by G in the first table under *Twenty curves*), NCV + CL2 1.0–6.9 pp. Signal strength changes little. Truth roughness matters for NCV + CL2: with the rough truth and smooth errors it covers β at 0.80–0.90 (smooth truth 0.90–0.94), REML + CL2 at 0.94–0.95.

Table 3: 5% quantile of pointwise coverage (minimum and median over main-factorial settings) and minimum grid-averaged coverage, by dependence, G , estimand and interval.

dependence	G	estimand	interval	q05 min	q05 median	min coverage
dependent	40	$\beta(s,t)$	NCV + CL2	0.600	0.861	0.785
dependent	40	$\beta(s,t)$	NCV + CL2, widened by NCV–REML difference	0.885	0.950	0.948
dependent	40	$\beta(s,t)$	REML + CL2	0.891	0.897	0.923
dependent	40	$\beta(s,t)$	REML, model-based	0.520	0.580	0.609
dependent	40	$E(Y X)$	NCV + CL2	0.685	0.827	0.842
dependent	40	$E(Y X)$	NCV + CL2, widened by NCV–REML difference	0.860	0.907	0.930
dependent	40	$E(Y X)$	REML + CL2	0.887	0.900	0.921
dependent	40	$E(Y X)$	REML, model-based	0.515	0.560	0.601
dependent	100	$\beta(s,t)$	NCV + CL2	0.775	0.861	0.896
dependent	100	$\beta(s,t)$	NCV + CL2, widened by NCV–REML difference	0.920	0.950	0.971
dependent	100	$\beta(s,t)$	REML + CL2	0.911	0.915	0.940
dependent	100	$\beta(s,t)$	REML, model-based	0.540	0.596	0.653
dependent	100	$E(Y X)$	NCV + CL2	0.740	0.863	0.876
dependent	100	$E(Y X)$	NCV + CL2, widened by NCV–REML difference	0.865	0.920	0.933
dependent	100	$E(Y X)$	REML + CL2	0.910	0.915	0.940
dependent	100	$E(Y X)$	REML, model-based	0.535	0.570	0.631
independent	40	$\beta(s,t)$	NCV + CL2	0.925	0.945	0.973
independent	40	$\beta(s,t)$	NCV + CL2, widened by NCV–REML difference	0.930	0.960	0.979
independent	40	$\beta(s,t)$	REML + CL2	0.915	0.965	0.973
independent	40	$\beta(s,t)$	REML, model-based	0.945	0.975	0.984
independent	40	$E(Y X)$	NCV + CL2	0.835	0.900	0.923
independent	40	$E(Y X)$	NCV + CL2, widened by NCV–REML difference	0.850	0.905	0.935
independent	40	$E(Y X)$	REML + CL2	0.825	0.925	0.925
independent	40	$E(Y X)$	REML, model-based	0.875	0.940	0.952
independent	100	$\beta(s,t)$	NCV + CL2	0.925	0.940	0.976
independent	100	$\beta(s,t)$	NCV + CL2, widened by NCV–REML difference	0.945	0.951	0.986
independent	100	$\beta(s,t)$	REML + CL2	0.955	0.965	0.979
independent	100	$\beta(s,t)$	REML, model-based	0.965	0.970	0.983
independent	100	$E(Y X)$	NCV + CL2	0.860	0.900	0.939
independent	100	$E(Y X)$	NCV + CL2, widened by NCV–REML difference	0.870	0.910	0.945
independent	100	$E(Y X)$	REML + CL2	0.885	0.930	0.949
independent	100	$E(Y X)$	REML, model-based	0.910	0.935	0.962

Twenty curves. The main factorial was also run with $G = 20$, the first 20 curves of each replicate, so the three numbers of curves share their draws (first table below; means over each family's settings for coverage, medians for relative error and detection). Under dependent errors REML + CL2 covers the mean, β and γ at 0.89–0.92 with 20 curves, 0.92–0.94 with 40 and 0.94–0.95 with 100, and $\alpha(t)$ at 0.85–0.91 with 20. NCV + CL2 falls to 0.78–0.90 with 20 curves, the model-based REML intervals to 0.58–0.77. Under independent errors every interval undercovers binary $\alpha(t)$ and $\gamma(t)$ with 20 curves, the model-based one included (0.88–0.93). REML's median relative error for β under dependence is 0.86–2.43 with 20 curves, NCV's 0.29–0.74; REML + CL2 detects 0.16–0.45 of the clearly non-zero grid points of β , NCV + CL2 0.40–0.84. With the Satterthwaite critical value (second table; default signal, as for 40 and 100 curves), REML + CL2 covers the mean, β and γ at 0.94–0.96 under dependence with 20 curves, against 0.89–0.92 with z and 0.91–0.93 with t_{G-1} , at 1.19–1.33 times the width of the z interval. Binary $\alpha(t)$ stays at 0.897 under dependence and 0.844 under independent errors, and under independent Gaussian and Poisson errors the intervals overcover (0.97–1.00).

Table 4: Main factorial by number of curves $G = 20, 40, 100$ (all signal levels; dependent: rough stationary and smooth errors): grid-averaged coverage (cov; mean over settings), median relative error of the estimate (rel; median over settings of the median over replicates; the model-based and CL2 intervals of a fit share the estimate) and detection rate (det; median over settings; $\beta(s, t)$ and $\gamma(t)$ only), per family, dependence, estimand and interval.

family	errors	estimand	interval	cov 20	cov 40	cov 100	rel 20	rel 40	rel 100	det 20	det 40	det 100
Gaussian	independent	$E(Y X)$	REML, model-based	0.966	0.967	0.966	0.193	0.135	0.091			
Gaussian	independent	$E(Y X)$	REML + CL2	0.949	0.957	0.960	0.193	0.135	0.091			
Gaussian	independent	$E(Y X)$	NCV + CL2	0.941	0.947	0.950	0.193	0.133	0.088			
Gaussian	independent	$E(Y X)$	NCV + CL2, widened by NCV-REML difference	0.949	0.954	0.958	0.193	0.133	0.088			
Gaussian	independent	$\beta(s, t)$	REML, model-based	0.992	0.991	0.988	0.185	0.146	0.109	0.925	0.959	0.982
Gaussian	independent	$\beta(s, t)$	REML + CL2	0.987	0.987	0.985	0.185	0.146	0.109	0.926	0.960	0.983
Gaussian	independent	$\beta(s, t)$	NCV + CL2	0.983	0.982	0.981	0.159	0.116	0.083	0.943	0.971	0.993
Gaussian	independent	$\beta(s, t)$	NCV + CL2, widened by NCV-REML difference	0.988	0.989	0.989	0.159	0.116	0.083	0.939	0.967	0.990
Gaussian	independent	$\alpha(t)$	REML, model-based	0.952	0.955	0.956	0.256	0.181	0.116			
Gaussian	independent	$\alpha(t)$	REML + CL2	0.934	0.945	0.953	0.256	0.181	0.116			
Gaussian	independent	$\alpha(t)$	NCV + CL2	0.932	0.939	0.945	0.259	0.182	0.118			
Gaussian	independent	$\alpha(t)$	NCV + CL2, widened by NCV-REML difference	0.937	0.942	0.949	0.259	0.182	0.118			
Gaussian	independent	$\gamma(t)$	REML, model-based	0.956	0.959	0.961	0.189	0.132	0.089	0.966	0.998	1.000
Gaussian	independent	$\gamma(t)$	REML + CL2	0.932	0.946	0.950	0.189	0.132	0.089	0.962	0.998	1.000
Gaussian	independent	$\gamma(t)$	NCV + CL2	0.927	0.937	0.944	0.191	0.133	0.090	0.964	0.997	1.000
Gaussian	independent	$\gamma(t)$	NCV + CL2, widened by NCV-REML difference	0.932	0.941	0.950	0.191	0.133	0.090	0.963	0.997	1.000
Gaussian	dependent	$E(Y X)$	REML, model-based	0.576	0.615	0.640	0.570	0.358	0.215			
Gaussian	dependent	$E(Y X)$	REML + CL2	0.909	0.931	0.943	0.570	0.358	0.215			
Gaussian	dependent	$E(Y X)$	NCV + CL2	0.875	0.912	0.928	0.402	0.270	0.178			
Gaussian	dependent	$E(Y X)$	NCV + CL2, widened by NCV-REML difference	0.943	0.954	0.957	0.402	0.270	0.178			
Gaussian	dependent	$\beta(s, t)$	REML, model-based	0.597	0.625	0.663	1.188	0.680	0.353	0.636	0.779	0.916
Gaussian	dependent	$\beta(s, t)$	REML + CL2	0.912	0.926	0.942	1.188	0.680	0.353	0.325	0.525	0.788
Gaussian	dependent	$\beta(s, t)$	NCV + CL2	0.882	0.918	0.932	0.301	0.216	0.152	0.842	0.922	0.973
Gaussian	dependent	$\beta(s, t)$	NCV + CL2, widened by NCV-REML difference	0.975	0.982	0.983	0.301	0.216	0.152	0.412	0.623	0.845
Gaussian	dependent	$\alpha(t)$	REML, model-based	0.580	0.623	0.650	0.585	0.385	0.232			
Gaussian	dependent	$\alpha(t)$	REML + CL2	0.904	0.937	0.950	0.585	0.385	0.232			
Gaussian	dependent	$\alpha(t)$	NCV + CL2	0.864	0.904	0.937	0.542	0.361	0.227			
Gaussian	dependent	$\alpha(t)$	NCV + CL2, widened by NCV-REML difference	0.920	0.931	0.946	0.542	0.361	0.227			
Gaussian	dependent	$\gamma(t)$	REML, model-based	0.589	0.637	0.657	0.435	0.284	0.177	0.879	0.968	0.998
Gaussian	dependent	$\gamma(t)$	REML + CL2	0.917	0.942	0.943	0.435	0.284	0.177	0.600	0.849	0.975
Gaussian	dependent	$\gamma(t)$	NCV + CL2	0.872	0.912	0.925	0.380	0.274	0.172	0.688	0.888	0.983
Gaussian	dependent	$\gamma(t)$	NCV + CL2, widened by NCV-REML difference	0.927	0.937	0.938	0.380	0.274	0.172	0.597	0.870	0.981
Poisson	independent	$E(Y X)$	REML, model-based	0.968	0.967	0.964	0.156	0.103	0.067			
Poisson	independent	$E(Y X)$	REML + CL2	0.950	0.956	0.958	0.156	0.103	0.067			
Poisson	independent	$E(Y X)$	NCV + CL2	0.934	0.943	0.945	0.160	0.103	0.065			
Poisson	independent	$E(Y X)$	NCV + CL2, widened by NCV-REML difference	0.945	0.951	0.955	0.160	0.103	0.065			
Poisson	independent	$\beta(s, t)$	REML, model-based	0.993	0.992	0.987	0.189	0.145	0.108	0.919	0.955	0.979
Poisson	independent	$\beta(s, t)$	REML + CL2	0.989	0.988	0.982	0.189	0.145	0.108	0.921	0.956	0.980
Poisson	independent	$\beta(s, t)$	NCV + CL2	0.982	0.983	0.979	0.162	0.118	0.082	0.936	0.968	0.990
Poisson	independent	$\beta(s, t)$	NCV + CL2, widened by NCV-REML difference	0.989	0.989	0.988	0.162	0.118	0.082	0.930	0.964	0.986
Poisson	independent	$\alpha(t)$	REML, model-based	0.956	0.961	0.952	0.270	0.189	0.123			
Poisson	independent	$\alpha(t)$	REML + CL2	0.939	0.952	0.950	0.270	0.189	0.123			
Poisson	independent	$\alpha(t)$	NCV + CL2	0.920	0.933	0.931	0.281	0.194	0.126			
Poisson	independent	$\alpha(t)$	NCV + CL2, widened by NCV-REML difference	0.935	0.946	0.942	0.281	0.194	0.126			
Poisson	independent	$\gamma(t)$	REML, model-based	0.955	0.955	0.959	0.200	0.134	0.085	0.962	0.995	1.000
Poisson	independent	$\gamma(t)$	REML + CL2	0.931	0.945	0.949	0.200	0.134	0.085	0.959	0.994	1.000
Poisson	independent	$\gamma(t)$	NCV + CL2	0.918	0.933	0.942	0.206	0.137	0.088	0.961	0.994	1.000
Poisson	independent	$\gamma(t)$	NCV + CL2, widened by NCV-REML difference	0.928	0.940	0.947	0.206	0.137	0.088	0.959	0.994	1.000
Poisson	dependent	$E(Y X)$	REML, model-based	0.650	0.657	0.663	0.470	0.280	0.160			
Poisson	dependent	$E(Y X)$	REML + CL2	0.907	0.932	0.943	0.470	0.280	0.160			
Poisson	dependent	$E(Y X)$	NCV + CL2	0.874	0.908	0.925	0.351	0.217	0.130			
Poisson	dependent	$E(Y X)$	NCV + CL2, widened by NCV-REML difference	0.934	0.950	0.955	0.351	0.217	0.130			
Poisson	dependent	$\beta(s, t)$	REML, model-based	0.685	0.682	0.695	0.865	0.571	0.322	0.676	0.817	0.930
Poisson	dependent	$\beta(s, t)$	REML + CL2	0.912	0.928	0.943	0.865	0.571	0.322	0.451	0.631	0.837
Poisson	dependent	$\beta(s, t)$	NCV + CL2	0.903	0.923	0.933	0.289	0.206	0.141	0.836	0.922	0.972
Poisson	dependent	$\beta(s, t)$	NCV + CL2, widened by NCV-REML difference	0.969	0.980	0.982	0.289	0.206	0.141	0.518	0.707	0.879
Poisson	dependent	$\alpha(t)$	REML, model-based	0.645	0.656	0.665	0.599	0.394	0.243			
Poisson	dependent	$\alpha(t)$	REML + CL2	0.906	0.938	0.949	0.599	0.394	0.243			
Poisson	dependent	$\alpha(t)$	NCV + CL2	0.864	0.903	0.924	0.549	0.375	0.236			
Poisson	dependent	$\alpha(t)$	NCV + CL2, widened by NCV-REML difference	0.923	0.943	0.948	0.549	0.375	0.236			
Poisson	dependent	$\gamma(t)$	REML, model-based	0.662	0.660	0.674	0.425	0.277	0.167	0.873	0.969	0.996
Poisson	dependent	$\gamma(t)$	REML + CL2	0.914	0.937	0.941	0.425	0.277	0.167	0.643	0.866	0.976
Poisson	dependent	$\gamma(t)$	NCV + CL2	0.862	0.904	0.924	0.397	0.277	0.166	0.704	0.899	0.983
Poisson	dependent	$\gamma(t)$	NCV + CL2, widened by NCV-REML difference	0.919	0.933	0.936	0.397	0.277	0.166	0.613	0.877	0.982
binary	independent	$E(Y X)$	REML, model-based	0.939	0.958	0.965	0.539	0.386	0.258			
binary	independent	$E(Y X)$	REML + CL2	0.903	0.937	0.954	0.539	0.386	0.258			
binary	independent	$E(Y X)$	NCV + CL2	0.907	0.930	0.941	0.566	0.393	0.260			
binary	independent	$E(Y X)$	NCV + CL2, widened by NCV-REML difference	0.926	0.940	0.948	0.566	0.393	0.260			
binary	independent	$\beta(s, t)$	REML, model-based	0.957	0.987	0.993	0.449	0.340	0.246	0.517	0.711	0.850
binary	independent	$\beta(s, t)$	REML + CL2	0.935	0.979	0.990	0.449	0.340	0.246	0.595	0.744	0.862
binary	independent	$\beta(s, t)$	NCV + CL2	0.947	0.975	0.981	0.440	0.317	0.229	0.585	0.760	0.886
binary	independent	$\beta(s, t)$	NCV + CL2, widened by NCV-REML difference	0.959	0.982	0.987	0.440	0.317	0.229	0.548	0.747	0.879
binary	independent	$\alpha(t)$	REML, model-based	0.885	0.930	0.950	0.805	0.592	0.389			
binary	independent	$\alpha(t)$	REML + CL2	0.847	0.899	0.936	0.805	0.592	0.389			
binary	independent	$\alpha(t)$	NCV + CL2	0.850	0.896	0.922	0.829	0.597	0.405			
binary	independent	$\alpha(t)$	NCV + CL2, widened by NCV-REML difference	0.872	0.909	0.929	0.829	0.597	0.405			
binary	independent	$\gamma(t)$	REML, model-based	0.928	0.937	0.946	0.598	0.433	0.300	0.313	0.582	0.851
binary	independent	$\gamma(t)$	REML + CL2	0.875	0.906	0.929	0.598	0.433	0.300	0.393	0.633	0.866

Table 4: Main factorial by number of curves $G = 20, 40, 100$ (all signal levels; dependent: rough stationary and smooth errors): grid-averaged coverage (cov; mean over settings), median relative error of the estimate (rel; median over settings of the median over replicates; the model-based and CL2 intervals of a fit share the estimate) and detection rate (det; median over settings; $\beta(s, t)$ and $\gamma(t)$ only), per family, dependence, estimand and interval. (*continued*)

family	errors	estimand	interval	cov 20	cov 40	cov 100	rel 20	rel 40	rel 100	det 20	det 40	det 100
binary	independent	$\gamma(t)$	NCV + CL2	0.874	0.893	0.917	0.632	0.461	0.304	0.399	0.629	0.862
binary	independent	$\gamma(t)$	NCV + CL2, widened by NCV-REML difference	0.893	0.907	0.922	0.632	0.461	0.304	0.363	0.613	0.858
binary	dependent	$E(Y X)$	REML, model-based	0.722	0.731	0.742	1.352	0.868	0.530			
binary	dependent	$E(Y X)$	REML + CL2	0.895	0.925	0.942	1.352	0.868	0.530			
binary	dependent	$E(Y X)$	NCV + CL2	0.831	0.861	0.895	0.971	0.704	0.468			
binary	dependent	$E(Y X)$	NCV + CL2, widened by NCV-REML difference	0.936	0.937	0.942	0.971	0.704	0.468			
binary	dependent	$\beta(s, t)$	REML, model-based	0.774	0.782	0.800	2.428	1.479	0.853	0.326	0.450	0.627
binary	dependent	$\beta(s, t)$	REML + CL2	0.912	0.934	0.951	2.428	1.479	0.853	0.161	0.237	0.412
binary	dependent	$\beta(s, t)$	NCV + CL2	0.783	0.844	0.913	0.743	0.577	0.359	0.399	0.593	0.779
binary	dependent	$\beta(s, t)$	NCV + CL2, widened by NCV-REML difference	0.954	0.962	0.976	0.743	0.577	0.359	0.097	0.232	0.470
binary	dependent	$\alpha(t)$	REML, model-based	0.680	0.703	0.735	2.001	1.182	0.734			
binary	dependent	$\alpha(t)$	REML + CL2	0.852	0.900	0.928	2.001	1.182	0.734			
binary	dependent	$\alpha(t)$	NCV + CL2	0.832	0.853	0.873	1.340	1.001	0.740			
binary	dependent	$\alpha(t)$	NCV + CL2, widened by NCV-REML difference	0.932	0.920	0.919	1.340	1.001	0.740			
binary	dependent	$\gamma(t)$	REML, model-based	0.721	0.719	0.741	1.340	0.845	0.521	0.389	0.547	0.769
binary	dependent	$\gamma(t)$	REML + CL2	0.895	0.922	0.936	1.340	0.845	0.521	0.182	0.278	0.514
binary	dependent	$\gamma(t)$	NCV + CL2	0.815	0.822	0.867	1.044	0.811	0.518	0.167	0.286	0.539
binary	dependent	$\gamma(t)$	NCV + CL2, widened by NCV-REML difference	0.913	0.899	0.911	1.044	0.811	0.518	0.069	0.177	0.466

Table 5: REML + CL2 with the normal (z), t_{G-1} and Satterthwaite critical values by number of curves, main factorial at the default signal: grid-averaged coverage (mean over the settings of each family and dependence class), with the 5% quantile of pointwise coverage (q05) and the width relative to z of the Satterthwaite interval with 20 curves.

family	errors	estimand	$z, 20$	$t(G-1), 20$	Satt., 20	q05 Satt., 20	width Satt./ $z, 20$	$z, 40$	Satt., 40	$z, 100$	Satt., 100
Gaussian	independent	$E(Y X)$	0.952	0.964	0.982	0.965	1.24	0.958	0.973	0.961	0.967
Gaussian	independent	$\beta(s, t)$	0.989	0.993	0.998	0.990	1.24	0.989	0.995	0.986	0.989
Gaussian	independent	$\alpha(t)$	0.939	0.952	0.963	0.915	1.13	0.947	0.956	0.956	0.959
Gaussian	independent	$\gamma(t)$	0.937	0.950	0.974	0.955	1.25	0.951	0.969	0.950	0.959
Gaussian	dependent	$E(Y X)$	0.909	0.927	0.958	0.935	1.25	0.931	0.953	0.943	0.951
Gaussian	dependent	$\beta(s, t)$	0.913	0.930	0.961	0.938	1.26	0.926	0.950	0.942	0.951
Gaussian	dependent	$\alpha(t)$	0.905	0.922	0.945	0.911	1.17	0.938	0.952	0.951	0.953
Gaussian	dependent	$\gamma(t)$	0.918	0.934	0.964	0.948	1.25	0.942	0.961	0.943	0.953
Poisson	independent	$E(Y X)$	0.949	0.961	0.981	0.960	1.27	0.958	0.974	0.960	0.967
Poisson	independent	$\beta(s, t)$	0.990	0.993	0.998	0.990	1.24	0.989	0.995	0.986	0.989
Poisson	independent	$\alpha(t)$	0.937	0.953	0.963	0.912	1.13	0.953	0.961	0.951	0.954
Poisson	independent	$\gamma(t)$	0.930	0.943	0.972	0.953	1.26	0.948	0.967	0.955	0.965
Poisson	dependent	$E(Y X)$	0.908	0.925	0.956	0.930	1.33	0.932	0.954	0.944	0.953
Poisson	dependent	$\beta(s, t)$	0.912	0.929	0.960	0.935	1.25	0.928	0.953	0.943	0.953
Poisson	dependent	$\alpha(t)$	0.904	0.923	0.941	0.891	1.15	0.934	0.945	0.949	0.952
Poisson	dependent	$\gamma(t)$	0.915	0.931	0.962	0.946	1.26	0.940	0.963	0.941	0.951
binary	independent	$E(Y X)$	0.879	0.898	0.927	0.810	1.21	0.925	0.946	0.949	0.957
binary	independent	$\beta(s, t)$	0.898	0.911	0.933	0.800	1.23	0.973	0.982	0.990	0.993
binary	independent	$\alpha(t)$	0.804	0.825	0.844	0.613	1.11	0.859	0.872	0.924	0.927
binary	independent	$\gamma(t)$	0.850	0.872	0.914	0.833	1.25	0.890	0.918	0.923	0.932
binary	dependent	$E(Y X)$	0.895	0.915	0.945	0.913	1.19	0.924	0.946	0.940	0.949
binary	dependent	$\beta(s, t)$	0.910	0.928	0.958	0.935	1.25	0.932	0.954	0.951	0.960
binary	dependent	$\alpha(t)$	0.856	0.879	0.897	0.821	1.14	0.894	0.909	0.923	0.928
binary	dependent	$\gamma(t)$	0.895	0.914	0.950	0.919	1.26	0.921	0.945	0.931	0.941

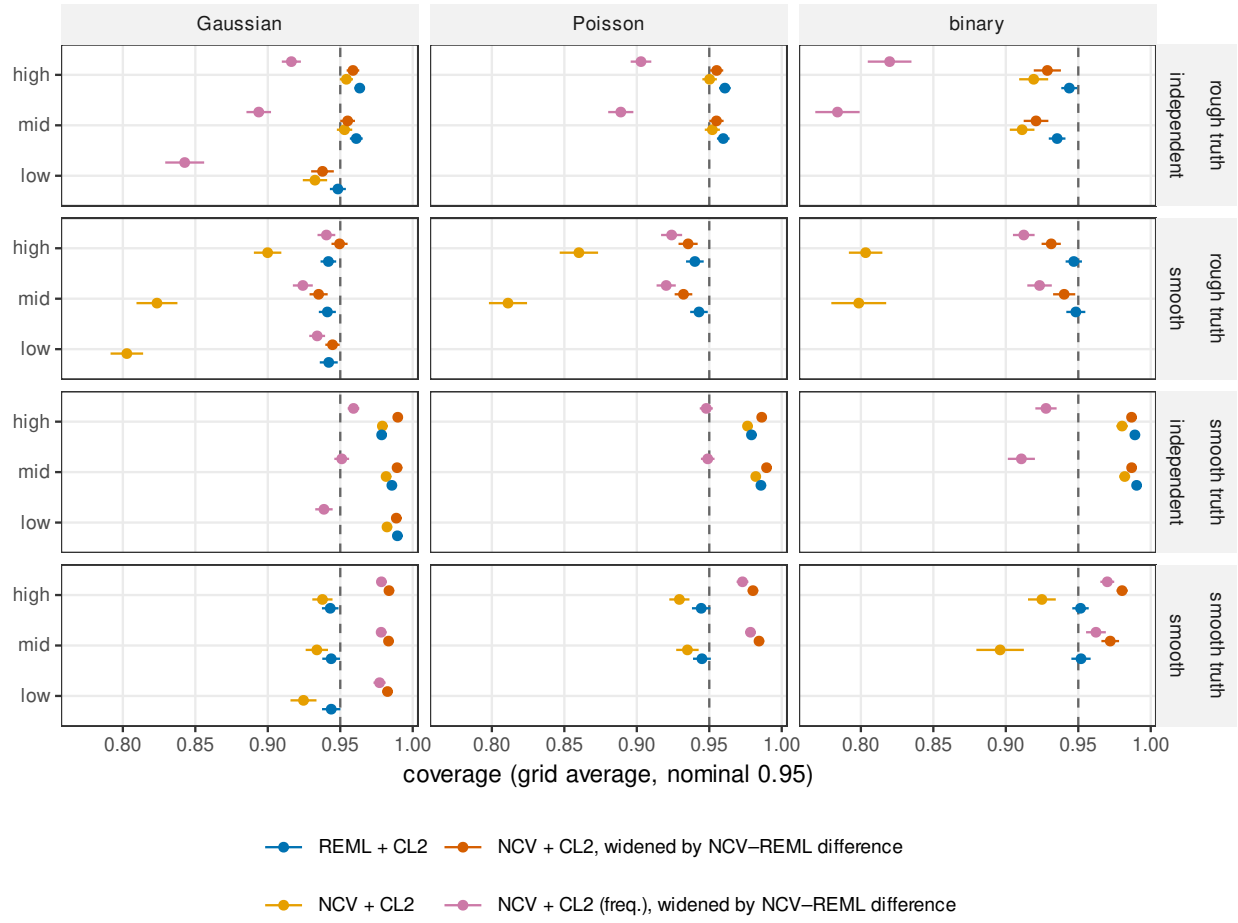


Figure 1: Coverage of $\beta(s, t)$ by signal level and truth roughness, $G = 100$, smooth and independent errors. Bars: ± 2 MC SE; dashed: nominal 0.95.

S1.3 Error structures and extended families

From $D = 61$ to $D = 241$ grid points per curve, the coverage of β by model-based intervals drops by 30.7–36.9 pp under dependent errors, while REML + CL2 changes by -1.3 to 0.0 pp. A variance profile along t changes no interval. Variance that depends on the scalar covariate breaks the model-based intervals for $\gamma(t)$ (0.49–0.51), not REML + CL2 (0.93–0.94). Under misregistration REML + CL2 covers the Gaussian estimands at 0.93–0.94; for misregistered Poisson responses every interval undercovers the mean (REML + CL2 0.90–0.91). Sign-changing correlation: model-based 0.65–0.88, REML + CL2 0.94–0.96. Extended families under smooth errors: model-based 0.63–0.67, REML + CL2 0.94–0.95.

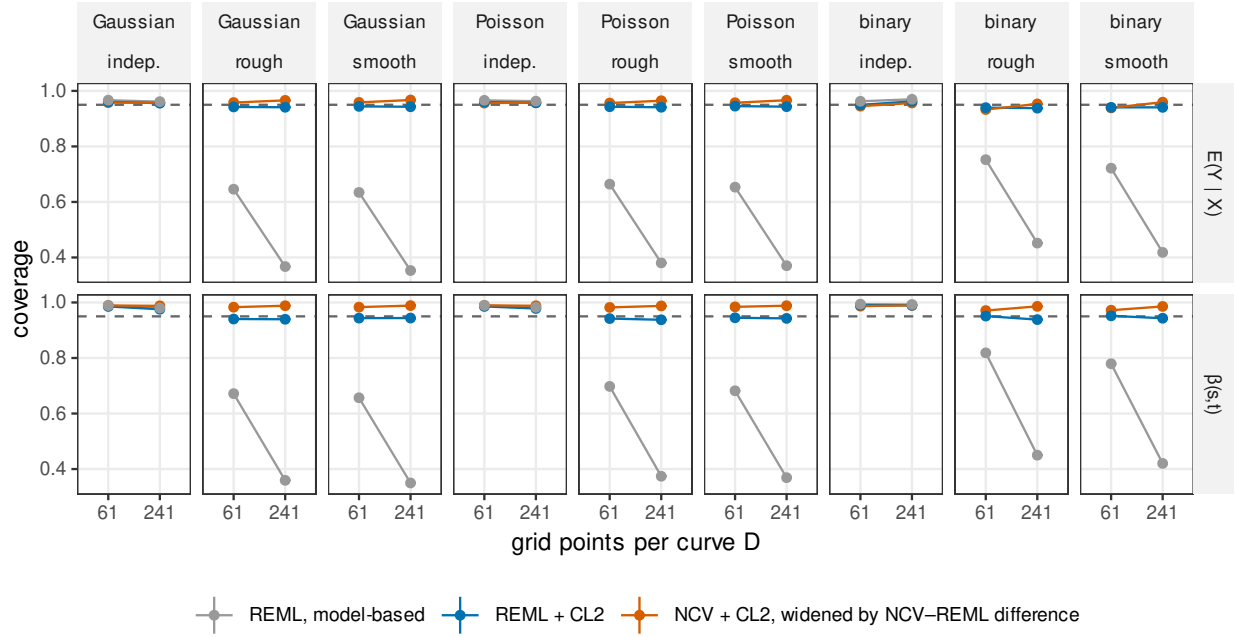


Figure 2: Coverage of $\beta(s, t)$ and the mean at $D = 61$ and $D = 241$ grid points per curve, $G = 100$. Bars: ± 2 MC SE.

Table 6: Heteroskedastic Gaussian errors next to the homoskedastic smooth process, coverage per interval.

error	G	estimand	NCV + CL2	NCV + CL2, widened by NCV-REML difference	REML + CL2	REML, model-based	AR(1) working model
smooth	40	$\beta(s, t)$	0.924	0.983	0.926	0.615	
smooth	40	$\gamma(t)$	0.927	0.946	0.946	0.621	
smooth	40	$E(Y X)$	0.922	0.959	0.931	0.606	
smooth	100	$\beta(s, t)$	0.934	0.983	0.944	0.657	0.993
smooth	100	$\gamma(t)$	0.935	0.946	0.948	0.655	0.967
smooth	100	$E(Y X)$	0.929	0.959	0.944	0.634	0.972
var(z)	40	$\beta(s, t)$	0.925	0.985	0.939	0.619	
var(z)	40	$\gamma(t)$	0.898	0.931	0.926	0.490	
var(z)	40	$E(Y X)$	0.912	0.955	0.934	0.588	
var(z)	100	$\beta(s, t)$	0.935	0.984	0.949	0.660	0.992
var(z)	100	$\gamma(t)$	0.920	0.945	0.938	0.494	0.879
var(z)	100	$E(Y X)$	0.923	0.955	0.944	0.606	0.949
var(t)	40	$\beta(s, t)$	0.918	0.979	0.926	0.619	
var(t)	40	$\gamma(t)$	0.925	0.945	0.944	0.632	
var(t)	40	$E(Y X)$	0.919	0.957	0.932	0.614	
var(t)	100	$\beta(s, t)$	0.933	0.983	0.944	0.662	0.991
var(t)	100	$\gamma(t)$	0.932	0.943	0.947	0.661	0.967
var(t)	100	$E(Y X)$	0.928	0.959	0.944	0.642	0.972
var(t, z)	40	$\beta(s, t)$	0.924	0.983	0.940	0.629	
var(t, z)	40	$\gamma(t)$	0.893	0.930	0.925	0.502	
var(t, z)	40	$E(Y X)$	0.911	0.955	0.936	0.596	
var(t, z)	100	$\beta(s, t)$	0.934	0.984	0.950	0.666	0.990
var(t, z)	100	$\gamma(t)$	0.920	0.944	0.940	0.510	0.889
var(t, z)	100	$E(Y X)$	0.922	0.955	0.944	0.615	0.951

Table 7: Misregistration settings (high signal): coverage of the observed-clock estimands; binary and Poisson settings score the mean only.

family	G	signal	truth	estimand	REML, model-based	REML + CL2	NCV + CL2	NCV + CL2, widened by NCV-REML difference	NCV + CL2 (freq.), widened by NCV-REML difference	AR(1) working model
binary	40	high	smooth	$E(Y X)$	0.945	0.944	0.919	0.932	0.884	
binary	40	high	rough	$E(Y X)$	0.924	0.925	0.892	0.908	0.854	
binary	100	high	smooth	$E(Y X)$	0.943	0.954	0.931	0.941	0.902	
binary	100	high	rough	$E(Y X)$	0.918	0.932	0.909	0.919	0.878	
Gaussian	40	high	smooth	$\beta(s, t)$	0.761	0.949	0.937	0.974	0.956	
Gaussian	40	high	smooth	$E(Y X)$	0.729	0.938	0.919	0.946	0.934	
Gaussian	40	high	rough	$\beta(s, t)$	0.739	0.945	0.820	0.928	0.901	
Gaussian	40	high	rough	$E(Y X)$	0.706	0.934	0.885	0.932	0.918	
Gaussian	100	high	smooth	$\beta(s, t)$	0.753	0.951	0.937	0.976	0.959	0.875
Gaussian	100	high	smooth	$E(Y X)$	0.730	0.942	0.930	0.952	0.941	0.844
Gaussian	100	high	rough	$\beta(s, t)$	0.729	0.944	0.863	0.934	0.908	0.860
Gaussian	100	high	rough	$E(Y X)$	0.706	0.937	0.911	0.938	0.926	0.838
Poisson	40	high	smooth	$E(Y X)$	0.634	0.902	0.792	0.910	0.900	
Poisson	40	high	rough	$E(Y X)$	0.614	0.899	0.773	0.910	0.902	
Poisson	100	high	smooth	$E(Y X)$	0.571	0.910	0.801	0.922	0.915	
Poisson	100	high	rough	$E(Y X)$	0.558	0.906	0.780	0.918	0.912	

Table 8: Sign-changing error correlation, $G = 100$: coverage per interval.

family	error	estimand	AR(1) working model	NCV + CL2	NCV + CL2, widened by NCV-REML difference	REML + CL2	REML, model-based
binary	sign-chg.	$\beta(s, t)$		0.927	0.977	0.959	0.885
binary	sign-chg.	$E(Y X)$		0.882	0.945	0.943	0.794
Gaussian	sign-chg.	$\beta(s, t)$	0.992	0.935	0.985	0.942	0.700
Gaussian	sign-chg.	$E(Y X)$	0.963	0.915	0.957	0.942	0.646
Poisson	sign-chg.	$\beta(s, t)$		0.935	0.986	0.942	0.728
Poisson	sign-chg.	$E(Y X)$		0.905	0.951	0.941	0.670

Table 9: Extended families at $G = 100$: coverage and NCV/REML MSE ratio (95% bootstrap interval). Negative-binomial data are also fitted as Poisson.

family	fitted family	error	estimand	REML, model-based	REML + CL2	NCV + CL2	NCV + CL2, widened by NCV-REML difference	MSE ratio NCV/REML
beta	beta	smooth	$\beta(s, t)$	0.659	0.941	0.929	0.982	0.18 [0.15, 0.22]
beta	beta	smooth	$E(Y X)$	0.635	0.941	0.924	0.956	0.70 [0.68, 0.72]
beta	beta	indep.	$\beta(s, t)$	0.989	0.984	0.979	0.988	0.75 [0.65, 0.91]
beta	beta	indep.	$E(Y X)$	0.966	0.958	0.946	0.956	0.98 [0.96, 0.99]
negative binomial	negative binomial	smooth	$\beta(s, t)$	0.669	0.943	0.932	0.983	0.18 [0.14, 0.21]
negative binomial	Poisson	smooth	$\beta(s, t)$	0.531	0.938	0.913	0.984	0.09 [0.07, 0.11]
negative binomial	negative binomial	smooth	$E(Y X)$	0.641	0.943	0.925	0.956	0.69 [0.64, 0.74]
negative binomial	Poisson	smooth	$E(Y X)$	0.529	0.939	0.913	0.957	0.55 [0.52, 0.59]
negative binomial	negative binomial	indep.	$\beta(s, t)$	0.991	0.988	0.982	0.990	0.89 [0.75, 1.05]
negative binomial	Poisson	indep.	$\beta(s, t)$	0.945	0.972	0.962	0.982	0.53 [0.49, 0.59]
negative binomial	negative binomial	indep.	$E(Y X)$	0.968	0.962	0.950	0.958	1.06 [1.03, 1.09]
negative binomial	Poisson	indep.	$E(Y X)$	0.898	0.952	0.934	0.950	0.90 [0.87, 0.93]
scaled t	scaled t	smooth	$\beta(s, t)$	0.671	0.944	0.940	0.985	0.22 [0.18, 0.26]
scaled t	scaled t	smooth	$E(Y X)$	0.644	0.947	0.938	0.962	0.71 [0.70, 0.74]
scaled t	scaled t	indep.	$\beta(s, t)$	0.988	0.985	0.982	0.990	0.60 [0.58, 0.63]
scaled t	scaled t	indep.	$E(Y X)$	0.965	0.960	0.950	0.959	0.95 [0.94, 0.97]

S1.4 Studentized errors

The studentized error $z = (\text{estimate} - \text{truth})/\text{SE}$ of the REML fit, pooled over grid points and replicates, is descriptive: an SD above one with mean near zero indicates standard errors that are too small; a shifted mean indicates bias. Pooling cannot separate a global scale error from local biases of opposite sign. In the dependent main-factorial settings at $G = 100$ the SD of z with the CL2 standard error is 0.99–1.03 for β (model-based standard error: 1.48–2.12). The intercept $\alpha(t)$ of binary responses behaves differently. Under independent errors at $G = 40$ its studentized CL2 error has mean 0.044 and SD 1.28 (model-based standard error: mean 0.064, SD 1.15). These are averages over the two signal levels: the SD is 1.49 at the default signal and 1.06 at the high signal. These values are compatible with a CL2 standard error that is too small for this term, but they do not exclude local bias. Under dependence the studentized error is in addition shifted (mean -0.45 to -0.39 at $G = 40$ and -0.24 to -0.23 at $G = 100$, per dependent error process, averaged over the signal levels; SD 1.06 at $G = 100$). The covariance-components runs (Appendix A.6 of the paper) add a diagnostic. Fixing the smoothing parameters at

their typical value raises binary $\gamma(t)$ under independent errors at $G = 40$ from 0.890 to 0.929, and binary $\alpha(t)$ under smooth errors from 0.885 to 0.925; binary $\alpha(t)$ under independent errors stays at 0.867 (0.859 with selected smoothing parameters). For $\gamma(t)$, and for $\alpha(t)$ under dependence, a large part of the shortfall therefore looks like variability of the selected smoothing parameters, which CL2 ignores; for $\alpha(t)$ under independent errors it remains unexplained. Satterthwaite degrees of freedom widen these intervals only slightly (Appendix A.7 of the paper), so they cannot compensate a standard error that is too small by about a fifth.

Table 10: Pooled studentized errors of the REML fit with the CL2 and the model-based standard error: mean, SD and share of $|z| > 1.96$ (means over settings), per estimand; the $\beta(s, t)$ columns hold $f(x, t)$ in the smooth-effect settings. Mean only for Gaussian responses.

block	family	error	G	basis	covariate	E(Y X)						$\beta(s,t)$ or $f(x,t)$						$\alpha(t)$						$\gamma(t)$					
						CL2			model			CL2			model			CL2			model			CL2			model		
						mean	SD	z >1.96	mean	SD	z >1.96	mean	SD	z >1.96	mean	SD	z >1.96	mean	SD	z >1.96	mean	SD	z >1.96	mean	SD	z >1.96	mean	SD	z >1.96
exact AR(1)	Gaussian	AR(1)	100	default	rich	-0.017	1.031	0.058	-0.039	2.219	0.375	0.003	1.032	0.060	0.010	2.130	0.349	-0.039	0.997	0.045	-0.084	2.155	0.364	-0.042	1.022	0.063	-0.098	2.167	0.355
basis size	binary	smooth	100	2* finer knots	rich							0.003	0.962	0.042	0.005	1.582	0.211	-0.382	1.062	0.085	-0.561	1.737	0.272	-0.016	1.014	0.058	-0.017	1.693	0.238
basis size	binary	smooth	100	4* finer knots	rich							0.001	0.959	0.040	0.004	1.555	0.205	-0.594	1.043	0.101	-0.869	1.603	0.287	0.003	1.007	0.054	0.017	1.643	0.221
basis size	binary	indep.	100	2* finer knots	rich							0.010	0.626	0.003	0.009	0.592	0.002	0.025	1.014	0.056	0.044	0.937	0.039	-0.047	1.056	0.064	-0.049	0.985	0.042
basis size	binary	indep.	100	4* finer knots	rich							0.007	0.525	0.001	0.007	0.498	0.000	0.026	0.962	0.044	0.043	0.892	0.031	-0.045	1.011	0.051	-0.046	0.945	0.034
basis size	Gaussian	smooth	100	2* finer knots	rich	-0.013	1.010	0.053	-0.027	2.018	0.327	0.000	1.015	0.055	0.002	2.060	0.329	-0.023	0.999	0.050	-0.044	1.918	0.303	-0.022	0.999	0.050	-0.046	1.896	0.297
basis size	Gaussian	smooth	100	4* finer knots	rich	-0.014	0.991	0.049	-0.028	1.872	0.294	0.001	0.983	0.046	0.004	2.005	0.327	-0.031	0.982	0.046	-0.056	1.756	0.261	-0.022	0.998	0.050	-0.040	1.762	0.266
basis size	Gaussian	indep.	100	2* finer knots	rich	-0.001	0.878	0.027	-0.002	0.861	0.022	0.002	0.650	0.003	0.002	0.637	0.002	-0.007	0.936	0.035	-0.006	0.925	0.032	-0.025	0.934	0.037	-0.026	0.908	0.029
basis size	Gaussian	indep.	100	4* finer knots	rich	-0.001	0.822	0.018	-0.001	0.810	0.015	0.001	0.535	0.000	0.001	0.526	0.000	-0.006	0.884	0.025	-0.005	0.875	0.022	-0.022	0.884	0.027	-0.022	0.864	0.024
basis size	Poisson	smooth	100	2* finer knots	rich							0.001	1.004	0.052	0.000	1.929	0.298	-0.169	0.986	0.051	-0.304	1.810	0.284	-0.027	0.995	0.050	-0.046	1.804	0.277
basis size	Poisson	smooth	100	4* finer knots	rich							0.000	0.978	0.045	0.003	1.843	0.285	-0.241	0.967	0.050	-0.404	1.639	0.248	-0.019	0.983	0.047	-0.023	1.646	0.231
basis size	Poisson	indep.	100	2* finer knots	rich							0.005	0.645	0.003	0.005	0.630	0.002	0.034	0.932	0.036	0.036	0.917	0.032	-0.035	0.931	0.036	-0.036	0.903	0.029
basis size	Poisson	indep.	100	4* finer knots	rich							0.003	0.529	0.000	0.003	0.518	0.000	0.037	0.878	0.025	0.039	0.864	0.022	-0.030	0.886	0.027	-0.032	0.862	0.022
main factorial	binary	smooth	40	default	rich							0.009	1.081	0.071	0.015	1.716	0.243	-0.454	1.134	0.106	-0.657	1.902	0.323	-0.045	1.119	0.080	-0.073	1.939	0.296
main factorial	binary	smooth	100	default	rich							0.001	0.987	0.048	0.004	1.608	0.219	-0.241	1.063	0.075	-0.355	1.844	0.284	-0.019	1.029	0.061	-0.025	1.818	0.266
main factorial	binary	indep.	40	default	rich							0.011	0.835	0.021	0.010	0.767	0.013	0.044	1.276	0.101	0.064	1.149	0.070	-0.068	1.203	0.094	-0.065	1.055	0.063
main factorial	binary	indep.	100	default	rich							0.011	0.751	0.010	0.011	0.713	0.007	0.033	1.060	0.064	0.049	0.993	0.050	-0.041	1.082	0.071	-0.043	1.018	0.054
main factorial	binary	rough	40	default	rich							0.012	1.048	0.061	0.019	1.511	0.194	-0.393	1.130	0.095	-0.521	1.778	0.271	-0.062	1.113	0.076	-0.111	1.728	0.265
main factorial	binary	rough	100	default	rich							0.006	0.987	0.049	0.010	1.479	0.181	-0.232	1.058	0.068	-0.317	1.712	0.246	-0.037	1.053	0.068	-0.059	1.714	0.251
main factorial	Gaussian	smooth	40	default	rich	-0.020	1.073	0.069	-0.040	2.334	0.394	0.003	1.094	0.075	0.006	2.336	0.387	-0.041	1.055	0.064	-0.080	2.315	0.385	-0.050	1.029	0.055	-0.103	2.219	0.377
main factorial	Gaussian	smooth	100	default	rich	-0.016	1.026	0.056	-0.034	2.201	0.366	0.003	1.021	0.056	0.009	2.123	0.344	-0.032	1.003	0.051	-0.062	2.144	0.359	-0.026	1.002	0.052	-0.062	2.111	0.347
main factorial	Gaussian	indep.	40	default	rich	-0.010	0.959	0.043	-0.008	0.919	0.033	0.003	0.769	0.013	0.003	0.743	0.009	-0.027	1.014	0.055	-0.018	0.981	0.045	-0.026	1.007	0.054	-0.030	0.956	0.041
main factorial	Gaussian	indep.	100	default	rich	-0.002	0.952	0.040	-0.001	0.927	0.034	0.005	0.794	0.015	0.005	0.773	0.012	-0.011	1.008	0.047	-0.008	0.993	0.044	-0.031	0.990	0.050	-0.033	0.959	0.039
main factorial	Gaussian	rough	40	default	rich	-0.017	1.075	0.069	-0.035	2.227	0.377	0.006	1.093	0.074	0.012	2.182	0.363	-0.042	1.048	0.062	-0.081	2.214	0.369	-0.060	1.035	0.062	-0.129	2.125	0.350
main factorial	Gaussian	rough	100	default	rich	-0.016	1.032	0.058	-0.036	2.127	0.354	0.004	1.029	0.059	0.010	2.049	0.329	-0.035	1.003	0.049	-0.075	2.062	0.341	-0.046	1.022	0.062	-0.100	2.066	0.339
main factorial	Poisson	smooth	40	default	rich							0.006	1.079	0.071	0.010	2.063	0.330	-0.202	1.031	0.063	-0.383	2.080	0.353	-0.050	1.038	0.060	-0.095	2.056	0.348
main factorial	Poisson	smooth	100	default	rich							0.001	1.015	0.055	0.004	1.977	0.311	-0.112	0.997	0.052	-0.215	2.037	0.338	-0.038	1.012	0.056	-0.077	2.025	0.334
main factorial	Poisson	indep.	40	default	rich							0.009	0.769	0.012	0.008	0.741	0.008	0.030	0.991	0.048	0.036	0.960	0.039	-0.026	1.026	0.055	-0.027	0.969	0.045
main factorial	Poisson	indep.	100	default	rich							0.008	0.810	0.018	0.009	0.785	0.013	0.033	1.008	0.050	0.035	0.991	0.048	-0.043	0.997	0.051	-0.044	0.963	0.041
main factorial	Poisson	rough	40	default	rich							0.008	1.081	0.072	0.014	1.938	0.306	-0.193	1.022	0.060	-0.353	1.982	0.334	-0.071	1.066	0.066	-0.145	1.994	0.332
main factorial	Poisson	rough	100	default	rich							0.004	1.029	0.059	0.010	1.925	0.299	-0.115	1.006	0.051	-0.219	1.978	0.331	-0.065	1.028	0.063	-0.124	1.962	0.318
dense grid	binary	smooth	100	default	rich							0.003	1.026	0.057	0.013	3.578	0.580	-0.421	1.034	0.080	-1.282	3.386	0.590	-0.005	1.013	0.057	-0.028	3.410	0.555
dense grid	binary	indep.	100	default	rich							0.005	0.739	0.010	0.004	0.713	0.007	-0.007	1.004	0.049	-0.001	0.970	0.043	-0.063	1.001	0.049	-0.063	0.962	0.040
dense grid	binary	rough	100	default	rich							0.004	1.047	0.062	0.013	3.278	0.550	-0.393	1.043	0.081	-1.064	3.076	0.553	-0.029	1.036	0.059	-0.087	3.111	0.526
dense grid	Gaussian	smooth	100	default	rich	-0.011	1.028	0.057	-0.051	4.408	0.647	0.000	1.024	0.056	0.007	4.468	0.650	-0.024	1.007	0.053	-0.098	4.203	0.627	-0.016	1.010	0.053	-0.085	4.168	0.631
dense grid	Gaussian	indep.	100	default	rich	-0.011	0.970	0.044	-0.010	0.948	0.039	0.002	0.854	0.024	0.002	0.837	0.020	-0.023	0.997	0.049	-0.022	0.990	0.048	-0.027	1.009	0.057	-0.030	0.979	0.049
dense grid	Gaussian	rough	100	default	rich	-0.012	1.035	0.059	-0.055	4.186	0.633	0.002	1.038	0.060	0.013	4.293	0.641	-0.031	0.996	0.047	-0.135	3.932	0.618	-0.036	1.027	0.064	-0.162	3.984	0.603
dense grid	Poisson	smooth	100	default	rich							0.000	1.029	0.057	0.004	4.226	0.631	-0.178	0.997	0.053	-0.690	3.958	0.620	-0.020	1.007	0.053	-0.084	3.943	0.613
dense grid	Poisson	indep.	100	default	rich							0.003	0.840	0.022	0.004	0.821	0.017	0.001	0.993	0.049	0.001	0.986	0.046	-0.041	1.000	0.052	-0.042	0.965	0.043
dense grid	Poisson	rough	100	default	rich							0.002	1.045	0.062	0.010	4.092	0.626	-0.178	0.993	0.048	-0.670	3.761	0.603	-0.046	1.024	0.064	-0.178	3.783	0.5

Table 10: Pooled studentized errors of the REML fit with the CL2 and the model-based standard error: mean, SD and share of $|z| > 1.96$ (means over settings), per estimand; the $\beta(s, t)$ columns hold $f(x, t)$ in the smooth-effect settings. Mean only for Gaussian responses. (continued)

block	family	error	G	basis	covariate	mean	SD	z >1.96	mean	SD	z >1.96	mean	SD	z >1.96	mean	SD	z >1.96	mean	SD	z >1.96	mean	SD	z >1.96	mean	SD	z >1.96			
sign-changing	Poisson	sign-chg.	100	default	rich							0.001	1.027	0.058	0.004	1.801	0.272	-0.117	1.015	0.053	-0.226	2.058	0.344	-0.023	1.023	0.061	-0.050	2.013	0.327
rough truth	binary	smooth	100	default	rich							0.000	1.007	0.052	0.002	1.654	0.234	-0.246	1.057	0.074	-0.362	1.833	0.282	-0.005	1.045	0.065	-0.026	1.847	0.268
rough truth	binary	indep.	100	default	rich							0.002	1.044	0.060	0.001	0.989	0.048	0.070	1.070	0.066	0.084	0.995	0.051	-0.023	1.205	0.103	-0.035	1.120	0.082
rough truth	Gaussian	smooth	100	default	rich	-0.016	1.029	0.057	-0.033	2.213	0.369	0.002	1.031	0.058	0.007	2.164	0.355	-0.030	1.004	0.051	-0.058	2.149	0.359	-0.021	1.010	0.054	-0.065	2.117	0.350
rough truth	Gaussian	indep.	100	default	rich	-0.003	0.995	0.050	-0.003	0.969	0.043	0.002	0.958	0.042	0.002	0.930	0.036	-0.011	1.008	0.046	-0.008	0.994	0.044	-0.031	1.019	0.056	-0.034	0.988	0.047
rough truth	Poisson	smooth	100	default	rich							0.001	1.032	0.059	0.004	2.029	0.326	-0.114	0.994	0.051	-0.214	2.024	0.333	-0.026	1.016	0.056	-0.069	2.017	0.326
rough truth	Poisson	indep.	100	default	rich							0.001	0.941	0.040	0.001	0.912	0.033	0.043	1.013	0.049	0.046	0.999	0.048	-0.042	1.025	0.056	-0.044	0.990	0.046
smooth effect	binary	smooth	100	default	rich							0.002	1.022	0.055	0.059	1.689	0.246						0.041	1.051	0.057	0.076	1.807	0.277	
smooth effect	binary	indep.	100	default	rich							-0.059	0.815	0.018	-0.057	0.778	0.012						-0.017	1.080	0.070	-0.016	0.998	0.048	
smooth effect	binary	rough	100	default	rich							-0.010	1.005	0.051	0.040	1.556	0.209						0.011	1.108	0.075	0.018	1.810	0.271	
smooth effect	Gaussian	smooth	100	default	rich	-0.020	1.054	0.063	-0.041	2.133	0.350	-0.006	1.035	0.058	-0.013	2.123	0.350						0.007	1.008	0.050	0.006	2.115	0.352	
smooth effect	Gaussian	indep.	100	default	rich	0.000	0.944	0.039	-0.001	0.915	0.032	-0.021	0.897	0.029	-0.022	0.878	0.024						-0.016	0.977	0.048	-0.019	0.951	0.038	
smooth effect	Gaussian	rough	100	default	rich	-0.017	1.047	0.062	-0.036	2.031	0.332	0.000	1.016	0.055	0.000	2.000	0.323						-0.024	1.023	0.062	-0.057	2.062	0.339	
smooth effect	Poisson	smooth	100	default	rich							0.009	1.031	0.057	0.027	2.027	0.329						0.015	1.017	0.054	0.025	2.035	0.336	
smooth effect	Poisson	indep.	100	default	rich							-0.048	0.884	0.027	-0.047	0.863	0.022						-0.004	0.994	0.051	-0.008	0.963	0.041	
smooth effect	Poisson	rough	100	default	rich							0.010	1.014	0.055	0.029	1.918	0.303						-0.012	1.024	0.062	-0.031	1.979	0.327	
misreg.	Gaussian	misreg.	40	default	rich	0.009	1.059	0.064	-0.001	2.032	0.282	0.001	1.011	0.053	0.002	1.842	0.250	0.009	1.001	0.054	-0.005	1.847	0.240	-0.013	1.123	0.078	-0.043	2.211	0.302
misreg.	Gaussian	misreg.	100	default	rich	0.009	1.040	0.060	0.008	2.040	0.282	0.001	1.009	0.052	0.005	1.887	0.259	0.015	1.012	0.049	0.021	1.847	0.251	-0.021	1.074	0.070	-0.051	2.170	0.294
misreg. (AR(1))	Gaussian	misreg.	100	default	rich	0.007	1.051	0.063	0.006	2.122	0.294	0.001	1.026	0.056	0.004	1.961	0.271	0.012	1.024	0.051	0.017	1.943	0.268	-0.020	1.087	0.072	-0.048	2.264	0.308

S1.5 Estimation

Under dependent errors the NCV/REML MSE ratio for β , geometric mean over the main-factorial settings of each error process, is 0.12–0.21 (independent errors: 0.69–1.05); for the mean 0.58–0.74 (same summary; per setting at $G = 100$ the mean's ratio is 0.59–0.80). NCV's advantage for β does not vanish under independent errors: for Gaussian responses the per-setting ratio is 0.55–0.94 there. Part of the gap therefore reflects the smooth truths, which favour more smoothing, and not only the dependence. With the rough truth the β ratio is 0.21–0.43. REML's excess error is variance: the squared-bias share of REML's β error in the dependent settings at $G = 100$ is 0.01, of NCV's 0.12–0.20; with the rough truth NCV's share rises to 0.26–0.42. REML's surface EDF exceeds NCV's by a factor 1.92–2.31 under dependence (1.09–1.31 under independence). Pointwise NCV, which leaves out single grid points, is no substitute for curve blocks: its β MSE is 2.86–4.86 times REML's under smooth and sign-changing errors. A low-rank covariate gives a β ratio of 0.09–0.12 under smooth errors.

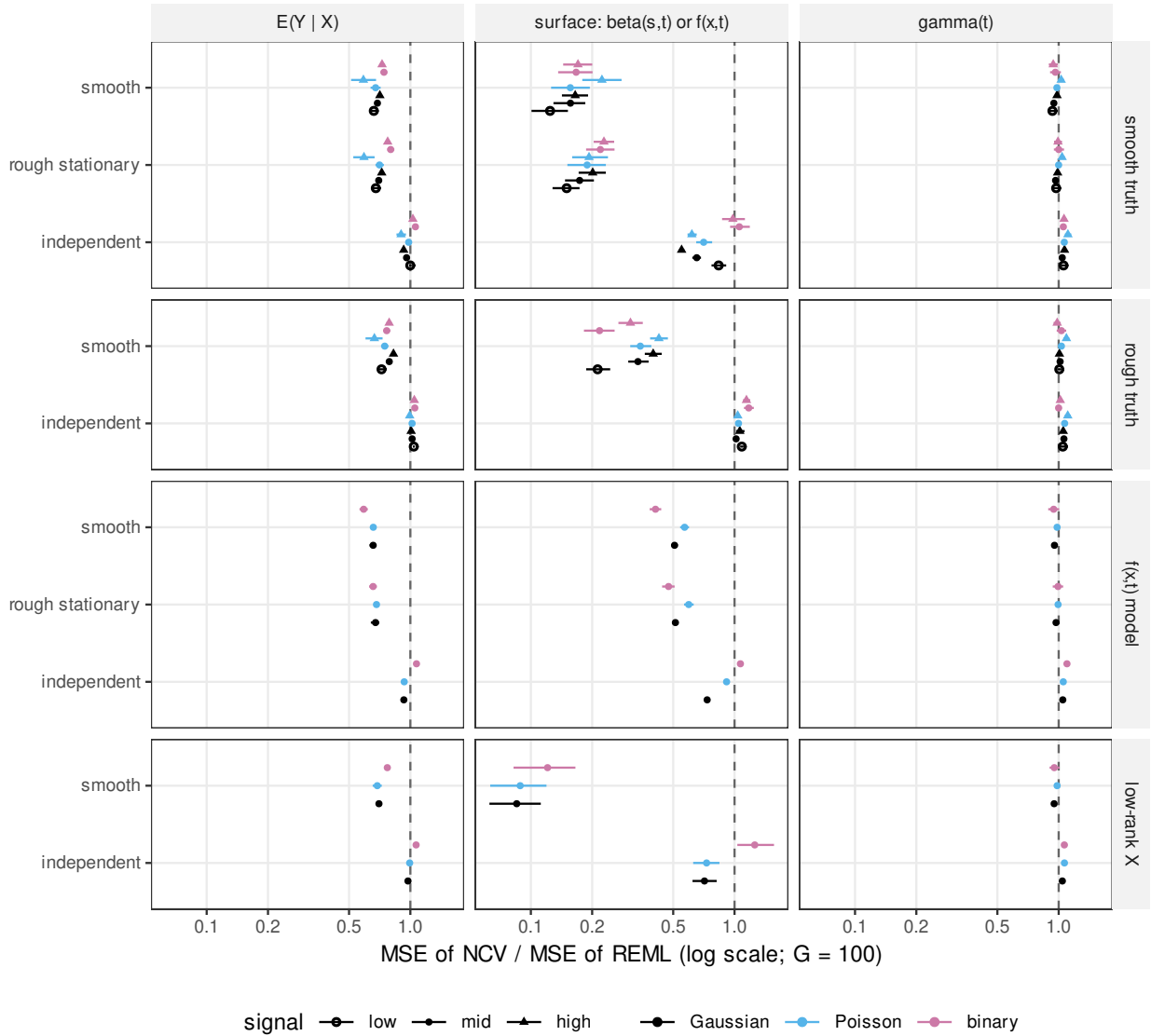


Figure 3: MSE of the NCV fit relative to the REML fit (log scale) with 95% replicate-bootstrap intervals, $G = 100$, by error process, family and signal level: smooth truth, rough truth, smooth-effect model and low-rank covariate. The surface column shows $\beta(s,t)$ or $f(x,t)$.

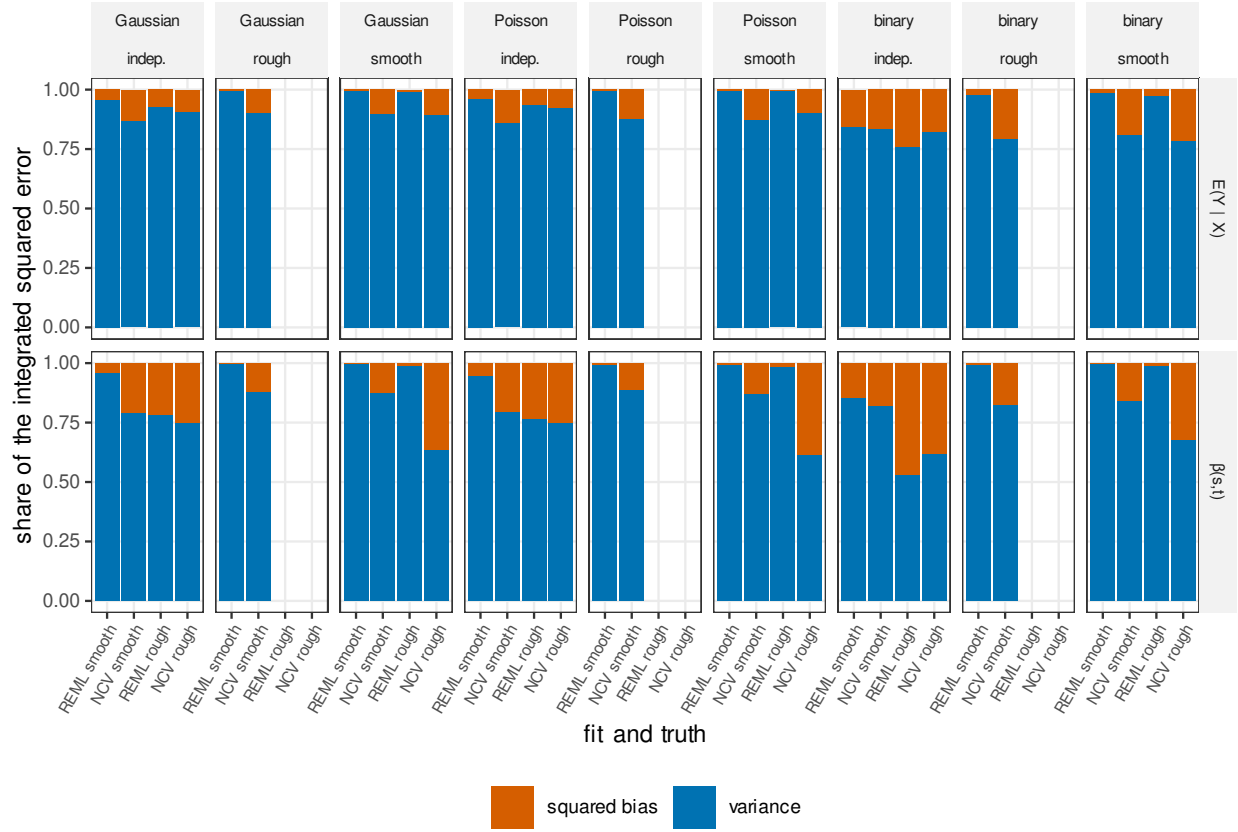


Figure 4: Share of squared bias and variance in the integrated squared error of $\beta(s, t)$ and the mean, REML and NCV fits, $G = 100$, default signal.

Table 11: NCV/REML MSE ratio, geometric mean over the main-factorial settings of each error process (all G and signal levels), 95% bootstrap intervals.

family	estimand	independent	rough stationary	smooth
Gaussian	$E(Y X)$	0.97 [0.96, 0.98]	0.66 [0.64, 0.68]	0.64 [0.62, 0.65]
Gaussian	$\beta(s, t)$	0.73 [0.69, 0.78]	0.15 [0.13, 0.16]	0.12 [0.10, 0.14]
Poisson	$E(Y X)$	0.97 [0.93, 1.01]	0.59 [0.53, 0.66]	0.58 [0.52, 0.64]
Poisson	$\beta(s, t)$	0.69 [0.66, 0.72]	0.17 [0.14, 0.19]	0.15 [0.13, 0.17]
binary	$E(Y X)$	1.06 [1.04, 1.07]	0.74 [0.72, 0.76]	0.68 [0.66, 0.70]
binary	$\beta(s, t)$	1.05 [0.96, 1.15]	0.21 [0.18, 0.24]	0.17 [0.14, 0.19]

Table 12: Mean EDF per term of the REML and NCV fits, main factorial at $G = 100$, default signal.

family	error	term	NCV	REML	ratio
binary	smooth	$\alpha(t)$	2.83	5.06	1.786
binary	indep.	$\alpha(t)$	4.70	4.47	0.951
binary	rough	$\alpha(t)$	2.74	5.03	1.832
Gaussian	smooth	$\alpha(t)$	7.26	8.46	1.166
Gaussian	indep.	$\alpha(t)$	8.09	8.33	1.030
Gaussian	rough	$\alpha(t)$	7.33	8.41	1.148
Poisson	smooth	$\alpha(t)$	6.63	8.09	1.220
Poisson	indep.	$\alpha(t)$	7.53	7.97	1.059
Poisson	rough	$\alpha(t)$	6.69	8.05	1.203
binary	smooth	$\beta(s,t)$	11.60	26.77	2.308
binary	indep.	$\beta(s,t)$	15.72	17.07	1.086
binary	rough	$\beta(s,t)$	11.99	25.39	2.117
Gaussian	smooth	$\beta(s,t)$	21.03	41.84	1.990
Gaussian	indep.	$\beta(s,t)$	25.48	33.50	1.315
Gaussian	rough	$\beta(s,t)$	21.35	41.09	1.925
Poisson	smooth	$\beta(s,t)$	19.64	39.37	2.004
Poisson	indep.	$\beta(s,t)$	24.30	31.42	1.293
Poisson	rough	$\beta(s,t)$	20.07	38.71	1.928
binary	smooth	$\gamma(t)$	3.70	5.93	1.602
binary	indep.	$\gamma(t)$	5.33	5.21	0.979
binary	rough	$\gamma(t)$	3.69	5.72	1.548
Gaussian	smooth	$\gamma(t)$	7.44	9.18	1.233
Gaussian	indep.	$\gamma(t)$	8.39	9.02	1.075
Gaussian	rough	$\gamma(t)$	7.49	9.12	1.218
Poisson	smooth	$\gamma(t)$	7.26	8.82	1.215
Poisson	indep.	$\gamma(t)$	8.15	8.68	1.065
Poisson	rough	$\gamma(t)$	7.25	8.75	1.207

Table 13: SD over replicates of the selected log smoothing parameters of the surface term (mean over its two penalties), main factorial at the default signal.

family	error	G	NCV	REML
binary	smooth	40	4.672	1.742
binary	smooth	100	1.912	1.340
binary	indep.	40	1.425	0.906
binary	indep.	100	1.123	0.267
binary	rough	40	4.257	1.502
binary	rough	100	1.850	1.130
Gaussian	smooth	40	1.014	1.508
Gaussian	smooth	100	1.063	1.281
Gaussian	indep.	40	0.885	0.128
Gaussian	indep.	100	0.661	0.097
Gaussian	rough	40	1.372	1.421
Gaussian	rough	100	0.972	1.168
Poisson	smooth	40	1.106	1.479
Poisson	smooth	100	1.045	1.268
Poisson	indep.	40	0.756	0.130
Poisson	indep.	100	0.894	0.106
Poisson	rough	40	1.448	1.367
Poisson	rough	100	1.128	1.157

Table 14: Pointwise (non-blocked) NCV relative to REML: MSE ratio with 95% bootstrap interval, Gaussian, $G = 100$.

error	estimand	MSE pointwise NCV / REML
indep.	$\beta(s,t)$	0.65 [0.62, 0.68]
indep.	$E(Y X)$	0.96 [0.95, 0.97]
smooth	$\beta(s,t)$	2.86 [2.57, 3.21]
smooth	$E(Y X)$	1.34 [1.32, 1.37]
sign-chg.	$\beta(s,t)$	4.86 [4.35, 5.38]
sign-chg.	$E(Y X)$	1.45 [1.42, 1.47]

Table 15: NCV/REML MSE ratio with the rich and the low-rank functional covariate, $G = 100$, default signal.

family	error	estimand	rich	low-rank
binary	smooth	$\alpha(t)$	0.88 [0.84, 0.93]	0.87 [0.82, 0.91]
binary	smooth	$\beta(s, t)$	0.17 [0.14, 0.20]	0.12 [0.08, 0.17]
binary	smooth	$\gamma(t)$	0.96 [0.91, 1.03]	0.95 [0.90, 1.01]
binary	smooth	$E(Y X)$	0.74 [0.71, 0.77]	0.77 [0.74, 0.80]
binary	indep.	$\alpha(t)$	1.10 [1.05, 1.17]	1.06 [1.03, 1.11]
binary	indep.	$\beta(s, t)$	1.06 [0.95, 1.19]	1.26 [1.03, 1.56]
binary	indep.	$\gamma(t)$	1.05 [1.03, 1.09]	1.07 [1.04, 1.10]
binary	indep.	$E(Y X)$	1.06 [1.04, 1.08]	1.07 [1.05, 1.09]
Gaussian	smooth	$\alpha(t)$	0.97 [0.94, 0.99]	0.97 [0.94, 0.99]
Gaussian	smooth	$\beta(s, t)$	0.16 [0.13, 0.19]	0.09 [0.06, 0.11]
Gaussian	smooth	$\gamma(t)$	0.95 [0.92, 0.97]	0.95 [0.93, 0.97]
Gaussian	smooth	$E(Y X)$	0.69 [0.67, 0.71]	0.70 [0.68, 0.72]
Gaussian	indep.	$\alpha(t)$	1.04 [1.02, 1.07]	1.04 [1.02, 1.07]
Gaussian	indep.	$\beta(s, t)$	0.65 [0.62, 0.68]	0.71 [0.62, 0.82]
Gaussian	indep.	$\gamma(t)$	1.04 [1.02, 1.07]	1.04 [1.02, 1.07]
Gaussian	indep.	$E(Y X)$	0.96 [0.95, 0.97]	0.97 [0.96, 0.99]
Poisson	smooth	$\alpha(t)$	0.94 [0.90, 0.97]	0.93 [0.90, 0.97]
Poisson	smooth	$\beta(s, t)$	0.16 [0.13, 0.19]	0.09 [0.06, 0.12]
Poisson	smooth	$\gamma(t)$	0.98 [0.95, 1.01]	0.98 [0.95, 1.01]
Poisson	smooth	$E(Y X)$	0.68 [0.64, 0.71]	0.69 [0.65, 0.72]
Poisson	indep.	$\alpha(t)$	1.04 [1.01, 1.07]	1.03 [1.00, 1.06]
Poisson	indep.	$\beta(s, t)$	0.71 [0.65, 0.78]	0.73 [0.63, 0.84]
Poisson	indep.	$\gamma(t)$	1.07 [1.04, 1.10]	1.07 [1.04, 1.10]
Poisson	indep.	$E(Y X)$	0.98 [0.96, 1.00]	0.99 [0.97, 1.01]

Table 16: Integrated MSE of the REML and NCV fits per setting and estimand, and their paired ratio NCV/REML (95% bootstrap interval); for the smooth-effect settings the $\beta(s, t)$ columns hold $f(x, t)$.

block	family	error	G	signal	truth	basis	E(Y X)			$\beta(s,t)$			$\alpha(t)$			$\gamma(t)$		
							REML	NCV	ratio	REML	NCV	ratio	REML	NCV	ratio	REML	NCV	ratio
main factorial	Gaussian	indep.	40	low	smooth	default	0.0619	0.0636	1.03 [1.01, 1.04]	0.1677	0.1575	0.94 [0.86, 1.04]	0.0129	0.0137	1.06 [1.03, 1.09]	0.0126	0.0134	1.06 [1.04, 1.09]
main factorial	Gaussian	indep.	100	low	smooth	default	0.0271	0.0271	1.00 [0.98, 1.01]	0.0930	0.0777	0.84 [0.77, 0.91]	0.0056	0.0059	1.05 [1.02, 1.08]	0.0054	0.0057	1.06 [1.03, 1.09]
main factorial	Gaussian	rough	40	low	smooth	default	0.4799	0.2958	0.62 [0.59, 0.64]	8.1491	0.9605	0.12 [0.09, 0.15]	0.0657	0.0668	1.02 [0.95, 1.08]	0.0625	0.0614	0.98 [0.92, 1.04]
main factorial	Gaussian	rough	100	low	smooth	default	0.1718	0.1163	0.68 [0.66, 0.70]	2.3958	0.3591	0.15 [0.13, 0.17]	0.0244	0.0236	0.97 [0.94, 1.00]	0.0250	0.0243	0.97 [0.94, 1.00]
main factorial	Gaussian	smooth	40	low	smooth	default	0.5272	0.3064	0.58 [0.56, 0.61]	10.4813	0.9421	0.09 [0.07, 0.11]	0.0728	0.0694	0.95 [0.90, 1.01]	0.0669	0.0615	0.92 [0.87, 0.98]
main factorial	Gaussian	smooth	100	low	smooth	default	0.1827	0.1209	0.66 [0.64, 0.68]	2.8137	0.3496	0.12 [0.10, 0.15]	0.0260	0.0255	0.98 [0.94, 1.03]	0.0255	0.0237	0.93 [0.91, 0.96]
main factorial	Gaussian	indep.	40	default	smooth	default	0.0196	0.0192	0.98 [0.96, 1.00]	0.0742	0.0633	0.85 [0.70, 1.13]	0.0039	0.0040	1.04 [1.01, 1.06]	0.0038	0.0040	1.04 [1.02, 1.06]
main factorial	Gaussian	indep.	100	default	smooth	default	0.0084	0.0081	0.96 [0.95, 0.97]	0.0412	0.0269	0.65 [0.62, 0.68]	0.0016	0.0017	1.04 [1.02, 1.07]	0.0016	0.0017	1.04 [1.02, 1.07]
main factorial	Gaussian	rough	40	default	smooth	default	0.1298	0.0799	0.62 [0.59, 0.64]	2.1864	0.2788	0.13 [0.10, 0.16]	0.0179	0.0166	0.93 [0.90, 0.96]	0.0172	0.0161	0.93 [0.90, 0.97]
main factorial	Gaussian	rough	100	default	smooth	default	0.0464	0.0325	0.70 [0.68, 0.72]	0.6473	0.1124	0.17 [0.15, 0.20]	0.0066	0.0063	0.96 [0.95, 0.98]	0.0068	0.0065	0.97 [0.94, 0.99]
main factorial	Gaussian	smooth	40	default	smooth	default	0.1415	0.0832	0.59 [0.57, 0.61]	2.7504	0.2633	0.10 [0.08, 0.12]	0.0198	0.0180	0.91 [0.87, 0.95]	0.0183	0.0163	0.89 [0.85, 0.93]
main factorial	Gaussian	smooth	100	default	smooth	default	0.0494	0.0340	0.69 [0.67, 0.71]	0.7524	0.1177	0.16 [0.13, 0.19]	0.0070	0.0068	0.97 [0.94, 0.99]	0.0069	0.0066	0.95 [0.92, 0.97]
main factorial	Gaussian	indep.	40	high	smooth	default	0.0061	0.0057	0.94 [0.92, 0.95]	0.0330	0.0205	0.62 [0.58, 0.67]	0.0011	0.0011	1.03 [1.01, 1.05]	0.0012	0.0012	1.04 [1.01, 1.06]
main factorial	Gaussian	indep.	100	high	smooth	default	0.0026	0.0024	0.93 [0.92, 0.94]	0.0177	0.0097	0.55 [0.53, 0.57]	0.0004	0.0005	1.03 [1.01, 1.05]	0.0005	0.0005	1.07 [1.04, 1.10]
main factorial	Gaussian	rough	40	high	smooth	default	0.0351	0.0221	0.63 [0.61, 0.65]	0.6178	0.0735	0.12 [0.10, 0.14]	0.0048	0.0044	0.93 [0.90, 0.96]	0.0047	0.0043	0.93 [0.90, 0.96]
main factorial	Gaussian	rough	100	high	smooth	default	0.0126	0.0091	0.72 [0.71, 0.74]	0.1895	0.0381	0.20 [0.17, 0.23]	0.0017	0.0017	0.98 [0.96, 1.00]	0.0018	0.0018	0.99 [0.97, 1.01]
main factorial	Gaussian	smooth	40	high	smooth	default	0.0381	0.0233	0.61 [0.59, 0.63]	0.7526	0.0795	0.11 [0.09, 0.13]	0.0052	0.0048	0.91 [0.88, 0.95]	0.0050	0.0045	0.91 [0.87, 0.94]
main factorial	Gaussian	smooth	100	high	smooth	default	0.0134	0.0095	0.71 [0.69, 0.73]	0.2135	0.0353	0.17 [0.14, 0.19]	0.0019	0.0018	0.97 [0.95, 1.00]	0.0018	0.0018	0.98 [0.96, 1.00]
main factorial	Poisson	indep.	40	default	smooth	default	0.0816	0.0817	1.00 [0.98, 1.03]	0.0235	0.0170	0.72 [0.68, 0.76]	0.0014	0.0015	1.04 [1.02, 1.06]	0.0014	0.0015	1.04 [1.02, 1.06]
main factorial	Poisson	indep.	100	default	smooth	default	0.0337	0.0330	0.98 [0.96, 1.00]	0.0130	0.0092	0.71 [0.65, 0.78]	0.0006	0.0007	1.04 [1.01, 1.07]	0.0006	0.0006	1.07 [1.04, 1.10]
main factorial	Poisson	rough	40	default	smooth	default	0.5989	0.3617	0.60 [0.55, 0.65]	0.6181	0.0854	0.14 [0.11, 0.17]	0.0072	0.0061	0.86 [0.81, 0.90]	0.0062	0.0060	0.97 [0.92, 1.03]
main factorial	Poisson	rough	100	default	smooth	default	0.1949	0.1375	0.71 [0.67, 0.74]	0.1986	0.0375	0.19 [0.15, 0.23]	0.0026	0.0024	0.92 [0.88, 0.95]	0.0023	0.0023	1.00 [0.97, 1.02]
main factorial	Poisson	smooth	40	default	smooth	default	0.6682	0.3816	0.57 [0.53, 0.61]	0.8135	0.0794	0.10 [0.08, 0.12]	0.0078	0.0065	0.83 [0.78, 0.89]	0.0064	0.0062	0.96 [0.92, 1.01]
main factorial	Poisson	smooth	100	default	smooth	default	0.2194	0.1481	0.68 [0.64, 0.71]	0.2277	0.0356	0.16 [0.13, 0.19]	0.0027	0.0026	0.94 [0.90, 0.97]	0.0024	0.0023	0.98 [0.95, 1.01]
main factorial	Poisson	indep.	40	high	smooth	default	0.5135	0.5123	1.00 [0.92, 1.08]	0.0539	0.0385	0.71 [0.66, 0.79]	0.0031	0.0034	1.08 [1.05, 1.12]	0.0027	0.0030	1.09 [1.06, 1.14]
main factorial	Poisson	indep.	100	high	smooth	default	0.1768	0.1592	0.90 [0.85, 0.95]	0.0292	0.0180	0.62 [0.59, 0.65]	0.0013	0.0014	1.06 [1.03, 1.10]	0.0010	0.0011	1.11 [1.07, 1.15]
main factorial	Poisson	rough	40	high	smooth	default	4.3810	2.1726	0.50 [0.37, 0.64]	0.9085	0.1389	0.15 [0.13, 0.18]	0.0136	0.0128	0.93 [0.87, 1.01]	0.0107	0.0104	0.97 [0.91, 1.03]
main factorial	Poisson	rough	100	high	smooth	default	1.0830	0.6411	0.59 [0.52, 0.67]	0.3030	0.0585	0.19 [0.16, 0.24]	0.0051	0.0049	0.96 [0.92, 1.00]	0.0037	0.0039	1.04 [1.01, 1.07]
main factorial	Poisson	smooth	40	high	smooth	default	5.0480	2.4766	0.49 [0.36, 0.63]	1.1357	0.1547	0.14 [0.11, 0.17]	0.0152	0.0134	0.88 [0.81, 0.96]	0.0107	0.0106	0.99 [0.93, 1.04]
main factorial	Poisson	smooth	100	high	smooth	default	1.2444	0.7319	0.59 [0.51, 0.68]	0.3238	0.0723	0.22 [0.18, 0.28]	0.0053	0.0051	0.96 [0.92, 1.01]	0.0038	0.0039	1.03 [0.99, 1.06]
main factorial	binary	indep.	40	default	smooth	default	0.0019	0.0020	1.08 [1.05, 1.12]	0.1508	0.1733	1.15 [0.99, 1.36]	0.0174	0.0186	1.07 [1.01, 1.13]	0.0168	0.0196	1.17 [1.10, 1.25]
main factorial	binary	indep.	100	default	smooth	default	0.0008	0.0009	1.06 [1.04, 1.08]	0.0754	0.0796	1.06 [0.95, 1.19]	0.0073	0.0080	1.10 [1.05, 1.17]	0.0075	0.0079	1.05 [1.03, 1.09]
main factorial	binary	rough	40	default	smooth	default	0.0094	0.0061	0.66 [0.62, 0.69]	3.7624	0.7505	0.20 [0.15, 0.25]	0.0711	0.0482	0.68 [0.63, 0.73]	0.0446	0.0523	0.81 [0.75, 0.87]
main factorial	binary	rough	100	default	smooth	default	0.0035	0.0028	0.80 [0.77, 0.83]	1.1349	0.2490	0.22 [0.19, 0.26]	0.0264	0.0241	0.91 [0.87, 0.97]	0.0242	0.0242	1.00 [0.95, 1.06]
main factorial	binary	smooth	40	default	smooth	default	0.0116	0.0069	0.59 [0.56, 0.62]	6.5792	1.0127	0.15 [0.11, 0.20]	0.0959	0.0541	0.56 [0.52, 0.61]	0.0793	0.0543	0.69 [0.64, 0.74]
main factorial	binary	smooth	100	default	smooth	default	0.0041	0.0031	0.74 [0.71, 0.77]	1.6613	0.2772	0.17 [0.14, 0.20]	0.0309	0.0272	0.88 [0.84, 0.93]	0.0268	0.0258	0.96 [0.91, 1.03]
main factorial	binary	indep.	40	high	smooth	default	0.0022	0.0023	1.06 [1.03, 1.09]	0.2715	0.2799	1.03 [0.90, 1.21]	0.0239	0.0261	1.09 [1.05, 1.13]	0.0251	0.0267	1.06 [1.03, 1.11]
main factorial	binary	indep.	100	high	smooth	default	0.0010	0.0010	1.03 [1.01, 1.05]	0.1475	0.1447	0.98 [0.87, 1.12]	0.0105	0.0113	1.08 [1.05, 1.12]	0.0108	0.0115	1.06 [1.04, 1.09]
main factorial	binary	rough	40	high	smooth	default	0.0098	0.0072	0.74 [0.71, 0.77]	4.4148	0.8701	0.20 [0.16, 0.24]	0.1037	0.0761	0.73 [0.68, 0.79]	0.0863	0.0785	0.91 [0.84, 0.99]
main factorial	binary	rough	100	high	smooth	default	0.0037	0.0028	0.77 [0.75, 0.80]	1.4186	0.3241	0.23 [0.20, 0.26]	0.0359	0.0308	0.86 [0.82, 0.90]	0.0314	0.0311	0.99 [0.95, 1.04]
main factorial	binary	smooth	40	high	smooth	default	0.0121	0.0080	0.66 [0.63, 0.70]	7.7267	1.3164	0.17 [0.12, 0.23]	0.1431	0.0844	0.59 [0.54, 0.64]	0.1101	0.0892	0.81 [0.75, 0.88]
main factorial	binary	smooth	100	high	smooth	default	0.0043	0.0032	0.73 [0.70, 0.75]	2.1147	0.3603	0.17 [0.14, 0.20]	0.0437	0.0364	0.83 [0.79, 0.88]	0.0364	0.0342	0.94 [0.89, 0.99]
misreg.	Gaussian	misreg.	40	high	smooth	default	0.0351	0.0288	0.82 [0.79, 0.85]	0.3887	0.1711	0.44 [0.33, 0.56]	0.0042	0.0043	1.02 [0.97, 1.06]	0.0063	0.0067	1.06 [1.00, 1.13]
misreg.	Gaussian	misreg.	40	high	rough	default	0.0475	0.0431	0.91 [0.88, 0.93]	0.5156	0.3330	0.65 [0.54, 0.78]	0.0052	0.0052	1.01 [0.96, 1.06]	0.0090	0.0101	1.12 [1.06, 1.18]
misreg.	Gaussian	misreg.	100	high	smooth	default	0.0140	0.0117	0.83 [0.82, 0.85]	0.1520	0.0712	0.47 [0.39, 0.55]	0.0017	0.0016	0.98 [0.96, 1.01]	0.0025	0.0025	1.02 [1.00, 1.04]
misreg.	Gaussian	misreg.	100	high	rough	default	0.0194	0.0175	0.91 [0.89, 0.92]	0.2543	0.1733	0.68 [0.62, 0.74]	0.0022	0.0022	0.98 [0.95, 1.01]	0.0034	0.0035	1.02 [1.01, 1.03]
misreg.	Poisson	misreg.	40	high	smooth	default	12.9278	8.2514	0.64 [0.54, 0.74]									
misreg.	Poisson	misreg.	40	high	rough	default	12.8997	7.8017	0.60 [0.46, 0.80]									

Table 16: Integrated MSE of the REML and NCV fits per setting and estimand, and their paired ratio NCV/REML (95% bootstrap interval); for the smooth-effect settings the $\beta(s, t)$ columns hold $f(x, t)$. (continued)

block	family	error	G	signal	truth	basis	REML	NCV	ratio	REML	NCV	ratio	REML	NCV	ratio	REML	NCV	ratio
dense grid	binary	smooth	100	default	smooth	default	0.0073	0.0030	0.41 [0.39, 0.42]	9.5703	0.2991	0.03 [0.03, 0.04]	0.0472	0.0263	0.56 [0.52, 0.60]	0.0369	0.0251	0.68 [0.63, 0.73]
rough truth	Gaussian	indep.	100	low	rough	default	0.0330	0.0343	1.04 [1.03, 1.05]	0.2073	0.2251	1.09 [1.07, 1.11]	0.0057	0.0060	1.05 [1.03, 1.09]	0.0067	0.0070	1.05 [1.02, 1.08]
rough truth	Gaussian	smooth	100	low	rough	default	0.1931	0.1394	0.72 [0.70, 0.74]	3.1309	0.6660	0.21 [0.19, 0.25]	0.0262	0.0256	0.98 [0.94, 1.02]	0.0279	0.0281	1.01 [0.98, 1.03]
rough truth	Gaussian	indep.	100	default	rough	default	0.0100	0.0102	1.02 [1.01, 1.03]	0.0835	0.0850	1.02 [1.01, 1.03]	0.0016	0.0017	1.05 [1.02, 1.07]	0.0018	0.0019	1.06 [1.04, 1.09]
rough truth	Gaussian	smooth	100	default	rough	default	0.0531	0.0418	0.79 [0.77, 0.81]	0.9297	0.3120	0.34 [0.30, 0.38]	0.0071	0.0069	0.98 [0.96, 1.01]	0.0074	0.0075	1.02 [1.00, 1.03]
rough truth	Gaussian	indep.	100	high	rough	default	0.0030	0.0030	1.01 [1.00, 1.02]	0.0307	0.0325	1.06 [1.02, 1.12]	0.0005	0.0005	1.04 [1.02, 1.06]	0.0005	0.0005	1.05 [1.04, 1.07]
rough truth	Gaussian	smooth	100	high	rough	default	0.0146	0.0121	0.83 [0.81, 0.84]	0.2927	0.1167	0.40 [0.36, 0.44]	0.0019	0.0019	0.99 [0.97, 1.01]	0.0019	0.0019	1.01 [0.99, 1.02]
rough truth	Poisson	indep.	100	default	rough	default	0.0391	0.0398	1.02 [1.01, 1.03]	0.0267	0.0279	1.05 [1.02, 1.07]	0.0007	0.0007	1.02 [0.99, 1.05]	0.0007	0.0007	1.07 [1.04, 1.10]
rough truth	Poisson	smooth	100	default	rough	default	0.2270	0.1698	0.75 [0.72, 0.78]	0.2706	0.0933	0.34 [0.31, 0.39]	0.0027	0.0026	0.93 [0.90, 0.96]	0.0025	0.0026	1.03 [1.01, 1.05]
rough truth	Poisson	indep.	100	high	rough	default	0.1632	0.1620	0.99 [0.97, 1.02]	0.0547	0.0568	1.04 [1.01, 1.06]	0.0014	0.0014	1.05 [1.01, 1.09]	0.0011	0.0012	1.11 [1.07, 1.15]
rough truth	Poisson	smooth	100	high	rough	default	1.0961	0.7299	0.67 [0.60, 0.73]	0.4569	0.1944	0.43 [0.39, 0.47]	0.0054	0.0050	0.92 [0.89, 0.96]	0.0039	0.0042	1.09 [1.06, 1.13]
rough truth	binary	indep.	100	default	rough	default	0.0010	0.0010	1.05 [1.03, 1.07]	0.1416	0.1663	1.17 [1.11, 1.24]	0.0073	0.0079	1.08 [1.03, 1.14]	0.0101	0.0101	1.00 [0.97, 1.03]
rough truth	binary	smooth	100	default	rough	default	0.0043	0.0033	0.77 [0.74, 0.80]	1.7833	0.3877	0.22 [0.18, 0.26]	0.0309	0.0268	0.87 [0.82, 0.92]	0.0293	0.0302	1.03 [0.98, 1.09]
rough truth	binary	indep.	100	high	rough	default	0.0012	0.0012	1.05 [1.03, 1.06]	0.3097	0.3544	1.14 [1.10, 1.19]	0.0106	0.0111	1.05 [1.03, 1.08]	0.0140	0.0143	1.02 [1.00, 1.04]
rough truth	binary	smooth	100	high	rough	default	0.0046	0.0036	0.79 [0.76, 0.81]	2.3240	0.7165	0.31 [0.27, 0.35]	0.0435	0.0368	0.85 [0.80, 0.90]	0.0400	0.0393	0.98 [0.94, 1.02]
smooth effect	Gaussian	indep.	100	default	smooth	default	0.0103	0.0095	0.93 [0.91, 0.95]	0.0052	0.0039	0.73 [0.72, 0.75]				0.0016	0.0017	1.05 [1.02, 1.08]
smooth effect	Gaussian	rough	100	default	smooth	default	0.0565	0.0381	0.67 [0.64, 0.70]	0.0303	0.0155	0.51 [0.49, 0.53]				0.0068	0.0066	0.97 [0.92, 0.99]
smooth effect	Gaussian	smooth	100	default	smooth	default	0.0638	0.0419	0.66 [0.63, 0.68]	0.0346	0.0175	0.51 [0.49, 0.53]				0.0070	0.0066	0.95 [0.93, 0.98]
smooth effect	Poisson	indep.	100	default	smooth	default	0.0311	0.0291	0.93 [0.92, 0.95]	0.0019	0.0017	0.91 [0.89, 0.94]				0.0006	0.0006	1.05 [1.03, 1.08]
smooth effect	Poisson	rough	100	default	smooth	default	0.1623	0.1107	0.68 [0.66, 0.70]	0.0111	0.0066	0.60 [0.57, 0.63]				0.0022	0.0022	0.99 [0.97, 1.02]
smooth effect	Poisson	smooth	100	default	smooth	default	0.1818	0.1195	0.66 [0.64, 0.68]	0.0128	0.0073	0.57 [0.54, 0.60]				0.0023	0.0023	0.98 [0.96, 1.01]
smooth effect	binary	indep.	100	default	smooth	default	0.0007	0.0008	1.07 [1.05, 1.10]	0.0140	0.0150	1.07 [1.04, 1.11]				0.0069	0.0076	1.10 [1.06, 1.14]
smooth effect	binary	rough	100	default	smooth	default	0.0038	0.0025	0.66 [0.63, 0.68]	0.0931	0.0442	0.47 [0.44, 0.51]				0.0245	0.0243	0.99 [0.93, 1.05]
smooth effect	binary	smooth	100	default	smooth	default	0.0048	0.0028	0.59 [0.56, 0.62]	0.1257	0.0515	0.41 [0.38, 0.44]				0.0280	0.0265	0.94 [0.89, 1.01]
families	scaled t	indep.	100	default	smooth	default	0.0063	0.0060	0.95 [0.94, 0.97]	0.0336	0.0203	0.60 [0.58, 0.63]	0.0012	0.0012	1.04 [1.02, 1.07]	0.0012	0.0012	1.05 [1.02, 1.09]
families	scaled t	smooth	100	default	smooth	default	0.0336	0.0240	0.71 [0.70, 0.74]	0.4734	0.1031	0.22 [0.18, 0.26]	0.0048	0.0047	0.97 [0.95, 1.00]	0.0048	0.0046	0.95 [0.93, 0.98]
families	beta	indep.	100	default	smooth	default	0.0001	0.0001	0.98 [0.96, 0.99]	0.0100	0.0075	0.75 [0.65, 0.91]	0.0004	0.0004	1.06 [1.03, 1.09]	0.0004	0.0004	1.08 [1.05, 1.12]
families	beta	smooth	100	default	smooth	default	0.0005	0.0003	0.70 [0.68, 0.72]	0.1764	0.0322	0.18 [0.15, 0.22]	0.0017	0.0016	0.97 [0.94, 1.00]	0.0016	0.0016	0.96 [0.94, 0.99]
families	negative binomial	indep.	100	default	smooth	default	0.0570	0.0604	1.06 [1.03, 1.09]	0.0173	0.0153	0.89 [0.75, 1.05]	0.0009	0.0010	1.06 [1.03, 1.10]	0.0009	0.0009	1.10 [1.06, 1.14]
families	negative binomial	smooth	100	default	smooth	default	0.4070	0.2795	0.69 [0.64, 0.74]	0.3909	0.0685	0.18 [0.14, 0.21]	0.0042	0.0039	0.92 [0.88, 0.96]	0.0037	0.0036	0.97 [0.94, 1.00]
families	negative binomial	indep.	100	default	smooth	default	0.0683	0.0615	0.90 [0.87, 0.93]	0.0249	0.0133	0.53 [0.49, 0.59]	0.0010	0.0010	1.00 [0.97, 1.04]	0.0009	0.0010	1.04 [1.01, 1.08]
families	negative binomial	smooth	100	default	smooth	default	0.5344	0.2957	0.55 [0.52, 0.59]	0.7490	0.0647	0.09 [0.07, 0.11]	0.0045	0.0039	0.87 [0.82, 0.92]	0.0041	0.0040	0.96 [0.92, 0.99]
exact AR(1)	Gaussian	AR(1)	100	default	smooth	default	0.0516	0.0346	0.67 [0.66, 0.69]	0.7611	0.1115	0.15 [0.13, 0.17]	0.0072	0.0068	0.95 [0.93, 0.97]	0.0075	0.0071	0.95 [0.92, 0.97]
sign-changing	Gaussian	sign-chg.	100	default	smooth	default	0.0485	0.0294	0.61 [0.59, 0.62]	0.4781	0.0877	0.18 [0.17, 0.20]	0.0073	0.0065	0.89 [0.84, 0.93]	0.0073	0.0063	0.86 [0.82, 0.90]
sign-changing	Poisson	sign-chg.	100	default	smooth	default	0.2034	0.1195	0.59 [0.56, 0.62]	0.1488	0.0287	0.19 [0.18, 0.21]	0.0029	0.0023	0.79 [0.75, 0.82]	0.0025	0.0023	0.92 [0.88, 0.97]
sign-changing	binary	sign-chg.	100	default	smooth	default	0.0033	0.0020	0.61 [0.58, 0.63]	0.7819	0.2016	0.26 [0.21, 0.32]	0.0270	0.0173	0.64 [0.60, 0.68]	0.0245	0.0179	0.73 [0.69, 0.77]
misreg. (AR(1))	Gaussian	misreg.	100	high	rough	default	0.0194	0.0175	0.91 [0.89, 0.92]	0.2543	0.1733	0.68 [0.62, 0.74]	0.0022	0.0022	0.98 [0.95, 1.01]	0.0034	0.0035	1.02 [1.01, 1.03]
heteroskedastic	Gaussian	var(t)	40	default	smooth	default	0.1444	0.0865	0.60 [0.58, 0.62]	2.8124	0.3693	0.13 [0.10, 0.17]	0.0197	0.0180	0.91 [0.88, 0.95]	0.0185	0.0167	0.90 [0.86, 0.94]
heteroskedastic	Gaussian	var(t)	100	default	smooth	default	0.0504	0.0345	0.68 [0.66, 0.70]	0.7678	0.1270	0.17 [0.14, 0.20]	0.0070	0.0068	0.96 [0.94, 0.99]	0.0071	0.0066	0.94 [0.92, 0.96]
heteroskedastic	Gaussian	var(z)	40	default	smooth	default	0.1638	0.1036	0.63 [0.61, 0.66]	2.6201	0.2672	0.10 [0.08, 0.13]	0.0188	0.0176	0.93 [0.89, 0.97]	0.0372	0.0309	0.83 [0.79, 0.88]
heteroskedastic	Gaussian	var(z)	100	default	smooth	default	0.0587	0.0424	0.72 [0.70, 0.74]	0.7612	0.1238	0.16 [0.14, 0.19]	0.0070	0.0067	0.97 [0.94, 0.99]	0.0140	0.0126	0.90 [0.87, 0.92]
heteroskedastic	Gaussian	var(tz)	40	default	smooth	default	0.1649	0.1044	0.63 [0.61, 0.66]	2.5654	0.2735	0.11 [0.09, 0.13]	0.0187	0.0174	0.93 [0.89, 0.97]	0.0376	0.0313	0.83 [0.79, 0.88]
heteroskedastic	Gaussian	var(tz)	100	default	smooth	default	0.0597	0.0424	0.71 [0.69, 0.73]	0.7931	0.1203	0.15 [0.13, 0.18]	0.0069	0.0067	0.96 [0.93, 0.98]	0.0141	0.0125	0.89 [0.85, 0.92]
low-rank covariate	Gaussian	indep.	100	default	smooth	default	0.0081	0.0079	0.97 [0.96, 0.99]	0.1017	0.0724	0.71 [0.62, 0.82]	0.0016	0.0017	1.04 [1.02, 1.07]	0.0016	0.0017	1.04 [1.02, 1.07]
low-rank covariate	Gaussian	smooth	100	default	smooth	default	0.0476	0.0334	0.70 [0.68, 0.72]	3.7471	0.3194	0.09 [0.06, 0.11]	0.0070	0.0068	0.97 [0.94, 0.99]	0.0069	0.0066	0.95 [0.93, 0.97]
low-rank covariate	Poisson	indep.	100	default	smooth	default	0.0323	0.0320	0.99 [0.97, 1.01]	0.0312	0.0228	0.73 [0.63, 0.84]	0.0006	0.0007	1.03 [1.00, 1.06]	0.0006	0.0006	1.07 [1.04, 1.10]
low-rank covariate	Poisson	smooth	100	default	smooth	default	0.2097	0.1442	0.69 [0.65, 0.72]	1.1068	0.0981	0.09 [0.06, 0.12]	0.0027	0.0026	0.93 [0.90, 0.97]	0.0023	0.0023	0.98 [0.95, 1.01]
low-rank covariate	binary	indep.	100	default	smooth	default	0.0008	0.0009	1.07 [1.05, 1.09]	0.1574	0.1978	1.26 [1.03, 1.56]	0.0073	0.0077	1.06 [1.03, 1.11]	0.0075	0.0080	1.07 [1.04, 1.10]
low-rank covariate	binary	smooth	100	default	smooth	default	0.0040	0.0031	0.77 [0.74, 0.80]	7.0153	0.8475	0.12 [0.08, 0.17]	0.0305	0.0264	0.87 [0.82, 0.91]	0.0267	0.0254	0.95 [0.90, 1.01]
basis size	Gaussian	indep.	100	default	smooth	2× finer knots	0.0114	0.0102	0.89 [0.88, 0.90]	0.0678	0.0377	0.56 [0.53, 0.60]	0.0020	0.0019	0.94 [0.93, 0.96]	0.0021	0.0021	1.01 [0.98, 1.03]
basis size	Gaussian	indep.	100	default	smooth	4× finer knots	0.0138	0.0117	0.85 [0.83, 0.86]	0.0895	0.0438	0.49 [0.47, 0.51]	0.0024	0.0021	0.89 [0.87, 0.91]	0.0024	0.0023	0.96 [0.94, 0.98]
basis size	Gaussian	smooth	100	default	smooth	2× finer knots	0.0801	0.0376	0.47 [0.45, 0.49]	3.6193	0.1201	0.03 [0.03, 0.04]	0.0083	0.0071	0.85 [0.82, 0.88]	0.0083	0.0072	0.87 [0.83, 0.90]
basis size	Gaussian	smooth	100	default	smooth	4× finer knots	0.1209	0.0400	0.33 [0.32, 0.34]	11.7721	0.1482	0.01 [0.01, 0.02]	0.0090	0.0073	0.81 [0.77, 0.85]	0.0092	0.0075	0.82 [0.79, 0.85]
basis size	Poisson	indep.	100	default	smooth	2× finer knots	0.0458	0.0418	0.91 [0.89, 0.93]	0.0204	0.0113	0.55 [0.53, 0.57]	0.0008	0.0007	0.96 [0.94, 0.99]	0.0007	0.0007	1.05 [1.03, 1.07]
basis size	Poisson	indep.	100	default	smooth	4× finer knots	0.0558	0.0488	0.87 [0.86, 0.89]	0.0261	0.0134	0.52 [0.49, 0.54]	0.0009	0.0008	0.92 [0.90, 0.95]	0.0008	0.0008	1.03 [1.01, 1.06]
basis size	Poisson	smooth	100	default	smooth	2× finer knots	0.3825	0.1637	0.43 [0.40, 0.46]	1.0140	0.0349	0.03 [0.03, 0.04]	0.0032	0.0026	0.82 [0.77, 0.86]	0.0027	0.0025	0.92 [0.88, 0.95]
basis size	Poisson	smooth																

S1.6 Relative error, half-width and detection per setting

Under dependent errors the median relative error of β over settings is 0.67 for REML and 0.21 for NCV; under independent errors 0.16 and 0.12. In the main factorial the REML/NCV ratio of relative errors is 2.1–3.7 under dependent and 1.1–1.4 under independent errors. Wrong-sign exclusions of the CL2 intervals stay rare (at most 0.023 of the grid for REML + CL2 and 0.040 for NCV + CL2); model-based REML intervals reach 0.20.

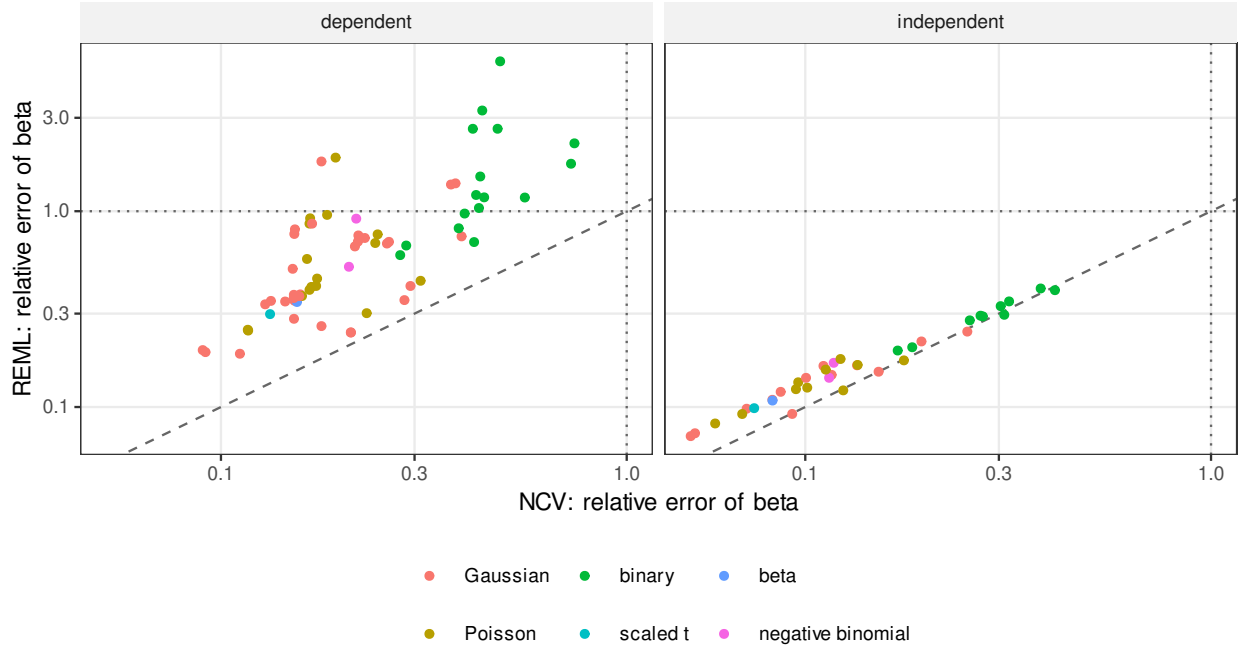


Figure 5: Relative error of the REML and NCV estimates of $\beta(s, t)$ per setting (median over replicates; log scales), by dependence. Dashed: equal error; dotted: error as large as the effect.

Table 17: Relative error and informativeness per estimand, dependence and interval (all synthetic settings): median coverage, median and maximum relative error, settings with relative error above 0.5, mean share of replicates with relative error above 1, median relative half-width (hw), median and minimum detection rate, maximum wrong-sign rate.

estimand	dependence	interval	settings	cov	rel	rel max	settings >0.5	reps >1	hw	det	det min	wrong max
E(Y X)	dependent	REML, model-based	83	0.65	0.28	1.19	18	0.04	0.27			
E(Y X)	dependent	REML + CL2	83	0.94	0.28	1.19	18	0.04	0.58			
E(Y X)	dependent	NCV + CL2	83	0.92	0.23	0.91	17	0.01	0.42			
E(Y X)	dependent	NCV + CL2, NCV-REML widening	83	0.95	0.23	0.91	17	0.01	0.53			
E(Y X)	dependent	AR(1) working model	9	0.95	0.18	0.21	0	0.00	0.38			
E(Y X)	independent	REML, model-based	40	0.97	0.11	0.49	0	0.00	0.24			
E(Y X)	independent	REML + CL2	40	0.96	0.11	0.49	0	0.00	0.24			
E(Y X)	independent	NCV + CL2	40	0.95	0.10	0.50	1	0.00	0.22			
E(Y X)	independent	NCV + CL2, NCV-REML widening	40	0.95	0.10	0.50	1	0.00	0.23			
$\beta(s, t)$	dependent	REML, model-based	69	0.67	0.67	5.83	40	0.28	0.57	0.83	0.16	0.20
$\beta(s, t)$	dependent	REML + CL2	69	0.94	0.67	5.83	40	0.28	1.35	0.54	0.04	0.02
$\beta(s, t)$	dependent	NCV + CL2	69	0.92	0.21	0.74	3	0.02	0.37	0.95	0.40	0.04
$\beta(s, t)$	dependent	NCV + CL2, NCV-REML widening	69	0.98	0.21	0.74	3	0.02	1.14	0.65	0.04	0.01
$\beta(s, t)$	dependent	AR(1) working model	9	0.99	0.19	0.21	0	0.00	0.50	0.94	0.93	0.00
$\beta(s, t)$	independent	REML, model-based	37	0.99	0.16	0.40	0	0.00	0.40	0.96	0.43	0.00
$\beta(s, t)$	independent	REML + CL2	37	0.99	0.16	0.40	0	0.00	0.40	0.96	0.47	0.00
$\beta(s, t)$	independent	NCV + CL2	37	0.98	0.12	0.41	0	0.00	0.31	0.97	0.56	0.00
$\beta(s, t)$	independent	NCV + CL2, NCV-REML widening	37	0.99	0.12	0.41	0	0.00	0.32	0.97	0.55	0.00
$\alpha(t)$	dependent	REML, model-based	69	0.66	0.31	1.60	18	0.11	0.31			
$\alpha(t)$	dependent	REML + CL2	69	0.95	0.31	1.60	18	0.11	0.65			
$\alpha(t)$	dependent	NCV + CL2	69	0.92	0.28	1.23	17	0.09	0.57			
$\alpha(t)$	dependent	NCV + CL2, NCV-REML widening	69	0.94	0.28	1.23	17	0.09	0.64			
$\alpha(t)$	dependent	AR(1) working model	9	0.95	0.23	0.24	0	0.00	0.49			
$\alpha(t)$	independent	REML, model-based	37	0.96	0.16	0.74	1	0.00	0.34			
$\alpha(t)$	independent	REML + CL2	37	0.95	0.16	0.74	1	0.00	0.35			
$\alpha(t)$	independent	NCV + CL2	37	0.94	0.16	0.74	4	0.01	0.32			
$\alpha(t)$	independent	NCV + CL2, NCV-REML widening	37	0.94	0.16	0.74	4	0.01	0.34			
$\gamma(t)$	dependent	REML, model-based	75	0.66	0.22	1.14	17	0.04	0.23	0.99	0.41	0.07
$\gamma(t)$	dependent	REML + CL2	75	0.94	0.22	1.14	17	0.04	0.46	0.94	0.15	0.01
$\gamma(t)$	dependent	NCV + CL2	75	0.92	0.22	1.00	17	0.03	0.42	0.96	0.12	0.01
$\gamma(t)$	dependent	NCV + CL2, NCV-REML widening	75	0.93	0.22	1.00	17	0.03	0.44	0.95	0.06	0.01
$\gamma(t)$	dependent	AR(1) working model	9	0.92	0.17	0.24	0	0.00	0.37	0.98	0.94	0.01
$\gamma(t)$	independent	REML, model-based	40	0.96	0.11	0.54	1	0.00	0.24	1.00	0.35	0.01
$\gamma(t)$	independent	REML + CL2	40	0.95	0.11	0.54	1	0.00	0.25	1.00	0.45	0.01
$\gamma(t)$	independent	NCV + CL2	40	0.94	0.12	0.58	1	0.00	0.23	1.00	0.44	0.01
$\gamma(t)$	independent	NCV + CL2, NCV-REML widening	40	0.94	0.12	0.58	1	0.00	0.23	1.00	0.42	0.01

Table 18: $\beta(s, t)$ per setting: relative error of the REML and NCV estimates, relative half-width (hw) and detection rate (det) of REML + CL2 and NCV + CL2.

block	family	signal	error	G	rel NCV	rel REML	hw NCV+CL2	hw REML+CL2	det NCV+CL2	det REML+CL2
main factorial	Gaussian	low	indep.	40	0.19	0.22	0.49	0.60	0.92	0.90
main factorial	Gaussian	low	indep.	100	0.13	0.16	0.34	0.44	0.96	0.95
main factorial	Gaussian	low	rough	40	0.37	1.37	0.66	2.40	0.77	0.27
main factorial	Gaussian	low	rough	100	0.26	0.68	0.47	1.36	0.89	0.49
main factorial	Gaussian	low	smooth	40	0.38	1.39	0.67	2.69	0.76	0.25
main factorial	Gaussian	low	smooth	100	0.26	0.70	0.47	1.44	0.90	0.47
main factorial	Gaussian	default	indep.	40	0.12	0.15	0.29	0.40	0.97	0.96
main factorial	Gaussian	default	indep.	100	0.08	0.11	0.20	0.29	0.99	0.98
main factorial	Gaussian	default	rough	40	0.21	0.66	0.42	1.24	0.92	0.54
main factorial	Gaussian	default	rough	100	0.15	0.35	0.29	0.71	0.97	0.80
main factorial	Gaussian	default	smooth	40	0.22	0.70	0.41	1.35	0.92	0.51
main factorial	Gaussian	default	smooth	100	0.15	0.36	0.29	0.75	0.97	0.78
main factorial	Gaussian	high	indep.	40	0.07	0.10	0.18	0.25	1.00	0.99
main factorial	Gaussian	high	indep.	100	0.05	0.07	0.12	0.18	1.00	1.00
main factorial	Gaussian	high	rough	40	0.13	0.33	0.25	0.64	0.99	0.83
main factorial	Gaussian	high	rough	100	0.09	0.19	0.17	0.39	1.00	0.95
main factorial	Gaussian	high	smooth	40	0.13	0.35	0.26	0.68	0.99	0.80
main factorial	Gaussian	high	smooth	100	0.09	0.20	0.17	0.40	1.00	0.94
main factorial	Poisson	default	indep.	40	0.13	0.16	0.34	0.45	0.96	0.95
main factorial	Poisson	default	indep.	100	0.09	0.12	0.23	0.32	0.98	0.97
main factorial	Poisson	default	rough	40	0.24	0.69	0.46	1.36	0.89	0.53
main factorial	Poisson	default	rough	100	0.17	0.40	0.32	0.80	0.96	0.77
main factorial	Poisson	default	smooth	40	0.24	0.76	0.46	1.48	0.90	0.49
main factorial	Poisson	default	smooth	100	0.17	0.41	0.32	0.84	0.96	0.75
main factorial	Poisson	high	indep.	40	0.10	0.13	0.24	0.33	0.98	0.97
main factorial	Poisson	high	indep.	100	0.07	0.09	0.16	0.23	1.00	0.99
main factorial	Poisson	high	rough	40	0.17	0.42	0.32	0.85	0.95	0.77
main factorial	Poisson	high	rough	100	0.12	0.25	0.22	0.51	0.99	0.91
main factorial	Poisson	high	smooth	40	0.17	0.45	0.33	0.89	0.95	0.73
main factorial	Poisson	high	smooth	100	0.12	0.25	0.22	0.52	0.98	0.91
main factorial	binary	default	indep.	40	0.38	0.40	0.93	0.99	0.66	0.66

Table 18: $\beta(s, t)$ per setting: relative error of the REML and NCV estimates, relative half-width (hw) and detection rate (det) of REML + CL2 and NCV + CL2. (continued)

block	family	signal	error	G	rel NCV	rel REML	hw NCV+CL2	hw REML+CL2	det NCV+CL2	det REML+CL2
main factorial	binary	default	indep.	100	0.27	0.29	0.66	0.78	0.84	0.81
main factorial	binary	default	rough	40	0.73	1.75	1.02	3.51	0.46	0.17
main factorial	binary	default	rough	100	0.43	1.04	0.77	2.16	0.68	0.31
main factorial	binary	default	smooth	40	0.74	2.22	1.04	4.16	0.40	0.15
main factorial	binary	default	smooth	100	0.45	1.18	0.79	2.41	0.66	0.27
main factorial	binary	high	indep.	40	0.25	0.28	0.61	0.74	0.86	0.83
main factorial	binary	high	indep.	100	0.18	0.20	0.44	0.55	0.93	0.92
main factorial	binary	high	rough	40	0.40	0.97	0.71	1.99	0.75	0.36
main factorial	binary	high	rough	100	0.28	0.60	0.51	1.25	0.88	0.57
main factorial	binary	high	smooth	40	0.43	1.21	0.74	2.37	0.72	0.30
main factorial	binary	high	smooth	100	0.29	0.67	0.52	1.38	0.88	0.52
misreg.	Gaussian	high	misreg.	40	0.15	0.28	0.30	0.59	0.96	0.84
misreg.	Gaussian	high	misreg.	40	0.28	0.35	0.42	0.72	0.90	0.77
misreg.	Gaussian	high	misreg.	100	0.11	0.19	0.21	0.40	0.99	0.95
misreg.	Gaussian	high	misreg.	100	0.21	0.24	0.34	0.49	0.97	0.91
dense grid	Gaussian	default	indep.	100	0.05	0.07	0.12	0.18	1.00	1.00
dense grid	Gaussian	default	rough	100	0.15	0.77	0.27	1.51	0.98	0.57
dense grid	Gaussian	default	smooth	100	0.15	0.81	0.27	1.65	0.98	0.52
dense grid	Poisson	default	indep.	100	0.06	0.08	0.14	0.20	1.00	0.99
dense grid	Poisson	default	rough	100	0.17	0.86	0.29	1.74	0.96	0.54
dense grid	Poisson	default	smooth	100	0.17	0.92	0.30	1.88	0.96	0.49
dense grid	binary	default	indep.	100	0.17	0.19	0.41	0.52	0.94	0.93
dense grid	binary	default	rough	100	0.42	2.64	0.68	5.32	0.75	0.11
dense grid	binary	default	smooth	100	0.44	3.27	0.72	6.64	0.71	0.09
rough truth	Gaussian	low	indep.	100	0.25	0.24	0.50	0.50	0.95	0.95
rough truth	Gaussian	low	smooth	100	0.39	0.74	0.54	1.47	0.81	0.42
rough truth	Gaussian	default	indep.	100	0.15	0.15	0.32	0.33	0.99	0.99
rough truth	Gaussian	default	smooth	100	0.29	0.42	0.42	0.80	0.94	0.75
rough truth	Gaussian	high	indep.	100	0.09	0.09	0.19	0.21	1.00	1.00
rough truth	Gaussian	high	smooth	100	0.18	0.26	0.31	0.48	0.99	0.95
rough truth	Poisson	default	indep.	100	0.18	0.17	0.37	0.37	0.99	0.99
rough truth	Poisson	default	smooth	100	0.31	0.44	0.43	0.90	0.91	0.72
rough truth	Poisson	high	indep.	100	0.12	0.12	0.26	0.27	0.99	0.99
rough truth	Poisson	high	smooth	100	0.23	0.30	0.36	0.61	0.97	0.90
rough truth	binary	default	indep.	100	0.41	0.40	0.76	0.80	0.73	0.74
rough truth	binary	default	smooth	100	0.56	1.17	0.80	2.50	0.58	0.24
rough truth	binary	high	indep.	100	0.31	0.30	0.61	0.60	0.88	0.89
rough truth	binary	high	smooth	100	0.42	0.70	0.57	1.46	0.78	0.47
families	scaled t	default	indep.	100	0.07	0.10	0.18	0.25	1.00	0.99
families	scaled t	default	smooth	100	0.13	0.30	0.26	0.62	0.98	0.86
families	beta	default	indep.	100	0.08	0.11	0.20	0.28	0.99	0.98
families	beta	default	smooth	100	0.15	0.34	0.28	0.72	0.97	0.80
families	negative binomial	default	indep.	100	0.11	0.14	0.28	0.38	0.97	0.96
families	negative binomial	default	smooth	100	0.21	0.52	0.39	1.08	0.93	0.62
families	negative binomial	default	indep.	100	0.12	0.17	0.25	0.40	0.98	0.96
families	negative binomial	default	smooth	100	0.22	0.92	0.37	1.81	0.93	0.46
exact AR(1)	Gaussian	default	AR(1)	100	0.16	0.38	0.29	0.76	0.97	0.77
sign-changing	Gaussian	default	sign-chg.	100	0.14	0.35	0.26	0.68	0.97	0.84
sign-changing	Poisson	default	sign-chg.	100	0.16	0.37	0.28	0.75	0.96	0.81
sign-changing	binary	default	sign-chg.	100	0.39	0.82	0.70	1.79	0.77	0.39
misreg. (AR(1))	Gaussian	high	misreg.	100	0.21	0.24	0.34	0.49	0.97	0.91
heteroskedastic	Gaussian	default	var(t)	40	0.23	0.73	0.41	1.39	0.91	0.50
heteroskedastic	Gaussian	default	var(t)	100	0.16	0.37	0.29	0.76	0.97	0.77
heteroskedastic	Gaussian	default	var(z)	40	0.22	0.75	0.41	1.46	0.92	0.52
heteroskedastic	Gaussian	default	var(z)	100	0.15	0.36	0.29	0.74	0.97	0.79
heteroskedastic	Gaussian	default	var(t,z)	40	0.22	0.74	0.41	1.39	0.92	0.52
heteroskedastic	Gaussian	default	var(t,z)	100	0.15	0.37	0.29	0.76	0.97	0.78
low-rank covariate	Gaussian	default	indep.	100	0.09	0.12	0.22	0.34	0.99	0.97
low-rank covariate	Gaussian	default	smooth	100	0.15	0.51	0.29	1.06	0.96	0.63
low-rank covariate	Poisson	default	indep.	100	0.10	0.13	0.25	0.38	0.98	0.96
low-rank covariate	Poisson	default	smooth	100	0.16	0.57	0.32	1.20	0.95	0.60
low-rank covariate	binary	default	indep.	100	0.27	0.29	0.70	0.84	0.81	0.78
low-rank covariate	binary	default	smooth	100	0.44	1.50	0.78	3.13	0.68	0.22
basis size	Gaussian	default	indep.	100	0.10	0.14	0.28	0.44	0.98	0.96
basis size	Gaussian	default	indep.	100	0.11	0.16	0.37	0.60	0.97	0.91
basis size	Gaussian	default	smooth	100	0.17	0.86	0.34	1.75	0.96	0.37
basis size	Gaussian	default	smooth	100	0.18	1.80	0.39	3.78	0.95	0.12
basis size	Poisson	default	indep.	100	0.11	0.16	0.31	0.48	0.97	0.94
basis size	Poisson	default	indep.	100	0.12	0.18	0.40	0.66	0.95	0.88
basis size	Poisson	default	smooth	100	0.18	0.96	0.37	1.94	0.95	0.35
basis size	Poisson	default	smooth	100	0.19	1.88	0.42	3.97	0.93	0.12
basis size	binary	default	indep.	100	0.30	0.33	0.85	1.04	0.71	0.64
basis size	binary	default	indep.	100	0.32	0.35	1.04	1.30	0.56	0.47
basis size	binary	default	smooth	100	0.48	2.64	0.87	5.69	0.59	0.08
basis size	binary	default	smooth	100	0.49	5.83	0.96	12.44	0.51	0.04

S1.7 Satterthwaite critical values beyond the main factorial

REML + CL2 intervals with the per-point Satterthwaite critical value of Appendix A.7 of the paper were computed from the stored REML fits for every synthetic setting outside the main factorial (110 settings, including those with 20 curves and the zero surface) and for the plasmode combinations with curve flips (S2.1). Paired with the normal critical value on the same replicates, the Satterthwaite interval never covered less: the smallest difference over all settings, plasmode combinations and estimands is 0.01 pp. Over the extension settings it adds 2.2 pp of coverage with 40 curves and 0.7 pp with 100 on average, at 1.10 and 1.04 times the width (first table). Coverage stays below 0.93 in 24 combinations of setting and estimand (second table): misregistered count and binary responses, and the binary intercept or scalar effect in the dense grid, rough truth, low-rank covariate, larger bases settings. The per-point degrees of freedom (third table) are about a third of $G - 1$ for the mean, β and γ in every synthetic setting and larger for $\alpha(t)$; on the plasmode data they spread more.

Table 19: REML + CL2 with the normal (z) and the Satterthwaite (Satt.) critical value in the extension settings: grid-averaged coverage per block, number of curves and estimand (mean over the block’s settings; β includes $f(x, t)$ of the smooth-effect settings; misregistered count and binary settings score the mean only), and the mean width ratio.

block	G	settings	$E(Y X)$ z	$E(Y X)$ Satt.	β z	β Satt.	α z	α Satt.	γ z	γ Satt.	width Satt./ z
misregistration	40	6	0.924	0.948	0.942	0.964	0.930	0.943	0.893	0.924	1.10
misregistration	100	6	0.930	0.940	0.943	0.953	0.919	0.923	0.906	0.922	1.04
dense grid	100	9	0.947	0.955	0.954	0.962	0.943	0.947	0.944	0.953	1.03
rough truth	100	14	0.944	0.953	0.948	0.957	0.945	0.948	0.936	0.946	1.04
smooth effect	100	9	0.945	0.960	0.955	0.964			0.941	0.951	1.06
extended families	100	8	0.950	0.958	0.962	0.968	0.948	0.951	0.948	0.957	1.03
exact AR(1)	100	1	0.942	0.950	0.940	0.949	0.955	0.957	0.937	0.947	1.03
sign-changing	100	3	0.942	0.950	0.947	0.956	0.944	0.948	0.940	0.950	1.03
misreg. (AR(1) fit)	100	1	0.937	0.946	0.944	0.953	0.949	0.951	0.928	0.940	1.03
heteroskedastic	40	3	0.934	0.955	0.935	0.958	0.935	0.946	0.932	0.956	1.09
heteroskedastic	100	3	0.944	0.952	0.948	0.956	0.948	0.951	0.942	0.952	1.03
low-rank covariate	100	6	0.949	0.956	0.970	0.976	0.942	0.945	0.943	0.953	1.03
larger bases	100	12	0.962	0.968	0.976	0.980	0.950	0.953	0.954	0.962	1.03

Table 20: Extension settings and estimands in which REML + CL2 with the Satterthwaite critical value covers less than 0.93: coverage with z and Satterthwaite critical value, and the 5% quantile of pointwise coverage of the Satterthwaite interval.

block	family	errors	G	signal	truth	estimand	z	Satt.	q05 Satt.
misregistration	binary	misreg.	40	high	rough	$\alpha(t)$	0.911	0.925	0.873
misregistration	binary	misreg.	40	high	smooth	$\gamma(t)$	0.897	0.928	0.885
misregistration	binary	misreg.	40	high	rough	$\gamma(t)$	0.867	0.900	0.740
misregistration	binary	misreg.	100	high	rough	$\alpha(t)$	0.913	0.917	0.818
misregistration	binary	misreg.	100	high	rough	$\gamma(t)$	0.881	0.896	0.743
misregistration	Poisson	misreg.	40	high	smooth	$\gamma(t)$	0.879	0.918	0.855
misregistration	Poisson	misreg.	40	high	rough	$\gamma(t)$	0.873	0.909	0.810
misregistration	Poisson	misreg.	40	high	rough	$E(Y X)$	0.899	0.930	0.865
misregistration	Poisson	misreg.	100	high	smooth	$\alpha(t)$	0.905	0.910	0.848
misregistration	Poisson	misreg.	100	high	rough	$\alpha(t)$	0.858	0.865	0.713
misregistration	Poisson	misreg.	100	high	smooth	$\gamma(t)$	0.894	0.915	0.840
misregistration	Poisson	misreg.	100	high	rough	$\gamma(t)$	0.885	0.906	0.802
misregistration	Poisson	misreg.	100	high	smooth	$E(Y X)$	0.910	0.924	0.850
misregistration	Poisson	misreg.	100	high	rough	$E(Y X)$	0.906	0.919	0.850
dense grid	binary	rough	100	default	smooth	$\alpha(t)$	0.919	0.925	0.887
dense grid	binary	smooth	100	default	smooth	$\alpha(t)$	0.920	0.925	0.890
rough truth	binary	indep.	100	default	rough	$\alpha(t)$	0.921	0.924	0.775
rough truth	binary	smooth	100	default	rough	$\alpha(t)$	0.920	0.925	0.875
rough truth	binary	indep.	100	default	rough	$\gamma(t)$	0.877	0.890	0.710
rough truth	binary	indep.	100	high	rough	$\gamma(t)$	0.916	0.928	0.868
low-rank covariate	binary	indep.	100	default	smooth	$\alpha(t)$	0.923	0.927	0.785
low-rank covariate	binary	smooth	100	default	smooth	$\alpha(t)$	0.922	0.926	0.885
larger bases	binary	smooth	100	default	smooth	$\alpha(t)$	0.915	0.920	0.865
larger bases	binary	smooth	100	default	smooth	$\alpha(t)$	0.899	0.904	0.853

Table 21: Per-point Satterthwaite degrees of freedom of REML + CL2: range over settings (plasmode: combinations) of the median over grid points and replicates, and that median relative to $G - 1$ (minimum, median and maximum over settings), by data source, number of curves (all: all subjects) and estimand; β includes $f(x, t)$.

data	G	estimand	settings	df min	df max	df/(G-1) min	df/(G-1) median	df/(G-1) max
plasmode	40	$E(Y X)$	60	11.43	19.11	0.29	0.35	0.49
plasmode	40	$\beta(s, t)$	60	10.65	13.52	0.27	0.31	0.35
plasmode	40	$\alpha(t)$	60	12.36	20.92	0.32	0.35	0.54
plasmode	40	$\gamma(t)$	60	7.03	25.35	0.18	0.41	0.65
plasmode	all	$E(Y X)$	60	21.45	50.84	0.28	0.35	0.51
plasmode	all	$\beta(s, t)$	60	19.81	47.94	0.20	0.30	0.35
plasmode	all	$\alpha(t)$	60	22.62	57.98	0.24	0.34	0.59
plasmode	all	$\gamma(t)$	60	13.56	96.92	0.18	0.46	0.83
synthetic	20	$E(Y X)$	21	6.40	7.52	0.34	0.35	0.40
synthetic	20	$\beta(s, t)$	21	6.00	6.94	0.32	0.33	0.37
synthetic	20	$\alpha(t)$	21	8.44	12.20	0.44	0.52	0.64
synthetic	20	$\gamma(t)$	21	6.10	6.58	0.32	0.33	0.35
synthetic	40	$E(Y X)$	22	12.68	18.25	0.33	0.36	0.47
synthetic	40	$\beta(s, t)$	22	11.36	13.77	0.29	0.34	0.35
synthetic	40	$\alpha(t)$	22	24.53	35.97	0.63	0.68	0.92
synthetic	40	$\gamma(t)$	22	11.35	13.45	0.29	0.33	0.34
synthetic	100	$E(Y X)$	85	21.85	46.26	0.22	0.36	0.47
synthetic	100	$\beta(s, t)$	85	24.32	44.31	0.25	0.33	0.45
synthetic	100	$\alpha(t)$	76	71.99	96.05	0.73	0.83	0.97
synthetic	100	$\gamma(t)$	85	23.97	34.01	0.24	0.33	0.34

S1.8 A zero coefficient surface

The zero-surface settings set $\beta \equiv 0$ in the smooth-truth setting at the default signal (Appendix B.1 of the paper). The share of grid points at which a pointwise 95% interval excludes zero estimates the pointwise false-positive rate, nominally 0.05 (first table). Under smooth errors REML + CL2 has 0.054–0.076 with z , 0.051–0.068 with t_{G-1} and 0.046–0.052 with the Satterthwaite critical value. The model-based intervals have 0.23–0.40 (REML) and 0.46–0.57 (NCV), NCV + CL2 0.11–0.15. NCV + CL2 widened by the NCV–REML difference has the lowest rate under smooth errors (0.017–0.018), but not because it is calibrated: its variance adds the squared difference between the NCV and the REML estimate, and under a zero surface the REML estimate is far from zero (second table), so its intervals are wide. Under independent errors the Gaussian setting with 40 curves stands out: REML + CL2 has 0.095 (MC SE 0.018), with the Satterthwaite critical value 0.075, against 0.050 with 100 curves. Its excluded points concentrate in a minority of replicates (0.18 have more than 5% of the grid excluding zero; the 90% quantile of the per-replicate rate is 0.24). In most replicates REML shrinks the surface towards the null space of its penalty; intervals for terms shrunk this far are known to behave poorly (Marra and Wood 2012).

The second table is descriptive. The share of replicates with at least one grid point, or with more than 5% of the grid, excluding zero is not an error rate of pointwise intervals; whether β is zero everywhere needs a simultaneous band or a global test, which the study does not evaluate. Under smooth errors REML’s estimate of the zero surface has a root-mean-square size of 0.39–2.25 times that of the non-zero surface of the corresponding main-factorial setting, NCV’s 0.06–0.45 times; under independent errors both are small (0.01–0.16). The zero surface hardly changes the other estimands (third table): REML + CL2 coverage of $\alpha(t)$ and $\gamma(t)$ differs from that in the matching non-zero setting by -0.7 to 1.6 pp.

Table 22: Zero surface: pointwise false-positive rate (MC SE), the share of grid points at which the 95% interval excludes zero, mean over replicates; nominal 0.05. NCV+CL2 widened: NCV + CL2 with the squared NCV–REML difference added to the variance.

family	errors	G	REML model	REML+CL2	REML+CL2, t(G-1)	REML+CL2, Satt.	NCV model	NCV+CL2	NCV+CL2 widened
Gaussian	smooth	40	0.399 (0.011)	0.076 (0.005)	0.068 (0.004)	0.052 (0.004)	0.516 (0.026)	0.130 (0.015)	0.017 (0.003)
Gaussian	smooth	100	0.361 (0.010)	0.058 (0.004)	0.055 (0.004)	0.049 (0.003)	0.573 (0.026)	0.151 (0.019)	0.017 (0.003)
Gaussian	indep.	40	0.024 (0.009)	0.095 (0.018)	0.086 (0.017)	0.075 (0.017)	0.066 (0.014)	0.096 (0.017)	0.079 (0.017)
Gaussian	indep.	100	0.009 (0.004)	0.050 (0.013)	0.049 (0.013)	0.046 (0.013)	0.040 (0.009)	0.047 (0.010)	0.035 (0.010)
binary	smooth	40	0.250 (0.011)	0.071 (0.005)	0.063 (0.005)	0.048 (0.004)	0.462 (0.025)	0.111 (0.014)	0.018 (0.003)
binary	smooth	100	0.228 (0.011)	0.054 (0.004)	0.051 (0.004)	0.046 (0.004)	0.476 (0.027)	0.118 (0.016)	0.018 (0.004)
binary	indep.	40	0.009 (0.004)	0.056 (0.013)	0.048 (0.012)	0.029 (0.009)	0.040 (0.009)	0.060 (0.012)	0.035 (0.010)
binary	indep.	100	0.015 (0.007)	0.042 (0.012)	0.034 (0.011)	0.031 (0.010)	0.033 (0.009)	0.048 (0.011)	0.037 (0.011)

Table 23: Zero surface, descriptive: share of replicates with at least one grid point / with more than 5% of the grid excluding zero, per interval (not error rates of pointwise intervals); root-mean-square size of the REML and NCV estimates over the grid (mean over replicates), absolute and relative to the non-zero surface of the corresponding main-factorial setting.

family	errors	G	REML+CL2	REML+CL2, Satt.	NCV+CL2	RMS REML	RMS NCV	rel. REML	rel. NCV
Gaussian	smooth	40	0.92 / 0.62	0.83 / 0.41	0.41 / 0.41	1.398	0.205	0.771	0.113
Gaussian	smooth	100	0.86 / 0.48	0.81 / 0.41	0.41 / 0.40	0.702	0.117	0.387	0.064
Gaussian	indep.	40	0.20 / 0.18	0.15 / 0.14	0.29 / 0.22	0.036	0.059	0.020	0.032
Gaussian	indep.	100	0.14 / 0.11	0.12 / 0.10	0.24 / 0.17	0.021	0.034	0.012	0.019
binary	smooth	40	0.83 / 0.54	0.76 / 0.33	0.37 / 0.35	2.043	0.410	2.255	0.452
binary	smooth	100	0.77 / 0.42	0.72 / 0.34	0.34 / 0.33	1.005	0.208	1.109	0.229
binary	indep.	40	0.14 / 0.14	0.11 / 0.09	0.27 / 0.20	0.084	0.142	0.093	0.157
binary	indep.	100	0.13 / 0.10	0.10 / 0.08	0.27 / 0.16	0.053	0.083	0.059	0.091

Table 24: Zero surface: coverage of the mean, $\alpha(t)$ and $\gamma(t)$ per interval, and of REML + CL2 in the matching non-zero setting (same α and γ ; the mean differs).

family	errors	G	estimand	REML model	REML+CL2	NCV+CL2	NCV+CL2 widened	REML+CL2, non-zero β
Gaussian	smooth	40	$E(Y X)$	0.600	0.930	0.918	0.965	
Gaussian	smooth	40	$\alpha(t)$	0.616	0.936	0.922	0.940	0.937
Gaussian	smooth	40	$\gamma(t)$	0.621	0.945	0.928	0.947	0.946
Gaussian	smooth	100	$E(Y X)$	0.624	0.943	0.926	0.967	
Gaussian	smooth	100	$\alpha(t)$	0.641	0.950	0.939	0.950	0.950
Gaussian	smooth	100	$\gamma(t)$	0.654	0.949	0.936	0.948	0.948
Gaussian	indep.	40	$E(Y X)$	0.961	0.944	0.935	0.944	
Gaussian	indep.	40	$\alpha(t)$	0.951	0.944	0.934	0.939	0.947
Gaussian	indep.	40	$\gamma(t)$	0.961	0.949	0.944	0.947	0.951
Gaussian	indep.	100	$E(Y X)$	0.961	0.948	0.941	0.949	
Gaussian	indep.	100	$\alpha(t)$	0.954	0.952	0.939	0.945	0.956
Gaussian	indep.	100	$\gamma(t)$	0.961	0.951	0.946	0.951	0.950
binary	smooth	40	$E(Y X)$	0.705	0.923	0.872	0.953	
binary	smooth	40	$\alpha(t)$	0.680	0.901	0.869	0.925	0.885
binary	smooth	40	$\gamma(t)$	0.699	0.930	0.845	0.912	0.922
binary	smooth	100	$E(Y X)$	0.716	0.938	0.873	0.943	
binary	smooth	100	$\alpha(t)$	0.714	0.925	0.860	0.917	0.920
binary	smooth	100	$\gamma(t)$	0.715	0.937	0.854	0.905	0.939
binary	indep.	40	$E(Y X)$	0.941	0.893	0.898	0.918	
binary	indep.	40	$\alpha(t)$	0.900	0.852	0.870	0.884	0.859
binary	indep.	40	$\gamma(t)$	0.938	0.888	0.878	0.897	0.890
binary	indep.	100	$E(Y X)$	0.953	0.925	0.926	0.936	
binary	indep.	100	$\alpha(t)$	0.943	0.923	0.923	0.930	0.924
binary	indep.	100	$\gamma(t)$	0.944	0.918	0.916	0.920	0.923

S1.9 A principal component term for a low-rank covariate

In the eight settings of the next table (all with smooth errors and REML fits), the functional term was also fitted with refund’s `ffpc()` term, on the same simulated data as the `ff()` fits. `ffpc()` estimates the leading functional

principal components of the covariate curves and regresses the response on their scores, with one coefficient function $b_k(t)$ per component in the basis and penalty used for $\alpha(t)$. The surface estimate is $\hat{\beta}(s, t) = \sum_{k=1}^K \psi_k(s) \hat{b}_k(t)$, where the ψ_k are the estimated eigenfunctions scaled so that the integrals of the centred covariate curves against them reproduce the scores. Its standard errors use the CL2 covariance of the b_k (full block, with $V_p - V_e$) and are conditional on the estimated eigenfunctions; the intervals use the normal critical value. The number of components K is fixed at 3 or 5, or is the smallest number that explains 95% or 90% of the covariate's variance. Accuracy is measured by the relative root-mean-square error $\{\sum_{r,j} (\hat{\beta}_{rj} - \beta_j)^2 / \sum_{r,j} \beta_j^2\}^{1/2}$, pooled over replicates r and grid points j (not the median over replicates of Section 2.4 of the paper); the $\text{ff}()$ columns are the study's REML + CL2 and NCV + CL2 on the same replicates, with the same measures.

With the low-rank covariate, the components that explain 95% of its variance (median $K = 3$) estimate β with relative RMSE 0.20, against 0.76 for $\text{ff}()$ with REML and 0.22 with NCV. Their CL2 intervals cover 0.915 (MC SE 0.005) at 0.32 times the half-width of $\text{ff}()$ REML + CL2; with five components they cover 0.948 at relative RMSE 0.42 and 0.73 times the half-width. Two components (90% of the variance) truncate the surface: they cover the true β at 0.788, its projection onto the span of the estimated components at 0.940.

With the default covariate, whose spectrum decays slowly, the variance-explained rules choose many components (median 11–11.5 for 95%, 6 for 90%), and the estimates are noisy: relative RMSE 1.0–9.7 at 95%, against 0.48–3.24 for $\text{ff}()$ with REML. A fixed $K = 3$ has 0.69–1.14 times the relative RMSE of $\text{ff}()$ with NCV, but it truncates the surface, and its intervals undercover: 0.84–0.93 for smooth truths and 0.64 for the rough truth. With $K = 5$ they cover 0.90–0.94 at 0.57–0.85 times the half-width of $\text{ff}()$ REML + CL2, with relative RMSE between those of $\text{ff}()$ with NCV and with REML in every setting. In a pilot with 20 replicates of three of the settings, refund's default threshold of 0.99 (at most 15 components) chose median K of 9–15 and gave relative RMSE 1.12–3.22 ($\text{ff}()$ with REML: 0.44–0.95), also with the low-rank covariate. The model-based intervals of the $\text{ffpc}()$ fits cover 0.34–0.68 in all settings and rules.

These results are synthetic only. The smooth truth varies along s mainly like the first cosine functions, which are the covariate's leading principal components, so a small K suits it. No rule that chooses K from the response, such as cross-validation, was evaluated, and there is no plasmode check, in which truths fitted with $\text{ff}()$ would not lie in the span of a few components.

Table 25: Principal component term (ffpc) and ff , smooth errors, REML fits, 200 replicates: relative RMSE of $\beta(s, t)$ pooled over replicates and grid (rel.) and grid-averaged coverage of nominal 95% CL2 intervals for the true $\beta(s, t)$ (cov.); ff with REML + CL2 and NCV + CL2; ffpc with a fixed number K of components or the number explaining 95% or 90% of the covariate's variance (K : median over replicates). Signal: default unless stated.

setting	ff, REML		ff, NCV		ffpc, K = 3		ffpc, K = 5		ffpc, 95%			ffpc, 90%		
	rel.	cov.	rel.	cov.	rel.	cov.	rel.	cov.	K	rel.	cov.	K	rel.	cov.
Gaussian, low-rank covariate, G = 100	0.76	0.945	0.22	0.936	0.20	0.915	0.42	0.948	3	0.20	0.915	2	0.15	0.788
Gaussian, G = 100	0.48	0.944	0.19	0.934	0.21	0.872	0.32	0.939	11.5	1.02	0.948	6	0.42	0.944
Gaussian, low signal, G = 40	1.79	0.923	0.54	0.902	0.52	0.913	0.93	0.932	11	3.70	0.929	6	1.37	0.932
Gaussian, G = 40	0.92	0.926	0.28	0.924	0.32	0.865	0.50	0.924	11	1.89	0.929	6	0.72	0.927
Gaussian, low signal, G = 20	3.24	0.909	1.07	0.839	0.74	0.892	1.42	0.914	11	9.65	0.904	6	2.23	0.913
Gaussian, G = 20	1.63	0.911	0.56	0.899	0.45	0.844	0.76	0.904	11	4.88	0.903	6	1.16	0.905
binary, G = 40	2.83	0.926	1.11	0.785	1.03	0.930	1.93	0.939	11	8.27	0.925	6	2.78	0.937
Gaussian, rough truth, G = 100	0.52	0.941	0.30	0.824	0.34	0.636	0.35	0.906	11.5	1.02	0.947	6	0.44	0.929

S2 Plasmode study

The datasets, the fitted models and the construction of the plasmode data are described in Section 5.1 and Appendices B.2 and B.3 of the paper. In a replicate each subject's error is its fitted residual curve with a random sign, so it

takes only two values: the construction is a Rademacher wild cluster bootstrap around the fitted residuals, which keeps each residual curve’s shape and within-curve covariance but not the full error distribution of new real curves. With curve flips (the main design) the simulated errors are independent across curves for every dataset, including the three whose source data have temporal or spatial dependence between curves (ocean, weather, electricity; “datasets with dependent source curves” below); only the block flips of S2.3 share one sign within a block of neighbouring curves. Coverage ranges are therefore also given over all nine datasets. All plasmode truths are fits of the model itself, so the plasmode study cannot evaluate misspecification of the mean, and the comparator results of S3 on plasmode data describe this generator, not real data in general.

S2.1 Coverage per dataset, truth source and residual source

On the datasets with independent source curves (means over each dataset’s truth $\times$ residual combinations, all subjects), model-based intervals cover β at 0.51–0.67 and the mean at 0.54–0.75; REML + CL2 covers them at 0.92–0.94 and 0.93–0.95; NCV + CL2 at 0.73–0.91 and 0.90–0.92. On the datasets with dependent source curves (curve flips, so independent simulated errors) REML + CL2 covers β at 0.88–0.93. Over all nine datasets it covers β at 0.88–0.94. The exception is weather: 0.882 on average (combinations 0.84–0.91) with all 73 stations and 0.835 (0.77–0.87) with 40, on a covariate of rank 4. With 40 subjects REML + CL2 loses 0.4–1.7 pp on the datasets with independent source curves. On the datasets with independent source curves the ratio of REML + CL2’s root mean square standard error to the replicate SD of the estimate is 0.90–1.01 (median over grid points; S4.1(a)). For a centred normal estimator a ratio r gives coverage of about $2\Phi(1.96r) - 1$, for example 0.922 at $r = 0.90$. The ratios are therefore consistent with a standard error a little short and with residual bias; the shortfall was not decomposed further. Its lower tail for β depends on the truth source: under the REML-fitted truth its 5% quantile on these datasets is 0.85–0.90, under the NCV-fitted truth 0.88–0.90.

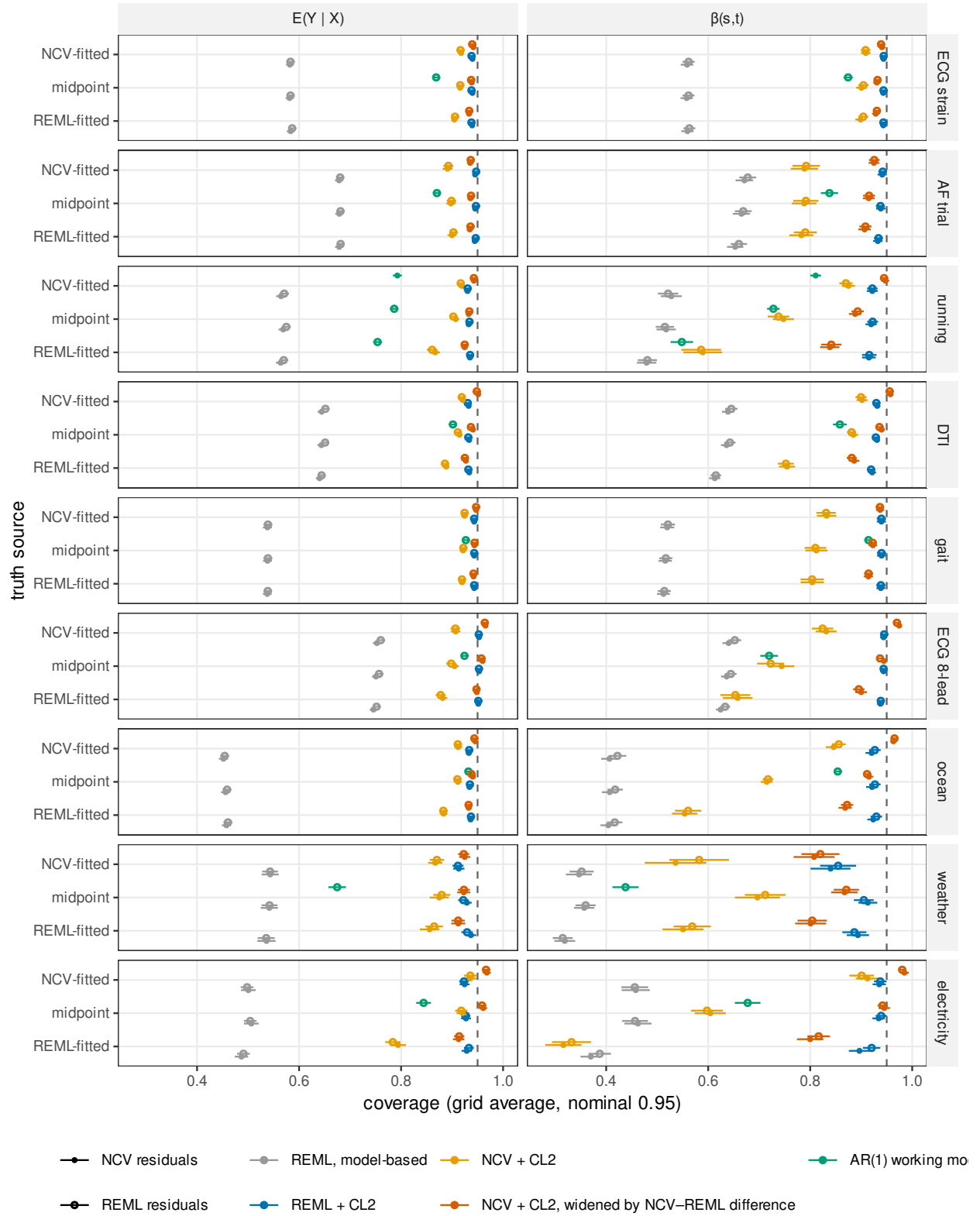


Figure 6: Plasmode coverage of the mean and of $\beta(s, t)$, all subjects, by dataset (rows), truth source (y axis), residual source (shape) and interval (colour). The AR(1) working model was fitted with the midpoint truth and REML residuals and in two further ECG strain and running combinations. Bars: ± 2 MC SE; dashed: nominal 0.95.

Table 26: 5% quantile of pointwise coverage: mean and minimum over each dataset’s combinations, all subjects.

dataset	role	estimand	q05 mean: NCV + CL2	q05 mean: NCV + CL2, NCV-REML widening	q05 mean: REML + CL2	q05 mean: REML, model-based	q05 min: NCV + CL2	q05 min: NCV + CL2, NCV-REML widening	q05 min: REML + CL2	q05 min: REML, model-based
ECG strain	indep. curves	$\beta(s, t)$	0.899	0.862	0.901	0.219	0.775	0.840	0.895	0.151
ECG strain	indep. curves	$E(Y X)$	0.820	0.876	0.888	0.238	0.765	0.850	0.880	0.190
AF trial	indep. curves	$\beta(s, t)$	0.485	0.772	0.879	0.363	0.455	0.745	0.870	0.360
AF trial	indep. curves	$E(Y X)$	0.782	0.856	0.896	0.385	0.760	0.845	0.895	0.385
running	indep. curves	$\beta(s, t)$	0.441	0.766	0.873	0.268	0.300	0.690	0.860	0.255
running	indep. curves	$E(Y X)$	0.770	0.875	0.887	0.268	0.655	0.860	0.880	0.260
DTI	indep. curves	$\beta(s, t)$	0.586	0.806	0.876	0.475	0.316	0.715	0.845	0.445
DTI	indep. curves	$E(Y X)$	0.789	0.871	0.892	0.469	0.705	0.840	0.890	0.465
gait	indep. curves	$\beta(s, t)$	0.547	0.833	0.903	0.231	0.512	0.810	0.900	0.225
gait	indep. curves	$E(Y X)$	0.833	0.895	0.910	0.243	0.830	0.895	0.910	0.240
ECG 8-lead	indep. curves	$\beta(s, t)$	0.449	0.840	0.873	0.282	0.301	0.740	0.870	0.265
ECG 8-lead	indep. curves	$E(Y X)$	0.714	0.887	0.907	0.329	0.635	0.860	0.905	0.325
ocean	dep. source	$\beta(s, t)$	0.281	0.794	0.887	0.224	0.055	0.720	0.880	0.210
ocean	dep. source	$E(Y X)$	0.815	0.891	0.899	0.234	0.755	0.880	0.895	0.230
weather	dep. source	$\beta(s, t)$	0.415	0.715	0.833	0.233	0.252	0.670	0.796	0.200
weather	dep. source	$E(Y X)$	0.728	0.869	0.865	0.342	0.662	0.785	0.835	0.335
electricity	dep. source	$\beta(s, t)$	0.398	0.835	0.881	0.266	0.075	0.695	0.820	0.230
electricity	dep. source	$E(Y X)$	0.756	0.899	0.887	0.295	0.505	0.830	0.880	0.285

Table 27: REML + CL2 for $\beta(s, t)$, all subjects, by truth source (mean over the two fitted residual sources): coverage and 5% quantile of pointwise coverage.

dataset	role	cov REML truth	cov midpoint	cov NCV truth	q05 REML truth	q05 midpoint	q05 NCV truth
ECG strain	indep. curves	0.943	0.943	0.944	0.901	0.900	0.903
AF trial	indep. curves	0.933	0.938	0.942	0.873	0.878	0.885
running	indep. curves	0.915	0.921	0.922	0.860	0.880	0.880
DTI	indep. curves	0.920	0.930	0.931	0.847	0.889	0.891
gait	indep. curves	0.939	0.940	0.940	0.903	0.905	0.900
ECG 8-lead	indep. curves	0.938	0.944	0.944	0.870	0.873	0.876
ocean	dep. source	0.927	0.924	0.924	0.888	0.890	0.883
weather	dep. source	0.890	0.909	0.847	0.830	0.860	0.808
electricity	dep. source	0.909	0.937	0.936	0.845	0.897	0.901

Table 28: Coverage with all subjects minus coverage with 40 subjects (mean of paired differences over truth $\times$ residual combinations).

dataset	role	estimand	NCV + CL2	NCV + CL2, widened by NCV-REML difference	REML + CL2	REML, model-based
ECG strain	indep. curves	$\beta(s,t)$	0.035	0.006	0.010	0.019
ECG strain	indep. curves	$E(Y \mid X)$	0.019	0.003	0.008	0.016
AF trial	indep. curves	$\beta(s,t)$	0.186	0.031	0.017	0.045
AF trial	indep. curves	$E(Y \mid X)$	0.047	0.009	0.012	0.034
running	indep. curves	$\beta(s,t)$	0.124	0.013	0.017	0.060
running	indep. curves	$E(Y \mid X)$	0.030	0.002	0.011	0.044
DTI	indep. curves	$\beta(s,t)$	0.089	0.018	0.010	0.013
DTI	indep. curves	$E(Y \mid X)$	0.044	0.013	0.014	0.019
gait	indep. curves	$\beta(s,t)$	0.185	0.020	0.004	0.048
gait	indep. curves	$E(Y \mid X)$	0.058	0.005	0.007	0.035
ECG 8-lead	indep. curves	$\beta(s,t)$	0.083	0.003	0.006	0.002
ECG 8-lead	indep. curves	$E(Y \mid X)$	0.031	0.006	0.008	0.022
ocean	dep. source	$\beta(s,t)$	0.141	0.010	-0.004	0.032
ocean	dep. source	$E(Y \mid X)$	0.024	-0.003	0.003	0.028
weather	dep. source	$\beta(s,t)$	0.213	0.059	0.047	0.038
weather	dep. source	$E(Y \mid X)$	0.041	0.007	0.008	0.038
electricity	dep. source	$\beta(s,t)$	0.016	0.008	0.014	0.072
electricity	dep. source	$E(Y \mid X)$	0.004	0.006	0.005	0.052

Satterthwaite critical values. With the per-point Satterthwaite critical value, REML + CL2 covers β at 0.939 on average over the nine datasets with all subjects and 0.940 with 40 (normal critical value: 0.927 and 0.914), at 1.05 and 1.13 times the width (next table; means over each dataset’s combinations of truth and residual source). Coverage stays below 0.93 only for $\alpha(t)$ of the gait data (0.925 with 40 subjects); β of the weather data (0.863 with 40 subjects, 0.899 with all subjects); $\gamma(t)$ of the weather data (0.921 with 40 subjects, 0.927 with all subjects); $\alpha(t)$ of the electricity data (0.879 with 40 subjects, 0.912 with all subjects). The curve flips make the simulated errors independent across curves, so these shortfalls do not come from dependence between curves. The paired gain over the normal critical value is positive in every combination (smallest 0.13 pp).

Table 29: Plasmode study, curve flips: grid-averaged coverage of REML + CL2 with the normal (z) and the Satterthwaite (Satt.) critical value per dataset and number of subjects G (mean over the dataset’s combinations of truth and residual source), and the width ratio for $\beta(s, t)$.

dataset	G	$E(Y X) z$	$E(Y X) \text{ Satt.}$	βz	$\beta \text{ Satt.}$	αz	$\alpha \text{ Satt.}$	γz	$\gamma \text{ Satt.}$	width Satt./ z
ECG strain	40	0.936	0.964	0.934	0.959	0.949	0.967	0.943	0.977	1.12
ECG strain	78	0.944	0.960	0.944	0.957	0.949	0.960	0.951	0.967	1.06
AF trial	40	0.934	0.955	0.920	0.947	0.923	0.938	0.936	0.949	1.12
AF trial	118	0.946	0.953	0.938	0.947	0.943	0.948	0.946	0.949	1.04
running	40	0.922	0.948	0.902	0.931	0.927	0.950	0.926	0.951	1.12
running	90	0.933	0.944	0.919	0.932	0.939	0.950	0.933	0.943	1.05
DTI	40	0.918	0.944	0.917	0.943	0.936	0.956	0.933	0.951	1.11
DTI	92	0.932	0.943	0.927	0.938	0.937	0.947	0.932	0.937	1.04
gait	40	0.936	0.958	0.935	0.958	0.892	0.925	0.946	0.959	1.11
gait	138	0.943	0.950	0.939	0.946	0.930	0.939	0.932	0.938	1.03
ECG 8-lead	40	0.944	0.960	0.936	0.963	0.956	0.967	0.951	0.961	1.15
ECG 8-lead	100	0.952	0.959	0.942	0.955	0.956	0.960	0.960	0.963	1.07
ocean	40	0.932	0.958	0.928	0.956	0.931	0.954	0.933	0.954	1.14
ocean	116	0.935	0.943	0.925	0.936	0.943	0.947	0.947	0.952	1.05
weather	40	0.916	0.940	0.835	0.863	0.939	0.963	0.890	0.921	1.12
weather	73	0.924	0.938	0.882	0.899	0.927	0.943	0.907	0.927	1.06
electricity	40	0.922	0.943	0.914	0.941	0.859	0.879	0.966	0.977	1.13
electricity	102	0.927	0.936	0.927	0.941	0.904	0.912	0.950	0.953	1.05

Table 30: Grid-averaged coverage (MC SE) per plasmode combination and interval, for $\beta(s, t)$ and the mean. REML model: model-based REML interval; NCV wid.: NCV + CL2 widened by the NCV–REML difference; AR(1): AR(1) working model.

dataset	truth	residual	flip	error	G	$\beta(s,t)$					$E(Y X)$					
						REML model	REML+CL2	NCV+CL2	NCV wid.	AR(1)	REML model	REML+CL2	NCV+CL2	NCV wid.	AR(1)	
ECG strain	midpoint	NCV residuals	curve	attached	78	0.558 (0.006)	0.943 (0.003)	0.900 (0.005)	0.931 (0.004)	0.874 (0.005)	0.581 (0.004)	0.940 (0.002)	0.916 (0.003)	0.938 (0.002)	0.869 (0.003)	
ECG strain	midpoint	REML residuals	curve	attached	78	0.561 (0.006)	0.944 (0.003)	0.904 (0.005)	0.932 (0.004)		0.583 (0.004)	0.938 (0.002)	0.916 (0.003)	0.937 (0.002)		
ECG strain	midpoint	beat differences (own scale)	curve	attached	78	0.534 (0.005)	0.945 (0.004)	0.933 (0.004)	0.946 (0.003)		0.570 (0.003)	0.950 (0.002)	0.940 (0.002)	0.951 (0.002)		
ECG strain	midpoint	beat differences (variance-matched)	curve	attached	78	0.524 (0.005)	0.945 (0.004)	0.910 (0.005)	0.940 (0.003)	0.837 (0.004)	0.564 (0.003)	0.950 (0.002)	0.923 (0.002)	0.947 (0.002)	0.847 (0.003)	
ECG strain	NCV-fitted	NCV residuals	curve	attached	78	0.559 (0.006)	0.944 (0.003)	0.909 (0.005)	0.941 (0.003)	0.582 (0.004)	0.940 (0.002)	0.918 (0.003)	0.942 (0.002)			
ECG strain	NCV-fitted	REML residuals	curve	attached	78	0.562 (0.006)	0.944 (0.003)	0.909 (0.005)	0.939 (0.003)	0.583 (0.004)	0.938 (0.003)	0.916 (0.003)	0.939 (0.002)			
ECG strain	NCV-fitted	beat differences (own scale)	curve	attached	78	0.532 (0.005)	0.945 (0.004)	0.928 (0.004)	0.946 (0.003)	0.569 (0.003)	0.950 (0.002)	0.935 (0.002)	0.950 (0.002)			
ECG strain	NCV-fitted	beat differences (variance-matched)	curve	attached	78	0.523 (0.005)	0.945 (0.004)	0.906 (0.005)	0.942 (0.003)	0.564 (0.003)	0.950 (0.002)	0.921 (0.002)	0.948 (0.002)			
ECG strain	REML-fitted	NCV residuals	curve	attached	78	0.559 (0.006)	0.943 (0.003)	0.899 (0.005)	0.929 (0.004)	0.583 (0.004)	0.939 (0.002)	0.904 (0.003)	0.933 (0.002)			
ECG strain	REML-fitted	REML residuals	curve	attached	78	0.563 (0.006)	0.943 (0.003)	0.904 (0.005)	0.931 (0.004)	0.586 (0.004)	0.938 (0.002)	0.906 (0.003)	0.933 (0.002)			
ECG strain	REML-fitted	beat differences (own scale)	curve	attached	78	0.536 (0.005)	0.945 (0.003)	0.932 (0.004)	0.944 (0.003)	0.579 (0.003)	0.950 (0.002)	0.940 (0.002)	0.948 (0.002)			
ECG strain	REML-fitted	beat differences (variance-matched)	curve	attached	78	0.526 (0.005)	0.944 (0.004)	0.910 (0.004)	0.939 (0.003)	0.571 (0.003)	0.949 (0.002)	0.921 (0.002)	0.945 (0.002)			
ECG strain	midpoint	NCV residuals	curve	attached	40	0.529 (0.006)	0.935 (0.004)	0.873 (0.006)	0.928 (0.004)	0.562 (0.004)	0.932 (0.002)	0.901 (0.003)	0.937 (0.002)			
ECG strain	midpoint	REML residuals	curve	attached	40	0.532 (0.006)	0.932 (0.004)	0.873 (0.006)	0.925 (0.004)	0.564 (0.004)	0.932 (0.003)	0.902 (0.003)	0.936 (0.003)			
ECG strain	midpoint	beat differences (own scale)	curve	attached	40	0.524 (0.006)	0.934 (0.004)	0.913 (0.005)	0.936 (0.004)	0.559 (0.004)	0.941 (0.002)	0.926 (0.003)	0.945 (0.002)			
ECG strain	midpoint	beat differences (variance-matched)	curve	attached	40	0.516 (0.006)	0.935 (0.004)	0.845 (0.011)	0.930 (0.004)	0.554 (0.004)	0.941 (0.003)	0.890 (0.006)	0.943 (0.002)			
ECG strain	NCV-fitted	NCV residuals	curve	attached	40	0.530 (0.006)	0.936 (0.004)	0.887 (0.006)	0.942 (0.004)	0.562 (0.004)	0.933 (0.002)	0.908 (0.003)	0.943 (0.002)			
ECG strain	NCV-fitted	REML residuals	curve	attached	40	0.533 (0.006)	0.933 (0.004)	0.887 (0.006)	0.941 (0.004)	0.564 (0.004)	0.932 (0.003)	0.908 (0.003)	0.942 (0.003)			
ECG strain	NCV-fitted	beat differences (own scale)	curve	attached	40	0.522 (0.006)	0.935 (0.004)	0.910 (0.005)	0.939 (0.004)	0.559 (0.004)	0.941 (0.002)	0.924 (0.003)	0.946 (0.002)			
ECG strain	NCV-fitted	beat differences (variance-matched)	curve	attached	40	0.515 (0.006)	0.935 (0.004)	0.854 (0.011)	0.938 (0.004)	0.554 (0.004)	0.941 (0.003)	0.896 (0.005)	0.947 (0.002)			
ECG strain	REML-fitted	NCV residuals	curve	attached	40	0.530 (0.006)	0.934 (0.004)	0.867 (0.006)	0.922 (0.004)	0.562 (0.004)	0.931 (0.002)	0.887 (0.003)	0.930 (0.002)			
ECG strain	REML-fitted	REML residuals	curve	attached	40	0.533 (0.006)	0.932 (0.004)	0.868 (0.006)	0.921 (0.004)	0.565 (0.004)	0.931 (0.003)	0.885 (0.003)	0.930 (0.003)			
ECG strain	REML-fitted	beat differences (own scale)	curve	attached	40	0.526 (0.006)	0.934 (0.004)	0.911 (0.004)	0.933 (0.004)	0.566 (0.004)	0.940 (0.002)	0.924 (0.003)	0.941 (0.002)			
ECG strain	REML-fitted	beat differences (variance-matched)	curve	attached	40	0.517 (0.006)	0.934 (0.004)	0.837 (0.012)	0.927 (0.004)	0.558 (0.004)	0.940 (0.002)	0.876 (0.006)	0.938 (0.002)			
AF trial	midpoint	NCV residuals	curve	attached	118	0.665 (0.008)	0.939 (0.005)	0.788 (0.012)	0.915 (0.006)	0.679 (0.004)	0.945 (0.002)	0.898 (0.004)	0.936 (0.003)			
AF trial	midpoint	REML residuals	curve	attached	118	0.669 (0.008)	0.938 (0.004)	0.792 (0.012)	0.915 (0.006)	0.838 (0.008)	0.681 (0.004)	0.947 (0.002)	0.899 (0.005)	0.937 (0.003)	0.870 (0.004)	
AF trial	NCV-fitted	NCV residuals	curve	attached	118	0.671 (0.009)	0.941 (0.004)	0.789 (0.013)	0.925 (0.005)	0.678 (0.004)	0.945 (0.003)	0.891 (0.005)	0.936 (0.003)			
AF trial	NCV-fitted	REML residuals	curve	attached	118	0.678 (0.008)	0.942 (0.004)	0.792 (0.013)	0.926 (0.005)	0.681 (0.004)	0.947 (0.002)	0.893 (0.005)	0.937 (0.003)			
AF trial	REML-fitted	NCV residuals	curve	attached	118	0.653 (0.008)	0.933 (0.005)	0.783 (0.012)	0.907 (0.006)	0.679 (0.004)	0.945 (0.002)	0.900 (0.004)	0.935 (0.003)			
AF trial	REML-fitted	REML residuals	curve	attached	118	0.660 (0.008)	0.934 (0.004)	0.790 (0.011)	0.908 (0.006)	0.681 (0.004)	0.946 (0.002)	0.903 (0.004)	0.936 (0.003)			
AF trial	midpoint	NCV residuals	curve	attached	40	0.620 (0.010)	0.920 (0.006)	0.596 (0.018)	0.883 (0.007)	0.644 (0.006)	0.933 (0.004)	0.849 (0.006)	0.926 (0.004)			
AF trial	midpoint	REML residuals	curve	attached	40	0.625 (0.010)	0.922 (0.006)	0.589 (0.018)	0.883 (0.007)	0.648 (0.006)	0.934 (0.004)	0.849 (0.006)	0.927 (0.004)			
AF trial	NCV-fitted	NCV residuals	curve	attached	40	0.628 (0.011)	0.925 (0.006)	0.640 (0.016)	0.904 (0.006)	0.645 (0.006)	0.934 (0.004)	0.853 (0.006)	0.930 (0.004)			
AF trial	NCV-fitted	REML residuals	curve	attached	40	0.631 (0.011)	0.926 (0.006)	0.649 (0.016)	0.906 (0.006)	0.647 (0.006)	0.935 (0.004)	0.856 (0.005)	0.931 (0.003)			
AF trial	REML-fitted	NCV residuals	curve	attached	40	0.609 (0.010)	0.913 (0.006)	0.570 (0.019)	0.868 (0.008)	0.644 (0.006)	0.932 (0.004)	0.845 (0.006)	0.924 (0.004)			
AF trial	REML-fitted	REML residuals	curve	attached	40	0.614 (0.010)	0.916 (0.006)	0.572 (0.018)	0.865 (0.008)	0.648 (0.005)	0.934 (0.004)	0.847 (0.006)	0.925 (0.004)			
running	midpoint	NCV residuals	curve	attached	90	0.518 (0.009)	0.919 (0.006)	0.747 (0.010)	0.888 (0.006)	0.569 (0.005)	0.933 (0.003)	0.907 (0.003)	0.933 (0.003)			
running	midpoint	REML residuals	curve	attached	90	0.515 (0.009)	0.922 (0.005)	0.738 (0.011)	0.893 (0.006)	0.728 (0.006)	0.575 (0.004)	0.934 (0.003)	0.902 (0.003)	0.934 (0.003)	0.787 (0.004)	
running	NCV-fitted	NCV residuals	curve	attached	90	0.528 (0.010)	0.922 (0.005)	0.875 (0.006)	0.948 (0.004)	0.811 (0.005)	0.565 (0.005)	0.930 (0.003)	0.919 (0.003)	0.943 (0.003)		0.793 (0.004)
running	NCV-fitted	NCV residuals	curve	detached	90	0.506 (0.011)	0.926 (0.006)	0.844 (0.010)	0.947 (0.005)	0.577 (0.005)	0.936 (0.003)	0.917 (0.004)	0.947 (0.003)			
running	NCV-fitted	REML residuals	curve	attached	90	0.521 (0.010)	0.922 (0.005)	0.870 (0.006)	0.945 (0.004)	0.571 (0.005)	0.931 (0.003)	0.917 (0.003)	0.943 (0.003)			
running	REML-fitted	NCV residuals	curve	attached	90	0.479 (0.010)	0.915 (0.007)	0.589 (0.019)	0.838 (0.009)	0.564 (0.005)	0.935 (0.003)	0.867 (0.005)	0.925 (0.003)			
running	REML-fitted	REML residuals	curve	attached	90	0.481 (0.009)	0.916 (0.007)	0.586 (0.020)	0.841 (0.010)	0.548 (0.011)	0.570 (0.004)	0.935 (0.003)	0.861 (0.005)	0.925 (0.003)	0.754 (0.004)	
running	REML-fitted	REML residuals	curve	detached	90	0.468 (0.008)	0.919 (0.006)	0.505 (0.018)	0.850 (0.009)	0.582 (0.005)	0.939 (0.003)	0.841 (0.006)	0.932 (0.003)			
running	midpoint	NCV residuals	curve	attached	40	0.452 (0.010)	0.903 (0.007)	0.636 (0.015)	0.880 (0.007)	0.527 (0.006)	0.922 (0.004)	0.871 (0.005)	0.930 (0.003)			
running	midpoint	REML residuals	curve	attached	40	0.467 (0.010)	0.912 (0.007)	0.638 (0.014)	0.889 (0.006)	0.531 (0.006)	0.923 (0.004)	0.871 (0.005)	0.932 (0.003)			
running	NCV-fitted	NCV residuals	curve	attached	40	0.463 (0.011)	0.910 (0.007)	0.760 (0.014)	0.927 (0.006)	0.522 (0.006)	0.921 (0.004)	0.885 (0.005)	0.937 (0.003)			
running	NCV-fitted	NCV residuals	curve	detached	40	0.453 (0.011)	0.926 (0.006)	0.747 (0.013)	0.929 (0.005)	0.533 (0.006)	0.926 (0.004)	0.888 (0.005)	0.941 (0.004)			
running	NCV-fitted	REML residuals	curve	attached	40	0.463 (0.011)	0.914 (0.006)	0.767 (0.013)	0.928 (0.005)	0.527 (0.006)	0.921 (0.004)	0.886 (0.005)	0.938 (0.003)			
running	REML-fitted	NCV residuals	curve	attached	40	0.414 (0.010)	0.888 (0.009)	0.432 (0.015)	0.823 (0.010)	0.520 (0.006)	0.922 (0.004)	0.841 (0.005)	0.924 (0.004)			
running	REML-fitted	REML residuals	curve	attached	40	0.421 (0.010)	0.887 (0.009)	0.427 (0.015)	0.829 (0.011)	0.522 (0.005)	0.921 (0.004)	0.837 (0.006)	0.925 (0.004)			
running	REML-fitted	REML residuals	curve	detached	40	0.422 (0.009)	0.909 (0.007)	0.412 (0.014)	0.836 (0.009)	0.536 (0.005)	0.928 (0.004)	0.839 (0.005)	0.929 (0.004)			
DTI	midpoint	NCV residuals	curve	attached	92	0.636 (0.005)	0.931 (0.003)	0.885 (0.005)	0.940 (0.003)	0.644 (0.004)	0.933 (0.002)	0.914 (0.003)	0.941 (0.002)			
DTI	midpoint	REML residuals	curve	attached	92	0.643 (0.005)	0.929 (0.003)	0.881 (0.005)	0.936 (0.003)	0.858 (0.007)	0.651 (0.004)	0.932 (0.002)	0.911 (0.003)	0.937 (0.002)	0.902 (0.004)	
DTI	NCV-fitted	NCV residuals	curve	attached	92	0.639 (0.006)	0.931 (0.003)	0.902 (0.005)	0.957 (0.003)	0.644 (0.004)	0.933 (0.002)	0.921 (0.003)	0.951 (0.002)			
DTI	NCV-fitted	REML residuals	curve	attached	92	0.645 (0.006)	0.930 (0.003)	0.900 (0.005)	0.956 (0.003)	0.652 (0.004)	0.931 (0.002)	0.919 (0.003)	0.948 (0.002)			
DTI	REML-fitted	NCV residuals	curve	attached	92	0.613 (0.005)	0.922									

Table 30: Grid-averaged coverage (MC SE) per plasmode combination and interval, for $\beta(s, t)$ and the mean. REML model: model-based REML interval; NCV wid.: NCV + CL2 widened by the NCV–REML difference; AR(1): AR(1) working model. (continued)

dataset	truth	residual	flip	error	G	REML model	REML+CL2	NCV+CL2	NCV wid.	AR(1)	REML model	REML+CL2	NCV+CL2	NCV wid.	AR(1)
gait	midpoint	NCV residuals	curve	attached	40	0.469 (0.006)	0.936 (0.004)	0.626 (0.014)	0.904 (0.004)		0.503 (0.004)	0.936 (0.003)	0.863 (0.006)	0.939 (0.003)	
gait	midpoint	REML residuals	curve	attached	40	0.470 (0.006)	0.936 (0.004)	0.625 (0.013)	0.905 (0.004)		0.504 (0.004)	0.936 (0.003)	0.864 (0.006)	0.940 (0.003)	
gait	NCV-fitted	NCV residuals	curve	attached	40	0.470 (0.007)	0.935 (0.004)	0.672 (0.014)	0.915 (0.004)		0.503 (0.004)	0.936 (0.003)	0.868 (0.006)	0.940 (0.003)	
gait	NCV-fitted	REML residuals	curve	attached	40	0.471 (0.007)	0.936 (0.004)	0.673 (0.014)	0.917 (0.004)		0.503 (0.004)	0.936 (0.003)	0.869 (0.006)	0.941 (0.003)	
gait	REML-fitted	NCV residuals	curve	attached	40	0.465 (0.006)	0.934 (0.004)	0.595 (0.013)	0.893 (0.004)		0.503 (0.004)	0.936 (0.003)	0.860 (0.006)	0.937 (0.003)	
gait	REML-fitted	REML residuals	curve	attached	40	0.466 (0.006)	0.935 (0.004)	0.595 (0.013)	0.893 (0.004)		0.504 (0.004)	0.936 (0.003)	0.860 (0.006)	0.938 (0.003)	
ECG 8-lead	midpoint	NCV residuals	curve	attached	100	0.636 (0.006)	0.944 (0.003)	0.744 (0.012)	0.945 (0.003)		0.751 (0.004)	0.953 (0.002)	0.905 (0.004)	0.960 (0.002)	
ECG 8-lead	midpoint	REML residuals	curve	attached	100	0.644 (0.006)	0.944 (0.003)	0.722 (0.013)	0.937 (0.004)	0.720 (0.009)	0.757 (0.004)	0.953 (0.002)	0.898 (0.005)	0.957 (0.002)	0.924 (0.003)
ECG 8-lead	NCV-fitted	NCV residuals	curve	attached	100	0.640 (0.006)	0.944 (0.003)	0.831 (0.010)	0.974 (0.002)		0.753 (0.004)	0.953 (0.002)	0.907 (0.005)	0.965 (0.002)	
ECG 8-lead	NCV-fitted	REML residuals	curve	attached	100	0.652 (0.006)	0.945 (0.003)	0.824 (0.010)	0.970 (0.003)		0.760 (0.004)	0.952 (0.002)	0.906 (0.005)	0.964 (0.002)	
ECG 8-lead	REML-fitted	NCV residuals	curve	attached	100	0.624 (0.005)	0.938 (0.003)	0.658 (0.014)	0.900 (0.006)		0.746 (0.004)	0.951 (0.002)	0.882 (0.005)	0.948 (0.002)	
ECG 8-lead	REML-fitted	REML residuals	curve	attached	100	0.633 (0.005)	0.938 (0.004)	0.653 (0.015)	0.895 (0.006)		0.752 (0.004)	0.951 (0.002)	0.878 (0.005)	0.948 (0.002)	
ECG 8-lead	midpoint	NCV residuals	curve	attached	40	0.639 (0.006)	0.937 (0.004)	0.637 (0.014)	0.940 (0.003)		0.729 (0.004)	0.944 (0.002)	0.863 (0.005)	0.951 (0.002)	
ECG 8-lead	midpoint	REML residuals	curve	attached	40	0.641 (0.006)	0.937 (0.004)	0.646 (0.014)	0.936 (0.004)		0.734 (0.005)	0.944 (0.003)	0.867 (0.005)	0.950 (0.003)	
ECG 8-lead	NCV-fitted	NCV residuals	curve	attached	40	0.640 (0.007)	0.936 (0.004)	0.770 (0.012)	0.965 (0.003)		0.730 (0.005)	0.944 (0.003)	0.881 (0.005)	0.959 (0.002)	
ECG 8-lead	NCV-fitted	REML residuals	curve	attached	40	0.645 (0.007)	0.937 (0.004)	0.774 (0.012)	0.965 (0.003)		0.736 (0.005)	0.944 (0.003)	0.884 (0.005)	0.957 (0.002)	
ECG 8-lead	REML-fitted	NCV residuals	curve	attached	40	0.625 (0.005)	0.934 (0.004)	0.550 (0.015)	0.900 (0.004)		0.724 (0.004)	0.944 (0.003)	0.846 (0.006)	0.945 (0.002)	
ECG 8-lead	REML-fitted	REML residuals	curve	attached	40	0.631 (0.005)	0.934 (0.004)	0.562 (0.016)	0.895 (0.005)		0.730 (0.004)	0.944 (0.003)	0.849 (0.006)	0.943 (0.003)	
ocean	midpoint	NCV residuals	curve	attached	116	0.407 (0.008)	0.921 (0.006)	0.715 (0.006)	0.915 (0.004)		0.456 (0.004)	0.934 (0.003)	0.911 (0.004)	0.941 (0.003)	
ocean	midpoint	REML residuals	curve	attached	116	0.418 (0.008)	0.927 (0.006)	0.717 (0.005)	0.912 (0.004)	0.854 (0.003)	0.459 (0.004)	0.935 (0.003)	0.910 (0.004)	0.938 (0.003)	0.932 (0.003)
ocean	NCV-fitted	NCV residuals	curve	attached	116	0.406 (0.008)	0.920 (0.006)	0.846 (0.008)	0.963 (0.003)		0.452 (0.004)	0.933 (0.003)	0.912 (0.004)	0.945 (0.003)	
ocean	NCV-fitted	NCV residuals	block	attached	116	0.420 (0.009)	0.934 (0.006)	0.855 (0.007)	0.966 (0.002)		0.448 (0.004)	0.931 (0.003)	0.902 (0.004)	0.937 (0.003)	
ocean	NCV-fitted	REML residuals	curve	attached	116	0.422 (0.009)	0.927 (0.006)	0.856 (0.007)	0.965 (0.002)		0.454 (0.004)	0.934 (0.003)	0.911 (0.004)	0.944 (0.003)	
ocean	REML-fitted	NCV residuals	curve	attached	116	0.404 (0.008)	0.924 (0.006)	0.554 (0.012)	0.869 (0.007)		0.458 (0.004)	0.936 (0.003)	0.883 (0.004)	0.933 (0.003)	
ocean	REML-fitted	REML residuals	curve	attached	116	0.417 (0.008)	0.930 (0.006)	0.561 (0.013)	0.872 (0.006)		0.461 (0.004)	0.937 (0.003)	0.883 (0.004)	0.932 (0.003)	
ocean	REML-fitted	REML residuals	block	attached	116	0.417 (0.007)	0.935 (0.006)	0.564 (0.014)	0.883 (0.006)		0.456 (0.004)	0.933 (0.003)	0.875 (0.004)	0.929 (0.003)	
ocean	midpoint	NCV residuals	curve	attached	40	0.376 (0.007)	0.927 (0.005)	0.585 (0.014)	0.914 (0.004)		0.427 (0.004)	0.931 (0.003)	0.882 (0.004)	0.942 (0.002)	
ocean	midpoint	REML residuals	curve	attached	40	0.387 (0.007)	0.933 (0.005)	0.589 (0.014)	0.914 (0.004)		0.432 (0.004)	0.933 (0.003)	0.882 (0.004)	0.942 (0.002)	
ocean	NCV-fitted	NCV residuals	curve	attached	40	0.375 (0.008)	0.929 (0.005)	0.689 (0.012)	0.944 (0.003)		0.425 (0.004)	0.931 (0.003)	0.885 (0.004)	0.946 (0.003)	
ocean	NCV-fitted	REML residuals	curve	attached	40	0.388 (0.008)	0.934 (0.005)	0.700 (0.012)	0.946 (0.003)		0.430 (0.004)	0.933 (0.003)	0.886 (0.004)	0.945 (0.003)	
ocean	REML-fitted	NCV residuals	curve	attached	40	0.375 (0.007)	0.922 (0.005)	0.416 (0.012)	0.860 (0.005)		0.428 (0.004)	0.932 (0.003)	0.867 (0.004)	0.940 (0.002)	
ocean	REML-fitted	REML residuals	curve	attached	40	0.383 (0.007)	0.926 (0.005)	0.420 (0.012)	0.858 (0.005)		0.432 (0.004)	0.934 (0.003)	0.868 (0.004)	0.939 (0.002)	
weather	midpoint	NCV residuals	curve	attached	73	0.357 (0.010)	0.913 (0.009)	0.697 (0.022)	0.867 (0.013)		0.542 (0.008)	0.929 (0.005)	0.875 (0.009)	0.923 (0.006)	
weather	midpoint	REML residuals	curve	attached	73	0.360 (0.010)	0.905 (0.010)	0.712 (0.020)	0.871 (0.012)	0.438 (0.013)	0.542 (0.008)	0.922 (0.005)	0.880 (0.008)	0.923 (0.006)	0.675 (0.009)
weather	NCV-fitted	NCV residuals	curve	attached	73	0.347 (0.013)	0.840 (0.020)	0.536 (0.030)	0.808 (0.020)		0.543 (0.008)	0.913 (0.006)	0.867 (0.007)	0.925 (0.005)	
weather	NCV-fitted	NCV residuals	block	attached	73	0.369 (0.011)	0.896 (0.016)	0.540 (0.031)	0.831 (0.017)		0.489 (0.006)	0.897 (0.004)	0.820 (0.008)	0.899 (0.005)	
weather	NCV-fitted	REML residuals	curve	attached	73	0.352 (0.012)	0.855 (0.018)	0.582 (0.029)	0.820 (0.019)		0.543 (0.008)	0.912 (0.005)	0.870 (0.007)	0.923 (0.005)	
weather	REML-fitted	NCV residuals	curve	attached	73	0.319 (0.010)	0.893 (0.011)	0.551 (0.020)	0.801 (0.015)		0.536 (0.009)	0.937 (0.005)	0.856 (0.010)	0.912 (0.007)	
weather	REML-fitted	REML residuals	curve	attached	73	0.315 (0.010)	0.887 (0.012)	0.569 (0.018)	0.804 (0.014)		0.536 (0.008)	0.929 (0.006)	0.865 (0.009)	0.912 (0.006)	
weather	REML-fitted	REML residuals	block	attached	73	0.350 (0.013)	0.911 (0.009)	0.605 (0.018)	0.814 (0.014)		0.522 (0.007)	0.914 (0.004)	0.842 (0.008)	0.898 (0.005)	
weather	midpoint	NCV residuals	curve	attached	40	0.310 (0.012)	0.866 (0.018)	0.454 (0.030)	0.795 (0.019)		0.496 (0.008)	0.918 (0.006)	0.818 (0.009)	0.909 (0.006)	
weather	midpoint	REML residuals	curve	attached	40	0.327 (0.012)	0.863 (0.017)	0.453 (0.030)	0.806 (0.018)		0.504 (0.008)	0.912 (0.006)	0.820 (0.009)	0.909 (0.007)	
weather	NCV-fitted	NCV residuals	curve	attached	40	0.298 (0.013)	0.798 (0.022)	0.357 (0.028)	0.741 (0.023)		0.505 (0.008)	0.915 (0.005)	0.859 (0.007)	0.923 (0.005)	
weather	NCV-fitted	REML residuals	curve	attached	40	0.305 (0.014)	0.769 (0.023)	0.351 (0.028)	0.730 (0.023)		0.511 (0.008)	0.907 (0.006)	0.855 (0.007)	0.919 (0.006)	
weather	REML-fitted	NCV residuals	curve	attached	40	0.288 (0.011)	0.861 (0.018)	0.386 (0.026)	0.769 (0.019)		0.496 (0.008)	0.925 (0.006)	0.809 (0.009)	0.908 (0.007)	
weather	REML-fitted	REML residuals	curve	attached	40	0.295 (0.011)	0.853 (0.017)	0.366 (0.025)	0.776 (0.018)		0.502 (0.008)	0.921 (0.006)	0.806 (0.009)	0.910 (0.006)	
electricity	midpoint	NCV residuals	curve	attached	102	0.462 (0.013)	0.934 (0.006)	0.604 (0.015)	0.945 (0.006)		0.506 (0.007)	0.927 (0.005)	0.921 (0.006)	0.961 (0.004)	
electricity	midpoint	REML residuals	curve	attached	102	0.457 (0.013)	0.939 (0.005)	0.598 (0.016)	0.942 (0.005)	0.677 (0.013)	0.504 (0.006)	0.927 (0.005)	0.918 (0.006)	0.959 (0.004)	0.844 (0.007)
electricity	NCV-fitted	NCV residuals	curve	attached	102	0.458 (0.014)	0.935 (0.007)	0.912 (0.011)	0.985 (0.004)		0.501 (0.007)	0.925 (0.005)	0.935 (0.006)	0.968 (0.004)	
electricity	NCV-fitted	NCV residuals	block	attached	102	0.452 (0.013)	0.935 (0.006)	0.915 (0.011)	0.991 (0.002)		0.490 (0.007)	0.926 (0.004)	0.925 (0.007)	0.966 (0.004)	
electricity	NCV-fitted	REML residuals	curve	attached	102	0.456 (0.013)	0.937 (0.006)	0.901 (0.012)	0.980 (0.004)		0.498 (0.006)	0.923 (0.005)	0.936 (0.006)	0.967 (0.004)	
electricity	REML-fitted	NCV residuals	curve	attached	102	0.370 (0.010)	0.897 (0.010)	0.316 (0.018)	0.800 (0.013)		0.487 (0.007)	0.929 (0.005)	0.794 (0.008)	0.913 (0.006)	
electricity	REML-fitted	REML residuals	curve	attached	102	0.387 (0.011)	0.920 (0.009)	0.332 (0.019)	0.817 (0.011)		0.491 (0.006)	0.933 (0.005)	0.784 (0.007)	0.914 (0.005)	
electricity	REML-fitted	REML residuals	block	attached	102	0.390 (0.010)	0.905 (0.011)	0.308 (0.020)	0.799 (0.014)		0.493 (0.007)	0.935 (0.005)	0.773 (0.008)	0.911 (0.005)	
electricity	midpoint	NCV residuals	curve	attached	40	0.385 (0.012)	0.936 (0.007)	0.672 (0.011)	0.967 (0.004)		0.451 (0.007)	0.924 (0.005)	0.905 (0.007)	0.953 (0.005)	
electricity	midpoint	REML residuals	curve	attached	40	0.378 (0.011)	0.932 (0.007)	0.656 (0.012)	0.959 (0.005)		0.455 (0.008)	0.923 (0.005)	0.908 (0.007)	0.953 (0.005)	
electricity	NCV-fitted	NCV residuals	curve	attached	40	0.381 (0.013)	0.936 (0.006)	0.869 (0.016)	0.982 (0.004)		0.446 (0.007)	0.923 (0.005)	0.910 (0.008)	0.957 (0.005)	
electricity	NCV-fitted	REML residuals	curve	attached	40	0.386 (0.012)	0.938 (0.006)	0.880 (0.014)	0.982 (0.004)		0.452 (0.008)	0.923 (0.005)	0.915 (0.007)	0.957 (0.005)	
electricity	REML-fitted	NCV residuals	curve	attached	40	0.312 (0.009)	0.861 (0.012)	0.252 (0.016)	0.752 (0.014)		0.433 (0.006)	0.917 (0.005)	0.816 (0.007)	0.911 (0.006)	
electricity	REML-fitted	REML residuals	curve	attached	40	0.319 (0.009)	0.878 (0.010)	0.239 (0.015)	0.778 (0.012)		0.438 (0.007)	0.924 (0.005)	0.808 (0.007)	0.915 (0.005)	

S2.2 Estimation by truth source; relative error and detection

The NCV/REML MSE ratio for β (geometric mean over residual sources, all subjects) is 0.04–0.78 under the NCV-fitted truth, 0.05–0.90 under the midpoint truth and 0.32–2.24 under the REML-fitted truth; NCV loses under the REML-fitted truth on 4 of 9 datasets (electricity, DTI, ocean, running). A truth taken from one fit has that fit’s smoothness, so each truth source favours its own selector. The midpoint truth does not remove this: its fit lies between the REML and NCV fits in EDF for every term on only 3 of 9 datasets, and its surface EDF exceeds both fits’ on running, ocean, weather, where it is the roughest of the three truths. The relative error of β is large for both fits on several datasets (next table): β is weakly determined by these covariates.

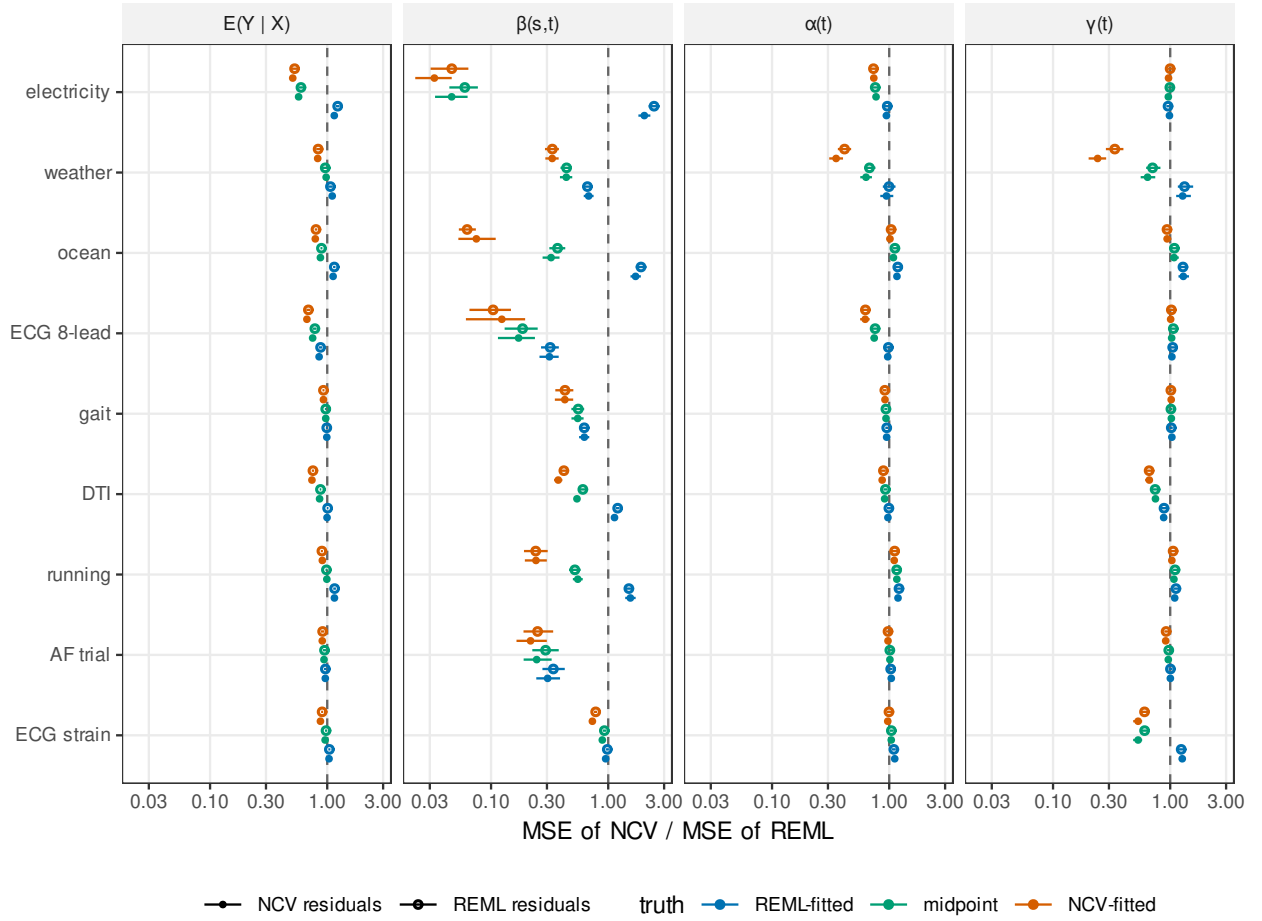


Figure 7: NCV/REML MSE ratio per dataset, truth source (colour) and residual source (shape), all subjects, with 95% paired bootstrap intervals; log scale.

Table 31: NCV/REML MSE ratio per dataset and truth source (geometric mean over residual sources, all subjects).

role	dataset	estimand	REML	midpoint	NCV
indep. curves	ECG strain	$\alpha(t)$	1.071	1.029	0.980
indep. curves	AF trial	$\alpha(t)$	1.035	1.013	0.977
indep. curves	running	$\alpha(t)$	1.201	1.160	1.108
indep. curves	DTI	$\alpha(t)$	0.984	0.921	0.881
indep. curves	gait	$\alpha(t)$	0.952	0.938	0.919
indep. curves	ECG 8-lead	$\alpha(t)$	0.977	0.751	0.625
dep. source	ocean	$\alpha(t)$	1.174	1.102	1.026
dep. source	weather	$\alpha(t)$	0.971	0.656	0.383
dep. source	electricity	$\alpha(t)$	0.955	0.766	0.736
indep. curves	ECG strain	$\beta(s,t)$	0.952	0.904	0.780
indep. curves	AF trial	$\beta(s,t)$	0.321	0.268	0.233
indep. curves	running	$\beta(s,t)$	1.523	0.534	0.241
indep. curves	DTI	$\beta(s,t)$	1.167	0.573	0.396
indep. curves	gait	$\beta(s,t)$	0.625	0.552	0.427
indep. curves	ECG 8-lead	$\beta(s,t)$	0.317	0.179	0.113
dep. source	ocean	$\beta(s,t)$	1.806	0.347	0.068
dep. source	weather	$\beta(s,t)$	0.672	0.440	0.333
dep. source	electricity	$\beta(s,t)$	2.237	0.052	0.039
indep. curves	ECG strain	$\gamma(t)$	1.184	0.634	0.628
indep. curves	AF trial	$\gamma(t)$	1.004	0.968	0.919
indep. curves	running	$\gamma(t)$	1.106	1.088	1.046
indep. curves	DTI	$\gamma(t)$	0.883	0.746	0.662
indep. curves	gait	$\gamma(t)$	1.028	1.018	1.017
indep. curves	ECG 8-lead	$\gamma(t)$	1.042	1.046	1.016
dep. source	ocean	$\gamma(t)$	1.290	1.082	0.941
dep. source	weather	$\gamma(t)$	1.301	0.673	0.284
dep. source	electricity	$\gamma(t)$	0.972	0.979	0.983
indep. curves	ECG strain	$E(Y X)$	1.016	0.966	0.907
indep. curves	AF trial	$E(Y X)$	0.964	0.943	0.910
indep. curves	running	$E(Y X)$	1.154	0.988	0.904
indep. curves	DTI	$E(Y X)$	1.002	0.868	0.748
indep. curves	gait	$E(Y X)$	0.991	0.971	0.928
indep. curves	ECG 8-lead	$E(Y X)$	0.865	0.767	0.681
dep. source	ocean	$E(Y X)$	1.136	0.884	0.796
dep. source	weather	$E(Y X)$	1.084	0.970	0.833
dep. source	electricity	$E(Y X)$	1.188	0.583	0.517

Table 32: $\beta(s, t)$, all subjects: coverage, median relative error of the estimate, median relative half-width (hw) and median detection rate (det) per interval and dataset (over truth $\times$ residual combinations); NCV–REML widening: NCV + CL2 widened by the NCV–REML difference (S4.1).

role	dataset	cov NCV+CL2	cov NCV-REML widening	cov REML+CL2	rel NCV+CL2	rel REML+CL2	hw NCV+CL2	hw NCV-REML widening	hw REML+CL2	det NCV+CL2	det NCV-REML widening	det REML+CL2
indep. curves	ECG strain	0.912	0.938	0.944	0.676	0.698	1.177	1.30	1.42	0.461	0.404	0.442
indep. curves	AF trial	0.789	0.916	0.938	0.714	0.865	0.963	1.53	1.82	0.533	0.303	0.372
indep. curves	running	0.734	0.892	0.919	0.661	0.825	0.678	1.52	1.85	0.735	0.578	0.637
indep. curves	DTI	0.846	0.926	0.927	0.469	0.633	0.760	1.08	1.17	0.772	0.562	0.638
indep. curves	gait	0.816	0.925	0.939	0.765	1.007	1.024	1.93	2.39	0.650	0.486	0.546
indep. curves	ECG 8-lead	0.739	0.937	0.942	0.806	2.427	0.851	4.37	4.88	0.568	0.210	0.216
dep. source	ocean	0.708	0.916	0.925	0.667	1.035	0.527	2.11	2.25	0.528	0.178	0.450
dep. source	weather	0.608	0.828	0.882	0.678	1.007	0.771	1.79	1.98	0.538	0.228	0.321
dep. source	electricity	0.611	0.911	0.927	0.815	3.418	0.564	5.62	5.72	0.512	0.094	0.076

Table 33: NCV/REML MSE ratio per plasmode combination and estimand (95% paired bootstrap intervals).

dataset	truth	residual	flip	error	G	alpha	beta	gamma	mean
ECG strain	midpoint	NCV residuals	curve	attached	78	1.04 [1.00, 1.09]	0.89 [0.86, 0.92]	0.53 [0.48, 0.58]	0.96 [0.94, 0.98]
ECG strain	midpoint	REML residuals	curve	attached	78	1.04 [1.00, 1.09]	0.93 [0.90, 0.96]	0.60 [0.56, 0.65]	0.98 [0.96, 1.00]
ECG strain	midpoint	beat differences (own scale)	curve	attached	78	1.01 [0.99, 1.04]	0.94 [0.93, 0.96]	0.72 [0.68, 0.75]	0.97 [0.97, 0.98]
ECG strain	midpoint	beat differences (variance-matched)	curve	attached	78	1.02 [0.98, 1.07]	0.86 [0.83, 0.88]	0.70 [0.66, 0.73]	0.95 [0.94, 0.97]
ECG strain	NCV-fitted	NCV residuals	curve	attached	78	0.97 [0.93, 1.02]	0.73 [0.71, 0.76]	0.53 [0.48, 0.58]	0.87 [0.86, 0.89]
ECG strain	NCV-fitted	REML residuals	curve	attached	78	1.00 [0.95, 1.04]	0.78 [0.76, 0.80]	0.60 [0.56, 0.64]	0.90 [0.89, 0.92]
ECG strain	NCV-fitted	beat differences (own scale)	curve	attached	78	1.00 [0.97, 1.04]	0.89 [0.87, 0.91]	0.70 [0.67, 0.73]	0.96 [0.95, 0.98]
ECG strain	NCV-fitted	beat differences (variance-matched)	curve	attached	78	0.95 [0.90, 1.00]	0.73 [0.71, 0.75]	0.69 [0.65, 0.72]	0.89 [0.87, 0.90]
ECG strain	REML-fitted	NCV residuals	curve	attached	78	1.11 [1.07, 1.16]	0.95 [0.92, 0.98]	1.27 [1.19, 1.36]	1.04 [1.01, 1.06]
ECG strain	REML-fitted	REML residuals	curve	attached	78	1.10 [1.06, 1.14]	0.98 [0.95, 1.01]	1.24 [1.17, 1.33]	1.04 [1.02, 1.06]
ECG strain	REML-fitted	beat differences (own scale)	curve	attached	78	1.01 [0.99, 1.03]	0.97 [0.96, 0.98]	1.05 [1.04, 1.06]	0.99 [0.99, 1.00]
ECG strain	REML-fitted	beat differences (variance-matched)	curve	attached	78	1.07 [1.02, 1.12]	0.91 [0.88, 0.93]	1.20 [1.13, 1.27]	0.99 [0.98, 1.01]
ECG strain	midpoint	NCV residuals	curve	attached	40	0.95 [0.90, 1.01]	0.74 [0.71, 0.77]	0.54 [0.49, 0.59]	0.87 [0.85, 0.90]
ECG strain	midpoint	REML residuals	curve	attached	40	0.96 [0.91, 1.01]	0.78 [0.75, 0.81]	0.58 [0.53, 0.64]	0.89 [0.87, 0.92]
ECG strain	midpoint	beat differences (own scale)	curve	attached	40	0.99 [0.94, 1.05]	0.89 [0.86, 0.91]	0.69 [0.64, 0.73]	0.95 [0.93, 0.97]
ECG strain	midpoint	beat differences (variance-matched)	curve	attached	40	0.94 [0.86, 1.03]	0.69 [0.66, 0.73]	0.68 [0.63, 0.73]	0.93 [0.88, 0.97]
ECG strain	NCV-fitted	NCV residuals	curve	attached	40	0.83 [0.79, 0.89]	0.59 [0.56, 0.62]	0.53 [0.47, 0.58]	0.77 [0.75, 0.80]
ECG strain	NCV-fitted	REML residuals	curve	attached	40	0.85 [0.80, 0.91]	0.62 [0.59, 0.65]	0.59 [0.53, 0.64]	0.80 [0.77, 0.82]
ECG strain	NCV-fitted	beat differences (own scale)	curve	attached	40	0.93 [0.87, 0.99]	0.77 [0.75, 0.80]	0.68 [0.64, 0.72]	0.89 [0.87, 0.91]
ECG strain	NCV-fitted	beat differences (variance-matched)	curve	attached	40	0.82 [0.75, 0.90]	0.56 [0.53, 0.59]	0.67 [0.62, 0.71]	0.81 [0.77, 0.85]
ECG strain	REML-fitted	NCV residuals	curve	attached	40	1.05 [1.00, 1.11]	0.81 [0.78, 0.84]	1.05 [0.99, 1.11]	0.95 [0.92, 0.97]

Table 33: NCV/REML MSE ratio per plasmode combination and estimand (95% paired bootstrap intervals). (*continued*)

dataset	truth	residual	flip	error	G	alpha	beta	gamma	mean
ECG strain	REML-fitted	REML residuals	curve	attached	40	1.07 [1.02, 1.14]	0.85 [0.82, 0.88]	1.13 [1.07, 1.20]	0.97 [0.95, 1.00]
ECG strain	REML-fitted	beat differences (own scale)	curve	attached	40	1.03 [0.98, 1.09]	0.94 [0.91, 0.97]	1.14 [1.09, 1.21]	1.00 [0.97, 1.02]
ECG strain	REML-fitted	beat differences (variance-matched)	curve	attached	40	1.07 [0.98, 1.17]	0.75 [0.71, 0.79]	1.27 [1.20, 1.36]	1.01 [0.97, 1.06]
AF trial	midpoint	NCV residuals	curve	attached	118	1.01 [0.97, 1.06]	0.25 [0.19, 0.33]	0.96 [0.93, 1.00]	0.94 [0.91, 0.97]
AF trial	midpoint	REML residuals	curve	attached	118	1.01 [0.97, 1.06]	0.29 [0.22, 0.38]	0.97 [0.93, 1.01]	0.95 [0.92, 0.98]
AF trial	NCV-fitted	NCV residuals	curve	attached	118	0.98 [0.94, 1.02]	0.22 [0.16, 0.30]	0.92 [0.88, 0.95]	0.91 [0.88, 0.94]
AF trial	NCV-fitted	REML residuals	curve	attached	118	0.98 [0.94, 1.02]	0.25 [0.19, 0.34]	0.92 [0.89, 0.96]	0.91 [0.88, 0.94]
AF trial	REML-fitted	NCV residuals	curve	attached	118	1.04 [1.00, 1.09]	0.30 [0.24, 0.39]	1.00 [0.97, 1.04]	0.96 [0.93, 0.99]
AF trial	REML-fitted	REML residuals	curve	attached	118	1.03 [0.99, 1.07]	0.34 [0.27, 0.43]	1.01 [0.97, 1.05]	0.96 [0.94, 0.99]
AF trial	midpoint	NCV residuals	curve	attached	40	0.97 [0.90, 1.04]	0.17 [0.10, 0.28]	0.84 [0.78, 0.89]	0.78 [0.74, 0.82]
AF trial	midpoint	REML residuals	curve	attached	40	0.97 [0.91, 1.05]	0.19 [0.11, 0.31]	0.84 [0.79, 0.90]	0.79 [0.75, 0.83]
AF trial	NCV-fitted	NCV residuals	curve	attached	40	0.93 [0.87, 1.01]	0.12 [0.07, 0.21]	0.81 [0.76, 0.86]	0.72 [0.68, 0.76]
AF trial	NCV-fitted	REML residuals	curve	attached	40	0.93 [0.88, 1.00]	0.13 [0.08, 0.23]	0.81 [0.76, 0.87]	0.73 [0.70, 0.77]
AF trial	REML-fitted	NCV residuals	curve	attached	40	0.99 [0.92, 1.06]	0.19 [0.12, 0.29]	0.86 [0.81, 0.92]	0.81 [0.77, 0.85]
AF trial	REML-fitted	REML residuals	curve	attached	40	1.01 [0.95, 1.10]	0.20 [0.13, 0.31]	0.87 [0.82, 0.93]	0.82 [0.79, 0.86]
running	midpoint	NCV residuals	curve	attached	90	1.16 [1.10, 1.23]	0.55 [0.50, 0.61]	1.08 [1.04, 1.12]	0.99 [0.98, 1.01]
running	midpoint	REML residuals	curve	attached	90	1.16 [1.09, 1.23]	0.52 [0.46, 0.58]	1.10 [1.06, 1.15]	0.98 [0.97, 1.00]
running	NCV-fitted	NCV residuals	curve	attached	90	1.10 [1.04, 1.17]	0.24 [0.19, 0.30]	1.03 [0.99, 1.08]	0.91 [0.89, 0.92]
running	NCV-fitted	NCV residuals	curve	detached	90	1.06 [1.00, 1.13]	0.15 [0.12, 0.20]	1.09 [1.03, 1.17]	0.87 [0.85, 0.90]
running	NCV-fitted	REML residuals	curve	attached	90	1.11 [1.05, 1.18]	0.24 [0.19, 0.30]	1.06 [1.02, 1.12]	0.90 [0.88, 0.92]
running	REML-fitted	NCV residuals	curve	attached	90	1.19 [1.12, 1.26]	1.55 [1.40, 1.71]	1.09 [1.05, 1.13]	1.15 [1.12, 1.18]
running	REML-fitted	REML residuals	curve	attached	90	1.21 [1.13, 1.29]	1.50 [1.37, 1.65]	1.12 [1.07, 1.17]	1.16 [1.13, 1.19]
running	REML-fitted	REML residuals	curve	detached	90	1.14 [1.07, 1.23]	1.35 [1.22, 1.51]	1.17 [1.09, 1.25]	1.20 [1.17, 1.24]
running	midpoint	NCV residuals	curve	attached	40	1.09 [1.02, 1.18]	0.31 [0.25, 0.38]	0.99 [0.92, 1.05]	0.92 [0.88, 0.96]
running	midpoint	REML residuals	curve	attached	40	1.11 [1.03, 1.21]	0.29 [0.24, 0.36]	0.99 [0.92, 1.06]	0.91 [0.87, 0.95]
running	NCV-fitted	NCV residuals	curve	attached	40	1.00 [0.93, 1.09]	0.15 [0.11, 0.20]	0.92 [0.87, 0.99]	0.83 [0.79, 0.86]
running	NCV-fitted	NCV residuals	curve	detached	40	1.00 [0.92, 1.08]	0.14 [0.10, 0.19]	0.82 [0.77, 0.89]	0.77 [0.74, 0.81]
running	NCV-fitted	REML residuals	curve	attached	40	1.02 [0.95, 1.10]	0.15 [0.11, 0.20]	0.93 [0.87, 0.99]	0.82 [0.79, 0.86]
running	REML-fitted	NCV residuals	curve	attached	40	1.13 [1.05, 1.24]	0.82 [0.72, 0.93]	0.99 [0.93, 1.06]	1.01 [0.97, 1.06]
running	REML-fitted	REML residuals	curve	attached	40	1.15 [1.06, 1.27]	0.80 [0.71, 0.91]	1.00 [0.93, 1.08]	1.01 [0.97, 1.05]
running	REML-fitted	REML residuals	curve	detached	40	1.10 [1.02, 1.20]	0.73 [0.65, 0.83]	0.97 [0.90, 1.06]	0.97 [0.94, 1.01]
DTI	midpoint	NCV residuals	curve	attached	92	0.91 [0.88, 0.95]	0.54 [0.51, 0.57]	0.75 [0.70, 0.79]	0.86 [0.85, 0.87]
DTI	midpoint	REML residuals	curve	attached	92	0.93 [0.89, 0.97]	0.61 [0.57, 0.64]	0.74 [0.70, 0.78]	0.88 [0.86, 0.89]
DTI	NCV-fitted	NCV residuals	curve	attached	92	0.87 [0.83, 0.91]	0.37 [0.35, 0.41]	0.66 [0.61, 0.71]	0.74 [0.72, 0.76]
DTI	NCV-fitted	REML residuals	curve	attached	92	0.89 [0.85, 0.93]	0.42 [0.39, 0.45]	0.66 [0.61, 0.70]	0.76 [0.74, 0.77]
DTI	REML-fitted	NCV residuals	curve	attached	92	0.98 [0.94, 1.01]	1.13 [1.08, 1.18]	0.88 [0.83, 0.93]	1.00 [0.98, 1.01]
DTI	REML-fitted	REML residuals	curve	attached	92	0.99 [0.96, 1.03]	1.20 [1.16, 1.25]	0.89 [0.84, 0.93]	1.01 [0.99, 1.02]
DTI	midpoint	NCV residuals	curve	attached	40	0.91 [0.86, 0.97]	0.49 [0.46, 0.53]	0.68 [0.62, 0.74]	0.81 [0.78, 0.84]
DTI	midpoint	REML residuals	curve	attached	40	0.91 [0.86, 0.95]	0.53 [0.49, 0.57]	0.68 [0.63, 0.73]	0.82 [0.80, 0.85]
DTI	NCV-fitted	NCV residuals	curve	attached	40	0.92 [0.86, 0.98]	0.37 [0.33, 0.40]	0.63 [0.57, 0.68]	0.71 [0.68, 0.73]
DTI	NCV-fitted	REML residuals	curve	attached	40	0.91 [0.85, 0.97]	0.40 [0.37, 0.44]	0.64 [0.59, 0.70]	0.73 [0.71, 0.76]
DTI	REML-fitted	NCV residuals	curve	attached	40	0.91 [0.86, 0.97]	0.84 [0.80, 0.89]	0.75 [0.70, 0.81]	0.87 [0.85, 0.90]
DTI	REML-fitted	REML residuals	curve	attached	40	0.92 [0.86, 0.97]	0.90 [0.85, 0.94]	0.75 [0.70, 0.81]	0.88 [0.86, 0.91]
gait	midpoint	NCV residuals	curve	attached	138	0.94 [0.91, 0.97]	0.55 [0.49, 0.62]	1.02 [0.99, 1.06]	0.97 [0.95, 0.99]
gait	midpoint	REML residuals	curve	attached	138	0.94 [0.91, 0.97]	0.55 [0.49, 0.62]	1.01 [0.98, 1.05]	0.97 [0.95, 0.99]
gait	NCV-fitted	NCV residuals	curve	attached	138	0.92 [0.88, 0.96]	0.43 [0.35, 0.50]	1.02 [0.98, 1.06]	0.93 [0.91, 0.95]
gait	NCV-fitted	REML residuals	curve	attached	138	0.92 [0.88, 0.95]	0.43 [0.35, 0.50]	1.01 [0.98, 1.05]	0.93 [0.91, 0.95]
gait	REML-fitted	NCV residuals	curve	attached	138	0.95 [0.92, 0.98]	0.63 [0.56, 0.69]	1.03 [1.00, 1.07]	0.99 [0.97, 1.01]
gait	REML-fitted	REML residuals	curve	attached	138	0.95 [0.93, 0.98]	0.62 [0.56, 0.69]	1.02 [0.99, 1.05]	0.99 [0.97, 1.01]
gait	midpoint	NCV residuals	curve	attached	40	0.86 [0.80, 0.92]	0.26 [0.20, 0.33]	0.88 [0.83, 0.95]	0.91 [0.87, 0.94]
gait	midpoint	REML residuals	curve	attached	40	0.86 [0.80, 0.92]	0.25 [0.19, 0.32]	0.89 [0.83, 0.96]	0.91 [0.87, 0.94]
gait	NCV-fitted	NCV residuals	curve	attached	40	0.85 [0.79, 0.91]	0.22 [0.16, 0.29]	0.87 [0.81, 0.93]	0.88 [0.85, 0.92]
gait	NCV-fitted	REML residuals	curve	attached	40	0.85 [0.79, 0.91]	0.22 [0.15, 0.29]	0.87 [0.81, 0.94]	0.88 [0.85, 0.92]
gait	REML-fitted	NCV residuals	curve	attached	40	0.86 [0.80, 0.92]	0.29 [0.23, 0.36]	0.89 [0.83, 0.96]	0.92 [0.89, 0.95]
gait	REML-fitted	REML residuals	curve	attached	40	0.87 [0.81, 0.93]	0.29 [0.23, 0.36]	0.89 [0.84, 0.97]	0.92 [0.89, 0.96]
ECG 8-lead	midpoint	NCV residuals	curve	attached	100	0.74 [0.70, 0.80]	0.17 [0.11, 0.24]	1.03 [0.98, 1.08]	0.75 [0.73, 0.78]
ECG 8-lead	midpoint	REML residuals	curve	attached	100	0.76 [0.71, 0.81]	0.19 [0.13, 0.25]	1.06 [1.02, 1.12]	0.78 [0.75, 0.81]
ECG 8-lead	NCV-fitted	NCV residuals	curve	attached	100	0.62 [0.56, 0.68]	0.12 [0.06, 0.20]	1.01 [0.95, 1.07]	0.67 [0.64, 0.70]
ECG 8-lead	NCV-fitted	REML residuals	curve	attached	100	0.63 [0.57, 0.68]	0.10 [0.07, 0.15]	1.02 [0.97, 1.09]	0.69 [0.66, 0.72]
ECG 8-lead	REML-fitted	NCV residuals	curve	attached	100	0.97 [0.91, 1.04]	0.31 [0.26, 0.38]	1.03 [0.99, 1.08]	0.85 [0.83, 0.88]
ECG 8-lead	REML-fitted	REML residuals	curve	attached	100	0.98 [0.92, 1.06]	0.32 [0.27, 0.38]	1.05 [1.01, 1.10]	0.88 [0.85, 0.91]
ECG 8-lead	midpoint	NCV residuals	curve	attached	40	0.67 [0.61, 0.74]	0.10 [0.07, 0.13]	1.00 [0.93, 1.07]	0.66 [0.62, 0.69]
ECG 8-lead	midpoint	REML residuals	curve	attached	40	0.68 [0.62, 0.75]	0.12 [0.08, 0.16]	0.98 [0.91, 1.05]	0.68 [0.64, 0.71]
ECG 8-lead	NCV-fitted	NCV residuals	curve	attached	40	0.57 [0.51, 0.64]	0.08 [0.05, 0.12]	0.93 [0.87, 1.01]	0.58 [0.55, 0.61]
ECG 8-lead	NCV-fitted	REML residuals	curve	attached	40	0.58 [0.52, 0.64]	0.09 [0.06, 0.13]	0.93 [0.86, 1.00]	0.60 [0.56, 0.63]
ECG 8-lead	REML-fitted	NCV residuals	curve	attached	40	0.83 [0.76, 0.92]	0.14 [0.11, 0.18]	0.99 [0.92, 1.05]	0.71 [0.68, 0.74]
ECG 8-lead	REML-fitted	REML residuals	curve	attached	40	0.83 [0.76, 0.91]	0.18 [0.14, 0.23]	1.00 [0.93, 1.07]	0.74 [0.71, 0.77]
ocean	midpoint	NCV residuals	curve	attached	116	1.09 [1.04, 1.13]	0.33 [0.28, 0.38]	1.08 [1.00, 1.18]	0.87 [0.86, 0.89]
ocean	midpoint	REML residuals	curve	attached	116	1.12 [1.07, 1.17]	0.37 [0.31, 0.43]	1.08 [1.00, 1.18]	0.89 [0.88, 0.91]
ocean	NCV-fitted	NCV residuals	curve	attached	116	1.01 [0.95, 1.07]	0.07 [0.05, 0.11]	0.94 [0.88, 1.02]	0.79 [0.77, 0.81]
ocean	NCV-fitted	NCV residuals	block	attached	116	1.00 [0.95, 1.05]	0.05 [0.04, 0.06]	0.84 [0.79, 0.89]	0.81 [0.79, 0.83]
ocean	NCV-fitted	REML residuals	curve	attached	116	1.04 [0.98, 1.09]	0.06 [0.05, 0.07]	0.94 [0.88, 1.02]	0.80 [0.78, 0.82]
ocean	REML-fitted	NCV residuals	curve	attached	116	1.16 [1.10, 1.23]	1.71 [1.55, 1.90]	1.29 [1.18, 1.45]	1.12 [1.10, 1.15]
ocean	REML-fitted	REML residuals	curve	attached	116	1.18 [1.12, 1.25]	1.90 [1.72, 2.13]	1.29 [1.17, 1.42]	1.15 [1.12, 1.18]
ocean	REML-fitted	REML residuals	block	attached	116	1.16 [1.09, 1.24]	1.82 [1.62, 2.07]	1.22 [1.11, 1.33]	1.15 [1.12, 1.18]
ocean	midpoint	NCV residuals	curve	attached	40	0.90 [0.83, 0.97]	0.13 [0.11, 0.15]	0.76 [0.70, 0.83]	0.76 [0.74, 0.79]
ocean	midpoint	REML residuals	curve	attached	40	0.93 [0.86, 1.00]	0.13 [0.11, 0.15]	0.76 [0.70, 0.82]	0.77 [0.75, 0.80]
ocean	NCV-fitted	NCV residuals	curve	attached	40	0.85 [0.79, 0.92]	0.04 [0.03, 0.06]	0.64 [0.59, 0.69]	0.70 [0.67, 0.73]
ocean	NCV-fitted	REML residuals	curve	attached	40	0.88 [0.81, 0.95]	0.04 [0.03, 0.05]	0.65 [0.60, 0.69]	0.71 [0.68, 0.74]
ocean	REML-fitted	NCV residuals	curve	attached	40	0.94 [0.87, 1.02]	0.64 [0.57, 0.72]	0.86 [0.78, 0.95]	0.85 [0.82, 0.88]
ocean	REML-fitted	REML residuals	curve	attached	40	0.97 [0.90, 1.05]	0.69 [0.61, 0.78]	0.86 [0.79, 0.95]	0.87 [0.83, 0.90]
weather	midpoint	NCV residuals	curve	attached	73	0.63 [0.57, 0.71]	0.44 [0.39, 0.49]	0.64 [0.56, 0.75]	0.98 [0.93, 1.03]
weather	midpoint	REML residuals	curve	attached	73	0.68 [0.61, 0.76]	0.44 [0.39, 0.49]	0.71 [0.62, 0.83]	0.96 [0.92, 1.01]
weather	NCV-fitted	NCV residuals	curve	attached	73	0.35 [0.31, 0.40]	0.33 [0.29, 0.38]	0.24 [0.20, 0.28]	0.83 [0.80, 0.87]
weather	NCV-fitted	NCV residuals	block	attached	73	0.32 [0.28, 0.37]	0.41 [0.36, 0.48]	0.20 [0.17, 0.24]	0.86 [0.83, 0.90]
weather	NCV-fitted	REML residuals	curve	attached	73	0.42 [0.36, 0.47]	0.33 [0.29, 0.38]	0.34 [0.28, 0.40]	0.84 [0.80, 0.87]
weather	REML-fitted	NCV residuals	curve	attached	73	0.95 [0.84, 1.08]	0.68 [0.62, 0.75]	1.28 [1.13, 1.50]	1.10 [1.05, 1.16]
weather	REML-fitted	REML residuals	curve	attached	73	1.00 [0.88, 1.13]	0.66 [0.60, 0.73]	1.32 [1.16, 1.57]	1.07 [1.02, 1.11]
weather	REML-fitted	REML residuals	block	attached	73	0.82 [0.71, 0.94]	0.83 [0.75, 0.90]	1.16 [1.02, 1.31]	1.10 [1.06, 1.14]
weather	midpoint	NCV residuals	curve	attached	40	0.50 [0.41, 0.61]	0.32 [0.28, 0.38]	0.61 [0.49, 0.76]	0.90 [0.85, 0.95]

Table 33: NCV/REML MSE ratio per plasmode combination and estimand (95% paired bootstrap intervals). (*continued*)

dataset	truth	residual	flip	error	G	alpha	beta	gamma	mean
weather	midpoint	REML residuals	curve	attached	40	0.59 [0.49, 0.69]	0.35 [0.31, 0.41]	0.75 [0.61, 0.94]	0.93 [0.88, 0.98]
weather	NCV-fitted	NCV residuals	curve	attached	40	0.30 [0.24, 0.38]	0.21 [0.17, 0.25]	0.32 [0.25, 0.40]	0.71 [0.67, 0.75]
weather	NCV-fitted	REML residuals	curve	attached	40	0.38 [0.31, 0.46]	0.22 [0.19, 0.27]	0.39 [0.31, 0.49]	0.74 [0.70, 0.78]
weather	REML-fitted	NCV residuals	curve	attached	40	0.64 [0.52, 0.78]	0.44 [0.39, 0.51]	1.00 [0.80, 1.23]	0.97 [0.91, 1.02]
weather	REML-fitted	REML residuals	curve	attached	40	0.77 [0.63, 0.91]	0.45 [0.39, 0.51]	1.18 [0.97, 1.44]	0.99 [0.94, 1.05]
electricity	midpoint	NCV residuals	curve	attached	102	0.77 [0.73, 0.82]	0.05 [0.03, 0.06]	0.96 [0.92, 1.01]	0.57 [0.54, 0.60]
electricity	midpoint	REML residuals	curve	attached	102	0.76 [0.72, 0.80]	0.06 [0.04, 0.08]	0.99 [0.94, 1.04]	0.60 [0.57, 0.62]
electricity	NCV-fitted	NCV residuals	curve	attached	102	0.74 [0.70, 0.78]	0.03 [0.02, 0.05]	0.97 [0.92, 1.02]	0.51 [0.48, 0.54]
electricity	NCV-fitted	NCV residuals	block	attached	102	0.74 [0.70, 0.78]	0.03 [0.02, 0.04]	0.95 [0.90, 0.99]	0.52 [0.49, 0.55]
electricity	NCV-fitted	REML residuals	curve	attached	102	0.73 [0.69, 0.77]	0.05 [0.03, 0.06]	1.00 [0.95, 1.05]	0.53 [0.50, 0.56]
electricity	REML-fitted	NCV residuals	curve	attached	102	0.95 [0.88, 1.01]	2.03 [1.81, 2.29]	0.99 [0.95, 1.02]	1.15 [1.10, 1.19]
electricity	REML-fitted	REML residuals	curve	attached	102	0.96 [0.90, 1.02]	2.46 [2.21, 2.76]	0.96 [0.93, 0.99]	1.23 [1.18, 1.28]
electricity	REML-fitted	REML residuals	block	attached	102	0.94 [0.89, 1.00]	2.38 [2.12, 2.71]	0.97 [0.94, 1.01]	1.26 [1.21, 1.32]
electricity	midpoint	NCV residuals	curve	attached	40	0.76 [0.71, 0.82]	0.03 [0.02, 0.04]	0.88 [0.81, 0.95]	0.49 [0.45, 0.53]
electricity	midpoint	REML residuals	curve	attached	40	0.78 [0.73, 0.85]	0.04 [0.02, 0.06]	0.89 [0.82, 0.96]	0.49 [0.45, 0.53]
electricity	NCV-fitted	NCV residuals	curve	attached	40	0.75 [0.69, 0.81]	0.02 [0.02, 0.04]	0.89 [0.82, 0.96]	0.46 [0.43, 0.50]
electricity	NCV-fitted	REML residuals	curve	attached	40	0.76 [0.71, 0.82]	0.04 [0.02, 0.06]	0.89 [0.82, 0.96]	0.46 [0.43, 0.50]
electricity	REML-fitted	NCV residuals	curve	attached	40	0.87 [0.81, 0.94]	0.84 [0.75, 0.94]	0.87 [0.81, 0.94]	0.76 [0.72, 0.80]
electricity	REML-fitted	REML residuals	curve	attached	40	0.82 [0.76, 0.89]	0.98 [0.88, 1.10]	0.86 [0.79, 0.93]	0.78 [0.74, 0.81]

S2.3 Truth source against residual source; heartbeat errors; block flips

On the 2×2 of REML/NCV truths and REML/NCV residuals (all subjects), the truth source shifts the log NCV/REML MSE ratio for β by -4.05 to -0.24 (NCV-fitted minus REML-fitted truth), the residual source by -0.27 to 0.08: the truth source dominates. With heartbeat difference curves as errors, REML + CL2's coverage changes by 0.1–1.2 pp (variance-matched) and 0.1–1.2 pp (own scale) relative to the REML residuals. For the three datasets with dependent source curves, block flips share one sign within each block of neighbouring curves, so the simulated errors are independent only across blocks (ocean 39 blocks, weather 16 blocks, electricity 51 blocks). Block flips change REML + curve-clustered CL2 coverage by -1.8 to 2.1 pp relative to curve flips; block-wise NCV changes the MSE by a factor 0.99–1.11.

Table 34: Truth-source and residual-source main effects and their interaction (paired, 95% bootstrap intervals) on the log NCV/REML MSE ratio and on coverage. Residual verdict: small = interval inside ± 2 points or log[0.95, 1.05].

dataset	estimand	quantity	truth	residual	interaction	residual verdict
ECG strain	$\beta(s,t)$	coverage NCV + CL2, NCV-REML widening	0.010 [0.007, 0.013]	-0.000 [-0.003, 0.002]	0.004 [0.002, 0.006]	small
AF trial	$\beta(s,t)$	coverage NCV + CL2, NCV-REML widening	0.018 [0.009, 0.026]	-0.001 [-0.004, 0.002]	0.001 [-0.004, 0.005]	small
running	$\beta(s,t)$	coverage NCV + CL2, NCV-REML widening	0.106 [0.090, 0.123]	-0.000 [-0.006, 0.005]	0.006 [-0.006, 0.017]	small
DTI	$\beta(s,t)$	coverage NCV + CL2, NCV-REML widening	0.073 [0.063, 0.082]	0.003 [0.000, 0.006]	-0.003 [-0.008, 0.002]	small
gait	$\beta(s,t)$	coverage NCV + CL2, NCV-REML widening	0.023 [0.017, 0.028]	0.000 [-0.002, 0.002]	0.001 [-0.002, 0.003]	small
ECG 8-lead	$\beta(s,t)$	coverage NCV + CL2, NCV-REML widening	0.075 [0.064, 0.085]	0.005 [0.001, 0.008]	-0.000 [-0.006, 0.006]	small
ocean	$\beta(s,t)$	coverage NCV + CL2, NCV-REML widening	0.094 [0.081, 0.106]	-0.003 [-0.007, 0.001]	0.002 [-0.007, 0.011]	small
weather	$\beta(s,t)$	coverage NCV + CL2, NCV-REML widening	0.011 [-0.015, 0.036]	-0.008 [-0.019, 0.004]	-0.009 [-0.029, 0.010]	small
electricity	$\beta(s,t)$	coverage NCV + CL2, NCV-REML widening	0.174 [0.153, 0.197]	-0.006 [-0.014, 0.002]	0.021 [0.005, 0.037]	small
ECG strain	E(Y X)	coverage NCV + CL2, NCV-REML widening	0.007 [0.006, 0.009]	0.001 [-0.001, 0.003]	0.003 [0.001, 0.004]	small
AF trial	E(Y X)	coverage NCV + CL2, NCV-REML widening	0.001 [-0.002, 0.003]	-0.001 [-0.002, 0.000]	0.000 [-0.001, 0.002]	small
running	E(Y X)	coverage NCV + CL2, NCV-REML widening	0.018 [0.014, 0.023]	0.000 [-0.002, 0.002]	0.000 [-0.003, 0.003]	small
DTI	E(Y X)	coverage NCV + CL2, NCV-REML widening	0.024 [0.021, 0.027]	0.002 [0.001, 0.004]	0.001 [-0.001, 0.003]	small
gait	E(Y X)	coverage NCV + CL2, NCV-REML widening	0.005 [0.004, 0.006]	0.000 [-0.000, 0.001]	0.000 [-0.001, 0.001]	small
ECG 8-lead	E(Y X)	coverage NCV + CL2, NCV-REML widening	0.017 [0.013, 0.020]	0.001 [-0.000, 0.003]	0.001 [-0.001, 0.003]	small
ocean	E(Y X)	coverage NCV + CL2, NCV-REML widening	0.012 [0.008, 0.016]	0.001 [-0.001, 0.003]	0.002 [-0.001, 0.004]	small
weather	E(Y X)	coverage NCV + CL2, NCV-REML widening	0.012 [0.006, 0.017]	0.001 [-0.002, 0.005]	0.002 [-0.003, 0.006]	small
electricity	E(Y X)	coverage NCV + CL2, NCV-REML widening	0.054 [0.046, 0.062]	0.000 [-0.004, 0.004]	0.003 [-0.003, 0.009]	small
ECG strain	$\beta(s,t)$	coverage REML + CL2	0.000 [-0.000, 0.001]	-0.000 [-0.002, 0.002]	-0.000 [-0.001, 0.001]	small
AF trial	$\beta(s,t)$	coverage REML + CL2	0.008 [0.005, 0.012]	-0.001 [-0.004, 0.001]	0.000 [-0.002, 0.002]	small
running	$\beta(s,t)$	coverage REML + CL2	0.006 [-0.002, 0.014]	-0.000 [-0.007, 0.006]	0.001 [-0.007, 0.008]	small
DTI	$\beta(s,t)$	coverage REML + CL2	0.010 [0.005, 0.015]	0.002 [-0.001, 0.005]	-0.000 [-0.003, 0.003]	small
gait	$\beta(s,t)$	coverage REML + CL2	0.001 [-0.000, 0.002]	0.000 [-0.001, 0.001]	0.000 [-0.001, 0.001]	small
ECG 8-lead	$\beta(s,t)$	coverage REML + CL2	0.006 [0.003, 0.010]	-0.001 [-0.004, 0.002]	-0.001 [-0.004, 0.003]	small
ocean	$\beta(s,t)$	coverage REML + CL2	-0.003 [-0.006, -0.000]	-0.006 [-0.011, -0.002]	-0.000 [-0.003, 0.003]	small
weather	$\beta(s,t)$	coverage REML + CL2	-0.043 [-0.071, -0.017]	-0.004 [-0.016, 0.007]	-0.022 [-0.043, -0.003]	small
electricity	$\beta(s,t)$	coverage REML + CL2	0.028 [0.013, 0.044]	-0.013 [-0.023, -0.003]	0.022 [0.005, 0.038]	undetermined
ECG strain	E(Y X)	coverage REML + CL2	0.000 [-0.000, 0.001]	0.002 [-0.000, 0.003]	0.000 [-0.001, 0.001]	small
AF trial	E(Y X)	coverage REML + CL2	0.001 [-0.000, 0.002]	-0.002 [-0.003, -0.001]	-0.000 [-0.001, 0.001]	small
running	E(Y X)	coverage REML + CL2	-0.005 [-0.006, -0.003]	-0.000 [-0.003, 0.002]	0.000 [-0.001, 0.001]	small
DTI	E(Y X)	coverage REML + CL2	-0.001 [-0.002, 0.000]	0.002 [0.000, 0.003]	-0.000 [-0.001, 0.001]	small
gait	E(Y X)	coverage REML + CL2	-0.000 [-0.001, 0.000]	0.000 [-0.000, 0.001]	-0.000 [-0.000, 0.000]	small
ECG 8-lead	E(Y X)	coverage REML + CL2	0.001 [0.000, 0.002]	0.000 [-0.001, 0.002]	0.000 [-0.001, 0.001]	small
ocean	E(Y X)	coverage REML + CL2	-0.003 [-0.004, -0.003]	-0.001 [-0.002, 0.001]	0.000 [-0.001, 0.001]	small

Table 34: Truth-source and residual-source main effects and their interaction (paired, 95% bootstrap intervals) on the log NCV/REML MSE ratio and on coverage. Residual verdict: small = interval inside ± 2 points or $\log[0.95, 1.05]$. (*continued*)

dataset	estimand	quantity	truth	residual	interaction	residual verdict
weather	E(Y X)	coverage REML + CL2	-0.021 [-0.027, -0.015]	0.005 [0.001, 0.009]	-0.006 [-0.011, -0.002]	small
electricity	E(Y X)	coverage REML + CL2	-0.007 [-0.011, -0.002]	-0.002 [-0.006, 0.003]	0.006 [0.001, 0.010]	small
ECG strain	$\beta(s, t)$	coverage REML, model-based	-0.001 [-0.002, 0.000]	-0.003 [-0.007, 0.000]	0.001 [-0.001, 0.002]	small
AF trial	$\beta(s, t)$	coverage REML, model-based	0.018 [0.010, 0.026]	-0.007 [-0.012, -0.002]	0.001 [-0.007, 0.008]	small
running	$\beta(s, t)$	coverage REML, model-based	0.044 [0.036, 0.053]	0.002 [-0.007, 0.013]	0.008 [-0.001, 0.017]	small
DTI	$\beta(s, t)$	coverage REML, model-based	0.028 [0.022, 0.034]	-0.004 [-0.009, 0.001]	-0.004 [-0.010, 0.001]	small
gait	$\beta(s, t)$	coverage REML, model-based	0.007 [0.005, 0.010]	-0.002 [-0.003, 0.000]	0.000 [-0.001, 0.001]	small
ECG 8-lead	$\beta(s, t)$	coverage REML, model-based	0.017 [0.012, 0.023]	-0.011 [-0.015, -0.006]	-0.003 [-0.010, 0.003]	small
ocean	$\beta(s, t)$	coverage REML, model-based	0.004 [-0.002, 0.009]	-0.014 [-0.021, -0.007]	-0.003 [-0.010, 0.003]	undetermined
weather	$\beta(s, t)$	coverage REML, model-based	0.032 [0.019, 0.047]	-0.001 [-0.013, 0.014]	-0.009 [-0.024, 0.006]	small
electricity	$\beta(s, t)$	coverage REML, model-based	0.078 [0.060, 0.099]	-0.008 [-0.022, 0.006]	0.020 [-0.001, 0.041]	undetermined
ECG strain	E(Y X)	coverage REML, model-based	-0.002 [-0.003, -0.001]	-0.002 [-0.004, 0.000]	0.002 [0.000, 0.003]	small
AF trial	E(Y X)	coverage REML, model-based	-0.001 [-0.002, 0.001]	-0.002 [-0.004, -0.000]	0.001 [-0.001, 0.002]	small
running	E(Y X)	coverage REML, model-based	0.001 [-0.001, 0.003]	-0.006 [-0.010, -0.002]	-0.000 [-0.002, 0.001]	small
DTI	E(Y X)	coverage REML, model-based	0.006 [0.003, 0.008]	-0.006 [-0.009, -0.002]	-0.003 [-0.006, -0.001]	small
gait	E(Y X)	coverage REML, model-based	0.000 [-0.000, 0.001]	-0.001 [-0.002, -0.000]	0.000 [-0.000, 0.001]	small
ECG 8-lead	E(Y X)	coverage REML, model-based	0.008 [0.006, 0.009]	-0.007 [-0.009, -0.005]	-0.001 [-0.002, -0.000]	small
ocean	E(Y X)	coverage REML, model-based	-0.006 [-0.008, -0.005]	-0.003 [-0.005, -0.000]	0.001 [-0.001, 0.002]	small
weather	E(Y X)	coverage REML, model-based	0.007 [0.003, 0.012]	0.000 [-0.005, 0.006]	-0.000 [-0.004, 0.004]	small
electricity	E(Y X)	coverage REML, model-based	0.010 [0.005, 0.015]	-0.001 [-0.008, 0.007]	0.006 [-0.000, 0.013]	small
ECG strain	$\beta(s, t)$	log MSE NCV/REML	-0.245 [-0.267, -0.225]	-0.049 [-0.059, -0.038]	-0.032 [-0.043, -0.022]	undetermined
AF trial	$\beta(s, t)$	log MSE NCV/REML	-0.323 [-0.566, -0.089]	-0.125 [-0.208, -0.046]	-0.027 [-0.153, 0.105]	undetermined
running	$\beta(s, t)$	log MSE NCV/REML	-1.844 [-2.042, -1.623]	0.017 [-0.060, 0.101]	-0.024 [-0.159, 0.113]	undetermined
DTI	$\beta(s, t)$	log MSE NCV/REML	-1.082 [-1.142, -1.017]	-0.086 [-0.118, -0.059]	-0.048 [-0.092, -0.009]	not small
gait	$\beta(s, t)$	log MSE NCV/REML	-0.381 [-0.503, -0.287]	-0.001 [-0.021, 0.018]	-0.004 [-0.041, 0.033]	small
ECG 8-lead	$\beta(s, t)$	log MSE NCV/REML	-1.029 [-1.469, -0.781]	0.078 [-0.113, 0.259]	0.186 [-0.173, 0.559]	undetermined
ocean	$\beta(s, t)$	log MSE NCV/REML	-3.273 [-3.480, -3.060]	0.038 [-0.135, 0.240]	0.287 [-0.013, 0.678]	undetermined
weather	$\beta(s, t)$	log MSE NCV/REML	-0.703 [-0.786, -0.625]	0.011 [-0.050, 0.072]	-0.027 [-0.104, 0.048]	undetermined
electricity	$\beta(s, t)$	log MSE NCV/REML	-4.051 [-4.404, -3.753]	-0.267 [-0.403, -0.125]	-0.151 [-0.386, 0.066]	not small
ECG strain	E(Y X)	log MSE NCV/REML	-0.155 [-0.168, -0.143]	-0.020 [-0.029, -0.012]	-0.027 [-0.038, -0.017]	small
AF trial	E(Y X)	log MSE NCV/REML	-0.057 [-0.077, -0.035]	-0.005 [-0.013, 0.004]	-0.006 [-0.019, 0.007]	small
running	E(Y X)	log MSE NCV/REML	-0.243 [-0.263, -0.223]	0.002 [-0.008, 0.013]	0.012 [-0.003, 0.025]	small
DTI	E(Y X)	log MSE NCV/REML	-0.293 [-0.312, -0.273]	-0.016 [-0.025, -0.007]	-0.008 [-0.020, 0.003]	small
gait	E(Y X)	log MSE NCV/REML	-0.066 [-0.075, -0.056]	0.000 [-0.002, 0.003]	-0.000 [-0.005, 0.004]	small
ECG 8-lead	E(Y X)	log MSE NCV/REML	-0.240 [-0.267, -0.212]	-0.029 [-0.044, -0.015]	-0.001 [-0.021, 0.020]	small
ocean	E(Y X)	log MSE NCV/REML	-0.355 [-0.375, -0.333]	-0.018 [-0.029, -0.008]	0.009 [-0.003, 0.022]	small
weather	E(Y X)	log MSE NCV/REML	-0.264 [-0.301, -0.229]	0.013 [-0.007, 0.033]	-0.043 [-0.073, -0.013]	small
electricity	E(Y X)	log MSE NCV/REML	-0.832 [-0.893, -0.767]	-0.051 [-0.078, -0.024]	0.030 [-0.015, 0.077]	undetermined

Table 35: ECG strain: coverage and NCV/REML MSE ratio with fitted residuals and with heartbeat differences (variance-matched and own scale).

truth	residual	estimand	NCV + CL2	NCV + CL2, NCV-REML widening	REML + CL2	REML, model-based	AR(1)	MSE NCV/REML
REML	NCV residuals	$\beta(s, t)$	0.899		0.929	0.943		0.951
REML	REML residuals	$\beta(s, t)$	0.904		0.931	0.943		0.982
REML	beat differences (own scale)	$\beta(s, t)$	0.932		0.944	0.945		0.970
REML	beat differences (variance-matched)	$\beta(s, t)$	0.910		0.939	0.944		0.907
midpoint	NCV residuals	$\beta(s, t)$	0.900		0.931	0.943		0.890
midpoint	REML residuals	$\beta(s, t)$	0.904		0.932	0.944	0.874	0.927
midpoint	beat differences (own scale)	$\beta(s, t)$	0.933		0.946	0.945		0.943
midpoint	beat differences (variance-matched)	$\beta(s, t)$	0.910		0.940	0.945	0.837	0.857
NCV	NCV residuals	$\beta(s, t)$	0.909		0.941	0.944		0.732
NCV	REML residuals	$\beta(s, t)$	0.909		0.939	0.944		0.782
NCV	beat differences (own scale)	$\beta(s, t)$	0.928		0.946	0.945		0.887
NCV	beat differences (variance-matched)	$\beta(s, t)$	0.906		0.942	0.945		0.728
REML	NCV residuals	E(Y X)	0.904		0.933	0.939		1.036
REML	REML residuals	E(Y X)	0.906		0.933	0.938		1.043
REML	beat differences (own scale)	E(Y X)	0.940		0.948	0.950		0.995
REML	beat differences (variance-matched)	E(Y X)	0.921		0.945	0.949		0.992
midpoint	NCV residuals	E(Y X)	0.916		0.938	0.940		0.959
midpoint	REML residuals	E(Y X)	0.916		0.937	0.938	0.869	0.977
midpoint	beat differences (own scale)	E(Y X)	0.940		0.951	0.950		0.974
midpoint	beat differences (variance-matched)	E(Y X)	0.923		0.947	0.950	0.847	0.953
NCV	NCV residuals	E(Y X)	0.918		0.942	0.940		0.875
NCV	REML residuals	E(Y X)	0.916		0.939	0.938		0.905
NCV	beat differences (own scale)	E(Y X)	0.935		0.950	0.950		0.962
NCV	beat differences (variance-matched)	E(Y X)	0.921		0.948	0.950		0.888

Table 36: Datasets with dependent source curves: block-NCV/NCV MSE ratio, mismatched-partition ratio and coverage changes (95% bootstrap intervals), under block flips and curve flips.

dataset	estimand	verdict	flips	MSE block-NCV/NCV	coverage loss	coverage loss (method)	MSE mismatched blocks/NCV
ocean	$\beta(s,t)$	inconclusive	block flips	1.105 [0.979, 1.278]	0.006 [0.005, 0.007]	0.006 [0.003, 0.009]	
ocean	$\beta(s,t)$	inconclusive	curve flips	1.168 [1.029, 1.290]	0.001 [-0.000, 0.002]	0.004 [0.001, 0.007]	
ocean	$E(Y X)$	cost only	block flips	0.995 [0.989, 1.001]	-0.006 [-0.007, -0.005]	-0.005 [-0.007, -0.004]	
ocean	$E(Y X)$	cost only	curve flips	1.003 [0.997, 1.010]	0.003 [0.001, 0.004]	0.003 [0.002, 0.005]	
weather	$\beta(s,t)$	inconclusive	block flips	1.017 [0.955, 1.090]	0.011 [0.009, 0.014]	0.014 [0.008, 0.019]	1.039 [0.957, 1.128]
weather	$\beta(s,t)$	inconclusive	curve flips	1.061 [1.002, 1.132]	0.010 [0.007, 0.014]	-0.007 [-0.014, -0.001]	
weather	$E(Y X)$	hurts	block flips	1.014 [0.995, 1.032]	-0.006 [-0.010, -0.002]	-0.003 [-0.007, 0.002]	0.999 [0.982, 1.015]
weather	$E(Y X)$	hurts	curve flips	1.009 [0.986, 1.030]	0.029 [0.023, 0.033]	0.026 [0.021, 0.031]	
electricity	$\beta(s,t)$	inconclusive	block flips	1.021 [1.003, 1.045]	0.000 [-0.000, 0.001]	-0.002 [-0.003, -0.000]	
electricity	$\beta(s,t)$	inconclusive	curve flips	1.055 [0.977, 1.161]	0.001 [-0.000, 0.001]	0.000 [-0.001, 0.002]	
electricity	$E(Y X)$	cost only	block flips	1.001 [0.995, 1.008]	-0.008 [-0.009, -0.006]	-0.008 [-0.009, -0.007]	
electricity	$E(Y X)$	cost only	curve flips	1.004 [0.995, 1.013]	0.001 [-0.000, 0.003]	0.001 [-0.001, 0.002]	

Table 37: Coverage of curve- and block-clustered intervals under curve flips and block flips. Widened: NCV + CL2 widened by the NCV-REML difference; alt. blocks: the alternative partition (weather).

dataset	estimand	flips	NCV + curve CL2, widened	block NCV + block CL2, widened	NCV + block CL2, widened	block NCV + curve CL2, widened	REML + curve CL2	REML + block CL2	mismatched blocks, widened	REML + alt. block CL2
ocean	$\beta(s,t)$	block flips	0.925	0.919	0.920	0.925	0.934	0.929		
ocean	$\beta(s,t)$	curve flips	0.916	0.912	0.916	0.913	0.925	0.922		
ocean	$E(Y X)$	block flips	0.933	0.939	0.940	0.933	0.932	0.936		
ocean	$E(Y X)$	curve flips	0.939	0.936	0.937	0.938	0.935	0.930		
weather	$\beta(s,t)$	block flips	0.823	0.809	0.813	0.821	0.903	0.889	0.806	0.883
weather	$\beta(s,t)$	curve flips	0.828	0.836	0.824	0.846	0.882	0.867		
weather	$E(Y X)$	block flips	0.898	0.901	0.908	0.895	0.906	0.902	0.900	0.904
weather	$E(Y X)$	curve flips	0.920	0.894	0.898	0.922	0.924	0.892		
electricity	$\beta(s,t)$	block flips	0.895	0.896	0.894	0.897	0.920	0.921		
electricity	$\beta(s,t)$	curve flips	0.911	0.911	0.911	0.912	0.927	0.925		
electricity	$E(Y X)$	block flips	0.938	0.946	0.946	0.939	0.931	0.938		
electricity	$E(Y X)$	curve flips	0.947	0.946	0.946	0.947	0.927	0.926		

S2.4 Studentized errors

On the datasets with independent source curves the pooled SD of the studentized REML error for β with the CL2 standard error is 1.01–1.14 (mean -0.00 to 0.03), on the datasets with dependent source curves 1.07–2.86; with the model-based standard error 2.20–3.95.

Table 38: Pooled studentized errors of the REML fit (all subjects, fitted residuals; means over residual sources).

dataset	role	truth	estimand	mean z (CL2)	SD z (CL2)	share $ z > 1.96$ (CL2)	mean z (model)	SD z (model)	share $ z > 1.96$ (model)
ECG strain	indep. curves	REML	$\beta(s,t)$	0.000	1.024	0.057	-0.001	3.24	0.439
ECG strain	indep. curves	midpoint	$\beta(s,t)$	-0.003	1.025	0.057	0.000	3.25	0.440
ECG strain	indep. curves	NCV	$\beta(s,t)$	-0.003	1.022	0.056	0.001	3.25	0.440
AF trial	indep. curves	REML	$\beta(s,t)$	0.032	1.067	0.067	0.077	2.46	0.343
AF trial	indep. curves	midpoint	$\beta(s,t)$	0.022	1.044	0.062	0.060	2.42	0.333
AF trial	indep. curves	NCV	$\beta(s,t)$	0.021	1.028	0.058	0.063	2.37	0.326
running	indep. curves	REML	$\beta(s,t)$	0.003	1.143	0.085	-0.003	3.65	0.520
running	indep. curves	midpoint	$\beta(s,t)$	0.002	1.115	0.079	-0.004	3.40	0.484
running	indep. curves	NCV	$\beta(s,t)$	0.001	1.115	0.078	-0.001	3.38	0.475
DTI	indep. curves	REML	$\beta(s,t)$	0.006	1.118	0.080	0.014	2.36	0.386
DTI	indep. curves	midpoint	$\beta(s,t)$	0.009	1.079	0.070	0.018	2.22	0.361
DTI	indep. curves	NCV	$\beta(s,t)$	0.012	1.078	0.069	0.024	2.20	0.358
gait	indep. curves	REML	$\beta(s,t)$	0.002	1.048	0.061	-0.006	3.95	0.487
gait	indep. curves	midpoint	$\beta(s,t)$	0.001	1.044	0.060	-0.006	3.94	0.484
gait	indep. curves	NCV	$\beta(s,t)$	0.000	1.045	0.060	-0.007	3.93	0.480
ECG 8-lead	indep. curves	REML	$\beta(s,t)$	0.011	1.046	0.062	0.025	2.87	0.371
ECG 8-lead	indep. curves	midpoint	$\beta(s,t)$	0.008	1.015	0.056	0.017	2.79	0.360
ECG 8-lead	indep. curves	NCV	$\beta(s,t)$	0.009	1.011	0.056	0.018	2.75	0.354
ocean	dep. source	REML	$\beta(s,t)$	0.001	1.095	0.073	0.004	4.46	0.589
ocean	dep. source	midpoint	$\beta(s,t)$	0.002	1.108	0.076	0.006	4.35	0.588
ocean	dep. source	NCV	$\beta(s,t)$	0.001	1.111	0.076	0.005	4.37	0.586
weather	dep. source	REML	$\beta(s,t)$	-0.040	2.809	0.110	-0.047	5.83	0.683
weather	dep. source	midpoint	$\beta(s,t)$	-0.018	1.866	0.091	-0.016	4.60	0.642
weather	dep. source	NCV	$\beta(s,t)$	-0.087	2.861	0.153	-0.083	5.47	0.651
electricity	dep. source	REML	$\beta(s,t)$	0.001	1.392	0.091	-0.016	4.70	0.621
electricity	dep. source	midpoint	$\beta(s,t)$	-0.005	1.077	0.063	-0.023	3.74	0.541
electricity	dep. source	NCV	$\beta(s,t)$	-0.002	1.068	0.064	-0.002	3.77	0.543
ECG strain	indep. curves	REML	$E(Y X)$	-0.006	1.040	0.061	-0.016	3.07	0.415
ECG strain	indep. curves	midpoint	$E(Y X)$	-0.006	1.040	0.061	-0.018	3.09	0.418
ECG strain	indep. curves	NCV	$E(Y X)$	-0.006	1.039	0.061	-0.017	3.08	0.417
AF trial	indep. curves	REML	$E(Y X)$	-0.027	1.015	0.055	-0.097	2.34	0.320
AF trial	indep. curves	midpoint	$E(Y X)$	-0.028	1.012	0.054	-0.098	2.34	0.320
AF trial	indep. curves	NCV	$E(Y X)$	-0.028	1.011	0.054	-0.094	2.34	0.320

Table 38: Pooled studentized errors of the REML fit (all subjects, fitted residuals; means over residual sources).
(continued)

dataset	role	truth	estimand	mean z (CL2)	SD z (CL2)	share $ z > 1.96$ (CL2)	mean z (model)	SD z (model)	share $ z > 1.96$ (model)
running	indep. curves	REML	$E(Y X)$	-0.011	1.062	0.065	-0.028	3.30	0.433
running	indep. curves	midpoint	$E(Y X)$	-0.010	1.068	0.067	-0.026	3.27	0.428
running	indep. curves	NCV	$E(Y X)$	-0.011	1.079	0.070	-0.028	3.29	0.432
DTI	indep. curves	REML	$E(Y X)$	0.010	1.067	0.067	0.013	2.20	0.358
DTI	indep. curves	midpoint	$E(Y X)$	0.012	1.069	0.067	0.019	2.17	0.353
DTI	indep. curves	NCV	$E(Y X)$	0.012	1.074	0.068	0.019	2.17	0.352
gait	indep. curves	REML	$E(Y X)$	-0.015	1.028	0.057	-0.069	3.62	0.462
gait	indep. curves	midpoint	$E(Y X)$	-0.014	1.027	0.057	-0.068	3.62	0.462
gait	indep. curves	NCV	$E(Y X)$	-0.014	1.029	0.057	-0.066	3.62	0.462
ECG 8-lead	indep. curves	REML	$E(Y X)$	-0.007	0.986	0.049	-0.006	2.21	0.251
ECG 8-lead	indep. curves	midpoint	$E(Y X)$	-0.009	0.979	0.047	-0.006	2.19	0.246
ECG 8-lead	indep. curves	NCV	$E(Y X)$	-0.011	0.980	0.047	-0.007	2.17	0.244
ocean	dep. source	REML	$E(Y X)$	0.002	1.059	0.063	0.003	3.86	0.541
ocean	dep. source	midpoint	$E(Y X)$	0.003	1.066	0.065	0.004	3.86	0.543
ocean	dep. source	NCV	$E(Y X)$	0.003	1.071	0.067	0.002	3.88	0.547
weather	dep. source	REML	$E(Y X)$	0.021	1.082	0.067	0.061	2.86	0.464
weather	dep. source	midpoint	$E(Y X)$	0.022	1.102	0.075	0.056	2.81	0.458
weather	dep. source	NCV	$E(Y X)$	0.006	1.148	0.088	0.050	2.80	0.457
electricity	dep. source	REML	$E(Y X)$	0.016	1.073	0.069	0.032	3.43	0.511
electricity	dep. source	midpoint	$E(Y X)$	0.004	1.089	0.073	-0.004	3.32	0.495
electricity	dep. source	NCV	$E(Y X)$	0.003	1.101	0.076	-0.001	3.34	0.501

S3 Comparators

S3.1 AR(1) working model

The AR(1) working model fits the REML model with an AR(1) residual correlation along t , ρ profiled (mgcv bam), and uses its model-based intervals. On stationary homoskedastic synthetic errors it covers 0.93–0.99 (Gaussian, $G = 100$; next table). It fails under misregistration (mean 0.84, β 0.86–0.88), for $\gamma(t)$ under covariate-dependent variance, and in the plasmode study: with the midpoint truth and REML residuals it covers the mean at 0.67–0.93 and β at 0.44–0.91 across the nine datasets. Median fit time: 39 s against 2.0 s for REML (Appendix C of the paper).

Table 39: AR(1) working model (Gaussian, $G = 100$): coverage next to REML intervals, interval score relative to REML + CL2, MSE relative to NCV (95% bootstrap interval).

error	G	truth	estimand	AR(1) working model	NCV + CL2, widened by NCV–REML difference	REML + CL2	REML, model-based	MSE AR(1) / NCV	IS AR(1) / REML+CL2
smooth	100	smooth	$\beta(s, t)$	0.993	0.983	0.944	0.657	0.98 [0.82, 1.18]	0.532
smooth	100	smooth	$E(Y X)$	0.972	0.959	0.944	0.634	1.04 [1.02, 1.05]	0.846
var(z)	100	smooth	$\beta(s, t)$	0.992	0.984	0.949	0.660	0.92 [0.76, 1.09]	0.550
var(z)	100	smooth	$E(Y X)$	0.949	0.955	0.944	0.606	1.05 [1.03, 1.07]	0.925
var(t)	100	smooth	$\beta(s, t)$	0.991	0.983	0.944	0.662	1.00 [0.80, 1.21]	0.558
var(t)	100	smooth	$E(Y X)$	0.972	0.959	0.944	0.642	1.10 [1.08, 1.12]	0.902
var(t, z)	100	smooth	$\beta(s, t)$	0.990	0.984	0.950	0.666	1.06 [0.88, 1.25]	0.580
var(t, z)	100	smooth	$E(Y X)$	0.951	0.955	0.944	0.615	1.12 [1.10, 1.14]	0.980
sign-chg.	100	smooth	$\beta(s, t)$	0.992	0.985	0.942	0.700	1.13 [1.01, 1.27]	0.551
sign-chg.	100	smooth	$E(Y X)$	0.963	0.957	0.942	0.646	1.03 [1.01, 1.06]	0.778
rough	100	smooth	$\beta(s, t)$	0.972	0.983	0.940	0.672	1.30 [1.11, 1.49]	0.521
rough	100	smooth	$E(Y X)$	0.934	0.958	0.942	0.646	1.04 [1.03, 1.06]	0.834
AR(1)	100	smooth	$\beta(s, t)$	0.989	0.984	0.940	0.651	1.12 [0.96, 1.27]	0.493
AR(1)	100	smooth	$E(Y X)$	0.961	0.959	0.942	0.625	1.02 [1.01, 1.04]	0.805
misreg.	100	smooth	$\beta(s, t)$	0.875	0.976	0.951	0.753	1.29 [1.00, 1.62]	1.059
misreg.	100	smooth	$E(Y X)$	0.844	0.952	0.942	0.730	1.09 [1.07, 1.11]	1.347
misreg.	100	rough	$\beta(s, t)$	0.860	0.934	0.944	0.729	0.84 [0.70, 0.98]	1.102
misreg.	100	rough	$E(Y X)$	0.838	0.938	0.937	0.706	1.00 [0.99, 1.01]	1.391

Table 40: AR(1) working model in the plasmode study: coverage, 5% quantile of pointwise coverage, REML + CL2 coverage on the same data, profiled ρ and fit time.

dataset	role	truth	residual	estimand	AR(1)	AR(1) q05	REML + CL2	rho	fit time s
ECG strain	indep. curves	midpoint	REML	$\beta(s, t)$	0.874	0.650	0.944	0.976	44.5
ECG strain	indep. curves	midpoint	REML	$\gamma(t)$	0.926	0.725	0.948	0.976	44.5
ECG strain	indep. curves	midpoint	REML	$E(Y X)$	0.869	0.620	0.938	0.976	44.5

Table 40: AR(1) working model in the plasmode study: coverage, 5% quantile of pointwise coverage, REML + CL2 coverage on the same data, profiled ρ and fit time. (*continued*)

dataset	role	truth	residual	estimand	AR(1)	AR(1) q05	REML + CL2	rho	fit time s
ECG strain	indep. curves	midpoint	beat differences (variance-matched)	$\beta(s,t)$	0.837	0.475	0.945	0.979	39.5
ECG strain	indep. curves	midpoint	beat differences (variance-matched)	$\gamma(t)$	0.943	0.782	0.957	0.979	39.5
ECG strain	indep. curves	midpoint	beat differences (variance-matched)	$E(Y X)$	0.847	0.497	0.950	0.979	39.5
AF trial	indep. curves	midpoint	REML	$\beta(s,t)$	0.838	0.520	0.938	0.940	27.6
AF trial	indep. curves	midpoint	REML	$\gamma(t)$	0.877	0.685	0.946	0.940	27.6
AF trial	indep. curves	midpoint	REML	$E(Y X)$	0.870	0.630	0.947	0.940	27.6
running	indep. curves	midpoint	REML	$\beta(s,t)$	0.728	0.445	0.922	0.990	52.8
running	indep. curves	midpoint	REML	$\gamma(t)$	0.806	0.540	0.934	0.990	52.8
running	indep. curves	midpoint	REML	$E(Y X)$	0.787	0.500	0.934	0.990	52.8
running	indep. curves	REML-fitted	REML	$\beta(s,t)$	0.548	0.165	0.916	0.990	52.5
running	indep. curves	NCV-fitted	NCV	$\beta(s,t)$	0.811	0.545	0.922	0.990	52.7
running	indep. curves	REML-fitted	REML	$\gamma(t)$	0.805	0.538	0.937	0.990	52.5
running	indep. curves	NCV-fitted	NCV	$\gamma(t)$	0.798	0.545	0.929	0.990	52.7
running	indep. curves	REML-fitted	REML	$E(Y X)$	0.754	0.470	0.935	0.990	52.5
running	indep. curves	NCV-fitted	NCV	$E(Y X)$	0.793	0.495	0.930	0.990	52.7
DTI	indep. curves	midpoint	REML	$\beta(s,t)$	0.858	0.470	0.929	0.817	42.7
DTI	indep. curves	midpoint	REML	$\gamma(t)$	0.900	0.855	0.934	0.817	42.7
DTI	indep. curves	midpoint	REML	$E(Y X)$	0.902	0.667	0.932	0.817	42.7
gait	indep. curves	midpoint	REML	$\beta(s,t)$	0.914	0.730	0.940	0.987	161.6
gait	indep. curves	midpoint	REML	$\gamma(t)$	0.919	0.770	0.933	0.987	161.6
gait	indep. curves	midpoint	REML	$E(Y X)$	0.927	0.760	0.943	0.987	161.6
ECG 8-lead	indep. curves	midpoint	REML	$\beta(s,t)$	0.720	0.170	0.944	0.910	120.4
ECG 8-lead	indep. curves	midpoint	REML	$\gamma(t)$	0.940	0.620	0.961	0.910	120.4
ECG 8-lead	indep. curves	midpoint	REML	$E(Y X)$	0.924	0.555	0.953	0.910	120.4
ocean	dep. source	midpoint	REML	$\beta(s,t)$	0.854	0.299	0.927	0.990	157.0
ocean	dep. source	midpoint	REML	$\gamma(t)$	0.934	0.815	0.945	0.990	157.0
ocean	dep. source	midpoint	REML	$E(Y X)$	0.932	0.775	0.935	0.990	157.0
weather	dep. source	midpoint	REML	$\beta(s,t)$	0.438	0.320	0.905	0.445	44.3
weather	dep. source	midpoint	REML	$\gamma(t)$	0.567	0.500	0.909	0.445	44.3
weather	dep. source	midpoint	REML	$E(Y X)$	0.675	0.440	0.922	0.445	44.3
electricity	dep. source	midpoint	REML	$\beta(s,t)$	0.677	0.220	0.939	0.979	40.7
electricity	dep. source	midpoint	REML	$\gamma(t)$	0.908	0.823	0.956	0.979	40.7
electricity	dep. source	midpoint	REML	$E(Y X)$	0.844	0.640	0.927	0.979	40.7

S3.2 Random curve effect, GLS and curve bootstrap

Three further methods address dependent residual curves. All are fitted to the same data as the study’s REML fit, with common random numbers, in the 4 Gaussian main-factorial settings with dependent errors at the default signal (rough stationary and smooth errors, $G = 40$ and 100 ; below: the *four synthetic dependent Gaussian settings*), the corresponding settings with independent errors, and the six plasmode datasets with independent source curves (midpoint truth, REML residuals, all subjects and 40: 12 plasmode combinations). The evidence on these methods is limited to these four Gaussian settings with dependent errors and these plasmode combinations.

Functional random effect per curve (`pcre()` in `refund`): the REML model plus a curve-level random function expanded in the leading eigenfunctions of the REML residual curves (95% of their variance), with model-based intervals.

Generalized least squares (GLS). Estimate the within-curve residual covariance from the REML residual curves, either as their raw covariance or as an FPCA estimate plus white noise; prewhiten every curve with the inverse square root of that covariance; refit by REML; use model-based intervals. This is the algorithm of `refund`’s former `pffr_gls()`, rebuilt from `refund`’s internal functions; a singular estimated covariance (fewer curves than grid points) is repaired by the nearest positive-definite matrix.

Curve bootstrap: resample whole curves, refit by REML (199 resamples), and form percentile intervals or Wald intervals with the bootstrap standard error around the REML estimate.

Results. In the four synthetic dependent Gaussian settings the random curve effect and GLS with the FPCA covariance cover β at 0.93–0.97 at 0.51–0.66 times REML + CL2’s width and 0.52–0.57 times its interval score. The random curve effect pays for this with the other estimands there: it covers the mean at 0.88–0.91 and $\alpha(t)$ at 0.51–0.59. On the plasmode data both fall to 0.59–0.86 for β , while REML + CL2 covers 0.91–0.94. A common working covariance that fits the synthetic error processes is only an approximation to the subject-specific two-point error law of the plasmode generator (S2), and the model-based intervals of these fits do not hold their level

there; this is a finding about that generator, not a general statement about real data. GLS with the raw covariance undercovers everywhere (0.59–0.89 even under independent errors), because the raw covariance of fewer curves than grid points is a poor, singular estimate. The percentile bootstrap covers β at 0.93–0.97 on the plasmode data with intervals 1.42 times as wide as REML + CL2's, at the cost of 199 refits (Appendix C of the paper).

As *estimators* the random curve effect and GLS (FPCA) remove most of REML's excess error. Their MSE for β relative to the NCV fit's, geometric mean over settings, is 1.75 and 1.79 in the four synthetic dependent Gaussian settings (REML 7.4), and 0.72 and 0.68 on the plasmode datasets (REML 2.73; per setting 0.23–1.25 and 0.21–1.20). Both thus improve much on REML, but in the synthetic settings they are worse than NCV. Their plasmode advantage over NCV rests on the midpoint truth with REML residuals only, and the GLS covariance is estimated from the same REML residual curves that generate the plasmode errors. The random curve effect is expensive: median 18 s per fit, up to 1342 s, against 2.4 s for REML.

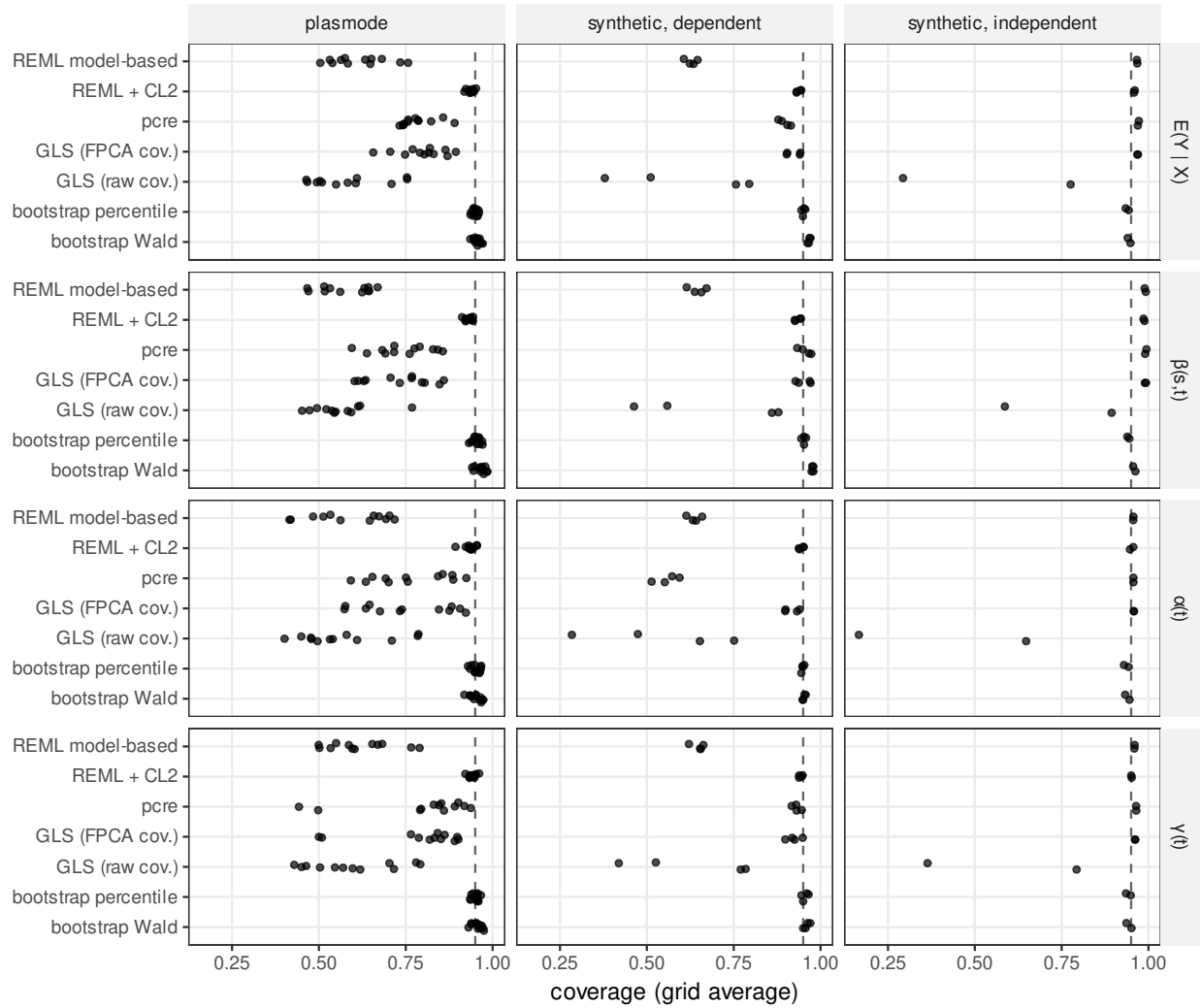


Figure 8: Random curve effect (pcre), GLS and curve bootstrap: grid-averaged coverage per setting or plasmode combination, by estimand and data source. Dashed: nominal 0.95.

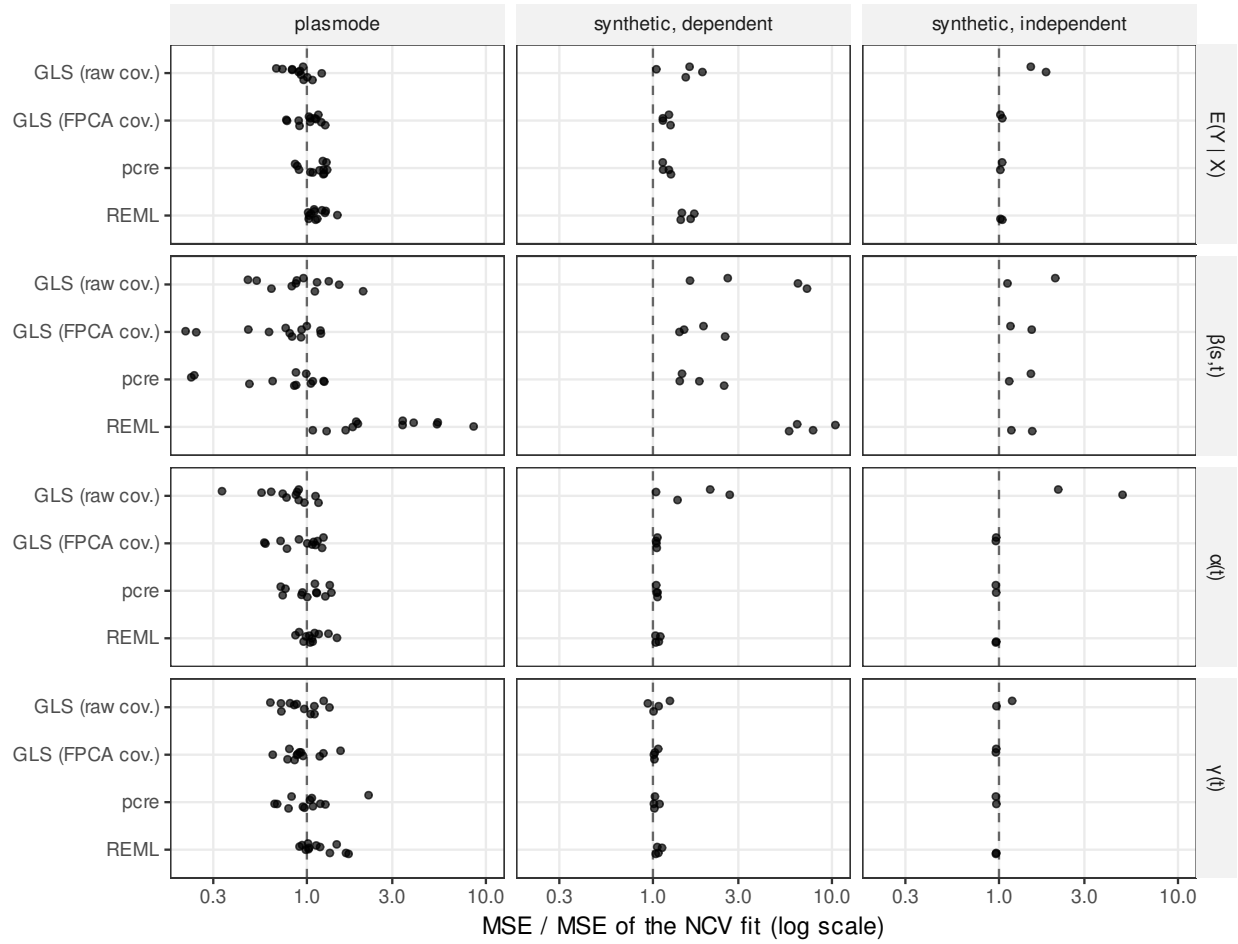


Figure 9: MSE of each estimator divided by the NCV fit's MSE (log scale; below 1 = better than NCV). The bootstrap intervals use the REML estimate.

Table 41: Comparators: mean and minimum coverage, mean 5% quantile of pointwise coverage, width and interval score relative to REML + CL2 (geometric means of per-setting ratios).

setting	estimand	interval	settings	coverage	min coverage	q05	width rel.	IS rel.
plasmode	$E(Y X)$	REML model-based	12	0.62	0.50	0.32	0.42	2.70
plasmode	$E(Y X)$	REML + CL2	12	0.94	0.92	0.88	1.00	1.00
plasmode	$E(Y X)$	pcre	12	0.78	0.73	0.54	0.63	1.46
plasmode	$E(Y X)$	GLS (FPCA cov.)	12	0.80	0.66	0.58	0.63	1.35
plasmode	$E(Y X)$	GLS (raw cov.)	12	0.58	0.46	0.43	0.38	1.99
plasmode	$E(Y X)$	bootstrap percentile	12	0.95	0.94	0.90	1.14	1.05
plasmode	$E(Y X)$	bootstrap Wald	12	0.96	0.94	0.91	1.12	1.00
plasmode	$\beta(s, t)$	REML model-based	12	0.58	0.47	0.32	0.39	2.90
plasmode	$\beta(s, t)$	REML + CL2	12	0.93	0.91	0.88	1.00	1.00
plasmode	$\beta(s, t)$	pcre	12	0.74	0.59	0.42	0.40	1.05
plasmode	$\beta(s, t)$	GLS (FPCA cov.)	12	0.73	0.60	0.45	0.38	1.03
plasmode	$\beta(s, t)$	GLS (raw cov.)	12	0.56	0.45	0.38	0.30	1.62
plasmode	$\beta(s, t)$	bootstrap percentile	12	0.95	0.93	0.91	1.42	1.22
plasmode	$\beta(s, t)$	bootstrap Wald	12	0.97	0.94	0.93	1.39	1.15
plasmode	$\alpha(t)$	REML model-based	12	0.58	0.42	0.33	0.39	2.83
plasmode	$\alpha(t)$	REML + CL2	12	0.93	0.89	0.89	1.00	1.00
plasmode	$\alpha(t)$	pcre	12	0.76	0.59	0.63	0.62	1.52
plasmode	$\alpha(t)$	GLS (FPCA cov.)	12	0.75	0.57	0.63	0.58	1.50
plasmode	$\alpha(t)$	GLS (raw cov.)	12	0.57	0.40	0.50	0.35	1.95
plasmode	$\alpha(t)$	bootstrap percentile	12	0.95	0.93	0.92	1.07	0.97
plasmode	$\alpha(t)$	bootstrap Wald	12	0.95	0.92	0.91	1.06	0.96
plasmode	$\gamma(t)$	REML model-based	12	0.62	0.50	0.37	0.43	2.67
plasmode	$\gamma(t)$	REML + CL2	12	0.94	0.92	0.90	1.00	1.00
plasmode	$\gamma(t)$	pcre	12	0.80	0.44	0.64	0.71	1.33
plasmode	$\gamma(t)$	GLS (FPCA cov.)	12	0.79	0.50	0.64	0.65	1.33
plasmode	$\gamma(t)$	GLS (raw cov.)	12	0.60	0.43	0.48	0.40	1.89

Table 41: Comparators: mean and minimum coverage, mean 5% quantile of pointwise coverage, width and interval score relative to REML + CL2 (geometric means of per-setting ratios). (*continued*)

setting	estimand	interval	settings	coverage	min coverage	q05	width rel.	IS rel.
plasmode	$\gamma(t)$	bootstrap percentile	12	0.95	0.94	0.92	1.13	1.05
plasmode	$\gamma(t)$	bootstrap Wald	12	0.96	0.93	0.91	1.11	1.02
syn. dep.	$E(Y X)$	REML model-based	4	0.63	0.61	0.54	0.47	2.05
syn. dep.	$E(Y X)$	REML + CL2	4	0.94	0.93	0.91	1.00	1.00
syn. dep.	$E(Y X)$	pcrc	4	0.90	0.88	0.84	0.76	0.91
syn. dep.	$E(Y X)$	GLS (FPCA cov.)	4	0.92	0.90	0.88	0.81	0.88
syn. dep.	$E(Y X)$	GLS (raw cov.)	4	0.61	0.38	0.55	0.43	2.03
syn. dep.	$E(Y X)$	bootstrap percentile	4	0.95	0.94	0.91	1.17	1.10
syn. dep.	$E(Y X)$	bootstrap Wald	4	0.97	0.96	0.93	1.16	1.02
syn. dep.	$\beta(s,t)$	REML model-based	4	0.65	0.61	0.56	0.49	2.06
syn. dep.	$\beta(s,t)$	REML + CL2	4	0.93	0.93	0.91	1.00	1.00
syn. dep.	$\beta(s,t)$	pcrc	4	0.96	0.93	0.92	0.59	0.55
syn. dep.	$\beta(s,t)$	GLS (FPCA cov.)	4	0.95	0.93	0.91	0.58	0.55
syn. dep.	$\beta(s,t)$	GLS (raw cov.)	4	0.69	0.46	0.63	0.40	1.33
syn. dep.	$\beta(s,t)$	bootstrap percentile	4	0.95	0.94	0.91	1.60	1.41
syn. dep.	$\beta(s,t)$	bootstrap Wald	4	0.98	0.97	0.94	1.57	1.28
syn. dep.	$\alpha(t)$	REML model-based	4	0.64	0.61	0.56	0.47	2.04
syn. dep.	$\alpha(t)$	REML + CL2	4	0.94	0.94	0.92	1.00	1.00
syn. dep.	$\alpha(t)$	pcrc	4	0.56	0.51	0.50	0.40	2.41
syn. dep.	$\alpha(t)$	GLS (FPCA cov.)	4	0.92	0.90	0.87	0.90	1.02
syn. dep.	$\alpha(t)$	GLS (raw cov.)	4	0.54	0.28	0.50	0.46	3.07
syn. dep.	$\alpha(t)$	bootstrap percentile	4	0.95	0.94	0.92	1.05	1.02
syn. dep.	$\alpha(t)$	bootstrap Wald	4	0.95	0.95	0.92	1.04	1.00
syn. dep.	$\gamma(t)$	REML model-based	4	0.65	0.62	0.58	0.47	1.96
syn. dep.	$\gamma(t)$	REML + CL2	4	0.94	0.94	0.92	1.00	1.00
syn. dep.	$\gamma(t)$	pcrc	4	0.93	0.92	0.90	0.92	0.97
syn. dep.	$\gamma(t)$	GLS (FPCA cov.)	4	0.92	0.90	0.89	0.90	0.98
syn. dep.	$\gamma(t)$	GLS (raw cov.)	4	0.62	0.42	0.58	0.45	2.00
syn. dep.	$\gamma(t)$	bootstrap percentile	4	0.96	0.94	0.92	1.07	1.01
syn. dep.	$\gamma(t)$	bootstrap Wald	4	0.96	0.95	0.93	1.05	0.97
syn. indep.	$E(Y X)$	REML model-based	2	0.97	0.97	0.94	1.01	0.98
syn. indep.	$E(Y X)$	REML + CL2	2	0.96	0.96	0.93	1.00	1.00
syn. indep.	$E(Y X)$	pcrc	2	0.97	0.97	0.94	1.03	0.99
syn. indep.	$E(Y X)$	GLS (FPCA cov.)	2	0.97	0.97	0.94	1.02	0.98
syn. indep.	$E(Y X)$	GLS (raw cov.)	2	0.53	0.29	0.47	0.42	3.05
syn. indep.	$E(Y X)$	bootstrap percentile	2	0.94	0.93	0.90	0.93	1.01
syn. indep.	$E(Y X)$	bootstrap Wald	2	0.94	0.94	0.90	0.92	0.98
syn. indep.	$\beta(s,t)$	REML model-based	2	0.99	0.99	0.97	1.02	1.01
syn. indep.	$\beta(s,t)$	REML + CL2	2	0.99	0.99	0.97	1.00	1.00
syn. indep.	$\beta(s,t)$	pcrc	2	0.99	0.99	0.98	1.04	1.02
syn. indep.	$\beta(s,t)$	GLS (FPCA cov.)	2	0.99	0.99	0.98	1.02	1.01
syn. indep.	$\beta(s,t)$	GLS (raw cov.)	2	0.74	0.59	0.65	0.49	1.50
syn. indep.	$\beta(s,t)$	bootstrap percentile	2	0.94	0.94	0.90	0.82	0.98
syn. indep.	$\beta(s,t)$	bootstrap Wald	2	0.96	0.96	0.92	0.81	0.90
syn. indep.	$\alpha(t)$	REML model-based	2	0.96	0.96	0.93	1.01	0.98
syn. indep.	$\alpha(t)$	REML + CL2	2	0.95	0.95	0.92	1.00	1.00
syn. indep.	$\alpha(t)$	pcrc	2	0.96	0.96	0.93	1.00	0.98
syn. indep.	$\alpha(t)$	GLS (FPCA cov.)	2	0.96	0.96	0.94	1.01	0.98
syn. indep.	$\alpha(t)$	GLS (raw cov.)	2	0.41	0.17	0.36	0.42	5.72
syn. indep.	$\alpha(t)$	bootstrap percentile	2	0.94	0.93	0.90	0.95	1.01
syn. indep.	$\alpha(t)$	bootstrap Wald	2	0.94	0.93	0.91	0.95	0.99
syn. indep.	$\gamma(t)$	REML model-based	2	0.96	0.96	0.93	1.01	0.97
syn. indep.	$\gamma(t)$	REML + CL2	2	0.95	0.95	0.92	1.00	1.00
syn. indep.	$\gamma(t)$	pcrc	2	0.96	0.96	0.94	1.04	0.98
syn. indep.	$\gamma(t)$	GLS (FPCA cov.)	2	0.96	0.96	0.94	1.02	0.97
syn. indep.	$\gamma(t)$	GLS (raw cov.)	2	0.58	0.36	0.51	0.41	2.40
syn. indep.	$\gamma(t)$	bootstrap percentile	2	0.94	0.94	0.89	0.95	1.00
syn. indep.	$\gamma(t)$	bootstrap Wald	2	0.94	0.94	0.90	0.94	0.97

Table 42: Comparators as estimators: median relative error, MSE relative to REML and to NCV (geometric means of per-setting ratios; range for NCV).

setting	estimand	fit	settings	rel. error	MSE/REML	MSE/NCV min	MSE/NCV max	MSE/NCV
plasmode	$E(Y X)$	REML	12	0.21	1.00	1.02	1.48	1.14
plasmode	$E(Y X)$	NCV	12	0.19	0.87	1.00	1.00	1.00
plasmode	$E(Y X)$	pcrc	12	0.21	0.97	0.86	1.30	1.11
plasmode	$E(Y X)$	GLS (FPCA cov.)	12	0.19	0.89	0.77	1.27	1.01
plasmode	$E(Y X)$	GLS (raw cov.)	12	0.18	0.79	0.67	1.21	0.91
plasmode	$\beta(s,t)$	REML	12	1.49	1.00	1.08	8.54	2.73
plasmode	$\beta(s,t)$	NCV	12	0.96	0.37	1.00	1.00	1.00
plasmode	$\beta(s,t)$	pcrc	12	0.78	0.26	0.23	1.25	0.72
plasmode	$\beta(s,t)$	GLS (FPCA cov.)	12	0.76	0.25	0.21	1.20	0.68
plasmode	$\beta(s,t)$	GLS (raw cov.)	12	0.89	0.35	0.47	2.06	0.94
plasmode	$\alpha(t)$	REML	12	0.64	1.00	0.86	1.47	1.07

Table 42: Comparators as estimators: median relative error, MSE relative to REML and to NCV (geometric means of per-setting ratios; range for NCV). (*continued*)

setting	estimand	fit	settings	rel. error	MSE/REML	MSE/NCV min	MSE/NCV max	MSE/NCV
plasmode	$\alpha(t)$	NCV	12	0.63	0.93	1.00	1.00	1.00
plasmode	$\alpha(t)$	pcr	12	0.65	0.95	0.71	1.37	1.01
plasmode	$\alpha(t)$	GLS (FPCA cov.)	12	0.66	0.86	0.58	1.24	0.92
plasmode	$\alpha(t)$	GLS (raw cov.)	12	0.61	0.73	0.34	1.16	0.78
plasmode	$\gamma(t)$	REML	12	0.85	1.00	0.91	1.71	1.17
plasmode	$\gamma(t)$	NCV	12	0.82	0.85	1.00	1.00	1.00
plasmode	$\gamma(t)$	pcr	12	0.89	0.86	0.66	2.21	1.01
plasmode	$\gamma(t)$	GLS (FPCA cov.)	12	0.87	0.80	0.64	1.54	0.94
plasmode	$\gamma(t)$	GLS (raw cov.)	12	0.75	0.79	0.63	1.34	0.93
syn. dep.	$E(Y X)$	REML	4	0.29	1.00	1.43	1.70	1.55
syn. dep.	$E(Y X)$	NCV	4	0.23	0.65	1.00	1.00	1.00
syn. dep.	$E(Y X)$	pcr	4	0.25	0.77	1.13	1.26	1.19
syn. dep.	$E(Y X)$	GLS (FPCA cov.)	4	0.25	0.77	1.14	1.25	1.19
syn. dep.	$E(Y X)$	GLS (raw cov.)	4	0.29	0.96	1.04	1.89	1.48
syn. dep.	$\beta(s, t)$	REML	4	0.65	1.00	5.76	10.45	7.41
syn. dep.	$\beta(s, t)$	NCV	4	0.24	0.13	1.00	1.00	1.00
syn. dep.	$\beta(s, t)$	pcr	4	0.31	0.24	1.41	2.50	1.75
syn. dep.	$\beta(s, t)$	GLS (FPCA cov.)	4	0.31	0.24	1.41	2.53	1.79
syn. dep.	$\beta(s, t)$	GLS (raw cov.)	4	0.52	0.51	1.61	7.25	3.75
syn. dep.	$\alpha(t)$	REML	4	0.32	1.00	1.03	1.10	1.06
syn. dep.	$\alpha(t)$	NCV	4	0.31	0.94	1.00	1.00	1.00
syn. dep.	$\alpha(t)$	pcr	4	0.32	0.99	1.04	1.06	1.05
syn. dep.	$\alpha(t)$	GLS (FPCA cov.)	4	0.32	0.99	1.04	1.06	1.05
syn. dep.	$\alpha(t)$	GLS (raw cov.)	4	0.41	1.58	1.04	2.69	1.68
syn. dep.	$\gamma(t)$	REML	4	0.24	1.00	1.04	1.12	1.07
syn. dep.	$\gamma(t)$	NCV	4	0.24	0.93	1.00	1.00	1.00
syn. dep.	$\gamma(t)$	pcr	4	0.24	0.97	1.01	1.09	1.04
syn. dep.	$\gamma(t)$	GLS (FPCA cov.)	4	0.24	0.96	1.01	1.07	1.03
syn. dep.	$\gamma(t)$	GLS (raw cov.)	4	0.25	0.99	0.94	1.24	1.06
syn. indep.	$E(Y X)$	REML	2	0.11	1.00	1.02	1.04	1.03
syn. indep.	$E(Y X)$	NCV	2	0.11	0.97	1.00	1.00	1.00
syn. indep.	$E(Y X)$	pcr	2	0.11	1.00	1.02	1.04	1.03
syn. indep.	$E(Y X)$	GLS (FPCA cov.)	2	0.11	1.00	1.02	1.04	1.03
syn. indep.	$E(Y X)$	GLS (raw cov.)	2	0.15	1.61	1.51	1.83	1.66
syn. indep.	$\beta(s, t)$	REML	2	0.13	1.00	1.17	1.53	1.34
syn. indep.	$\beta(s, t)$	NCV	2	0.11	0.75	1.00	1.00	1.00
syn. indep.	$\beta(s, t)$	pcr	2	0.13	0.98	1.14	1.51	1.31
syn. indep.	$\beta(s, t)$	GLS (FPCA cov.)	2	0.13	0.99	1.16	1.52	1.33
syn. indep.	$\beta(s, t)$	GLS (raw cov.)	2	0.14	1.13	1.12	2.06	1.52
syn. indep.	$\alpha(t)$	REML	2	0.15	1.00	0.96	0.96	0.96
syn. indep.	$\alpha(t)$	NCV	2	0.16	1.04	1.00	1.00	1.00
syn. indep.	$\alpha(t)$	pcr	2	0.15	1.00	0.96	0.96	0.96
syn. indep.	$\alpha(t)$	GLS (FPCA cov.)	2	0.15	1.00	0.96	0.96	0.96
syn. indep.	$\alpha(t)$	GLS (raw cov.)	2	0.30	3.37	2.14	4.90	3.24
syn. indep.	$\gamma(t)$	REML	2	0.12	1.00	0.96	0.96	0.96
syn. indep.	$\gamma(t)$	NCV	2	0.12	1.04	1.00	1.00	1.00
syn. indep.	$\gamma(t)$	pcr	2	0.12	1.00	0.96	0.97	0.96
syn. indep.	$\gamma(t)$	GLS (FPCA cov.)	2	0.12	1.00	0.96	0.96	0.96
syn. indep.	$\gamma(t)$	GLS (raw cov.)	2	0.12	1.11	0.97	1.18	1.07

CL2 on the GLS and pcr fits. GLS whitens each curve with $\hat{\Sigma}^{-1/2}$. The transform acts within curves, so curves stay independent clusters, and CL2 (Bayesian form, curves as clusters) applies to the whitened fit; it does not need $\hat{\Sigma}$ to be correct. pcr + CL2 is CL2 on the pcr fit. The comparison uses replicates 1–100, where all constructions exist; one pcr + CL2 replicate failed in a singular value decomposition and is left out for that construction only. Under dependent synthetic errors GLS-FPCA + CL2 covers β at 0.959 at 0.61 times REML + CL2’s width (interval score 0.55 times). On the plasmode data it covers β at 0.823 (by dataset 0.65–0.90; GLS-FPCA without CL2 0.731, REML + CL2 0.931), at 0.46 times the width. Its standard error is close to the spread of the GLS estimate there (geometric mean SE/SD 0.98). This is consistent with residual smoothing bias of the GLS estimate under the plasmode generator, but an aggregate ratio near one does not exclude local or tail errors of the standard error, and no single cause is established. pcr + CL2 behaves alike (β 0.831) and repairs pcr’s intervals for $\alpha(t)$ on synthetic data (0.56 to 0.95). GLS with the raw residual covariance fails with CL2 as well (0.65 under dependent synthetic errors): with more grid points than curves, the raw covariance absorbs the residuals themselves.

Table 43: CL2 on the GLS and pcre fits (replicates 1–100): mean and minimum coverage over settings or plasmode combinations, mean 5% quantile of pointwise coverage, width and interval score relative to REML + CL2, MSE relative to the NCV fit (geometric means), and detection rate, by data source, estimand and construction.

setting	estimand	interval	coverage	min cov.	q05	width rel.	IS rel.	MSE/NCV	detection
plasmode	$E(Y X)$	REML + CL2	0.94	0.92	0.87	1.00	1.00	1.15	0.97
plasmode	$E(Y X)$	NCV + CL2	0.89	0.85	0.74	0.81	1.04	1.00	0.99
plasmode	$E(Y X)$	GLS (FPCA cov.)	0.80	0.66	0.57	0.63	1.34	1.01	0.99
plasmode	$E(Y X)$	GLS (FPCA cov.) + CL2	0.89	0.83	0.74	0.79	1.02	1.01	0.99
plasmode	$E(Y X)$	GLS (raw cov.) + CL2	0.84	0.51	0.73	0.66	1.10	0.90	0.99
plasmode	$E(Y X)$	pcre	0.78	0.73	0.54	0.63	1.45	1.10	0.99
plasmode	$E(Y X)$	pcre + CL2	0.91	0.86	0.74	0.86	1.04	1.10	0.99
plasmode	$\beta(s,t)$	REML + CL2	0.93	0.91	0.87	1.00	1.00	2.75	0.38
plasmode	$\beta(s,t)$	NCV + CL2	0.76	0.60	0.46	0.46	1.13	1.00	0.54
plasmode	$\beta(s,t)$	GLS (FPCA cov.)	0.73	0.60	0.45	0.38	1.02	0.67	0.55
plasmode	$\beta(s,t)$	GLS (FPCA cov.) + CL2	0.82	0.63	0.56	0.46	0.83	0.67	0.48
plasmode	$\beta(s,t)$	GLS (raw cov.) + CL2	0.79	0.52	0.63	0.48	0.95	0.90	0.57
plasmode	$\beta(s,t)$	pcre	0.74	0.59	0.42	0.40	1.03	0.71	0.49
plasmode	$\beta(s,t)$	pcre + CL2	0.83	0.66	0.53	0.49	0.85	0.71	0.42
plasmode	$\alpha(t)$	REML + CL2	0.94	0.91	0.88	1.00	1.00	1.06	0.76
plasmode	$\alpha(t)$	NCV + CL2	0.89	0.79	0.77	0.85	1.12	1.00	0.79
plasmode	$\alpha(t)$	GLS (FPCA cov.)	0.75	0.57	0.62	0.58	1.52	0.93	0.86
plasmode	$\alpha(t)$	GLS (FPCA cov.) + CL2	0.87	0.77	0.77	0.74	1.15	0.93	0.80
plasmode	$\alpha(t)$	GLS (raw cov.) + CL2	0.83	0.56	0.74	0.61	1.10	0.77	0.86
plasmode	$\alpha(t)$	pcre	0.77	0.59	0.63	0.62	1.54	1.02	0.85
plasmode	$\alpha(t)$	pcre + CL2	0.89	0.80	0.77	0.83	1.15	1.02	0.78
plasmode	$\gamma(t)$	REML + CL2	0.94	0.92	0.89	1.00	1.00	1.19	0.41
plasmode	$\gamma(t)$	NCV + CL2	0.90	0.86	0.79	0.84	1.01	1.00	0.45
plasmode	$\gamma(t)$	GLS (FPCA cov.)	0.79	0.50	0.63	0.65	1.34	0.96	0.52
plasmode	$\gamma(t)$	GLS (FPCA cov.) + CL2	0.85	0.65	0.74	0.73	1.10	0.96	0.46
plasmode	$\gamma(t)$	GLS (raw cov.) + CL2	0.82	0.52	0.73	0.64	1.16	0.94	0.55
plasmode	$\gamma(t)$	pcre	0.80	0.45	0.65	0.71	1.33	1.02	0.49
plasmode	$\gamma(t)$	pcre + CL2	0.87	0.54	0.74	0.83	1.13	1.02	0.40
syn. dep.	$E(Y X)$	REML + CL2	0.94	0.93	0.90	1.00	1.00	1.59	0.98
syn. dep.	$E(Y X)$	NCV + CL2	0.93	0.92	0.87	0.76	0.79	1.00	0.99
syn. dep.	$E(Y X)$	GLS (FPCA cov.)	0.93	0.91	0.87	0.81	0.87	1.19	0.99
syn. dep.	$E(Y X)$	GLS (FPCA cov.) + CL2	0.94	0.93	0.90	0.86	0.87	1.19	0.99
syn. dep.	$E(Y X)$	GLS (raw cov.) + CL2	0.59	0.28	0.51	0.39	2.11	1.50	1.00
syn. dep.	$E(Y X)$	pcre	0.90	0.88	0.84	0.76	0.89	1.19	0.99
syn. dep.	$E(Y X)$	pcre + CL2	0.98	0.97	0.95	1.05	0.91	1.19	0.98
syn. dep.	$\beta(s,t)$	REML + CL2	0.94	0.93	0.89	1.00	1.00	7.72	0.66
syn. dep.	$\beta(s,t)$	NCV + CL2	0.93	0.92	0.87	0.36	0.37	1.00	0.95
syn. dep.	$\beta(s,t)$	GLS (FPCA cov.)	0.95	0.93	0.91	0.58	0.55	1.84	0.87
syn. dep.	$\beta(s,t)$	GLS (FPCA cov.) + CL2	0.96	0.94	0.92	0.61	0.55	1.84	0.86
syn. dep.	$\beta(s,t)$	GLS (raw cov.) + CL2	0.65	0.36	0.57	0.36	1.43	3.87	0.91
syn. dep.	$\beta(s,t)$	pcre	0.96	0.93	0.92	0.59	0.55	1.81	0.87
syn. dep.	$\beta(s,t)$	pcre + CL2	0.98	0.96	0.96	0.71	0.58	1.81	0.82
syn. dep.	$\alpha(t)$	REML + CL2	0.95	0.94	0.91	1.00	1.00	1.07	0.98
syn. dep.	$\alpha(t)$	NCV + CL2	0.93	0.93	0.86	0.93	0.98	1.00	0.99
syn. dep.	$\alpha(t)$	GLS (FPCA cov.)	0.92	0.89	0.84	0.90	1.03	1.06	1.00
syn. dep.	$\alpha(t)$	GLS (FPCA cov.) + CL2	0.93	0.92	0.84	0.95	1.02	1.06	0.99
syn. dep.	$\alpha(t)$	GLS (raw cov.) + CL2	0.52	0.21	0.46	0.42	3.08	1.65	1.00
syn. dep.	$\alpha(t)$	pcre	0.56	0.51	0.47	0.40	2.37	1.05	1.00
syn. dep.	$\alpha(t)$	pcre + CL2	0.95	0.94	0.92	1.00	0.99	1.05	0.99
syn. dep.	$\gamma(t)$	REML + CL2	0.95	0.95	0.92	1.00	1.00	1.05	0.92
syn. dep.	$\gamma(t)$	NCV + CL2	0.93	0.92	0.87	0.91	0.99	1.00	0.94
syn. dep.	$\gamma(t)$	GLS (FPCA cov.)	0.93	0.91	0.89	0.90	0.98	1.02	0.94
syn. dep.	$\gamma(t)$	GLS (FPCA cov.) + CL2	0.94	0.93	0.90	0.95	0.99	1.02	0.93
syn. dep.	$\gamma(t)$	GLS (raw cov.) + CL2	0.61	0.33	0.54	0.41	2.06	1.04	0.99
syn. dep.	$\gamma(t)$	pcre	0.94	0.93	0.89	0.92	0.97	1.03	0.93
syn. dep.	$\gamma(t)$	pcre + CL2	0.98	0.98	0.96	1.22	1.08	1.03	0.87
syn. indep.	$E(Y X)$	REML + CL2	0.96	0.96	0.92	1.00	1.00	1.03	1.00
syn. indep.	$E(Y X)$	NCV + CL2	0.95	0.95	0.90	0.92	0.97	1.00	1.00
syn. indep.	$E(Y X)$	GLS (FPCA cov.)	0.97	0.97	0.94	1.02	0.98	1.03	1.00
syn. indep.	$E(Y X)$	GLS (FPCA cov.) + CL2	0.96	0.96	0.92	1.00	1.00	1.03	1.00
syn. indep.	$E(Y X)$	GLS (raw cov.) + CL2	0.46	0.14	0.38	0.28	3.45	1.66	1.00
syn. indep.	$E(Y X)$	pcre	0.97	0.97	0.94	1.03	0.99	1.03	1.00
syn. indep.	$E(Y X)$	pcre + CL2	0.97	0.97	0.94	1.02	1.01	1.03	1.00
syn. indep.	$\beta(s,t)$	REML + CL2	0.99	0.99	0.97	1.00	1.00	1.27	0.97
syn. indep.	$\beta(s,t)$	NCV + CL2	0.98	0.98	0.94	0.75	0.78	1.00	0.98
syn. indep.	$\beta(s,t)$	GLS (FPCA cov.)	0.99	0.99	0.97	1.02	1.01	1.26	0.97
syn. indep.	$\beta(s,t)$	GLS (FPCA cov.) + CL2	0.99	0.99	0.97	1.00	1.00	1.26	0.97
syn. indep.	$\beta(s,t)$	GLS (raw cov.) + CL2	0.64	0.38	0.54	0.37	1.84	1.44	0.99
syn. indep.	$\beta(s,t)$	pcre	0.99	0.99	0.98	1.03	1.02	1.25	0.97
syn. indep.	$\beta(s,t)$	pcre + CL2	0.99	0.99	0.97	1.01	1.01	1.25	0.97
syn. indep.	$\alpha(t)$	REML + CL2	0.95	0.95	0.92	1.00	1.00	0.98	1.00
syn. indep.	$\alpha(t)$	NCV + CL2	0.95	0.94	0.90	0.98	1.01	1.00	1.00
syn. indep.	$\alpha(t)$	GLS (FPCA cov.)	0.96	0.96	0.92	1.01	0.98	0.98	1.00
syn. indep.	$\alpha(t)$	GLS (FPCA cov.) + CL2	0.95	0.95	0.92	1.00	1.00	0.98	1.00
syn. indep.	$\alpha(t)$	GLS (raw cov.) + CL2	0.36	0.06	0.30	0.25	6.22	3.36	1.00

Table 43: CL2 on the GLS and pcre fits (replicates 1–100): mean and minimum coverage over settings or plasmode combinations, mean 5% quantile of pointwise coverage, width and interval score relative to REML + CL2, MSE relative to the NCV fit (geometric means), and detection rate, by data source, estimand and construction. (*continued*)

setting	estimand	interval	coverage	min cov.	q05	width rel.	IS rel.	MSE/NCV	detection
syn. indep.	$\alpha(t)$	pcre	0.96	0.96	0.92	1.00	0.98	0.98	1.00
syn. indep.	$\alpha(t)$	pcre + CL2	0.96	0.95	0.92	1.00	1.00	0.98	1.00
syn. indep.	$\gamma(t)$	REML + CL2	0.96	0.95	0.92	1.00	1.00	0.96	1.00
syn. indep.	$\gamma(t)$	NCV + CL2	0.95	0.95	0.88	0.96	1.01	1.00	1.00
syn. indep.	$\gamma(t)$	GLS (FPCA cov.)	0.97	0.96	0.93	1.02	0.98	0.96	1.00
syn. indep.	$\gamma(t)$	GLS (FPCA cov.) + CL2	0.96	0.96	0.92	1.00	1.00	0.96	1.00
syn. indep.	$\gamma(t)$	GLS (raw cov.) + CL2	0.48	0.16	0.42	0.26	2.87	1.06	1.00
syn. indep.	$\gamma(t)$	pcre	0.97	0.97	0.94	1.04	0.99	0.96	1.00
syn. indep.	$\gamma(t)$	pcre + CL2	0.96	0.96	0.92	1.03	1.01	0.96	1.00

S3.3 Functional GEE

Their simulation design

We reran the function-on-scalar simulation of Loewinger et al. (2025, sec. 4) with their data generator unchanged: $Y_{ij}(s) = \beta_0(s) + X_{1i}\beta_1(s) + X_{2i}\beta_2(s) + \varepsilon_{ij}(s)$ on 50 grid points, with Gaussian errors (variance 10) or Poisson and binary responses generated through a Gaussian copula whose latent correlation is AR(1) (0.75) along s and exchangeable (0.75) between the curves of a subject. The design crosses the three families with 25, 100, or 500 subjects and 5 or 1 curves per subject (5 as in their study, 1 as in ours), with 200 replicates per setting; 1 one-step fit failed (singular system). The initial fit is their pffr call (bam with fREML, discretisation, P-spline bases with 16 functions); all intervals labelled “initial fit” in Table 44 are computed on that fit, the REML row on a refit of the same model and bases by gam with REML, and all clusters are subjects. Every interval is scored through one evaluation matrix per coefficient function; these matrices reproduce fastFGEE’s fitted curves and standard errors, so centring and basis differences between methods do not enter. As in their Table 3, coverage is the pointwise 95% coverage averaged over the grid, over replicates and over β_0 , β_1 and β_2 ; its Monte Carlo standard error is at most 0.010.

Table 44: Their simulation design: ranges over the nine settings (three families, three numbers of subjects) of the coverage of nominal 95% pointwise intervals, averaged over the grid and the three coefficient functions, by number of curves per subject. Initial fit: their pffr call (bam). fGEE: one-step update with the stated functional and longitudinal working correlations. Calibrated crit.: fastFGEE’s wild-bootstrap critical value with its t adjustment; z: normal critical value.

Intervals	5 curves per subject	1 curve per subject
Initial fit, model-based	0.43–0.57	0.65–0.79
Initial-fit sandwich of fastFGEE, z	0.60–0.87	0.61–0.80
Initial-fit sandwich of fastFGEE, calibrated crit.	0.94–1.00	0.95–1.00
Initial-fit sandwich of fastFGEE, penalty rescaled, z	0.90–0.95	0.87–0.94
Initial fit, cluster sandwich, frequentist, centred scores	0.90–0.95	0.87–0.94
Initial fit, CR1	0.90–0.95	0.89–0.95
Initial fit, CL2	0.92–0.95	0.92–0.95
REML (gam) fit, CL2	0.92–0.95	0.92–0.95
fGEE, independence $\times$ independence, z	0.74–0.94	0.75–0.94
fGEE, independence $\times$ independence, calibrated crit.	0.93–0.96	0.90–0.95
fGEE, AR(1) $\times$ exchangeable, z	0.75–0.94	0.71–0.94
fGEE, AR(1) $\times$ exchangeable, calibrated crit.	0.94–0.96	0.88–0.95

With 5 curves per subject, CL2 clustered by subject on the initial fit covers 0.94–0.95 with 100 or 500 subjects and 0.92–0.93 with 25. Refitting by gam and REML changes the pooled CL2 coverage by at most 0.013. fastFGEE’s one-step intervals with its calibrated critical values cover 0.93–0.96 in this design; with a normal critical value they cover 0.74–0.92 with at most 100 subjects and 0.92–0.94 with 500. The one-step estimate of β_1 has an RMSE of 0.85–1.01 (independence) and 0.81–1.04 (AR(1) $\times$ exchangeable) times that of the initial fit.

The initial-fit sandwich and penalty-scale conventions

mgcv’s smoothing parameters are on the scale of the summed log-likelihood: the penalized estimate maximizes $\sum_{i=1}^n \ell_i(\theta) - \theta^\top S_\lambda \theta / (2\phi)$, with $S_\lambda = \sum_j \lambda_j S_j$, n clusters (subjects) and the sum over all observations of cluster i in ℓ_i (for the Gaussian, the residual sum of squares). A sandwich written with averaged quantities therefore needs the bread

$$\bar{A} = \frac{1}{n} \sum_{i=1}^n \mathbf{X}_i^\top \mathbf{W}_i \mathbf{X}_i + \frac{S_\lambda}{n\phi} = (n \mathbf{V}_p)^{-1},$$

with $\mathbf{W}_i$ the Fisher weights of the main text, and $\bar{A}^{-1} \bar{M} \bar{A}^{-1} / n$ with the averaged meat $\bar{M}$ is then the cluster sandwich $\mathbf{V}_p(\sum_i \mathbf{U}_i \mathbf{U}_i^\top) \mathbf{V}_p$ up to degrees-of-freedom factors.

The initial-fit sandwich of fastFGEE 0.2.2 (`gee.fit = FALSE`), which the simulation code of Loewinger et al. (2025) calls for their pffr baseline, differs in the penalty scale: it adds S_λ to the averaged bread, a penalty larger by the factor $n\phi$ (for the Gaussian, approximately $n\sigma^2$). The estimate is the unchanged pffr estimate, but the inverse bread, and with it the standard errors, shrink. In their design the ratio of mean standard error to the empirical standard deviation of $\hat{\beta}_1$ is 0.44–0.76 for this sandwich and does not improve with the number of subjects (0.45–0.69 with 25, 0.44–0.76 with 500), because the penalty-scale difference does not vanish. With 5 curves per subject its pooled coverage is 0.60–0.64 for Gaussian, 0.84–0.87 for Poisson and 0.72–0.76 for binary responses (Table 44); its coverage here is in line with their Table 3 for the pffr baseline. With its wild-bootstrap critical value, which widens the intervals instead, coverage is 0.97–1.00 with 100 or more subjects.

Dividing the penalty by $n\phi$ and leaving everything else unchanged gives 0.87–0.95 over all 18 settings. The rescaled sandwich then coincides with refund’s frequentist cluster sandwich (no $\mathbf{V}_p - \mathbf{V}_e$ term) with centred cluster scores: pooled coverages agree to three decimals in every Poisson and binary setting and differ by at most 0.008 in the Gaussian settings, where fastFGEE estimates the working variance per grid point instead of using one σ^2 . CL2 on the same fit covers 0.94–0.95 with 100 or 500 subjects and 0.92–0.93 with 25; the lowest single coefficient function is 0.901 (β_0 , binary, 25 subjects, 1 curve per subject). Its gain over the rescaled sandwich at 25 subjects, 0.025–0.048, combines uncentred scores, the $\mathbf{V}_p - \mathbf{V}_e$ term and the leverage adjustment.

The main factorial settings of our simulation

fastFGEE was started from the REML fit of the main study in its 18 main factorial settings at the default signal level (Gaussian, Poisson and binary responses; independent, rough stationary and smooth errors; $G = 40$ and 100; model $\text{ff}(X) + \gamma(t)$; one curve per subject; 200 replicates, the main study’s data sets), with curves as clusters, independent longitudinal correlation, functional working correlation independence, AR(1) or FPCA, and the package defaults otherwise (cluster cross-validation of the smoothing parameters, wild bootstrap). Its interval interface covers the functional intercept and scalar-covariate effects but not ff surfaces or fitted means. For $\beta(s, t)$ and the mean we therefore score Wald intervals from its covariance with the normal critical value. For $\alpha(t)$ and $\gamma(t)$ we also transfer its calibrated critical value for the term (wild bootstrap with t adjustment) to our evaluation points and apply it to the same estimate and standard error; the calibration on the package’s own term grid need not carry over to other points. Monte Carlo standard errors of coverage are at most 0.018.

Table 45: Main factorial settings of our simulation: ranges of grid-averaged coverage of nominal 95% pointwise intervals over the 12 settings with dependent errors (rough stationary, smooth) and the 6 settings with independent errors, and the ratio SE/SD (grid average of mean standard error over the empirical SD of the estimate) under dependence. fGEE intervals: Wald intervals from its covariance, or with its transferred calibrated critical value (calibrated crit.). Mean on the link scale.

Estimand	Intervals	Cov. dependent	Cov. independent	SE/SD dependent
$\beta(s,t)$	REML, model-based	0.61–0.82	0.98–0.99	0.41–0.62
	REML + CL2	0.93–0.95	0.97–0.99	0.83–0.94
	fGEE, independence	0.73–0.89	0.83–0.90	0.46–0.76
	fGEE, AR(1)	0.72–0.89	0.83–0.90	0.45–0.76
	fGEE, FPCA	0.67–0.86	0.82–0.89	0.41–0.68
mean (link)	REML, model-based	0.61–0.75	0.95–0.97	0.44–0.59
	REML + CL2	0.92–0.95	0.93–0.96	0.92–1.00
	fGEE, independence	0.81–0.91	0.83–0.90	0.70–0.91
	fGEE, AR(1)	0.81–0.90	0.83–0.90	0.70–0.91
	fGEE, FPCA	0.71–0.88	0.83–0.90	0.55–0.84
$\alpha(t)$	REML, model-based	0.61–0.75	0.90–0.96	0.45–0.59
	REML + CL2	0.88–0.95	0.86–0.96	0.89–1.02
	fGEE, independence	0.79–0.92	0.79–0.91	0.72–0.95
	fGEE, AR(1)	0.79–0.92	0.79–0.91	0.72–0.94
	fGEE, FPCA	0.66–0.89	0.79–0.91	0.55–0.88
	fGEE, independence, calibrated crit.	0.91–0.95	0.83–0.94	
	fGEE, AR(1), calibrated crit.	0.90–0.95	0.83–0.94	
	fGEE, FPCA, calibrated crit.	0.82–0.94	0.84–0.94	
$\gamma(t)$	REML, model-based	0.62–0.74	0.93–0.96	0.47–0.59
	REML + CL2	0.92–0.95	0.89–0.96	0.93–1.03
	fGEE, independence	0.77–0.91	0.78–0.91	0.70–0.94
	fGEE, AR(1)	0.76–0.90	0.78–0.91	0.70–0.94
	fGEE, FPCA	0.70–0.87	0.78–0.90	0.56–0.86
	fGEE, independence, calibrated crit.	0.89–0.93	0.83–0.93	
	fGEE, AR(1), calibrated crit.	0.89–0.93	0.83–0.93	
	fGEE, FPCA, calibrated crit.	0.82–0.91	0.84–0.93	

The Wald intervals for β and the mean use the sandwich covariance fastFGEE 0.2.2 returns for its one-step estimate under the stated working correlation. These surface results therefore do not test fastFGEE’s calibrated bootstrap procedure, which it provides only for intercepts and scalar-covariate effects, nor its performance in its own function-on-scalar setting (studied above). Under dependent errors the fGEE Wald intervals for β cover 0.67–0.89 against 0.93–0.95 for REML + CL2 (Table 45). Smoothing bias alone does not explain the shortfall: the root mean squared bias of the fGEE surface estimate is 0.11–0.35 of its empirical standard deviation (REML 0.09–0.13), while SE/SD is 0.41–0.76. With calibrated critical values, the intervals for α and γ cover 0.82–0.95 under dependence.

Table 46: Main factorial settings: paired MSE ratios on identical data sets (ratio of mean squared errors over replicates, grid-averaged), ranges over settings. fGEE: independence working correlation. In brackets: number of settings whose 95% replicate-bootstrap interval lies entirely below one / entirely above one.

Estimand	Errors	Settings	fGEE / REML	NCV / REML	fGEE / NCV
$\beta(s,t)$	dependent	12	0.22–0.53 [12/0]	0.10–0.22 [12/0]	1.32–3.70 [0/12]
	independent	6	0.71–2.21 [2/3]	0.65–1.15 [3/0]	1.01–1.93 [0/5]
$\alpha(t)$	dependent	12	0.65–0.99 [10/0]	0.56–0.97 [12/0]	0.97–1.16 [0/9]
	independent	6	1.07–1.10 [0/6]	1.04–1.10 [0/6]	0.99–1.05 [0/4]
$\gamma(t)$	dependent	12	0.75–1.01 [5/0]	0.69–1.00 [6/0]	0.98–1.10 [0/6]
	independent	6	1.10–1.20 [0/6]	1.04–1.17 [0/6]	1.03–1.08 [0/5]

Under dependent errors the one-step update reduces the MSE of β relative to REML (0.22–0.53), by less than curve-blocked NCV (0.10–0.22): fGEE/NCV is 1.32–3.70, with every bootstrap interval above one (lowest lower bound 1.09). Choosing, after the fact, the working correlation with the lowest MSE in each setting gives 1.27–3.70. Under independent errors fGEE/REML for β is 0.71–2.21, above one in the settings with $G = 40$ (1.06–2.21). For α and γ , fGEE/NCV is 0.97–1.16.

Plasmode datasets

The plasmode comparison uses the six datasets with independent curves, the midpoint truth (refit at the geometric mean of the REML and NCV smoothing parameters), the REML residual curves attached to their own subjects with signs flipped per curve, all subjects and 40 subjects, 200 replicates per combination, and the plasmode study’s own data sets; methods and interval definitions are those of the main factorial settings (Gaussian responses). For the two cardiology datasets only aggregate results are reported. Monte Carlo standard errors of coverage are at most 0.025.

Table 47: Plasmode datasets (midpoint truth, REML residuals), $\beta(s,t)$: grid-averaged coverage of nominal 95% pointwise intervals (Model, CL2: REML model-based and REML + CL2; Indep., AR(1), FPCA: fGEE Wald intervals from its covariance under that working correlation), and paired MSE ratios with 95% replicate-bootstrap intervals (fGEE with independence working correlation).

Dataset	G	Model	CL2	Indep.	AR(1)	FPCA	fGEE / REML	fGEE / NCV
ECG strain	78	0.56	0.94	0.87	0.83	0.87	0.87 [0.85, 0.90]	0.94 [0.92, 0.96]
	40	0.53	0.93	0.81	0.79	0.78	0.74 [0.71, 0.76]	0.95 [0.92, 0.97]
AF trial	118	0.67	0.94	0.78	0.78	0.69	0.39 [0.21, 0.70]	1.35 [0.75, 2.47]
	40	0.62	0.92	0.59	0.64	0.67	0.88 [0.35, 1.48]	4.72 [1.92, 9.29]
running	90	0.52	0.92	0.75	0.75	0.70	0.54 [0.51, 0.59]	1.05 [0.97, 1.15]
	40	0.47	0.91	0.69	0.73	0.71	0.59 [0.52, 0.67]	2.04 [1.61, 2.59]
DTI	92	0.64	0.93	0.85	0.84	0.82	0.66 [0.63, 0.69]	1.08 [1.05, 1.12]
	40	0.63	0.92	0.76	0.76	0.72	0.67 [0.63, 0.71]	1.25 [1.16, 1.37]
gait	138	0.52	0.94	0.83	0.87	0.82	0.61 [0.55, 0.66]	1.09 [0.99, 1.21]
	40	0.47	0.94	0.74	0.80	0.76	0.53 [0.46, 0.59]	2.09 [1.63, 2.75]
ECG 8-lead	100	0.64	0.94	0.81	0.83	0.83	0.45 [0.36, 0.55]	2.42 [1.66, 3.68]
	40	0.64	0.94	0.77	0.79	0.75	0.61 [0.50, 0.75]	5.25 [3.73, 7.48]

Over the 12 dataset and G combinations, REML + CL2 covers β at 0.91–0.94, the mean at 0.92–0.95, α at 0.89–0.95

and γ at 0.92–0.96; model-based REML intervals cover β at 0.47–0.67 (Table 47). The fastFGEE 0.2.2 Wald intervals cover β at 0.59–0.87 with SE/SD 0.43–0.87 (REML + CL2: 0.80–1.05), the mean at 0.74–0.93, α at 0.64–0.93 and γ at 0.49–0.92, the lowest values for the randomised treatment effect in the AF trial (0.49–0.67). With calibrated critical values the intervals for α and γ cover 0.80–1.00.

For β , the MSE ratio fGEE/REML is 0.39–0.88, below one in 12 of 12 combinations, with the interval below one in 11. Against curve-blocked NCV it is worse in 7 combinations, better in 2 (ECG strain) and not resolved in the remaining 3; the best working correlation per combination, chosen after the fact, gives fGEE/NCV 0.89–4.06. For γ , fGEE/NCV is 0.59–1.10, with the interval below one in 7 combinations (gait, ECG 8-lead, ECG strain, and AF trial) and above one in 2; for α it is 0.72–1.14.

S4 Approaches that did not work

Each approach is given as method, derivation, result and reason. The notation is that of §2 and Appendix A of the paper: $\mathbf{W}$ holds the working weights including $1/\phi$, $\mathbf{S} = \sum_k (\lambda_k/\phi) \mathbf{S}_k$, $\mathbf{V}_p = (\mathbf{X}^\top \mathbf{W} \mathbf{X} + \mathbf{S})^{-1}$ and $\mathbf{V}_e = \mathbf{V}_p \mathbf{X}^\top \mathbf{W} \mathbf{X} \mathbf{V}_p$. In addition, $\mathbf{F} = \mathbf{X}^\top \mathbf{W} \mathbf{X} + \mathbf{S} = \mathbf{V}_p^{-1}$ is the penalized Hessian, $\rho = \log \lambda$ and $\dot{\mathbf{S}}_k = \partial \mathbf{S} / \partial \rho_k = (\lambda_k/\phi) \mathbf{S}_k$. For a canonical link the working weights are $w_{gd} = (d\mu/d\eta)_{gd}/\phi$, and the score of curve g is $\mathbf{u}_g = \mathbf{X}_g^\top (\mathbf{y}_g - \mu_g)/\phi$. The CL2 influence of curve g is $\mathbf{k}_g = \mathbf{V}_p \mathbf{U}_g$ with $\mathbf{U}_g$ as in Appendix A, so that $\hat{\mathbf{V}}_{\text{CL2}} = \mathbf{V}_f + (\mathbf{V}_p - \mathbf{V}_e)$ with the frequentist form $\mathbf{V}_f = \frac{G}{G-1} \sum_g \mathbf{k}_g \mathbf{k}_g^\top$. $\mathbf{L}$ maps the coefficients to an estimand's values on its grid.

S4.1 Intervals centred at the NCV fit

Under dependent errors curve-blocked NCV estimates β with a fraction of REML's error (S1.6), but NCV + CL2 undercovers. Below: why, and six constructions that try to repair it. None of the constructions as implemented here gives intervals around the NCV estimate that hold their level across the plasmode datasets; the conclusions refer to these constructions, not to every possible correction.

(a) Why NCV + CL2 undercovers

Method. Two diagnostics. *Bias-eliminated coverage* replaces each interval's centre by the estimate minus its pointwise bias (estimated from the replicates) and shows the coverage without bias. A refit with every smoothing parameter fixed at the median NCV value of its combination removes the variability of the selected smoothing parameters. It also moves every replicate to the typical amount of smoothing, so it changes the bias as well as the variability; it uses the replicates themselves and is a diagnostic, not a decomposition of the shortfall and not a method.

Result. In the dependent synthetic settings at $G = 100$, NCV + CL2 covers β at 0.927 (smooth truth) and 0.828 (rough truth); without bias it would cover 0.948 and 0.939. On the plasmode datasets with independent source curves the ratio of the root mean square standard error to the replicate SD of the estimate (SE/SD, median over grid points) is 0.75–0.98 for NCV + CL2; with fixed smoothing parameters it is 1.03–1.15, but coverage stays at 0.69–0.92.

Reason. On synthetic data the shortfall is smoothing bias at the amount of smoothing NCV selects. On the plasmode data the diagnostics suggest two contributions: the standard error is short, consistent with the variability of the selected smoothing parameters that it ignores, and coverage stays short at a fixed typical amount of smoothing, consistent with smoothing bias. Fixing the smoothing parameters changes both the centre and its sampling distribution, so the two cannot be separated exactly.

Table 48: Plasmode, $\beta(s, t)$, all subjects: coverage (5% quantile of pointwise coverage) and SE/SD for NCV + CL2, its bias-eliminated coverage, NCV + CL2 after a refit at the median NCV smoothing parameters, and REML + CL2 (100 replicates per combination).

dataset	NCV + CL2		bias removed	NCV, λ fixed		REML + CL2	
ECG strain	0.91	(0.80) 0.98	0.93	0.92	(0.85) 1.03	0.94	(0.89) 1.01
AF trial	0.81	(0.52) 0.84	0.87	0.87	(0.57) 1.07	0.94	(0.86) 0.96
running	0.71	(0.40) 0.75	0.82	0.69	(0.29) 1.04	0.91	(0.84) 0.90
DTI	0.85	(0.58) 0.95	0.92	0.86	(0.57) 1.07	0.93	(0.86) 0.99
gait	0.81	(0.53) 0.90	0.88	0.88	(0.62) 1.07	0.94	(0.89) 1.00
ECG 8-lead	0.76	(0.48) 0.79	0.85	0.85	(0.46) 1.15	0.94	(0.88) 0.95
ocean	0.71	(0.29) 0.89	0.88	0.71	(0.24) 1.03	0.93	(0.87) 0.93
weather	0.59	(0.39) 0.68	0.70	0.69	(0.29) 1.14	0.88	(0.82) 0.93
electricity	0.62	(0.40) 0.69	0.76	0.64	(0.30) 1.06	0.93	(0.88) 0.89

(b) Widening by the NCV–REML difference

Method. Widen the NCV + CL2 interval in quadrature by the difference of the two estimates, $\hat{L}\hat{\theta}_N \pm z\{\text{se}_{\text{CL2},N}^2 + (\hat{L}\hat{\theta}_N - \hat{L}\hat{\theta}_R)^2\}^{1/2}$, with the difference standing in for the bias of the NCV estimate. If the bias b were known, an interval $\pm z(\sigma^2 + b^2)^{1/2}$ would never undercover at the 95% level:

with $r = b/\sigma$ its coverage is $c(r) = \Phi\{z(1+r^2)^{1/2} - r\} - \Phi\{-z(1+r^2)^{1/2} - r\}$, and $c(r) - c(0) = z(z^2 - 3)\varphi(z)r^4/6 + O(r^6)$. Numerically $c(r) \geq c(0)$ for every r exactly when $z \geq \sqrt{3}$, that is for nominal levels of at least 0.917.

Derivation. The interval covers when $|Z + r| \leq z(1 + r^2)^{1/2}$ with $Z \sim N(0, 1)$. Write $u = z(1 + r^2)^{1/2} = z + zr^2/2 - zr^4/8 + O(r^6)$. Since $\Phi(-u - r) = 1 - \Phi(u + r)$,

$$c(r) = \Phi(u - r) + \Phi(u + r) - 1 = 2\Phi(u) - 1 + \varphi'(u)r^2 + \varphi'''(u)r^4/12 + O(r^6),$$

an even function of r . With $\varphi'(z) = -z\varphi(z)$, $\varphi''(z) = (z^2 - 1)\varphi(z)$ and $\varphi'''(z) = (3z - z^3)\varphi(z)$, expanding u around z gives $2\Phi(u) = 2\Phi(z) + z\varphi(z)r^2 - (z + z^3)\varphi(z)r^4/4 + O(r^6)$ and $\varphi'(u)r^2 = -z\varphi(z)r^2 + z(z^2 - 1)\varphi(z)r^4/2 + O(r^6)$. The terms in r^2 cancel, and the terms in r^4 add up to $\varphi(z)\{-(z + z^3)/4 + z(z^2 - 1)/2 + (3z - z^3)/12\} = z(z^2 - 3)\varphi(z)/6$. For small biases the widened interval therefore overcovers if $z > \sqrt{3}$ and undercovers if $z < \sqrt{3}$; the global statement for all r is the numerical check above.

The plug-in replaces b by the estimated difference. For a contrast $\mathbf{a}$ let $\hat{\delta} = \mathbf{a}^\top(\hat{\theta}_N - \hat{\theta}_R)$, with biases b_N, b_R and variances σ_N^2, σ_R^2 , covariance σ_{NR} of the two contrasts. Then

$$E\hat{\delta}^2 = (b_N - b_R)^2 + \sigma_N^2 + \sigma_R^2 - 2\sigma_{NR},$$

so $\text{se}_{\text{CL2},N}^2 + \hat{\delta}^2$ estimates $\sigma_N^2 + b_N^2$ only if $b_R \approx 0$ and $\sigma_R^2 \approx 2\sigma_{NR}$. Under within-curve dependence σ_R^2 is large (REML undersmooths), which inflates the allowance, while a bias shared by both fits cancels in $b_N - b_R$.

Result. In the dependent synthetic settings with a smooth truth the widened interval overcovers β (0.981); under the REML-fitted plasmode truth it covers 0.873 with a lower tail of 0.74 (next table and figure).

Reason. The estimated difference is not the bias. It carries REML's sampling variance, which is large under dependence, so the interval is too wide where REML is noisy; and it cannot see bias that both fits share, which concentrates where the truth is most curved.

(c) NCV estimate with the REML + CL2 width

Method. Centre REML + CL2's pointwise half-width at the NCV estimate.

Result. Its detection rate for β is lower than REML + CL2's on 9 of 9 plasmode datasets (0.05–0.51 against 0.22–0.62). Under the REML-fitted plasmode truth it covers 0.906 with a lower tail of 0.67 (REML + CL2: 0.924, 0.87).

Reason. Its width equals REML + CL2's by construction, so a lower interval score only says that the NCV estimate is closer to the truth, and the wide band around a smoother centre excludes zero less often. Its coverage relies on REML's standard error absorbing NCV's bias, which fails when the truth is rougher than the NCV fit.

Table 49: $\beta(s, t)$: coverage (5% quantile of pointwise coverage) and width / interval score relative to REML + CL2 (geometric means of per-setting ratios). Synthetic: main factorial at $G = 100$ and rough-truth settings with smooth errors; plasmode: all subjects, fitted residuals, by truth source. NCV–REML widening: NCV + CL2 widened in quadrature by the NCV–REML difference.

setting	combinations	REML + CL2	NCV + CL2	NCV–REML widening	NCV estimate, REML + CL2 width
dependent, rough truth	7	0.943 (0.92) 1.00 / 1.00	0.828 (0.54) 0.47 / 0.78	0.938 (0.83) 0.86 / 0.82	0.988 (0.96) 1.00 / 0.83
dependent, smooth truth	14	0.945 (0.92) 1.00 / 1.00	0.927 (0.85) 0.40 / 0.45	0.981 (0.94) 0.82 / 0.69	0.997 (0.99) 1.00 / 0.81
independent	7	0.985 (0.96) 1.00 / 1.00	0.981 (0.94) 0.77 / 0.79	0.988 (0.95) 0.85 / 0.84	0.991 (0.97) 1.00 / 1.00
plasmode, NCV-fitted truth	18	0.925 (0.88) 1.00 / 1.00	0.828 (0.66) 0.37 / 0.68	0.938 (0.86) 0.91 / 0.83	0.970 (0.92) 1.00 / 0.82
plasmode, REML-fitted truth	18	0.924 (0.87) 1.00 / 1.00	0.659 (0.36) 0.51 / 2.17	0.873 (0.74) 1.02 / 1.23	0.906 (0.67) 1.00 / 1.04
plasmode, midpoint truth	18	0.932 (0.89) 1.00 / 1.00	0.765 (0.45) 0.44 / 1.09	0.918 (0.80) 0.91 / 0.93	0.962 (0.88) 1.00 / 0.86

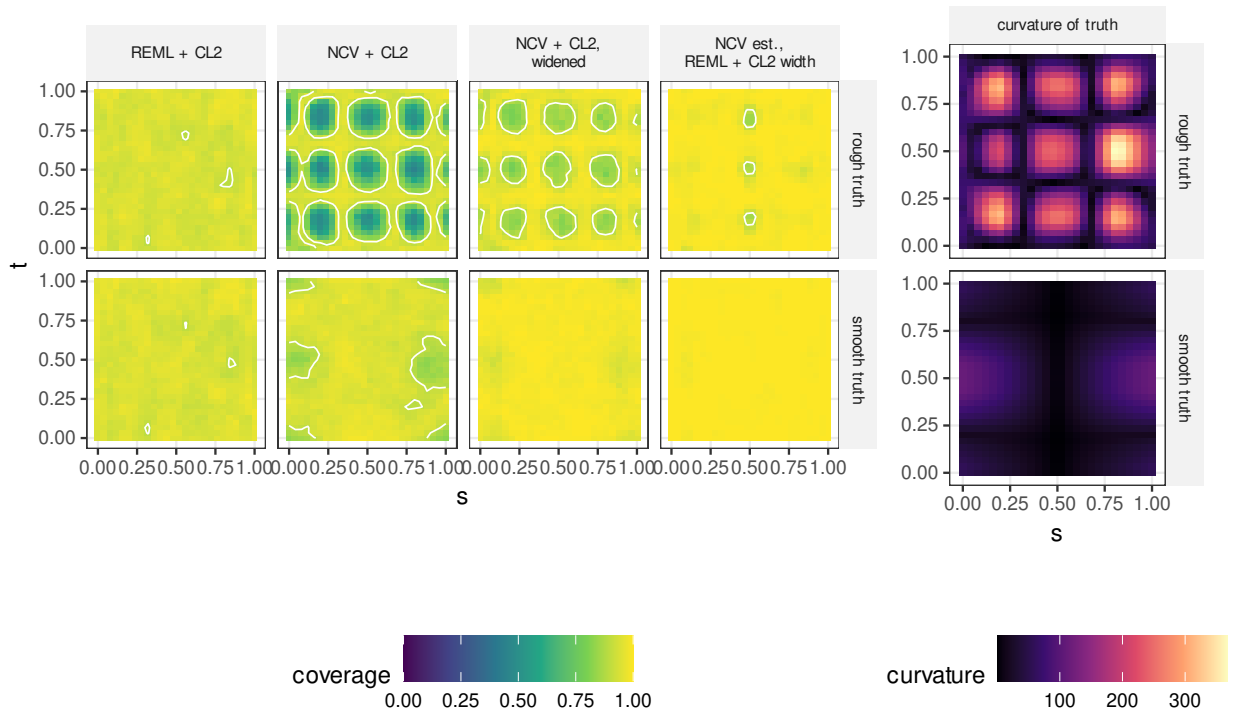


Figure 10: Pointwise coverage of $\beta(s, t)$ for REML + CL2 and three intervals centred at the NCV estimate, Gaussian, default signal, $G = 100$, smooth errors, rough truth (top) and smooth truth (bottom); white contour at 0.90. Widened: NCV + CL2 widened by the NCV–REML difference; NCV est.: NCV estimate with REML + CL2's half-width. Right: curvature of the truth.

(d) One-step bias correction

Method. For a penalized fit at fixed λ and Gaussian responses, $E\hat{\theta} = A\theta$ with $A = V_e V_p^{-1} = V_p X^T W X$. Directions in which NCV shrinks the estimate almost completely ($A \approx 0$) carry no information from the data, so no correction of this kind can restore them. The corrected estimate $\tilde{\theta} = (2I - A)\hat{\theta}$ (“twicing”, Tukey (1977)) has bias $-(I - A)^2\theta$ instead of $-(I - A)\theta$. Its covariance propagates the frequentist CL2 covariance V_f through the same map and keeps the Bayesian smoothing-bias term, $MV_f M^T + (I - A)(V_p - V_e)(I - A)^T$ with $M = 2I - A$, as in robust bias correction (Calonico, Cattaneo, and Farrell 2018). It needs no refit.

Derivation: mean and bias. For Gaussian responses with identity link and fixed λ , $\hat{\theta} = F^{-1}X^T W y$, so $E\hat{\theta} = F^{-1}X^T W X\theta = A\theta$ with

$$A = V_p X^T W X = (V_p X^T W X V_p) V_p^{-1} = V_e V_p^{-1}, \quad I - A = V_p (F - X^T W X) = V_p S,$$

and the bias is $-(I - A)\theta = -V_p S\theta$ (Appendix A.3 of the paper). Twicing applies the estimated bias map to the estimate, $\tilde{\theta} = \hat{\theta} + (I - A)\hat{\theta} = (2I - A)\hat{\theta}$, with $E\tilde{\theta} = (2I - A)A\theta = \{I - (I - A)^2\}\theta$. Repeating the step m times gives $M_m = \sum_{j=0}^m (I - A)^j$ and $E M_m \hat{\theta} = \{I - (I - A)^{m+1}\}\theta$; two steps give $M_2 = 3I - 3A + A^2$ with bias $-(I - A)^3\theta$. If $X^T W X$ is invertible, $M_m \rightarrow A^{-1}$ as $m \rightarrow \infty$, and $A^{-1}\hat{\theta}$ is the unpenalized estimate. For GLMs the same algebra holds for the working linear model at the converged weights, to first order; A is computed as $V_e V_p^{-1}$ from the fit.

Derivation: covariance. At fixed λ , M is fixed, so $\text{Var}(\tilde{\theta}) = M \text{Var}(\hat{\theta}) M^T$, estimated by $MV_f M^T$ with the frequentist CL2. The smoothing-bias allowance follows from the Bayesian reading of Appendix A.3: under $\theta \sim N(0, S^-)$ the bias $-(I - A)\theta$ of $\hat{\theta}$ has covariance $V_p S S^- S V_p = V_p S V_p = V_p - V_e$, which is CL2’s allowance. The bias of $\tilde{\theta}$ is $-(I - A)\{(I - A)\theta\}$, with prior covariance $(I - A)(V_p - V_e)(I - A)^T$; after two steps it is $(I - A)^2(V_p - V_e)\{(I - A)^2\}^T$. The interval uses

$$MV_f M^T + (I - A)(V_p - V_e)(I - A)^T,$$

not $M\hat{V}_{\text{CL2}} M^T$, which propagates the allowance through the correction as if it were sampling variance.

Derivation: directions. Let $F^{-1/2}X^T W X F^{-1/2} = Q\Lambda Q^T$ with $\Lambda = \text{diag}(a_j)$, $0 \leq a_j \leq 1$ ($a_j = 1$ in the null space of the penalty), and $B = F^{-1/2}Q$. Then $AB = B\Lambda$, and

$$V_e = B\Lambda B^T, \quad V_p - V_e = B(I - \Lambda)B^T, \quad (I - A)(V_p - V_e)(I - A)^T = B(I - \Lambda)^3 B^T, \quad M(V_p - V_e)M^T = B(2I - \Lambda)^2(I - \Lambda)B^T.$$

In the coordinates $\vartheta = B^{-1}\theta$, component j of the estimate has mean $a_j\vartheta_j$, working-model variance a_j and bias $-(1 - a_j)\vartheta_j$; after twicing the mean is $(2 - a_j)a_j\vartheta_j$, the variance $(2 - a_j)^2 a_j$ and the bias $-(1 - a_j)^2 \vartheta_j$. The correctly propagated allowance is $(1 - a_j)^3$, smaller than the original $1 - a_j$; propagating the Bayesian CL2 through M gives $(2 - a_j)^2(1 - a_j)$, up to four times the original. In directions with $a_j \approx 0$, which NCV shrinks almost completely, the estimate carries almost no information about ϑ_j : any linear correction $p(A)\hat{\theta}$ has mean $p(a_j)a_j\vartheta_j$ and variance $p(a_j)^2 a_j$, so recovering a fraction κ of ϑ_j needs $p(a_j) = \kappa/a_j$ and a variance κ^2/a_j that diverges as $a_j \rightarrow 0$. For M_m , $p(a) = \{1 - (1 - a)^{m+1}\}/a \leq m + 1$, and the bias $-(1 - a_j)^{m+1}\vartheta_j$ stays close to $-\vartheta_j$.

Result. On 200 fresh synthetic replicates it covers β at 0.932 (smooth truth, Gaussian and Poisson) and 0.899 (rough truth; NCV + CL2 0.849, REML + CL2 0.940), and binary β at 0.873. On the plasmode datasets it adds 1.1–7.0 pp to NCV + CL2 (table in (f)).

Reason. One step removes only part of the bias, none of it in the directions NCV shrinks completely, and leaves the smoothing-parameter variability of (a) untouched.

Table 50: One-step bias correction, $\beta(s, t)$, 200 fresh replicates per setting: coverage, 5% quantile of pointwise coverage, width relative to REML + CL2 (geometric mean), detection rate and median relative error of the centre.

group	interval	settings	cov	q05	width rel.	detection	rel. error
binary	NCV + CL2	6	0.848	0.659	0.358	0.598	0.498
binary	bias correction	6	0.873	0.763	0.435	0.587	0.547
binary	REML + CL2	6	0.941	0.914	1.000	0.267	1.199
independent errors	NCV + CL2	13	0.959	0.877	0.913	0.914	0.155
independent errors	bias correction	13	0.951	0.909	0.975	0.916	0.166
independent errors	REML + CL2	13	0.968	0.907	1.000	0.913	0.165
rough truth, Gaussian/Poisson	NCV + CL2	5	0.849	0.609	0.509	0.922	0.294
rough truth, Gaussian/Poisson	bias correction	5	0.899	0.784	0.617	0.902	0.306
rough truth, Gaussian/Poisson	REML + CL2	5	0.940	0.914	1.000	0.746	0.416
smooth truth, Gaussian/Poisson	NCV + CL2	8	0.930	0.872	0.380	0.933	0.193
smooth truth, Gaussian/Poisson	bias correction	8	0.932	0.901	0.467	0.910	0.228
smooth truth, Gaussian/Poisson	REML + CL2	8	0.937	0.911	1.000	0.656	0.509

(e) Smoothing-parameter uncertainty term

Method. The NCV criterion with curve blocks has analytic derivatives with respect to a weight w_g on each curve. With the curve scores $s_g = \partial^2 \text{NCV} / \partial \rho \partial w_g$ ($\rho = \log \lambda$) and the Hessian $\mathbf{H}$ of the criterion in ρ , the infinitesimal jackknife over curves gives the influence $\mathbf{d}_g = -\mathbf{H}^{-1}(\mathbf{s}_g - \bar{\mathbf{s}})$ of curve g on $\hat{\rho}$ and the curve-robust variance $\hat{\mathbf{V}}_\rho = \sum_g \mathbf{d}_g \mathbf{d}_g^\top$. It enters the coefficient covariance as $\mathbf{J} \hat{\mathbf{V}}_\rho \mathbf{J}^\top$ with $\mathbf{J} = \partial \hat{\theta} / \partial \rho$ (“ λ term”), optionally with the first-order cross term between the coefficient and the smoothing-parameter influences.

Derivation: the first-order term. Let $\hat{\theta}(\rho)$ be the penalized fit at fixed ρ and $\hat{\rho}$ the selected value with limit ρ_0 . Implicit differentiation of the estimating equation $\sum_g \mathbf{u}_g(\hat{\theta}) = \hat{\mathbf{S}}\hat{\theta}$ (Appendix A.1 of the paper) in ρ_k gives $-\mathbf{F} \partial \hat{\theta} / \partial \rho_k - \dot{\mathbf{S}}_k \hat{\theta} = \mathbf{0}$, so the columns of $\mathbf{J} = \partial \hat{\theta} / \partial \rho^\top$ are $\mathbf{J}_k = -\mathbf{V}_p \dot{\mathbf{S}}_k \hat{\theta}$ (mgcv’s `db.drho`). To first order

$$\hat{\theta}(\hat{\rho}) - \hat{\theta}(\rho_0) \approx \mathbf{J}(\hat{\rho} - \rho_0), \quad \text{Var} \hat{\theta}(\hat{\rho}) \approx \text{Var} \hat{\theta}(\rho_0) + \mathbf{J} \text{Var}(\hat{\rho}) \mathbf{J}^\top + \mathbf{C} + \mathbf{C}^\top,$$

with the cross covariance $\mathbf{C} = \text{Cov}\{\hat{\theta}(\rho_0), \hat{\rho}\} \mathbf{J}^\top$. The terms left out (the dependence of $\mathbf{V}_p$ and $\mathbf{J}$ on ρ , the curvature of $\hat{\theta}(\rho)$) are of order G^{-1} relative to the first-order term when $\hat{\rho}$ is $\sqrt{G}$ -regular. For the surface penalty under dependence it is not: the sampling distribution of $\log \hat{\lambda}$ is flat or bimodal at $G = 40$, and its SD shrinks by a factor of only 0.84 from $G = 40$ to $G = 100$ (REML; $\sqrt{40/100} = 0.63$).

Derivation: mgcv’s $\mathbf{V}_c$. Wood, Pya, and Säfken (2016) argue in the Bayesian model: given $\rho, \theta \mid \mathbf{y}, \rho \sim N\{\hat{\theta}(\rho), \mathbf{V}_p(\rho)\}$, and with a flat prior the Laplace approximation gives $\rho \mid \mathbf{y} \approx N(\hat{\rho}, \mathbf{H}^{-1})$, with $\mathbf{H}$ the Hessian of the negative log restricted likelihood in the outer parameters (block $\rho\rho$ of its inverse, if the scale is estimated). The law of total variance and the linearization $\hat{\theta}(\rho) \approx \hat{\theta}(\hat{\rho}) + \mathbf{J}(\rho - \hat{\rho})$ give $\mathbf{V}_c = \mathbf{V}_p + \mathbf{J} \mathbf{H}^{-1} \mathbf{J}^\top$, to which mgcv adds a second-order term. Two working assumptions are built in. First, $\mathbf{H}^{-1}$ is the variance of $\hat{\rho}$ only if the information identity holds, i.e. if the variance of the gradient of the criterion equals its expected Hessian, which is the case when the N rows are independent with variance $\phi V(\mu)$. For Gaussian responses the data enter the ρ_k -gradient through $\frac{1}{2} \hat{\theta}^\top \dot{\mathbf{S}}_k \hat{\theta} = \frac{1}{2} \mathbf{y}^\top \mathbf{Q}_k \mathbf{y}$ with $\mathbf{Q}_k = \mathbf{W} \mathbf{X} \mathbf{V}_p \dot{\mathbf{S}}_k \mathbf{V}_p \mathbf{X}^\top \mathbf{W}$; for $\mathbf{E} \mathbf{y} = \mathbf{m}$ and $\text{Var} \mathbf{y} = \Sigma$ its variance is $\frac{1}{2} \text{tr}(\mathbf{Q}_k \Sigma \mathbf{Q}_k \Sigma) + \mathbf{m}^\top \mathbf{Q}_k \Sigma \mathbf{Q}_k \mathbf{m}$, whereas $\mathbf{H}$ corresponds to $\Sigma = \phi \mathbf{I}$. In the dependent main-factorial settings mgcv’s $\mathbf{H}^{-1}$ gives standard deviations of the surface penalty’s $\log \hat{\lambda}$ of 0.26–0.52 times its Monte Carlo SD (median over replicates; REML). Second, the law of total variance has no cross term: in the Bayesian model θ given ρ is the posterior. Over repeated samples with a fixed truth, $\hat{\theta}$ and the gradient are correlated: for symmetric errors $\text{Cov}(\mathbf{y}^\top \mathbf{Q}_k \mathbf{y}, \mathbf{a}^\top \mathbf{y}) = 2 \mathbf{m}^\top \mathbf{Q}_k \Sigma \mathbf{a}$, so $\text{Cov}(\hat{\theta}^\top \dot{\mathbf{S}}_k \hat{\theta}, \hat{\theta}) = 2 \mathbf{V}_p \mathbf{X}^\top \mathbf{W} \Sigma \mathbf{W} \mathbf{X} \mathbf{V}_p \dot{\mathbf{S}}_k \mathbf{E} \hat{\theta}$, which vanishes only if $\dot{\mathbf{S}}_k \mathbf{E} \hat{\theta} = \mathbf{0}$.

Derivation: curve contributions. Give curve g a weight c_g in the selection criterion $\mathcal{V}(\omega; \mathbf{c})$ (in its log-likelihood for REML; in its loss term and in every fit for NCV), so that $\mathbf{c} = \mathbf{1}$ is the data and $c_g = 0$ deletes the curve; ω

collects the outer parameters (ρ , and $\tau = \log \phi$ for a Gaussian REML fit). The selected value solves $\nabla_{\omega} \mathcal{V}(\hat{\omega}; \mathbf{c}) = \mathbf{0}$, and the implicit-function theorem gives

$$\left. \frac{d\hat{\omega}}{dc_g} \right|_{c=1} = -\mathbf{H}_{\omega}^{-1} \mathbf{s}_g, \quad \mathbf{s}_g = \left. \frac{\partial \nabla_{\omega} \mathcal{V}(\hat{\omega}; \mathbf{c})}{\partial c_g} \right|_{c=1}, \quad \mathbf{H}_{\omega} = \nabla_{\omega}^2 \mathcal{V},$$

where $\mathbf{s}_g$ is a total derivative that includes the dependence of $\hat{\theta}$, $\mathbf{W}$ and the leave-out fits on $\mathbf{c}$. Neither criterion is a sum of curve terms (log determinants; leave-out fits that use all other curves), so the decomposition $\hat{\omega} - \omega_0 \approx \sum_g \mathbf{d}_g$ holds only to first order: it is the influence-function expansion of $\hat{\omega}$ in the curves. The penalty does not scale with $\mathbf{c}$, so $\sum_g \mathbf{s}_g \neq \mathbf{0}$. Perturbing the empirical distribution of the curves at fixed total weight, $\mathbf{c} = \mathbf{1} + \varepsilon(\mathbf{e}_g - \mathbf{1}/G)$, gives the curve influence $\mathbf{d}_g = -\mathbf{H}_{\omega}^{-1}(\mathbf{s}_g - \bar{\mathbf{s}})$.

Derivation: estimators of $\text{Var}(\hat{\omega})$. (i) The infinitesimal jackknife (IJ),

$$\hat{\mathbf{V}}_{\omega} = \frac{G}{G-1} \sum_g \mathbf{d}_g \mathbf{d}_g^{\top} = \mathbf{H}_{\omega}^{-1} \hat{\mathbf{M}} \mathbf{H}_{\omega}^{-1}, \quad \hat{\mathbf{M}} = \frac{G}{G-1} \sum_g (\mathbf{s}_g - \bar{\mathbf{s}})(\mathbf{s}_g - \bar{\mathbf{s}})^{\top},$$

a sandwich for $\hat{\omega}$ with the curves as clusters. It differs from mgcv's $\mathbf{H}^{-1}$ in the meat and agrees with it when $\hat{\mathbf{M}} = \mathbf{H}_{\omega}$. (ii) The one-step leave-curve-out (OS) estimator: at $\hat{\rho}$, $\hat{\theta}_{(-g)} = \hat{\theta} - (\mathbf{F} - \tilde{\mathbf{X}}_g^{\top} \tilde{\mathbf{X}}_g)^{-1} \mathbf{u}_g$ (exact for Gaussian responses; for GLMs the start of penalized IRLS at fixed $\hat{\rho}$, iterated to convergence), then one Newton step in the outer parameters, $\Delta_g = -\mathbf{H}_{\omega}^{-1} \{\nabla \mathcal{V}_{(-g)}(\hat{\omega}) - \nabla \mathcal{V}(\hat{\omega})\}$, with the leave-out gradient evaluated exactly (re-evaluated weights, $\mathbf{F}$ and number of observations), and $\hat{\mathbf{V}}_{\omega} = \frac{G-1}{G} \sum_g (\Delta_g - \bar{\Delta})(\Delta_g - \bar{\Delta})^{\top}$. Since $\nabla \mathcal{V}_{(-g)} - \nabla \mathcal{V} \approx -\mathbf{s}_g$, $\Delta_g - \bar{\Delta} \approx -\mathbf{d}_g$: IJ and OS agree to first order, and OS carries the leave-out inflation of the jackknife, as CR3 does relative to CR0. (iii) The exact curve jackknife, the reference: $\Delta_g = \hat{\omega}_{(-g)} - \hat{\omega}$ from a refit without curve g (same basis, constraints and penalties, warm start), with the same formula; it needs G refits. (iv) Regularized variants replace $\mathbf{H}_{\omega}^{-1}$ by $(\mathbf{H}_{\omega,+} + \mathbf{I}/10)^{-1}$, where $\mathbf{H}_{\omega,+}$ has its non-positive eigenvalues set to zero and $1/10$ is the prior precision of mgcv's second-order term. When a smoothing parameter drifts towards a boundary, $\mathbf{H}_{\omega}$ has a small eigenvalue along it and $\mathbf{H}_{\omega}^{-1} \mathbf{s}_g$ explodes while $\mathbf{J}$ is nearly zero there; the product is numerically uncontrolled and produced heavy upper tails of $\hat{\mathbf{V}}_{\rho}$ over replicates.

Derivation: adding the term to CL2. The influence of curve g on $\hat{\theta}(\hat{\rho})$ is $\psi_g = \mathbf{k}_g + \mathbf{J} \mathbf{d}_{g,\rho}$, where $\mathbf{d}_{g,\rho}$ is the ρ -block of $\mathbf{d}_g$ ($\hat{\theta}$ does not depend on τ). CL2 replaces the coefficient part by its leverage-adjusted version $\mathbf{k}_g$; the ρ -part is not adjusted. Without the cross term, $\hat{\mathbf{V}} = \hat{\mathbf{V}}_{\text{CL2}} + \mathbf{J} \hat{\mathbf{V}}_{\rho} \mathbf{J}^{\top}$ with $\hat{\mathbf{V}}_{\rho}$ the $\rho\rho$ -block of $\hat{\mathbf{V}}_{\omega}$. With the cross term, the sandwich of the stacked influences,

$$\hat{\mathbf{V}}_{\times} = \frac{G}{G-1} \sum_g (\mathbf{k}_g + \mathbf{J} \mathbf{d}_{g,\rho})(\mathbf{k}_g + \mathbf{J} \mathbf{d}_{g,\rho})^{\top} + (\mathbf{V}_{\rho} - \mathbf{V}_e) = \hat{\mathbf{V}}_{\text{CL2}} + \mathbf{J} \hat{\mathbf{V}}_{\rho} \mathbf{J}^{\top} + \frac{G}{G-1} \sum_g (\mathbf{k}_g \mathbf{d}_{g,\rho}^{\top} \mathbf{J}^{\top} + \mathbf{J} \mathbf{d}_{g,\rho} \mathbf{k}_g^{\top}),$$

the plug-in for $\mathbf{C} + \mathbf{C}^{\top}$ above. For leave-out changes $\mathbf{d}_{g,\rho}$ is replaced by $-(\Delta_{g,\rho} - \bar{\Delta}_{\rho})$ rescaled by $(G-1)/G$, so that the $\rho\rho$ -part equals the jackknife variance in both forms. The result is a CL2-based hybrid, not a sandwich derived from one estimating equation. Assumptions: independent curves, a correct mean model, a canonical link with unit prior weights and no offsets, smoothing parameters that are neither linked nor fixed, a converged fit at an interior optimum with positive definite $\mathbf{F}$, and, for the first-order term to be accurate, $\sqrt{G}$ -regular $\hat{\rho}$.

Derivation: curve-blocked NCV, Gaussian responses. With curve weights, $\mathbf{F}_{\mathbf{c}} = \sum_h c_h \tilde{\mathbf{X}}_h^{\top} \tilde{\mathbf{X}}_h + \mathbf{S}$, the fit without curve g is

$$\mathbf{t}_g = \mathbf{F}_{(-g)}^{-1} \sum_{h \neq g} c_h \mathbf{X}_h^{\top} \mathbf{W}_h \mathbf{y}_h, \quad \mathbf{F}_{(-g)} = \mathbf{F}_{\mathbf{c}} - c_g \tilde{\mathbf{X}}_g^{\top} \tilde{\mathbf{X}}_g,$$

and the curve-blocked NCV criterion of §2 of the paper, divided by ϕ , is $\mathcal{V}(\rho; \mathbf{c}) = \sum_g c_g \mathbf{r}_g^{\top} \mathbf{W}_g \mathbf{r}_g$ with $\mathbf{r}_g = \mathbf{y}_g - \mathbf{X}_g \mathbf{t}_g$. For Gaussian responses mgcv's Newton-step leave-out fit is exact, so this is mgcv's criterion; the factor $1/\phi$ cancels in $\mathbf{H}_{\rho}^{-1} \mathbf{s}_g$, and there is no scale parameter among the outer parameters. A zero weight removes

curve g from the loss and from every other leave-out fit. With $\mathbf{v}_{gk} = \mathbf{F}_{(-g)}^{-1} \dot{\mathbf{S}}_k \mathbf{t}_g$, $\partial \mathbf{t}_g / \partial \rho_k = -\mathbf{v}_{gk}$, and with $\mathbf{u}_g^{\text{cv}} = \mathbf{F}_{(-g)}^{-1} \mathbf{X}_g^\top \mathbf{W}_g \mathbf{r}_g$ and $q_{gk} = \mathbf{u}_g^{\text{cv}\top} \dot{\mathbf{S}}_k \mathbf{t}_g$,

$$\frac{\partial \mathcal{V}}{\partial \rho_k} = 2 \sum_g c_g q_{gk}, \quad H_{\rho,kl} = 2 \sum_g c_g [(\mathbf{X}_g \mathbf{v}_{gl})^\top \mathbf{W}_g \mathbf{X}_g \mathbf{v}_{gk} + \mathbb{1}_{k=l} q_{gk} - \mathbf{u}_g^{\text{cv}\top} \dot{\mathbf{S}}_l \mathbf{v}_{gk} - \mathbf{u}_g^{\text{cv}\top} \dot{\mathbf{S}}_k \mathbf{v}_{gl}],$$

using $\partial \mathbf{u}_g^{\text{cv}} / \partial \rho_l = \mathbf{F}_{(-g)}^{-1} (\mathbf{X}_g^\top \mathbf{W}_g \mathbf{X}_g \mathbf{v}_{gl} - \dot{\mathbf{S}}_l \mathbf{u}_g^{\text{cv}})$. For the curve scores, $\mathbf{t}_h$ does not depend on c_h , and for $g \neq h$ $\partial \mathbf{F}_{(-g)} / \partial c_h = \tilde{\mathbf{X}}_h^\top \tilde{\mathbf{X}}_h$ and $\partial \mathbf{t}_g / \partial c_h = \mathbf{F}_{(-g)}^{-1} \mathbf{X}_h^\top \mathbf{W}_h \mathbf{e}_{h|g}$, with $\mathbf{e}_{h|g} = \mathbf{y}_h - \mathbf{X}_h \mathbf{t}_g$ the residual of curve h under the fit without g . Differentiating q_{gk} gives, with $\mathbf{m}_{1,gk} = \mathbf{F}_{(-g)}^{-1} \mathbf{X}_g^\top \mathbf{W}_g \mathbf{X}_g \mathbf{v}_{gk}$ and $\mathbf{m}_{2,gk} = \mathbf{F}_{(-g)}^{-1} \dot{\mathbf{S}}_k \mathbf{u}_g^{\text{cv}}$,

$$s_{hk} = \left. \frac{\partial^2 \mathcal{V}}{\partial \rho_k \partial c_h} \right|_{\mathbf{c}=\mathbf{1}} = 2q_{hk} + 2 \sum_{g \neq h} [\mathbf{e}_{h|g}^\top \mathbf{W}_h \mathbf{X}_h (\mathbf{m}_{2,gk} - \mathbf{m}_{1,gk}) - (\mathbf{X}_h \mathbf{u}_g^{\text{cv}})^\top \mathbf{W}_h \mathbf{X}_h \mathbf{v}_{gk}].$$

The first term is curve h 's own loss, the sum its effect on every other leave-out fit. The cost is $O(GKNp + Gp^3)$. Since $\mathcal{V}(\rho; \mathbf{c}) = c \mathcal{V}(\rho - \log c; \mathbf{1})$, $\sum_h s_h = \nabla_\rho \mathcal{V} - \mathbf{H}_\rho \mathbf{1}$, a check of signs and centring that the implementation meets to 10^{-12} ; scores and Hessian also agree with central differences. NCV is a prediction loss, not a log-likelihood, so $\mathbf{H}_\rho^{-1}$ alone is not a covariance of $\hat{\rho}$; the IJ sandwich $\mathbf{H}_\rho^{-1} \hat{\mathbf{M}} \mathbf{H}_\rho^{-1}$ is, to first order, and it is invariant to rescaling the criterion. The regularized variants are not (the prior $1/10$ does not rescale), so they serve only as a sensitivity check. J is the same as for REML, because the penalized fit at given ρ is the same.

Derivation: curve-blocked NCV, Poisson and binary responses. With a canonical link and $\phi = 1$, $\mathbf{W} = \text{diag}(d\mu/d\eta)$ depends on $\hat{\theta}$, and mgcv approximates the fit without curve g by one Newton step from $\hat{\theta}$: $\mathbf{F}_{(-g)} = \mathbf{F} - c_g \mathbf{X}_g^\top \mathbf{W}_g \mathbf{X}_g$, $\delta_g = \mathbf{F}_{(-g)}^{-1} \mathbf{X}_g^\top c_g (\mathbf{y}_g - \mu_g)$, $\eta_{(-g)} = \mathbf{X}_g (\hat{\theta} - \delta_g)$, and $\mathcal{V} = \sum_g c_g \sum_d \text{dev}\{y_{gd}, \mu(\eta_{(-g),d})\}$. The derivatives in a direction $(\dot{\rho}, \dot{\mathbf{c}})$ follow by implicit differentiation of the score equation $(d\{c(\mathbf{y} - \mu)\})/d\eta = -c\mathbf{w}$ for a canonical link); with $\dot{\mathbf{S}} = \sum_k \dot{\rho}_k \dot{\mathbf{S}}_k$, $\mathbf{w}' = d\mathbf{w}/d\eta$ and $\circ$ the elementwise product,

$$\dot{\theta} = \mathbf{F}^{-1} \{\mathbf{X}^\top (\dot{\mathbf{c}} \circ (\mathbf{y} - \mu)) - \dot{\mathbf{S}} \hat{\theta}\}, \quad \dot{\eta} = \mathbf{X} \dot{\theta}, \quad (\mathbf{c}\mathbf{w})' = \dot{\mathbf{c}} \circ \mathbf{w} + \mathbf{c} \circ \mathbf{w}' \circ \dot{\eta}, \quad \{\mathbf{c}(\mathbf{y} - \mu)\}' = \dot{\mathbf{c}} \circ (\mathbf{y} - \mu) - \mathbf{c} \circ \mathbf{w} \circ \dot{\eta},$$

$$\dot{\mathbf{F}} = \mathbf{X}^\top \text{diag}\{(\mathbf{c}\mathbf{w})'\} \mathbf{X} + \dot{\mathbf{S}}, \quad \dot{\delta}_g = \mathbf{F}_{(-g)}^{-1} [\mathbf{X}_g^\top \{\mathbf{c}(\mathbf{y} - \mu)\}'_g - \{\dot{\mathbf{F}} - \mathbf{X}_g^\top \text{diag}\{(\mathbf{c}\mathbf{w})'\}_g \mathbf{X}_g\} \delta_g], \quad \dot{\eta}_{(-g)} = \mathbf{X}_g (\dot{\theta} - \dot{\delta}_g),$$

and, since $\partial \text{dev} / \partial \eta = -2(\mathbf{y} - \mu)$, $\dot{\mathcal{V}} = \sum_g [\dot{c}_g \sum_d \text{dev}_{gd} - 2c_g \sum_d (y_{gd} - \mu_{(-g),d}) \dot{\eta}_{(-g),d}]$. The K directions in ρ give the gradient (equal to mgcv's), the G directions in $\mathbf{c}$ give $\partial \mathcal{V} / \partial c_h$; $\mathbf{H}_\rho$ and the scores are central differences in ρ (step 10^{-3}) of these exact first derivatives. Compared with the Gaussian case, the weights, the Newton step and the leave-out predictions now move with ρ and $\mathbf{c}$ through $\hat{\eta}$, there is no scale parameter, and $\mathbf{J}_k = -\mathbf{V}_p \dot{\mathbf{S}}_k \hat{\theta}$ at the converged weights. A smoothing parameter whose curvature is below 10^{-3} times the largest diagonal entry of $\mathbf{H}_\rho$ lies on a flat stretch of the criterion ($\lambda_k \rightarrow \infty$, the fit no longer depends on it); as in mgcv's Newton iteration it is treated as fixed ($d_{g,k} = 0$) and the IJ uses the sub-matrices of the free components. This rule cannot represent jumps of $\hat{\rho}$ between interior and boundary solutions.

Result. In the main factorial (200 replicates, dependent errors) the term without the cross term raises NCV + CL2's coverage of β from NA to NA for Gaussian responses, at NA times the width; for binary responses with 40 curves it reaches . On the plasmode datasets with independent source curves it raises the mean coverage of β from 0.816 to 0.863 (range over combinations 0.58–0.94). Without the cross term the variance is closer to the replicate variance (SE/SD 0.86–1.29 per plasmode dataset, against 0.99–1.38 with it) and coverage is lower by 0.5–2.6 pp per dataset (next table in (f)).

Reason. The term addresses only the variance part of (a), not the smoothing bias. For binary responses the selected smoothing parameters jump between interior and boundary solutions, which a first-order term cannot represent.

Table 51: Smoothing-parameter uncertainty term for curve-blocked NCV, $\beta(s, t)$, dependent main-factorial settings at the default signal (200 replicates): coverage (5% quantile of pointwise coverage) and width relative to NCV + CL2. λ fixed: refit at the geometric-mean NCV smoothing parameters of the setting, computed from the same replicates (a diagnostic).

family	G	NCV + CL2	+ λ term	λ fixed (diagnostic)
binary	40	0.787 (0.61) 1.00	0.859 (0.72) 1.27	0.704 (0.23) 0.70
binary	100	0.899 (0.78) 1.00	0.933 (0.85) 1.16	0.949 (0.88) 0.96
Gaussian	40	0.923 (0.86) 1.00	0.948 (0.90) 1.13	0.949 (0.91) 1.00
Gaussian	100	0.934 (0.86) 1.00	0.951 (0.89) 1.10	0.950 (0.90) 1.00
Poisson	40	0.924 (0.88) 1.00	0.951 (0.91) 1.21	0.952 (0.92) 1.01
Poisson	100	0.935 (0.88) 1.00	0.953 (0.90) 1.09	0.953 (0.91) 1.00

(f) Bias correction and smoothing-parameter term combined

Method. Propagate the smoothing-parameter term of (e) through the bias-correction map of (d) (Gaussian responses only), with or without the cross term. The map $\mathbf{M} = 2\mathbf{I} - \mathbf{A}$ is held fixed at $\hat{\rho}$: the construction is a *frozen-map approximation* that ignores $\partial\mathbf{A}/\partial\rho$ (derivation below).

Derivation. The corrected estimate is $\tilde{\theta}(\hat{\rho}) = \mathbf{M}(\hat{\rho})\hat{\theta}(\hat{\rho})$. For Gaussian responses $\partial\mathbf{V}_p/\partial\rho_k = -\mathbf{V}_p\dot{\mathbf{S}}_k\mathbf{V}_p$, so $\partial\mathbf{A}/\partial\rho_k = \partial(\mathbf{I} - \mathbf{V}_p\mathbf{S})/\partial\rho_k = -\mathbf{V}_p\dot{\mathbf{S}}_k\mathbf{A}$ and

$$\frac{\partial\tilde{\theta}}{\partial\rho_k} = \mathbf{M}\mathbf{J}_k - \frac{\partial\mathbf{A}}{\partial\rho_k}\hat{\theta} = \mathbf{M}\mathbf{J}_k + \mathbf{V}_p\dot{\mathbf{S}}_k\mathbf{A}\hat{\theta}.$$

The second term, $-(\partial\mathbf{A}/\partial\rho_k)\hat{\theta}$, is first order in the same smoothing-parameter perturbation as the first. The frozen-map construction keeps the first term only: the curve influences on $\tilde{\theta}$ are taken as $\mathbf{M}(\mathbf{k}_g + \mathbf{J}\mathbf{d}_{g,\rho})$, so the coefficient-smoothing-parameter cross term also uses $\mathbf{J} = \partial\hat{\theta}/\partial\rho$ and not $\partial\tilde{\theta}/\partial\rho$. The covariance is $\mathbf{M}\mathbf{V}_{f,\rho}\mathbf{M}^\top + (\mathbf{I} - \mathbf{A})(\mathbf{V}_p - \mathbf{V}_e)(\mathbf{I} - \mathbf{A})^\top$, where $\mathbf{V}_{f,\rho}$ is the frequentist CL2 plus the term of (e), with or without the cross term, and the allowance is propagated as in (d). With two steps, $\mathbf{M}_2$ and $(\mathbf{I} - \mathbf{A})^2$ replace $\mathbf{M}$ and $\mathbf{I} - \mathbf{A}$.

Result. On 100 fresh synthetic replicates per setting the version with the cross term covers β at 0.971 on smooth truths and 0.941 on rough truths, at 0.59 and 0.76 times REML + CL2's width, with SE/SD 1.21–1.27; without the cross term 0.961 and 0.929. On the plasmode data the combination was run on replicates 1–50 of each combination; REML + CL2 comes from replicates 1–100 of the bias-correction run on the same combinations. The comparison is therefore only partly paired, and small differences are descriptive, not evidence of equivalence or of a precise ranking. It covers β at 0.83–0.94 on the datasets with independent source curves (REML + CL2 0.91–0.94) at 0.73–1.09 times the width, with lower tails 0.55–0.87 (REML + CL2 0.84–0.89), and 0.66–0.84 on the datasets with dependent source curves. By truth source: 0.81 (REML-fitted), 0.88 (midpoint), 0.91 (NCV-fitted). Without the cross term it covers 0.81–0.94 on the datasets with independent source curves and stays below 0.85 on running, ocean, weather, electricity. Of the 3000 plasmode replicate fits, 8 failed the internal consistency check of the smoothing-parameter derivatives; they are reported as missing, not replaced.

Reason. On synthetic data the excess variance of the term offsets the bias the correction leaves (SE/SD above 1). On the plasmode data the remaining bias is larger, and this construction offers no width advantage over REML + CL2 while its lower tails are worse. The conclusion is about the frozen-map approximation evaluated here; whether the full derivative, including $-(\partial\mathbf{A}/\partial\rho)\hat{\theta}$, would change it was not studied.

Table 52: Synthetic, Gaussian, dependent errors, $\beta(s, t)$, 100 fresh replicates: coverage (5% quantile of pointwise coverage) and width relative to REML + CL2 for NCV + CL2, with the smoothing-parameter term, with the bias correction, and with both.

group	settings	NCV + CL2	+ λ term	bias correction	bias correction + λ term	REML + CL2
rough truth, Gaussian	3	0.846 (0.59) 0.50	0.889 (0.68) 0.59	0.894 (0.77) 0.61	0.941 (0.84) 0.76	0.940 (0.92)
smooth truth, Gaussian	4	0.925 (0.87) 0.38	0.957 (0.90) 0.46	0.928 (0.89) 0.47	0.971 (0.94) 0.59	0.938 (0.91)

Table 53: Plasmode, $\beta(s, t)$, all subjects, mean over truth $\times$ residual combinations: coverage (5% quantile of point-wise coverage) and width relative to REML + CL2 for the four NCV-centred intervals (50 replicates), and REML + CL2 (100 replicates). Datasets with independent source curves first.

dataset	NCV + CL2	+ λ term	bias correction	bias correction + λ term	REML + CL2
ECG strain	0.89 (0.76) 0.84	0.92 (0.81) 0.94	0.91 (0.82) 0.94	0.94 (0.87) 1.09	0.94 (0.89)
AF trial	0.81 (0.53) 0.53	0.89 (0.70) 0.77	0.84 (0.66) 0.63	0.94 (0.81) 1.04	0.94 (0.86)
running	0.72 (0.42) 0.52	0.78 (0.48) 0.71	0.76 (0.48) 0.59	0.83 (0.55) 0.88	0.91 (0.84)
DTI	0.84 (0.54) 0.64	0.88 (0.64) 0.76	0.88 (0.68) 0.75	0.93 (0.77) 0.95	0.93 (0.86)
gait	0.81 (0.53) 0.58	0.88 (0.66) 0.82	0.84 (0.61) 0.67	0.92 (0.75) 1.06	0.94 (0.89)
ECG 8-lead	0.76 (0.47) 0.32	0.87 (0.64) 0.53	0.80 (0.56) 0.38	0.92 (0.75) 0.73	0.94 (0.88)
ocean	0.72 (0.30) 0.34	0.78 (0.38) 0.43	0.78 (0.39) 0.41	0.84 (0.47) 0.56	0.93 (0.87)
weather	0.53 (0.30) 0.33	0.61 (0.44) 0.51	0.60 (0.43) 0.42	0.66 (0.53) 0.69	0.88 (0.82)
electricity	0.62 (0.37) 0.21	0.70 (0.46) 0.31	0.63 (0.38) 0.24	0.73 (0.50) 0.41	0.93 (0.88)

Table 54: Plasmode, $\beta(s, t)$, all subjects, mean over truth $\times$ residual combinations: coverage (5% quantile of point-wise coverage) and SE/SD for the smoothing-parameter term without and with the cross term, alone and combined with the one-step bias correction (50 replicates), and REML + CL2 (100 replicates). Datasets with independent source curves first.

dataset	+ λ term	+ λ term, cross	bias corr. + λ term	bias corr. + λ term, cross	REML + CL2
ECG strain	0.92 (0.80) 1.02	0.92 (0.81) 1.05	0.94 (0.86) 1.06	0.94 (0.87) 1.10	0.94 (0.89) 1.01
AF trial	0.88 (0.65) 1.12	0.89 (0.70) 1.25	0.92 (0.79) 1.27	0.94 (0.81) 1.44	0.94 (0.86) 0.96
running	0.77 (0.47) 0.95	0.78 (0.48) 1.06	0.81 (0.53) 1.07	0.83 (0.55) 1.20	0.91 (0.84) 0.90
DTI	0.87 (0.61) 1.09	0.88 (0.64) 1.15	0.92 (0.75) 1.16	0.93 (0.77) 1.24	0.93 (0.86) 0.99
gait	0.87 (0.64) 1.07	0.88 (0.66) 1.13	0.91 (0.74) 1.21	0.92 (0.75) 1.31	0.94 (0.89) 0.99
ECG 8-lead	0.86 (0.62) 1.29	0.87 (0.64) 1.38	0.90 (0.73) 1.52	0.92 (0.75) 1.64	0.94 (0.88) 0.94
ocean	0.76 (0.36) 1.03	0.78 (0.38) 1.10	0.83 (0.45) 1.11	0.84 (0.47) 1.20	0.93 (0.87) 0.94
weather	0.59 (0.41) 0.94	0.61 (0.44) 1.08	0.65 (0.50) 1.07	0.66 (0.53) 1.23	0.88 (0.82) 0.93
electricity	0.68 (0.44) 0.86	0.70 (0.46) 0.99	0.71 (0.47) 0.96	0.73 (0.50) 1.12	0.93 (0.88) 0.90

(g) Truths outside the spline space

Method. NCV + CL2 and the corrections of (d)–(f), including their two-step versions, on the misspecified truths of S5.1 (default basis, 100 replicates).

Result. For T2, whose missing part loads on directions the covariates barely excite, every NCV-centred interval covers β at 0.28–0.45. For T1 (a local bump) under independent errors they cover 0.85–0.90 with lower tails 0.11–0.25.

Reason. A bias that the spline space cannot represent is invisible to every estimate the corrections are built from, including the frozen-map combination of (f).

S4.2 mgcv’s neighbourhood covariance for curve-blocked NCV

Method. mgcv offers a covariance for NCV fits that accounts for the neighbourhood structure (help page ?NCV, Wood (2024)). Computed correctly for curve neighbourhoods, it was evaluated on the same NCV fits as NCV +

CL2: in three main-factorial settings (Gaussian, Poisson, binary; smooth errors, $G = 100$, default signal; replicates 1–100) and in the 42 plasmode combinations of the six datasets with independent source curves (all subjects, curve flips, every truth $\times$ residual pair; replicates 1–100).

Derivation: what mgcv computes. The description follows ?NCV, Wood (2024) and the source of mgcv 1.9-5 (functions `gam`, `outer` and `nei_cov`). mgcv specifies NCV per observation: each observation i is predicted from the fit with a set $\alpha(i)$ of observations dropped. For each i mgcv computes the change of the coefficients when i alone is left out, $\mathbf{d}_i = \hat{\theta}_{-i} - \hat{\theta}$, by one Newton step at the selected λ , multiplied by $\{(N-1)/N\}^{0.4}$. To first order $\mathbf{d}_i \approx -\mathbf{V}_p \tilde{\mathbf{x}}_i \mathbf{z}_i$ (Appendix A.1 of the paper). The neighbourhood-aware estimator is

$$\hat{\mathbf{V}}_{\text{nei}} = \frac{N}{N-\tau} \sum_{i=1}^N \tilde{\mathbf{d}}_i \left(\sum_{j \in \alpha(i)} \tilde{\mathbf{d}}_j \right)^\top \quad (\text{symmetrized}), \quad \tilde{\mathbf{d}}_i = \mathbf{d}_i \frac{r_i^{\text{cv}}}{r_i},$$

with τ the effective degrees of freedom and r_i^{cv}/r_i the ratio of the cross-validated to the ordinary deviance residual. mgcv returns $\hat{\mathbf{V}}_{\text{nei}} + a(\mathbf{V}_p - \mathbf{V}_e)$, projected onto the positive semidefinite matrices, where $a = \max\{\text{tr} \hat{\mathbf{V}}_{\text{nei}}^{(0)} / \text{tr} \mathbf{V}_e, 1\}$ and $\hat{\mathbf{V}}_{\text{nei}}^{(0)}$ is the same sum with $\mathbf{d}_i$ in place of $\tilde{\mathbf{d}}_i$. If $\alpha(i)$ is the curve of observation i , the sum equals $\frac{N}{N-\tau} \sum_g \mathbf{D}_g \mathbf{D}_g^\top$ with $\mathbf{D}_g = \sum_{i \in g} \tilde{\mathbf{d}}_i \approx -\mathbf{V}_p \tilde{\mathbf{X}}_g^\top \mathbf{z}_g$: a curve-cluster sandwich of CR0 type with the penalized bread and without CL2's leverage adjustment, plus a rescaled smoothing-bias allowance $a(\mathbf{V}_p - \mathbf{V}_e)$. With `nei$jackknife = TRUE` mgcv instead returns the jackknife over the folds, $\sum_g \{(N - D_g)/N\}^{0.8} \mathbf{d}_g \mathbf{d}_g^\top + a'(\mathbf{V}_p - \mathbf{V}_e)$, with $\mathbf{d}_g$ the one-step change when curve g is left out and $a' = \text{tr}(\cdot) / \text{tr} \mathbf{V}_e$ of the first sum.

The block specification. `refund` specifies the same NCV criterion in block form: G folds, each dropping and predicting a whole curve. Both forms select the same smoothing parameters and give the same coefficients, but mgcv 1.9-5 computes its covariance only for the per-observation form. For the block form it returns the Bayesian $\mathbf{V}_p$ when `jackknife` is unset. With `jackknife = TRUE` its routine `nei_cov()` reads the $N \times p$ matrix of the changes $\mathbf{d}_i$ with row stride G , so it uses the wrong entries: the sum has a trace of about 10^{-3} times $\text{tr} \mathbf{V}_e$ or less, $a = 1$, and the returned covariance is $\mathbf{V}_p - \mathbf{V}_e$, the smoothing-bias allowance alone. Its standard errors relative to the Bayesian ones are $\{a^\top (\mathbf{V}_p - \mathbf{V}_e) a / a^\top \mathbf{V}_p a\}^{1/2} < 1$ for every contrast a with $a^\top \mathbf{V}_e a > 0$. This explains the result of that setting in the three synthetic settings (20 replicates): standard errors 0.49–0.77 times the Bayesian ones and coverage of the mean and β of 0.32–0.58. The correction will be part of a forthcoming mgcv release.

Corrected computation. The function `jackfix_vcov()` (code archive, `ncv-jack-recheck/jackfix-lib.R`) computes from the block fit what mgcv returns for the per-observation form with $\alpha(i)$ = the curve of i : the single-observation Newton changes, the residual ratios at the leave-curve-out fit, both sums, a and a' . It needs no refit and takes about 1 s per study fit (median), about as long as CL2. In small examples it reproduces mgcv's covariance for the per-observation form to a relative difference of 10^{-14} (Gaussian) and 10^{-12} (Poisson); its leave-curve-out steps agree with exact leave-one-curve-out refits at fixed smoothing parameters to 10^{-13} for Gaussian responses and to within a few percent for Poisson responses (one Newton step). At study scale (setting 12, replicate 1), a `pffr` fit with per-observation neighbourhoods selects the same smoothing parameters (largest difference in $\log \lambda$: 0), and `jackfix_vcov()` of the block fit reproduces its $\mathbf{V}_p$ and $\mathbf{V}_c$ to a relative difference of $1.3\text{e-}13$; the per-observation fit took 801 s against 32 s for the block fit (one fit each, timed in a separate run, not comparable with the benchmark times of Appendix C of the paper). The NCV and REML estimates and CL2 standard errors of the re-evaluation reproduce the study's records (largest absolute difference $2\text{e-}06$). No replicate fit failed (0 failures).

Result. On synthetic data (next table) the corrected covariance is conservative: it covers β at 0.96–0.98 (NCV + CL2 0.90–0.94, REML + CL2 0.95) with SE/SD 1.13–1.24, at 1.30–1.32 times NCV + CL2's width and 0.46–0.51 times REML + CL2's. Its sampling part alone behaves like CL2 (SE/SD for β 0.78–0.94). The median of a is 4.1–5.1 per setting, so the bias allowance $a(\mathbf{V}_p - \mathbf{V}_e)$ adds a large term. For binary responses it covers the mean, $\alpha(t)$ and $\gamma(t)$ at 0.89–0.93, below REML + CL2 (0.93–0.94). The variant with the $\hat{\lambda}$ correction and mgcv's `jackknife = TRUE` estimate differ from it by at most 0.013 in coverage. On the plasmode data, over the 42 combinations, it covers β

at 0.87 on average (5% quantile 0.65, lowest combination 0.62), against 0.82 (0.59) for NCV + CL2 and 0.935 (0.87) for REML + CL2, at 0.68 times REML + CL2's width (geometric mean). Per dataset it raises the coverage of NCV + CL2 for β by 0.01–0.08 and stays below REML + CL2 on every dataset (by 0.004–0.164).

Reason. Computed correctly, mgcv's neighbourhood covariance is a curve-cluster sandwich plus a smoothing-bias allowance that is several times $V_p - V_e$. On synthetic data this allowance makes the intervals conservative. On the plasmode data they still undercover. Residual smoothing bias of the NCV estimate is a plausible contributor (S4.1(a)), which a covariance alone does not remove; variance and tail-calibration errors are not excluded (SE/SD 0.85 on ECG 8-lead and 0.93 on running; second table below). The failure to reach nominal coverage on these data stands without a unique cause.

Table 55: Intervals for the curve-blocked NCV fit (synthetic, smooth errors, $G = 100$, 100 replicates): coverage (5% quantile of pointwise coverage), SE/SD [width relative to NCV + CL2 / REML + CL2]. Bayesian Vp, NCV + CL2 and the corrected neighbourhood covariance use the same NCV fit; REML + CL2 the REML fit to the same data.

estimand	interval	Gaussian	Poisson	binary
E(Y X)	Bayesian Vp	0.618 (0.49) 0.46 [0.47/0.36]	0.620 (0.49) 0.47 [0.48/0.36]	0.612 (0.44) 0.49 [0.53/0.38]
E(Y X)	NCV + CL2	0.936 (0.87) 1.02 [1.00/0.78]	0.929 (0.86) 1.00 [1.00/0.77]	0.886 (0.76) 0.93 [1.00/0.71]
E(Y X)	corrected neighbourhood covariance	0.963 (0.92) 1.14 [1.12/0.87]	0.960 (0.92) 1.13 [1.14/0.87]	0.929 (0.83) 1.05 [1.14/0.81]
E(Y X)	... with λ correction	0.964 (0.92) 1.15 [1.13/0.88]	0.961 (0.92) 1.14 [1.15/0.88]	0.936 (0.84) 1.08 [1.16/0.83]
E(Y X)	its sampling part	0.936 (0.87) 1.03 [1.01/0.78]	0.930 (0.86) 1.02 [1.02/0.78]	0.874 (0.74) 0.90 [0.97/0.69]
E(Y X)	mgcv per-point jackknife = TRUE	0.963 (0.92) 1.14 [1.12/0.87]	0.960 (0.91) 1.13 [1.13/0.87]	0.933 (0.84) 1.06 [1.15/0.82]
E(Y X)	REML + CL2	0.949 (0.90) 1.02 [1.29/1.00]	0.948 (0.91) 1.02 [1.31/1.00]	0.943 (0.90) 1.00 [1.40/1.00]
$\beta(s,t)$	Bayesian Vp	0.703 (0.52) 0.50 [0.53/0.21]	0.710 (0.57) 0.51 [0.55/0.21]	0.715 (0.54) 0.53 [0.63/0.22]
$\beta(s,t)$	NCV + CL2	0.940 (0.86) 0.97 [1.00/0.39]	0.936 (0.87) 0.96 [1.00/0.38]	0.900 (0.79) 0.87 [1.00/0.35]
$\beta(s,t)$	corrected neighbourhood covariance	0.981 (0.95) 1.24 [1.30/0.51]	0.981 (0.95) 1.23 [1.30/0.49]	0.960 (0.89) 1.13 [1.32/0.46]
$\beta(s,t)$	... with λ correction	0.982 (0.95) 1.25 [1.30/0.51]	0.982 (0.95) 1.24 [1.31/0.50]	0.964 (0.90) 1.15 [1.34/0.47]
$\beta(s,t)$	its sampling part	0.925 (0.84) 0.93 [0.95/0.37]	0.921 (0.85) 0.94 [0.95/0.36]	0.858 (0.73) 0.78 [0.89/0.31]
$\beta(s,t)$	mgcv per-point jackknife = TRUE	0.983 (0.95) 1.26 [1.32/0.51]	0.983 (0.95) 1.25 [1.33/0.51]	0.965 (0.89) 1.16 [1.35/0.47]
$\beta(s,t)$	REML + CL2	0.947 (0.90) 0.97 [2.56/1.00]	0.946 (0.91) 0.96 [2.63/1.00]	0.952 (0.91) 0.96 [2.86/1.00]
$\alpha(t)$	Bayesian Vp	0.617 (0.48) 0.46 [0.46/0.44]	0.607 (0.46) 0.45 [0.48/0.43]	0.557 (0.46) 0.47 [0.51/0.41]
$\alpha(t)$	NCV + CL2	0.939 (0.85) 0.99 [1.00/0.94]	0.928 (0.85) 0.96 [1.00/0.91]	0.866 (0.69) 0.92 [1.00/0.80]
$\alpha(t)$	corrected neighbourhood covariance	0.955 (0.87) 1.07 [1.09/1.03]	0.950 (0.89) 1.05 [1.09/1.00]	0.894 (0.72) 1.02 [1.11/0.88]
$\alpha(t)$	... with λ correction	0.958 (0.89) 1.08 [1.10/1.03]	0.956 (0.90) 1.06 [1.11/1.01]	0.902 (0.74) 1.05 [1.14/0.91]
$\alpha(t)$	its sampling part	0.941 (0.86) 0.99 [1.01/0.95]	0.928 (0.85) 0.97 [1.00/0.91]	0.860 (0.69) 0.90 [0.98/0.78]
$\alpha(t)$	mgcv per-point jackknife = TRUE	0.954 (0.87) 1.07 [1.08/1.02]	0.950 (0.89) 1.05 [1.09/1.00]	0.896 (0.72) 1.03 [1.12/0.89]
$\alpha(t)$	REML + CL2	0.950 (0.91) 1.04 [1.06/1.00]	0.949 (0.92) 1.00 [1.10/1.00]	0.935 (0.89) 0.99 [1.25/1.00]
$\gamma(t)$	Bayesian Vp	0.639 (0.58) 0.48 [0.46/0.43]	0.629 (0.56) 0.49 [0.48/0.44]	0.568 (0.47) 0.47 [0.51/0.41]
$\gamma(t)$	NCV + CL2	0.936 (0.87) 1.09 [1.00/0.92]	0.932 (0.84) 1.05 [1.00/0.92]	0.849 (0.74) 0.93 [1.00/0.79]
$\gamma(t)$	corrected neighbourhood covariance	0.962 (0.94) 1.19 [1.11/1.02]	0.961 (0.94) 1.19 [1.12/1.03]	0.889 (0.79) 1.05 [1.12/0.89]
$\gamma(t)$	... with λ correction	0.963 (0.94) 1.20 [1.12/1.03]	0.962 (0.94) 1.22 [1.13/1.04]	0.902 (0.81) 1.09 [1.16/0.92]
$\gamma(t)$	its sampling part	0.939 (0.88) 1.10 [1.02/0.94]	0.936 (0.85) 1.07 [1.02/0.94]	0.840 (0.73) 0.91 [0.98/0.78]
$\gamma(t)$	mgcv per-point jackknife = TRUE	0.962 (0.94) 1.18 [1.10/1.02]	0.960 (0.93) 1.18 [1.12/1.03]	0.894 (0.80) 1.07 [1.14/0.90]
$\gamma(t)$	REML + CL2	0.954 (0.93) 1.09 [1.08/1.00]	0.951 (0.90) 1.09 [1.09/1.00]	0.940 (0.91) 1.02 [1.26/1.00]

Table 56: Plasmode, six datasets with independent source curves, all subjects, 100 replicates per combination: coverage (5% quantile of pointwise coverage) SE/SD [width relative to NCV + CL2 / REML + CL2], means over each dataset's truth $\times$ residual combinations (SE/SD: median; widths: geometric mean).

estimand	dataset	combinations	Bayesian Vp	NCV + CL2	corrected neighbourhood covariance	mgcv per-point jackknife = TRUE	REML + CL2
$\beta(s,t)$	AF trial	6	0.54 (0.21) 0.39	0.81 (0.52) 0.84	0.891 (0.63) 1.11 [1.38/0.72]	0.893 (0.64) 1.12 [1.41/0.74]	0.94 (0.86) 0.96
$\beta(s,t)$	DTI	6	0.56 (0.29) 0.46	0.85 (0.58) 0.95	0.904 (0.68) 1.12 [1.21/0.79]	0.905 (0.68) 1.12 [1.22/0.79]	0.93 (0.86) 0.99
$\beta(s,t)$	ECG strain	12	0.49 (0.20) 0.29	0.91 (0.80) 0.98	0.937 (0.84) 1.10 [1.12/0.95]	0.933 (0.83) 1.08 [1.10/0.93]	0.94 (0.89) 1.01
$\beta(s,t)$	ECG 8-lead	6	0.50 (0.18) 0.33	0.76 (0.48) 0.79	0.777 (0.49) 0.85 [1.05/0.32]	0.760 (0.48) 0.82 [1.01/0.31]	0.94 (0.88) 0.95
$\beta(s,t)$	gait	6	0.44 (0.15) 0.30	0.81 (0.53) 0.90	0.845 (0.60) 0.98 [1.10/0.63]	0.846 (0.60) 0.98 [1.10/0.63]	0.94 (0.89) 1.00
$\beta(s,t)$	running	6	0.44 (0.14) 0.32	0.71 (0.40) 0.75	0.788 (0.46) 0.93 [1.25/0.65]	0.792 (0.47) 0.94 [1.28/0.66]	0.91 (0.84) 0.90
E(Y X)	AF trial	6	0.61 (0.31) 0.41	0.90 (0.79) 0.94	0.944 (0.84) 1.10 [1.22/1.04]	0.944 (0.84) 1.10 [1.23/1.05]	0.95 (0.89) 1.01
E(Y X)	DTI	6	0.61 (0.39) 0.47	0.91 (0.79) 0.97	0.938 (0.84) 1.08 [1.12/0.95]	0.937 (0.84) 1.08 [1.12/0.95]	0.93 (0.88) 0.99
E(Y X)	ECG strain	12	0.53 (0.22) 0.34	0.92 (0.81) 1.01	0.947 (0.86) 1.12 [1.12/1.02]	0.942 (0.85) 1.09 [1.09/1.00]	0.94 (0.88) 1.03
E(Y X)	ECG 8-lead	6	0.77 (0.30) 0.78	0.90 (0.73) 0.97	0.910 (0.75) 1.01 [1.04/0.78]	0.899 (0.73) 0.98 [1.01/0.76]	0.95 (0.89) 1.04
E(Y X)	gait	6	0.53 (0.23) 0.33	0.93 (0.82) 1.01	0.943 (0.85) 1.08 [1.06/0.97]	0.941 (0.85) 1.07 [1.05/0.96]	0.95 (0.91) 1.03
E(Y X)	running	6	0.54 (0.23) 0.37	0.89 (0.75) 0.96	0.932 (0.82) 1.07 [1.13/1.03]	0.930 (0.81) 1.06 [1.12/1.02]	0.93 (0.88) 0.99

S4.3 Smoothing-parameter uncertainty term for REML

Method. The analogous term for REML uses curve weights in mgcv's REML criterion: the implicit-function theorem gives each curve's influence on $(\log \hat{\lambda}, \log \hat{\phi})$ through analytic curve scores, and the infinitesimal jackknife over curves gives a curve-robust variance that enters as $\mathbf{J}\hat{\mathbf{V}}_\rho\mathbf{J}^\top$ on top of CL2.

Derivation: criterion and gradient. With curve weights c_g on the log-likelihood contributions (including their normalizing constants), $\hat{\theta}(\rho, \mathbf{c})$ maximizes $\ell_{\mathbf{c}}(\theta) - \frac{1}{2}\theta^\top \mathbf{S}\theta$ with $\ell_{\mathbf{c}} = \sum_g c_g \ell_g$, and mgcv's Laplace-approximate REML criterion (Wood 2011) is, up to a constant that depends on neither ω nor $\mathbf{c}$,

$$\mathcal{V}(\omega; \mathbf{c}) = -\ell_{\mathbf{c}}(\hat{\theta}) + \frac{1}{2}\hat{\theta}^\top \mathbf{S}\hat{\theta} + \frac{1}{2}\log|\mathbf{X}^\top \mathbf{W}_{\mathbf{c}}\mathbf{X} + \mathbf{S}| - \frac{1}{2}\log|\mathbf{S}|_+,$$

with $\mathbf{W}_{\mathbf{c}} = \text{diag}(c_{g(i)}\mathbf{w}_i)$, M_p the dimension of the null space of the penalty, $\omega = (\rho, \tau)$ with $\tau = \log \phi$ for Gaussian responses and $\omega = \rho$, $\phi = 1$ for Poisson and binary responses. For Gaussian responses $-\ell_{\mathbf{c}} = \mathcal{D}_{\mathbf{c}}/(2\phi) + (N_{\mathbf{c}}/2)\log(2\pi\phi)$ with the weighted residual sum of squares $\mathcal{D}_{\mathbf{c}}$ and $N_{\mathbf{c}} = \sum_g c_g D_g$. Write $\mathbf{F}_{\mathbf{c}} = \mathbf{X}^\top \mathbf{W}_{\mathbf{c}}\mathbf{X} + \mathbf{S}$. Because $\hat{\theta}$ maximizes the penalized log-likelihood, its derivative does not enter the first two terms (envelope theorem), and

$$\frac{\partial \mathcal{V}}{\partial \rho_k} = \frac{1}{2}\hat{\theta}^\top \dot{\mathbf{S}}_k \hat{\theta} + \frac{1}{2}\text{tr}(\mathbf{F}_{\mathbf{c}}^{-1} \dot{\mathbf{F}}_k) - \frac{1}{2}\text{tr}(\mathbf{S}^+ \dot{\mathbf{S}}_k), \quad \dot{\mathbf{F}}_k = \dot{\mathbf{S}}_k + \mathbf{X}^\top \text{diag}(c_{g(i)}\mathbf{w}'_i \mathbf{x}_i^\top \mathbf{J}_k)\mathbf{X},$$

$$\frac{\partial \mathcal{V}}{\partial \tau} = -\frac{1}{2}\left(\frac{\mathcal{D}_{\mathbf{c}}}{\phi} + \hat{\theta}^\top \mathbf{S}\hat{\theta}\right) + \frac{N_{\mathbf{c}} - M_p}{2} \quad (\text{Gaussian}),$$

where $\mathbf{w}'_i = d\mathbf{w}_i/d\eta_i$ (zero for Gaussian responses), $\mathbf{x}_i^\top$ is row i of $\mathbf{X}$ and $\mathbf{J}_k = -\mathbf{F}_{\mathbf{c}}^{-1}\dot{\mathbf{S}}_k\hat{\theta}$. In the τ -derivative $\mathbf{S} = \sum_k (\lambda_k/\phi)\mathbf{S}_k$ moves with ϕ at fixed λ , and the two log-determinants contribute $-p/2$ and $(p - M_p)/2$. $\mathbf{H}_\omega$ is mgcv's Hessian of this criterion.

Derivation: curve scores. At $\mathbf{c} = \mathbf{1}$ the fixed- λ influence of curve g on $\hat{\theta}$ is $\delta_g = \partial\hat{\theta}/\partial c_g = \mathbf{V}_p \mathbf{u}_g$ (implicit differentiation of $\sum_h c_h \mathbf{u}_h = \mathbf{S}\hat{\theta}$), and $\partial\mathbf{F}/\partial c_g = \mathbf{X}^\top \text{diag}(\mathbf{b}_g)\mathbf{X}$ with $b_{g,i} = \mathbb{1}[i \in g] \mathbf{w}_i + \mathbf{w}'_i \mathbf{x}_i^\top \delta_g$. With $h_i = \mathbf{x}_i^\top \mathbf{V}_p \mathbf{x}_i$, $q_{ki} = \mathbf{x}_i^\top \mathbf{V}_p \dot{\mathbf{F}}_k \mathbf{V}_p \mathbf{x}_i$ and $\tilde{\mathbf{a}} = \mathbf{V}_p \mathbf{X}^\top (\mathbf{w}' \circ \mathbf{h})$, differentiating the three parts of $\partial \mathcal{V}/\partial \rho_k$ in c_g gives

$$s_{g,\rho_k} = \hat{\theta}^\top \dot{\mathbf{S}}_k \delta_g + \frac{1}{2}\left[-\sum_i b_{g,i} q_{ki} + \sum_{i \in g} \mathbf{w}'_i (\mathbf{x}_i^\top \mathbf{J}_k) h_i + \sum_i \mathbf{w}''_i (\mathbf{x}_i^\top \delta_g) (\mathbf{x}_i^\top \mathbf{J}_k) h_i - \sum_i b_{g,i} (\mathbf{x}_i^\top \tilde{\mathbf{a}}) (\mathbf{x}_i^\top \mathbf{J}_k) - \tilde{\mathbf{a}}^\top \dot{\mathbf{S}}_k \delta_g\right],$$

$$s_{g,\tau} = -\frac{\|\mathbf{y}_g - \mu_g\|^2}{2\phi} + \frac{D_g}{2} \quad (\text{Gaussian}).$$

The first term of s_{g,ρ_k} is the route through $\hat{\theta}^\top \dot{\mathbf{S}}_k \hat{\theta}$. The bracket is the route through $\log|\mathbf{F}|$: the trace term changes with $\mathbf{c}$ through $\mathbf{F}^{-1}$ (first sum), through the weights of curve g and their change with $\hat{\eta}$ (second and third sums), and through $\mathbf{J}_k$, whose derivative is $\partial\mathbf{J}_k/\partial c_g = -\mathbf{V}_p \mathbf{X}^\top \text{diag}(\mathbf{b}_g)\mathbf{X}\mathbf{J}_k - \mathbf{V}_p \dot{\mathbf{S}}_k \delta_g$ (last two terms). For Gaussian responses only the first sum remains; it varies over curves because their designs $\mathbf{X}_g$ differ (functional covariates). The scale enters through the curve's share of the residual sum of squares and of N . The cost is $O(GNp + KNp^2)$ per fit. The penalty and M_p do not scale with $\mathbf{c}$, so the scores are centred as in S4.1(e), and the estimators (IJ, OS, exact curve jackknife, regularized) and the two ways of adding the term to CL2 are those of S4.1(e). For Poisson and binary responses OS evaluates the leave-out gradient at the converged penalized IRLS solution without curve g .

Derivation: non-linear variants. Because $\text{Var}(\log \hat{\lambda})$ of the surface penalty is large and far from normal under dependence, $\mathbf{J}\Delta_{g,\rho}$ can be a poor approximation of the change of $\hat{\theta}$. The non-linear variants replace it by $\hat{\theta}_{(-g)}(\hat{\rho} + \Delta_{g,\rho}) - \hat{\theta}_{(-g)}(\hat{\rho})$, refitted by penalized IRLS at fixed smoothing parameters, for the one-step and the exact jackknife changes.

Result. In the dependent main-factorial settings it raises REML + CL2's coverage of β from to ($G = 40$) and from to ($G = 100$) at NA times the width, with interval score ratios NA. The within-setting rank correlation between a replicate's standard-error inflation and CL2's share of missed grid points in that replicate is -0.24 at $G = 40$ and -0.11 at $G = 100$. As a diagnostic, a refit with the smoothing parameters fixed at the setting's geometric-mean REML value, computed from the same replicates, is narrower than CL2 and covers better (, interval score ratio at $G = 40$); it changes both the typical amount of smoothing and its variability, so it does not separate the two. On the plasmode data the term raises REML + CL2's coverage of β by 0.0–1.8 pp per combination at interval score ratios 0.96–1.11.

Reason. The term raises coverage but not the interval score: the added width is not concentrated in the replicates where CL2 misses (the rank correlation above is negative or near zero), so it buys coverage at about the price of width. These results do not exclude a variance contribution to CL2's shortfall; they show that this first-order term does not target it.

Table 57: Smoothing-parameter terms for REML + CL2, dependent main-factorial settings at the default signal (200 replicates): coverage, then width / interval score relative to CL2. mgcv's term: mgcv's smoothing-parameter correction added to CL2. λ fixed: refit at the setting's geometric-mean REML smoothing parameters, computed from the same replicates (a diagnostic).

estimand	G	CL2	CL2 + mgcv's term	CL2 + curve-robust λ term	λ fixed (diagnostic)
$\beta(s,t)$	40	0.929 1.00 / 1.000	0.936 1.02 / 0.984	0.945 1.09 / 0.996	0.945 0.99 / 0.880
$\beta(s,t)$	100	0.946 1.00 / 1.000	0.951 1.02 / 0.993	0.958 1.07 / 1.009	0.955 1.00 / 0.936
$E(Y X)$	40	0.929 1.00 / 1.000	0.933 1.01 / 0.991	0.938 1.04 / 0.992	0.941 1.01 / 0.951
$E(Y X)$	100	0.943 1.00 / 1.000	0.945 1.01 / 0.997	0.947 1.02 / 1.000	0.949 1.01 / 0.978

S5 Misspecified truths and basis size

S5.1 Truths outside the spline space

Three truths are scored against the exact functions, not their basis projections (Gaussian, $R^2 = 0.5$, $D = 61$, 200 replicates): T0 the study's smooth β (control); T1 = T0 plus a two-dimensional Gaussian bump (SD 0.05, centre (0.3, 0.6)); T2 = T0 plus $\sum_{j=12}^{16} \sqrt{2} \cos(j\pi s) \sin(\pi t)$, directions the covariates barely excite. Each feature carries 20% of the β -term signal. Settings: truth $\times$ {independent $G = 100$, smooth $G = 100$, smooth $G = 40$ } $\times$ basis {default; 4 $\times$ finer knots: surface 23×31 , $k = 39$ along t }. The bump region is the set of grid points within two SDs of the bump centre.

With the default basis, approximation bias breaks every interval under T2 (REML + CL2 covers β , over the three error and G settings, at 0.30–0.56, NCV + CL2 at 0.27–0.33) and inside T1's bump (REML + CL2 0.17–0.58). With the basis with 4 $\times$ finer knots and dependent (smooth) errors, REML + CL2 recovers: β 0.93–0.95, bump region 0.93–0.94. NCV + CL2 stays biased (T2, smooth errors: 0.41–0.79). REML's coverage with the finer basis is bought with noisy estimates: its β MSE under smooth errors is 11.8–66.9 against 0.12–39.47 for NCV. Under independent errors the finer basis does not restore REML + CL2 (T2 β 0.86). α and γ are unaffected (REML + CL2 0.93–0.98).

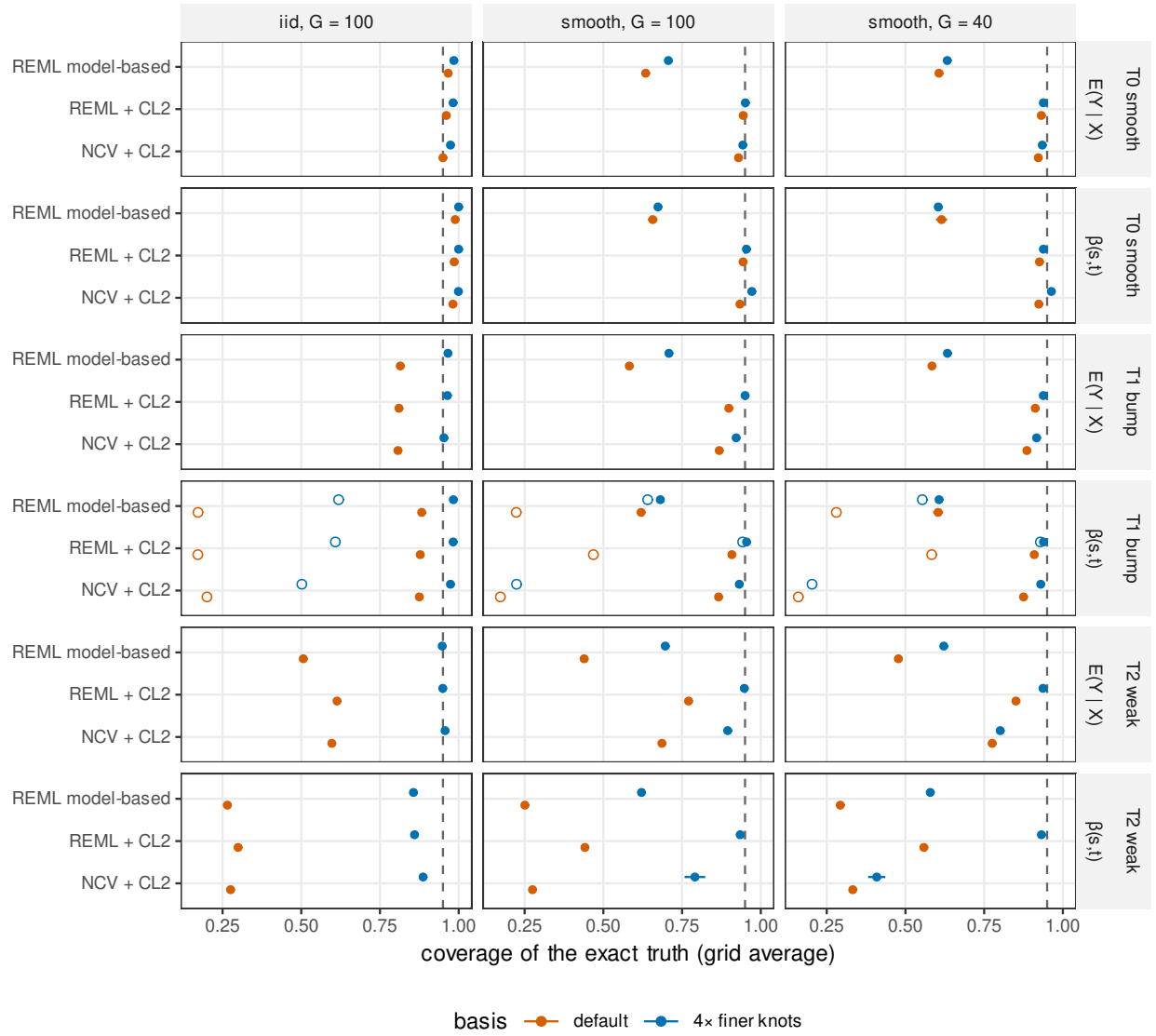


Figure 11: Misspecified truths: coverage of the exact $\beta(s, t)$ and mean by truth (rows), errors and G (columns), basis (colour) and interval; open symbols in the T1 row: coverage inside the bump region. Bars: ± 2 MC SE.

Table 58: Misspecified truths: feature amplitude, the feature's share of $\int \beta^2$, and the relative L_2 error and residual signal share after projecting β onto the default and surface bases with 4× finer knots.

truth	amplitude	share $\int \beta^2$	rel. L_2 err. default	resid. signal default	rel. L_2 err. 4× finer knots	resid. signal 4× finer knots
T0 smooth	0.00	0.000	0.000	0.000	0.000	0.000
T1 bump	9.29	0.206	0.261	0.023	0.004	0.000
T2 weak	3.75	0.925	0.891	0.222	0.097	0.001

Table 59: Misspecified truths: coverage (MC SE) of the exact truth per estimand, truth, setting, basis and interval.

estimand	truth	setting	basis	REML model-based	REML + CL2	REML + CL2 (freq.)	NCV + CL2
$\alpha(t)$	T0 smooth	iid, $G = 100$	default	0.956 (0.004)	0.956 (0.004)	0.941 (0.005)	0.945 (0.005)
$\alpha(t)$	T0 smooth	iid, $G = 100$	4× finer knots	0.978 (0.002)	0.975 (0.002)	0.940 (0.004)	0.967 (0.004)
$\alpha(t)$	T0 smooth	smooth, $G = 100$	default	0.641 (0.011)	0.950 (0.005)	0.948 (0.005)	0.941 (0.006)

Table 59: Misspecified truths: coverage (MC SE) of the exact truth per estimand, truth, setting, basis and interval.
(continued)

estimand	truth	setting	basis	REML model-based	REML + CL2	REML + CL2 (freq.)	NCV + CL2
$\alpha(t)$	T0 smooth	smooth, G = 100	4 × finer knots	0.739 (0.010)	0.954 (0.004)	0.942 (0.005)	0.944 (0.007)
$\alpha(t)$	T0 smooth	smooth, G = 40	default	0.614 (0.013)	0.937 (0.007)	0.933 (0.007)	0.919 (0.008)
$\alpha(t)$	T0 smooth	smooth, G = 40	4 × finer knots	0.640 (0.012)	0.938 (0.006)	0.926 (0.007)	0.932 (0.009)
$\alpha(t)$	T1 bump	iid, G = 100	default	0.953 (0.004)	0.954 (0.004)	0.941 (0.005)	0.947 (0.005)
$\alpha(t)$	T1 bump	iid, G = 100	4 × finer knots	0.978 (0.002)	0.974 (0.003)	0.942 (0.004)	0.966 (0.004)
$\alpha(t)$	T1 bump	smooth, G = 100	default	0.645 (0.011)	0.949 (0.005)	0.947 (0.005)	0.937 (0.006)
$\alpha(t)$	T1 bump	smooth, G = 100	4 × finer knots	0.739 (0.010)	0.955 (0.004)	0.942 (0.005)	0.943 (0.007)
$\alpha(t)$	T1 bump	smooth, G = 40	default	0.616 (0.013)	0.935 (0.007)	0.932 (0.007)	0.919 (0.008)
$\alpha(t)$	T1 bump	smooth, G = 40	4 × finer knots	0.641 (0.012)	0.938 (0.006)	0.926 (0.007)	0.927 (0.009)
$\alpha(t)$	T2 weak	iid, G = 100	default	0.842 (0.012)	0.945 (0.007)	0.937 (0.007)	0.941 (0.007)
$\alpha(t)$	T2 weak	iid, G = 100	4 × finer knots	0.975 (0.003)	0.973 (0.003)	0.939 (0.004)	0.968 (0.004)
$\alpha(t)$	T2 weak	smooth, G = 100	default	0.634 (0.013)	0.946 (0.006)	0.944 (0.006)	0.936 (0.006)
$\alpha(t)$	T2 weak	smooth, G = 100	4 × finer knots	0.732 (0.010)	0.955 (0.004)	0.941 (0.005)	0.941 (0.007)
$\alpha(t)$	T2 weak	smooth, G = 40	default	0.597 (0.013)	0.936 (0.007)	0.932 (0.008)	0.929 (0.008)
$\alpha(t)$	T2 weak	smooth, G = 40	4 × finer knots	0.629 (0.013)	0.936 (0.006)	0.926 (0.007)	0.939 (0.009)
$\beta(s,t)$	T0 smooth	iid, G = 100	default	0.989 (0.001)	0.986 (0.001)	0.940 (0.003)	0.982 (0.001)
$\beta(s,t)$	T0 smooth	iid, G = 100	4 × finer knots	1.000 (0.000)	1.000 (0.000)	0.942 (0.002)	0.999 (0.000)
$\beta(s,t)$	T0 smooth	smooth, G = 100	default	0.657 (0.008)	0.944 (0.003)	0.935 (0.003)	0.934 (0.004)
$\beta(s,t)$	T0 smooth	smooth, G = 100	4 × finer knots	0.673 (0.006)	0.954 (0.002)	0.939 (0.002)	0.972 (0.003)
$\beta(s,t)$	T0 smooth	smooth, G = 40	default	0.615 (0.009)	0.926 (0.004)	0.915 (0.004)	0.924 (0.005)
$\beta(s,t)$	T0 smooth	smooth, G = 40	4 × finer knots	0.604 (0.006)	0.938 (0.003)	0.923 (0.003)	0.963 (0.003)
$\beta(s,t)$	T1 bump	iid, G = 100	default	0.882 (0.001)	0.878 (0.001)	0.823 (0.002)	0.875 (0.002)
$\beta(s,t)$	T1 bump	iid, G = 100	4 × finer knots	0.983 (0.000)	0.982 (0.000)	0.907 (0.001)	0.974 (0.001)
$\beta(s,t)$	T1 bump	smooth, G = 100	default	0.620 (0.006)	0.908 (0.003)	0.899 (0.003)	0.866 (0.003)
$\beta(s,t)$	T1 bump	smooth, G = 100	4 × finer knots	0.681 (0.005)	0.955 (0.002)	0.939 (0.002)	0.932 (0.002)
$\beta(s,t)$	T1 bump	smooth, G = 40	default	0.604 (0.008)	0.909 (0.003)	0.898 (0.004)	0.875 (0.004)
$\beta(s,t)$	T1 bump	smooth, G = 40	4 × finer knots	0.607 (0.006)	0.938 (0.003)	0.924 (0.003)	0.930 (0.003)
$\beta(s,t)$	T2 weak	iid, G = 100	default	0.265 (0.002)	0.299 (0.002)	0.273 (0.002)	0.275 (0.005)
$\beta(s,t)$	T2 weak	iid, G = 100	4 × finer knots	0.856 (0.002)	0.860 (0.002)	0.769 (0.003)	0.887 (0.004)
$\beta(s,t)$	T2 weak	smooth, G = 100	default	0.251 (0.002)	0.441 (0.003)	0.433 (0.004)	0.275 (0.004)
$\beta(s,t)$	T2 weak	smooth, G = 100	4 × finer knots	0.621 (0.005)	0.935 (0.002)	0.920 (0.002)	0.791 (0.016)
$\beta(s,t)$	T2 weak	smooth, G = 40	default	0.293 (0.003)	0.559 (0.004)	0.548 (0.004)	0.333 (0.004)
$\beta(s,t)$	T2 weak	smooth, G = 40	4 × finer knots	0.579 (0.005)	0.932 (0.003)	0.918 (0.003)	0.409 (0.014)
$\gamma(t)$	T0 smooth	iid, G = 100	default	0.960 (0.004)	0.950 (0.005)	0.937 (0.005)	0.946 (0.005)
$\gamma(t)$	T0 smooth	iid, G = 100	4 × finer knots	0.976 (0.002)	0.973 (0.003)	0.936 (0.004)	0.966 (0.004)
$\gamma(t)$	T0 smooth	smooth, G = 100	default	0.655 (0.012)	0.948 (0.005)	0.945 (0.006)	0.935 (0.006)
$\gamma(t)$	T0 smooth	smooth, G = 100	4 × finer knots	0.734 (0.011)	0.950 (0.005)	0.939 (0.005)	0.935 (0.007)
$\gamma(t)$	T0 smooth	smooth, G = 40	default	0.621 (0.012)	0.946 (0.006)	0.941 (0.006)	0.927 (0.007)
$\gamma(t)$	T0 smooth	smooth, G = 40	4 × finer knots	0.643 (0.013)	0.932 (0.007)	0.921 (0.007)	0.929 (0.010)
$\gamma(t)$	T1 bump	iid, G = 100	default	0.956 (0.004)	0.949 (0.004)	0.937 (0.005)	0.944 (0.005)
$\gamma(t)$	T1 bump	iid, G = 100	4 × finer knots	0.976 (0.002)	0.975 (0.002)	0.936 (0.004)	0.968 (0.004)
$\gamma(t)$	T1 bump	smooth, G = 100	default	0.657 (0.012)	0.950 (0.005)	0.947 (0.005)	0.936 (0.006)
$\gamma(t)$	T1 bump	smooth, G = 100	4 × finer knots	0.736 (0.011)	0.950 (0.005)	0.937 (0.005)	0.939 (0.007)
$\gamma(t)$	T1 bump	smooth, G = 40	default	0.630 (0.013)	0.947 (0.006)	0.942 (0.006)	0.927 (0.007)
$\gamma(t)$	T1 bump	smooth, G = 40	4 × finer knots	0.640 (0.013)	0.932 (0.007)	0.922 (0.007)	0.924 (0.010)
$\gamma(t)$	T2 weak	iid, G = 100	default	0.839 (0.012)	0.946 (0.008)	0.935 (0.009)	0.930 (0.009)
$\gamma(t)$	T2 weak	iid, G = 100	4 × finer knots	0.975 (0.003)	0.974 (0.003)	0.935 (0.004)	0.966 (0.004)
$\gamma(t)$	T2 weak	smooth, G = 100	default	0.622 (0.013)	0.944 (0.007)	0.940 (0.007)	0.933 (0.007)
$\gamma(t)$	T2 weak	smooth, G = 100	4 × finer knots	0.729 (0.011)	0.950 (0.005)	0.937 (0.005)	0.931 (0.008)
$\gamma(t)$	T2 weak	smooth, G = 40	default	0.604 (0.013)	0.945 (0.007)	0.940 (0.007)	0.936 (0.007)
$\gamma(t)$	T2 weak	smooth, G = 40	4 × finer knots	0.635 (0.013)	0.934 (0.007)	0.922 (0.007)	0.931 (0.010)
$E(Y X)$	T0 smooth	iid, G = 100	default	0.967 (0.001)	0.961 (0.002)	0.939 (0.002)	0.950 (0.002)
$E(Y X)$	T0 smooth	iid, G = 100	4 × finer knots	0.985 (0.001)	0.982 (0.001)	0.938 (0.001)	0.974 (0.001)
$E(Y X)$	T0 smooth	smooth, G = 100	default	0.634 (0.005)	0.944 (0.002)	0.940 (0.002)	0.929 (0.003)
$E(Y X)$	T0 smooth	smooth, G = 100	4 × finer knots	0.706 (0.003)	0.951 (0.002)	0.939 (0.002)	0.943 (0.003)
$E(Y X)$	T0 smooth	smooth, G = 40	default	0.606 (0.005)	0.931 (0.003)	0.926 (0.003)	0.922 (0.004)
$E(Y X)$	T0 smooth	smooth, G = 40	4 × finer knots	0.633 (0.004)	0.938 (0.002)	0.926 (0.002)	0.935 (0.004)
$E(Y X)$	T1 bump	iid, G = 100	default	0.815 (0.001)	0.810 (0.002)	0.781 (0.002)	0.807 (0.002)
$E(Y X)$	T1 bump	iid, G = 100	4 × finer knots	0.966 (0.001)	0.964 (0.001)	0.914 (0.001)	0.953 (0.001)
$E(Y X)$	T1 bump	smooth, G = 100	default	0.582 (0.004)	0.898 (0.002)	0.894 (0.002)	0.868 (0.002)
$E(Y X)$	T1 bump	smooth, G = 100	4 × finer knots	0.708 (0.003)	0.950 (0.001)	0.938 (0.002)	0.922 (0.003)
$E(Y X)$	T1 bump	smooth, G = 40	default	0.584 (0.005)	0.912 (0.003)	0.906 (0.003)	0.886 (0.004)
$E(Y X)$	T1 bump	smooth, G = 40	4 × finer knots	0.633 (0.004)	0.938 (0.002)	0.926 (0.002)	0.916 (0.004)
$E(Y X)$	T2 weak	iid, G = 100	default	0.506 (0.002)	0.614 (0.003)	0.591 (0.003)	0.597 (0.004)
$E(Y X)$	T2 weak	iid, G = 100	4 × finer knots	0.948 (0.001)	0.949 (0.001)	0.896 (0.002)	0.957 (0.002)
$E(Y X)$	T2 weak	smooth, G = 100	default	0.439 (0.003)	0.771 (0.003)	0.765 (0.003)	0.686 (0.003)
$E(Y X)$	T2 weak	smooth, G = 100	4 × finer knots	0.697 (0.003)	0.948 (0.002)	0.937 (0.002)	0.895 (0.007)
$E(Y X)$	T2 weak	smooth, G = 40	default	0.478 (0.004)	0.851 (0.003)	0.844 (0.004)	0.775 (0.004)
$E(Y X)$	T2 weak	smooth, G = 40	4 × finer knots	0.622 (0.004)	0.937 (0.002)	0.926 (0.002)	0.801 (0.005)

Table 60: Misspecified truths: MSE and squared-bias share of the MSE for the REML and NCV estimates.

estimand	truth	setting	basis	mse_REML	mse_NCV	bias_share_REML	bias_share_NCV
$\beta(s,t)$	T0 smooth	iid, G = 100	default	0.0412	0.0269	0.0439	0.2129
$\beta(s,t)$	T0 smooth	iid, G = 100	4 × finer knots	0.0895	0.0409	0.0427	0.2603

Table 60: Misspecified truths: MSE and squared-bias share of the MSE for the REML and NCV estimates. (*continued*)

estimand	truth	setting	basis	mse_REML	mse_NCV	bias_share_REML	bias_share_NCV
$\beta(s,t)$	T0 smooth	smooth, G = 100	default	0.7524	0.1177	0.0075	0.1284
$\beta(s,t)$	T0 smooth	smooth, G = 100	4× finer knots	11.7721	0.1183	0.0026	0.1722
$\beta(s,t)$	T0 smooth	smooth, G = 40	default	2.7504	0.2633	0.0094	0.1114
$\beta(s,t)$	T0 smooth	smooth, G = 40	4× finer knots	52.7704	0.2551	-0.0011	0.1687
$\beta(s,t)$	T1 bump	iid, G = 100	default	0.3312	0.3636	0.8356	0.6942
$\beta(s,t)$	T1 bump	iid, G = 100	4× finer knots	0.2338	0.2371	0.4852	0.5810
$\beta(s,t)$	T1 bump	smooth, G = 100	default	1.1114	0.5116	0.2153	0.6274
$\beta(s,t)$	T1 bump	smooth, G = 100	4× finer knots	12.5268	0.4226	0.0024	0.6029
$\beta(s,t)$	T1 bump	smooth, G = 40	default	3.1364	0.6563	0.0820	0.5801
$\beta(s,t)$	T1 bump	smooth, G = 40	4× finer knots	54.3339	0.6555	-0.0011	0.5126
$\beta(s,t)$	T2 weak	iid, G = 100	default	39.4160	38.2757	0.9942	0.9867
$\beta(s,t)$	T2 weak	iid, G = 100	4× finer knots	7.7238	9.8557	0.8074	0.5261
$\beta(s,t)$	T2 weak	smooth, G = 100	default	38.1191	39.3778	0.9588	0.9904
$\beta(s,t)$	T2 weak	smooth, G = 100	4× finer knots	20.5659	18.9438	0.1660	0.4700
$\beta(s,t)$	T2 weak	smooth, G = 40	default	40.5430	39.8773	0.8820	0.9836
$\beta(s,t)$	T2 weak	smooth, G = 40	4× finer knots	66.8557	39.4673	0.0453	0.8759
E(Y X)	T0 smooth	iid, G = 100	default	0.0084	0.0081	0.0425	0.1337
E(Y X)	T0 smooth	iid, G = 100	4× finer knots	0.0138	0.0113	0.0455	0.2120
E(Y X)	T0 smooth	smooth, G = 100	default	0.0494	0.0340	0.0049	0.1024
E(Y X)	T0 smooth	smooth, G = 100	4× finer knots	0.1209	0.0374	0.0031	0.1464
E(Y X)	T0 smooth	smooth, G = 40	default	0.1415	0.0832	0.0048	0.0964
E(Y X)	T0 smooth	smooth, G = 40	4× finer knots	0.4706	0.0900	0.0006	0.1455
E(Y X)	T1 bump	iid, G = 100	default	0.0306	0.0311	0.7026	0.6733
E(Y X)	T1 bump	iid, G = 100	4× finer knots	0.0185	0.0177	0.1959	0.3172
E(Y X)	T1 bump	smooth, G = 100	default	0.0719	0.0629	0.2754	0.4329
E(Y X)	T1 bump	smooth, G = 100	4× finer knots	0.1270	0.0498	0.0088	0.2677
E(Y X)	T1 bump	smooth, G = 40	default	0.1671	0.1151	0.1259	0.3041
E(Y X)	T1 bump	smooth, G = 40	4× finer knots	0.4819	0.1070	0.0018	0.2242
E(Y X)	T2 weak	iid, G = 100	default	0.1408	0.1354	0.8539	0.8460
E(Y X)	T2 weak	iid, G = 100	4× finer knots	0.0354	0.0340	0.2890	0.1854
E(Y X)	T2 weak	smooth, G = 100	default	0.1742	0.1655	0.6219	0.7483
E(Y X)	T2 weak	smooth, G = 100	4× finer knots	0.1471	0.1115	0.0253	0.2664
E(Y X)	T2 weak	smooth, G = 40	default	0.3001	0.2349	0.3586	0.5629
E(Y X)	T2 weak	smooth, G = 40	4× finer knots	0.5287	0.2455	0.0086	0.5093

S5.2 Basis size with representable truths

With nested larger bases (2× and 4× finer knots; the truths stay representable), REML’s surface EDF under smooth errors rises from 27–42 (default) to 113–176 (4× finer knots) across families, NCV’s from 12–21 to 18–41. REML + CL2 stays near nominal (0.95–0.96 with 4× finer knots) but its intervals widen with REML’s variance: mean width for Gaussian β grows from 2.81 to 13.15. The NCV/REML MSE ratio for β falls to 0.01.

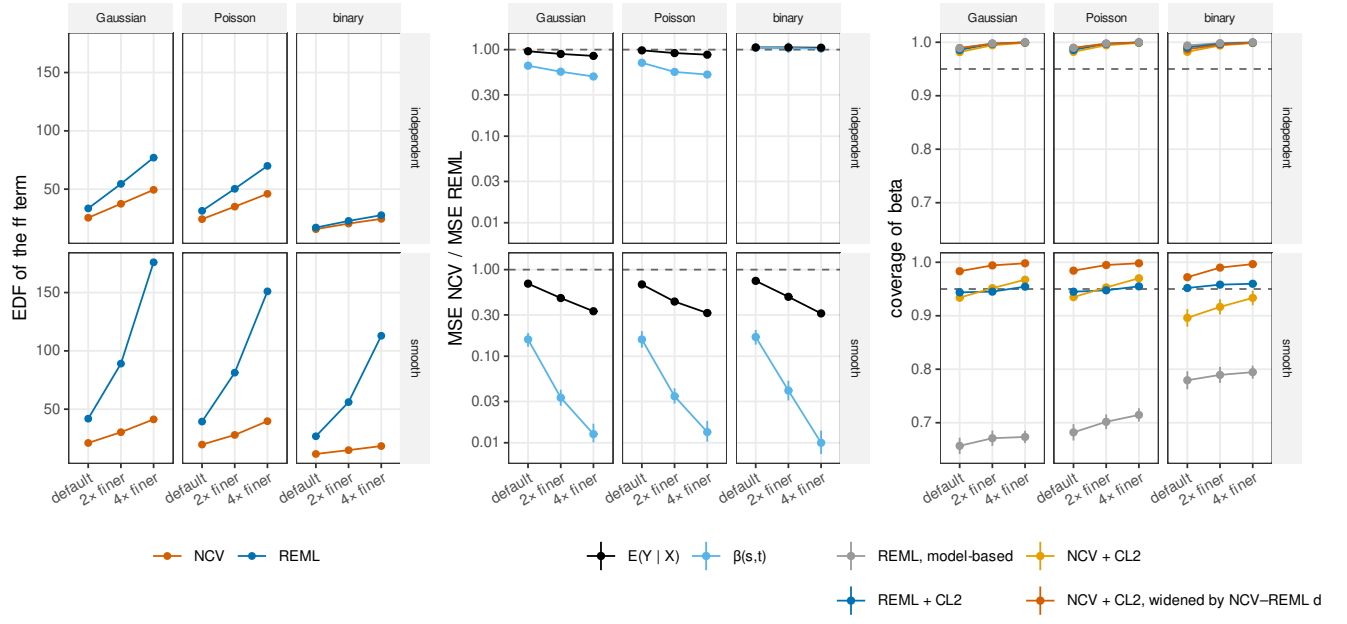


Figure 12: Basis size ($G = 100$, default signal, smooth truth): EDF of the surface term under REML and NCV, NCV/REML MSE ratio, and coverage of $\beta(s, t)$.

Table 61: Basis size: coverage (cov), interval score (IS) and mean width of $\beta(s, t)$ intervals per interval construction, and the NCV/REML MSE ratio.

family	error	basis	cov: NCV + CL2	cov: NCV + CL2, widened by NCV-REML difference	cov: REML + CL2	cov: REML, model-based	IS: NCV + CL2	IS: NCV + CL2, widened by NCV-REML difference	IS: REML + CL2	IS: REML, model-based	width: NCV + CL2	width: NCV + CL2, widened by NCV-REML difference	width: REML + CL2	width: REML, model-based	MSE ratio NCV/REML (β)
Gaussian	independent	default	0.982		0.989	0.986	0.989	0.764		1.016	1.023	0.724	0.796	0.982	0.652
Gaussian	independent	2+ finer knots	0.994		0.997	0.997	0.998	1.044		1.141	1.547	1.566	1.029	1.132	0.556
Gaussian	independent	4+ finer knots	0.999		0.999	1.000	1.000	1.358		1.471	2.154	2.183	1.355	1.539	0.490
Gaussian	smooth	default	0.934		0.983	0.941	0.657	1.480		2.418	3.493	7.226	1.094	2.314	0.156
Gaussian	smooth	2+ finer knots	0.952		0.994	0.945	0.671	1.557		5.187	7.803	15.488	1.266	5.150	0.033
Gaussian	smooth	4+ finer knots	0.967		0.998	0.954	0.673	1.695		10.327	15.510	29.378	1.479	10.313	0.013
Poisson	independent	default	0.982		0.990	0.986	0.989	0.438		0.467	0.573	0.578	0.413	0.553	0.705
Poisson	independent	2+ finer knots	0.994		0.997	0.997	0.998	0.577		0.627	0.854	0.867	0.568	0.622	0.553
Poisson	independent	4+ finer knots	0.999		0.999	1.000	1.000	0.747		0.805	1.173	1.192	0.745	0.804	0.515
Poisson	smooth	default	0.935		0.984	0.945	0.682	0.801		1.345	1.941	3.741	0.597	1.294	0.156
Poisson	smooth	2+ finer knots	0.953		0.995	0.948	0.702	0.833		2.806	4.212	7.704	0.687	2.790	0.034
Poisson	smooth	4+ finer knots	0.970		0.998	0.955	0.715	0.910		5.445	8.180	13.982	0.806	5.436	0.013
binary	independent	default	0.982		0.987	0.990	0.994	1.500		1.318	1.410	1.464	1.227	1.274	1.055
binary	independent	2+ finer knots	0.994		0.996	0.997	0.998	1.638		1.676	1.856	1.943	1.611	1.659	1.046
binary	independent	4+ finer knots	0.999		0.999	0.999	1.000	2.020		2.060	2.325	2.439	2.015	2.057	1.029
binary	smooth	default	0.896		0.972	0.952	0.779	2.545		3.742	5.227	7.995	1.512	3.465	0.167
binary	smooth	2+ finer knots	0.917		0.990	0.958	0.789	2.533		7.752	11.569	16.642	1.702	7.659	0.040
binary	smooth	4+ finer knots	0.953		0.997	0.960	0.795	2.593		16.870	25.640	35.337	1.896	16.838	0.010

S6 Data and code

Recommended analysis. The procedures recommended in Section 6 of the paper will be the defaults of forthcoming releases of refund (version X.Y or later) and mgcv (version X.Y or later). With these releases and the bases of the study, the analysis reads as follows. The evaluated interval for $\alpha(t)$ is that of the full intercept curve, the scalar intercept plus the centred Intercept(yindex) term.

```
library(refund)
bs_t <- list(bs = "ps", k = 12, m = c(2, 1))
bs_ff <- list(bs = "ps", k = c(8, 10), m = list(c(2, 1), c(2, 1)))
form <- Y ~ ff(X, xind = s, splinepars = bs_ff) + z
# inference: REML fit, CL2 with Satterthwaite critical values (the defaults)
fit_reml <- pffr(form, yind = t, data = d, bs.yindex = bs_t, bs.int = bs_t)
ci_reml <- coef(fit_reml, ci = "pointwise")
# shape of the surface: curve-blocked NCV estimate
fit_ncv <- pffr(form, yind = t, data = d, bs.yindex = bs_t, bs.int = bs_t,
               method = "NCV")
est_ncv <- coef(fit_ncv)
# covariate with few principal components: ffpc() term with a small K
fit_pc <- pffr(Y ~ ffpc(X, xind = s, splinepars = bs_t,
                      decomppars = list(pve = 0.95, useSymm = TRUE)) + z,
              yind = t, data = d, bs.yindex = bs_t, bs.int = bs_t)
ci_pc <- coef(fit_pc, ci = "pointwise")
# several curves per subject: cluster by subject
fit_rm <- pffr(form, yind = t, data = d, bs.yindex = bs_t, bs.int = bs_t,
              cluster = subject)
```

Software. The study library ran R 4.6.1 with 101 packages (full list with versions in paper/data/restart-lib-versions.txt); the key versions are in the next table. refund is the development branch pffr-refactor on GitHub (refunders/refund). The study build is commit 79a346fb (reported version 0.1-40), installed by its immutable commit with remotes::install_github("refunders/refund@79a346fb").

mgcv is a pre-release build of version 1.9-5, in two builds. Both exact source builds are deposited in the code archive (software/mgcv_1.9-5.tar.gz, unpatched; software/mgcv_1.9-5-patched.tar.gz, patched), together with the patch (software/mgcv-1.9-5-ncv-vc-and-cxwyd.patch); their SHA-256 checksums are in the table below. The unpatched build ran the main synthetic factorial and the heteroskedastic and low-rank-covariate extension settings. The patched build ran the basis-size extension (study settings 112–123), the settings with 20 curves and with a zero surface (study settings 124–152), the whole plasmode study and all later experiments (covariance components and Satterthwaite degrees of freedom, Appendices A.6 and A.7 of the paper, S1.7; the smoothing-parameter terms and NCV-centred corrections, S4; the comparators, S3.2 and S3.3; the misspecified truths, S5.1; the principal component term, S1.9). The patch changes V_c of NCV fits (in the unpatched build it holds only the smoothing-parameter correction and is not a valid covariance; with the patch it is V_p plus that correction) and two bam() paths (NCV, and non-Gaussian discrete fits with AR(1) errors). The REML, NCV and CL2 computations evaluated here use none of these; the study uses V_p as the model-based covariance of NCV fits throughout, so only the row “... with $\hat{\lambda}$ correction” in S4.2, which uses V_c of the NCV fit, depends on the patch. Runs on the patched build reproduced the stored results of the unpatched build to about 10^{-9} . Functional GEE: fastFGEE 0.2.2 from CRAN. No container or renv lockfile is provided; the package list and checksums are the record of the computing environment.

Reproducing one synthetic setting. From the root directory of the code archive, with R 4.6.1, single-threaded BLAS, mgcv installed from the deposited source build and refund from commit 79a346fb, the following runs the 200 replicates of study setting 12 (Gaussian responses, smooth errors, $G = 100$, default signal) and summarises them:

```
export OPENBLAS_NUM_THREADS=1 OMP_NUM_THREADS=1
export REFUND_SHA=79a346fb8cc1220a2a6fd2dc8dd3e33166a21061
R CMD INSTALL software/mgcv_1.9-5.tar.gz
Rscript -e 'remotes::install_github("refunders/refund@79a346fb")'
for run in 1 2; do RESTART_CELLS=12 SLURM_ARRAY_TASK_ID=1 Rscript scripts/run-
array.R 100 $run results-12; done
Rscript -e 'source("scripts/setup.R"); readr::write_csv(summarise_study("results-
12")$cells, "cells-12.csv")'
```

scripts/run-array.R (arguments: chunk size, half of the replicates, output directory) runs the tasks of the selected setting. REFUND_SHA declares the refund commit to the study's software check, because a GitHub installation does not record it. The last line summarises the setting's replicates; analysis/run-analysis.R writes the summary tables for the whole study, from which writeup/RESULTS.qmd and this supplement compute their exhibits. The route was checked: one task run with scripts/run-array.R completes, and the summary of the 200 stored replicates of setting 12 reproduces that setting's rows of the study's summary table, mean squared errors identical to all digits.

Table 62: Key package versions of the study library.

Package	Version
R	4.6.1
dplyr	1.2.1
gamm4	0.3-0
ggplot2	4.0.3
lme4	2.0-6
Matrix	1.7-6
mgcv	1.9-5
nlme	3.1-171
refund	0.1-40

Table 63: SHA-256 checksums of the mgcv source archives and the patch.

file	sha256
mgcv 1.9-5, patched build	ad56a0a02f8cf2ab84abe6f0d5271ff6662764e7e5d4eb4afff24143e728cb5d
mgcv 1.9-5, unpatched build	b20330037cdcd78f4c2018a3d64aed946c3ab20de9bd09c8677070d3b9df6fb4
patch	4684be40034776fae7bb402e8c324eee0f11d737032d9ff6cbfc350fb78be306

Exhibits and files. All exhibits are computed from summary files of the study's code archive: the synthetic study from summaries/final/ (coverage, MSE, relative error, EDF, timing) and the studentized-error and design-effect tables computed from the replicate records; the plasmode study from summaries/plasmode/; NCV-centred intervals widened by the NCV-REML difference from summaries/hybrid/; Appendix A.6 of the paper (covariance components) from ablation/summaries/; S1.9 from ffpc-pilot/ffpc-1rz-summary.csv and ffpc-pilot/ffpc-pilot-summary.csv; Appendix A.7 of the paper, S1.7 and the Satterthwaite results of S1.2 and S2.1 from satterthwaite/summaries/; the 20-curve and zero-surface settings (S1.2, S1.8) from small-g-null/summaries/; S3.2 from comparators/summaries/; S3.3 from competitors/fgee/summaries/; S4.1 (d)-(g) from ncv-rbc/summaries/ (including the combo2-* files

for (e)–(g)) and `lambda-correction/summaries-ncv*/`; S4.2 from `ncv-jack-recheck/` and `ncv-jack-recheck/summaries/`; S4.3 from `lambda-correction/summaries/` and `summaries-plasmode/`; S5.1 from `misspec/summaries/`.

Data. The DTI, running, gait, 8-lead ECG, ocean, weather and electricity data are public (sources in Appendix B.2 of the paper). The two cardiology datasets are patient and volunteer data that cannot be shared; only aggregate results are reported, so their plasmode results cannot be reproduced from public data.

Use of generative AI. All code and text, including this supplement, were generated by large language models (details in the paper’s statement on the use of generative AI); the author takes full responsibility for the content.

References

- Calonico, Sebastian, Matias D. Cattaneo, and Max H. Farrell. 2018. “On the Effect of Bias Estimation on Coverage Accuracy in Nonparametric Inference.” *Journal of the American Statistical Association* 113 (522): 767–79. <https://doi.org/10.1080/01621459.2017.1285776>.
- Loewinger, Gabriel, Alex W. Levis, Erjia Cui, and Francisco Pereira. 2025. “Fast Penalized Generalized Estimating Equations for Large Longitudinal Functional Datasets.” <https://arxiv.org/abs/2506.20437v3>.
- Marra, Giampiero, and Simon N. Wood. 2012. “Coverage Properties of Confidence Intervals for Generalized Additive Model Components.” *Scandinavian Journal of Statistics* 39 (1): 53–74. <https://doi.org/10.1111/j.1467-9469.2011.00760.x>.
- Tukey, John W. 1977. *Exploratory Data Analysis*. Reading, MA: Addison-Wesley.
- Wood, Simon N. 2011. “Fast Stable Restricted Maximum Likelihood and Marginal Likelihood Estimation of Semiparametric Generalized Linear Models.” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 73 (1): 3–36. <https://doi.org/10.1111/j.1467-9868.2010.00749.x>.
- . 2024. “On Neighbourhood Cross Validation.” <https://arxiv.org/abs/2404.16490>.
- Wood, Simon N., Natalya Pya, and Benjamin Säfken. 2016. “Smoothing Parameter and Model Selection for General Smooth Models.” *Journal of the American Statistical Association* 111 (516): 1548–63. <https://doi.org/10.1080/01621459.2016.1180986>.