Biometrika (2016), **xx**, x, *pp*. 1–7
© 2007 Biometrika Trust
Printed in Great Britain

Supplementary material: A pairwise interaction model for multivariate functional and longitudinal data

By JENG-MIN CHIOU

Institute of Statistical Science, Academia Sinica, 128 Section 2 Academia Road, Nankang, Taipei 11529, Taiwan

jmchiou@stat.sinica.edu.tw

and HANS-GEORG MÜLLER

Department of Statistics, University of California, Davis, One Shields Avenue, Davis, California 95616, U.S.A.

hgmueller@ucdavis.edu

S.1. Additional details on estimation

The data model for sparsely sampled functional data in (11) is

$$W_{ijl} = X_{ij}(T_{ijl}) + \varepsilon_{ijl}, \quad (i = 1, \ldots, n;\ j = 1, \ldots, p;\ l = 1, \ldots, N_{ij}),$$

with independent zero mean errors ε_{ijl}, where measurements W_{ijl} are made on a sparse grid of measurements T_{ijl} for each realization i of the jth component process X_{ij}.

Eestimators $\hat{\mu}_j$ of mean functions μ_j, $\hat{C}_j$ of auto-covariance functions C_j and $\hat{C}_{ij}$ of cross-covariance functions C_{ij} are obtained by pooling the data across subjects, thus borrowing information across the sample, as described in previous work (Yao et al., 2005; Müller & Yao, 2010; Yang et al. 2011). Specifically, we apply local linear fitting, using a univariate density function κ_1 as kernel function, with the Epanechnikov kernel as default. Pooling observations across subjects, the goal is to minimize

$$\sum_{i=1}^{n} \sum_{l=1}^{N_{ij}} \kappa_1\left(\frac{T_{ijl} - t}{h_{\mu_j}}\right)\left[W_{ijl} - \{\alpha_0 + \alpha_1(T_{ijl} - t)\}\right]^2, \tag{S1}$$

for each t with respect to α_0, α_1. This yields the estimators $\hat{\mu}_j(t) = \alpha_0^*$. Here α_0^* is the minimizer at t and h_{μ_j} is the bandwidth employed by the local linear smoother.

Analogously, specifying a bandwidth h_{C_j} and a kernel κ_2 that corresponds to a bivariate distribution, for which the default is the product of two one-dimensional Epanechnikov kernels, one obtains the estimators $\tilde{C}_j$ by local linear surface smoothing. Defining raw auto-covariances $G_{ij}(T_{ijl}, T_{ijl'}) = \{W_{ijl} - \hat{\mu}(T_{ijl})\}\{W_{ijl'} - \hat{\mu}(T_{ijl'})\}$, pooling the raw auto-covariances obtained from the repeated measurements, while omitting the diagonal elements where $m = m'$ to remove the variance effects of the measurement errors, the goal is to minimize

$$\sum_{i=1}^{n} \sum_{1 \le l \ne l' \le N_{ij}} \kappa_2\left(\frac{T_{ijl} - t}{h_{C_j}}, \frac{T_{ijl'} - s}{h_{C_j}}\right)\Big[G_{ij}(T_{ijl}, T_{ijl'}) - \{\alpha_0 + \alpha_1(T_{ijl} - s) + \alpha_2(T_{ijl'} - t)\}\Big]^2, \quad \text{(S2)}$$

for fixed (s, t) with respect to $\alpha_0, \alpha_1, \alpha_2$. One then sets $\tilde{C}_j(s, t) = \alpha_0^*$ and obtains $\hat{C}_{jk}$ as described in step 4 of the estimation procedure in §3.

Finally, the same idea is applied to obtain the estimators $\hat{C}_{ij}$ of the cross-covariance surfaces. Here we pool raw cross-covariances $G_{ijk}(T_{ijl}, T_{ikl'}) = \{W_{ijl} - \hat{\mu}(T_{ijl})\}\{W_{ikl'} - \hat{\mu}(T_{ikl'})\}$ and obtain the minimizers

$$\sum_{i=1}^{n} \sum_{1 \le l \le N_{ij}, 1 \le l' \le N_{ik}} \kappa_2\left(\frac{T_{ijl} - t}{h_{C_{jk}}}, \frac{T_{ikl'} - s}{h_{C_{jk}}}\right)\Big[G_{ijk}(T_{ijl}, T_{ikl'}) - \{\alpha_0 + \alpha_1(T_{ijl} - s) + \alpha_2(T_{ikl'} - t)\}\Big]^2, \quad \text{(S3)}$$

for fixed (s, t) with respect to $\alpha_0, \alpha_1, \alpha_2$. Denoting the minimizers as $\alpha_0^*, \alpha_1^*, \alpha_2^*$, one then sets $\hat{C}_{jk}(s, t) = \alpha_0^*$.

S.2. Proofs and additional results

Proof of theorem 1. The proof is essentially the same as that given in Müller & Yao (2010), making use of Lemma 1 of that paper for general locally weighted estimators, and for detailed derivations we refer to that paper. Specifically for the result for $\hat{C}_j$, one first proves the corresponding convergence for $\tilde{C}_j$, as defined in (S2), whence the stated result for $\hat{C}_j$ follows from Theorem 1 of Hall et al. (2008).

Specific rates of convergence. Perusing relations (19), if we assume that $\lambda_m = m^{-\alpha}$ for $\alpha \ge 2$, we find $\delta_m \le \alpha m^{-\alpha-1}$ and if the optimal bandwidths given right before (16) are used, one has $\|\hat{\phi}_m - \phi_m\| = O_p(n^{-2/3} m^{\alpha+1})$. Then if we consider all eigenfunction estimators for $m =$

$1, \ldots, M = M(n)$ simultaneously, we obtain consistency with a polynomial rate of convergence if $M(n) \leq n^{2/9}$.

If instead we consider exponentially declining eigenvalues $\lambda_m = e^{-\alpha m}$, we have $\|\hat{\phi}_m - \phi_m\| = O_p(n^{-2/3}e^{\alpha m})$, whence polynomial rates of convergence for all $m = 1, \ldots, M = M(n)$ can be achieved if $\alpha < 2/3$ and $M(n) \leq \log n$.

Consistency of imputed trajectories. Theorem 1 also implies consistency of estimated conditional expectations $\hat{E}\{X_{ij}(t) \mid W_{ij1}, \ldots, W_{ijN_{ij}}\}$, which we refer to as imputed trajectories, towards the true conditional expectations $E\{X_{ij}(t) \mid W_{ij1}, \ldots, W_{ijN_{ij}}\}$. These conditional expectations are assumed to be linear in the conditioning variables. This is automatically the case in the Gaussian situations and in non-Gaussian situations this assumption will lead to the targeting of best linear predictors. Here the W_{ijl} are the sparse measurements available for the ith subject. Explicit formulas for these conditional expectations and their estimators are in Yao et al. (2005) and Müller & Yao (2010) and these formulas can be directly applied in the current setting. For estimating the conditional expectations, the error variance, covariances and eigenvalues that determine the true best linear predictors are replaced by the corresponding estimators. These estimators are then used to obtain the fitted trajectories conditional on the observations for a given subject.

Proof of theorem 2. Since $\tilde{C}_j(t,t)$ is bounded below by $(\log n)^{-1}$ and $C_j(t,t)$ is uniformly bounded below, the denominator yields a $\log(n)$ factor that goes in front and in addition a factor of the type $f_{1jn} = |\int_{\mathcal{T}} \hat{C}_j(t,t)dt - \int_{\mathcal{T}} C_j(t,t)dt|$, observing that the convergence behavior of $\tilde{C}_j$ is the same as that of $\hat{C}_j$. The numerator difference for (21) corresponds to a factor $f_{2jn} = |\int_{\mathcal{T}} \hat{H}_j(t,t)dt - \int_{\mathcal{T}} H_j(t,t)dt|$, while that for (22) corresponds to $f_{3jkn} = |\int_{\mathcal{T}} \hat{G}_{jk}(t,t)dt - \int_{\mathcal{T}} G_{jk}(t,t)dt|$. We observe that

$$f_{1n} = |\int_{\mathcal{T}} \sum_{m=1}^{M(n)} \hat{\lambda}_{Hjm}\hat{\phi}_{Hjm}(t)^2 dt - \int_{\mathcal{T}} \sum_{m=1}^{\infty} \lambda_{Hjm}\phi_{Hjm}(t)^2 dt| = \sum_{m=1}^{M(n)} |\hat{\lambda}_{Hjm} - \lambda_{Hjm}| + \sum_{m=M(n)+1}^{\infty} \lambda_{Hjm},$$

where we have used the orthonormality property of underlying and estimated eigenfunctions. Analogous expressions hold for f_{2jn} and f_{3jkn}. Applying (20), one finds

$$\sum_{m=1}^{M(n)} |\hat{\lambda}_{Hjm} - \lambda_{Hjm}| = O_p\{M(n)\|\hat{H}_j - H_j\|\} = O_p\{M(n)\eta_n\},$$

using (16). Again, the arguments for the other terms are analogous. Putting the factors together yields the rates as stated.

S.3. ADDITIONAL SIMULATION RESULTS

We provide more detailed results for the proposed functional pairwise interaction model, investigating the effect of the sample sizes $n = 100, 200, 400$ on the estimators of the fractions of variance $R^2_{j\cdot k}$ and on the estimators of the varying coefficient functions β_{jk} that yield the representation of Z_{jk} in (17).

Figures S4 and S2 for the sparse and the dense design, respectively, show the effect of increasing sample sizes on the estimators of the various fractions of variance given in (11) and (12). In both the sparse and the dense designs, the interquartile ranges of the boxplots decrease as the sample sizes increase. All the median values of the boxplots are close to the true values of $R^2_{j\cdot k}$, with reasonable ranges of dispersion. The estimation of the fractions of variance $R^2_{j\cdot k}$ appears to work well in the simulation.

Figure S3 displays the samples of the estimators (25) of the varying coefficient functions $\beta_{jk1}(t)$ that are defined in (24). These results demonstrate how the variance of these estimators decreases as the sample size increases.

The performance of the estimated regression coefficient functions can be further quantified by Average Integrated Squared Error, $\text{AISE}(j) = \sum_{k\neq j} n^{-1}_{sim} \sum_{sim=1}^{n_{sim}} \int \left\{\beta_{jk}(t) - \hat{\beta}^{sim}_{jk}(t)\right\}^2 dt$. The simulation results for AISE are displayed in Table S1. As expected, the AISE of the estimated functions $\beta_{jk1}(t)$ decreases as the sample size increases and is smaller for the dense design in comparison with the sparse designs.

Table S1: *Average mean integrated squared errors,* AISE$\times 10^{-2}$*, for the estimators of* β_{jk1} *under Scenario 1 for sparse and dense designs*

	Scenario 1, sparse			Scenario 1, dense		
n	$j=1$	$j=2$	$j=3$	$j=1$	$j=2$	$j=3$
100	0·96	1·03	0·95	0·59	0·61	0·61
200	0·50	0·55	0·46	0·31	0·33	0·29
400	0·26	0·50	0·27	0·14	0·17	0·14

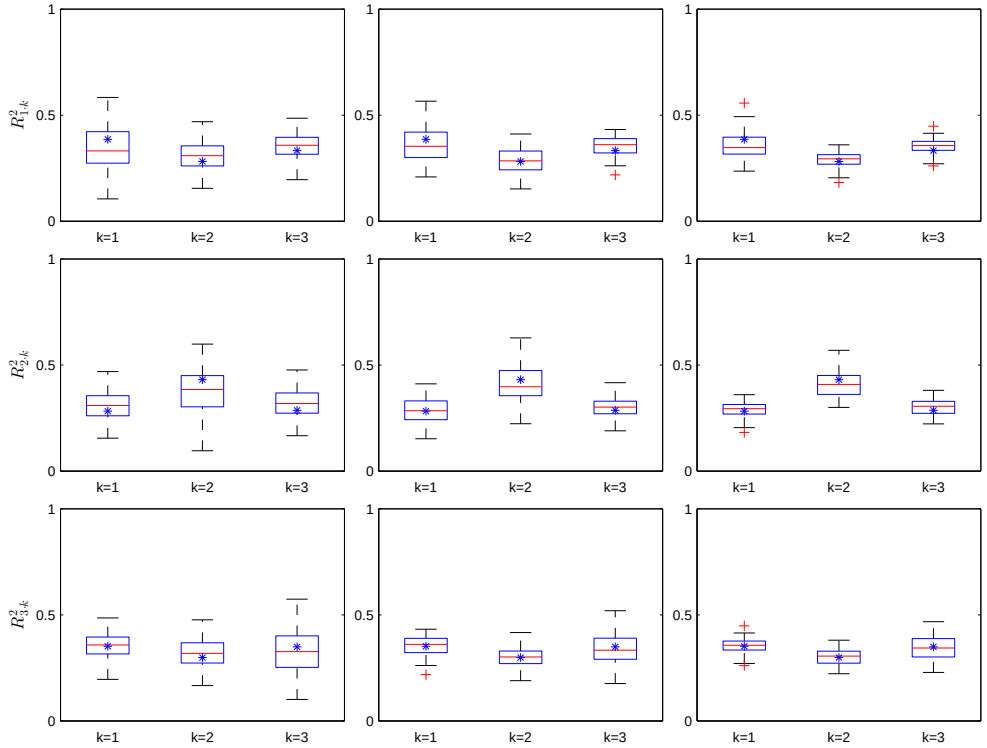


Fig. S1: Boxplots of the estimated $R^2_{j\cdot k}$ for $j = 1, 2, 3$ (row-wise from top to bottom) and for $k = 1, 2, 3$ (within each panel from left to right) under simulation Scenario 1, sparse design, for sample sizes $n = 100, 200, 400$ (column-wise from left to right). The stars in the boxplots indicate the true values of the $R^2_{j\cdot k}$.

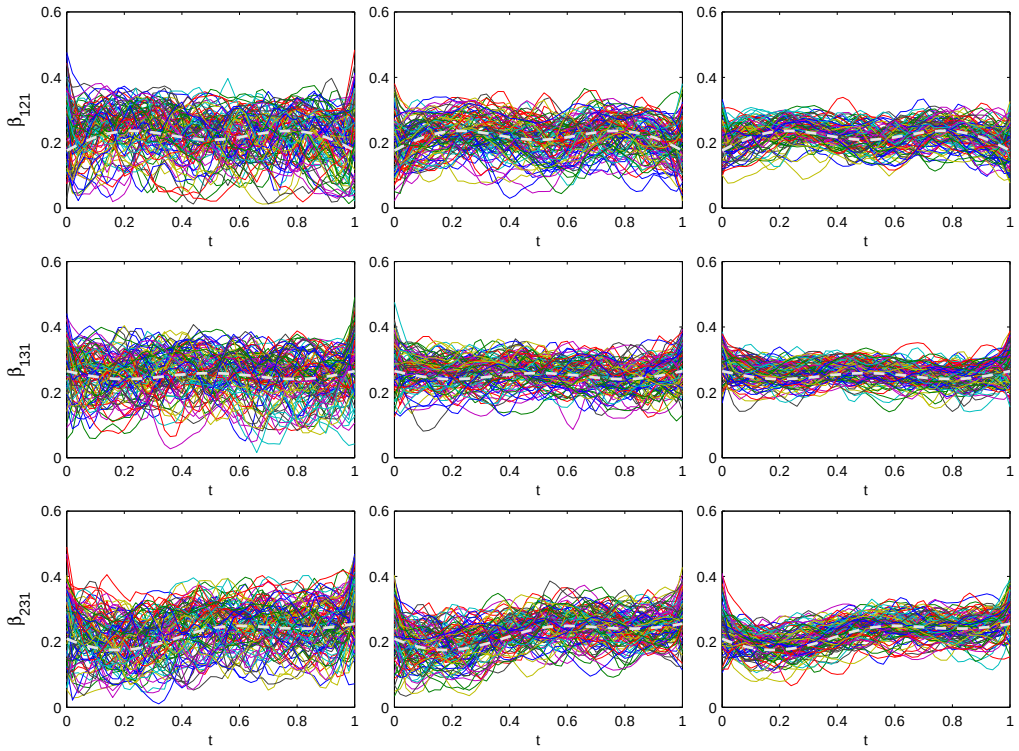


Fig. S3: Samples of the estimated $\beta_{jk1}(t)$ $(1 \leq j < k \leq 3)$ under Scenario 1, sparse design, for sample sizes $n = 100$, 200, and 400 (column-wise from left to right), based on 100 simulation data sets. The white dashed curves indicate the true functions.

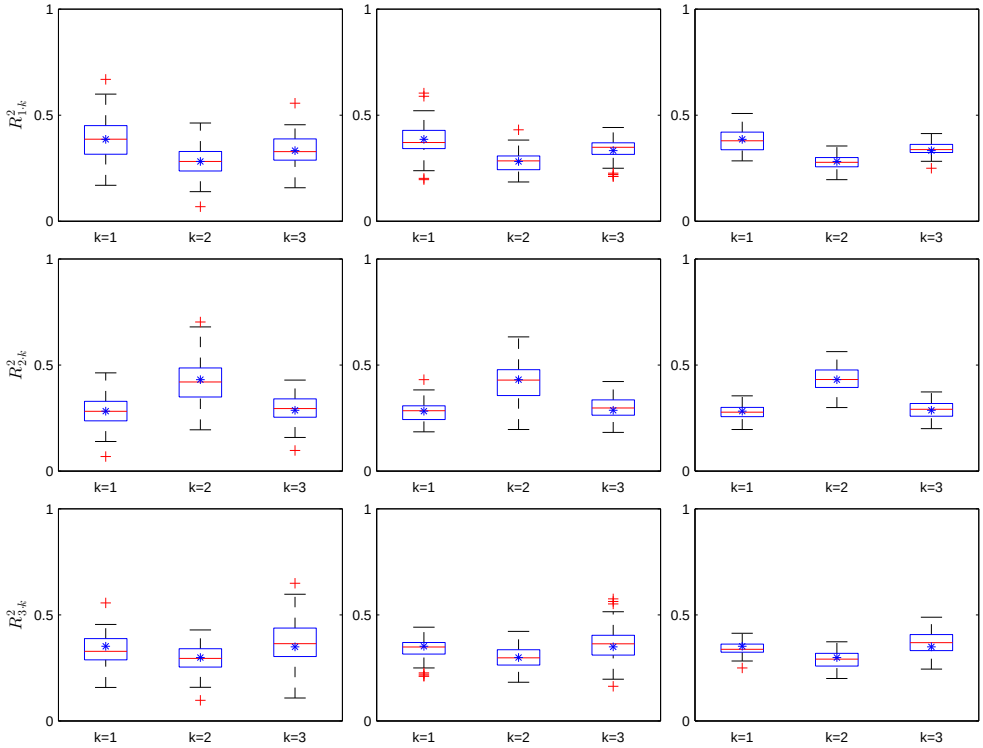


Fig. S2: Boxplots of the estimated $R^2_{j\cdot k}$ for $j = 1, 2, 3$ (row-wise from top to bottom) and for $k = 1, 2, 3$ (within each panel from left to right) under simulation Scenario 1, dense design, with sample sizes $n = 100, 200, 400$ (column-wise from left to right). The stars in the boxplots indicate the true values of the $R^2_{j\cdot k}$.

S.4. SIMULATION RESULTS FOR $p = 5$

Next we provide additional simulation details and results for the case of the sparse design for the proposed functional pairwise interaction model for the case of multivariate longitudinal trajectories with $p = 5$. All the previously reported simulation results are for the three-dimensional case $p = 3$, as in the the application example. Following the notations in the section of simulation, we describe the additional settings for the case of $p = 5$.

$\tilde{\boldsymbol{\theta}}^Y_1 = [9\ 4\ 1]$, $\tilde{\phi}^Y_{11}(t) = P^0_0(2t-1)$, $\tilde{\phi}^Y_{12}(t) = P^0_1(2t-1)$, $\tilde{\phi}^Y_{13}(t) = P^0_2(2t-1)$,

$\tilde{\boldsymbol{\theta}}^Y_2 = [10\ 5]$, $\tilde{\phi}^Y_{21}(t) = 1$, $\tilde{\phi}^Y_{22}(t) = \cos(\pi t)$,

$\tilde{\boldsymbol{\theta}}^Y_3 = [9\ 3]$, $\tilde{\phi}^Y_{31}(t) = \exp(t)/\{1+\exp(t)\}$, $\tilde{\phi}^Y_{32}(t) = \log(t+1)$,

$\tilde{\boldsymbol{\theta}}^Y_4 = [8\ 5\ 2]$, $\tilde{\phi}^Y_{41}(t) = P^1_1(2t-1)$, $\tilde{\phi}^Y_{42}(t) = P^1_2(2t-1)$, $\tilde{\phi}^Y_{43}(t) = P^1_3(2t-1)$,

$\tilde{\boldsymbol{\theta}}^Y_5 = [10\ 6]$, $\tilde{\phi}^Y_{51}(t) = P^2_2(2t-1)$, $\tilde{\phi}^Y_{52}(t) = P^2_3(2t-1)$,

$\tilde{\boldsymbol{\theta}}^Z_{12} = [7\ 3]$, $\tilde{\phi}^Z_{12} = \tilde{\phi}^Y_3$, $\tilde{\boldsymbol{\theta}}^Z_{13} = [7\ 4\ 2]$, $\tilde{\phi}^Z_{13} = \tilde{\phi}^Y_1$, $\tilde{\boldsymbol{\theta}}^Z_{14} = [8\ 3]$, $\tilde{\phi}^Z_{14} = \tilde{\phi}^Y_2$, $\tilde{\boldsymbol{\theta}}^Z_{15} = [8\ 5]$, $\tilde{\phi}^Z_{15} = \tilde{\phi}^Y_5$,

$\tilde{\boldsymbol{\theta}}^Z_{23} = [6\ 3\ 1]$, $\tilde{\phi}^Z_{23} = \tilde{\phi}^Y_4$, $\tilde{\boldsymbol{\theta}}^Z_{24} = [7\ 4\ 2]$, $\tilde{\phi}^Z_{24} = \tilde{\phi}^Y_1$, $\tilde{\boldsymbol{\theta}}^Z_{25} = [6\ 3]$, $\tilde{\phi}^Z_{25} = \tilde{\phi}^Y_5$,
$\tilde{\boldsymbol{\theta}}^Z_{34} = [6\ 3\ 1]$, $\tilde{\phi}^Z_{34} = \tilde{\phi}^Y_4$, $\tilde{\boldsymbol{\theta}}^Z_{35} = [8\ 4]$, $\tilde{\phi}^Z_{35} = \tilde{\phi}^Y_2$, $\tilde{\boldsymbol{\theta}}^Z_{45} = [6\ 2]$, $\tilde{\phi}^Z_{45} = \tilde{\phi}^Y_3$.

Table S2 reports the simulation results with the sample mean of $\text{RMSE}(X_j)$ under various sample sizes for the proposed functional pairwise interaction model for $p = 5$. The results indicate that the prediction errors in reconstructing the component functions through the interaction model decrease as the sample size increases.

Table S2: *Sample means (with standard errors in parentheses) of* $\text{RMSE}(X_j)$ *in (29) under Scenario 1 of the sparse design with* $n = 100, 200$*, and 400, based on 100 simulation replicates*

n	$\text{RMSE}(X_1)$	$\text{RMSE}(X_2)$	$\text{RMSE}(X_3)$	$\text{RMSE}(X_4)$	$\text{RMSE}(X_5)$
100	3·12 (6·68)	2·82 (6·44)	2·83 (6·40)	3·02 (6·42)	2·76 (8·22)
200	2·49 (5·32)	2·21 (4·02)	2·30 (4·40)	2·45 (5·05)	2·14 (5·99)
400	2·25 (3·71)	1·88 (2·74)	2·01 (3·79)	2·21 (3·75)	1·88 (4·32)

For the estimated $R^2_{j\cdot k}$ as shown in Fig. S4, all the median values of the boxplots are close to the corresponding true values. Table S3 summarises the performance of the estimated $\beta_{jk1}(t)$, and Fig. S5 displays the samples of the estimated $\beta_{jk1}(t)$. All simulation results for the case $p = 5$ indicate that the conclusions of the results are very similar to those for the case $p = 3$.

Table S3: *Average integrated squared errors,* $\text{AISE}\times 10^{-2}$*, for* β_{jk1} *under Scenario 1 for sparse designs*

n	$j = 1$	$j = 2$	$j = 3$	$j = 4$	$j = 5$
100	2·67	2·88	2·66	2·79	2·98
200	1·53	1·57	1·58	1·66	1·78
400	0·91	0·89	0·90	1·02	1·11

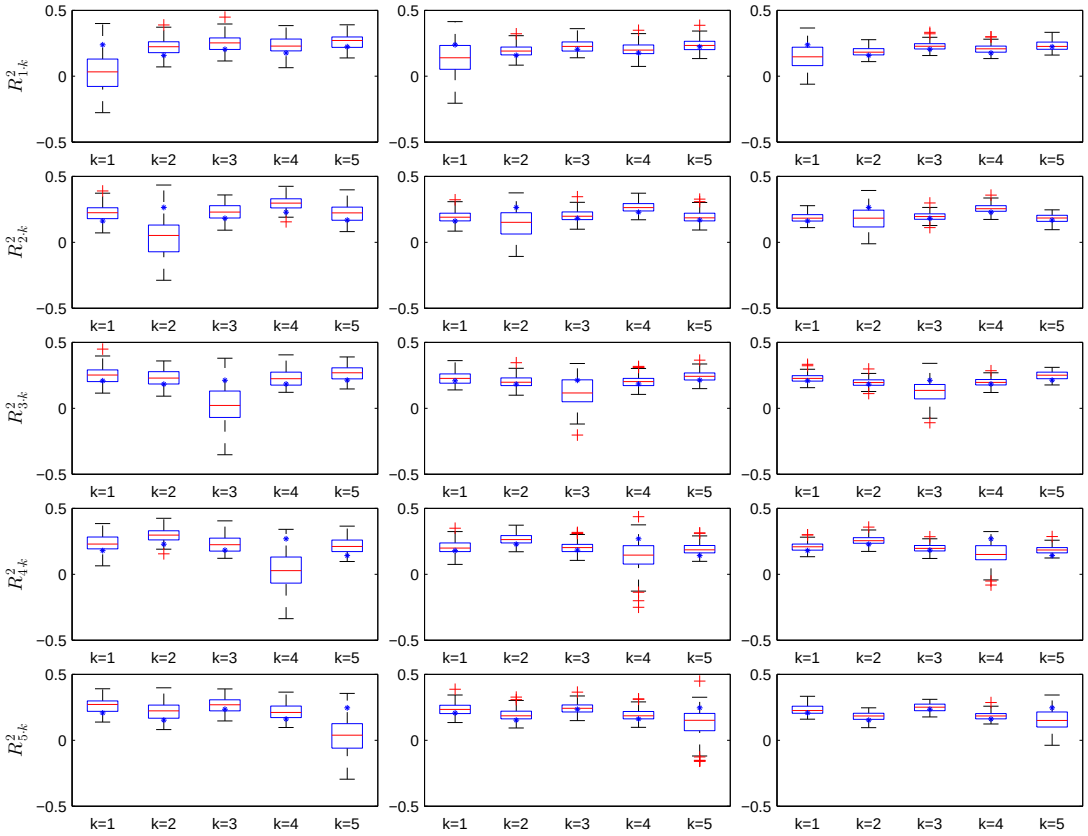


Fig. S4: Boxplots of the estimated $R^2_{j\cdot k}$ for $j = 1, \ldots, 5$ (row-wise from top to bottom) and for $k = 1, \ldots, 5$ (within each panel from left to right) under simulation Scenario 1, sparse design, for sample sizes $n = 100, 200, 400$ (column-wise from left to right). The stars in the boxplots indicate the true values of the $R^2_{j\cdot k}$.

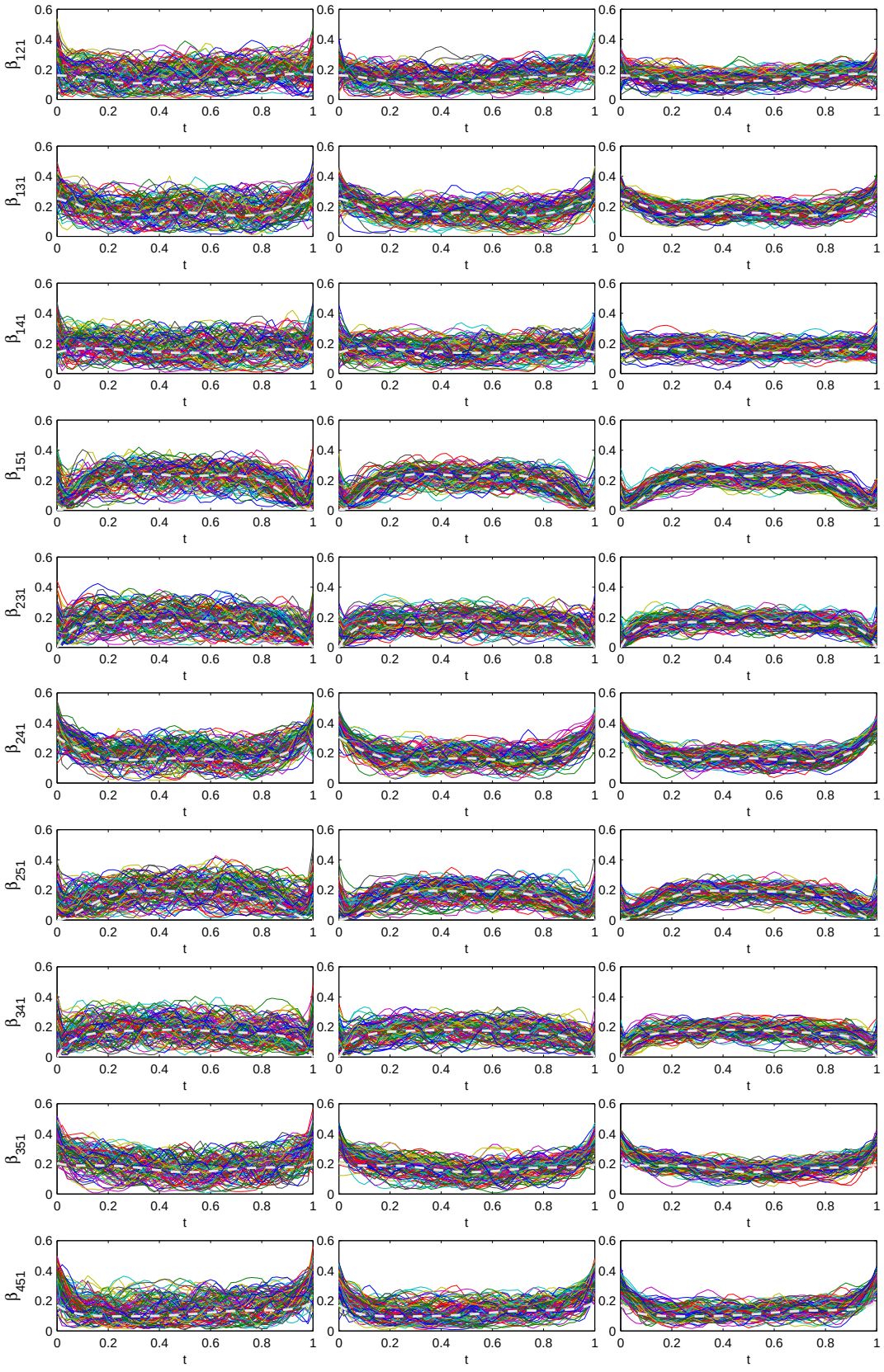


Fig. S5: Samples of the estimated $\beta_{jk1}(t)$ $(1 \leq j < k \leq 5)$ under Scenario 1, sparse design, for sample sizes $n = 100$, 200, and 400 (column-wise from left to right), based on 100 simulated data sets. The white dashed curves indicate the true functions.