Hybrid Partial Least Sq

Axiom [ax

rem]Lemma [theoren

Fact Remark

BIOMETRICS 000, 000–000
DOI: 000 000 0000

Hybrid Partial Least Squares Regression with Multiple Functional and Scalar Predictors

Keywords: dimension reduction, func-

1. Introduction

Biomedical studies often collect multimodal data, as in the Emory University renal study (Chang et al., 2020; Jang, 2021), which records multiple renogram curves (functional

tional data analysis, multiple data modalities, multivariate data analysis, multivariate functional data, partial least square

and scalar-valued i.i.d. covariates in a linear regression model:

$$Y_i = \boldsymbol{\beta}^\top \mathbf{Z}_i + \sum_{k=1}^{K} \int \beta_k(t) X_{ik}(t) dt + \epsilon_i,$$

where Y_i is a scalar response, $X_{i1}(t), \ldots, X_{iK}(t) \in \mathbb{L}^2[0,1]$ are functional predictors, $\mathbf{Z}_i = (Z_{i1}, \ldots, Z_{ip})^\top$ is

data) along with scalar side information for each kidney. Such distinct yet related physiological signals can be represented as functional-a Euclidean vector covariate, and ϵ_i denotes observational noise. For notational convenience, we assume throughout that the responses and predictors are centered, so the intercept term can be omitted. In addition to the ill-posedness of the infinite-dimensional slope functions $\beta_k(t)$ and the high dimensionality of the predictors $(K+p)$, a further challenge comes from the strong correlations between functional and scalar predictors. This paper addresses all three issues in a unified manner by introducing a hybrid partial least squares (PLS) regression framework defined on a novel hybrid Hilbert space.

1.1 *Previous works and limitations*

To address the ill-posedness stemming from the infinite-dimensionality of the functional components, common remedies include basis expansion, using power-series (Goldsmith et al., 2011), B-splines (Cardot et al., 2003; Cai and Hall, 2006), or wavelets (Zhao et al., 2012), and structure-aware regurection is principal component analysis (PCA) regression, which applies PCA separately to the functional- (by, for example, Happ and Greven

larization, such as roughness penalties For the high-dimensionality, a major solution is using derived inputs. A straightforward approach in this diear regression with functional predictors alone (Hall and Horowitz, 2007; Reiss and Ogden, 2007; Febrero-Bande et al., 2017). However, since the PCA-derived inputs are not in-

2018) and scalar predictors, and then performs a classical multivariate regression on the combined scores. PCA regression is well-studied for lin-

In terms of predictive power, partial least squares (PLS) regression is a powerful alternative. It iteratively constructs a set of or-

formed by the response variable, they are not guaranteed to capture the core regression relationship with a small number of derived inputs.

for regression. PLS for linear regression in the context of functional predictor alone was introduced by Preda and Saporta (2005) for the case of

thogonal latent components from the predictors that have the maximal covariance with the response variable, and use the resulting scores the framework by incorporating basis approximations and roughness penalties to promote smoothness. Delaigle and Hall (2012) provided the first

a single predictor. Motivated by regression on chemometric spectra, Reiss and Ogden (2007), Aguilera et al. (2010), and Aguilera et al. (2016) extended onalization. For a comprehensive review, see Febrero-Bande et al. (2017). More recently, Beyaztas and Lin Shang (2022) extended the framework to accommodate

thorough theoretical analysis, while Saricam et al. (2022) proposed a computationally efficient procedure based on Golub-Kahan bidiagmultiple functional predictors. However, these existing approaches overlook the potentially strong correlations between functional and scalar components, which can lead to multicollinearity and sub-optimal predictive performance. Multimodal correlations have mostly been addressed within an unsudata, while Geng et al. (2020) considered joint precision matrix estimation for brain measurements and confounding scalar covariate. Jang (2021) proposed

pervised learning framework. For instance, Kolar et al. (2014) studied the estimation of joint undirected graphical models for functional and vector a joint PCA method that accounts for correlations between functional and scalar data. However, the resulting components are still not informed by the response and may fail to capture correlation with the outcome.

1.2 *Our contributions*

To address the gaps mentioned above, we unify the remedies pro-

propose a hybrid PLS regression framework that integrates functional and vector predictors in a principled and coherent manner. To

posed for these three issues: basis expansion, partial least squares, and accounting for correlations between the functional and scalar components. We extract predictive structure from these jointly observed and potentially correlated data types, we define a Hilbert space that treats the tuple of functional and vector components as a single hybrid object, equipped with a suitable inner product. The hybrid PLS direction is then obtained by is readily applicable to dense and irregular functional data and supports regularization techniques to prevent overfitting and re-

iteratively maximizing the empirical covariance with the response, subject to a unit-norm constraint in this Hilbert space. Our framework

2. Background on partial least squares and its extension to hybrid predictors

For intuition, let us return to the high-dimensional Euclidean predictor set-

duce variance. We also provide the mathematical properties that justify our algorithm.

vector $\left(\widehat{\rho}_1^{[1]}, \ldots, \widehat{\rho}_1^{[L]}\right)^\top \in \mathbb{R}^L$. To retain the regression relationship, the l-th PLS direction $\hat{\boldsymbol{\xi}}_l$ solves:

$$\max_{\mathbf{h}} \widehat{\mathrm{Cov}}^2\left(\{\langle \mathbf{h}, \mathbf{Z}_i \rangle, Y_i\}_{i=1}^n\right) \text{ s.t. } \|\mathbf{h}\|_2 =$$

where the two $\widehat{\mathrm{Cov}}^2$ denote sample cross-

ting $Y_i = \beta^\top \mathbf{Z}_i + \epsilon_i$. A common way to address ill-posedness and correlation is to approximate the high-dimensional vector $\mathbf{Z}_i$ using a low-dimensional covariance and sample covariance, respectively. A standard algorithm for solving this optimization problem, called nonlinear iterative partial least squares (NIPALS) is presented in Algorithm 1.

To extend Algorithm 1 to the hybrid regression model (1), we must compute hybrid objects in both the PLS direction estimation and predictor residualization steps (lines 4 and 15). These computations cannot

Algorithm 1 Scalar partial least squares regression

1: Standardize each $\mathbf{Z}_1, \ldots, \mathbf{Z}_n$ so that each feature have mean zero and variance one. Standardize $Y_1, \ldots, Y_n$.
2: **for** $l = 1, 2, \ldots, L$ **do**
3: **PLS direction and score estimation:**
4: $\hat{\boldsymbol{\xi}}^{[l]} \leftarrow \arg\max_{\boldsymbol{\alpha}} \widehat{\mathrm{Cov}}^2(\{$ 1 ▷ PLS direction
5: $\widehat{\rho}_i^{[l]} \leftarrow \langle \hat{\boldsymbol{\xi}}^{[l]}, \mathbf{Z}_i^{[l]} \rangle$, $i = 1, \ldots, n$ ▷ PLS score
6: **Residualization:**
7: $\nu^{[l]} \leftarrow \frac{\sum_{i=1}^n Y_i^{[l]} \widehat{\rho}_i^{[l]}}{\sum_{i=1}^n \widehat{\rho}_i^{[l]2}}$ ▷

tion points t to lines 4 and 15, treating $\big(X_{i1}(t), \ldots, X_{iK}(t), Z_{i1}, \ldots, Z_{ip}\big) \in \mathbb{R}$ as a Euclidean predictor, applying steps 4 and 15, and aggregating the results. This

be separated into functional and scalar parts, as their correlation must be addressed. A naive approach is to add loops over the observed evaluavariability across the domain, resulting in overfitting and unstable predictions. The main challenge of extension to the hybrid setting

pointwise method, however, is computationally prohibitive for densely observed functional predictors, infeasible for irregular data, and prone to lies in the following points. First, independent variables consist of multiple highly structured images and scalar predictors. Second, our sample size is small compared to the dimension and number of functional and scalar predictors. Third, existing partial least squares (PLS) methods can only acet al., 2017; Beyaztas and Shang, 2020); or (ii) a univariate functional predictor with other scalar predictors (Wang, 2018). The next section proposes a new ex-

commodate (i) univariate or multivariate functional predictors without any scalar predictors (Preda and Saporta, 2005; Delaigle and Hall, 2012; Febrero-Bande are arithmetic operations in an inner product space. We therefore formally define the Hilbert space of hybrid random predictors along with its

tension framework that addresses these issues.

3. Proposed PLS Algorithm

Our new framework builds on the fact that all computations in Algorithm 1 defined as the Cartesian product of K copies of the space of square-integrable functions on $[0, 1]$ and the p-dimensional Euclidean space: $\mathbb{H} := (\mathbb{L}^2[0,1])^K \times \mathbb{R}^p$. An element $h \in$

corresponding complete inner product, and leverage them to extend Algorithm 1.

definition

[. Hybrid space] Let $\mathbb{H}$ be a product space cation. We define an inner product on $\mathbb{H}$ for any two elements $h_1 = (f_1, \ldots, f_K, \mathbf{u})$ and $h_2 = (g_1, \ldots, g_K, \mathbf{v})$ as:

$$\langle h_1, h_2 \rangle_{\mathbb{H}} := \sum_{k=1}^{K} \int f_k(t) g_k(t)\, dt + \omega \mathbf{u}^{\top}$$

where ω is a positive constant

$\mathbb{H}$ is an ordered tuple $h = (f_1, \ldots, f_K, \mathbf{u})$, where $f_k \in \mathbb{L}^2[0, 1]$ for $k = 1, \ldots, K$ and $\mathbf{u} \in \mathbb{R}^p$. This space is equipped with element-wise vector addition and scalar multipliis a positive weight that needs to be prespecified or estimated. It is mainly used to take into account heterogeneity between functional and scalar

$(\omega > 0)$. The inner product induces a norm $\|\cdot\|_{\mathbb{H}}$ on the space, defined as $\|h\|_{\mathbb{H}} := \langle h, h \rangle_{\mathbb{H}}^{1/2}$, and a corresponding metric $d(h_1, h_2) = \|h_1 - h_2\|_{\mathbb{H}}$.

In (3), ω

parts in terms of measurement scale and/or amount of variation (see Section 3.2). Without loss of generality and for the clarity of illustration, all the following theoretical results will be derived for $\omega = 1$. The results remain valid for any positive weights.

lemma

[

Predictor] For the Hilbert space $\mathbb{H}$ defined in Definition 3.1, The Borel σ-field on $\mathbb{H}$, denoted $\mathfrak{B}(\mathbb{H})$, is the smallest σ-field containing the

. Hybrid Hilbert space] The hybrid space $\mathbb{H}$ is a separable Hilbert space.

Proof of Lemma 3.2 is provided in Appendix **??**.

definition

[

. Hybrid

ing and Eubank 2015). A hybrid predictor $W_i = (X_{i1}(t), \ldots, X_{iK}(t), \mathbf{Z}_i)$ is a measurable mapping from a probability space $(\Omega, \mathfrak{F}, P)$ into $\big(\mathbb{H}, \mathfrak{B}(\mathbb{H})\big)$.

Then the joint regres-

class $\mathfrak{M}$ of all sets of the form $\{q \in \mathbb{H} \mid \langle q, h \rangle \in O\}$, for any $h \in \mathbb{H}$ and any open subset $O \subseteq \mathbb{R}$ (details can be found in Theorem 7.1.1 of Hsto a different Hilbert space, possibly defined over a distinct compact domain in $\mathbb{R}^d$, for arbitrary d, and observed at different time

4)

sion model (1) can be concisely written as

$$Y_i = \langle \beta, W_i \rangle_{\mathbb{H}} + \epsilon_i, \text{ wh}$$

remark

O. ur method is applicable to settings where each functional predictor belongs provides an efficient and robust means of producing PLS components and scores in the presence of multiple dense and/or irregular func-

points. However, for notational simplicity, our discussion assumes a common Hilbert space over domain $[0, 1]$ for all functional predictors.

Our approach

ploit structural relationships within and between the functions to avoid overfitting of the PLS components and to improve the generalizability

tional predictors and scalar predictors. It also incorporates a regularization scheme that enables the algorithm to borrow strength and extion of smoothed PLS components and orthogonalization, detailed in Sections 3.3.1 and 3.3.2, respectively. After a suitable number of iterations,

ity and interpretability of the predictive model. Each iteration of our approach consists of two subroutines: regularized estimathe hybrid regression coefficient is estimated, as described in Section 3.4. For notational simplicity, we omit the iteration index l in the following discussion, with the understanding that the subroutines apply to any iteration. The complete algorithm linearly independent, for example, cubic B-splines, the Fourier basis, or an orthonormal polynomial basis of degree greater than three. Us-

is summarized in Algorithm 2.

3.1 *Preliminary step 1: finite-basis approximation*

Let $b_m(t)$ be a twice-differentiable basis of $\mathbb{L}^2([0,1])$ whose second derivatives are also coefficient (with iteration indices suppressed) are represented as follows:

$$X_{ij}(t) = \sum_{m=1}^{\infty} \theta_{ijm}\, b_m(t),\ \beta_j(t) = \sum_{m=1}^{\infty}$$

In practice, the full set of coefficients can not

ing this basis, the jth functional predictor, regression coefficient, PLS component direction, and orthogonalization regression be obtained with finite sample size, as functional data are measured on a finite grid. Thus, we truncate the expansion at M terms. We choose a moderately large M (e.g., 15 or 20) to capture functional variation without fine-tuning, as smoothness is handled via penalization (see Section 3.3.1).

rection and orthogonalization regression coefficient to admit the following expansion:

$$\xi_j(t) = \sum_{m=1}^{M} \gamma_{jm}\, b_m(t), \quad \delta_j(t) = \sum_{m=1}^{M} \gamma$$

This implies that all compu-

The truncated expansions of the predictor and coefficient are denoted as

$$\widetilde{X}_{ij}(t) := \sum_{m=1}^{M} \theta_{ijm}\, b_m$$

and our suggest method restricts each PLS component ditations in this paper are carried out entirely within the subspace

$$\widetilde{\mathbb{H}} := \mathrm{span}\big(b_1(t), \ldots,$$

6)

The ith hybrid predictor, projected on $\widetilde{\mathbb{H}}$, is represented by the tuple

$$\widetilde{W}_i := (\widetilde{X}_{i1}, \ldots, \widetilde{X}_{iK}, \boldsymbol{Z}_i).$$

Let $\boldsymbol{\theta}_{ij}$, $\boldsymbol{\eta}_j$, $\boldsymbol{\gamma}_j$, and $\boldsymbol{\pi}_j$ denote the M-dimensional vectors of coefficients:

$$\boldsymbol{\theta}_{ij} := (\theta_{ij1}, \ldots, \theta_{ijM})^\top, \ \boldsymbol{\eta}_j := (\eta_{j1}, \ldots$$

For the predictors, we stack the coefficients ficient vectors across observations into the matrix

$$\Theta_j := (\boldsymbol{\theta}_{1j}, \ldots, \boldsymbol{\theta}_{nj})^\top$$ (8)

and construct the full design matrix

$$\Theta := (\Theta_1, \ldots, \Theta_K, \mathbf{Z}$$ (9)

Let us denote the response vector as $\mathbf{y} := (y_1, \ldots, y_n)^\top$. Let $B, B'' \in \mathbb{R}^{M\times M}$ denote the Gram matrices of the basis functions and their second derivatives, with entries

$$B_{m,m'} := \int_0^1 b_m(t)\, b_{m'}(t)\, dt, \quad B''_{m,m'}$$

for $m, m' = 1, \ldots, M$. We then

define the block-diagonal matrices

$\mathbb{B} := \mathrm{blkdiag}(B, \ldots,$ (1)

Then the full data for the hybrid PLS problem at the l-th iteration can be represented by the tuple

$(\mathbb{B}, \mathbb{B}'', \Theta, \mathbf{y}) \in \mathbb{R}^{(MK+p)\times(MK+p)} \times \mathbb{R}^{(}$

with the index l omitted for brevity.

remark

W. hile different bases could be used for each functional predictor, we adopt a common basis for simplicity. The definitions of $\mathbb{B}$ and $\mathbb{B}''$ remain general enough to accommodate distinct bases if needed.

3.2 *Preliminary step 2: data pre-processing*

Functional and scalar elements of the hybrid predictors often have incompatible units and/or exhibit difchance of contributing to the predictor/response structure; and ii) a predictor with high correlation to Y but relatively low variance may be

ferent amounts of variation. This can be problematic for our PLS framework which is not scale invariant as: i) each predictor has different

The first step is to account for discrepancies *within* respective functional and scalar parts, if needed. If the functional parts $\widetilde{X}_{i1}(t), \ldots, \widetilde{X}_{iK}(t)$ are measured in differ-

overlooked. To obtain PLS components that have a meaningful interpretation, we standardize the predictor data via the following steps.

dictors $\mathbf{Z}_i = (Z_{i1}, \ldots, Z_{ip})^\top$ exhibit different amounts of variation, one can standardize them to have mean zero and unit variance. The second step is to elim-

ent units or have quite different domains, one can standardize them to have mean zero and integrated variance of one. If multivariate scalar preproduct (3) that ensures functional and vector parts have comparable variance. A sensible data-driven approach to choosing an appropriate

inate the discrepancies *between* functional and scalar parts. To accomplish this aim, we choose an appropriate weight ω in the hybrid inner weight is to set

$$\omega = \frac{\sum_{i=1}^{n} \sum_{k=1}^{K} \|\widetilde{X}_{i}}{\sum_{i=1}^{n} \|\mathbf{Z}_i\|}$$

In practice, instead of using inner product weighted by $\omega^{1/2}$, one can implement this weighting scheme by formulating the hybrid object as $\widetilde{W}_i = \big(\widetilde{X}_{i1}(t), \ldots, \widetilde{X}_{iK}(t), \omega^{1/2}\mathbf{Z}_i\big)$, whose vector part has been scaled by a factor of $\omega^{1/2}$.

3.3 *Iterative steps*

The iterative process presented here yields an tion of the PLS component direction (Section 3.3.1) and residualization (Section 3.3.2). The proofs are deferred to Appendix ??. The properties

orthonormal hybrid basis that effectively captures the predictor-response relationships. It proceeds through two intermediate steps: the estima-

nent direction.

We begin by formally introducing the core optimization problem pertinent to the PLS direction estima-

of the resulting estimates are introduced in Section 4.

3.3.1 *Iterative step 1: regularized estimation of PLS compo-*

PLS component direction estimation step that promotes smoothness (Proposition 3.5). Furthermore, we detail an efficient computational

tion, which is formulated as a generalized Rayleigh quotient (Proposition 3.4). Building upon this foundational concept, we present our regularized rection. It fully leverages the continuous nature of the functional components and the function-scalar hybrid structure. We describe the strategy at

scheme (Proposition 3.6).

Core optimization problem. We present the core optimization problem that directly estimates the hybrid PLS component dithe l-th iteration. The PLS component direction is estimated by the unit norm direction $\xi^{[l]} \in \widetilde{\mathbb{H}}$ that maximizes the squared em-

pirical covariance, which quantifies the linear dependence between the PLS scores $\langle \widetilde{W}_1, \xi \rangle_{\mathbb{H}}, \ldots, \langle \widetilde{W}_n, \xi \rangle_{\mathbb{H}}$ and the responses $y_1, \ldots, y_n$, defined as

$$\widehat{\mathrm{Cov}}(\langle \widetilde{W}, \xi \rangle_{\mathbb{H}}, Y) := \frac{1}{n} \sum_{i=1}^{n} y_i \langle \widetilde{W}_i, \xi \rangle_{\mathbb{H}}.$$

We detaining these coefficients is equivalent to solving the maximization problem (14). The following proposition formulates this coef-

4)

note the estimated PLS component direction as

$$\hat{\xi} := \arg\max_{\xi \in \widetilde{\mathbb{H}}} \widehat{\mathrm{Cov}}^2 \big(\langle \widehat{W}$$

Here, $\hat{\xi} \in \widetilde{\mathbb{H}}$ is an ordered pair expanded as:

$$\hat{\xi} = \big(\hat{\xi}_1(t), \ldots, \hat{\xi}_K(t),$$

where $\hat{\zeta}$ is the scalar part. Ob-

iteration of the PLS algorithm, the coefficients of the squared covariance maximizer defined in (14), *is obtained as*

$$\left(\hat{\gamma}_{11}, \ldots, \hat{\gamma}_{1M}, \ldots, \hat{\gamma}_{K1}, \ldots, \hat{\gamma}_{KM}, \hat{\boldsymbol{\zeta}}^{\top}\right)$$

ficients obtaining procedure as a generalized Rayleigh quotient:

Proposition 1. *Let* $(\mathbb{B}, \mathbb{B}'', \Theta, \mathbf{y})$ *denote the observed data defined in* (12). *At the* l*-th*

6)

offers an efficient way to estimate the PLS component direction $\hat{\xi}$, its functional components, $\hat{\xi}_1, \ldots, \hat{\xi}_K$, may not be smooth. which complicates

where

$$V := \frac{1}{n^2}(\mathbb{B}\Theta^\top \mathbf{y})(\mathbb{B}\Theta$$

The proof of Proposition 3.4 is provided in Appendix **??**.

Proposed regularized estimation procedure. Although Proposition 3.4

balances predictive performance with smoothness. Specifically, we penalize the roughness of each $\hat{\xi}_j$ using its integrated squared second deriva-

interpretation and can lead to overfitting and unstable predictions. To address this, we propose a regularized extension that

7)

control the complexity of the estimated functional components. One possible approach is to extend the smoothed functional PCA framework of Rice

tive

$$\mathrm{PEN}(\hat{\xi}_j) := \int_0^1 \{\hat{\xi}_j''(t$$

Instead of solely maximizing the squared empirical covariance $\widehat{\mathrm{Cov}}^2\big(\langle \widetilde{W}, \xi\rangle_{\mathbb{H}}, Y\big)$, we incorporate this roughness penalty to simultaneously maximizing covariance and penalizing roughness. However, this approach Rice and Silverman (1991) assumes that the functional predictors ad-

and Silverman (1991) by modifying the objective in (14) to

$$\widehat{\mathrm{Cov}}^2\big(\langle \widetilde{W}, \xi\rangle_{\mathbb{H}}, Y\big) - \sum_{j=}^{K}$$

where the smoothing parameters $\{\lambda_j\}_{j=1}^K$ control the trade-off between strategy of Silverman (1996), which replaces the standard orthonormality constraint with a weaker one based on a modified inner product that in-

mit an orthogonal expansion in the $\mathbb{L}^2$ sense. To avoid the orthogonal basis assumption of Rice and Silverman (1991), we adopt the corporates roughness. Accordingly, our estimation procedure at the l-th iteration, iteration index omitted and assuming the observations have been residualized in previous steps, solves the following optimization problem:

$$\hat{\xi} := \arg\max_{\xi \in \widetilde{\mathbb{H}}} \widehat{\mathrm{Cov}}^2\left(\langle \widetilde{W}, \xi \rangle_{\mathbb{H}}, Y\right) \; s.t. \; \|\xi$$

Here, $\hat{\xi} \in \widetilde{\mathbb{H}}$ is an taining these coefficients is equivalent to solving the maximization problem (18). The following proposition formulates this coef-

ordered pair expanded as:

$$\hat{\xi} = (\hat{\xi}_1(t), \ldots, \hat{\xi}_K(t),$$

where $\hat{\zeta}$ is the scalar part. This formulation maximizes the squared covariance over a class of smooth functions. Ob-

Let $(\mathbb{B}, \mathbb{B}'', \Theta, \mathbf{y})$ *denote the data given at the* l*-th iteration, as defined in* (12). *Recall from* (16) *that* $V = n^{-2}(\mathbb{B}\Theta^\top \mathbf{y})(\mathbb{B}\Theta^\top \mathbf{y})^\top$. *Let* $\Lambda \in \mathbb{R}^{(MK+p)\times(MK+p)}$ *be defined*

ficients obtaining procedure as a generalized Rayleigh quotient:

Proposition 2 (Regularized estimation of PLS component direction)**.**

as

$$\left(\hat{\gamma}_{11}, \ldots, \hat{\gamma}_{1M}, \ldots, \hat{\gamma}_{K1}, \ldots, \hat{\gamma}_{KM}\right) = \hat{\zeta}^\top, \quad (9)$$

The proof of Proposition 3.5 is provided in Appendix **??**. The constraint $\boldsymbol{\xi}^\top(\mathbb{B} + \Lambda\mathbb{B}'')\boldsymbol{\xi} = 1$ enforces the orthonormality of the estimated

as:

$$\Lambda = \mathrm{blkdiag}(\lambda_1 I, \ldots$$

Here, $0_{p \times p}$ *denotes the* $p \times p$ *zero matrix. The coefficients of the squared covariance maximizer defined in* (18)*, are obtained*

PLS component directions with respect to a modified inner product (see Section 4.2 for details). The smoothing parameter λ_k balances goodness of fit and smoothness in $\hat{\xi}_j$. Smaller λ_k yields components that better fit the data but risks overfitting; setting $\lambda_k = 0$ recovers the unregularized solution in Proposition 3.4. Larger λ_k enforces greater smoothness, and in the limit $\lambda_k \to \infty$, $\hat{\xi}_j(t)$ approaches a linear form $a+bt$. In practice, both $\{\lambda_k\}$ and the number of components L can be selected via cross-validation using a predictive criterion such as mean squared error.

one structure of the matrix V Proposition 3.6 derives a closed-form solution that requires only the solution of linear systems.

Proposition 3

Computation. The generalized eigenproblem presented in Proposition 3.5 may be computationally unstable in practice. However, by leveraging the rank-

(Closed–form solution). *Consider the optimization problem described in Proposition 3.5. Define the following quantities, which depend on the observed data but are not decision variables:*

$$\mathbf{u}_j := B\Theta_j^\top \mathbf{y} \in \mathbb{R}^M \quad \textit{for } j = 1, \ldots, K,$$

Let

$$q := \sum_{j=1}^{K} \mathbf{u}_j^\top (B+\lambda_j B'')^{-1} \mathbf{u}_j + \mathbf{v}^\top \mathbf{v}.$$

Then the unique (up to sign) solution to the regularized maximization prob-

the functional and scalar components separately, followed by normalization by a common factor. Although the unnormalized coefficients

lem is given in closed form by

$$\hat{\gamma}_j = \frac{1}{\sqrt{q}}(B+\lambda_j B'')^{-}$$

The proof of Proposition 3.6 is provided in Appendix **??**. The expressions above involve solving linear systems for enables the procedure to capture the correlation between the functional and scalar components of the PLS direction.

3.3.2 *Iterative*

are obtained independently, the normalization step couples the functional and scalar parts, allowing them to influence one another. This coupling dictors and responses. We first compute the individual PLS score:

$$\hat{\rho}_i^{[l]} := \langle \widetilde{W}_i^{[l]}, \hat{\xi}^{[l]} \rangle_{\mathbb{H}},$$

using the estimated PLS component direction $\widehat{\xi}^{[l]}$ obtained from Propo-

step 2: residualization via hybrid-on-scalar regression.

The l-th iteration's second step involves residualization of both pre-

1)-th iteration's responses and hybrid predictors, we regress the (l)-th iteration's responses and hybrid predictors on these PLS scores

sitions 3.5 and 3.6. Since $\widetilde{W}_i^{[l]}$ are assumed to have a sample mean of zero, these PLS scores will also have a sample mean of zero. To obtain the $(l+$

model:

$$\widetilde{W}_i^{[l]} = \widehat{\rho}_i^{[l]} \hat{\delta}^{[l]} + \epsilon_i,$$

where $\hat{\delta}^{[l]} \in \widetilde{\mathbb{H}}$ is the regression coefficient. In the same spirit as the PLS component direction estimation step, rather than

by least squares and then residualize. Specifically, the $(l+1)$-th predictor is computed as a residual of hybrid-on-scalar linear regression expansion approach to fit the entire hybrid object in one step. Therefore, our method is computationally efficient, and applicable

treating the hybrid object as a long vector of concatenated function evaluations at time points and scalar vectors, we employ a basis

β seems to soley rely on the smoothness of ξ. See Section 3.4.

for dense or irregular functional data. Consequently, $\delta^{[l]}$ is obtained by minimizing a least squares criterion:

$$\hat{\delta}^{[l]} := \arg\min_{\delta \in \widetilde{\mathbb{H}}} \sum_{i=1}^{n} \| \widetilde{W}_i^{[l]} - \hat{\rho}_i^{[l]} \, \delta \|_{\mathbb{H}}^2.$$

On the other hand, the $(l+1)$-th response is computed as a residual

impossible. Smoothness of

ply, analogous to scalar PLS.

lemma [. Closed-form solution] Let us denote $\hat{\boldsymbol{\rho}}^{[l]} := (\hat{\rho}_1^{[l]}, \dots, \hat{\rho}_n^{[l]})^\top$. The $(l+1)$-th iteration's predictors and re-

of a scalar-on-scalar linear regression model:

$$Y_i^{[l]} = \hat{\nu}^{[l]} \widehat{\rho}_i^{[l]} + \epsilon_i.$$

The following proposition demonstrates that this residualization step can be performed simsponses are computed as

$$\widetilde{W}_i^{[l+1]} := \widetilde{W}_i^{[l]} - \widehat{\rho}_i^{[l]} \hat{\delta}^{[}$$

and

$$Y_i^{[l+1]} = Y_i^{[l]} - \hat{\nu}^{[l]} \widehat{\rho}_i^{[l]},$$

3)

Proof of Lemma 3.7 is provided in Appendix **??**. Since $\widetilde{W}_i^{[l]}$ and $Y_i^{[l]}$ are assumed to have a sample mean of zero, their respective residuals, $\widetilde{W}_i^{[l+1]}$ and $Y_i^{[l+1]}$, also maintain a zero sample mean.

a linear combination of PLS directions:

lemma

. lemma:recursive Let use define $\widehat{\iota}^{[1]} := \widehat{\xi}^{[1]}$. For $l \geq 2$, we recursively define:

$$\widehat{\iota}^{[l]} = \widehat{\xi}^{[l]} - \sum_{u=1}^{l-1} \langle \widehat{\delta}^{[u]}, \widehat{\xi}^{[l]} \rangle_{\mathbb{H}} \widehat{\iota}^{[u]}.$$

Then

3.4 *Final step: estimating the hybrid regression coefficient*

The hybrid regression coefficient β in model (4) can be written as as:

$$Y_i = \sum_{l=1}^{L} \widehat{\nu}^{[l]} \langle W_i^{[l]}, \widehat{\xi}^{[l]} \rangle_{\mathbb{H}} + \epsilon_i = \langle W_i, \sum_{l=1}^{L}$$

which, given the uniqueness of β, leads to

$$\widehat{\beta} = \sum_{l=1}^{L} \widehat{\nu}^{[l]} \widehat{\iota}^{[l]} \sum_{l=1}^{L} \Big(\widehat{\nu}^{[l]} - \sum_{k=l+1}^{L} \widehat{\nu}^{[k]} \langle \widehat{\delta}$$

Thus the number of PLS components L to be estimated can be chosen by

we have:

$$\widehat{\rho}_i^{[l]} = \langle W_i^{[l]}, \widehat{\xi}^{[l]} \rangle_{\mathbb{H}} = \langle$$

Proof of Lemma 3.8 is provided in Appendix ??.

Next, (23) leads to the following model:

$$Y_i = \sum_{l=1}^{L} \widehat{\nu}^{[l]} \widehat{\rho}_i^{[l]} + \epsilon_i.$$

This model lets us to express Y_i

Algorithm 2 Hybrid partial least squares regression

1: **Initialize:** $(\mathbb{B}, \mathbb{B}'', \Theta^{[1]}, \mathbf{y}^{[1]})$ as the data objects after basis expansion, following Section 3.1.

2: $\widetilde{W}_1^{[1]}, \ldots, \widetilde{W}_n^{[1]}, Y_1^{[1]}, \ldots, Y_n^{[1]} \leftarrow$ standardized versions of $W_1, \ldots, W_n, Y_1, \ldots, Y_n$, following Section 3.2

3: **for** $l = 1, 2, \ldots, L$ **do**

4: **PLS direction and score estimation (Proposition 3.6):**

5: $\mathbf{u}^{[l]} \leftarrow$

cross-validation.

4. Properties of the hybrid PLS

This section provides the mathematical properties that support the algorithm suggested in Section

3. Section 4.1 shows that the core optimization problem for the Partial Least Squares (PLS) direction estimation step, presented in Proposition

3.4, is well-defined under mild conditions. Section 4.2 demonstrates that our algorithm preserves the core properties of PLS, namely the orthonormality gorithm 2 and establish its key mathematical properties, namely the orthogonality of the PLS directions and scores, as well as con-

of the derived directions and the orthogonality of the scores.

4.1 *Tucker's Criterion*

We now derive the population analogue of Alvergence to the true response. At the population level, where the random objects are fully observable, we may drop the sample index i and simply work with Y and $W = (X_1, \dots, X_K, Z_1, \dots, Z_p)$. Throughout this section we assume $\mathbb{E}[Y] = 0 \in \mathbb{R}$ and $\mathbb{E}[W] = 0 \in \mathbb{H}$. The core optimization problem in PLS is formulated in terms of maximizing the squared cross-covariance. At the population level, this quantity is naturally characterized by cross-covariance operators, which we define next. For notational simplicity, we suppress the iteration index l, with the understanding that for $l =$

$\mathbb{E}[W \otimes W]$, so that for $u, v \in \mathbb{H}$, we have

$$\Sigma_W u = \mathbb{E}[\langle W, u\rangle_{\mathbb{H}} W], \quad \langle \Sigma_W u, v\rangle_{\mathbb{H}} =$$

The cross-covariances between the response Y and the predictors are

$$\sigma_{YX} := \big(\mathbb{E}[YX_1], \ldots, \mathbb{E}[YX_K]\big) \in \mathcal{F}^K$$

The cross-covariance between Y and

1 the predictors correspond to the original variables, while for $l \geq 2$ they represent the residualized versions.

definition

•

def:cross$_cov_terms The$

the hybrid predictor W is then

$$\Sigma_{YW} := \mathbb{E}[YW] = (c$$

At the population level, Algorithm 2 without regularizaiton can be expressed as follows.

Let $W^{[1]} := W$ and $Y^{[1]} := Y$. For any integer $l \geq 2$, to be the residuals of the

where

and

which

is attained by the leading eigenfunction of the operator $\mathcal{U}^{[l-1]}$, defined analogously to Lemma 4.6 using $W^{[l-1]}$ and $Y^{[l-1]}$, provided the conditions in Lemma 4.4 are satisfied.

definition

[. Population PLS] The population analogue of the PLS optimization problem (2) in the hybrid space is formulated as the constrained gen-

Proposition 4 (Properties of population PLS)**.** *The solution of optimization problem* (24) *of Definition* 4.2 *is equivalent to that*

eralized eigenproblem:

$$\xi^{[l]} = \arg\max_{h \in \mathbb{H}} \mathrm{Cov}^2(\langle$$

4)

As in scalar case, this problem can be solved by an iterative algorithm, and yields orthogonal scores:

is provided in Appendix **??**.

We show that problem (24) and Algorithm 3 are well-defined under mild conditions by establishing that line

of population NIPALS (Algorithm 3). For every $s \geq 1$ *and every* $j \leq s-1$*, we have* $\mathbb{E}[\rho^{[j]}\rho^{[s]}] = 0$ *and* $\mathbb{E}[Y^{[s]}\rho^{[j]}] = 0$.

Proof of Proposition 4.3

3 of Algorithm 3 is well-defined under mild conditions. For clarity, we define the relevant operators while omitting the iter-

Algorithm 3 Population NIPALS

1: $(W^{[1]}, Y^{[1]} \leftarrow (W, Y)$
2: **for** $l = 1, 2, \ldots, L$ **do**
3: $\xi^{[l]} \leftarrow \arg\max_{h \in \mathbb{H}} \mathrm{Cov}^2$ 1, ▷ PLS direction
4: $\rho^{[l]} \leftarrow \langle \xi^{[l]}, W^{[l]} \rangle$ ▷ PLS score
5: $\delta^{[l]} \leftarrow \frac{1}{\mathbb{E}[(\rho^{[l]})^2]} \mathbb{E}[W^{[l]} \rho^{[l]}$ ▷ Linear regression of $W^{[l]}$ on $\rho^{[l]}$
6: $\nu^{[l]} \leftarrow \frac{1}{\mathbb{E}[(\rho^{[l]})^2]} \mathbb{E}[Y^{[l]} \rho^{[l]}]$ ▷ Linear regression of $Y^{[l]}$ on $\rho^{[l]}$
7: $W^{[l+1]} \leftarrow W^{[l]} - \rho^{[l]} \delta^{[l]}$ ▷ Residualized predictor
8: $Y^{[l]} \leftarrow$

are bounded in the sense that there exist finite constants Q_1 and Q_2 such that

a- tion index for the residualized predictors and responses.

lemma

$$\max_{k=1,\ldots,K} \sup_{t\in[0,1]} \mathbb{E}[YX_k](t)^2 < Q_1 \quad \text{and}$$

.

the operator $\mathcal{C}_{YW}$ is a compact operator.

lemma:cross$_cov_function$

C_{YW} such that for any $h = (f_1, \ldots, f_K, \mathbf{v}) \in \mathbb{H}$,

$$\mathcal{C}_{YW}h := \mathbb{E}\left[\langle W, h\rangle_{\mathbb{H}} Y\right.$$

If σ_{YX} and σ_{YZ}

tor of $\mathcal{C}_{YW}$. That is, $\mathcal{C}_{WY} = \mathcal{C}^*_{YW}$.

The proof of Lemma 4.5 is provided in Appendix **??**.

lemma

[. Composite cross-covariance operator] Define $\mathcal{U} :=$

Proof of Lemma 4.4 is provided in Appendix **??**.

lemma

.

lemma:cross$_{c}ov_{o}perat$ $\mathbb{E}[Y \otimes W]$, such that for any $d \in \mathbb{R}$,

$$\mathcal{C}_{WY} d := \mathbb{E}\big[\langle Y, d\rangle W\big]$$

Then $\mathcal{C}_{WY}$ is an adjoint operasemidefinite operator. Under the conditions of Lemma 4.4, $\mathcal{U}$ is a compact operator.

The proof of Lemma 4.6 is provided in Appendix ??.

$\mathcal{C}_{WY} \circ \mathcal{C}_{YW} : \mathbb{H} \to \mathbb{H}$ as an operator which performs the following mapping:

$$\mathcal{U}h = \mathcal{C}_{WY}(\mathcal{C}_{YW}h) =$$

In other words, $\mathcal{U} = \Sigma_{YW} \otimes \Sigma_{YW}$. Then $\mathcal{U}$ is a self-adjoint and positive-

By the Hilbert-Schmidt theorem (e.g., Theorem 4.2.4 in Hsing and Eubank, 2015), Lemma 4.6 guarantees the existence of a complete orthonormal system of eigenfunctions $\{\xi_{(u)}\}_{u\in\mathbb{N}}$ of $\mathcal{U}$ in $\mathbb{H}$ such that $\mathcal{U}\xi_{(u)} = \kappa_{(u)}\xi_{(u)}$, where $\{\kappa_{(u)}\}_{u\in\mathbb{N}}$ are the corresponding sequence of eigenvalues that goes to zero as $u \to$

∞, that is, $\kappa_{(1)} \geq \kappa_{(2)} \geq \cdots \geq 0$.

The following theorem presents population-level PLS and Tucker's Criterion for our scalar-on-hybrid regression model and ensures that line 3 of Algorithm 3 is well-defined under mild conditions.

theoremPLS and Tucker's Criterion

[. Tucker's criterion] Under the conditions of Lemma 4.4, the orem 4.7 is provided in Appendix **??**. Finally, the following theorem states that the fitted values from hybrid PLS converge to those of or-

constrained maximum

$$\max_{\|\xi\|_{\mathbb{H}}=1} \mathrm{Cov}^2\left(\langle W, \xi\rangle_{\mathbb{H}},\right.$$

is attained by the eigenfunction associated with the largest eigenvalue of the operator $\mathcal{U}$.

The proof of The-of a full linear model when a sufficient number of components is used.

theorem

[. L^2 Convergence of Hybrid PLS at Popu- la-

dinary linear regression in the mean-squared sense. This shows that hybrid PLS, as a dimension-reduction technique, can effectively capture the predictive power sense:

$$\lim_{L\to\infty} \mathbb{E}\left[\|Y_{\mathrm{PLS},L} - Y\|_2^2\right] = 0$$

The proof of Theorem 4.8 is provided in Appendix **??**.

4.2 *Geometric properties*

A fundamental property of partial least squares, as seem

tion Level] Let $Y_{\mathrm{PLS},L} = \sum_{l=1}^{L} \nu^{[l]} \rho^{[l]}$ be the Hybrid PLS fitted value with L components. Then, the Hybrid PLS fitted value converges to Y in the mean-squared onal. Our regularized estimates preserve this property, with respect to a modified inner product that incorporates the roughness

in Definition 4.2 and Proposition 4.3, is that between iterations, its derived directions are orthonormal and PLS scores are orthogpenalty, defines as follows:

definition [. Roughness-sensitive inner product. Given two hybrid predictors $W_1 = (X_{11}, \ldots, X_{1K}, \mathbf{Z}_1)$ and $W_2 = (X_{21}, \ldots, X_{2K}, \mathbf{Z}_2)$, both elements of $\mathbb{H}$ as defined in Definition 3.3, and a roughness penalty matrix $\Lambda = \text{blkdiag}(\lambda_1 I_M, \ldots, \lambda_K I_M, 0_{p\times p})$, the roughness-sensitive inner product between W_1 and W_2 is defined as:

$$\langle W_1, W_2 \rangle_{\mathbb{H},\Lambda} := \sum_{k=1}^{K} \int_0^1 X_{1k}(t) X_{2k}(t)$$

thonormal.

Proposition 5 **(Orthonormality of estimated PLS component directions).** *The PLS component directions* $\widehat{\xi}^{[1]}, \widehat{\xi}^{[2]}, \ldots, \widehat{\xi}^{[L]}$, *estimated via Proposition*

Based on this inner product, the following proposition states that the PLS component directions estimated from Proposition 3.5 are or-sition 4.10 is provided in Appendix ??. The next proposition states that the vectors of estimated PLS scores for different iteration num-

3.5 with a roughness penalty matrix $\Lambda = \mathrm{blkdiag}(\lambda_1 I_M, \ldots, \lambda_P$ *are mutually orthonormal with respect to the inner product* $\langle \cdot, \cdot \rangle_{\mathbb{H},\Lambda}$. *That is,*

$$\langle \widehat{\xi}^{[l_1]}, \widehat{\xi}^{[l_2]} \rangle_{\mathbb{H},\Lambda} = \mathbb{1}(l_1$$

The proof of Propo-

estimated PLS scores ($l = 1, \ldots, L$) of n observations. The vectors $\widehat{\boldsymbol{\rho}}^{[1]}, \widehat{\boldsymbol{\rho}}^{[2]}, \ldots, \widehat{\boldsymbol{\rho}}^{[L]}$ are mutually orthogonal in the sense that

$$\widehat{\boldsymbol{\rho}}^{[l_1]\top}\widehat{\boldsymbol{\rho}}^{[l_2]} = 0 \quad \text{for} \quad l_1, l_2 \in \{1, \ldots, L$$

The proof of Propo-

bers are mutually orthogonal.

Proposition 6. *Recall from Lemma 3.7 that $\widehat{\boldsymbol{\rho}}^{[l]}$ denote the n-dimensional vector whose elements consist of the l-th* tures — specifically, dependencies among functional predictors, among scalar predictors, and between scalar and functional predictors —

5.1 *Synthetic data*

To gen-

sition 4.11 is provided in Appendix ??.

5. Simulation studies

To evaluate the superiority of our method under complex dependency struc-

$3\times$ 100 ma- trix, where each row rep- re- sents the ith ob- ser- va- tion of a func- tional pre- dic- tor with 100 eval- u- a- tion points. This method gives us pre-

er- ate func- tional pre- dic- tors, we use a matrix- normal dis- tri- bu- tion, also known as a Kronecker- separable co- vari- ance model. For each of $i =$ $1, \ldots, n,$ we cre- ate a insight behind Kronecker separability is that if $\mathbf{Y} \sim \mathcal{MN}(\mathbf{M}, \mathbf{R}, \mathbf{C})$, then its vectorized form follows a multivariate normal distribution:

cise control over cross-modal (between-row) dependence and smoothness of the curves (between-column dependence).

$$\mathbf{X} \sim \mathcal{MN}_{m\times n}(\mathbf{M}; \mathbf{R}$$

and its log-density is given by

$$\log p(\mathbf{X} \mid \mathbf{M}, \mathbf{R}, \mathbf{C}) =$$

The key

work leveraging the matrix-normal distribution, an approach previously suggested by our advisor. The goal was to create a dataset with a spe-

$\mathrm{vec}(\mathbf{Y}) \sim \mathcal{N}(\mathrm{vec}(\mathbf{M}), \mathbf{C} \otimes \mathbf{R})$, where $\otimes$ denotes the Kronecker product and vec is the vectorization operator. We developed a data generation frame-

To introduce a structured dependence within each row, the row covariance matrix was set to mimic an AR(1) process, which is a special case of

cific, controllable correlation structure between functional and scalar predictors. corresponds to the number of evaluation points for each curve.

a Gaussian process. This ensures that each row exhibits the smooth, curve-like characteristics of a functional observation. The first two rows of each matrix were treated as functional data. The third row was converted into a set of scalar predictors. This was accomplished by partitioning the third tors, we would partition the row into (1,3), (2,4), and (3,4), yielding the scalar values 2.0,3.0, and 3.5. Building on the functional graphical model simulation

row into blocks and then averaging the values within each block. For instance, if the third row was (1,3,2,4,3,4) and we wanted to generate three scalar predicas Gaussian processes using a truncated Karhunen-Loève expansion, where the eigenbasis consists of Fourier basis functions with a fixed number of ba-

of Zhu et al. (2016), we generate a mixed graphical model with five nodes, described as follows:

- Nodes F1 and F2: Two functional predictors modeled the functional predictors, these scalar predictors are modeled directly without basis expansion.

To capture dependencies among predictors, we

sis functions, $M = 9$.

- Nodes S1, S2, and S3: Three scalar predictors, each following an s-dimensional multivariate normal distribution. Unlike connected graph. In the Gaussian process framework, the precision matrix $\mathbf{R}_0^{-1}$ encodes conditional independence relationships, while its inverse,

introduce a graph structure that governs their conditional correlations. We consider two types of graph structures: a weakly connected graph and a strongly tions between FPC scores and scalar predictor values. Each block (i, j) of $\mathbf{R}$ is given by $(R_0)_{ij}\mathbf{I}_{M_i,M_j}$, where $\mathbf{I}_{M_i,M_j}$ is a rectangular identity matrix. Here, $M_i =$

$\mathbf{R}_0$, represents marginal covariances. This structure extends to a blockwise correlation matrix $\mathbf{R} \in \mathbb{R}^{(2M+3s)\times(2M+3s)}$, where off-diagonal blocks represent correlations tor, we assgin M reference eigenvalues (or FPC score) drawn independently from gamma distributions with decreasing means. These reference eigen-

9 if node i corresponds to a functional predictor and $M_j =$ s if node j corresponds to a scalar predictor. For each functional predic-

are fixed over all samples and all independent repetition of the experiment. We draw it one randomly just to ensure the differ-

values will later be multiplies by randomly multiplier drawn from correlated multivaraite Normal distribution. These reference eigenvalues tivariate normal data from the covariance matrix $\mathbf{R}$. The last $3s$ components are assigned as scalar predictors, while the first $2M$ components,

entiation between the two functional predictors. we independently draw for each functional predictor from . We then sample zero-mean muling a Fourier basis. Finally, we introduce structured variability by adding a common mean function, defined as a scaled and shifted sine function. To vi-

scaled by their corresponding eigenvalues, serve as FPC scores. These scores are then expanded into functional data, evaluated over 100 equally spaced points on $[0, 1]$ us-graph structures, illustrating the distinct dependency patterns induced by the precision matrices.

5.2 *Simulation shown in february 2025 meeting*

To ef-

sualize the generated data, Figure **??** plots functional predictor realizations and heatmaps of sample correlation matrices for both fectively manage the dependencies among functional predictors, scalar covariates, and between functional and scalar predictors, we utilize a Gaussian Markov random field (GMRF) within the framework of Gaussian undirected graphical models. In a GMRF, the off-diagonal elements of the precision matrix we adopt the simulation setup proposed by Kolar et al. (2014), which focuses on mixed attribute Gaussian graphical models. We construct a graph con-

capture the conditional correlations between the corresponding components. Given that our setting involves both functional and scalar covariates, tion process for each node.

Functional Predictors. We consider two functional predictors that share the same precision matrix. Denoted as $\boldsymbol{\Theta} := (\theta_{t\tilde{t}}) \in \mathbb{R}^{p\times p}$, this

sisting of two functional predictor nodes and three vector predictor nodes. Below, we sequentially describe the graph structure and the generaa smooth functional trajectory. More formally, each functional predictor is evaluated at p equally spaced points on $[0, 1]$, with values defined recursively

precision matrix follows an AR(1) structure with a white noise variance of 0.1^2 and an autoregressive coefficient of $=$ 0.95, ensuring as: $X_i^{(k)}(0)$ $t = 0$ $1, p$ $X_i^{(k)}($

In this setting, the precision matrix $\boldsymbol{\Theta}$ exhibits a tridiagonal Toeplitz structure, where the diagonal entries are given by:

$$\theta_{tt} = \begin{cases} \dfrac{1}{\sigma^2(1-0.95^2)}, & t = 1, p \\ \dfrac{1+0.95^2}{\sigma^2(1-0.95^2)}, & 2 \leq t \leq p-1 \end{cases}$$

The off-diagonal entries are:

$$\theta_{t,t+1} = \theta_{t+1,t} = -\frac{\rho}{\sigma^2(1-\rho^2)}, \quad 1 \leq$$

Finally, the functional predictors are smoothed using a B-spline basis with 15 basis functions, as described in Section 3.3.1.

Vector Predictors. We consider s vector predictors, each following a d-dimensional zero-mean Gaussian distribution with a shared precision matrix. This precision matrix, denoted as $\Gamma := (\gamma_{ij}) \in \mathbb{R}^{d \times d}$, follows a Toeplitz structure with exponentially decaying entries, to the first, as illustrated in Figure ??. An edge in the graph indicates that the connected nodes remain correlated when the values of all other

given by:

$$\gamma_{ij} := 0.5^{|i-j|}.$$

Graph Structure. We arrange five nodes in a chain structure, where each node follows a sequential order, and the last node connects back nodes are fixed. The resulting marginal dependence structure is significantly more complex than the chain structure itself. We designate nodes F_1 and F_2 as functional predictors and nodes V_1, V_2, and V_3 as vector predictors. To introduce conditional dependencies between the components, we set the off-diagonal blocks of the precision matrix to be $0.5\mathbf{1}$ if the corresponding components are connected in the graph, and zero otherwise. Here, $\mathbf{1}$ denotes a matrix of appropriate dimensions where all elements are equal to 1. Overall, we generate a $2p+3d$-

observation drawn from this multivariate Gaussian distribution, we process

The

regression coefficients for the first functional predic-

dimensional multivariate Gaussian distribution with mean zero and a precision matrix Ω, structured as follows:

$$\Omega = \begin{pmatrix} \Omega_F & 0.5\mathbf{1} \\ 0.5\mathbf{1} & \Omega_F \\ 0 & 0.5\mathbf{1} \\ 0 & 0 \\ 0.5\mathbf{1} & 0 \end{pmatrix}$$

For

each

tor, β_{F_1}, are drawn from a multivariate normal distribution $N(0, 5\mathbf{I}_p)$, with a fixed random seed. The coefficients for the second functional predictor, β_{F_2}, are drawn independently from the same distribution and are also smoothed using a B-spline basis with 10 basis functions. For the vector covariates, the regression coefficients are sampled from $N(0, I_d)$. After calculating the inner product of the covariates and their corresponding coefficients, independent Gaussian noise from $N(0, 0.1^2)$ is added to the generated responses to simulate measurement noise.

predictors Happ and Greven (2018) and scalar PCA and run OLS on the PC scores.

We compare our method with these baseline methods using $p =$ 100. We consider sce-

The baseline methods are:

- Penalized functional regression (pfr) Goldsmith et al. (2011)
- Principal component regression (fpcr): run both of PCA for multple functional

narios with $d = 1, 2, 3, 4, 5$ and $n = 100, 200, 300, 400.$ For each scenario, we use 70% of the data for training and evaluate prediction performance on the remaining 30% test set, using the root prediction mean squared error as the evaluation metric. For our method and PCR, the maximum number of strate that our method consistently outperforms the baseline methods across all scenarios.

6. Data Application

Renal study data. We applied our proposed hy-

components are set as 20. The number of components is chosen by 5-fold cross validation. The results, summarized in Table **??**, demon-

113 subjects, including: (i) baseline renogram curves; (ii) post-furosemide renogram curves; (iii) ordinal ratings of kidney obstruction status (non-obstructed, equivocal,

brid functional PLS regression, along with other regression methods, to the Emory renal study data. The study collected data on 226 kidneys (left and right) from or obstructed) independently assessed by three nuclear medicine experts; (iv) eight kidney-level pharmacokinetic variables derived from radionuclide imaging; and (v) two subject-level variables (age and gender). The subjects had a mean age of 57.8 years (SD = 15.5; range = 18–83), with 54 males (48%) and 59 females (52%).

The two renogram curves, (i) and (ii), were treated as functional predictors and smoothed using a B-spline basis of order 15. The remaining variables, ex-

The three experts unanimously classified 153 kidneys as non-obstructed, 5 as equivocal, and 40 as obstructed, while 28 kidneys had discrepant ratings.

not entirely redundant, as they may contain additional useful information. Finally, the diagnoses provided by the three experts were av-

cluding the diagnosis, were treated as scalar predictors. Given the nature of these variables, we assume they are correlated with the renogram curves but prediction perforamnec by root mean squared error on the test data, normalized by the range of the test data response.

6.1 ...

eraged and transformed using a min-max logit transformation. We splitted the data into 70% of training data and 30% of testing data, and evalauted the

References

Aguilera, A. M., Aguilera-Morillo, M. C., and Preda, C. (2016). Penalized versions of functional PLS regression. *Chemometrics and Intelligent Laboratory Systems* **154,** 80–92.

Aguilera, A. M., Escabias, M., Preda, C., and Saporta, G. (2010). Using basis expansions for estimating functional PLS regression:

bust functional partial least squares for scalar-on-multiple-function regression. *Journal of Chemometrics* **36,** e3394.

Beyaztas, U. and Shang, H. L. (2020). On function-on-function re-

Applications with chemometric data. *Chemometrics and Intelligent Laboratory Systems* **104,** 289–305.

Beyaztas, U. and Lin Shang, H. (2022). A rotional linear regression. *The Annals of Statistics* **34,** 2159–2179.

Cardot, H., Ferraty, F., and Sarda, P. (2003). Spline estimators for the functional lin-

gression: Partial least squares approach. *Environmental and Ecological Statistics* **27,** 95–114.

Cai, T. T. and Hall, P. (2006). Prediction in funcear model. *Statistica Sinica* **13,** 571–591.

Chang, C., Jang, J. H., Manatunga, A., Taylor, A. T., and Long, Q. (2020). A bayesian latent class model to predict kidney obstruction in the absence of gold standard. *Journal of the American Statistical Association* **115,** 1645–1663.

Delaigle, A. and Hall,

González-Manteiga, W. (2017). Functional principal component regression and functional partial least-squares regression: An overview and a comparative study. *In-*

P. (2012). Methodology and theory for partial least squares applied to functional data. *The Annals of Statistics* **40,** 322–352.

Febrero-Bande, M., Galeano, P., and

tion with Confounding. In *Proceedings of The 35th Uncertainty in Artificial Intelligence Conference*, pages 378–388. PMLR.

Goldsmith, J., Bobb,

ternational Statistical Review **85,** 61–83.

Geng, S., Kolar, M., and Koyejo, O. (2020). Joint Nonparametric Precision Matrix Estima-

Hall, P. and Horowitz, J. L. (2007). Methodology and convergence rates for functional linear regression. *The Annals of Statistics* **35,** 70–91.

Happ, C. and

J., Crainiceanu, C. M., Caffo, B., and Reich, D. (2011). Penalized functional regression. *Journal of Computational and Graphical Statistics* **20,** 830–851.

American Statistical Association **113,** 649–659.

Hsing, T. and Eubank, R. (2015). *Theoretical Foundations of Functional Data*

Greven, S. (2018). Multivariate functional principal component analysis for data observed on different (dimensional) domains. *Journal of the* sis of hybrid functional and vector data. *Statistics in Medicine* **40,** 5152–5173.

Kolar, M., Liu, H., and Xing, E. P. (2014). Graph estimation from multi-attribute data.

Analysis, with an Introduction to Linear Operators. Wiley, 1st edition edition.

Jang, J. H. (2021). Principal component analy-

& Data Analysis **48,** 149–158.

Reiss, P. T. and Ogden, R. T. (2007). Functional principal component regression and functional partial least squares. *Jour-*

Journal of Machine Learning Research **15,** 1713–1750.

Preda, C. and Saporta, G. (2005). PLS regression on a stochastic process. *Computational Statistics* variance structure nonparametrically when the data are curves. *Journal of the Royal Statistical Society: Series B* **53,** 233–243.

Saricam, S.,

nal of the American Statistical Association **102,** 984–996.

Rice, J. A. and Silverman, B. W. (1991). Estimating the mean and co-

Silverman, B. W. (1996). Smoothed functional principal components by choice of norm. *The Annals of Statistics* **24,** 1–24.

Wang, Y. (2018). *Partial least squares meth-*

Beyaztas, U., Asikgil, B., and Shang, H. L. (2022). On partial least-squares estimation in scalar-on-function regression models. *Journal of Chemometrics* **36,** e3452.

in functional linear regression. *Journal of Computational and Graphical Statistics* **21,** 600–617.

Zhu, H., Strawn, N., and Dunson, D. B. (2016). Bayesian

ods for functional regression models. PhD thesis, University of North Carolina at Chapel Hill.

Zhao, Y., Ogden, R. T., and Reiss, P. T. (2012). Wavelet-based LASSO

Tables and Figures

Table 1: Example Table Caption.

Col1	Col2	Col3
a	b	c

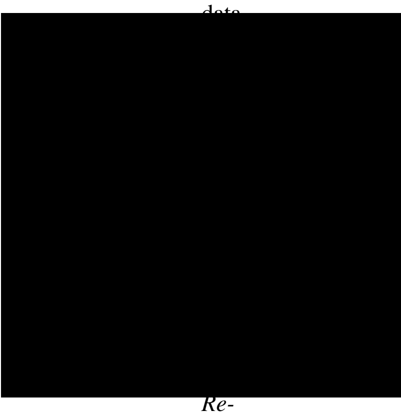

Figure 1: Example Figure Caption.

graphical models for multivariate functional data [redacted] *Research* **17,** 1–27.