ABSTRACT

ISLAM, MD NAZMUL. Functional Data Analysis with Applications in Animal Science and Biomedical Engineering. (Under the direction of Ana-Maria Staicu and Jonathan Richards-Stallings.)

In the second part of the dissertation, we develop flexible methodology to study the association between scalar outcomes and functional predictors observed over time, at many instances, in longitudinal studies. We propose a parsimonious modeling framework to study time-varying regression that leads to superior prediction properties and allows to reconstruct full trajectories of the response. The idea is to model the time-varying functional predictors using orthogonal basis functions and expand the time-varying regression coefficient using the same basis. Numerical investigation through simulation studies and data analysis show excellent performance in terms of accurate prediction and efficient computations, when compared with existing alternatives. The methods are inspired and applied to an animal science application, where of interest is to study the association between the feed intake of lactating sows and the minute-by-minute temperature throughout the 21st days of their lactation period. R code and an R illustration are provided at https://github.com/mnazmulislam/LDFR.

In the third part of the dissertation, we consider the problem of testing functional effects in the context of scalar-on-function regression in a longitudinal setting where the effects of functional covariates vary over time. Based on a sample, we are interested in making inference about: 1) the adequacy of functional structure, 2) the necessity of time-varying coefficient, 3) the pattern of change in the temporal structure of smooth coefficient, and 4) null effect of a predictor. We apply the functional principal component analysis on the response-adjusted functional predictors, and use the leading eigenbasis to model both the functional predictor and smooth coefficient. We represent the model using penalized splines in a linear mixed model framework and adopt a pseudo likelihood ratio test for making inference. Extensive simulations exhibit excellent numerical performance in terms of attaining the size and power properties for the tests in both single and multiple samples. Methods are applied to an animal science application where the primary objective is to infer about the effect of thermal environment on sows’ feeding behavior.

The fourth part of the dissertation consists of proposing a novel functional linear varying coefficient model with an application in Biomedical engineering field. State-of-the-art robotic hand prosthetics generate finger and wrist movement through pattern recognition (PR) al-
gorithms using features of forearm electromyogram (EMG) signals, but requires extensive training and is prone to poor predictions for conditions outside the training data (Scheme et al., 2010; Peerdeman et al., 2011). We propose a novel approach to develop a dynamic robotic limb by utilizing the recent history of EMG signals in a model that accounts for physiological features of hand movement which are ignored by PR algorithms. We do this by viewing EMG signals as functional covariates and develop a functional linear model that quantifies the effect of the EMG signals on finger/wrist velocity through a bivariate coefficient function that is allowed to vary with current finger/wrist position. The model is made parsimonious and interpretable through a two-step variable selection procedure, called Sequential Adaptive Functional Empirical group LASSO (SAFE-gLASSO). Numerical studies show excellent selection and prediction properties of SAFE-gLASSO compared to popular alternatives. For our motivating dataset, the method correctly identifies the few EMG signals that are known to be important for an able-bodied subject with negligible false positives and the model can be directly implemented in a robotic prosthetic.
© Copyright 2018 by Md Nazmul Islam

All Rights Reserved
Functional Data Analysis with Applications in Animal Science and Biomedical Engineering

by
Md Nazmul Islam

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Statistics

Raleigh, North Carolina
2018

APPROVED BY:

____________________________________________________________________
Brian Reich
He Huang

____________________________________________________________________
Ana-Maria Staicu
Co-chair of Advisory Committee
Jonathan Richards-Stallings
Co-chair of Advisory Committee
DEDICATION

To my family.
BIOGRAPHY

The author was born and raised in Chittagong, the port city of Bangladesh. He completed his secondary and higher secondary education from a high school named as Ispahani Public School & College (IPSC). After graduation from IPSC, he went to University of Dhaka, Bangladesh where he earned his Bachelor of Science degree in Statistics in May 2010. He completed his Master of Business Administration with a major in Finance from the Institute of Business Administration, University of Dhaka in 2012. He joined the Department of Statistics in 2013 and earned his Master of Science in Statistics with the concentration in Biostatistics in 2016. He continued to pursue a Doctor of Philosophy degree in Statistics from 2015.
ACKNOWLEDGEMENTS

I would like to take this opportunity to show my heartfelt gratitude to my advisors Dr. Ana-Maria Staicu and Dr. Jonathan Richards-Stallings for their endless support, knowledge, and guidance throughout my graduate life. Their relentless efforts and strive for improvements in research have shaped my scientific thought process for which I am indebted to them. I am grateful to them for having their faith on me. I would like to thank Dr. Stallings for financial support and teaching me how to collaborate with non-statisticians during multiple consulting projects.

I am also thankful to Dr. Brian Reich and Dr. He Huang for being in my advisory committee. I am thankful to Dr. Anastasios Tsiatis for being my student advisor in my earlier graduate period and paving my ways smooth for the later period. I am very much thankful to Dr. Kim Weems for supporting me and backing me up while I was struggling to get adjusted in the new environment. I want to thank Dr. Donald Martin and Dr. Charles Smith for their continuous support and inspiration. I am thankful to Alison McCoy and Terry Byron for their endless support. I want to thank to all the faculty members, staffs, and fellow graduate students in the department of Statistics at North Carolina State University. I am thankful to Arnab Hazra, Alexandra Larsen, Dr. Arkaprava Roy, Dr. Cai Li, Dr. Debraj Das, Haoyu Wang, Indranil Sahoo, Dr. Marcela Alfaro Cordoba, Merve Tekbudak, Dr. So-Young Park, Dr. Suhyun Kang, Saebitna Oh, Meredith King, Palpasa Manandhar, Stephanie Chen, and Susheela Singh for always being supportive.

I would like to acknowledge my brothers, Saiful Islam, Zahedul Islam, Towhidul Islam, and my elder sister Shamim Ara for encouraging and supporting me since my childhood. I would like to thank my beloved wife and best friend, Kona, for her unconditional love and care since the beginning of this journey, and for motivating me whenever I get distraught. My words are so few to express how grateful I am to my parents, Nurul Islam and Sajeda Begum, for their sacrifices, care, and for instilling the moral principles in my life. I want to thank my nephews and nieces, Shaan, Naisha, Norah, Aarish, Omar, and Umayyah. I want to take this opportunity to thank all of my other family members and friends who have encouraged me. I am really grateful to have all of you in my life; I could not have finished this journey without you.
TABLE OF CONTENTS

LIST OF TABLES ........................................................................ viii

LIST OF FIGURES ...................................................................... xii

Chapter 1 Introduction ............................................................. 1
  1.1 Outlines and contributions ............................................... 3

Chapter 2 Longitudinal Dynamic Functional Regression .............. 6
  2.1 Introduction ..................................................................... 6
  2.2 Proposed methodology ................................................... 9
    2.2.1 Selection of the orthogonal basis .............................. 12
    2.2.2 Statistical modeling of the non-linear effects .......... 13
  2.3 Estimation ....................................................................... 13
    2.3.1 Estimation of the covariates-related components .... 14
    2.3.2 Estimation of the response-related components ....... 15
  2.4 Prediction and inference .................................................. 17
    2.4.1 Prediction of response trajectories ......................... 17
    2.4.2 Asymptotic prediction bands ................................. 18
  2.5 Simulation ..................................................................... 19
    2.5.1 Description of the settings ..................................... 19
    2.5.2 Evaluation criteria .................................................. 21
    2.5.3 Prediction performance assessment and comparison ... 22
  2.6 Data application ............................................................. 30
    2.6.1 Fit assessment ......................................................... 32
    2.6.2 Estimation of the model components. Prediction of response trajectories 35
    2.6.3 Validation of the results for data application via simulation study .... 38
  2.7 Discussion ...................................................................... 40

Chapter 3 Significance Testing for Functional Effects in the Longitudinal
  Time-Varying Functional Regression Model .......................... 41
  3.1 Introduction ..................................................................... 41
  3.2 Proposed methodology ................................................... 44
    3.2.1 Preliminaries .......................................................... 44
    3.2.2 Testing for time-varying functional effect .............. 45
    3.2.3 Testing procedure .................................................... 49
  3.3 Selection of basis functions ............................................. 50
  3.4 Testing different classes of hypotheses ............................. 52
    3.4.1 Testing for multiple predictors ............................... 54
  3.5 Simulation ..................................................................... 55
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.5.1</td>
<td>Description of the settings</td>
<td>56</td>
</tr>
<tr>
<td>3.5.2</td>
<td>Computational implementation</td>
<td>58</td>
</tr>
<tr>
<td>3.5.3</td>
<td>Performance assessments and comparisons</td>
<td>59</td>
</tr>
<tr>
<td>3.6</td>
<td>Data application</td>
<td>66</td>
</tr>
<tr>
<td>3.6.1</td>
<td>Significance testing</td>
<td>70</td>
</tr>
<tr>
<td>3.6.2</td>
<td>Estimation of regression surfaces</td>
<td>71</td>
</tr>
<tr>
<td>3.7</td>
<td>Discussion</td>
<td>74</td>
</tr>
</tbody>
</table>

**Chapter 4 Functional Variable Selection for EMG-based Control of a Robotic Hand Prosthetic**

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1</td>
<td>Introduction</td>
<td>75</td>
</tr>
<tr>
<td>4.2</td>
<td>Biomechanics of hand movement</td>
<td>78</td>
</tr>
<tr>
<td>4.3</td>
<td>Data collection and processing</td>
<td>80</td>
</tr>
<tr>
<td>4.3.1</td>
<td>Data collection procedure</td>
<td>82</td>
</tr>
<tr>
<td>4.3.2</td>
<td>Post-collection data processing for analysis</td>
<td>82</td>
</tr>
<tr>
<td>4.4</td>
<td>Variable selection and inferential framework</td>
<td>83</td>
</tr>
<tr>
<td>4.4.1</td>
<td>Approximation to linear model</td>
<td>84</td>
</tr>
<tr>
<td>4.4.2</td>
<td>Penalized criterion for variable selection</td>
<td>85</td>
</tr>
<tr>
<td>4.4.3</td>
<td>Adaptive penalized criterion</td>
<td>87</td>
</tr>
<tr>
<td>4.4.4</td>
<td>Selection of the tuning parameters, $\lambda$ and $\phi$</td>
<td>88</td>
</tr>
<tr>
<td>4.4.5</td>
<td>Sequential Adaptive Functional Empirical (SAFE) Selection</td>
<td>88</td>
</tr>
<tr>
<td>4.4.6</td>
<td>Post-selection inference and prediction</td>
<td>89</td>
</tr>
<tr>
<td>4.5</td>
<td>Data Application: EMG selection for finger/wrist movement</td>
<td>90</td>
</tr>
<tr>
<td>4.5.1</td>
<td>Computational Details</td>
<td>92</td>
</tr>
<tr>
<td>4.5.2</td>
<td>Variable selection, prediction, and implication</td>
<td>93</td>
</tr>
<tr>
<td>4.5.3</td>
<td>Estimation of smooth coefficients</td>
<td>95</td>
</tr>
<tr>
<td>4.6</td>
<td>Simulation study</td>
<td>97</td>
</tr>
<tr>
<td>4.6.1</td>
<td>Simulation mimicking the data application</td>
<td>97</td>
</tr>
<tr>
<td>4.6.2</td>
<td>Numerical experiment</td>
<td>100</td>
</tr>
<tr>
<td>4.7</td>
<td>Discussion</td>
<td>103</td>
</tr>
</tbody>
</table>

**Bibliography**

<table>
<thead>
<tr>
<th>Appendix</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Additional details for Chapter 2</td>
<td>121</td>
</tr>
<tr>
<td>A.1</td>
<td>Additional simulation results</td>
<td>121</td>
</tr>
<tr>
<td>A.1.1</td>
<td>Predictor functions</td>
<td>121</td>
</tr>
<tr>
<td>A.1.2</td>
<td>Gaussian response</td>
<td>122</td>
</tr>
<tr>
<td>A.1.3</td>
<td>Binary response</td>
<td>130</td>
</tr>
<tr>
<td>A.1.4</td>
<td>Comparison of results using B-splines and truncated linear splines</td>
<td>131</td>
</tr>
<tr>
<td>A.2</td>
<td>Additional results for the sows data application</td>
<td>132</td>
</tr>
<tr>
<td>A.2.1</td>
<td>Autoregressive model</td>
<td>132</td>
</tr>
</tbody>
</table>
A.2.2 Validation of residual assumption and model selection . . . . . . . 135
A.2.3 Additional results for the model presented in the main article . . 137
A.2.4 Fitting simpler model . . . . . . . . . . . . . . . . . . . . . . . . . . 145
A.2.5 Simulated data . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 145
A.3 Technical details . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 146
Appendix B Additional details for Chapter 3 . . . . . . . . . . . . . . . . . . . . . 147
   B.1 Validation of covariance assumptions . . . . . . . . . . . . . . . . . . . 147
   B.2 Additional simulation results . . . . . . . . . . . . . . . . . . . . . . . . 149
   B.3 Mixed model representation . . . . . . . . . . . . . . . . . . . . . . . . . 151
Appendix C Additional details for Chapter 4 . . . . . . . . . . . . . . . . . . . . . 153
   C.1 Generalized additive model . . . . . . . . . . . . . . . . . . . . . . . . . 153
   C.2 Additional results of EMG selection for movement data . . . . . . . 154
      C.2.1 Data restructuring of EMG signals . . . . . . . . . . . . . . . . . . . 154
      C.2.2 Variation in EMG curves . . . . . . . . . . . . . . . . . . . . . . . . . 154
      C.2.3 Variable selection for wrist movements . . . . . . . . . . . . . . . . 155
      C.2.4 Residual diagnostics . . . . . . . . . . . . . . . . . . . . . . . . . . . 156
      C.2.5 Regression surfaces for finger movements . . . . . . . . . . . . . . 157
      C.2.6 Regression surfaces for wrist movements . . . . . . . . . . . . . . . 161
      C.2.7 Regression surfaces based on competing approaches . . . . . . . . 165
   C.3 Additional simulation results for the data mimicking application . . 166
   C.4 Numerical experiment . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 168
   C.5 Algorithm of split conformal prediction band . . . . . . . . . . . . . . 172
Table 2.1 Gaussian responses with CS dependence structure ($D_{1ii}$), when the longitudinal design is sparse ($C_1$) and moderately sparse (mod sparse, $C_2$); the functional covariates are observed with high noise variance ($B_1$) and effect $E_1$. Model is fitted assuming CS type dependence structure. Median prediction errors and IQR in parenthesis are reported for 1000 simulations.

Table 2.2 Gaussian responses with CS dependence structure ($D_{1ii}$), when the longitudinal design is sparse ($C_1$) for $B_1$ and $B_2$ with effect $E_1$. Model is fitted assuming CS type dependence structure. Median prediction errors and IQR in parenthesis are reported.

Table 2.3 Binary responses ($D_2$) when the longitudinal design is sparse/moderately sparse (mod sparse); the functional covariates are observed with high noise variance ($B_1$) and effect $E_1$. Fitted model assumes subject-specific random intercept. We report the median of prediction errors of the linear predictor trajectories and, in parenthesis, the median of true-positive-rates.

Table 2.4 Gaussian responses with CS dependence structure ($D_{1ii}$), when the longitudinal design is moderately sparse; the functional covariates are observed with high ($B_1$) and low ($B_2$) noise variance with effect $E_1$. Fitted model assumes CS covariance structure. Reported are the average coverage probabilities of 95% and 90% pointwise prediction bands, standard errors (in parenthesis), average length (in square bracket), for the observed ($Y_i$) and unobserved subjects ($Y_{i*}$). Results are based on 1000 MC simulations.

Table 2.5 Median prediction accuracy, computing time (in seconds), and marginal AIC for the proposed LDFR and the competitive approaches LPFR and LPEER for the sows application; corresponding IQRs are reported in parenthesis.

Table 2.6 Numerical results based on 100 Monte Carlo simulations mimicking sow data. Reported are the median prediction errors, marginal AICs, RMPE$_{trj}$, IQR (in parenthesis), average coverage probabilities of 95% and 90% pointwise prediction bands, standard errors (in parenthesis with superscript †), and average length of intervals (in square bracket) for the existing ($Y_i$) and new sows ($Y_{i*}$).
Table 3.1 Testing for the adequacy of time-varying effect \((F1)\) for the case \(H1\). The leading \(K\) FPCs explain at least 90\% variation of marginal covariance function. Reported are the observed type I error rates and standard errors (in parenthesis) based on 5000 Monte Carlo simulations. The functional covariate is observed as in \(A1\) with high \((B2)\) SNR; the smooth functional coefficient is represented as in \(G1\). Responses are generated assuming normality \((C)\) and with CS type covariance structure \((D)\). Both moderate \((E1)\) and low \((E2)\) sparse sampling designs are considered.

Table 3.2 Testing for linear pattern of time-varying effect \((F2)\) for the case \(H1\). The leading \(K\) FPCs explain at least 90\% variation of marginal covariance function. Reported are the observed type-I errors and the corresponding standard errors (in parenthesis) based on 5000 Monte Carlo simulations. The functional covariate is observed as in \(A1\) with high \((B2)\) SNR; the smooth functional coefficient is represented as in \(G2\). Responses are generated assuming normality \((C)\) and with CS type covariance structure \((D)\). Both moderate \((E1)\) and low \((E2)\) sparse sampling designs are considered.

Table 3.3 Testing for the adequacy of time-varying effect \((F1)\) for the case \(H2\). The leading \(K\) FPCs explain at least 90\% variation of marginal covariance function. Reported are the observed type I error rates and standard errors (in parenthesis) based on 5000 Monte Carlo simulations. The functional covariate is observed as in \(A1\) with high \((B2)\) SNR; the smooth functional coefficient is represented as in \(G4\). Responses are generated assuming normality \((C)\) and with CS type covariance structure \((D)\). Both moderate \((E1)\) and low \((E2)\) sparse sampling designs are considered.

Table 3.4 \(p\)-values for testing the null hypotheses associated with the scientific questions. Model with \(\xi_V\) corresponds to the eigenbasis obtained from the spectral decomposition of the covariance operator of the response-adjusted functional covariates. In contrast, model with \(\xi_W\) uses the eigenbasis of the marginal covariance function of the functional predictor without accounting for the correlation between the responses and predictors.

Table 4.1 Finger movements EMG signal selection for consistent (top three rows) and random (bottom three row) patterns at different postures and results with superscript \(\dagger\) correspond to the first stage of SAFE-gLASSO. RSPs, number of variables (square brackets), and the percentages of TPRs and FPRs are presented.

Table 4.2 Analysis of finger movements with fixed motion. Data is generated assuming noise variance \(C1\) and \(C2\) with different dominant processes \((A1-A3)\) for high correlation coefficient \((B2)\). Reported are the SPs (%), model size (in square brackets), TPRs (%), and FPRs (%) averaged over 100 simulations. Results with superscript \(\dagger\) correspond to the first stage of SAFE-gLASSO.
Table 4.3  Average pointwise coverage of 80%, 90%, 95%, and 99% prediction bands for
the in-sample ($y_i$) and future ($y_i^*$) observations using ROO and split conformal
prediction inference, respectively; standard errors (in parenthesis) and average
lengths (in square bracket) of prediction bands are reported. Results are based
on 100 simulations.

Table A.1  $SNR_y$ levels for different simulation settings; values are rounded to one decimal
place.

Table A.2  Gaussian responses with independent covariance structure ($D1i$), when the
longitudinal design is sparse ($C1$) and moderately sparse (mod sparse, $C2$);
the functional covariates are observed with high noise variance ($B1$) with
effect $E1$. Model is fitted assuming CS type dependence structure. Median
prediction errors and IQR in parenthesis are reported for 1000 simulations.

Table A.3  Gaussian responses with independent covariance structure ($D1i$), when the
longitudinal design is sparse ($C1$) and moderately sparse (mod sparse, $C2$);
the functional covariates are observed with low noise variance ($B2$) with
effect $E1$. Model is fitted assuming CS type dependence structure. Median
prediction errors and IQR in parenthesis are reported for 1000 simulations.

Table A.4  Gaussian responses with dependent covariance structure ($D1ii$), when the
longitudinal design is sparse ($C1$) and moderately sparse (mod sparse, $C2$);
the functional covariates are observed with low noise variance ($B2$) with
effect $E1$. Model is fitted assuming CS type dependence structure. Median
prediction errors and IQR in parenthesis are reported for 1000 simulations.

Table A.5  Gaussian responses with dependent covariance structure ($D1iii$), when the
longitudinal design is sparse ($C1$) and moderately sparse (mod sparse, $C2$);
the functional covariates are observed with high noise variance ($B1$) with
effect $E1$. Model is fitted assuming both subject-specific random intercept
and slope. Median prediction errors and IQR in parenthesis are reported
for 1000 simulations.

Table A.6  Gaussian responses with dependent covariance structure ($D1iii$), when the
longitudinal design is sparse ($C1$) and moderately sparse (mod sparse, $C2$);
the functional covariates are observed with low noise variance ($B2$) with
effect $E1$. Model is fitted assuming both subject-specific random intercept
and slope. Median prediction errors and IQR in parenthesis are reported
for 1000 simulations.

Table A.7  Gaussian responses with CS dependence structure ($D1ii$), when the lon-
gitudinal design is sparse ($C1$) for $B1$ and $B2$ with effect $E2$. Model is
fitted assuming CS type dependence structure. Median prediction errors
and IQR in parenthesis are reported.
Table A.8  Binary responses $(D2)$, when the longitudinal design is sparse $(C1)$ and moderately sparse $(C2)$; the functional covariates are observed with low noise variance $(B2)$ and with effect $E1$. Model is fitted assuming subject-specific random intercept. We report the median of prediction errors of the linear predictor trajectories and, in parenthesis, the median of true-positive-rates. ................................................................. 131

Table A.9  Modeling the unknown smooth parameters by B-splines and truncated linear splines. Reported are the median and IQR of in-sample and out-of-sample prediction errors. Results are based on 100 independent samples. 132

Table A.10 Prediction accuracy of the lagged models as assessed by RMPE. .......................... 135

Table A.11 Model comparison. Reported are the median prediction accuracy, marginal AIC, and BIC for model 10 with independent and dependent covariance assumption of random intercept and slope; IQR are reported in parenthesis. 137

Table A.12 Relative importance of model components. ................................................................. 142

Table A.13 Comparison of prediction accuracy and run time for the model with subject-specific random intercept and random block effect. ................................. 145

Table B.1  Testing for nullity and time-invariant effect $(F3)$ simultaneously for the case $H1$. The leading $K$ FPCs explain at least 90% variation of marginal covariance function. Reported are the observed type-I errors and the corresponding standard errors (in parenthesis) based on 5000 Monte Carlo simulations. The functional covariates are observed as in $A1$ and $A2$ with low $(B1)$ SNR$_w$; the smooth functional coefficients are represented as in $G3$. Responses are generated assuming normality $(C)$ and with CS type covariance structure $(D)$. Both moderate $(E1)$ and low $(E2)$ sparse sampling designs are considered. ................................. 150

Table C.1  Wrist movements EMG signal selection. RSPs, model size (square brackets), and the percentages of TPRs and FPRs are presented for consistent (top three rows) and random (bottom three rows) patterns at different postures. Results with superscript † correspond to the first stage of SAFE-gLASSO. ................................. 156

Table C.2  Analysis of finger movements with fixed motion. Data is generated assuming noise variance $C1$ and $C2$ with different dominant processes $(A1$-$A3)$ for low correlation coefficient $(B1)$. Reported are the SPs (%), model size (in square brackets), TPRs (%), and FPRs (%) averaged over 100 simulations. Results with superscript † correspond to the first stage of SAFE-gLASSO. ................................. 167

Table C.3  Variable selection for the toy example. Reported are the SPs (%), model size (in square brackets), TPRs (%), and FPRs (%) averaged across 100 simulations. Results with superscript † correspond to the first stage of SAFE-gLASSO. ................................. 170
Figure 2.1 Data for a randomly chosen sow; feed-intake (kg) and temperature (°C) profiles from left to right respectively. Two randomly chosen curves are highlighted in the right panel and the corresponding feed-intake values are filled with colors in the left panel, respectively. .................................................. 7

Figure 2.2 Gaussian responses with CS dependence structure (D1ii), when the longitudinal design is sparse (left) and moderately sparse (right); the functional covariates are observed with high noise variance (B1) and effect E1. Fitted model assumes CS covariance structure. Reported is RMPE_{trj} for observed (white boxplot) and unobserved (gray boxplot) subjects based on 1000 simulations. Reference lines are drawn at RMPE values 1, 2, and 3 for convenience. ................................................................. 27

Figure 2.3 Parameter estimates in the lactation sow application. Left panel depicts the estimated intercept function for the old (solid line) and young (dashed) sows with 95% pointwise confidence intervals. Right panel shows the estimated regression coefficient \( \hat{\gamma}(\cdot,t) \) for each lactation day \( t = 1, 2, \ldots, 21 \). .................................................. 36

Figure 2.4 Predicted full feed intake trajectories for two young (left and middle) and one old (right) sows. Shaded regions correspond to 95% pointwise prediction bands based on LDFR. ................................................................. 37

Figure 3.1 Observed power function for testing the adequacy of time-invariant effect (F1) for low (top-left-to-right) and high (bottom-left-to-right) SNR_y. Reported are the proportion of rejections based on 1000 simulations for \( \alpha = 0.05 \) at \( \delta = \{0.05, 0.20, 0.50, 1.00, 2.00\} \). The functional covariate is observed as in A1 with high \( \text{SNR}_w \) (B2) while the smooth effect is represented by G1. .................................................. 64

Figure 3.2 Observed power function for testing linear pattern of time-varying effect (F2) for low (top-left-to-right) and high (bottom-left-to-right) SNR_y. Reported are the proportion of rejections based on 1000 simulations for \( \alpha = 0.05 \) at \( \delta = \{0.05, 0.20, 0.50, 1.00, 2.00\} \). The functional covariate is observed as in A2 with high \( \text{SNR}_w \) (B2) while the smooth effect is represented by G2. .................................................. 65

Figure 3.3 Observed power function for testing the adequacy of time-invariant effect (F1) for the case H2 for low (top-left-to-right) and high (bottom-left-to-right) SNR_y. Reported are the proportion of rejections based on 1000 simulations for \( \alpha = 0.05 \) at \( \delta = \{0.05, 0.10, 0.25, 0.50, 1.00\} \). The functional covariate is observed as in A1 with high \( \text{SNR}_w \) (B2) while the smooth effect is represented by G4. .................................................. 66
Figure 3.4 Data for a randomly chosen sow: feed-intake (in kg) (left), temperature (in Celsius) (middle), and relative humidity (in %) (right) which are recorded at 5-minutes interval for 24 hours period starting from 2:00PM on each day inside the farrowing facility. Color (“red” and “blue”) correspondence is provided for convenience.

Figure 3.5 Estimated regression coefficients associated the air temperature (left) $\widehat{\gamma}_T(\cdot, \cdot)$ and relative humidity (right) $\widehat{\gamma}_H(\cdot, \cdot)$ for each lactation day ($X$-axis) $t = 1, \ldots, 21$ over 24-hours period ($Y$-axis). The corresponding mean effects averaged over hours (red solid line at the top of the heat map) and lactation days (blue solid line at the right of the heat map) are plotted for the air temperature and relative humidity. Reported are the plots based on the model with $\xi_V$.

Figure 4.1 Visualization of biomechanical system for hand movement for AB subject and TRA. (A) Internal limb representation in the motor cortex. (B) Neural signal sends motor commands to forearm muscles. (C) Forearm muscles contract to perform desired movement, which, for an AB subject, results in direct hand movement through tendon connections (red lines) to the hand. (D) For a TRA, a prosthesis controller projects real time hand movements using forearm EMG information, and (E) the robotic limb performs the intended movement.

Figure 4.2 Top panel corresponds to finger and wrist movement: finger extension, finger flexion, wrist extension, and wrist flexion (from left to right). Bottom panel illustrates the arm postures at which data is collected during the experiment.

Figure 4.3 Depicted are the joint finger angle (radians) for flexion and extension movements, in black, and two normalized EMG signals, in red and green. Left and right Y-axes correspond to finger angle and EMG signals, respectively. Event (a) demonstrates restricted movement due to a physical constraint, in this case maximal finger extension. Event (b) demonstrates movement due to passive forces, lacking active EMG signals.

Figure 4.4 Reference forearm muscles for finger (top) and wrist (bottom) movements.

Figure 4.5 Comparison of MSEs based on alternative approaches. Results correspond to finger (left) and wrist (right) movements with different patterns and postures.

Figure 4.6 Predicted velocity corresponding to consistent finger movements based on SAFE-gLASSO (red dashed line) and agLASSO (blue dashed line). Vertical reference lines are drawn at 1.45 and 1.88 second at which movements due to passive forces occur in the absence of muscle contractions. Shaded regions correspond to pointwise 95% split conformal prediction bands for SAFE-gLASSO.
Figure 4.7 Regression surfaces for finger flexion (top-left), finger extension (top-right), wrist flexion (bottom-left), and wrist extension (bottom-right). The selected forearm muscles are pointed out by red circles.

Figure 4.8 Comparison of MSEs based on competing approaches for high (left) and low (right) SNRs with high correlation coefficient ($B_2$). Simulated data corresponds to the consistent finger movements. Reported are the results based on 100 simulations.

Figure A.1 Functional observations for a randomly chosen subject for the initial simulation design with SNR = 0.5 (left) and with SNR = 2.5 (middle) and for the simulation mimicking the data application, SNR = 1.3 (right). The solid lines correspond to the smoothed profiles.

Figure A.2 Gaussian responses with CS dependence structure ($D_{1ii}$), when the longitudinal design is sparse (left) and moderately sparse (right); the functional covariates are observed with low noise variance ($B_2$) and effect $E_1$. Fitted model assumes CS covariance structure. Reported is $RMPE_{trj}$ for observed (white boxplot) and unobserved (gray boxplot) subjects based on 1000 simulations. Reference lines are drawn at RMPE values 1, 2, and 3 for convenience.

Figure A.3 Residuals versus predicted values for different models: model (10) in the current submission (left panel) and lagged model (A.1) in the right.

Figure A.4 ACF (left) and PACF (right) function for residuals.

Figure A.5 Diagnostics of the residual assumption of model (10); scatter plot of the residuals against the fitted values (left), histogram of the residuals (middle), and normal Q-Q plot (right).

Figure A.6 Daily mean temperature $\bar{T}(\cdot,\cdot)$ during the lactations days.

Figure A.7 Estimated basis functions $\phi_k(\cdot)$ corresponding to functional PCs.

Figure A.8 Observed (left) and smooth (right) temperature profiles for a randomly chosen sow.

Figure A.9 Full trajectory of time-varying basis coefficients corresponding to first (left of first row), second (right of first row), third (left of second row), fourth (right of second row), fifth (left of third row), sixth (right of third row), and seventh (bottom row) functional PCs.

Figure A.10 Prediction of sow-group specific random coefficients.

Figure A.11 Predicted responses with LDFR (red), LPEER (blue), and LPFR (purple) for two young (left and middle) an old (right) sows; the actual observations are given in black. Shaded regions correspond to 95% pointwise prediction bands based on LDFR.

Figure A.12 Illustration of simulated response for a randomly chosen subject.
Figure B.1 Observed power function for simultaneously testing the adequacy of time-varying functional effect and nullity ($F_4$) for low (top-left-to-right) and high (bottom-left-to-right) SNR$_y$. Reported are the proportion of rejections based on 1000 simulations for $\alpha = 0.05$ at $\delta = \{0.05, 0.20, 0.50, 1.00, 2.00\}$. The functional covariates are observed as in $A1$ and $A2$ with low SNR$_w$ ($B1$) while the smooth effects are represented by $G3$.  

Figure C.1 Visualization of data restructuring of EMG signals and joint finger angles at 0.9 second and 2.3 second. Time at current position (blue dot on black line) used to extract concurrent and previous $\delta$ values of the two EMG signals, shown in red and green in (a). The time domain for each set of past $\delta + 1$ EMG signal measurements is rescaled to $[-\delta, 0]$. Reconstructed EMG curves are plotted in gray lines on the top-right and bottom-right panels, and two curves corresponding to 0.9 second and 2.3 second are highlighted in (b) and (c) for EMG-7 (green solid line) and EMG-12 (red solid line) respectively.

Figure C.2 Restructured EMG curves corresponding to the extensor digitorum muscle and the first three eigenfunctions from left to right.

Figure C.3 Diagnostics of the residual assumption of the statistical model for kinematic data; scatter plot of the residuals against predicted values (left), histogram of the residuals (middle), and normal Q-Q plot (right).

Figure C.4 Autocorrelation (left) and partial autocorrelation (right) of the residuals corresponding to the varying finger movements.

Figure C.5 Postures for finger/wrist extension and flexion.

Figure C.6 Regression surface for finger flexion (left) and extension (right). Reported are the estimates corresponding to consistent finger movements with neutral posture depicted in the left panel of Figure C.5.

Figure C.7 Regression surfaces for finger flexion (left) and extension (right). Reported are the estimates corresponding to consistent finger movements with posture depicted in the right panel of Figure C.5.

Figure C.8 Regression surfaces for finger flexion (left) and extension (right). Reported are the estimates corresponding to varying finger movements with neutral posture depicted in the left panel of Figure C.5.

Figure C.9 Regression surface for finger flexion (left) and extension (right). Reported are the estimates corresponding to varying finger movements with posture depicted in the middle panel of Figure C.5.

Figure C.10 Regression surfaces for finger flexion (left) and extension (right). Reported are the estimates corresponding to varying finger movements with posture depicted in the right panel of Figure C.5.

Figure C.11 Regression surfaces for wrist flexion (left) and extension (right). Reported are the estimates corresponding to consistent wrist movements with neutral posture depicted in the left panel of Figure C.5.
Figure C.12 Regression surfaces for wrist flexion (left) and extension (right). Reported are the estimates corresponding to consistent wrist movements with posture depicted in the right panel of Figure C.5. ......................... 163
Figure C.13 Regression surfaces for wrist flexion (left) and extension (right). Reported are the estimates corresponding to varying wrist movements with neutral posture depicted in the left panel of Figure C.5. ......................... 163
Figure C.14 Regression surfaces for wrist flexion (left) and extension (right). Reported are the estimates corresponding to varying wrist movements with posture depicted in the middle panel of Figure C.5. ......................... 164
Figure C.15 Regression surfaces for wrist flexion (left) and extension (right). Reported are the estimates corresponding to varying wrist movements with posture depicted in the right panel of Figure C.5. ......................... 164
Figure C.16 Estimates of smooth coefficients for the selected muscles: flexor digitorum (left) and extensor digitorum (right) corresponding to agLASSO (top panels) and LAD-gLASSO (bottom panels). ......................... 165
Figure C.17 Simulated velocity across different simulation settings at SNR = \{875, 80, 8, 4, 0.8\}. Data correspond to non-varying finger movements. ......................... 166
Figure C.18 Comparison of MSEs based on competing approaches for high (left) and low (right) SNRs with high correlation coefficient (B2). Simulated data corresponds to the non-varying finger movements. Reported are the results based on 100 simulations. ......................... 167
Figure C.19 Functional measurements corresponding to the first two predictors; three curves are chosen randomly and displayed as black solid lines. ......................... 168
Figure C.20 $\beta_1$ (top) and $\beta_2$ (bottom) for $C = \{0, 2, 5, 10\}$ from left to right respectively.169
Figure C.21 Reported is MSEs based on competing approaches for different SNRs at $C = 0$ (left), $C = 2$ (middle), and $C = 5$ (right). Results are based on 100 simulations. Reference lines are drawn at MSE = 0.04 (left), MSE = 0.20 (middle), and MSE = 0.75 (right) for convenience. ......................... 171
Functional data analysis (FDA) technique is a statistical approach which has been extensively used in analyzing high dimensional data in various fields. The term functional refers to the inherent structure of the data; it is to say that the data are being generated by some underlying unknown functions and the discrete observed measurements are treated as snapshots of that function at different grid of points. In general, functional data is comprised of a sample of curves (smooth functions) which are typically assumed to be independent, identically distributed realizations of an underlying stochastic process. The key idea in FDA is smoothness which implies that the values at successive points in a neighborhood are not independent and they do not differ significantly (Ramsay & Silverman, 2005). Unlike multivariate data, time ordering is crucial in functional data and any reordering will disrupt FDA completely. In contrast to both longitudinal and multivariate data, FDA can extract additional information from the underlying smooth functions and their derivatives. Due to its increasing popularity and applicability, FDA has been extensively studied in the literature in the recent past years; we refer to Ramsay & Silverman (2005, 2002); Horváth & Kokoszka (2012a); Ferraty & Vieu (2006) for comprehensive review of FDA.

Let the observed data be \{y_i, (W_{ir}, s_{ir}) : r = 1, \ldots, r_i; i = 1, \ldots, I\}, where \( y_i \) is a scalar response corresponding to subject \( i \), \( W_{ir} \) is the \( r \)th repeated measurement for subject \( i \) at point \( s_{ir} \) such that \( s_{ir} \in S \) where \( S \) is a closed compact set, and \( r_i \) is assumed to be large. We say that the observed values \( W_{ir} \)'s arise from an unknown smooth function with some noise such that \( W_{ir} = X_i(s_{ir}) + \epsilon_{ir} \), where \( X_i(\cdot) \) is a squared integrable latent process in \( L^2(S) \) with unknown mean \( \mu(\cdot) \) and covariance function \( \Sigma(\cdot, \cdot) \), and \( \epsilon_{ir} \) is a white noise with mean zero and variance \( \sigma^2 \). FDA can accommodate two sampling designs: 1) dense design where
the set of sampling points \( \{s_{i1}, \ldots, s_{ir_i}\} \) is dense in \( S \) and 2) \textit{sparse} design where the set of sampling points \( \{s_{i1}, \ldots, s_{ir_i}\} \) is irregular and random; however, it is assumed that the pooled set of sampling points across all subjects is dense in \( S \).

In spite of the rapid development in theoretical and application of FDA, most of the developed methodologies allow one functional measurements for each subject. Recently, longitudinally observed functional predictors are taken into account in the FDA literature where both the between-curves and within-curve correlation is strong for each subject. We refer to Cardot \textit{et al.} (2003a); Yao \textit{et al.} (2005); James \textit{et al.} (2009); Di \textit{et al.} (2009); Crainiceanu \textit{et al.} (2009); Staicu \textit{et al.} (2010); Greven \textit{et al.} (2010); Goldsmith \textit{et al.} (2011, 2012a); Gertheiss \textit{et al.} (2013a); Park & Staicu (2015); Fan \textit{et al.} (2015); Islam \textit{et al.} (2017) for details.

The main interest of this dissertation is to assess the systematic relationship between responses and functional predictors using regression models. There has been considerably rich literature for functional regression models in various fields such as chemometrics (Müller & Yao, 2010), gene expression (Yao \textit{et al.}, 2005), linguistic (Hastie & Tibshirani, 1993), medical research (Greven \textit{et al.}, 2010; Goldsmith \textit{et al.}, 2011), demographic forecasting (Hyndman & Shang, 2009), climate variation forecasting (Shang & Hyndman, 2011), brain imaging (Reiss & Ogden, 2010) and many others. In this dissertation, we apply functional regression techniques in animal science and biomedical engineering.

Popular functional regression models can be classified into various categories depending on the nature of predictor and responses; for instance, generalized functional linear models (Cardot \textit{et al.}, 1999; Reiss & Ogden, 2007; James, 2002; Goldsmith \textit{et al.}, 2011), functional quadratic regression models (Horváth & Kokoszka, 2012b), functional additive models (Müller & Yao, 2012; McLean \textit{et al.}, 2014), functional response models (Chiou \textit{et al.}, 2004, 2003), functional regression models (Scheipl \textit{et al.}, 2015; Ivanescu \textit{et al.}, 2015), functional multivariate regression models (Matsui \textit{et al.}, 2008), and varying functional regression models (Cardot & Sarda, 2008; Davenport, 2013; Wu \textit{et al.}, 2010). In this dissertation, we consider functional linear regression and varying functional regression models for scalar responses where functional data consists of two scenarios: 1) single functional measurement observed for each instance and 2) multiple functional observations collected at multiple instances (visit times) for each subject.
1.1 Outlines and contributions

In chapter 2, we develop a statistical methodology to study the association between scalar outcomes and functional measurements that are observed over longitudinal time. We introduce a parsimonious modeling framework to study time-varying regression that accounts for the effect of predictor(s) in such a way so that it captures the effect of a predictor varying over both functional and longitudinal time domain. The state-of-the-art longitudinal functional regression models consider non-varying smooth coefficient assuming that the effect of a predictor is invariant over time. However, such assumption is inflexible in many situations when the predictor evolves over time. We apply functional principal component analysis on an appropriate covariance function of predictor to extract the leading eigencomponents. Then we model the time-varying functional predictors using these orthonormal basis functions and expand the time-varying regression coefficient using the same bases. Using the orthonormality condition, the functional model can be reduced to a varying coefficient model where the time-varying scores play the role of regressors. It is often of interest to reconstruct the full response trajectory which is not possible using the existing procedures. To tackle this problem, we model the time-varying loadings associated with the leading eigenbasis and apply sparse functional principal component analysis on them. Such approach borrows information across subjects, captures the dependence of predictor over longitudinal time within each subject, and estimates the scores for all time-points even at the points at which covariate information is missing or not being observed. We adopt a penalized likelihood based approach for the estimation of smooth coefficients and random effects. The method is inspired by and applied to an animal science application where the main interest is to study the association between the feed intake of lactating sows and the minute-by-minute temperature throughout the 21-day lactation period. We found that due to the prolonged exposure to temperature a sow is likely to experience behavioral changes across longitudinal time; such phenomenon is reflected through the change in their feeding habit. This finding has potential in developing an optimum, cost-effective feeding strategy for sows by following the pattern of changes. The proposed methodology can also be applied in other real world applications with longitudinal setting where the functional covariate is indeed time-varying. An R function LDFR will be included in widely used refund package (Crainiceanu et al., 2012a).

Chapter 3 discusses the problem of testing functional effects in the context of scalar-on-function regression in a longitudinal setting where the effects of functional covariates vary over time. In particular, we focus on testing: 1) the adequacy of functional structure; 2) the
necessity of time-varying coefficient, 3) the pattern of change in the temporal structure of smooth coefficient, and 4) null effect of a predictor. We propose a novel, data-driven basis functions selection strategy using the cross-covariance function of the response-adjusted functional predictors; where the key idea is to account for the correlation between the responses and functional predictors relaxing the conventional assumption in the functional principal component regression (FPCR) which dictates that the leading eigenbasis functions are most informative to predict responses. We borrow the idea of modeling both the functional predictor and smooth coefficient using same orthonormal bases. We cast the model using penalized splines in a linear mixed model framework. We denoise the responses and adopt a pseudo likelihood ratio test for testing fixed and variance parameters. Extensive simulations exhibit excellent numerical performance in terms of attaining the size and power properties for the proposed test in various settings. Methods are motivated by and applied to an animal science application where the primary objective is to infer about the effect of thermal environment on sows’ feeding behavior.

Chapter 4 corresponds to a biomechanical engineering study where the main objective is to develop multi-functional robotic prosthetic limb for the transradial amputees. The commercially available artificial limbs are based on pattern recognition (PR) approach which requires extensive training of the algorithm; this method is time-consuming and hard to operate. Moreover, the rejection rate of the PR-based prostheses by users is very high. We aim to develop the prosthetic so that it can perform daily activities such as opening, closing, squeezing, and pinching at various postures and patterns using only important muscle information. In this pursuit, we propose a novel approach to develop a dynamic robotic limb by utilizing the recent history of EMG signals in a model that accounts for physiological features of hand movement which are ignored by PR algorithms. We do this by viewing EMG signals as functional covariates and develop a functional linear model that quantifies the effect of the EMG signals on finger/wrist velocity through a bivariate coefficient function that is allowed to vary with current finger/wrist position. The model is made parsimonious and interpretable through a two-step variable selection procedure, called Sequential Adaptive Functional Empirical selection by group LASSO (SAFE-gLASSO). The proposed method selects EMG signals as “all-in-all-out” manner and ensures interpretability of the regression surfaces via smoothness penalty. For our motivating dataset, the method correctly identifies the few EMG signals that are known to be important for an able-bodied subject with negligible false positives and the model can be directly implemented in a robotic prosthetic.
We also conduct post-selection predictive inference to quantify the uncertainty in prediction based on “data-splitting”. Numerical studies show excellent selection and prediction properties of SAFE-gLASSO compared to popular alternatives. Since the model has potential to accurately identify the important predictors, biomedical engineers can use this approach as a screening method to select variables for the amputees and focus only on those for subsequent analysis. This also significantly reduces the cost and time of data collection from amputees. The proposed approach can also be applied in a setting where the functional predictors are observed with or without noise having sparse or dense sampling design in a general functional variable selection problem.
Chapter 2

Longitudinal Dynamic Functional Regression

2.1 Introduction

Functional regression has attracted a lot of interest in recent years; see Ramsay & Silverman (2005); Fan & Zhang (2000); Cardot et al. (1999, 2003a); Müller (2005); Cai et al. (2006); Morris & Carroll (2006); Reiss & Ogden (2007); Ivanescu et al. (2015); Scheipl et al. (2015) to name a few. In this paper we consider longitudinal scalar-on-function regression for scalar outcomes and functional predictors observed repeatedly. This research is motivated by an animal science study of the effect of daily ambient air temperature on feed intake of sows during their lactation period. To be specific, a number of sows are observed for several days during their lactation period and for each day, the total daily feed intake, as well as, the minute-by-minute daily temperature for the respective day are recorded. Figure 2.1 illustrates the data for a randomly chosen sow: daily feed-intake for each lactation day (left panel) and temperature profiles (right).

Functional linear model (FLM) for scalar-on-function regression is a popular regression model and assumes that the effect of the functional predictor is captured by the integral of the predictor weighted by a smooth regression coefficient. Three estimation approaches are quite common: both the functional predictor and smooth coefficient are expanded using the empirical eigenbasis of the predictors covariance (Cardot et al. (1999)); both the functional predictor and smooth coefficient are expanded using B-spline basis and penalties are employed to control the smoothness of the parameter function (Ramsay & Silverman
or a mixture of these approaches, predefined basis function is used to represent the smooth parameter, the empirical eigenbasis of the predictors covariance is used to expand the functional predictors, and in addition penalties are used to control the smoothness of the parameter function (Cardot et al. (2003a); Goldsmith et al. (2011)). Extensions of these approaches to accommodate additional covariates or more flexible relationships have been discussed previously by Ramsay & Silverman (2005); Morris & Carroll (2006); Cardot & Sarda (2008); Müller & Yao (2012); McLean et al. (2014).

Figure 2.1: Data for a randomly chosen sow; feed-intake (kg) and temperature (°C) profiles from left to right respectively. Two randomly chosen curves are highlighted in the right panel and the corresponding feed-intake values are filled with colors in the left panel, respectively.

The popular scalar-on-function regression model has been extended to analyze longitudinal functional data where the scalar response and the functional predictor are observed repeatedly for each subject. Goldsmith et al. (2012a) introduced longitudinal penalized functional regression (LPFR), which assumes that the effect of the functional predictor is constant over time. The authors model the time-invariant regression coefficient via the truncated polynomial basis and use a penalized-based approach for estimation. In a similar spirit, Gertheiss et al. (2013a) introduced longitudinal functional principal component regression, where the response is regressed onto the functional principal components of the various structures that compose the functional predictors, which are obtained using longitudinal functional principal component analysis (Greven et al. (2010)). The main limitation of these methods is the assumption that the effect of the predictor is invariant over time. Such assumption may be viewed as strong and unrealistic in some situations. For example, in the motivating animal science study, a lactating sow’s body can adjust to the prolonged exposure to heat and, thus,
the effect of the heat on their feed intake behavior is expected to change throughout their lactation period. Therefore, assuming a time-varying relationship between the temperature and the feed intake is more appropriate.

Recently, Kundu et al. (2016) considered a time-varying functional coefficient in this longitudinal functional model framework. The authors model the bivariate regression coefficient as a linear combination of user specified time-dependent basis functions with unknown time-invariant coefficient functions. While these time-invariant functions are estimated via a penalty operator that is informed by the scientific context of the problem, the time-dependent basis functions are chosen arbitrarily and their number is then selected using Akaike information criterion or pointwise confidence intervals for the time-invariant coefficient functions. In simulations and data application, the authors use two temporal functions $f_1(t) = 1$ and $f_2(t) = t$; their methodology is implemented solely for polynomial temporal functions. In general it is not clear how to select an optimal temporal basis. In addition, the methodology is limited to Gaussian responses and it does not seem trivial to extend it to non-Gaussian cases. Our numerical experience with this method indicated that, when applicable, the approach is rather computationally expensive. Furthermore, like LPFR and longitudinal functional principal component regression, this method too does not consider reconstruction of full response trajectory, which is often a major goal in longitudinal studies involving a repeatedly measured response.

In this paper, we propose the longitudinal dynamic functional regression (in short LDFR) for longitudinal scalar-on-function regression that accounts for a time-varying smooth effect of the functional predictors. There are three key novelties of this paper. The first novelty is the proposed model for both Gaussian and non-Gaussian longitudinal responses. Only Kundu et al. (2016) studied regression models for longitudinal functional covariates that allow a time-varying effect of the covariates, but their study is limited to Gaussian responses. The second novelty is the use of a combination of ideas from functional data analysis and longitudinal data analysis that has important advantages: 1) it allows to tackle a challenging problem that has not previously been solved in this generality; 2) it takes into account the dependence among the functional predictors; and 3) it allows prediction of the full trajectory of the response. The third novelty is the way the smoothness of the time-varying regression coefficient is modeled, by using a mixture of popular approaches in non-parametric regression. The main advantages of this approach are: (i) it enjoys a parsimonious modeling framework; (ii) it does not require any information about the temporal behavior of the
functional coefficient; (iii) it provides excellent numerical performance in terms of prediction accuracy; (iv) it is computationally efficient; and (v) it can be easily implemented using well-developed freely available software.

The structure of this paper is as follows. Section 2.2 introduces the proposed methodology for responses in the exponential family. Section 2.3 describes the estimation procedure of the parameters of interest. Section 2.4 details the response trajectory prediction. Numerical assessment of the proposed method is described in Section 2.5 in a simulation study and in Section 2.6, via our motivating application. Finally, Section 2.7 concludes our paper.

2.2 Proposed methodology

Let the observed data be \([t_{ij}, Y_{ij}, \{(W_{ijr}, s_{ijr}) : r = 1, \ldots, r_{ij}, i = 1, \ldots, I, j = 1, 2, \ldots, n_i\} : i = 1, \ldots, I, j = 1, 2, \ldots, n_i]\); where \(i\) indexes the subject, \(j\) indexes the repeated observations, \(Y_{ij}\) is the response measured at time \(t_{ij}\), and \(W_{ijr}\)'s are the noisy functional covariates observed at points \(s_{ijr}\). It is assumed that \(t_{ij} \in \mathcal{T}\) and \(s_{ijr} \in \mathcal{S}\) for closed and compact sets \(\mathcal{T}\) and \(\mathcal{S}\), respectively. We assume that for each \(i\) and \(j\), \(r_{ij}\) is large, and furthermore that the set \(\{s_{ij1}, \ldots, s_{ijr_{ij}}\}\) is dense in \(\mathcal{S}\). Also we assume that, while \(n_i\) may be small for all \(i\), the set \(\{t_{ij} : j = 1, \ldots, n_i, i = 1, \ldots, I\}\) is dense in \(\mathcal{T}\). We consider that \(W_{ijr} = X_{ij}(s_{ijr}) + \epsilon_{ij}^w(s_{ijr})\), where \(X_{ij}(\cdot)\) is the latent functional predictor corresponding to the subject \(i\) and (visit) time \(t_{ij}\), and \(\epsilon_{ij}^w(\cdot)\) is the zero-mean error process corresponding to the time \(t_{ij}\). By an abuse of notation, we write \(X_{ij}(\cdot) = X_i(\cdot, t_{ij})\), by suppressing the time dependence of the latent subject-specific signal. It is assumed that \(X_i(\cdot, \cdot)\)'s are independent and identically distributed (IID) over \(i\), and are comprised of the subject-specific deviation. In contrast \(\epsilon_{ij}(\cdot)\)'s are IID over \(i\) and \(j\), and characterize the time-specific deviation from the subject-specific trend. For convenience of illustration, we assume that the points \(s_{ijr}\) are same across \(i\) and \(j\); however the method is applicable for different \(s_{ijr}\)'s without loss of generality.

Our objective is to develop association models for time-varying responses \(Y_{ij} = Y_i(t_{ij})\) and true smooth time-varying functional covariate \(X_{ij}(\cdot) = X_i(\cdot, t_{ij})\). Specifically, we consider longitudinal dynamic functional regression models that account for time-varying smooth effect:

\[
Y_{ij} | X_{ij}(\cdot) \sim \text{EF}(\mu_{ij}, \eta),
\]

\[
g(\mu_{ij}) = \alpha(t_{ij}) + \int_{\mathcal{S}} X_{ij}(s) \gamma(s, t_{ij}) ds + Z_{b_{ij}} b_i,
\]

where \(\text{EF}(\mu_{ij}, \eta)\) denotes the exponential family with mean \(\mu_{ij} = \mu_i(t_{ij})\) and dispersion
parameter $\eta$ and $g(\cdot)$ is a known, monotone link function. Here, $\alpha(\cdot)$ is an unknown intercept function, and $\gamma(\cdot, t_{ij})$ is an unknown regression coefficient function that quantifies the time-varying effect of $X_{ij}(\cdot)$ on the conditional mean response of $Y_{ij}$ at time $t_{ij}$ and is the main object of interest. Also $b_i$ is a subject specific $q$-dimensional vector and $Z_{k,ij}$ is its associated $1 \times q$-dimensional random design matrix. It is assumed that $b_i$ is IID as $N_q(0_q, D)$ where $0_q$ is the $q$-dimensional vector of zeros and $D$ is $q \times q$ covariance matrix. Both the intercept function $\alpha(\cdot)$ and the regression coefficient function $\gamma(\cdot, \cdot)$ are assumed to be smooth. Remark that, just like the functional linear model, the regression coefficient $\gamma(\cdot, t)$ is only identifiable up to the space spanned by the $X_{ij}(\cdot)$’s. Equivalently, if $h(\cdot)$ is in the orthogonal complement of this space then both $\gamma(s, \cdot)$ and $\gamma(s, \cdot) + h(s)$ yield the same association with the response; nevertheless the integral term is identifiable. Model (2.1) has been introduced in Kundu et al. (2016). The key difference is that Kundu et al. (2016) assume that $\gamma(s, \cdot)$ is a linear combination of known, but arbitrarily chosen, temporal functions; for example $\gamma(s, t) = \gamma_0(s) + t \gamma_1(s) + t^2 \gamma_2(s)$ with unknown functions $\gamma_0(\cdot), \gamma_1(\cdot),$ and $\gamma_2(\cdot).$ We impose no such limitation and propose a parsimonious modeling framework that results in very competitive prediction performance and computational time. For convenience, assume $s_{ijr} = s_r.$

Our modeling approach consists of two parts. First, let $\{\phi_k(\cdot) : k \geq 1\}$ be a time-invariant orthonormal basis in $L^2(S);$ i.e. $\int_S \phi_k(s) \phi_{k'}(s) ds = 1$ if $k = k'$ and 0 otherwise. Consider the expansion of the functional regression coefficient $\gamma(\cdot, t_{ij})$ using this same basis function $\gamma(s, t) = \sum_{k=1}^{\infty} \beta_k(t) \phi_k(s),$ where $\beta_k(\cdot)$ is an unknown smooth function of time and defined uniquely by $\beta_k(t) = \int_S \gamma(s, t) \phi_k(s) ds.$ Often, the infinite summation is truncated at some finite level, say $K$ so that $\gamma(s, t) \approx \sum_{k=1}^{K} \beta_k(t) \phi_k(s).$ In some sense the truncation $K$ quantifies the smoothness of the function $\gamma(s, t)$ in the $s$-direction: a small value of $K$ results in over smooth function in this direction, while a large value gives wiggly behavior in this direction. Assuming that $\gamma(\cdot, \cdot)$ is a second order differentiable function, we can describe the smoothness in the $t$-direction by the total curvature in this direction, $\|\partial^2 \gamma(\cdot, \cdot)/\partial t^2\|^2 = \int \int \{\partial^2 \gamma(s, t)/\partial t^2\} ds dt.$ A direct extension to the popular non-parametric estimation approach of an unknown smooth is to control the amount of smoothness in $t,$ that is to estimate $\gamma(\cdot, \cdot)$ via the minimization of a penalized criterion $-2 \sum_{i=1}^{L} \sum_{j=1}^{n_i} \log f(Y_{ij}|b_i) - 2 \sum_{i=1}^{L} \log f_{b}(b_i) + \lambda_0 \|\alpha''(\cdot)\|^2 + \lambda \|\partial^2 \gamma(\cdot, \cdot)/\partial t^2\|^2,$ where $f(Y_{ij}|b_i)$ is the density specified by the exponential family model (2.1) and $f_{b}(b_i)$ is the density of the random terms. Typically $f_{b}(b) = \exp(-b^T D^{-1} b/2),$ corresponding to a multivariate normal with mean vector 0 and variance-covariance matrix $D,$ and by ignoring the multiplicative
term \( \{ \text{det}(2\pi D) \}^{-1/2} \) where \( \text{det}(D) \) is the determinant of \( D \). Here \( \lambda_0 > 0 \) is a smoothing parameter that controls the amount of smoothing of the unknown function \( \alpha(\cdot) \) and \( \lambda > 0 \) is an unknown parameter that controls the smoothness of \( \gamma(\cdot, \cdot) \) in direction \( t \), relative to the goodness of fit, quantified by the likelihood term. Using the representation of \( \gamma(\cdot, \cdot) \) via \( K \) orthogonal basis functions \( \phi_k(\cdot) \)'s, the penalized criterion is approximated by

\[
-2 \sum_{i=1}^{I} \sum_{j=1}^{n_i} \log f(Y_{ij}|b_i) - 2 \sum_{i=1}^{I} \log f(b_i) + \lambda_0 \| \alpha''(\cdot) \|^2 + \lambda \sum_{k=1}^{K} \| \beta_k''(\cdot) \|^2,
\]

as \( \| \partial^2 \gamma(\cdot, \cdot) / \partial t^2 \|^2 \approx \sum_{k=1}^{K} \| \beta_k''(\cdot) \|^2 \) due to the orthogonality of the basis \( \phi_k(\cdot) \)'s. The penalized criterion (2.2) essentially says that we control the smoothness of the unknown bivariate function \( \gamma(\cdot, \cdot) \) through two parameters - \( K \) and \( \lambda \) - but in a distinct way from the common practice (Ruppert et al., 2003; Wood, 2017, 2006b). Combining different approaches to control the smoothness of a multivariate function is inspired from Kim et al. (2017) in the context of function-on-function regression.

Using the orthogonal time-invariant basis \( \{ \phi_k(\cdot) \} \)'s we represent the longitudinal functional covariates as \( X_{ij}(s) = \sum_{k=1}^{\infty} \xi_{ijk} \phi_k(s), \) where \( \xi_{ijk} = \int_{S} X_{ij}(s) \phi_k(s) ds \). The time-varying basis coefficients \( \xi_{ijk} \) are dependent over \( j \), due to the dependence of the functional covariates within the same subject. It follows that \( \int_{S} X_{ij}(s) \gamma(s, t_{ij}) ds = \sum_{k=1}^{\infty} \xi_{ijk} \beta_k(t_{ij}) \), which yields the following more convenient representation of the model (2.1): \( g(\mu_{ij}) = \alpha(t_{ij}) + \sum_{k=1}^{\infty} \xi_{ijk} \beta_k(t_{ij}) + Z_{b,ij} b_i \). Corresponding to the truncation \( K \), the model (2.1) is approximated by:

\[
g(\mu_{ij}) = \alpha(t_{ij}) + \sum_{k=1}^{K} \xi_{ijk} \beta_k(t_{ij}) + Z_{b,ij} b_i,
\]

which is a well researched model in the longitudinal literature, if \( \xi_{ijk} \) were known. The model parameters of (2.3) can be estimated using the penalized criterion (2.2), which assumes that the coefficient functions \( \{ \beta_k(\cdot) \} \)'s have all the same type of smoothness.

Incorporating additional covariates via the modeling framework (2.1) carries on in a straightforward manner to (2.2) and (2.3), irrespective whether the covariates are modeled using a linear or a smooth dependence. There are two key challenges in this approach: 1) selection of the orthogonal basis \( \{ \phi_k(\cdot) \}_{k=1} \) as this directly impacts the selection of the truncation \( K \) and 2) the estimation of the basis coefficients \( \xi_{ijk} \) from the observed noisy functional covariates \( W_{ij}(\cdot) \).
2.2.1 Selection of the orthogonal basis

There are several approaches to select $\phi_k(\cdot)$'s. One approach is using a pre-specified basis similar to Zhou et al. (2008). Another approach is using the eigenbasis of some appropriately chosen covariance function; see Park & Staicu (2015). Specifically we consider the marginal covariance function induced by the observed functional covariates and select the basis as the eigenbasis of this covariance function.

Recall that the functional covariate is viewed as the sum of two independent processes $W_{ij}(s) = X_{ij}(s) + \epsilon_{ij}^w(s)$, where by an abuse of notation we write $W_{ij}(s_r) = W_{ijr}$ and $X_{ij}(\cdot) = X_i(\cdot, t_{ij})$. Both $X_i(\cdot, \cdot)$ and $\epsilon_{ij}^w(\cdot)$ are assumed to be zero-mean processes. Define $\Sigma(s, s') = \int_T E[X_i(s, t)X_i(s', t)]h(t)dt$, where $h(\cdot)$ is the sampling density of the time points $t_{ij}$'s; see Park & Staicu (2015) for justification that this function is a proper covariance function (positive semidefinite and symmetric function). We call this the ”marginal covariance function” induced by the latent signal $X_i$. Assume the covariance of the error process can be written as the sum between a smooth covariance function and a nugget effect such as $\text{cov}(\epsilon_{ij}^w(s), \epsilon_{ij}^w(s')) = \Gamma(s, s') + \sigma^2_w \mathbf{1}(s = s')$. Essentially this assumption means that the error process can be represented as the sum between an error component with smooth covariance function and an IID white noise component. Let $\Xi(s, s') = \Sigma(s, s') + \Gamma(s, s')$, which is too a proper covariance function, and denote by $\{\phi_k(\cdot), \lambda_k\}_k$ its eigen-components. Using this basis we represent $X_{ij}(\cdot)$ by $X_{ij}(\cdot) = \sum_{k=1}^K \xi_{ijk} \phi_k(\cdot)$ where the basis coefficients are $\xi_{ijk} = \int_S X_{ij}(s) \phi_k(s)ds$. Let $K$ be a finite truncation; this approach implicitly assumes that the $K$ main eigenbasis functions are the most informative to predict the response. The assumption, that the components with the largest variation are most predictive of the dependent variable, is rooted in the principal component regression literature (Mardia et al., 1979), and has been commonly employed in functional regression (Reiss & Ogden, 2007; Crainiceanu et al., 2009; Febrero-Bande et al., 2017). Nevertheless it may be viewed as a strong limitation, and future research is needed to investigate alternative approaches to select the orthogonal basis in a manner that appropriately accounts for the correlation between the functional predictor and response. Specifically, the truncation value $K$ is determined based on the pre-specified percentage of variance explained (PVE) value where the main idea is to choose the smallest integer $K$ such that $\frac{\sum_{k=1}^K \lambda_k}{\sum_{k=1}^\infty \lambda_k}$ is larger than the pre-specified PVE value.
2.2.2 Statistical modeling of the non-linear effects

The univariate functions in model (2.3), $\alpha(t)$ and $\beta_1(t) \ldots \beta_K(t)$ are unknown smooth functions. Assume for now that $X_{ij}(\cdot)$’s and furthermore $\xi_{ijk}$’s are known. The implied approximating model is known in the statistical literature as a time-varying coefficient model (Hastie & Tibshirani, 1993; Hoover et al., 1998). We briefly review it next and focus on how we ensure that the $K$ regression coefficients have the same smoothness.

We use basis expansions - the truncated polynomial splines, B-splines or Fourier basis etc - to model the smooth parameter functions in (2.3). Let $\{B_{0l}(t)\}_l$’s and $\{B_{kl}(t)\}_l$’s be such bases and let $\alpha(t) = \sum_{l=0}^{L} \beta_{0l} B_{0l}(t)$ and $\beta_k(t) = \sum_{l=1}^{L} \beta_{kl} B_{kl}(t)$. For simplicity of exposition, we illustrate on truncated polynomial spline basis and take the bases to be the same, that is $L_{\alpha 1} = \ldots = L_{\alpha K}$ and $B_{0l}(\cdot) = B_{kl}(\cdot) = B_l(\cdot)$ for all $l \geq 1$. Let $\alpha(t) = \beta_{00} + \beta_{01} t + \ldots + \beta_{0p} t^p + \sum_{i=1}^{L} u_{0l}(t - \kappa_i)^p$ and $\beta_k(t) = \beta_{k0} + \beta_{k1} t + \ldots + \beta_{kp} t^p + \sum_{i=1}^{L} u_{kl}(t - \kappa_i)^p$, where $\beta_{k0}, \ldots, \beta_{kp}$’s are unknown fixed parameters, $u_{kl}$’s are independent random variables. Here $\kappa_1, \ldots, \kappa_L$ are knots in $\mathcal{T}$ and $(x)^p = \max(0, x^p)$. Typically, the coefficients of the non-polynomial functions are assumed to vary according to $u_{kl} \sim N(0, \sigma_k^2)$ for $k = 0, 1, \ldots, K$, where the variance $\sigma_k^2$ controls the smoothing of the unknown functions, $\alpha(\cdot)$ or $\beta(\cdot)$’s; see Ruppert et al. (2003). As argued in Section 2.2, estimating the unknown function using the penalized criterion (2.2) entails assuming same smoothness for the functions $\beta_k(\cdot)$’s for $k = 1, \ldots, K$ or equivalently $\sigma_k^2 = \sigma^2$ for $k \geq 1$, $\sigma^2$ denotes their common value.

This yields the following mixed effects representation of $g(\mu_{ij}) = V_{ij} \beta + Z_{ij,0} u_0 + \xi_{ij1} Z_{ij,1} u_1 + \ldots + \xi_{ijK} Z_{ij,K} u_K$, where $V_{ij} = [1, t_{ij}, \ldots, t_{ij}^p, \xi_{ij1}, t_{ij} \xi_{ij1}, \ldots, t_{ij}^p \xi_{ijK}]$ is a $(p + 1)(K + 1)$-dimensional row vector, $\beta = (\beta_{00}, \beta_{01}, \ldots, \beta_{0p}, \beta_{10}, \beta_{11}, \ldots, \beta_{kp})^T$ is the full vector of fixed effects. Also let $Z_{ij,k}$ is the $L$-dimensional row vector of $(t_{ij} - \kappa_i)$’s and $u_k = (u_{k1}, \ldots, u_{KL})^T$ be the vector of random effects. Then $u_0 \sim N(0_L, \sigma_0^2 I_L)$ and $u_k \sim N(0_L, \sigma^2 I_L)$ for $k = 1, \ldots, K$, where $I_L$ is the $L \times L$ identity matrix. By an abuse of notation let $\xi_{ij0} = 1$ for all $i$ and $j$; then (2.3) becomes

$$g(\mu_{ij}) = V_{ij} \beta + \sum_{k=0}^{K} \xi_{ijk} Z_{ij,k} u_k + Z_{b,ij} b_i. \quad (2.4)$$

2.3 Estimation

In this section we detail the estimation of the model components, which is separated into covariates-related components, such as $\phi_k(\cdot)$ and $\xi_{ik}(\cdot)$, and response-related components, such as $\beta$, $u$, and $b$. Prediction of the response trajectory $Y_i(\cdot)$ is detailed in Section 2.4.
2.3.1 Estimation of the covariates-related components

Modeling the functional covariates is done as in Park & Staicu (2015). We briefly describe it here for completeness. We model the mean of $W_{ij}(s)$ at time $t_{ij}$ as a bivariate smooth function, and fit a bivariate smoother to estimate it, under a working independence assumption (Wood, 2017). We then demean the observed functional predictor; denote the demeaned data by $\tilde{W}_{ij}()$. Next $\tilde{W}_{ij}()$’s is used to estimate the marginal covariance function $\Xi(s,s') = \Sigma(s,s') + \Gamma(s,s')$. The pooled sample covariance, defined as $\tilde{\Xi}(s_r,s_{r'}) = \sum_{i=1}^{I} \sum_{j=1}^{n_i} \tilde{W}_{ijr} \tilde{W}_{ijr'}/(\sum_{i=1}^{I} n_i)$, is a method of moments estimator of $\Sigma(s_r,s_{r'}) + \Gamma(s_r,s_{r'}) + \sigma^2_w I(r = r')$. This estimator is not smooth and may be viewed as a raw estimator of $\Xi(s_r,s_{r'})$. The diagonal terms of $\tilde{\Xi}(s_r,s_{r'})$ are possibly inflated. One can ignore the diagonal terms and smooth the off-diagonal terms using a bivariate smoother (Yao et al., 2005; Xiao et al., 2013); denote the covariance estimator by $\tilde{\Xi}(s,s')$. We use (Xiao et al., 2013) for our numerical investigation. We estimate the eigen-components of $\tilde{\Xi}(s,s')$ by the eigen-components of $\tilde{\Xi}(s,s')\{\tilde{\phi}_k(\cdot), \tilde{\lambda}_k\}_k$, where $\int S \tilde{\phi}_k(s) \tilde{\phi}_k'(s) ds = 1$ if $k = k'$ and 0 otherwise, and $\tilde{\lambda}_1 \geq \tilde{\lambda}_2 \geq \cdots \geq 0$. Let $K$ be so that the first $K$ pairs provide a low-rank approximation of $\tilde{\Xi}(s,s')$: $\tilde{\Xi}(s,s') \approx \sum_{k=1}^{K} \tilde{\lambda}_k \tilde{\phi}_k(s) \tilde{\phi}_k(s')$. Using numerical integration, the time-varying loadings $\tilde{\xi}_{W,ijk}'$’s are estimated as $\tilde{\xi}_{W,ijk} = \int S \tilde{W}_{ij}(s) \tilde{\phi}_k(s) ds$ for $k = 1, \ldots, K$. Nevertheless these quantities are noisy estimates of $\xi_{ijk}$ and regressing the response directly onto them would lead to increased bias in the estimates. In addition, they correspond to the times $t_{ij}$ solely and would not be suitable to be used when predicting the response trajectory $Y_i(t)$ for any time point $t$ is of interest. We propose to model $\tilde{\xi}_{W,ijk}$ in a way that explicitly recognizes the dependence on the time $t_{ij}$.

Consider the working model $\tilde{\xi}_{W,ijk} = \xi_{ik}(t_{ij}) + \epsilon_{W,ijk}$, where $\xi_{ik}(\cdot)$ is a random curve with zero mean and covariance function $G_k(\cdot,\cdot)$ such that $G_k(t_{ij},t_{ij'}) = \text{cov}\{\xi_{ijk}, \xi_{ij'k}\}$, and $\epsilon_{W,ijk}$ is white noise with zero mean and finite variance $\sigma^2_{w,k}$. Here $\xi_{ijk} = \xi_{ik}(t_{ij})$. Recovering the trajectories $\xi_{ik}(\cdot)$ requires modeling and estimation of their covariance function. In this regard, we use the pseudo data $\{\tilde{\xi}_{W,ijk}(t_{ij}) : j = 1, \ldots, n_i\}_i$, separately for each $k$. One simple approach is to assume a parametric covariance model, such as exponential, or Matérn, or random effects based models; see Park & Staicu (2015) for more discussion. Standard methods in longitudinal data analysis can be used to estimate the covariance model. Here we consider a flexible nonparametric covariance model as it is common in functional data analysis and employ common techniques in sparse functional principal components analysis (Yao et al., 2005) to estimate it; this approach was described by Park & Staicu (2015).
The spectral decomposition of $G_k(\cdot, \cdot)$ is $G_k(t, t') = \sum_{l \geq 1} \eta_{kl} \psi_{kl}(t) \psi_{kl}(t')$, where $\{\eta_{kl}, \psi_{kl}(\cdot)\}$ is the pair of eigenvalues and eigenfunctions for $\eta_{k1} \geq \eta_{k2} \geq \cdots > 0$, and $\{\psi_{kl}(\cdot)\}_{l \geq 1}$ are mutually orthogonal, and have unit norm in $L^2(T)$. Using the truncated Karhunen-Loève (KL) expansion $\xi_{ik}(t) = \sum_{l \geq 1} \zeta_{ikl} \psi_{kl}(t)$, where $\zeta_{ik} = \int_T \xi_{ik}(t) \psi_{kl}(t) dt$ is random, with zero-mean and variance equal to $\eta_{kl}$.

Let $\hat{G}_k(\cdot, \cdot)$ be a covariance estimator of $G_k(\cdot, \cdot)$ obtained as Yao et al. (2005); Crainiceanu et al. (2009). The spectral decomposition of $\hat{G}_k(t, t')$, $\hat{G}_k(t, t') \approx \sum_{l \geq 1} \hat{\eta}_{kl} \hat{\psi}_{kl}(t) \hat{\psi}_{kl}(t')$ yields orthogonal functions that have unit norm, $\hat{\psi}_{kl}(\cdot)$'s and non-negative eigenvalues, $\hat{\eta}_{kl}$'s. Here $L_k$'s are truncation values that are chosen in similar style as $K$. The time-varying loadings are estimated using the truncated Karhunen-Loève expansion $\hat{\xi}_{ik}(t) = \sum_{l \geq 1} \hat{\zeta}_{ikl} \hat{\psi}_{kl}(t)$; the scores $\hat{\zeta}_{ikl}$ are obtained via conditional expectation $E[\zeta_{ikl}|\tilde{\xi}_{W,ik}, \ldots, \tilde{\xi}_{W,in,k}]$ in the associated mixed effects model $\tilde{\xi}_{W,ijk} = \sum_{l \geq 1} \zeta_{ikl} \psi_{kl}(t) + e_{W,ijk}$ using a working Gaussian response assumption.

### 2.3.2 Estimation of the response-related components

The estimation of the model parameters $\alpha(\cdot)$ and $\beta_k(\cdot)$'s for $k = 1, \ldots, K$ in (2.3), using the basis representation described in the Section 2.2.2, entails estimation of $\beta$'s and $u_k$'s in (2.4), where $\xi_{ijk}$ are replaced by $\tilde{\xi}_{ik}(t_{ij})$. For exposition simplicity denote by $\tilde{Z}_{ij,k} = \tilde{\xi}_{ik}(t_{ij}) Z_{ij,k}$ for all $i, j$ and $k$ and let $\tilde{V}_{ij}$ be the vector $V_{ij}$ with $\tilde{\xi}_{ik}(t_{ij})$'s used in place of $\xi_{ijk}$'s.

It follows that $g(\mu_{ij}) = \tilde{V}_{ij}\beta + \sum_{k=0}^{K} \tilde{Z}_{ij,k} u_k + Z_{b,ij} b_i$; remark that the subject specific effects $b_i$'s account for the dependence across repeated observations within the same subject. Then the model for $Y$ can be approximated by the following generalized mixed effects model

$$
Y \sim EF(\mu, \eta),
\mu_g = \tilde{V} \beta + \tilde{Z} u + Z_{b} b,
\quad u \sim N \left( 0_{L+LK}, \text{diag} \{\sigma^2_0, \sigma^2_1, \ldots, \sigma^2_n \} \otimes I_L \right) \quad \text{and} \quad b \sim N \left( 0_{Iq}, I_I \otimes D \right)
$$

(2.5)

$u$ and $b$ are mutually independent

where $\mu_g$ is obtained by columnwise stacking $(g(\mu_{i1}), \ldots, g(\mu_{in}))^T$, $\tilde{V}$ is obtained by stacking columnwise $\tilde{V}_{ij}$ first over $j$ and then over $i$. Here $u = (u_0^T | u_1^T | \ldots | u_K^T)^T$ and $b = (b_1^T | \ldots | b_f^T)^T$, $\tilde{Z} = (\tilde{Z}_0 | \tilde{Z}_1 | \ldots | \tilde{Z}_K)$ with $\tilde{Z}_k$ obtained like $\tilde{V}$ by stacking columnwise $Z_{ij,k}$ over $j$ and $i$ and $Z_{b} = \text{diag} \{Z_{b,1}, \ldots, Z_{b,I} \}$ and $Z_{b,ij}$ is obtained by stacking columnwise $Z_{b,ij}$ over $j = 1, \ldots, n_i$.

Once the model is represented in this form, parameter estimation and quantification of their estimation uncertainty follows easily.
For given values of the covariance parameters, $\sigma_0^2$, $\sigma^2$, and $D$, the estimates of $\beta$, $u_k$’s and $b_i$’s are the same as the minimizers of the following penalized criterion:

$$pl(\beta, u_0, \ldots, u_K, b, \eta) = -2 \sum_{i=1}^{I} \sum_{j=1}^{n_i} \ell_{ij}(\beta, u_0, \ldots, u_K, b, \eta) + b^T (I_I \otimes D^{-1}) b + \lambda_0 \|u_0\|^2 + \lambda \sum_{k=1}^{K} \|u_k\|^2,$$

(2.6)

where $\ell_{ij}(\beta, u_0, \ldots, u_K, b, \eta)$ is the log-likelihood function corresponding to the assumed conditional model for $Y_{ij}$, $\lambda_0 = 1/\sigma_0^2$, and $\lambda = 1/\sigma^2$. Modeling the smoothness of the unknown functions explicitly, as the inverse of a variance component, allows us to clearly describe the corresponding generalized mixed effects model (2.5). The criterion (2.6) can be easily modified to account for other bases and associated penalties: other choices of bases will modify the term $\tilde{V}_{ij} + \sum_{k=0}^{K} \tilde{Z}_{ij,k} u_k$ that appears in the expression of $\ell_{ij}(\beta, u_0, \ldots, u_K, b, \eta)$, while the associated penalties will modify the term $\lambda_0 \|u_0\|^2 + \lambda \sum_{k=1}^{K} \|u_k\|^2$. The smoothing parameters $\lambda_0$ and $\lambda$ will continue to have the same interpretation; see Wood (2017) and Ivanescu et al. (2015).

To estimate the variance parameters, a Bayesian perspective where the parameters are estimated using the log of a corresponding marginal likelihood, is more appealing. The ideas are described in Wood (2011) and Wood et al. (2016) and rely on using the Laplace approximation to calculate the desired marginal log-likelihood. When the responses are Gaussian, the implied marginal log-likelihood corresponds to the restricted maximum likelihood (REML). For both our simulation study and data application we used REML to select the smoothing parameters.

Using the parameter estimates from (2.6) we obtain estimates of the intercept function $\hat{\alpha}(t) = \hat{\beta}_{00} + \ldots + \hat{\beta}_{0p} t^p + \sum_{l=1}^{L} \hat{u}_{0l} (t - \kappa_l)$, and of the regression bivariate function

$$\hat{\gamma}(s, t) = \sum_{k=1}^{K} \hat{\phi}_k(s) \hat{\beta}_k(t),$$

(2.7)

where $\hat{\beta}_k(t) = \hat{\beta}_{k0} + \ldots + \hat{\beta}_{kp} t^p + \sum_{l=1}^{L} \hat{u}_{kl}(t - \kappa_l)$, for $k = 1, \ldots, K$. Then, conditional on the estimates from the pre-processing of the functional covariates, uncertainty quantification for $\hat{\alpha}(\cdot)$ and $\hat{\gamma}(\cdot, \cdot)$ is readily available. However, inference for $\gamma(s, t)$ is meaningless, due to the lack of parameter identifiability in the model (2.1).

In this paper we focus on response prediction and quantify the prediction uncertainty. There are many sources of uncertainty: the estimation of the basis functions $\phi_k(\cdot)$, of the basis functions coefficients $\xi_{ijk}$’s and of their covariance estimators, as well as the various
truncations $K$ and $L_k$'s. It is not clear how to account for all these sources in estimating the uncertainty of prediction. In the next section we discuss prediction and its associated inference, conditional on all these quantities.

2.4 Prediction and inference

2.4.1 Prediction of response trajectories

One of our main aim is to predict response trajectories in two settings: for an existing data subject, $i$, who has been observed at few sparse time points $t_{ij}$'s, and for a new subject, $i^*$ whose only functional observations are available. In general, the response for an existing data subject at an observed time $t_{ij}, Y_{ij}$, can be predicted by directly substituting the estimates of the parameters and the predicted random effects, $\hat{\beta}_i$ into equation (2.5); the specifics depend on the form of the link function. Irrespective of whether prior response data has been observed for a subject, prediction of a subject’s response trajectory $Y_i(t)$ for all $t$ requires estimation of the subject’s time-varying basis coefficients trajectories $\xi_{ik}(t)$; recall $\xi_{ik}(t_{ij}) = \xi_{ijk}$.

Let $\{W_i(\cdot, t_{ij}), j = 1, \ldots, n_i\}$ be the noisy functional covariate for a subject already in the data, and denote by $Y_{i1}, \ldots, Y_{in_i}$ the associated responses; it is assumed that $Y_{ij} = Y_i(t_{ij})$, $W_i(\cdot, t)$ is a noisy measurement of $X_i(\cdot, t)$, and that $Y_i(t)$ relates to $X_i(\cdot, t)$ through the model $Y_i(t) \sim EF(\mu_i(t), \eta)$, here $g(\mu_i(t))$ is a monotone link function for a generic time $t$ and is defined as before.

The subject mean response trajectory for an existing data subject, $\mu_i(t)$, is predicted by:

$$\bar{\mu}_i(t) = g^{-1}\{\bar{\alpha}(t) + \sum_{k=1}^{K} \bar{\xi}_{ik}(t)\bar{\beta}_k(t) + Z_{b,lt}\hat{\beta}_l\}, \tag{2.8}$$

where $g^{-1}$ is the inverse function of $g$. For Gaussian responses, $g(\mu) = \mu$, expression (2.8) can be used to predict subject trajectories $\bar{Y}_i(t) = \bar{\mu}_i(t)$. We conjecture that this predicted subject trajectory is a consistent estimator of $\bar{Y}_i(t) = E[Y_i(t)|W_i]$. In the case considered here, $\bar{Y}_i(t) = \alpha(t) + \sum_{k=1}^{K} \bar{\xi}_{ik}(t)\beta_k(t) + Z_{b,lt}\hat{\beta}_l$, where $\bar{\xi}_{ik}(t) = \sum_{l=1}^{L_k} \psi_{kl}(t)\tilde{\zeta}_{ikl}$ and $\tilde{\zeta}_{ikl} = E[\zeta_{ikl}|\xi_{ijk}: j = 1, \ldots, n_i]$. When the responses are non-Gaussian, prediction of the subject specific trajectories is not always clearly defined. For Bernoulli responses a common approach is to predict $\bar{Y}_i(t) = 1$ if $\bar{\mu}_i(t) \geq 0.5$ and predict $\bar{Y}_i(t) = 0$, if $\bar{\mu}_i(t) < 0.5$.

For a new subject $i^*$, conditional on the functional covariates $\{W_{i^*j}(\cdot) = W_{i^*}(\cdot, t_{i^*j}): j\}$,
the mean response trajectory is predicted as \( \hat{\mu}_i(t) = g^{-1}\{\hat{\alpha}(t) + \sum_{k=1}^K \hat{\xi}_{i,k}(t) \hat{\beta}_k(t)\} \), where the time-varying trajectories \( \hat{\xi}_{i,k}(t) \) are obtained as presented in Section 2.3.1 using the noisy pseudo-data \( \tilde{\xi}_{W,i,jk} = \int \tilde{W}_{i,j}(s) \tilde{\gamma}_k(s) ds \) and assuming a working model \( \tilde{\xi}_{W,i,jk} = \sum_{l=1}^{L_k} \tilde{\psi}_{kl}(t) \zeta_{i,kl} + \epsilon_{i,jk} \), where \( \zeta_{i,kl} \sim N(0, \tilde{\eta}_{kl}) \) and \( \epsilon_{i,jk} \sim N(0, \sigma^2_{\epsilon_{jk}}) \). The response trajectory does not involve subject-specific effects \( b_i \), as their estimation requires availability of response data at repeated times.

### 2.4.2 Asymptotic prediction bands

For Gaussian responses, we can construct asymptotic prediction bands for the individual response trajectory, conditional on the underlying predictor function. The prediction bands do not account for the variability associated with the estimation of the basis functions \( \{\hat{\phi}_k(\cdot) : k = 1, \ldots, K\} \), the truncation \( K \), \( \{\hat{\psi}_{kl}(\cdot) : l \geq 1\} \), \( \{\hat{\eta}_{kl} : l \geq 1\} \), and \( L_k \) for \( k = 1, \ldots, K \).

Let \( \{W_{ij}(\cdot), t_{ij} : j = 1, \ldots, n_i\} \) be the observed functional covariates for a subject \( i \). The uncertainty in prediction is measured by the prediction error (Ruppert et al., 2003) \( \{\hat{Y}_i(t) - Y_i(t)\} \); for a new subject, or in the case of an existing subject for \( t \notin \{t_{i1}, \ldots, t_{im_i}\} \) we have

\[
\text{Var}\{\hat{Y}_i(t) - Y_i(t)\} = \text{Var}\{\hat{Y}_i(t)\} + \text{Var}\{\epsilon_{it}\}.
\]

The variance of white noise \( \epsilon_{it} \) is estimated using REML along with the other variance parameters; see Section 2.3.2. The variance of \( \{\hat{Y}_i(t)\} \) is estimated with standard approaches in longitudinal data analysis (Ruppert et al., 2003; Wood, 2017, 2006b); the estimation is implemented in various computer packages and we discuss it in the Supplementary Material, Section A.3. Thus a 100(1 - \( \alpha \))% pointwise prediction interval for \( Y_i(t) \) is \( \hat{Y}_i(t) \pm z_{\alpha/2} \hat{SE}\{\hat{Y}_i(t) - Y_i(t)\} \), where \( z_{\alpha/2} \) is the \( \alpha/2 \) upper quantile of the standard normal distribution. Here \( \hat{SE}\{\hat{Y}_i(t) - Y_i(t)\} \) is the estimated standard error of \( \{\hat{Y}_i(t) - Y_i(t)\} \) and is calculated as the square root of the estimated variance of \( \{\hat{Y}_i(t) - Y_i(t)\} \). The terms \( \hat{Y}_i(t) \) and \( \hat{SE}\{\hat{Y}_i(t) - Y_i(t)\} \) have different expression according to whether the subject is an existing data subject, or is a new subject, along similar lines as detailed in the previous subsection. In particular, in the case of new subject, they do not include estimates of the random subject effects \( b_i \) and their estimation variability. In Section 2.5.3 we assess the performance of the response trajectory as well as that of the proposed pointwise prediction intervals for both existing data subjects and new subjects.
2.5 Simulation

2.5.1 Description of the settings

We use Monte Carlo simulations to assess the numerical performance of the proposed method (LDFR) and compare it with two other competing approaches: LPFR (Goldsmith et al., 2012a) and LPEER (Kundu et al., 2016). The data \([t_{ij}, Y_{ij}, \{W_{ijr}, s_r\} : r = 1, \ldots R] : j = 1, \ldots, n_i]^{n_i}_{i=1}\) are generated according to the following scenarios:

\(A\) \(X_i(s, t) = \tau(s, t) + \sqrt{2} \zeta_{i11} \cos(2\pi t) \cos(2\pi s) + \sqrt{2} \zeta_{i12} \sin(2\pi t) \cos(2\pi s) + \sqrt{2} \zeta_{i21} \cos(4\pi t) \sin(2\pi s) + \sqrt{2} \zeta_{i22} \sin(4\pi t) \sin(2\pi s),\)

where \(\tau(s, t) = 1 + 2s + 3t + 4st\). Moreover, \(\zeta_{i11}, \zeta_{i12}, \zeta_{i21}, \text{ and } \zeta_{i22}\) are assumed to be mutually independent and identically distributed (IID) such as \(\mathcal{N}(0,3.5), \mathcal{N}(0,2), \mathcal{N}(0,3), \text{ and } \mathcal{N}(0,1.5)\) respectively. Let \(W_{ijr} = W_i(s_r, t_{ij}) = X_i(s_r, t_{ij}) + \epsilon_{ijr}(s_r). \) We define the error term as \(\epsilon_{ijr}(s_r) = \sqrt{2} \cos(2\pi s_r) \epsilon_{1,ij} + \sqrt{2} \sin(2\pi s_r) \epsilon_{2,ij} + \epsilon_{3,ij}(s_r). \)

Here, \(\epsilon_{1,ij}, \epsilon_{2,ij}, \text{ and } \epsilon_{3,ij}(s_r)\) are IID such as \(\mathcal{N}(0,\sigma_{\epsilon_1}^2), \mathcal{N}(0,\sigma_{\epsilon_2}^2), \text{ and } \mathcal{N}(0,\sigma_{\epsilon_3}^2)\); where, \(\sigma_{\epsilon_1}^2 = 0.3, \sigma_{\epsilon_2}^2 = 0.7, \text{ and } \sigma_{\epsilon_3}^2\) are calculated using signal-to-noise-ratio (SNR) which is defined as

\[SNR = \frac{\int \int \text{var}\{W_i(s, t)\} ds dt}{\sigma_{\epsilon_1}^2 + \sigma_{\epsilon_2}^2 + \sigma_{\epsilon_3}^2} - 1.\]

\(B1\) Large noise variance: \(\sigma_{\epsilon_3}^2 = 9\) (i.e. \(SNR = 0.5\)).

\(B2\) Small noise variance: \(\sigma_{\epsilon_3}^2 = 1\) (i.e. \(SNR = 2.5\)).

We consider a dense design for \(s\): \(\{s_1, \ldots, s_R\}\) is taken as a grid of 101 equidistant points in \([0,1]\) and two sampling designs for \(t_{ij}\)’s depending on number of repeated measurements \(n_i:\)

\(C1\) Sparse design when the number of repeated responses per subject is: \(n_i \in \{11, \ldots, 15\}\),

\(C2\) Moderately sparse design when \(n_i \in \{21, \ldots, 25\}\).

In each case \(\{t_{i1}, \ldots, t_{in_i}\}\) are randomly chosen from a set of 41 equidistant points in \([0,1]\).

We generate \(Y_{ij}\) in the exponential family as follows:

\(D1\) Gaussian responses with mean \(\mu_{ij} = \mu_i(t_{ij})\) defined by \(\mu_i(t) = \alpha(t) + \int X_i(s, t) \gamma(s, t) ds,\) where \(\alpha(t) = 7 \sin(3\pi t)\) and random deviation \(\epsilon_{ij}. \) Consider three dependence structures:
(i) **Independent covariance structure:** $\varepsilon_{ij}$ is distributed as IID $N(0,2)$.

(ii) **Compound symmetric (CS) structure:** $\varepsilon_{ij} = b_{i0} + e_{ij}$, where, $b_{i0}$ and $e_{ij}$ are distributed as IID $N(0,1)$ and $N(0,0.5)$ respectively, and are mutually independent.

(iii) **Random effect model (REM):** $\varepsilon_{ij} = b_{i0} + b_{i1}t_{ij} + e_{ij}$, where, $b_{i0}$ and $b_{i1}$ are distributed as IID $N(0,1)$ and $N(0,0.5)$ respectively with $\text{cov}(b_{i0}, b_{i1}) = 0.1$. Also $e_{ij}$ is distributed as IID $N(0,0.3)$, and is independent from both $b_{i0}$ and $b_{i1}$.

**(D2)** Binary-valued responses. We consider $P(Y_{ij} = 1) = \exp(\omega_{ij})/[1 + \exp(\omega_{ij})]$, where $\omega_{ij} = \alpha(t_{ij}) + \int_X X_i(s, t_{ij}) \gamma^\delta(s, t_{ij}) ds + b_{i0} + b_{i1}t_{ij}$, and the subject-specific random effects $b_{i0}, b_{i1}$ are used to model the dependence across the repeated measurements and are generated as in (iii) above. The choice for $\alpha(\cdot)$ described in (D1) varies with time; while a time-varying intercept presents no issues for our proposed method, such choice does not seem to be accommodated by LPFR. Thus, for binary-valued responses, in order to compare the results of our method to LPFR, we consider $\alpha(\cdot) = 2$.

We consider the functional coefficients as below:

**(E1)** Mixture of trigonometric and exponential functions of $t$: $\gamma^\delta(s, t) = \sqrt{2} \exp(-\delta t) \cos(2\pi s) + \sqrt{2} \delta t \sin(\delta t) \sin(2\pi s)$.

**(E2)** Polynomial function of $t$: $\gamma^\delta(s, t) = \sqrt{2}(1 + \delta t) \cos(2\pi s) + \sqrt{2}(1 - \delta t + \delta t^2) \sin(2\pi s)$.

In both cases the parameter $\delta$ controls the departure from a time-invariant effect. For example in (E1) when $\delta = 0$ we have that $\gamma^\delta(s, t) = \sqrt{2} \cos(2\pi s)$, while when $\delta \neq 0$, $\gamma^\delta(s, t)$ varies with time $t$. We investigate the cases $\delta = \{0, 1, 2, 5, 10\}$. The various settings amount to a signal to noise ratio of the response varying between 0.1 (B1, $\delta = 1$, and any $D1i - D1iii$) to 3.5 (B2, $\delta = 5$ and any $D1i - D1iii$) as defined in Section A.1.1 of the Supplementary Material.

We study the performance of our proposed method for varying sample sizes $I \in \{100, 200, 300\}$. The implementation of the methodology consists of two main steps. First pre-process the noisy longitudinal functional covariates using Park & Staicu (2015); implemented in the function `fpca.lfda` in the R package `refund` (Huang et al. (2015)). For transparency we describe it briefly: (1) Estimate the bivariate mean function $\hat{\gamma}(\cdot, \cdot)$ using the fast bivariate smoother based on tensor product of cubic splines with 35 knots in each direction and second order difference penalty (Xiao et al., 2013), and by selecting the smoothing parameter using GCV. (2) Demean the observed functional predictors and estimate the
smooth marginal covariance using the sandwich bivariate smoother. (3) Perform eigenanalysis of the estimated smooth covariance and obtain the pairs of eigenvalues and eigenfunctions \( \{ \tilde{\lambda}_k, \tilde{\phi}_k(\cdot) \} \); here \( K \) is chosen using 95% PVE value. (4) Estimate the time-varying loadings as \( \tilde{\xi}_{ijk} = \int_S \{ W(s,t_{ij}) - \bar{\tau}(s,t_{ij}) \} \tilde{\phi}_k(s) ds \) using numerical integration. (5) For each \( k \), consider \( \{ \tilde{\xi}_{ijk}, t_{ij} : j = 1, \ldots, n_i \} \), and assume a nonparametric covariance structure for the dependence of \( \tilde{\xi}_{ijk} \) across \( j \) as described in Section 2.3.1. Furthermore, estimate the eigen-components \( \{ \tilde{\eta}_{kl}, \tilde{\psi}(\cdot)_{kl} \} \) where \( \eta_{kl} \) are the non-decreasing non-negative eigenvalues; \( L_k \) is chosen by 95% PVE value. Predict the scores \( \xi_{ik} \) using the associated truncated mixed effects model discussed in Section 2.3.1 and calculate the estimated time-varying loadings for any time \( t \in \mathcal{T} \); i.e. \( \hat{\xi}_{ik}(t) = \sum_{l=1}^{L_k} \xi_{ikl} \hat{\psi}_{kl}(t) \). The second step uses \( \tilde{\xi}_{ijk} = \hat{\xi}_{ik}(t_{ij}) \) in (2.4) and fits the approximated generalized mixed model with penalties and assuming independent random effects; identity link for Gaussian responses and logit link for binary responses. The function \texttt{lme} in the \texttt{R} package \texttt{nlme} is used at this step.

For LPFR the time-invariant regression coefficient \( \gamma(s) \) is modeled using the truncated linear basis with 30 functions (default choice) and the smoothing parameter is estimated by REML; the model is fitted using the function \texttt{lpfr} in the \texttt{R} package \texttt{refund}. For LPEER, the time-varying coefficient \( \gamma(s,t) \) is modeled using a polynomial basis in time \( t \), with coefficients that are smooth functions in \( s \) and which are estimated using a penalized criterion with a second-order difference penalty; the degree of the polynomial basis is selected using the Akaike information criterion (AIC) (Akaike (1974)) and the smoothing parameters of the smooth terms are selected using REML. The model is fitted using the function \texttt{lpeer} in the \texttt{R} package \texttt{refund}.

### 2.5.2 Evaluation criteria

To assess the prediction performance of the method, we divide each simulated dataset into a training and test set. Both sets contain information for the \( I \) subjects; recall \( I \) is the total number of subjects. The test set is formed as follows: for each subject \( i \) in the dataset, we randomly select five instances without replacement from the available \( n_i \) instances, say \( \{ t_{ij_1}, t_{ij_2}, t_{ij_3}, t_{ij_4}, t_{ij_5} \} \), and include the corresponding information \( \{ t_{ij}, Y_{ij}, \{ W_{ijr,s} : r = 1, \ldots, R \} \} \) in the test data. The remaining observations for each subject are included in the training set. We fit the model using the training data; then we predict both the responses in the training set (IN) and the responses in the test set (OUT) using the estimates obtained from the fit on training data. To evaluate the performance of the competing
models for normal responses, we compute the root-mean-prediction-error for the training set (INPE) and for the test set (OUTPE); i.e. \( \text{INPE} = \sqrt{\frac{1}{\sqrt{I}} \sum_{i=1}^{I} (\sum_{j \in \{j_1, ..., j_5\}} (Y_{ij} - \hat{Y}_{ij})^2)/(n_i - 5)} \) and \( \text{OUTPE} = \sqrt{\frac{1}{\sqrt{I}} \sum_{i=1}^{I} (\sum_{j \in \{j_1, ..., j_5\}} (Y_{ij} - \hat{Y}_{ij})^2)/5} \). For binary-valued responses, we assess the numerical performance in estimating the linear predictor trajectory \( g(\cdot) \), and with respect to sensitivity or true positive rate (TPR); where TPR is defined as the proportion of successes \( (\hat{Y}_{ij} = 1) \) that are correctly identified.

The prediction of the entire trajectory is assessed using the root mean prediction error, \( \text{RMPE}_{\text{trj}} \) of \( \hat{Y}_i(\cdot) \) which is defined as \( \text{RMPE}_{\text{trj}} = \sqrt{\frac{1}{\sqrt{I}} \sum_{i=1}^{I} \left[ \frac{1}{n} \sum_{j=1}^{n} \{Y_i(t_j) - \hat{Y}_i(t_j)\}^2 \right]} \), where \( \{t_1, ..., t_n\} \) is an equally spaced grid of 41 points in \([0, 1]\) and \( Y_i(t_j) \) is obtained using the generating model. For this part, the model parameters are estimated using the entire data set, and not just the training data set.

The accuracy of the pointwise prediction bands is evaluated in two cases. First, we assess the performance of the prediction bands for all the existing data subjects, that is subjects whose data are used to estimate the model parameters. Second, assess the performance for prediction bands of new subjects responses, whose functional predictor information is available solely. In the latter case, we construct a new set of 100 subjects and for each set we generate data according to our model; the data for these subjects are not used in the estimation of the model parameters. In both cases the performance of the 100(1 - \( \alpha \))% pointwise prediction band, say \( PB_t = (PB_{tl}, PB_{tu}) \) specified in terms of its endpoints and which is constructed as detailed in Section 2.4.2, is evaluated using the average pointwise coverage defined as \( \frac{1}{I} \sum_{i=1}^{I} \sum_{j=1}^{n} 1\{Y_i(t_j) \in PB_{tl}\}/n \), where \( 1(x \in A) \) equals 1 if \( x \in A \) and 0 otherwise. We also calculate the expected length of the constructed prediction bands as \( \frac{1}{I} \sum_{i=1}^{I} \sum_{j=1}^{n} (PB_{tl} - PB_{tu})/n \).

The results are based on 1000 independent samples for each combination of the simulation settings. In our numerical investigation, we use Intel(R) Core(TM) i7-4770, 3.40 GHz processor with 8.0 GB RAM in 64-bit operating system.

### 2.5.3 Prediction performance assessment and comparison

**Prediction accuracy.** First, we consider Gaussian responses (D1) and compare our longitudinal dynamic functional regression method (LDFR) with LPFR. Table 2.1 displays the median of IN-and-OUT sample prediction errors for different \( \delta \) values for 1000 simulations along with their respective interquartile ranges (IQR) in parenthesis. The results correspond to data generated using CS dependence structure (D1ii) and fitted by assuming a model.
with CS type covariance structure. We observe that both types of prediction errors (IN and OUT) are similar for the two approaches when \( \delta = 0 \). However, as \( |\delta| > 0 \) the functional coefficient is time-dependent and the prediction results with the proposed method are superior relative to LPFR. For example, when \( \delta = 5 \) our method yields improvement in prediction accuracy by more than 40% over LPFR. Furthermore, the numerical study shows that the accuracy of our method increases with the number of repeated measurements per subject; this is expected as in this case, the estimation of the within subject covariance improves. In the Supplementary Material, Section A.1.2 we investigate mild misspecification of the dependence structure and observe similar findings.

Next, we compare the performance of LDFR with LPEER. Because of the heavy computational burden of the latter approach, we limit our investigation to 100 Monte Carlo samples per setting; see Table 2.2. Here, we fit the competing model without assuming prior knowledge about the structure of the bivariate regression coefficient \( \gamma(s,t) \). Table 2.2 illustrates the prediction performance when data are generated from CS type covariance structure \((D1\text{ii})\). We fit the model by using a random subject intercept model (correct covariance model). When \( |\delta| > 0 \) the departure of \( \gamma(s,t) \) from a time-invariant coefficient is stronger. The numerical results show improved performance for our method as \( |\delta| > 0 \). When \( \delta = 1 \), LDFR and LPEER show similar prediction accuracy. However, LDFR is computationally over an order of magnitude faster than LPEER; see the third and sixth pairs of columns in Table 2.2. Furthermore, for \( \delta = 5 \), LDFR outweighs LPEER in nearly all the cases considered. This is possibly due to the fact that LPEER, in its implementation models \( \gamma(s,\cdot) \) using a polynomial basis in \( t \) and selects the number of basis functions from few choices; for the case \((E1)\), a much richer polynomial basis in \( t \) is needed to approximate \( \gamma(s,\cdot) \), than the bases considered. In contrast, the proposed method does not rely on such assumption. As an anonymous reviewer suggested, we also compare the prediction accuracy of the two approaches when the true regression coefficient \( \gamma(\cdot,\cdot) \) is a linear combination of polynomial functions in \( t \), case \((E2)\). The results are shown in Table A.7 in the Supplementary Material, Section A.1.2, and are consistent to the ones reported in Table 2.2; the major difference is the improved computing time for LPEER, but still much higher compared to our method. Also, there seems to be some numerical stability issues with LPEER; in our simulation study we experienced convergence problems in few cases where the sample size is small.

We consider binary responses with logit link \((D2)\) and evaluate the prediction accuracy of the proposed method with LPFR, which is the only existing alternative. We fit the model
with subject-specific random intercept, while data are generated assuming both random intercept and slope (model misspecification). Table 2.3 shows the prediction error of the linear predictors $\hat{g}(\mu_{ij})$ for both sparse and moderately sparse longitudinal designs, when the functional covariates are observed with large noise. The results are consistent to the previous ones. As the magnitude of $\delta$ increases the LDFR results also show an improved performance over LPFR with respect to the true positive rate. Additional simulation results for both Gaussian response and binary response cases are included in the Supplementary Material, Section A.1.3.

**Prediction accuracy of response trajectory.** Figure 2.2 illustrates the prediction error ($RMPE_{trj}$) for the entire trajectory in the case of Gaussian responses that are correlated using CS structure ($D_{2ii}$) and are observed in each of the two sampling designs considered, sparse and moderately sparse. As expected, the accuracy improves both as the number of repeated measurements per subject increases and when the sample size increases, the former factor having higher impact. As the magnitude of $\delta$ increases, the difficulty of the problem increases and the prediction accuracy for all cases suffers.
Table 2.1: Gaussian responses with CS dependence structure ($D_{1ii}$), when the longitudinal design is sparse ($C_1$) and moderately sparse (mod sparse, $C_2$); the functional covariates are observed with high noise variance ($B_1$) and effect $E_1$. Model is fitted assuming CS type dependence structure. Median prediction errors and IQR in parenthesis are reported for 1000 simulations.

\begin{center}
\begin{tabular}{cccccccccc}
\hline
 & \multicolumn{2}{c}{$\delta = 0$} & \multicolumn{2}{c}{$\delta = 2$} & \multicolumn{2}{c}{$\delta = 5$} \\
 & $\hat{IN}_{PE}$ & $\hat{OUT}_{PE}$ & $\hat{IN}_{PE}$ & $\hat{OUT}_{PE}$ & $\hat{IN}_{PE}$ & $\hat{OUT}_{PE}$ \\
\hline
LDFR & LPFR & LDFR & LPFR & LDFR & LPFR & LDFR & LPFR & LDFR & LPFR \\
\hline
\multirow{2}{*}{sparse} & 0.77 & 0.89 & 0.98 & 1.00 & 0.87 & 1.11 & 1.11 & 1.53 & 1.30 & 3.32 & 1.73 & 3.54 \\
& (0.04) & (0.04) & (0.11) & (0.05) & (0.05) & (0.08) & (0.11) & (0.09) & (0.13) & (0.30) & (0.25) & (0.32) \\
\multirow{2}{*}{$I = 100$} & 0.74 & 0.91 & 0.81 & 0.96 & 0.79 & 1.40 & 0.86 & 1.48 & 1.02 & 3.36 & 1.15 & 3.49 \\
& (0.03) & (0.03) & (0.05) & (0.04) & (0.03) & (0.06) & (0.06) & (0.09) & (0.08) & (0.26) & (0.12) & (0.32) \\
\multirow{2}{*}{mod sparse} & 0.76 & 0.88 & 0.94 & 1.00 & 0.85 & 1.37 & 1.05 & 1.53 & 1.20 & 3.34 & 1.60 & 3.54 \\
& (0.03) & (0.03) & (0.09) & (0.04) & (0.03) & (0.06) & (0.08) & (0.07) & (0.09) & (0.22) & (0.17) & (0.23) \\
\multirow{2}{*}{sparse} & 0.73 & 0.90 & 0.79 & 0.95 & 0.78 & 1.41 & 0.84 & 1.48 & 0.99 & 3.38 & 1.09 & 3.50 \\
& (0.02) & (0.02) & (0.04) & (0.03) & (0.02) & (0.05) & (0.04) & (0.07) & (0.05) & (0.18) & (0.08) & (0.25) \\
\multirow{2}{*}{$I = 200$} & 0.76 & 0.88 & 0.91 & 0.99 & 0.84 & 1.37 & 1.02 & 1.53 & 1.24 & 3.35 & 1.54 & 3.54 \\
& (0.03) & (0.02) & (0.08) & (0.03) & (0.03) & (0.05) & (0.06) & (0.05) & (0.07) & (0.18) & (0.15) & (0.19) \\
\multirow{2}{*}{mod sparse} & 0.73 & 0.90 & 0.79 & 0.95 & 0.77 & 1.41 & 0.83 & 1.48 & 0.98 & 3.38 & 1.07 & 3.49 \\
& (0.02) & (0.01) & (0.04) & (0.02) & (0.02) & (0.04) & (0.03) & (0.05) & (0.04) & (0.15) & (0.06) & (0.20) \\
\hline
\end{tabular}
\end{center}
Table 2.2: Gaussian responses with CS dependence structure ($D_{1ii}$), when the longitudinal design is sparse ($C_1$) for $B_1$ and $B_2$ with effect $E_1$. Model is fitted assuming CS type dependence structure. Median prediction errors and IQR in parenthesis are reported.

<table>
<thead>
<tr>
<th>SNR</th>
<th>$\delta$</th>
<th>INPE</th>
<th>OUTPE</th>
<th>Run time (sec)</th>
<th>INPE</th>
<th>OUTPE</th>
<th>Run time (sec)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>1</td>
<td>0.74 (0.03)</td>
<td>0.91 (0.07)</td>
<td>0.94 (0.05)</td>
<td>14.74</td>
<td>380.75</td>
<td>2.05 (0.13)</td>
</tr>
<tr>
<td>2.5</td>
<td>1</td>
<td>0.73 (0.03)</td>
<td>0.90 (0.07)</td>
<td>0.91 (0.03)</td>
<td>12.92</td>
<td>210.95</td>
<td>2.00 (0.14)</td>
</tr>
<tr>
<td>0.5</td>
<td>2</td>
<td>0.73 (0.02)</td>
<td>0.86 (0.05)</td>
<td>0.92 (0.03)</td>
<td>97.00</td>
<td>1270.51</td>
<td>2.02 (0.06)</td>
</tr>
<tr>
<td>2.5</td>
<td>2</td>
<td>0.72 (0.02)</td>
<td>0.85 (0.04)</td>
<td>0.90 (0.03)</td>
<td>90.30</td>
<td>1230.92</td>
<td>1.96 (0.09)</td>
</tr>
<tr>
<td>0.5</td>
<td>5</td>
<td>1.31 (0.15)</td>
<td>1.78 (0.24)</td>
<td>1.72 (0.15)</td>
<td>10.60</td>
<td>703.65</td>
<td></td>
</tr>
<tr>
<td>2.5</td>
<td>5</td>
<td>1.30 (0.14)</td>
<td>1.73 (0.28)</td>
<td>1.75 (0.14)</td>
<td>11.28</td>
<td>805.15</td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>5</td>
<td>1.25 (0.06)</td>
<td>1.78 (0.07)</td>
<td>1.56 (0.13)</td>
<td>63.75</td>
<td>4542.27</td>
<td></td>
</tr>
<tr>
<td>2.5</td>
<td>5</td>
<td>1.20 (0.09)</td>
<td>1.71 (0.06)</td>
<td>1.51 (0.15)</td>
<td>71.55</td>
<td>4367.22</td>
<td></td>
</tr>
</tbody>
</table>
Table 2.3: Binary responses ($D2$) when the longitudinal design is sparse/moderately sparse (mod sparse); the functional covariates are observed with high noise variance ($B1$) and effect $E1$. Fitted model assumes subject-specific random intercept. We report the median of prediction errors of the linear predictor trajectories and, in parenthesis, the median of true-positive-rates.

<table>
<thead>
<tr>
<th></th>
<th>$\delta = 0$</th>
<th></th>
<th></th>
<th>$\delta = 5$</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>INPE</td>
<td>OUTPE</td>
<td>INPE</td>
<td>OUTPE</td>
<td>INPE</td>
<td>OUTPE</td>
</tr>
<tr>
<td></td>
<td>LDFR</td>
<td>LPFR</td>
<td>LDFR</td>
<td>LPFR</td>
<td>LDFR</td>
<td>LPFR</td>
</tr>
<tr>
<td>I = 100 sparse</td>
<td>1.27 (0.96)</td>
<td>1.30 (0.96)</td>
<td>1.34 (0.96)</td>
<td>1.30 (0.96)</td>
<td>1.89 (0.90)</td>
<td>2.88 (0.87)</td>
</tr>
<tr>
<td>mod sparse</td>
<td>1.19 (0.96)</td>
<td>1.29 (0.96)</td>
<td>1.20 (0.96)</td>
<td>1.29 (0.96)</td>
<td>1.55 (0.91)</td>
<td>2.87 (0.86)</td>
</tr>
<tr>
<td>I = 200 sparse</td>
<td>1.23 (0.96)</td>
<td>1.29 (0.96)</td>
<td>1.28 (0.96)</td>
<td>1.29 (0.96)</td>
<td>1.81 (0.90)</td>
<td>2.89 (0.86)</td>
</tr>
<tr>
<td>mod sparse</td>
<td>1.17 (0.96)</td>
<td>1.28 (0.96)</td>
<td>1.18 (0.96)</td>
<td>1.29 (0.96)</td>
<td>1.46 (0.92)</td>
<td>2.87 (0.86)</td>
</tr>
<tr>
<td>I = 300 sparse</td>
<td>1.21 (0.96)</td>
<td>1.28 (0.96)</td>
<td>1.25 (0.96)</td>
<td>1.29 (0.96)</td>
<td>1.77 (0.91)</td>
<td>2.89 (0.87)</td>
</tr>
<tr>
<td>mod sparse</td>
<td>1.17 (0.96)</td>
<td>1.28 (0.96)</td>
<td>1.17 (0.96)</td>
<td>1.28 (0.96)</td>
<td>1.43 (0.92)</td>
<td>2.88 (0.86)</td>
</tr>
</tbody>
</table>

Figure 2.2: Gaussian responses with CS dependence structure ($D1ii$), when the longitudinal design is sparse (left) and moderately sparse (right); the functional covariates are observed with high noise variance ($B1$) and effect $E1$. Fitted model assumes CS covariance structure. Reported is $RMPE_{trj}$ for observed (white boxplot) and unobserved (gray boxplot) subjects based on 1000 simulations. Reference lines are drawn at RMPE values 1, 2, and 3 for convenience.
Accuracy of the prediction bands. We examine the performance of the prediction bands in terms of actual coverage and expected length for two nominal levels 90% and 95%. Table 2.4 shows the results for few choices of $\delta$. In general the average coverage stays around the nominal levels for both the observed and unobserved subjects across different settings, and irrespective of the complexity of the signal as defined by $\delta$; see the results for $\delta = 1$ and $\delta = 10$. As expected, the width of the prediction bands is larger for the new subjects compared to the existing one and it increases with the complexity of the regression coefficient (large $|\delta|$). For all the settings, the results improve for larger sample size and increased signal to noise ratio (SNR).
Table 2.4: Gaussian responses with CS dependence structure \((D1\text{ii})\), when the longitudinal design is moderately sparse; the functional covariates are observed with high \((B1)\) and low \((B2)\) noise variance with effect \(E1\). Fitted model assumes CS covariance structure. Reported are the average coverage probabilities of 95\% and 90\% pointwise prediction bands, standard errors (in parenthesis), average length (in square bracket), for the observed \((Y_i)\) and unobserved subjects \((Y_i^\ast)\). Results are based on 1000 MC simulations.

\[
\begin{array}{cccccccc}
\text{SNR} = 0.5 & \text{I} = 100 & Y_i & (1 - \alpha) = 0.95 & \delta = 1 & \delta = 5 & \delta = 10 & Y_i^\ast & (1 - \alpha) = 0.95 & \delta = 1 & \delta = 5 & \delta = 10 \\
\hline
\hline
\delta = 1 & 0.96 (0.01) & [2.99] & 0.95 (0.01) & [4.18] & 0.95 (0.01) & [5.29] & 0.96 (0.01) & [2.97] & 0.95 (0.01) & [4.12] & 0.95 (0.01) & [5.30] \\
Y_i & 0.95 (0.01) & [4.91] & 0.94 (0.01) & [5.85] & 0.94 (0.01) & [6.96] & 0.90 (0.01) & [4.11] & 0.90 (0.01) & [4.84] & 0.90 (0.01) & [5.84] \\
\hline
\text{I} = 300 & Y_i & 0.95 (0.01) & [2.97] & 0.95 (0.01) & [4.12] & 0.95 (0.01) & [5.30] & 0.95 (0.01) & [4.89] & 0.95 (0.01) & [5.77] & 0.94 (0.01) & [6.86] \\
Y_i^\ast & 0.95 (0.01) & [4.91] & 0.94 (0.01) & [5.85] & 0.94 (0.01) & [6.96] & 0.90 (0.01) & [4.11] & 0.90 (0.01) & [4.84] & 0.90 (0.01) & [5.76] \\
\hline
\text{SNR} = 2.5 & \text{I} = 100 & Y_i & 0.96 (0.01) & [2.97] & 0.95 (0.01) & [4.07] & 0.95 (0.01) & [5.20] & 0.95 (0.01) & [4.91] & 0.94 (0.01) & [5.75] & 0.94 (0.01) & [6.80] \\
Y_i^\ast & 0.95 (0.01) & [4.89] & 0.95 (0.01) & [5.67] & 0.94 (0.01) & [6.57] & 0.90 (0.01) & [4.10] & 0.90 (0.01) & [4.77] & 0.90 (0.01) & [5.70] \\
\hline
\text{I} = 300 & Y_i & 0.96 (0.01) & [2.95] & 0.95 (0.01) & [4.01] & 0.95 (0.01) & [5.08] & 0.95 (0.01) & [4.89] & 0.95 (0.01) & [5.67] & 0.94 (0.01) & [6.57] \\
Y_i^\ast & 0.95 (0.01) & [4.89] & 0.95 (0.01) & [5.67] & 0.94 (0.01) & [6.57] & 0.90 (0.01) & [4.10] & 0.90 (0.01) & [4.77] & 0.90 (0.01) & [6.00] \\
\end{array}
\]


2.6 Data application

Our motivating application is a lactating sow study where the primary objective is to investigate the effect of thermal environment (i.e. temperature ($T$)) on the feed-intake of the lactating sows. This study is very important for several reasons: (1) ambient temperatures above the evaporative critical temperature decreases the amount of food-intake which, as a result, deteriorates the reproductive performance and hinders the growth rate of piglets of lactating sows (Black et al. (1993). (2) Also, poor feed-intake of the lactating sows leads to increased body weight loss during lactation and reduced milk yield, and is further associated with compromised weight gain of their litter (Johnston et al. (1999), Renaudeau & Noblet (2001), Renaudeau & Noblet (2001)). (3) Thirdly, heat-stress results in a reduction of farrowing rate (the percentage of sows that become pregnant and farrow a litter of piglets) and total number of pigs born in sows (Bloemhof et al. (2013)) which in turn has a negative effect on the total production of pork meat per year. (4) Fourth, pigs from sows raised in an unfavorable thermal environment will be fatter than the ones reared in favorable cooler environments and this fact makes pork meat fattier (Baumgard (2015)). (5) Fifth, because of heat stress associated with hot climactic thermal environment, the swine industry in US incurs a total estimated loss worth of $300 million per year on average (St-Pierre et al. (2003)). Therefore, insight into how the feeding behavior changes over time due to the prolonged exposure to a hot environment will assist in proposing more economical and efficient feeding strategies for lactating sows.

The experimental study was carried during July to October in 2013 in a 2,600-sow commercial research unit in Oklahoma (Rosero et al. (2016)) and involved 480 PIC Camborough sows. The sows were kept in the farrowing facility where they gave birth to piglets. Depending on the number of previous pregnancies (parity levels), sows were classified into younger (parity equal to zero or one) or older (parity equal to two or higher). The sows are brought to the farrowing crates when they are approximately five days before they are due to give birth. They arrive in groups; the study involves 21 groups of about 21-23 sows. Sows are observed during their 20-21 day lactation period and their respective food-consumption is monitored. Each sow is provided food individually with a computerized feeding system (Howema, Big Dutchman, Germany). The amount of food-offered ($FO$) was recorded at 2.00PM on each day and feed-refusal ($FR$) was measured the following day prior to any subsequent food addition; feed-intake ($FI$) was calculated as $FI = FO - FR$ in kg. Minute-by-minute information about the ambient air temperature ($^\circ C$) and humidity (%) of the farrowing facility were
recorded by data loggers (LogTag, MicroDAQ Ltd., Contoocook, NH). The experimenters removed information of five sows due to unreliable measurements and thus we had available information for 475 sows. The facility ambient temperature was controlled by a ventilation system; the barns have cool cells that pull fresh air through wet corrugated material to provide further cooling of air. There are some missing observations for temperature profiles due to machine failure which qualifies the pattern of missingness as missing completely at random. Our objective is to study the effect of temperature on the feed intake of sows.

Let $i$ index the sows, $j$ index the repeated instances for the same sow, and $t_{ij}$ to denote the lactation day of the $i$th sow, corresponding to the $j$th instance at which the sow is observed; for many sows we have $t_{ij} = j$ for $j = 1, \ldots, 21$, but this is not always the case. The “time” is defined as the 24 hours time window from 2:00PM to 1:59PM. Furthermore let $g_i$ index the group of the $i$th sow, $g_i = 1, \ldots, 21$; typically, the sows within the same group give birth closer to one another. Denote by $nTemp_{ij}(\cdot) = nTemp_i(\cdot, t_{ij})$ the daily temperature profile observed at the $t_{ij}$ lactation day of the $i$th sow; the measurements typically include noise, hence the prefix “n”. Later we use notation $Temp_{ij}(\cdot)$ for the true temperature profile corresponding to $nTemp_{ij}(\cdot)$. The right panel of Figure 2.1 shows the daily temperature corresponding to the first 21st days for a random sow. Let $FI_{ij}$ be the $FI$ of the $i$th sow at its $j$th repeated occasion, lactation day $t_{ij}$.

We assume that the relationship between the feed intake and the temperature is described by the LDFR model:

$$FI_{ij} = \beta_{p_i}(t_{ij}) + \int S Temp_{ij}(s) \gamma(s, t_{ij}) ds + b_{g_i} + b_{0i(g_i)} + b_{1i(g_i)} t_{ij} + \varepsilon_{ij}, \quad (2.10)$$

where $\beta_{p_i}(\cdot)$ is the mean feed intake for group $p_i$ where $p_i = 0$ (young sows) and $p_i = 1$ (old sows) and $\gamma(s, \cdot)$ quantifies the time-varying effect of the temperature on $FI$; the integral reflects the aggregated effect during the course of the 24 hours. The term $b_{g_i} + b_{0i(g_i)} + b_{1i(g_i)} t_{ij}$ models the dependence of the responses within the same sow as well as the dependence of the responses of the sows who are in the same group. The random terms $b_{g_i}$ is a group-specific effect and $b_{0i(g_i)}$ and $b_{1i(g_i)}$ are sow within group-specific intercept and slope. We assume that $b_{g_i} \sim N(0, \sigma^2_{g_0})$, $b_{0i(g_i)} \sim N(0, \sigma^2_{0i})$, and $b_{1i(g_i)} \sim N(0, \sigma^2_{1i})$ are all mutually independent. Finally, it is assumed that the measurement errors $\varepsilon_{ij}$ are independent and distributed as $N(0, \sigma^2_e)$. Model (3.19) does not account for the previous day feed intake, which may be viewed as an important predictor for the current day feed intake. Such approach is discussed later this section.
The steps for fitting the model (2.10) are similar to the ones described in Section 2.5. One important specific is that the fast bivariate spline smoothing (Xiao et al. (2013)) uses cubic splines with 35 knots in s-direction and 19 knots in t-direction and second order difference penalty is used to control the smoothness of the fit and the tuning parameters are estimated by REML. Also the parameter functions $\beta_k(\cdot)$ are modeled using 15 truncated linear splines with knots placed uniformly in the time domain; the smoothing parameters are selected using REML.

2.6.1 Fit assessment

We first assess the prediction performance of the proposed method and compare it with the available competitors, LPFR and LPEER. In this regard we split the data into a train set, which is used to build the model and a test set on which the prediction performance is evaluated; we replicate the test-train split for 25 times. We consider two ways of forming the test data: (a) randomly select 350 sows of the total of 475 sows and include only the measurements corresponding to their last 10 lactation days; and (b) take all the 475 sows and include only the measurements corresponding to about 20% of their lactation days that are selected at random. The remaining data form the training set. Approach (a) involves data on fewer sows than approach (b). At the same time, the approach (a) assesses the performance of prediction at “future” lactation days, while the approach (b) evaluates the prediction performance at random lactation days within the 1 to 21 days at which the sows are observed.

For completeness we describe the implementation of the competitive approaches. LPFR assumes that temperature has a constant effect across the lactation days of a sow and models this time-invariant effect using 30 truncated linear splines basis functions and the tuning parameters are estimated by REML. The covariance structure is specified as in model (3.19) and the model is fitted using `lpfr()` function available in `refund` package (Huang et al. (2015)). For LPEER, we consider polynomial functions of time of degree $d = 0, 1, \ldots, 4$, $\gamma(s, t) = \gamma_1(s) + t\gamma_1(s) + \cdots + t^d\gamma_d(s)$ and select the optimal $d$ by AIC as described earlier; the model is fitted using `lpeer` in the same package. The covariance structure is specified using a subject-specific random intercept and random group effect; it is not clear how to modify the existing code to accommodate a subject-specific slope effect.

Table 2.5 reports the results. The findings show that LDIFR and LPEER perform relatively similar in terms of in-sample and out-of-sample prediction in the two situations.
considered. Nevertheless our methodology yields to computations that are an order of magnitude faster and a better fitting criteria as assessed by marginal AIC. While we use marginal AIC as an ad-hoc criterion for overall model performance, it is more appropriate to use conditional AIC in a longitudinal setting as marginal AIC is a biased estimator of the Akaike information; see Vaida & Blanchard (2005); Liang et al. (2008); Greven & Kneib (2010) for details. In contrast, the prediction results with LPFR are inferior, and they indicate a lack of appropriateness of a time-invariant effect model for our lactating sow application, based on the observations from our simulation investigation.
Table 2.5: Median prediction accuracy, computing time (in seconds), and marginal AIC for the proposed LDFR and the competitive approaches LPFR and LPEER for the sows application; corresponding IQRs are reported in parenthesis.

<table>
<thead>
<tr>
<th></th>
<th>IN $\phi$</th>
<th></th>
<th>OUT $\phi$</th>
<th></th>
<th>AIC</th>
<th></th>
<th>Computing time</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>LDFR</td>
<td>LPEER</td>
<td>LPFR</td>
<td>LDFR</td>
<td>LPEER</td>
<td>LPFR</td>
<td>LDFR</td>
</tr>
<tr>
<td>(a)</td>
<td>1.22</td>
<td>1.32</td>
<td>1.35</td>
<td>1.69</td>
<td>1.61</td>
<td>1.91</td>
<td>19348.32</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.02)</td>
<td>(0.02)</td>
<td>(0.06)</td>
<td>(56.11)</td>
</tr>
<tr>
<td>(b)</td>
<td>1.32</td>
<td>1.39</td>
<td>1.44</td>
<td>1.43</td>
<td>1.47</td>
<td>1.54</td>
<td>23942.34</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.03)</td>
<td>(0.02)</td>
<td>(0.02)</td>
<td>(90.71)</td>
</tr>
</tbody>
</table>

34
2.6.2 Estimation of the model components. Prediction of response trajectories

We fit the model (2.10) to the entire data. First, we examine the residuals: the auto-correlation (ACF) and the partial ACF (PACF) plots in Section A.2.1 of the Supplementary Material show no evidence of auto-regressive dependence. In fact in Section A.2.1 we investigated further dependence of the current feed intake onto the previous days feed intake and found no evidence of lagged auto-regressive dependence. Section B3 considers diagnostic plots for the other random components assumed by our model. Fitting model (2.10) yields the following estimates of the random effects: $\hat{\sigma}_0 = 0.71$, which quantifies the variability of the mean feed intake (intercept) across sows, $\hat{\sigma}_1 = 0.80$, which measures the variability of the sows rate of change of feed intake (slope), and $\hat{\sigma}_g = 0.21$, which indicates the amount of variability of the group-level mean feed intake.

Figure 2.3 shows the estimates of the model parameter functions. Specifically the left panel depicts the estimated mean feed intake for old sows (solid line), $\hat{\beta}_1(t)$, and for young sows (dashed line), $\hat{\beta}_0(t)$ along with 95% pointwise confidence bands. It appears that the feed intake is about the same in the first couple of days, for both young and old sows, but shortly afterwards the older sows eat more than their younger relatives. By the fourth lactation day the older sows have an advantage of feed intake of up to 1-1.5 kg per day and they maintain this advantage for the remaining duration of the lactation period.

Figure 2.3, in the right panel, shows the estimated regression coefficient $\hat{\gamma}(s,t)$, which quantifies the minute-by-minute ($s$) effect of the temperature on the feed intake during the first 21st lactation days ($t$). As this regression coefficient function is identifiable only up to a function of $s$, we focus mainly on the changes across $t$. The association between the temperature and the feed intake changes during the lactation duration. For example, lower temperature levels around 7:30AM - 8:30AM are associated with much lower feed intake during the middle of the lactation period, say days 8 through 16, relative to the feed intake at the beginning or end of the lactation period. Also higher early evening (5:45PM - 7PM) temperature levels are associated with larger feed intake at the beginning of the lactation period, compared to the feed intake later on. These findings are based on the point estimates solely and do not account for uncertainty in the estimation; thus should be interpreted with caution. Additionally, using a version of R–square for this setting, the proposed model accounts for about 33% of the variation in the data and the signal to noise ratio is estimated to be about 0.1. Sections A.2.1-A.2.3 of the Supplementary Material include additional results.
Figure 2.3: Parameter estimates in the lactation sow application. Left panel depicts the estimated intercept function for the old (solid line) and young (dashed) sows with 95% pointwise confidence intervals. Right panel shows the estimated regression coefficient $\hat{\gamma}(\cdot, t)$ for each lactation day $t = 1, 2, \ldots, 21$.

for the data analysis, while Section A.2.4 provides prediction results for LDFR approach when a covariance model based on random group and subject random intercept is used.

Figure 2.4 shows the predicted full trajectories of the feed intake for two young sows (left and middle panels) and one old sow (rightmost) selected at random from different groups along with their pointwise prediction bands constructed as detailed in Section 2.4. The predicted trajectories are obtained from $\hat{FI}_i(t) = \hat{\beta}_p(t) + \int \hat{Temp}_i(s, t) \hat{\gamma}(s, t) ds + \hat{b}_g + \hat{b}_{0i(g_i)} + \hat{b}_{1i(g_i)} t$ for every day $t = 1, \ldots, 21$, where $\hat{Temp}_i(\cdot, t)$ are the smooth and demeaned temperature profiles observed in relation to sow $i$, and $\hat{b}_g, \hat{b}_{0i(g_i)}$ and $\hat{b}_{1i(g_i)}$ are predicted effects. These results indicate too greater feed intake for older sows relative to their younger counterparts; the prediction intervals are wider to account for estimated measurement error.
Figure 2.4: Predicted full feed intake trajectories for two young (left and middle) and one old (right) sows. Shaded regions correspond to 95% pointwise prediction bands based on LDFR.
2.6.3 Validation of the results for data application via simulation study

In this section we consider a simulation study mimicking the sow data structure. In particular, we generate feed-intake (kg) using the model (2.10) with the estimated smooth effects $\hat{\beta}_p(\cdot)$ and $\hat{\gamma}(\cdot, \cdot)$. The dependence across measurements are modeled through $b_{gi} + b_{hi(gi)} + b_{li(gi)}t_{ij} + \varepsilon_{ij}$ resembling the covariance structure of the data; here all terms bear the usual meaning as before and are generated as $b_{gi} \sim N(0, \hat{\sigma}_g^2)$, $b_{hi(gi)} \sim N(0, \hat{\sigma}_0^2)$, $b_{li(gi)} \sim N(0, \hat{\sigma}_1^2)$, and $\varepsilon_{ij} \sim N(0, \hat{\sigma}_e^2)$. Temperature profiles are constructed as $Temp_{ij}(s) = \hat{\tau}(s, t_{ij}) + \sum_{k=1}^{7} \xi_{ik}(t_{ij})\hat{\phi}_k(s)$ where the mean $\hat{\tau}(s, t_{ij})$ and the fPCs $\{\hat{\phi}_k(\cdot); k = 1, \ldots, 7\}$ are obtained from the data, and the scores $\xi_{ik}(\cdot)$ are generated as a zero-mean random process with covariance $\hat{G}_k(\cdot, \cdot)$ for each $k$; see Section A.2.3 of the Supplementary Material. Denote the observed temperature profiles by $nTemp_{ij}(s) = Temp_{ij}(s) + \varepsilon_{1,ij}(s) + \varepsilon_{2,ij}(s)$; where $\varepsilon_{1,ij}(s)$ is a smooth error process with zero mean and covariance $\hat{\Gamma}(s, s')$, and $\varepsilon_{2,ij}(s)$ is a white noise with zero-mean and covariance $\hat{\sigma}_w^21(s = s')$. We consider $I = 475$, $m_i = \{7, \ldots, 21\}$, and $g_i = 1, \ldots, 21$ as same as that of data. We simulate the data for 100 times, and split each dataset into a training and test set on which prediction performance is evaluated. We also assess the prediction coverage for both existing ($i$) and new ($i^*$) sows feed-intake. For the latter case, we construct a new set of 125 sows and for each of them we simulate temperature profiles according to our model; the data for these 125 sows is not used in the estimation of the model parameters.

Table 2.6 compare the prediction performance of the three approaches. LDFR and LPEER show similar accuracy, while LPFR remains inferior. However LDFR exhibits better fitting criteria as assessed by AIC. In addition, the average prediction coverage stays around the nominal levels (i.e. 95% and 90%) for both the observed and unobserved sows while having larger prediction band width for the unobserved ones. These findings are in agreement with the results demonstrated in Table 2.4 and 2.5.
Table 2.6: Numerical results based on 100 Monte Carlo simulations mimicking sow data. Reported are the median prediction errors, marginal AICs, RMPE_{trj}, IQR (in parenthesis), average coverage probabilities of 95\% and 90\% pointwise prediction bands, standard errors (in parenthesis with superscript \( \dagger \)), and average length of intervals (in square bracket) for the existing \( (Y_i) \) and new sows \( (Y_{i*}) \).

<table>
<thead>
<tr>
<th>LDFR</th>
<th>IN_{PE}</th>
<th>(a)</th>
<th>OUT_{PE}</th>
<th>(a)</th>
<th>AIC</th>
<th>RMPE_{trj}</th>
<th>Y_{0.95}</th>
<th>Y_{0.90}</th>
<th>Cov_{trj}^{0.95}</th>
<th>Cov_{trj}^{0.90}</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1.29</td>
<td>(0.02)</td>
<td>1.31</td>
<td>(0.02)</td>
<td>1.43</td>
<td>19669.47</td>
<td>23899.41</td>
<td>1.34</td>
<td>1.61</td>
<td>0.96 (0.02)( \dagger )</td>
</tr>
<tr>
<td></td>
<td>1.33</td>
<td>(0.01)</td>
<td>1.64</td>
<td>(0.05)</td>
<td>0.95 (0.02)( \dagger )</td>
<td>0.92 (0.03)( \dagger )</td>
<td>0.90 (0.03)( \dagger )</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>LPEER</th>
<th>IN_{PE}</th>
<th>(a)</th>
<th>OUT_{PE}</th>
<th>(a)</th>
<th>AIC</th>
<th>RMPE_{trj}</th>
<th>Y_{0.95}</th>
<th>Y_{0.90}</th>
<th>Cov_{trj}^{0.95}</th>
<th>Cov_{trj}^{0.90}</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1.35</td>
<td>(0.02)</td>
<td>1.38</td>
<td>(0.02)</td>
<td>1.47</td>
<td>19999.92</td>
<td>24207.91</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>1.33</td>
<td>(0.01)</td>
<td>1.61</td>
<td>(0.03)</td>
<td>0.96 (0.02)( \dagger )</td>
<td>0.95 (0.02)( \dagger )</td>
<td>0.90 (0.03)( \dagger )</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>LPFR</th>
<th>IN_{PE}</th>
<th>(a)</th>
<th>OUT_{PE}</th>
<th>(a)</th>
<th>AIC</th>
<th>RMPE_{trj}</th>
<th>Y_{0.95}</th>
<th>Y_{0.90}</th>
<th>Cov_{trj}^{0.95}</th>
<th>Cov_{trj}^{0.90}</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1.42</td>
<td>(0.02)</td>
<td>1.42</td>
<td>(0.02)</td>
<td>1.53</td>
<td>20138.64</td>
<td>24262.62</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>1.42</td>
<td>(0.02)</td>
<td>1.69</td>
<td>(0.05)</td>
<td>0.96 (0.02)( \dagger )</td>
<td>0.95 (0.02)( \dagger )</td>
<td>0.90 (0.03)( \dagger )</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

NA = not applicable
2.7 Discussion

In this paper we consider longitudinal dynamic functional regression for scalar responses and functional covariates observed in a longitudinal design. We propose a flexible way to model the time-varying bivariate regression coefficient function by combining ideas from functional data analysis and longitudinal data analysis. As one anonymous reviewer asserted, this clever combination allows one to tackle a challenging problem that has not previously been solved in this generality. The methodology relies on the assumptions that the leading eigenbasis functions of the functional predictor are most predictive of the response and that the latent predictor signals are relatively smooth. The approach is applicable to Gaussian as well as non-Gaussian responses and can directly accommodate additional vector covariates, non-linear effects of vector covariates, as well as multiple functional covariates observed on diverse sampling designs and with measurement error. The methodology can be easily implemented using the existing freely available software.

Numerical results show that the prediction performance of our approach is superior to existing alternative approaches when the regression coefficient function is indeed time varying, and is very competitive with the existing alternatives when the regression coefficient function is time-invariant. In spite of the increased flexibility, this method is computationally efficient; in fact it is orders of magnitude faster than its closest competitor. We discuss an approach to reconstruct the full response trajectory. We applied the method to the animal science application and found that the effect of the temperature on the feed intake of the lactating sows varies with the days since they gave birth.

One limitation of our methodology is that it relies on the implicit assumption that the current response is related to the current functional predictor only i.e. $E[Y_{ij}|X_{i1}(\cdot), \ldots, X_{im}(\cdot)] = E[Y_{ij}|X_{ij}(\cdot)]$. While this assumption makes sense for our application, it may not be reasonable for other situations. One possible approach to account for the past functional covariates is by considering a regression model inspired by the historical functional linear model (see Malfait & Ramsay (2003); Scheipl et al. (2015); Pomann et al. (2016) and Kim et al. (2011)).
Chapter 3

Significance Testing for Functional Effects in the Longitudinal Time-Varying Functional Regression Model

3.1 Introduction

Functional linear regression (FLR) has received a lot of attention recently; see for example Ramsay & Silverman (2002); Malfait & Ramsay (2003); Reiss & Ogden (2007); Crainiceanu et al. (2009); Goldsmith et al. (2011); Ferraty et al. (2012b); McLean et al. (2014); Ivanescu et al. (2015); Febrero-Bande et al. (2017); Islam et al. (2017, 2018) and references therein. In this article, we consider a functional linear regression with longitudinally observed scalar outcomes and multiple functional observations for the same subject. The purpose of this paper is to make statistical inference about the effect of a functional predictor on the mean response. Most of the past inferential approaches in FLR conduct testing either through the use of tests for zero-variance components (McLean et al., 2014; Swihart et al., 2014; Staicu et al., 2014) or via the use of norm based test with re-sampling technique (Park et al., 2016; Choudhury et al., 2017). While the latter approach is relatively more computationally intensive, selecting between these two approaches depends on the matter of practicality and convenience.

In spite of the work related to estimation in scalar-on-function regression, the inference
for coefficient function estimates has received little attention. To date, most inferential procedures in scalar-on-function regression are concerned about testing the lack of fit of a functional effect. For example, Cardot et al. (2003b) developed test statistics based on the norm of the empirical cross-covariance operator of functional predictor and response, and used asymptotic normality for the approximation of null distribution of the test statistic. Later Cardot et al. (2004) developed an alternative approach based on a direct approximation of the distribution of the cross-covariance operator; the authors proposed a pseudo-likelihood test statistic where the null distribution is approximated by a Fisher-Snedecor distribution. In a similar spirit, Kong et al. (2016) developed inferential procedures where the authors adopted functional principal component (FPC) analysis and transformed the functional linear regression (FLR) in a standard multivariate regression model in which FPC scores play the role of regressors. The authors investigated the performance of four traditional tests such as Wald, score, likelihood ratio, and $F$ test. Recently, McLean et al. (2014) considered a testing procedure to investigate the linear dependence between scalar responses and a functional covariate in the class of functional generalized additive models (FGAM) (Ramsay & Silverman, 2005; Müller & Yao, 2012); here the authors used the restricted likelihood ratio test (RLRT) for making inference. To achieve the same goal, García-Portugués et al. (2014) proposed bootstrap based test statistic assuming that the functional covariates are observed without noise and recorded on a dense grid of points. However these methods do not extend for the cases with multiple predictors or in a longitudinal setting.

Swihart et al. (2014) extended the inferential procedure of the functional effect in longitudinal framework for scalar-on-function regression; specifically, the authors used penalized spline mixed model and used RLRT for testing the zero-variance component to test the hypothesis of interest. Indeed, Swihart et al. (2014) assumed that the predictor is invariant over time; however such assumption is strong and violated in many practical situations. Kundu et al. (2016) and Islam et al. (2017) relaxed this assumption, and proposed longitudinal functional regression model with time-varying smooth coefficients. But both authors primarily focus on modeling and do not investigate the hypothesis testing. In many situations, it is of interest to investigate the adequacy of using time-varying effect in the context of functional linear regression models. This type of inference is important for two reasons: 1) it is possible in many situations that a simpler time-invariant coefficient is sufficient to quantify the relationship between responses and functional predictors; indeed this invariant coefficient function is far simpler to estimate and interpret; 2) Most of the existing literature on
longitudinal functional regression models present point-wise, rather than the simultaneous, confidence bands of the estimated smooth regression coefficients. While such approach appeals to quantify the uncertainty in parameter estimation, this approach does not investigate the statistical significance of the estimates rigorously. In addition, inverting such point-wise confidence bands to construct a statistical test, as is usually done in multivariate situation, is not meaningful either. To the best of our knowledge, the inferential procedures for testing the time-varying smooth coefficients are yet to be developed in full generality, and is the main purpose of this paper.

Additionally, inferences about the mean effects of functional outcomes has been studied recently. In this direction, Zhang et al. (2010) developed $L^2$ norm based and bootstrap based test statistics for making inference about the equality of two curves assuming that the curves are mutually independent and observed without error. Recently, Choudhury et al. (2017) proposed a permutation based approach to make inference about the difference in the mean profiles after adjustment for covariates. Other inferential procedures regarding the mean profiles are discussed in Zhang et al. (2007); Crainiceanu et al. (2012b); Horváth & Kokoszka (2012b); Staicu et al. (2015, 2014); Park et al. (2016). While most approaches focus mainly on the equality of mean profiles, the equality of the functional effects on responses in a longitudinal setting is yet to be studied.

Most of the available literature on testing the functional regression coefficient adopt a penalized approach which casts the FLR model in a mixed effect nonparametric framework, and derives a likelihood ratio test (LRT) statistic. In particular, such approach considers the problem of testing null hypothesis with restrictions on single or multiple variance components; see for example Swihart et al. (2014); Staicu et al. (2014); Tekbudak et al. (2017). There is plethora of literature that describes the testing of zero variance components in nonparametric and semi-parametric literature; see for example Ruppert et al. (2003); Crainiceanu et al. (2005); Crainiceanu & Ruppert (2004); Liang & Self (1996). Here the main idea relies on the approximation of the smooth effect via penalized splines represented in a linear mixed model (LMM) framework. In particular, (R)LRTs have been theoretically developed and computationally implemented to test the higher-order polynomial and/or random splines terms (Crainiceanu & Ruppert, 2004; Crainiceanu, 2008; Greven et al., 2012; Scheipl et al., 2008; Staicu et al., 2014). Other approaches include Wald-type tests (Wood, 2012b) or score tests (Molenberghs & Verbeke, 2007; Zhang & Lin, 2008). Swihart et al. (2014) argued (R)LRTs and score tests are asymptotically equivalent while (R)LRTs are widely adopted by the prac-
titioners due to the availability of accessible software RLRsim (Scheipl & Bolker, 2016) and the possibility of testing multiple variance components simultaneously. Additionally, Scheipl et al. (2008) observed superior numerical performance of (R)LRTs over Wald-type tests in situations of multiple variance components.

We organize the paper as follows. Section 3.2 introduces the proposed methodology for Gaussian responses for testing a specific hypothesis. Section 3 describes the inferential procedure. Section 4 describes the selection of basis functions. Section 5 extends the idea for testing various classes of hypotheses. Numerical assessment of the methodology is described in Section 6 which is followed by a data application in Section 7. We conclude with discussion in Section 8.

3.2 Proposed methodology

3.2.1 Preliminaries

Let the observed data be \([t_{ij}, Y_{ij}, \{(W_{ijr}, s_{ijr}) : r = 1, \ldots, r_{ij} ; i = 1, \ldots, n, j = 1, 2, \ldots, m_i\} ; i = 1, \ldots, n, j = 1, 2, \ldots, m_i]\); where \(i\) indexes the subject and \(j\) the repeated observations. \(Y_{ij}\) is the scalar response measured at time \(t_{ij}\), and \(W_{ijr}\)'s are the evaluations of the noisy functional covariates observed at \(s_{ijr}\)'s. It is assumed that \(t_{ij} \in T\) and \(s_{ijr} \in S\) for closed and compact sets \(T\) and \(S\) respectively. We assume that for each \(i\) and \(j\), \(r_{ij}\) is large, and furthermore, the set \(\{s_{ij1}, \ldots, s_{ijr_{ij}}\}\) is a finite grid in \(S\). Define \(W_{ijr} = W_{ij}(s_r) = W_i(s_r, t_{ij})\) and \(W_{ij}(s_{ijr}) = X_{ij}(s_{ijr}) + \epsilon_{ijr}\) where \(X_{ij}(\cdot)\) is the latent functional predictor corresponding to the subject \(i\) at time \(t_{ij}\), and \(\epsilon_{ijr} = \epsilon_{ij}(s_{ijr})\) is the corresponding error evaluated at \(s_{ijr}\) for any \(r\). For the simplicity of exposition, assume \(s_{ijr} = s_r\) for all \(i\) and \(j\); the proposed methodology can also be applied when functional measurements are observed at different grid of points \(s_{ijr}\)'s. With the abuse of notations, we write \(X_{ij}(s_r) = X_i(s_r, t_{ij})\).

Our objective is to draw inference about the time-varying effect of the functional predictor. Assuming Gaussian responses, we posit the following longitudinal time-varying functional regression model:

\[
Y_{ij} = \alpha(t_{ij}) + \int_S X_{ij}(s)\gamma(s, t_{ij})ds + \varepsilon_{ij}. \quad (3.1)
\]

Here, \(\alpha(\cdot)\) is an unknown intercept function, \(\gamma(\cdot, t_{ij})\) is the smooth time-varying coefficient function characterizing the association between \(X_{ij}(\cdot)\) and the conditional mean response

44
of $Y_{ij}$ at $t_{ij}$, and is the main object of interest. We assume $\varepsilon_{ij} = Z_{b,ij}b_i + e_{ij}$ and $\varepsilon_{ij}$ is the zero-mean residual error and describes the dependence across the measurements of the same subject, where $b_i$ is a $q$-dimensional subject-specific random vector such that $b_i \sim N_q(0, D)$, $Z_{b,ij}$ is the corresponding random design matrix, and $e_{ij}$ is the white noise such that $e_{ij} \sim N(0, \sigma^2_e)$. Assume $b_i$ and $e_{ij}$ are mutually independent. We demean the responses and remove $\alpha(\cdot)$ from the rest of the paper; denote the population mean function by

$$\mu_{ij} = \int_s X_{ij}(s)\gamma(s, t_{ij})ds,$$

and let $\mu_{ij} = \mu_i(t_{ij})$. Model (3.1) has been discussed in Islam et al. (2017) and Kundu et al. (2016). Multiple functional predictors and/or scalar covariates can be easily accommodated in the model (3.1) without loss of any generality.

The intuitive assumption in (3.1) is that the functional structure of the predictor is indeed time-varying and contains useful information to explore the association with scalar responses. However, it is possible that the relevant effects are captured by a much simpler time-invariant coefficient function as described in Goldsmith et al. (2012a) and Gertheiss et al. (2013a). Therefore, testing for the goodness of fit in (3.1) is of practical interest in many situations and we describe it next.

3.2.2 Testing for time-varying functional effect

We investigate whether a time-invariant coefficient function would be sufficient to describe the association between the mean response and functional covariate as opposed to a time-varying one. Formally, we write the hypothesis of interest for testing the time-varying functional effect for a generic time $t$ by

$$H_0 : \gamma(s, t) = \gamma(s), \text{ for all } t$$

$$H_A : \gamma(s, t) \neq \gamma(s), \text{ for some } t.$$

Note that under null hypothesis, the resulting model reduces to the model described by Goldsmith et al. (2012a) and Gertheiss et al. (2013a); where the effect of a functional predictor is assumed to be constant over time. In our motivating data application, a lactating sow’s body can adjust to the prolonged exposure to thermal environment; as a result its physiological behavior such as feeding habit might also change over time. Therefore, testing whether the effect changes over time has the potential to help understanding the change in
the physiological behavior of sows.

Let \( \{ \phi_k(\cdot) \}_{k \geq 1} \) be the time-invariant orthonormal basis functions in \( L^2(\mathcal{S}) \) such that \( \int_{\mathcal{S}} \phi_k(s) \phi_{k'}(s') ds = 1 \) for \( k = k' \) and zero otherwise. We express the functional predictor as \( X_i(s, t_{ij}) = \sum_{k \geq 1} \xi_{ik}(t_{ij}) \phi_k(s) \), where \( \xi_{ik}(t_{ij}) = \xi_{ijk} \)'s are the \( k \)th time-varying basis coefficients at time \( t_{ij} \) such that \( \xi_{ijk} = \int_{\mathcal{S}} X_i(s, t_{ij}) \phi_k(s) \). We use the same basis \( \phi_k(\cdot)'s \) to approximate the corresponding regression coefficient; i.e. \( \gamma(s, t_{ij}) = \sum_{k \geq 1} \beta_k(t_{ij}) \phi_k(s) \) where \( \{ \beta_k(\cdot) \}_{k \geq 1} \)'s are unknown smooth functions of time. Using the orthonormal properties of \( \phi_k(\cdot)'s \), we write \( \int_{\mathcal{S}} X_i(s, t_{ij}) \gamma(s, t_{ij}) ds = \sum_{k \geq 1} \xi_{ijk} \beta_k(t_{ij}) \). It follows that an approximating mean model of (3.2) can be written as

\[
\mu_{ij} \approx \sum_{k=1}^{K} \xi_{ijk} \beta_k(t_{ij}). \tag{3.4}
\]

Define the first-order partial derivative of \( \gamma(\cdot, \cdot) \) by \( \partial \gamma / \partial t = \gamma_t \); in general, \( \partial^p \gamma / \partial t^p = \gamma_{t^p} \) is the \( p \)th-order partial derivative of \( \gamma(\cdot, \cdot) \). Likewise, \( \beta_{t,k}(\cdot) \) and \( \beta_{t-t,k}(\cdot) \) are the first and \( p \)th order partial derivatives of \( \beta_k(\cdot) \). If the effect is invariant over \( t \), then the first derivative of \( \gamma(\cdot, \cdot) \) with respect to \( t \) should be zero. Therefore, we express the hypothesis of interest (3.3) as \( H_0 : \gamma_t(s, t) = 0 \) or equivalently, \( \| \gamma_t \|^2 = 0 \), where \( \| \gamma_t \|^2 = \int_{\mathcal{S}} \int_{\mathcal{T}} \gamma_t(s, t) ds dt \). The latter part is equal to

\[
0 = \| \gamma_t \|^2 = \int_{\mathcal{T}} \int_{\mathcal{S}} \left[ \sum_{k \geq 1} \beta_{t,k}(t) \phi_k(s) \right] \left[ \sum_{k' \geq 1} \beta_{t,k'}(t) \phi_k'(s) \right] ds dt
= \sum_{k, k' \geq 1} \left\{ \int_{\mathcal{S}} \phi_k(s) \phi_{k'}(s) ds \right\} \left\{ \int_{\mathcal{T}} \beta_{t,k}(t) \beta_{t,k'}(t) dt \right\} = \sum_{k \geq 1} \int_{\mathcal{T}} \beta_{t,k}(t)^2 dt = \sum_{k \geq 1} \| \beta_{t,k} \|^2,
\]

since \( \int_{\mathcal{S}} \phi_k(s) \phi_{k'}(s) ds = 1 \) if \( k = k' \) and 0 otherwise. Thus, hypothesis (3.3) is equivalent to

\[
H_0 : \beta_{t,k}(t) = 0, \quad \forall k, \tag{3.5}
\]

versus the alternative \( H_A : \) for at least one \( \beta_{t,k}(t) \) is non-zero for some \( t \).

The infinite summation in the above expression is intractable; instead we study the hypothesis (3.3) in an approximating model with a finite truncation value; say \( K \). The optimal number \( K \) needs to be selected empirically and we discuss it further in Section 3.3. In general \( K \) provides a means of identifying the number of major sources of variation in a set of functional measurements. Therefore, the truncation value \( K \) controls the smoothness of the coefficient function \( \gamma(s, t) \) in the \( s \)-direction; specifically, a small value of \( K \) leads to over-smooth function in this direction while a large value \( K \) results in a wiggly surface in this
direction. Assuming that $\gamma(\cdot, \cdot)$ is a second order differentiable function, we can characterize the smoothness in the $t$-direction by the total curvature in this direction, $\|\partial^2 \gamma/\partial t^2\|^2 = \int_T \int_S \partial^2 \gamma/\partial t^2 ds dt$. Let $Y_i$ be a $m_i$-dimensional vector of responses $Y_{ij}$’s associated with the $i$th subject and $\mu_i$ be a $m_i$-dimensional vector where the $j$th element is equal to the mean function $\mu_{ij}$ as in (3.2). We estimate $\gamma(\cdot, \cdot)$ via the minimization of the penalized criterion

$$\sum_{i=1}^n \|Y_i - \mu_i\|^2_{\Gamma_i} + \lambda \text{Pen}_t(\gamma),$$

where $\|U\|^2_A = U^T A^{-1} U$ and $\lambda$ is a non-negative smoothing parameter that controls the goodness of fit and smoothness of $\gamma(\cdot, \cdot)$. Moreover $\text{Pen}_t(\gamma)$ is the penalty in direction $t$ of $\gamma(\cdot, \cdot)$. Since $\gamma(\cdot, \cdot)$ is represented using the orthogonal bases over $S$, then it makes sense to assume that $\text{Pen}_t(\gamma) = \sum_{k=1}^K \text{Pen}_t(\beta_k)$. For example, $\text{Pen}_t(\beta_k) = \int_T \beta_k''(t) dt$ if $\beta_k(\cdot)$ is modeled using a finite B-splines basis, or $\text{Pen}_t(\beta_k) = ||u||^2$ if $\beta_k(\cdot)$ is modeled using a truncated polynomial splines where $u$ is a vector of random parameters associated with the random splines terms. It follows that the above penalized criterion can be approximated by:

$$\sum_{i=1}^n \|Y_i - \mu_i,\xi\|^2_{\Gamma_i} + \lambda \sum_{k=1}^K \text{Pen}_t(\beta_k); \quad (3.6)$$

where $\mu_i,\xi$ is a $m_i$-dimensional vector with the $j$th element obtained from (3.4) using the scores $\xi_{ijk}$’s; $\mu_i,\xi$ is an approximation of $\mu_i$. The penalized criterion (3.6) controls the smoothness of the unknown bivariate function $\gamma(\cdot, \cdot)$ through two parameters, $K$ in the $s$-direction and $\lambda$ in the $t$-direction. This approach of smoothing in two directions is a variant of the conventional approaches (Ruppert et al., 2003; Wood, 2017) and is inspired by Kim et al. (2017); Islam et al. (2017).

We use finite basis expansion to approximate smooth coefficients and represent the approximating model of (3.1) in a linear mixed model (LMM) framework. Such approach is commonly adopted in nonparametric and semiparametric regression; see for details Verbyla (1999); Pinheiro & Bates (2000); Ruppert et al. (2003). Popular choices for bases are, but not limited to, natural cubic splines, truncated polynomial splines, piecewise constant splines, and so on. For simplicity of exposition, we use truncated polynomial splines for basis expansions. We represent each smooth coefficient by

$$\beta_k(t) \approx \beta_{0k} + \beta_{1k} t + \ldots + \beta_{pk} t^p + \sum_{r=1}^{R_k} u_{rk}(t - \kappa_{rk})^p_+; \quad (3.7)$$
where \( \{ \beta_{lk} \}_{l=0}^p \) are fixed parameters, \( \{ u_{rk} \}_{r=1}^{R_k} \)'s are random parameters, and \( \kappa_{r1}, \ldots, \kappa_{rR_k} \) are \( R_k \) equispaced knots in \( T \). Here \( (x)^p \) is \( \max(0, x^p) \). Denote by \( \beta_k = (\beta_{0k}, \beta_{1k}, \ldots, \beta_{pk})^T \) the vector of fixed parameters and by \( u_k = (u_{1k}, \ldots, u_{R_k})^T \) the vector of random parameters. The choice of \( R_k \) is selected to ensure the desired flexibility; the main intuition of treating \( u_k \) as random parameter is to avoid overfitting. In view of the penalty term in (3.6), the random vector \( u_k \) is assumed to be independent and distributed as \( u_k \sim \mathcal{N}(0_{R_k}, \sigma^2_R I_{R_k}) \) with \( I_{R_k} \) being the \((R_k \times R_k)\) identity matrix and \( \sigma^2_R = \lambda^{-1} \).

Assume \( R_1 = \ldots = R_K = R \) and \( \kappa_{rk} = \kappa_r \) \( \forall k \). Define \( X_{ijk} = [\xi_{ijk}, t_{ij}\xi_{ijk}, \ldots, t_{ij}^p\xi_{ijk}] \) and \( X_{ij} = [X_{ij1} \ldots X_{ijK}] \) which is obtained by stacking columnwise a \((p + 1)K\)-dimensional row vector \( X_{ijk} \)'s. Let \( \beta = (\beta_{1T}, \ldots, \beta_{K}^T)^T \) be the corresponding vector of fixed effects. Denote \( Z_{ijk} = [\xi_{ijk}(t_{ij} - \kappa_1)^t, \ldots, \xi_{ijk}(t_{ij} - \kappa_R)^t] \) the \( R \)-dimensional vector that constructs the random design matrix for the \( k \)th smooth parameter. Write \( Z_{ij} = [Z_{ij1} \ldots Z_{ijK}] \) and \( u = (u_1^T, \ldots, u_K^T)^T \) by appropriately stacking \( Z_{ijk} \)'s and \( u_k \)'s.

Let \( X_i \) be a \( m_i \times (p + 1)K \) dimensional matrix of \( X_{ij} \)'s and \( Z_i \) be a \( m_i \times RK \) dimensional matrix of \( Z_{ij} \)'s. Let \( Z_{b,i} \) be a \( n_i \times q \) matrix with the \( j \)th row equal to \( Z_{b,ij} \). If \( Y_i \) is a \( m_i \)-dimensional vector of demeaned responses \( Y_{ij} \)'s and \( e_i \) is a \( m_i \)-dimensional vector of \( e_{ij} \)'s, then our modeling framework corresponding to (3.1) is

\[
\begin{align*}
Y_i &\sim X_i \beta + Z_i u + Z_{b,i} b_i + e_i, \\
u &\sim \mathcal{N}(0, \sigma^2(I_K \otimes I_R)), \\
b_i &\sim \mathcal{N}(0, D), \\
e_i &\sim \mathcal{N}(0, \sigma^2_e I_{m_i}), \\
u, b_i, \text{ and } e_i &\text{ mutually independent;}
\end{align*}
\] (3.8)

here \( \otimes \) indicates the Kronecker product. Define the overall covariance structure of \( Y_i \) conditioning on \( u \) by \( \Gamma_i = Z_{b,i} D Z_{b,i}^T + \sigma^2_e I_{m_i} \) which is a positive-definite matrix. Note that, any other basis functions, say natural cubic splines, can also be used to represent the model in a mixed model framework with minor changes in basis construction; see Supplementary Material, Section B.3 for details. Next we relate the null hypothesis in terms of the model parameters of (3.8) and detail the testing procedure.
3.2.3 Testing procedure

The representation (3.8) allows for testing the null hypothesis (3.3) using (R)LRTs (Greven et al., 2012; Crainiceanu, 2008; Crainiceanu et al., 2005; Scheipl et al., 2008; Staicu et al., 2014; Wiencierz et al., 2011) which is one of the key advantages of using LMM framework. Many hypotheses of interest about the structure of the regression parameter $\gamma(\cdot,\cdot)$ can be formulated equivalently as hypotheses about the fixed effects and the variance parameter $\lambda^{-1} = \nu$. For example, for testing the adequacy of a time-varying functional coefficient as in (3.3) or equivalently (3.5) can be formulated as

$$H_0 : \beta_{11} = \ldots = \beta_{1K} = \beta_{p1} = \ldots = \beta_{pK} = 0 \text{ and } \sigma_s^2 = 0 \quad (3.9)$$

against $H_A : \text{at least one of the } \beta_{11}, \ldots, \beta_{1K}, \ldots, \beta_{p1}, \ldots, \beta_{pK} \text{ is non-zero and/or } \sigma_s^2 > 0$. Under null hypothesis (3.9) the null values of some parameters are on the boundary of the parameter space (i.e., $\sigma_s^2 = 0$), while for others (e.g., fixed parameters) they are not, and thus a LRT would be inappropriate. Inferentially, the $p$-values associated with the hypothesis can be calculated from the nonstandard (R)LRTs following the procedures described in Crainiceanu et al. (2005); Crainiceanu & Ruppert (2004) when $\Gamma_i = \sigma_s^2 I_{m_i}$. To handle the additional variance in the data due to the random effects $b_i$, we consider a pseudo LRT (pLRT) as in Wiencierz et al. (2011) and Staicu et al. (2014). Briefly, the idea of pLRT follows the same intuition of LRT with the difference that the covariance is replaced by a consistent estimator of $\Gamma_i$; Staicu et al. proved that the null distribution of the pLRT is the same as the null distribution of the LRT using the true covariance.

Let $N = \sum_{i=1}^n m_i$ be the total number of observations and $Y$ be a $N \times 1$ vector of responses. Let $\Gamma$ be a $N \times N$ block diagonal matrix having the $i$th block equal to $\Gamma_i$. For known covariance $\Gamma$, the corresponding LRT is defined as

$$\text{LRT}_N = 2 \sup_{H_A} \left\{ \log L_Y(\beta, \sigma_s^2) \right\} - 2 \sup_{H_0} \left\{ \log L_Y(\beta, \sigma_s^2) \right\};$$

where $2 \log L_Y(\beta, \sigma_s^2)$ is twice the log-likelihood for the demeaned response $Y$. However, $\Gamma$ is unknown in practice and requires estimation. In the similar spirit to Wiencierz et al. (2011) and Staicu et al. (2014), obtain the covariance estimate $\hat{\Gamma}$ using restricted maximum likelihood (REML) approach in the mixed effects model (3.8). Transform $Y$ by premultiplying with $\hat{\Gamma}^{-1/2}$; i.e. $\tilde{Y} = \hat{\Gamma}^{-1/2}Y$. Therefore the pseudo LRT statistic for testing the null hypothesis is as follows

$$\text{pLRT}_N = 2 \sup_{H_A} \left\{ \log L_{\tilde{Y}}(\beta, \sigma_s^2) \right\} - 2 \sup_{H_0} \left\{ \log L_{\tilde{Y}}(\beta, \sigma_s^2) \right\}. $$

49
The asymptotic distribution of \( p_{\text{LRT}}^{N} \) under the null hypothesis (3.9) with \( pK \) fixed parameters and a variance parameter is

\[
p_{\text{LRT}}^{N} \xrightarrow{D} \sup_{\nu \geq 0} \text{LRT}_{\infty}(\nu) + \sum_{u=1}^{pK} v_{u}^{2},
\]  

(3.10)

where \( \xrightarrow{D} \) denotes the convergence in distribution, \( \text{LRT}_{\infty}(\nu) = \sum_{m=1}^{M} \frac{\nu}{\Gamma(\nu \kappa_{m})} w_{m}^{2} - \sum_{m=1}^{M} \log(1 + \nu \kappa_{m}) \) with \( w_{m} \sim \mathcal{N}(0, \zeta_{m}) \) for \( m = 1, \ldots, M \), \( v_{u} \sim \mathcal{N}(0, 1) \) for \( u = 1, \ldots, pK \), and \( \zeta_{m} \) and \( \kappa_{m} \) are the nonnegative eigenvalues for some matrices, and additionally \( w_{m} \)’s and \( v_{u} \)’s are mutually independent; see Staicu et al. (2014) for details. Moreover, the right hand side is the null distribution of the corresponding LRT based on the true model covariance \( \Gamma \) (Crainiceanu & Ruppert, 2004). The asymptotic null distribution of \( p_{\text{LRT}}^{N} \) is non-standard and can be simulated easily following Crainiceanu & Ruppert (2004). For completeness, we review the algorithm as below.

(S1) For large \( Q \), define a grid of possible values of the smoothing parameter \( \nu \) as \( 0 = \nu_{1} < \ldots < \nu_{Q} \). In general, the grid of \( \nu \) needs to be very fine in the neighborhood of 0.

(S2) Simulate independent \( \mathcal{N}(0, \zeta_{m}) \) associated with random variables \( w_{m} \).

(S3) Compute \( \text{LRT}_{\infty}(\nu) \) in (3.10) and choose its maximizer \( \nu_{\text{max}} \) on the grid.

(S4) Obtain \( p_{\text{LRT}}^{N} = \text{LRT}_{\infty}(\nu_{\text{max}}) + \sum_{u=1}^{pK} v_{u}^{2} \); where \( v_{u} \) is generated as \( v_{u} \sim \mathcal{N}(0, 1) \).

(S5) Repeat Steps 2-4 until the desired number of simulations is achieved.

Inferential properties like the \( p \)-values or the critical values of the distribution of \( p_{\text{LRT}}^{N} \) can be obtained efficiently by 100,000 simulations which is implemented in the \texttt{RLRsim} (Scheipl & Bolker, 2016) package in \texttt{R} (Team, 2017).

### 3.3 Selection of basis functions

The selection of orthogonal bases is critical as it impacts the selection of the finite truncation \( K \); in particular an empirically selected basis tends to be more parsimonious compared to a pre-specified basis. We consider two data-driven approaches that are based on the eigenbasis of different covariance functions: one that is based entirely on the latent signal, and the other uses both the response and the latent signal.
Eigenbasis of the covariance of the latent functional predictor. Let $E[X_i(s, t_{ij})] = 0$ for all $i$ and $j$. Assume that $X_i(\cdot, t_{ij})$ consists of the subject-specific deviation that varies over time. Define the marginal covariance function induced by the latent signal $X_i$ by $\Sigma(s, s') = \int_T E[X_i(s, t)X_i(s', t)]h(t)dt$, where $h(\cdot)$ is the sampling density of $t_{ij}$’s; see Park & Staicu (2015) for justification that this function is a proper covariance function so that it is nonsingular, positive semidefinite, and symmetric. Let $\{\phi_k(\cdot), \eta_k\}_k$ be the eigen-components of $\Sigma(s, s')$; where $\phi_k(\cdot)$’s are the time-invariant orthonormal eigen functions in $L^2(S)$ such as $\int_S \phi_k(s)\phi_{k'}(s') = 1$ for $k = k'$ and zero otherwise, and $\eta_k$’s are the corresponding eigenvalues.

Recall that the true signal is observed with error and the observed functional covariate is assumed to be modeled as $W_{ij}(s) = X_i(s, t_{ij}) + \epsilon_{ij}^w(s)$. Assume that $\epsilon_{ij}^w(\cdot)$ is independent over both $i$ and $j$, and is a white noise with mean $E[\epsilon_{ij}^w(s)] = 0$ and covariance $cov(\epsilon_{ij}^w(s), \epsilon_{ij}^w(s')) = \sigma_w^2 \mathds{1}(s = s')$. In practice, the estimation of $\Sigma(s, s')$ is done in two steps. In the first step, using the demeaned $\bar{W}_{ijr}$’s, the pooled sample covariance is obtained as $\bar{\Sigma}(s_r, s_r) = \sum_{i=1}^n \sum_{j=1}^m \bar{W}_{ijr}\bar{W}_{ijr}'/\sum_{i=1}^n m_i$; where $\bar{\Sigma}(s_r, s_r)$ is a method of moments estimator of $\Sigma(s_r, s_r)$. This estimator is not smooth and may be viewed as a raw estimator of $\Sigma(s_r, s_r)$. The diagonal terms of $\bar{\Sigma}(s_r, s_r)$ are possibly inflated. One can ignore the diagonal terms and smooth the off-diagonal terms using a bivariate smoother (Yao et al., 2005; Xiao et al., 2013). Denote the consistent smooth covariance estimator of $\Sigma(s, s')$ by $\hat{\Sigma}(s, s')$ which is estimated by Xiao et al. (2013); we refer to Yao et al. (2005); Park & Staicu (2015) for details. Let $\{\hat{\phi}_k(\cdot), \hat{\eta}_k\}_{k=1}^K$ be the pairs of eigenfunctions and eigenvalues obtained from the spectral decomposition of $\hat{\Sigma}(s, s')$; where $\int_S \hat{\phi}_k(s)\hat{\phi}_k(s)ds = 1$ if $k = k'$ and 0 otherwise, and $\hat{\eta}_1 \geq \cdots \geq \hat{\eta}_K \geq 0$. The scores $\xi_{ijk}$’s can be estimated either through numerical integration or as random effects in a mixed model (Yao et al., 2005; Crainiceanu et al., 2009; Di et al., 2009; Xiao et al., 2013; Swihart et al., 2014; Goldsmith et al., 2012b); e.g., $\hat{\xi}_{ijk} = \int_S \hat{W}_{ij}(s)\hat{\phi}_k(s)ds$.

Let the truncation value $K$ be such that the first $K$ pairs of eigenfunctions/eigenvalues provide a low-rank approximation of $\hat{\Sigma}(s, s')$: $\hat{\Sigma}(s, s') \approx \sum_{k=1}^K \hat{\eta}_k \hat{\phi}_k(s)\hat{\phi}_k(s')$. Here the optimum choice of $K$ can be guided by the proportion of variance explained (PVE); such approach implicitly assumes that the $K$ leading eigenbasis functions are the most informative to predict the responses. While such technique of dimension reduction has been widely employed in functional regression (Reiss & Ogden, 2007; Crainiceanu et al., 2009; Febrero-Bande et al., 2017; Di et al., 2009; Staicu et al., 2014, 2010), it is deemed as a strong limitation since it does not consider the correlation between the functional predictor and response (Islam et al., 2017). The next approach accounts for this dependence.
Eigenbasis of the covariance between the latent functional predictor and the response.

Instead of using the marginal covariance function of the predictor, we propose using the cross-covariance function of the covariate and response; this idea is motivated from Cardot et al. (2003b). Such approach borrows information across functional predictor and response, and guides to have the components that are most strongly associated with the outcome variable. For a generic time \( t \in T \), denote the signal by \( \tilde{Y}(t) = \int_S X_i(s, t) \gamma(s, t) ds \). Transform the functional measurement at \( t_{ij} \) by multiplying with the response as \( V_i(s, t_{ij}) = \tilde{Y}(t_{ij})X_i(s, t_{ij}) \). Assume that \( V_i(\cdot, t_{ij}) = V_{ij}(\cdot) \) is a square integrable process defined in \( S \). Let \( E[V_i(s, t)] = \tau(s, t) \) and define the marginal covariance function induced by \( V_i \) by \( \Sigma_v(s, s') = \int_T E[\{V_i(s, t) - \tau(s, t)\}\{V_i(s', t) - \tau(s', t)\}]h(t)dt \), where \( h(\cdot) \) is the sampling density of \( t_{ij} \)'s. We show that \( \Sigma_v(s, s') \) is a proper covariance function by justifying that \( \Sigma_v(s, s') \) is symmetric, positive-definite, and the sum of eigenvalues of \( \Sigma_v(s, s') \) is finite; see the Supplementary Material, Section B.1. The spectral decomposition of \( \Sigma_v(s, s') \) is

\[
\Sigma_v(s, s') = \sum_{k \geq 1} \eta_k \phi_k(s)\phi_k(s')
\]

where \( \phi_k(\cdot)'s \) are the orthonormal eigenfunctions and \( \eta_k \)'s are the corresponding nonnegative eigenvalues.

In reality, \( X_i(s, t_{ij}) \) is observed with measurement error, so as the responses \( Y_i(t_{ij}) \). Define \( \tilde{V}_{ij}^w(s, t_{ij}) = \tilde{Y}(t_{ij})W_i(s, t_{ij}) \) where \( \tilde{Y}(t_{ij}) \)'s are the demeaned responses. We model the mean of \( V_i^w(s, t_{ij}) \) as a bivariate smooth function and fit a bivariate smoother to estimate it under a working independence assumption (Wood, 2017). We then demean the response-adjusted functional predictor and denote the demeaned data by \( \tilde{V}_{ij}^w(\cdot) \). Next \( \tilde{V}_{ij}^w(\cdot)'s \) are used to estimate the covariance function \( \Sigma_v(s, s') \). As before, using a method of moments approach, we first estimate the sample covariance estimator \( \check{\Sigma}_v(s, s') \) and then smooth the off-diagonal elements of \( \check{\Sigma}_v(s, s') \) to remove the effect of measurement error. Denote the smooth covariance estimator by \( \hat{\Sigma}_v(s, s') \) using Xiao et al. (2013). We perform eigenanalysis of the estimated smooth covariance, and obtain the pairs of eigenvalues and eigenfunctions \( \{\hat{\phi}_k(\cdot), \hat{\eta}_k\}_{k=1}^K \); here \( K \) is chosen using PVE value. The scores can be estimated as numerical integration \( \check{S}_{ijk} = \int_S \tilde{W}_{ij}(s)\hat{\phi}_k(s) ds \) or via mixed effects model; see Crainiceanu et al. (2009); Goldsmith et al. (2012a); Di et al. (2009).

### 3.4 Testing different classes of hypotheses

The framework we proposed allows to study the more general null hypotheses than the one discussed in Section 3.2. In this section, we detail the procedure of testing other null
hypotheses such as testing the nullity, constancy, and pattern of change of time-varying coefficient function.

- **Testing for constancy.** In many situations, it is of interest to investigate whether the functional structure as in (3.1) is needed or a simple summary statistics of these functional measurements is sufficient to explain the association between covariates and responses. Assume $\gamma(s, t) = \beta_x$ for all $s$ and $t$; where $\beta_x$ is an unknown parameter capturing the effect of summary statistics $\int_S X_i(s, t) ds = X^*_i$ on scalar responses. Therefore, the hypothesis to test the adequacy of the structure of a functional form can be written equivalently as

$$ H_0 : \gamma(s, t) = \beta_x, \text{ for all } s \text{ and } t, $$

$$ H_A : \gamma(s, t) \neq \beta_x, \text{ for some } s \text{ and } t. $$

(3.11)

In a similar vein, Swihart et al. (2014) too considered testing of the adequacy of functional structure for a scalar-on-function regression. However, their method is not designed to test for the time-varying coefficient function.

- **Testing for linear time effect.** In the situation, when the effect of a functional predictor evolves over time, it is of interest to investigate whether the effect follows any specific pattern such as linear, quadratic, or cubic. Consider the model (3.1) and recall that $\gamma_{tt}(s, t)$ is the second order partial derivative of $\gamma(\cdot, \cdot)$ with respect to $t$. Testing whether $\gamma(\cdot, \cdot)$ changes linearly over time is equivalent to testing the hypothesis

$$ H_0 : \gamma(s, t) = \gamma_1(s) + t\gamma_2(s), \text{ for all } s, $$

$$ H_A : \gamma(s, t) \neq \gamma_1(s) + t\gamma_2(s), \text{ for some } s. $$

(3.12)

Under null hypothesis, $\gamma_1(\cdot)$ and $\gamma_2(\cdot)$ denote the time-invariant coefficient functions and the model is reduced to the one described in Kundu et al. (2016) with linear temporal functions. Therefore, this test can be used efficiently to select the degree of polynomial basis functions in Kundu et al. (2016) along with the AIC-based selection criterion. Following the same argument as before and using the LMM framework (3.8), it follows that the hypothesis of interest can be formulated as

$$ H_0 : \beta_{21} = \ldots = \beta_{2K} = \beta_{p1} = \ldots = \beta_{pK} = 0 \text{ and } \sigma_s^2 = 0, \quad (3.13) $$
against \( H_A : \) at least one of \( \beta_{21}, \ldots, \beta_{2K}, \ldots, \beta_{p1}, \ldots, \beta_{pK} \) is non-zero and/or \( \sigma_s^2 > 0 \). In addition, one can also test whether the smooth coefficient is well approximated by a \( p \)th degree polynomial function of time. In this case, one only needs to test \( H_0 : \sigma_s^2 = 0 \) versus the alternative \( H_A : \sigma_s^2 > 0 \).

### 3.4.1 Testing for multiple predictors

Consider \( \mu_i(t_{ij}) = \alpha(t_{ij}) + \sum_{l=1}^2 \int_s X_{l,i}(s) \gamma_l(s, t_{ij}) ds \), where \( l \) indexes the functional predictor. Let \( k_l \) index the FPCs such that \( k_l = 1, \ldots, K_l \) where \( K_l \) is the total number of FPCs associated with the \( l \)th predictor. Denote the corresponding time-varying parameters by \( \beta_{lk_l} \)'s. Let \( \beta_{lk_l} = (\beta_{0k_l}, \beta_{1k_l}, \ldots, \beta_{pk_l})^T \) and \( \beta_{lk_l}^c = \beta_{lk_l} \setminus \{ \beta_{0k_l} \} \) for \( l = 1, 2 \) and \( k_l = 1, \ldots, K_l \). Let the variance of splines terms be \( \sigma_{s1}^2 \) and \( \sigma_{s2}^2 \) for the first and second functional predictor respectively.

- **Testing for nullity.** Before determining whether a complicated functional linear model with multiple predictors is necessary, it is of interest to assess whether there is any effect of a particular predictor on response at all. Such hypothesis testing can eliminate a predictor out of the model as the test becomes statistically insignificant; therefore, this test can be viewed as an alternative to the functional variable selection problem discussed in Gertheiss et al. (2013b); Pannu & Billor (2017); Islam et al. (2018). For example, we write the hypothesis of nullity of the second predictor as

\[
H_0 : \gamma_1(s, t) \neq 0 \text{ and } \gamma_2(s, t) = 0, \text{ for all } s \text{ and } t,
\]

\[
H_A : \gamma_1(s, t) \neq 0 \text{ and } \gamma_2(s, t) \neq 0, \text{ for some } s \text{ and } t.
\]

Using the representation of (3.8), we formulate the hypothesis of interest in terms fixed and variance parameters as

\[
H_0 : \beta_{21} = \cdots = \beta_{2K_2} = 0 \text{ and } \sigma_{s2}^2 = 0,
\]

against \( H_A : \) at least one parameter in \( \beta_{21}, \ldots, \beta_{2K_2} \) is non-zero and/or \( \sigma_{s2}^2 > 0 \).

- **Simultaneous testing for nullity and time-invariant effect.** This might be of interest to test a time-invariant effect for one predictor while a null effect for the other covariate
simultaneously. We write formally the hypothesis as
\[ H_0: \gamma_1(s,t) = \gamma_1(s) \text{ and } \gamma_2(s,t) = 0, \text{ for all } s \text{ and } t, \]
\[ H_A: \gamma_1(s,t) \neq \gamma_1(s) \text{ and } \gamma_2(s,t) \neq 0, \text{ for some } s \text{ and } t. \]  
(3.16)

Using the similar argument as before, the hypothesis can be written in terms of fixed and variance parameter(s) as follows
\[ H_0: \beta_{c11} = \cdots = \beta_{c1K_1} = 0, \beta_{21} = \cdots = \beta_{2K_2} = 0, \text{ and } \sigma^2_{s1} = 0, \sigma^2_{s2} = 0, \]
(3.17)

against \( H_A: \) at least one parameter in \( \beta_{c11}, \ldots, \beta_{c1K_1} \) is non-zero, and/or \( \sigma^2_{s1} > 0 \), and/or \( \sigma^2_{s2} > 0 \). We assume same smoothness for both functional predictor which entails \( \sigma^2_{s1} = \sigma^2_{s2} = \sigma^2_{s} \); such assumption reduces to testing one variance component for equality to zero in (3.17) using pLRT (Staicu et al., 2014).

- **Simultaneous testing for linear time-varying effect and time-invariant effect.** We may also want to simultaneously test quadratic time-varying effect for one and time-invariant effect for another covariate in a model. We therefore wish to infer about
\[ H_0: \gamma_1(s,t) = \gamma_1(s) \text{ and } \gamma_2(s,t) = \gamma_1(s) + t\gamma_2(s) + t^2\gamma_3(s), \text{ for all } s \text{ and } t, \]
\[ H_0: \gamma_1(s,t) \neq \gamma_1(s) \text{ and } \gamma_2(s,t) \neq \gamma_1(s) + t\gamma_2(s) + t^2\gamma_3(s), \text{ for some } s \text{ and } t. \]  
(3.18)

Under null hypothesis, \( \gamma_1(\cdot), \gamma_2(\cdot), \text{ and } \gamma_3(\cdot) \) are time-invariant unknown smooth coefficients such as \( \gamma_1(s) + t\gamma_2(s) + t^2\gamma_3(s) = \gamma(s,t) \) which changes quadratically over \( t \). Using the similar arguments as before, the hypothesis (3.18) can be written alternatively in terms of fixed and variance parameters, and thus we omit it here.

### 3.5 Simulation

In this section, we investigate the theoretical properties of the inferential procedure developed in Section 2 and 3 for testing different classes of hypotheses by simulations. We generate the data \([t_{ij}, Y_{ij}, \{(W_{ijr}, s_r) : r = 1, \ldots R\}; i = 1, \ldots, n, j = 1, 2, \ldots, m_i]\) according to the following scenarios.
3.5.1 Description of the settings

(A1) Let \( X_i(s, t) = \tau(s, t) + \sum_{k=1}^{3} \xi_{ik}(t) \phi_k(s) \). Denote the FPCs by \( \phi_1(s) = \sqrt{2} \cos(\pi s) \), \( \phi_2(s) = \sqrt{2} \sin(3\pi s) \), and \( \phi_3(s) = \sqrt{2} \sin(5\pi s) \) and define the corresponding time-varying scores by \( \xi_{i1}(t) = \sqrt{2} \zeta_{i11} \cos(3\pi t) + \sqrt{2} \zeta_{i12} \sin(3\pi t) \), \( \xi_{i2}(t) = \sqrt{2} \zeta_{i21} \sin(6\pi t) + \sqrt{2} \zeta_{i22} \cos(6\pi t) \), and \( \xi_{i3}(t) = \sqrt{2} \zeta_{i31} \cos(9\pi t) + \sqrt{2} \zeta_{i32} \sin(9\pi t) \). Define the mean function by \( \tau(s, t) = 0.5 + \sin(\pi s) \). Denote the noisy predictors by \( W_i(s, t) = X_i(s, t) + \epsilon_i(s, t) \) where \( \epsilon_i(\cdot, \cdot) \) is the error term defined as \( \epsilon_i(s, t) = \sqrt{2} \cos(\pi s) \epsilon_{i1} + \sqrt{2} \sin(3\pi s) \epsilon_{i2} + \epsilon_{i3}(s) \).

(A2) Let \( X_i(s, t) = \tau(s, t) + \sum_{k=1}^{3} \xi_{ik}(t) \phi_k(s) \). Define the corresponding FPCs by \( \phi_1(s) = \sqrt{2} \cos(\pi s) \), \( \phi_2(s) = \sqrt{2} \sin(\pi s) \), and \( \phi_3(s) = \sqrt{2} \sin(3\pi s) \) and define the corresponding time-varying scores by \( \xi_{i1}(t) = \sqrt{2} \zeta_{i11} \cos(2\pi t) + \sqrt{2} \zeta_{i12} \sin(2\pi t) \), \( \xi_{i2}(t) = \sqrt{2} \zeta_{i21} \sin(4\pi t) + \sqrt{2} \zeta_{i22} \cos(4\pi t) \), and \( \xi_{i3}(t) = \sqrt{2} \zeta_{i31} \cos(6\pi t) + \sqrt{2} \zeta_{i32} \sin(6\pi t) \). Let the mean function be \( \tau(s, t) = 0.5 - s \). Write the corresponding noisy predictor by \( W_i(s, t) = X_i(s, t) + \epsilon_i(s, t) \); where \( \epsilon_i(\cdot, \cdot) \) denotes the measurement error represented as \( \epsilon_i(s, t) = \sqrt{0.5s} \epsilon_{i1} + \epsilon_{i2} + \epsilon_{i3}(s) \). Note that setting (A2) is different from (A1) in terms of basis functions and error structure.

Here, \( \zeta_{i11}, \zeta_{i12}, \zeta_{i21}, \zeta_{i22}, \zeta_{i31}, \) and \( \zeta_{i32} \) are assumed to be IID as \( \mathcal{N}(0, 3.5) \), \( \mathcal{N}(0, 3.5) \), \( \mathcal{N}(0, 2.5) \), \( \mathcal{N}(0, 2.5) \), \( \mathcal{N}(0, 0.5) \), and \( \mathcal{N}(0, 0.5) \) respectively. Here, \( \epsilon_{i1}, \epsilon_{i2}, \) and \( \epsilon_{i3}(s) \)'s are IID such as \( \mathcal{N}(0, \sigma_{\epsilon_1}^2) \), \( \mathcal{N}(0, \sigma_{\epsilon_2}^2) \), and \( \mathcal{N}(0, \sigma_{\epsilon_3}^2) \); where, \( \sigma_{\epsilon_1}^2 = 0.3, \sigma_{\epsilon_2}^2 = 0.7, \) and \( \sigma_{\epsilon_3}^2 \) is calculated using signal-to-noise-ratio (SNRw) defined as

\[
\text{SNR}_w = \frac{\int_T \int_S \text{var}(W_i(s, t)) ds dt}{\int_T \int_S \text{var}(\epsilon_i(s, t)) ds dt} - 1.
\]

(B1) Large noise variance \( \sigma_{\epsilon_1}^2 \): SNRw = 4.

(B2) Small noise variance \( \sigma_{\epsilon_2}^2 \): SNRw = 8.

We consider a dense functional design \( s: \{s_1, \ldots, s_H\} \) is taken as a grid of 101 equidistant points in \([0, 1]\).

(C) Generate scalar responses such as \( Y_i(t) = \int_S X_i(s, t) \gamma(s, t) ds + \varepsilon_{it} \). In all cases, \( \varepsilon_{it} \) follows Gaussian distribution.

(D) Compound symmetric (CS) structure: Consider the dependence structure as \( \varepsilon_{it} = b_{i0} + e_{it} \), where, \( b_{i0} \) and \( e_{it} \) are distributed as IID \( \mathcal{N}(0, \sigma_{b0}^2) \) and \( \mathcal{N}(0, \sigma_{e}^2) \) respectively, and
are mutually independent. We consider \( \sigma_{b_0}^2 = \{1, 5\} \) and \( \sigma_c^2 = \{1, 4\} \). \( \text{SNR}_y \) is calculated as
\[
\text{SNR}_y = \frac{\int_T \int_S \text{var}(X_i(s, t)) \, ds \, dt}{\int_T \text{var}(\varepsilon_{it}) \, dt}.
\]
We consider two sampling designs for longitudinal time points \( t_{ij} \)'s corresponding to the number of repeated measurements per subject \( m_i \):

\((E1)\) Moderate sparsity: \( m_i \in \{8, \ldots, 13\} \).

\((E2)\) Low sparsity: \( m_i \in \{14, \ldots, 19\} \).

In each case \{\( t_{i1}, \ldots, t_{im_i} \)\} are randomly chosen from a dense set of 41 equidistant points; i.e. \( t_{ij} \in [0, 1] \). We focus on testing the following hypotheses in the longitudinal framework:

\((F1)\) Test for time-invariant effect. \( H_0 : \gamma(s, t) = \gamma(s) \).

\((F2)\) Test for linear time-varying effect. \( H_0 : \gamma(s, t) = \gamma_1(s) + t\gamma_2(s) \); where \( \gamma_0(\cdot) \) and \( \gamma_1(\cdot) \) are time-invariant smooth coefficients.

\((F3)\) Test for time-invariant and null effect. \( H_0 : \gamma_1(s, t) = \gamma_1(s) \) and \( \gamma_2(s, t) = 0 \).

Let \( |\delta| \geq 0 \). Define the smooth coefficients \( \gamma(\cdot, \cdot) \)'s corresponding to each hypotheses described above as follows.

\((G1)\) Let \( \gamma(s, t) = \sqrt{2}[1 + \delta(t + \sin(2t))] \sin(5\pi s) \). Under null hypothesis \((F1)\) for \( \delta = 0 \), the covariate effect \( \gamma(s, t) = \sqrt{2} \sin(5\pi s) \) is time-invariant; similar to the one considered in Goldsmith et al. (2012a); Gertheiss et al. (2013a).

\((G2)\) Define \( \gamma(s, t) = \sqrt{2}[0.5 - t + \delta(2t^2 + \exp(-\pi t/2))] \sin(3\pi s) \). Under null hypothesis \((F3)\) for \( \delta = 0 \), we write \( \gamma(s, t) = \sqrt{2}(1 - t) \sin(3\pi s) \); where the effect changes linearly over \( t \) as in Kundu et al. (2016).

\((G3)\) Define \( \gamma_1(s, t) = \sqrt{2}[1 + \delta(t + \sin(2t))] \sin(5\pi s) \) and \( \gamma_2(s, t) = \sqrt{2}\delta(t^3 + \cos(3t)) \sin(3\pi s) \). Under null hypothesis \((F4)\) for \( \delta = 0 \), we have \( \gamma_1(s, t) = \sqrt{2} \sin(5\pi s) \) and \( \gamma_2(s, t) = 0 \) where the first effect is time-invariant and the second one is a null effect.

\((G4)\) Let \( \gamma(s, t) = \sqrt{2}[1 + \delta(t + \sin(2t))] \sin(3\pi s) \). Under null hypothesis \((F1)\) for \( \delta = 0 \), \( \gamma(s, t) = \sqrt{2} \sin(3\pi s) \) resembles a time-invariant effect.
We examine the cases under two assumptions:

(H1) Leading $K$ eigenbasis functions selected by 90% PVE are not associated with responses; this is investigated in $G1-G3$.

(H2) Leading $K$ eigenbasis functions selected by 90% PVE are most predictive of responses; this is investigated in $G4$.

The performance of the testing procedure is evaluated in terms of the empirical type-I error rates and power probabilities corresponding to the nominal levels $\alpha = \{0.01, 0.02, 0.05, 0.10\}$ and for increasing sample sizes $n = 100$ and $n = 200$. Our testing procedure relies on the true scores $\xi_{ijk}$’s of the latent functional covariate which we do not observe and instead need to predict. Thus we assess the performance of the testing procedure with the estimated scores relative to when the true scores are observed and furthermore when they are estimated directly from the latent functional signals.

(I1) $\xi_{ijk}$’s are known; we refer it to $\xi_T$.

(I2) $\xi_{ijk}$’s are estimated from the observed functional predictor; we refer it to $\xi_W$.

(I3) $\xi_{ijk}$’s are estimated from the response-adjusted functional covariate; we refer it to $\xi_V$.

3.5.2 Computational implementation

We estimate the mean function of the functional predictor using the fast bivariate smoother (Xiao et al., 2013) using the $\text{fbps}$ function with 35 knots in the $s$ and $t$ directions from the $\text{refund}$ (Crainiceanu et al., 2012a) package. We estimate the smooth marginal covariance function for the response-adjusted functional predictors using the sandwich bivariate smoother and perform the eigenanalysis on the estimated smooth covariance to obtain the empirical eigen components $\{\hat{\phi}_k(\cdot), \hat{\eta}_k\}_{k=1}^K$. We use the $\text{fpca.face}$ function from the $\text{refund}$ package; alternatively $\text{fpca.sc}$ can also be used from the same package. The truncation value $K$ is chosen using 90% PVE. We estimate the scores using numerical integration. We generate the basis functions to approximate the smooth coefficients by the $\text{smspline}$ function from the $\text{lmeSplines}$ (Ball, 2003) package where 39 knots are placed uniformly in the time domain. We fit the model using the $\text{lme}$ function from the $\text{nlme}$ (Pinheiro et al., 2009) package and estimate the covariance parameters $\hat{\Gamma}$ using REML. After de-noising the responses, we implement the pLRT test by the $\text{exactLRT}$ function from the $\text{RLRsim}$ (Scheipl & Bolker, 2016) package. Smoothing parameters are selected using REML throughout the analysis.
3.5.3 Performance assessments and comparisons

**Empirical type-I error rate.** Table 3.1 reports the probability of rejecting the null hypothesis and the corresponding standard errors for testing the time-invariant effect (F1) at different simulation settings under the case when indeed the null hypothesis is true. The results are based on 5,000 simulations and data is generated assuming high SNR (B2). Note that the testing procedure with true scores $\xi_T$ is an oracle testing procedure. The empirical type-I error rates are around the nominal levels across different settings and irrespective of whether the scores are estimated directly from the covariates ($\xi_W$) or the response-adjusted functional measurements ($\xi_V$). As the sample size increases, we observe that the type-I error rate decreases and remains closer to the nominal levels for both $\xi_W$ and $\xi_V$. We also observe that the numerical performance is slightly affected by estimating the scores; compare the results at 2% nominal level under $\xi_T$ and $\xi_V$. However the performance improves considerably as SNR increases which is not surprising indeed.

Table 3.2 provides the empirical type-I error rate for different simulation settings for testing the linear pattern of time-varying effect (F2). As previously, the method attains the nominal levels at most cases irrespective of the type of estimated scores (i.e., $\xi_W$ and $\xi_V$). However, at low sparsity with larger sample size, the test with respect to both approaches has slightly inflated rejection probabilities; see the columns associated with $\alpha = 0.10$ in Table 3.2. In contrast, the magnitude of inflation is relatively lower in the case when time-varying scores are known ($\xi_T$). We notice that inflated type-I error rate decreases with larger SNR; compare the columns at 5% nominal level under $\xi_T$ and $\xi_V$. Though not reported here, inflated type-I error rate decreases with simpler error process in the functional predictor (i.e., having only white noise with large SNR).

Table 3.3 illustrates the empirical type-I error rates for testing the time-invariant effect (F1) at different simulation settings under the assumption $H2$ where the leading eigencomponents are the most predictive of responses. The testing procedure based on $\xi_W$ and $\xi_V$ performs satisfactorily across different simulation settings. As SNR increases, the numerical performance improves; compare the last two rows for $n = 100$ and $n = 200$ for any nominal level. Based on our preliminary investigation, though not reported here, the test with respect to $\xi_V$ remains unaffected by increasing PVE say at 95% while the type-I error rate with respect to $\xi_W$ gets slightly inflated. We also investigate the numerical performances for F3 and observe similar findings in terms attaining the nominal size; the results are shown in Table B.1 in the Supplementary Material.
Power of the tests. Figure 3.1 illustrates the power properties of the test for testing the time-invariant effect (\(F1\)). As \(\delta\) departs from 0, the smooth coefficient \(\gamma(\cdot, \cdot)\) becomes more time-dependent and the method exhibits superior power performance across different settings. As expected, the power curves for the model with \(\xi_T\) (oracle) get closer to 1 for higher \(\delta\). Interestingly, we notice that the model with \(\xi_W\) fails consistently in producing the correct power; see the power curves corresponding to \(\xi_W\) for any setting in Figure 3.1. This happens because this model chooses basis functions without taking into account for correlation between the responses and predictors; under assumption \(H1\), the association between the responses and predictors are not through the leading eigenbasis functions of \(X_i\) and such assumption is more realistic in reality. Moreover, the rejection rates are not even improving as the sample size increases and/or the sparsity decreases for the test with \(\xi_W\). In contrast, we observe substantially higher power for the test with \(\xi_V\); indeed this approach uses the data-driven basis functions obtained from the response-adjusted functional covariate. As sample size increases, the performance of this method improves. We observe substantial improvement as \(\text{SNR}_y\) increases and the power curves get closer to that of the oracle test; compare the top and bottom panels for any setting. We also notice subtle improvement as the number of repeated instances per subject increases.

Figure 3.2 plots the power curves for testing linear time-varying effect (\(F2\)) for different sparsity, variance parameters, and size. As \(\delta\) departs from zero, the test with respect to \(\xi_V\) provides good power properties across different simulation settings; the difference between the power curves based on the estimated scores \(\xi_V\) and oracle scores \(\xi_T\) shrink for larger \(\delta\). Moreover, the empirical power performance improves as the sample size increases. In contrast, the power performance with respect to \(\xi_W\) suffers and does not improve even by changing the size, sparsity level, or choice of basis.

We also observe similar phenomenon for testing \(F1\) under \(H2\) in Figure 3.3. In contrast to the case depicted in Figure 3.1, the power curves are similar for both \(\xi_W\) and \(\xi_V\). This is not surprising since under the assumption \(H2\), the association between \(Y_{ij}\)'s and \(X_{ij}(\cdot)\)'s is through \(\gamma\) which is defined on the leading eigenbasis functions associated with the functional predictor. Moreover, the numerical performance improves as the number of repeated instances per subject increases; compare the first and third panels at the top for \(n = 100\). The power performance also improves with smaller subject-specific variance (smaller noise) which is not surprising. Additional simulation results related to the empirical power performances of the tests for scenarios \(F3\) are presented in the Supplementary Material, Figure B.1.
Table 3.1: Testing for the adequacy of time-varying effect ($F_1$) for the case $H_1$. The leading $K$ FPCs explain at least 90% variation of marginal covariance function. Reported are the observed type I error rates and standard errors (in parenthesis) based on 5000 Monte Carlo simulations. The functional covariate is observed as in $A1$ with high ($B2$) SNR$_{uv}$; the smooth functional coefficient is represented as in $G1$. Responses are generated assuming normality ($C$) and with CS type covariance structure ($D$). Both moderate ($E1$) and low ($E2$) sparse sampling designs are considered.

<table>
<thead>
<tr>
<th>size $m$</th>
<th>sparsity $m_s$</th>
<th>SNR$_y$</th>
<th>$\xi_F$</th>
<th>level of significance $\alpha$</th>
<th>$\xi_W$</th>
<th>level of significance $\alpha$</th>
<th>$\xi_V$</th>
<th>level of significance $\alpha$</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>mod</td>
<td>2</td>
<td>0.011</td>
<td>0.021</td>
<td>0.055</td>
<td>0.116</td>
<td>0.013</td>
<td>0.025</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.001)</td>
<td>(0.002)</td>
<td>(0.003)</td>
<td>(0.005)</td>
<td>(0.002)</td>
<td>(0.002)</td>
</tr>
<tr>
<td></td>
<td>low</td>
<td>2</td>
<td>0.009</td>
<td>0.020</td>
<td>0.050</td>
<td>0.107</td>
<td>0.016</td>
<td>0.031</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.001)</td>
<td>(0.002)</td>
<td>(0.003)</td>
<td>(0.004)</td>
<td>(0.002)</td>
<td>(0.002)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6</td>
<td>0.011</td>
<td>0.020</td>
<td>0.052</td>
<td>0.106</td>
<td>0.018</td>
<td>0.031</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.001)</td>
<td>(0.002)</td>
<td>(0.003)</td>
<td>(0.004)</td>
<td>(0.002)</td>
<td>(0.002)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6</td>
<td>0.009</td>
<td>0.021</td>
<td>0.062</td>
<td>0.111</td>
<td>0.019</td>
<td>0.034</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.001)</td>
<td>(0.002)</td>
<td>(0.003)</td>
<td>(0.004)</td>
<td>(0.002)</td>
<td>(0.003)</td>
</tr>
<tr>
<td>200</td>
<td>mod</td>
<td>2</td>
<td>0.011</td>
<td>0.021</td>
<td>0.059</td>
<td>0.114</td>
<td>0.015</td>
<td>0.028</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.001)</td>
<td>(0.002)</td>
<td>(0.003)</td>
<td>(0.005)</td>
<td>(0.002)</td>
<td>(0.002)</td>
</tr>
<tr>
<td></td>
<td>low</td>
<td>6</td>
<td>0.010</td>
<td>0.019</td>
<td>0.054</td>
<td>0.104</td>
<td>0.015</td>
<td>0.029</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.001)</td>
<td>(0.002)</td>
<td>(0.003)</td>
<td>(0.004)</td>
<td>(0.002)</td>
<td>(0.002)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>0.009</td>
<td>0.023</td>
<td>0.055</td>
<td>0.107</td>
<td>0.020</td>
<td>0.035</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.001)</td>
<td>(0.002)</td>
<td>(0.003)</td>
<td>(0.004)</td>
<td>(0.002)</td>
<td>(0.003)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6</td>
<td>0.011</td>
<td>0.020</td>
<td>0.053</td>
<td>0.102</td>
<td>0.019</td>
<td>0.034</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.001)</td>
<td>(0.002)</td>
<td>(0.003)</td>
<td>(0.004)</td>
<td>(0.002)</td>
<td>(0.003)</td>
</tr>
</tbody>
</table>
Table 3.2: Testing for linear pattern of time-varying effect \((F2)\) for the case \(H1\). The leading \(K\) FPCs explain at least 90% variation of marginal covariance function. Reported are the observed type-I errors and the corresponding standard errors (in parenthesis) based on 5000 Monte Carlo simulations. The functional covariate is observed as in \(A1\) with high \((B2)\) SNR\(_w\); the smooth functional coefficient is represented as in \(G2\). Responses are generated assuming normality \((C)\) and with CS type covariance structure \((D)\). Both moderate \((E1)\) and low \((E2)\) sparse sampling designs are considered.

<table>
<thead>
<tr>
<th>size (n)</th>
<th>sparsity (m_s)</th>
<th>SNR(_y)</th>
<th>(\xi_T) level of significance (\alpha)</th>
<th>(\xi_W) level of significance (\alpha)</th>
<th>(\xi_V) level of significance (\alpha)</th>
</tr>
</thead>
<tbody>
<tr>
<td>100 mod</td>
<td>2</td>
<td></td>
<td>0.013 0.025 0.058 0.113</td>
<td>0.014 0.023 0.054 0.103</td>
<td>0.013 0.023 0.054 0.104</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.002) (0.002) (0.003) (0.004)</td>
<td>(0.002) (0.002) (0.003) (0.004)</td>
<td>(0.002) (0.002) (0.003) (0.004)</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td></td>
<td>0.012 0.024 0.051 0.109</td>
<td>0.012 0.023 0.053 0.101</td>
<td>0.012 0.022 0.053 0.103</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.002) (0.002) (0.003) (0.004)</td>
<td>(0.002) (0.002) (0.003) (0.004)</td>
<td>(0.002) (0.002) (0.003) (0.004)</td>
</tr>
<tr>
<td></td>
<td>low</td>
<td>2</td>
<td>0.014 0.025 0.058 0.114</td>
<td>0.015 0.027 0.061 0.114</td>
<td>0.014 0.027 0.060 0.114</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.002) (0.002) (0.003) (0.004)</td>
<td>(0.002) (0.002) (0.003) (0.004)</td>
<td>(0.002) (0.002) (0.003) (0.004)</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td></td>
<td>0.009 0.021 0.054 0.109</td>
<td>0.013 0.023 0.056 0.112</td>
<td>0.013 0.023 0.056 0.112</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.001) (0.002) (0.003) (0.004)</td>
<td>(0.002) (0.002) (0.003) (0.004)</td>
<td>(0.002) (0.002) (0.003) (0.005)</td>
</tr>
<tr>
<td>200 mod</td>
<td>2</td>
<td></td>
<td>0.014 0.025 0.058 0.107</td>
<td>0.014 0.025 0.059 0.107</td>
<td>0.014 0.026 0.061 0.108</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.002) (0.002) (0.003) (0.004)</td>
<td>(0.002) (0.002) (0.003) (0.004)</td>
<td>(0.002) (0.002) (0.003) (0.005)</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td></td>
<td>0.011 0.024 0.057 0.110</td>
<td>0.012 0.023 0.058 0.109</td>
<td>0.012 0.023 0.058 0.109</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.001) (0.002) (0.003) (0.004)</td>
<td>(0.002) (0.002) (0.003) (0.004)</td>
<td>(0.002) (0.002) (0.003) (0.004)</td>
</tr>
<tr>
<td></td>
<td>low</td>
<td>2</td>
<td>0.013 0.026 0.063 0.114</td>
<td>0.016 0.027 0.061 0.116</td>
<td>0.016 0.028 0.061 0.116</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.002) (0.002) (0.003) (0.004)</td>
<td>(0.002) (0.002) (0.003) (0.005)</td>
<td>(0.002) (0.002) (0.003) (0.005)</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td></td>
<td>0.011 0.019 0.052 0.106</td>
<td>0.012 0.024 0.057 0.112</td>
<td>0.012 0.024 0.057 0.114</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.001) (0.002) (0.003) (0.004)</td>
<td>(0.002) (0.002) (0.003) (0.004)</td>
<td>(0.002) (0.002) (0.003) (0.004)</td>
</tr>
</tbody>
</table>
Table 3.3: Testing for the adequacy of time-varying effect ($F1$) for the case $H2$. The leading $K$ FPCs explain at least 90% variation of marginal covariance function. Reported are the observed type I error rates and standard errors (in parenthesis) based on 5000 Monte Carlo simulations. The functional covariate is observed as in $A1$ with high ($B2$) SNR; the smooth functional coefficient is represented as in $G4$. Responses are generated assuming normality ($C$) and with CS type covariance structure ($D$). Both moderate ($E1$) and low ($E2$) sparse sampling designs are considered.

<table>
<thead>
<tr>
<th>size</th>
<th>sparsity</th>
<th>SNR$_y$</th>
<th>level of significance $\alpha$</th>
<th>level of significance $\alpha$</th>
<th>level of significance $\alpha$</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>mod</td>
<td>0.010</td>
<td>0.021</td>
<td>0.056</td>
<td>0.112</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.001)</td>
<td>(0.002)</td>
<td>(0.003)</td>
<td>(0.004)</td>
</tr>
<tr>
<td></td>
<td>low</td>
<td>0.012</td>
<td>0.024</td>
<td>0.058</td>
<td>0.108</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.002)</td>
<td>(0.002)</td>
<td>(0.003)</td>
<td>(0.004)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.011</td>
<td>0.021</td>
<td>0.054</td>
<td>0.105</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.001)</td>
<td>(0.002)</td>
<td>(0.003)</td>
<td>(0.004)</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.011</td>
<td>0.025</td>
<td>0.057</td>
<td>0.104</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.001)</td>
<td>(0.002)</td>
<td>(0.003)</td>
<td>(0.004)</td>
</tr>
<tr>
<td>200</td>
<td>mod</td>
<td>0.011</td>
<td>0.021</td>
<td>0.055</td>
<td>0.105</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.001)</td>
<td>(0.002)</td>
<td>(0.003)</td>
<td>(0.004)</td>
</tr>
<tr>
<td></td>
<td>low</td>
<td>0.012</td>
<td>0.024</td>
<td>0.059</td>
<td>0.118</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.002)</td>
<td>(0.002)</td>
<td>(0.003)</td>
<td>(0.004)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.011</td>
<td>0.025</td>
<td>0.055</td>
<td>0.108</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.001)</td>
<td>(0.002)</td>
<td>(0.003)</td>
<td>(0.004)</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.009</td>
<td>0.019</td>
<td>0.049</td>
<td>0.103</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.001)</td>
<td>(0.002)</td>
<td>(0.003)</td>
<td>(0.004)</td>
</tr>
</tbody>
</table>
Figure 3.1: Observed power function for testing the adequacy of time-invariant effect ($F_1$) for low (top-left-to-right) and high (bottom-left-to-right) SNR$_y$. Reported are the proportion of rejections based on 1000 simulations for $\alpha = 0.05$ at $\delta = \{0.05, 0.20, 0.50, 1.00, 2.00\}$. The functional covariate is observed as in $A1$ with high SNR$_w$ ($B2$) while the smooth effect is represented by $G1$. 


Figure 3.2: Observed power function for testing linear pattern of time-varying effect ($F_2$) for low (top-left-to-right) and high (bottom-left-to-right) SNR$_y$. Reported are the proportion of rejections based on 1000 simulations for $\alpha = 0.05$ at $\delta = \{0.05, 0.20, 0.50, 1.00, 2.00\}$. The functional covariate is observed as in $A_2$ with high SNR$_w$ ($B_2$) while the smooth effect is represented by $G_2$. 
Figure 3.3: Observed power function for testing the adequacy of time-invariant effect ($F_1$) for the case $H_2$ for low (top-left-to-right) and high (bottom-left-to-right) SNR$_y$. Reported are the proportion of rejections based on 1000 simulations for $\alpha = 0.05$ at $\delta = \{0.05, 0.10, 0.25, 0.50, 1.00\}$. The functional covariate is observed as in $A1$ with high SNR$_w$ ($B2$) while the smooth effect is represented by $G_4$.

3.6 Data application

We apply the proposed methodology in an animal science study where the main objective is to assess the relationship between the thermal environment (e.g., temperature and humidity) and feeding behavior of lactating sows. The experimental study was carried during July to October in 2013 in a 2,600-sow commercial research unit in Oklahoma and involved 480 PIC Camborough sows; for details of the study we refer to Benavides et al. (2016). The sows were kept in the farrowing facility where they gave birth to piglets. Each sow was observed longitudinally for a 21-day lactation period spanning both the pre-and-post birth-giving phase. Sows were classified as young and old depending on the history of the number of pregnancies. On each day, minute-by-minute information about the relative humidity (%) and ambient air temperature ($^\circ$C) inside the facility were recorded by data loggers (LogTag, MicroDAQ Ltd., Contoocook, NH). In addition, the amount of feed-intake (kg) was measured and recorded by a computerized system for each sow. Figure 3.4 illustrates the observed data for a randomly chosen sow: daily feed-intake, minute-by-minute air temperature and
humidity measurements from left to right. The experimenters discarded information of five sows due to the erroneous measurements at data collection step.

Figure 3.4: Data for a randomly chosen sow: feed-intake (in kg) (left), temperature (in Celsius) (middle), and relative humidity (in %) (right) which are recorded at 5-minutes interval for 24 hours period starting from 2:00PM on each day inside the farrowing facility. Color (“red” and “blue”) correspondence is provided for convenience.

Since sows were exposed to the heating environment for an extended period of time; the scientists argue that a sows’ body might get adjusted to heat and may behave differently over the period. It is expected that any such behavioral change is reflected through the change in its feeding habit. Therefore it is critical to investigate the statistical significance of the effect of thermal environment on sows feeding behavior. The data were first analyzed by Park et al. (2016); Islam et al. (2017). The former approach primarily focuses on modeling the feed-intake using the temperature. The latter approach considers a quantile regression model to investigate the association between the temperature and feed-intake at various quantile levels. However neither of these approaches studied statistical inference of the association of the temperature and/or humidity with the feed-intake.

Let \( i \) index the sows, \( j \) index the repeated instances for each sow, \( t_{ij} \) denote the lactation day of the \( i \)th sow. Denote by \( g_i \) the group of the \( i \)th sow; \( g_i = 1, \ldots, 21 \). The groups are homogeneous in the sense that the sows within the same group gave birth at adjacent calender dates. Let \( FI_{ij} \) be the feed-intake (FI) by the \( i \)th sow at its \( t_{ij} \)th lactation day. Denote by \( nATemp_{ij}(\cdot) = nTemp_{i}(\cdot, t_{ij}) \) and \( nHum_{ij}(\cdot) = nHum_{i}(\cdot, t_{ij}) \) the observed air temperature and relative humidity profiles at \( t_{ij} \). We assume that the profiles are contaminated by the measurement errors. The prefix “\( n \)” is used to distinguish the noisy profiles from the truth. Each sow belongs to a parity group denoted by \( p_i \) such that \( p_i = 1 \) for old sows and \( p_i = 0 \) for
young sows.

Assuming FI follows Gaussian distribution, we investigate the strength of association between the feeding behavior and the thermal environment by considering the full model

\[ FI_{ij} = \beta_{pi}(t_{ij}) + \int S Temp_{ij}(s) \gamma_r(s, t_{ij}) ds + \int S Hum_{ij}(s) \gamma_n(s, t_{ij}) ds + b_{gi} + b_{0i(g_i)} + b_{1i(g_i)} t_{ij} + \varepsilon_{ij}. \] (3.19)

Here \( \beta_{pi}(\cdot) \) is the mean feed intake for parity \( p_i \), \( \gamma_r(s, \cdot) \) quantifies the time-varying effect of the air temperature on mean FI keeping the relative humidity constant, \( \gamma_n(s, \cdot) \) quantifies the time-varying effect of the relative humidity on mean FI with the air temperature remains constant. Here the integral reflects the aggregated effect over the 24 hours period. The term \( b_{gi} + b_{0i(g_i)} + b_{1i(g_i)} t_{ij} \) models the dependence of the responses for two major sources of variation: group specific and sow specific. The random term \( b_{gi} \) is a group effect, and \( b_{0i(g_i)} \) and \( b_{1i(g_i)} \) are the sow specific intercept and slope effects within the group \( g_i \). Let \( b_{i(g_i)} = (b_{0i(g_i)}, b_{1i(g_i)})^T \) and assume \( b_{i(g_i)} \sim \mathcal{N}(0, D) \); where \( D \) is a 2 \( \times \) 2 positive definite matrix. Let \( b_{gi} \sim \mathcal{N}(0, \sigma^2_g) \), and assume that \( b_{i(g_i)} \) and \( b_{gi} \) are mutually independent. Finally, it is assumed that the measurement errors \( \varepsilon_{ij} \)'s are independent and distributed as \( \mathcal{N}(0, \sigma^2_e) \). We define the population mean model by \( \mu_{ij} = \beta_{pi}(t_{ij}) + \int S Temp_{ij}(s) \gamma_r(s, t_{ij}) ds + \int S Hum_{ij}(s) \gamma_n(s, t_{ij}) ds \).

We are interested in four scientific questions: (1) Does the relative humidity affect the feeding behavior of the lactating sows in the presence of the temperature in the model? (2) Are the smooth effects of the air temperature and relative humidity indeed time-varying? (3) Do the effects of the air temperature and relative humidity on the feed-intake change linearly over the 21-day lactation period? (4) Are the functional effects constant over the 24-hour period? Of course this last problem makes sense to investigate only if we fail to reject the null hypothesis stated in question 2 and question 3. We write the research questions formally for a generic time \( t \) as below.

**Hypothesis 1.** The first question can be formulated as testing the following null hypothesis.

\[ H_0 : \gamma_r(s, t) \neq 0 \text{ and } \gamma_n(s, t) = 0, \text{ for all } s \text{ and } t, \]
\[ H_A : \gamma_r(s, t) \neq 0 \text{ and } \gamma_n(s, t) \neq 0, \text{ for some } s \text{ and } t. \] (3.20)

**Hypothesis 2.** The hypothesis corresponding to the second question is as follows. This test
is valid if we reject the null hypothesis (3.20).

\[
H_0 : \gamma_T(s, t) = \gamma_T(s) \text{ and } \gamma_u(s, t) = \gamma_u(s), \text{ for all } t,
\]

\[
H_A : \gamma_T(s, t) \neq \gamma_T(s) \text{ and } \gamma_u(s, t) \neq \gamma_u(s), \text{ for some } t.
\]

(3.21)

Here \( \gamma_T(\cdot) \) and \( \gamma_u(s) \) are the unknown smooth time-invariant effects of temperature and humidity, respectively.

**Hypothesis 3.** The third question can be formulated mathematically as follows; this test makes sense if we reject the null hypothesis (3.21).

\[
H_0 : \gamma_T(s, t) = \gamma_T(s) + t\gamma_T(s) \text{ and } \gamma_u(s, t) = \gamma_u(s) + t\gamma_u(s), \text{ for all } s,
\]

\[
H_A : \gamma_T(s, t) \neq \gamma_T(s) + t\gamma_T(s) \text{ and } \gamma_u(s, t) \neq \gamma_u(s) + t\gamma_u(s), \text{ for some } s.
\]

(3.22)

where \{\gamma_T(\cdot), \gamma_T(\cdot)\} and \{\gamma_u(\cdot), \gamma_u(\cdot)\} are the time-invariant smooth functions associated with the temperature and relative humidity, respectively.

**Hypothesis 4.** This test might be of interest if we reject the null hypothesis (3.21) but fail to reject the null hypothesis (3.22).

\[
H_0 : \gamma_T(s, t) = \gamma_T(s) + t\gamma_T(s) \text{ and } \gamma_u(s, t) = \gamma_u(s), \text{ for all } s \text{ and } t,
\]

\[
H_A : \gamma_T(s, t) \neq \gamma_T(s) + t\gamma_T(s) \text{ and } \gamma_u(s, t) \neq \gamma_u(s), \text{ for some } s \text{ and } t.
\]

(3.23)

**Hypothesis 5.** This test is valid if we fail to reject the null hypotheses (3.20) and (3.21).

\[
H_0 : \gamma_T(s) = \beta, \text{ for all } s,
\]

\[
H_A : \gamma_T(s) \neq \beta, \text{ for some } s.
\]

(3.24)

Here \( \text{Temp}_t^* = \int_S \text{Temp}_i(s, t)ds \) and \( \beta \) is some unknown parameter measuring the effect of mean temperature profiles on FI such as \( \gamma_T(s) = \beta \). Under alternative, the model is similar to the one described in Goldsmith et al. (2012a) for which inferential procedure is detailed in Swihart et al. (2014).

The steps for fitting the model (3.19) are similar to the ones described in Section 3.5.2. We estimate the mean functions for the observed and transformed functional predictors by the fast bivariate smoothing splines with 20 knots in \( t \)-direction and 35 knots \( s \)-direction for both temperature and humidity. REML is used to estimate the variance parameters in the mixed model (3.19). The basis functions to approximate the smooth coefficients \( \beta_k(\cdot) \)'s
are generated by \texttt{smspline} function with 19 knots. The estimated covariances $\hat{\Sigma}(s,s')$ and $\hat{\Sigma}_v(s,s')$, and the corresponding eigencomponents are obtained for both temperature and humidity. We carry each of the above tests for the two choices of scores $\hat{\xi}_W$ and $\hat{\xi}_V$ which are obtained using the basis functions of $\hat{\Sigma}(s,s')$ and $\hat{\Sigma}_v(s,s')$, respectively. Smoothing parameters are selected using REML.

### 3.6.1 Significance testing

Table 3.4 compares the $p$-values for two different choices of bases; while one approach selects the orthonormal bases from the marginal covariance functions of the functional predictors, the other focuses on the selection of eigenbasis using the response-adjusted functional covariates. Interestingly, we can draw different conclusions depending on the selection of basis functions. Specifically, we notice that the testing procedure associated with $\xi_W$ is more conservative than the one with $\xi_V$; this observation matches with the findings in Section 3.5.

Based on the $p$-value associated with $\xi_V$ for testing (3.20), we conjecture that there is significant association between the feed-intake and relative humidity; this justifies the inclusion of the relative humidity in the model (3.19). However the test with respect to $\xi_W$ fails to reject the null (3.20) and favors the model with $\textit{Temp}_{ij}(\cdot)$ only.

Since the $p$-value is very small for testing (3.21) with respect to $\xi_V$, we conclude that the model favors the time-varying smooth coefficients over the time-invariant ones for both the temperature and relative humidity. On the other hand, the test associated with $\xi_W$ fails to reject the null; $p$-value is marginally insignificant at 5% nominal level.

Based on the $p$-values for testing (3.22) and (3.23) with respect to $\xi_V$, we imply that the pattern of change of the functional effects $\gamma_T(\cdot,\cdot)$ and $\gamma_H(\cdot,\cdot)$ are non-linear over the 21-day lactation period; we also observe similar phenomenon in Figure 3.5.

Since the test with $\xi_W$ fails to reject the null hypothesis (3.20) and favors the model with a single functional predictor, we are interested to investigate whether the functional structure modeled for the temperature is indeed justified. Since the $p$-value associated with this test is very small, it implies that the temperature effect is not constant over $\mathcal{S}$ and varies over the 24-hour period.

Based on this analysis, we infer that the test with respect to $\xi_W$ favors the model with time-invariant smooth coefficients as in Goldsmith \textit{et al.} (2012a) and Gertheiss \textit{et al.} (2013a); in contrast, the test with respect to $\xi_V$ suggests using time-varying smooth coefficients as in Kundu \textit{et al.} (2016) and Islam \textit{et al.} (2017).
Table 3.4: $p$-values for testing the null hypotheses associated with the scientific questions. Model with $\xi_V$ corresponds to the eigenbasis obtained from the spectral decomposition of the covariance operator of the response-adjusted functional covariates. In contrast, model with $\xi_W$ uses the eigenbasis of the marginal covariance function of the functional predictor without accounting for the correlation between the responses and predictors.

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>Model $\xi_W$</th>
<th>Model $\xi_V$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hypothesis 1</td>
<td>0.086</td>
<td>0.000*</td>
</tr>
<tr>
<td>Hypothesis 2</td>
<td>0.055</td>
<td>0.000*</td>
</tr>
<tr>
<td>Hypothesis 3</td>
<td>0.095</td>
<td>0.000*</td>
</tr>
<tr>
<td>Hypothesis 4</td>
<td>0.344</td>
<td>0.000*</td>
</tr>
<tr>
<td>Hypothesis 5</td>
<td>0.000*</td>
<td>0.000*</td>
</tr>
</tbody>
</table>

Note: *Statistical significance at 5% level.

3.6.2 Estimation of regression surfaces

The time-invariant orthogonal bases of the functional coefficients are obtained from the spectral decomposition of the covariance operator of the transformed temperature and humidity profiles which take into account for the correlation between the feed-intake and respective functional measurements. Note that the functional coefficients are identified only up to a space spanned by the time-invariant bases (Islam et al., 2017); and therefore we primarily focus on interpreting the surface across lactation days.

The association between the temperature and the feed intake changes across 21-day lactation period. In the early days, say from day 1 to day 6, the sows primarily take their food around 10:00AM-7:00PM. During the middle of the lactation period, say from day 10 to day 17, the amount of feed-intake is positively associated with high temperature around 12:00AM-9:00AM implying no adverse effect of the temperature during these hours. In contrast, during the same hours of a day, the sows exhibit different feeding behavior specially in the early and later phase of the lactation period. Moreover, the effect of the temperature averaged over hours exhibits downward trend towards the end of the lactation period; see the red solid line at the top of the left heatmap. The effect of the humidity on the feed-intake behaves similar to the temperature during the middle of the lactation period at 12:00AM-9:00AM; this is not so surprising as sows seem to take food more at night. In addition, the effect of the humidity is negatively associated with the amount of feed-intake during 7:00PM-12:00AM
and 10:00AM-1:30PM throughout the lactation period except for the first few days. The mean humidity effect averaged over hours also has downward pattern during the end of the period; see the red solid line at the top of the right heatmap. In addition, both the temperature and humidity has opposite effect (averaged over lactation days) over the 24 hours period; follow the blue solid lines to the right of each heatmap.
Figure 3.5: Estimated regression coefficients associated the air temperature (left) $\hat{\gamma}_T(\cdot,\cdot)$ and relative humidity (right) $\hat{\gamma}_H(\cdot,\cdot)$ for each lactation day (X-axis) $t = 1, \ldots, 21$ over 24-hours period (Y-axis). The corresponding mean effects averaged over hours (red solid line at the top of the heat map) and lactation days (blue solid line at the right of the heat map) are plotted for the air temperature and relative humidity. Reported are the plots based on the model with $\xi_V$. 
3.7 Discussion

In this article, we developed a rigorous hypothesis testing procedure based on pLRT (Staicu et al., 2010) for making inference about the time-varying effect of the functional predictor in a longitudinal framework. We proposed a novel, data-driven basis functions selection strategy using an appropriate covariance function of the response-adjusted functional predictors. Here the main idea is to account for the correlation between the responses and functional predictors. The proposed inferential procedure relaxes the assumption that the leading eigencomponents are the most informative of the responses which is deeply rooted in the functional principal component regression (FPCR) techniques (Reiss & Ogden, 2007; Febrero-Bande et al., 2017; Di et al., 2009). Our method relies on the assumption that the latent predictor is smooth and observed with small noise. The method can easily accommodate testing the effect of single/multiple scalar covariates.

Numerical results show that the power performance of the proposed inferential method is far superior to an existing alternative approach where the eigenbasis functions are selected directly from the marginal covariance function of a functional predictor; however both approaches perform satisfactorily in attaining the desired size of the test. The method is computationally efficient and can be easily implemented using the existing software in R. We applied the method to an animal science application where we found that the effect of the temperature and relative humidity on the feed-intake of sows is statistically significant.

Apparently there are several directions for future work. The approach is illustrated and developed assuming only Gaussian responses; however the testing procedure for the generalized longitudinal responses in the context of functional data is important and we leave it for our future study. We also acknowledge the fact that the regression coefficient, like other FPCR techniques, is only identifiable up to the space spanned by the functional predictor which limits the interpretability of the regression surface; this shortcoming also requires extensive future research (Islam et al., 2017).
Chapter 4

Functional Variable Selection for EMG-based Control of a Robotic Hand Prosthetic

4.1 Introduction

More than 160,000 Americans are transradial (i.e. below-elbow) amputees, henceforth TRAs, and must relearn how to perform tasks that typically require an intact hand (Ziegler-Graham et al., 2008). Passive hand prostheses and related devices are useful, but cannot emulate the full functionality of an intact limb. Multifunctional robotic prosthetics, such as the FDA-approved DEKA arm system (Linda Resnik et al., 2011; Resnik et al., 2011), have become very popular with recent advancements in their mechanical systems. The available software that provides user control of the hardware, however, is often nonintuitive to operate. For example, one approach is to have the user control the prosthetic with their foot. Forearm muscle contractions are known to cause hand movement for an able-bodied subject, henceforth AB subject. A TRA’s residual forearm, which once caused hand movement, will still contract and these contractions exhibit measurable EMG signals that are somewhat consistent with hand movement even though they no longer use that hand. A reliable mapping of a TRA’s forearm electromyogram (EMG) signals to hand movement would then provide a more intuitive approach for prosthesis control.

Figure (4.1) shows the process of movement intention to movement production for both an AB subject and TRA. For an AB subject, tendons are attached to the forearm muscles,
travel through the wrist and connect to bones in the hand. The muscle contractions transmit force through tendons to the bones, generating finger or wrist movement. For a TRA, this physical connection among muscles, tendons, and bones no longer exists, but they may still sense movement in their phantom limb, accompanied by observable muscle contractions in their residual limb. Their generated EMG signal data feed into a prosthesis controller that processes the data and predicts movement, which is then produced by the robotic limb.

![Diagram](image)

Figure 4.1: Visualization of biomechanical system for hand movement for AB subject and TRA. (A) Internal limb representation in the motor cortex. (B) Neural signal sends motor commands to forearm muscles. (C) Forearm muscles contract to perform desired movement, which, for an AB subject, results in direct hand movement through tendon connections (red lines) to the hand. (D) For a TRA, a prosthesis controller projects real time hand movements using forearm EMG information, and (E) the robotic limb performs the intended movement.

Direct myoelectric control, a traditional approach for prosthesis control, uses EMG signals from two antagonistic muscles and assigns to each a specific type of hand movement, such as finger flexion or extension. The EMG signal can only control that class of movement and, if activated for any reason, will produce that movement with speed proportional to the magnitude of the signal. Direct control is limited in its ability to direct multiple degrees of freedom of hand movement. The user must switch manually between functions (e.g. from wrist pronation/supination to hand open/close), but then the user’s attempted movement may not match with the prosthesis action. Since it is cumbersome for controlling multiple limb functions, direct control prostheses have shown a 75% rejection rate by users (Biddiss & Chau, 2007; Peerdeman et al., 2011). State-of-the-art prosthetics overcome the restrictions of direct myoelectric control using data-driven pattern recognition (PR) algorithms capable
of synthesizing a large number of EMG signals (potentially over 100) into classes of intended movement (Englehart & Hudgins, 2003; Zhou et al., 2007; Huang et al., 2008).

PR is more flexible than direct myoelectric control, but has its own set of limitations that prohibit effective, intuitive user control. PR does not incorporate knowledge of human neurophysiology and biomechanics and relies heavily on the representativeness of the data used to train the algorithm in identifying EMG patterns associated with classes of hand movements. This leads to poor prediction performance for conditions not directly observed in the training data. It assumes that a constant, repeatable pattern of muscle contractions leads to a specific movement class and users may find it difficult to perfectly repeat this pattern. PR also significantly reduces the EMG information, using summarizing features of the EMG signals such as the mean absolute value and number of slope sign changes instead of the original signals. The loss of important information due to data reduction as well as redundancy of predictor information leads to overfitting to the training data, as indicated by Scheme et al. (2010).

Recently, Crouch & Huang (2016) and Crouch & Huang (2017) proposed and implemented an EMG-based controller using a planar link-segment dynamic model that directly incorporates features of the neurophysiological and biomechanical system. Using EMG data from only four forearm muscles, they were able to accurately predict wrist and finger movement. The model they use, however, is nonlinear and it is not clear how well it will translate to a TRA, since their model requires the subset of muscles to be pre-specified. Their results do show that incorporating the neurophysiological and biomechanical properties can yield movement predictions that are more typical of biological movement than movements predicted by PR. This suggests that incorporating such properties in a data-driven approach may help improve movement predictions.

In this paper, we propose to model two degrees of freedom for hand movement, finger and wrist flexion/extension, using recent past behavior of the EMG signals. Our statistical model uses either finger or wrist velocity as its response and allows the effect of each EMG signal on the response to vary with the current finger or wrist position. This model retains much of the EMG information and respects known biomechanical constraints, thus allowing it to explain a broader range of hand movements using fewer EMG signals. We fit the model using a novel algorithm for functional regression which we call Sequential Adaptive Functional Empirical group LASSO (SAFE-gLASSO) that selects only a subset of the EMG to be included in the model while simultaneously requiring that the estimated functional coefficients are smooth
and interpretable. SAFE-gLASSO pairs a generalization of the adaptive sparse-smoothness penalty from Gertheiss et al. (2013b) with an efficient, sequential variable selection and fitting approach inspired by the relaxed LASSO (Meinshausen, 2007). We also extend the ideas of Lei et al. (2017) to generate post-selection prediction confidence bands.

The paper is organized as follows. Section 4.2 reviews the neurophysical and biomechanical relationships involved in hand movement that we address in our proposed statistical model. Section 4.3 describes the collection process for collecting EMG and hand movement data. Section 4.4 details the analysis methodology starting with the proposed functional linear model and explains how it accounts for known underlying biomechanical relationships. We then describe the variable selection and post-selection fit procedures. Section 4.5 presents the analysis results from data collected from an AB subject, where the underlying truth is known. An extensive simulation study is performed in sections 4.6.1 and 4.6.2, showing the robustness of our method to varying assumptions. Section 4.7 concludes the paper with a discussion of limitations and possible avenues of extension.

4.2 Biomechanics of hand movement

This section describes the underlying biomechanical process that generates intentional hand movement for an AB subject and highlights the unique challenges that will need to be addressed by the proposed model. Incorporating features of this process in the statistical analysis and the robotic prosthetic software should significantly aid user control of robotic prosthetics by TRAs. In particular, this section motivates the use of recent past behavior of the EMG signals as functional predictors and justifies the need for position-dependent effects to map the relationship between these EMG signals and velocity. Velocity was chosen as the desired response instead of position because almost all the upper limb prostheses on the market have direction and velocity control as their outputs (Scheme & Englehart, 2011).

The biological process of hand movement begins with initiation of action potentials that are first conducted along motor neurons from the central nervous system. At the neuromuscular junction, the action potential causes the release of neurotransmitters from the nerve that initiates an action potential on the muscle fiber membrane. That motor unit action potential (MUAP) conducts along the surface of the muscle fiber membrane that extends over and within the muscle fiber. This stimulates the release of calcium ions within myofibrils and initiates cross-bridge cycling of overlapping protein chains that cause the entire muscle
fiber to shorten (contract), generating mechanical energy. The output force of an entire muscle varies depending on the proportion of fibers in the muscle that are contracted and the frequency at which they are stimulated. The EMG signal measured from surface electrodes represents the sum of individual MUAPs and thus conveys information about the magnitude and duration of muscle contraction.

The mechanical aspect of hand movement explains how the mechanical energy generated by the muscle contractions lead to specific hand movements. For AB subjects, tendons connect forearm muscles to bones in the hand and it is through this connection that movements are generated. That is, a muscle contraction will pull its tendon which in turn moves the hand depending on the path of the tendon relative to the joint(s) that it crosses. As mentioned earlier, AB subjects have an internal representation that maps muscle contractions to hand movements, and this mapping is maintained for TRAs following surgery, although it may become distorted. Our goal is to build and implement a statistical model that decodes this internal biomechanical representation, specifically for relating finger and wrist movement to EMG signal data.

The top panel of Figure 4.2 demonstrates what is meant by finger and wrist movement. For example, finger flexion and extension refers to the simultaneous opening and closing of all finger digits, respectively, except for the thumb. The bottom panel of Figure 4.2 shows a range of arm postures considered in this paper. A PR-based controller trained on one posture setting would likely perform poorly in one of the other postures because EMG patterns change with limb posture due to, for example, a shift in electrode locations relative to the underlying muscles (Scheme et al., 2010). A biomechanically motivated model should behave consistently across postures because the biomechanical process does not change significantly with posture.

The known biomechanical process of AB subjects suggests that these two movement degrees of freedom can be explained by relatively few forearm muscles. The nonlinear model of Crouch & Huang (2016) is capable of accurate predictions with few EMG signals but does not clearly admit a variable selection procedure to decide which muscles to use. This presents a problem for its use by a TRA for whom the intended biomechanical actions of the EMG are not directly observable and may be altered following amputation. Thus we require a statistical model that is capable of accurate predictions using few EMG signals and the model must lend itself to a variable selection procedure.

There are many challenges and constraints that the proposed statistical model must
Figure 4.2: Top panel corresponds to finger and wrist movement: finger extension, finger flexion, wrist extension, and wrist flexion (from left to right). Bottom panel illustrates the arm postures at which data is collected during the experiment.

accommodate to successfully approximate this biomechanical process. First, the fingers and the wrist are limited in how far they can be extended/flexed. Second, muscles generate passive movement forces when stretched, which may produce movement in the absence of EMG information. For instance, if one relaxes their forearm muscles following a contraction, the tendons will return back to their resting length, generating a passive force, and the hand will return to a neutral configuration. Therefore, we may observe movement without observing concurrent EMG activation. While this tendon connection no longer exists for a TRA, they may still anticipate these passive forces.

Figure 4.3 illustrates some finger position data (in radians, shown in black) taken from an AB subject for a short time-window and overlays two concurrent EMG signals, labeled EMG 7 (green) and EMG 12 (red), known to contribute to finger movement. Figure 4.3(a) is an event window with an active EMG 7 signal but no movement; the result of a physical constraint. Figure 4.3(b) demonstrates finger movement in the absence of active, concurrent EMG signals, due to passive forces. This figure demonstrates the opportunity for past EMG behavior to predict passive movement, as seen by the declining green line preceding Figure 4.3(b).

### 4.3 Data collection and processing

This section and the analysis in section 4.5 focuses on the collection and analysis of data from an AB subject. The process of calibrating a robotic arm is subject-specific and we are not
interested in estimating a population model. The data collection procedure described here and the analysis methodology described in Section 4.4 could be modified to perform inference across multiple subjects. A data collection strategy is given at the end of Section 4.3.1 for TRAs consistent with our analysis approach, requiring measured hand movements.

Surface EMG electrodes (Biometrics, Newport, UK) were placed at different locations of the forearm. The subject was then asked to perform specific types of hand movement and the EMG signals were recorded continuously at 960 Hz. Electrodes were placed over 16 specific forearm muscles considered to be potentially important muscles for generating hand movements of interest. Raw EMG data were high-pass filtered at 40 Hz, rectified, and low-pass filtered at 6 Hz using a 4th order Butterworth zero-phase filter (Crouch & Huang, 2016). Prior to data collection, the subject was asked to perform movements exhibiting maximal muscle contraction, allowing normalization of the subject’s EMG signals to be between 0 (no contraction) and 1 (maximal contraction).

Movement data were collected by reflective markers placed on 9 anatomical locations on the forearm, wrist, and hand. Three-dimensional marker positions were recorded at 120 Hz using an infrared motion capture system (Vicon Motion Systems Ltd., UK), and processed (filtered) at 6 Hz using a 4th order Butterworth filter (Butterworth, 1930). The joint angles
(in radians), which we call positions, were then calculated from filtered marker data through a musculoskeletal model (Holzbaur et al., 2005) in OpenSim (Delp et al., 2007). EMG and joint angle data were collected synchronously for a given period of time during which a subject was asked to perform a set of finger and/or wrist movement. Figure 4.3 provides a snapshot of this data while a subject is performing finger flexion and extension with consistent movements in a fixed posture.

4.3.1 Data collection procedure

The subject was asked to perform basic finger and wrist movements in each arm posture as shown in the bottom panel of Figure 4.2. For a given posture, the subject was asked to perform single degree-of-freedom movements for their fingers/wrist following either a consistent or random pattern. For example, the subject performed consistent finger movement that continuously alternated between extending (opening) and flexing (closing) their fingers, while keeping their wrist in a neutral position. Such a protocol has been followed by others; see for example Kawashima & Mita (2016). The subject was also asked to perform a series of random movements, which was chosen solely by the subject, providing a challenging benchmark to determine how robust our fitted models are to different movements.

For a TRA, there would be no movement data for the corresponding hand. Instead, movement data can be collected from the other, intact limb while the TRA performs (attempts) mirrored movements with both limbs; see Scheme & Englehart (2011).

4.3.2 Post-collection data processing for analysis

To accommodate the biomechanical constraints described in section 4.2, we utilize recent past behavior of EMG signals, as well as current finger/wrist position, to predict finger/wrist velocity. For example, a concurrent EMG value of 0.50 leads to different movements depending on the historical EMG trend. We briefly describe here the data processing that was done that allowed us to fit the model described in the next section.

The velocity values were estimated from a penalized smoother of the entire series of recorded position data across time using the R package fda. In particular, we used a second-order smooth regularization penalty to control the goodness of fit and smoothness of the fitted curve, where the smoothing parameter is selected by cross-validation. We then generated a velocity estimate for each observed position data point. For each velocity data point, we
extracted the previous EMG observations that ended with the concurrent EMG value to the velocity value. In this application, we chose a past time window of roughly 1/3 seconds. The value was chosen based on observed passive force movement (see Figure 4.3). This was done for all EMG signals, so each velocity estimate had associated with it a position value and equally-spaced past measurements across 1/3 seconds for all 16 EMG signals collected. A visualization of this data restructuring may be found in the Supplementary Materials, Section B. We propose to model the current velocity of the finger/wrist movement as a function of the recent history of the EMG signals which is to be discussed next.

### 4.4 Variable selection and inferential framework

Denote the observed data by \( [y_i, z_i, \{X_{k,i}(s_r); r = 1, \ldots, R \}, k = 1, \ldots, K, i = 1, \ldots, N] \) where \( i \) indexes the instance at which data are collected, \( y_i \) is the scalar response, \( z_i \) is the continuous scalar covariate, \( k \) indexes the functional predictors, and \( X_{k,i}(s_r) \) is the \( k \)th functional predictor observed at point \( s_r \) such that \( s_r \in S \). It is assumed that \( z_i \in Z \), and both \( S \) and \( Z \) are closed compact sets. In our application, \( y_i \) is the current velocity, \( z_i \) is the current position, and \( X_{k,i}(\cdot) \) is the recent history of the \( k \)th EMG signal and \( S \) is a window of time that depicts the “recent” past.

We consider a functional linear model with varying smooth effects:

\[
E[y_i|X_{1,i}, \ldots, X_{K,i}] = \alpha + \sum_{k=1}^{K} \int_{S} X_{k,i}(s) \gamma_k(s, z_i) ds, \tag{4.1}
\]

where \( \alpha \) is an intercept and \( \gamma_k(\cdot, \cdot) \) is an unknown bivariate function defined on \( S \times Z \) that quantifies the effect of the \( k \)th functional predictor on the mean response of \( y_i \) conditional on \( z_i \). Model (4.1) is a direct extension of the functional linear model (FLM) described in Ramsay & Silverman (2005); Cardot et al. (2003a); James et al. (2009); Goldsmith et al. (2011); Ferraty et al. (2012a); McLean et al. (2014) who assume nonvarying functional coefficients.

In our application accounting for positions in the statistical model is crucial for modeling of passive forces since the effect of the EMG signals on finger/wrist movement heavily depends on the current state of the hand as argued in section 4.2. Varying coefficients for scalar covariates characterize the effect of a covariate on response through another covariate and have been studied extensively in nonparametric and semiparametric literature; see, for example, Fan & Zhang (2008); Maity & Huang (2012); Fan et al. (2014); Davenport et al. (2015);
Bandyopadhyay & Maity (2017). Recently, the varying-coefficient model has been extended to analyze functional data in Cardot & Sarda (2008); Wu et al. (2010); Davenport (2013).

Our primary objective is to select the functional predictors whose effects on the mean response as described by (4.1) is non-zero. For our application this would allow us to select the most important forearm muscles in explaining finger and wrist movements. Let \( \mathcal{K} \subseteq \{1, \ldots, K\} \) denote the true index set where \( \gamma_k(\cdot, \cdot) \neq 0 \), and let \( \mathcal{B} = \{ \gamma_k(\cdot, \cdot); k \in \mathcal{K}\} \) denote the set of true effects. Established fitting approaches focus on the estimation of the smooth coefficients rather than variable selection. Variable selection in scalar-on-function regression with invariant smooth coefficients, \( \gamma_k(\cdot) \), has been discussed in Matsui & Konishi (2011); Gertheiss et al. (2013b); Fan et al. (2015); Pannu & Billor (2017). In this paper, we extend these approaches to the varying functional linear model (4.1) that is motivated by our data application.

### 4.4.1 Approximation to linear model

Following the procedure described in Wood (2006a) and Eilers & Marx (2003), we approximate \( \gamma_k(\cdot, \cdot) \) using a tensor product of two univariate basis functions. Let \( \{ \omega_l(s) \}_{l=1}^L \) and \( \{ \tau_m(z) \}_{m=1}^M \) be two truncated univariate bases defined in \( S \) and \( Z \), respectively; and let \( \gamma_k(s, z) \approx \sum_{l=1}^L \sum_{m=1}^M \omega_l(s) \tau_m(z) \beta_{klm} \), where \( \beta_{klm} \)'s are the unknown basis coefficients associated with the \( k \)th functional predictor. For simplicity of exposition, we use the same bases for all \( \gamma_k(\cdot, \cdot); \ k = 1, \ldots, K \). Generalization to different bases for different predictors is straightforward.

It is more convenient to rewrite the tensor product in matrix form \( \gamma_k(s, z) \approx \omega^T(s) \mathbf{B}_k \tau(z) \) where \( \omega(s) = (\omega_1(s), \ldots, \omega_L(s))^T \), \( \tau(z) = (\tau_1(z), \ldots, \tau_M(z))^T \), and \( \mathbf{B}_k = (\beta_{klm})_{l,m} \) is the \( L \times M \) matrix of basis coefficients, which is our target for inference. The generic summand in (4.1) can be approximated by Riemann sum approximation as

\[
\int_S X_{k,i}(s) \gamma_k(s, z_i) ds \approx \sum_{l=1}^L \left( \sum_{r=1}^R \Delta_r X_{k,i}(s_r) \omega_l(s_r) \right) \mathbf{B}_{lk} \tau(z_i) = X_{ki}^T \mathbf{B}_k \tau(z_i),
\]  

where \( X_{ki}^T = \left( \sum_{r=1}^R \Delta_r X_{k,i}(s_r) \omega_l(s_r), \ldots, \sum_{r=1}^R \Delta_r X_{k,i}(s_r) \omega_L(s_r) \right) \) is a \( 1 \times L \) row vector and \( \Delta_r = s_r - s_{r-1}, \). By an abuse of notation, define the \( 1 \times LM \) vector \( X_{ki}^T \mathbf{B}_k \tau(z_i) \), where \( \otimes \) indicates the Kronecker product. For notational simplicity, we omit indices \( \omega \) and \( \tau \) in the subscript of \( X_{ki}^T \mathbf{B}_k \tau(z_i) \) and denote it by \( \mathbf{B}_k \).

Vectorize \( \mathbf{B}_k \) by appropriately stacking its \( M \) columns vertically and denote the \( LM \times 1 \) vector of basis coefficients \( \beta_{klm} \)'s by \( \beta_k \). This
yields an approximating linear model to (4.1) with unknown regression coefficients $\beta_k$'s

$$E[y_i|\bar{X}_{1i}, \ldots, \bar{X}_{Ki}] \approx \alpha + \sum_{k=1}^{K} \bar{X}_{ki}^T \beta_k. \quad (4.3)$$

### 4.4.2 Penalized criterion for variable selection

We allow $L$ and $M$ to be sufficiently large to capture the complexity of the regression surfaces and penalize the degree of smoothness. We adopt the penalized least squares approach for estimation that simultaneously induces group sparsity and controls smoothness of the corresponding regression surfaces.

Let $\|\gamma_k\|^2 = \int_S \int_Z \{\gamma_k(s, z)\}^2 dz ds$ be the $L^2$ norm of $\gamma_k$, and let $\gamma''_k$ and $\gamma'''_k$ be the second partial derivatives of $\gamma_k$ with respect to $s$ and $z$, respectively. Prior to analysis, we center the response $y_i$ and functional predictors $X_{k,i}$'s, and remove $\alpha$ from the model. By an abuse of notation, we use the same notation for the centered model. The estimates of the $\gamma_k$'s are then selected to be the minimizers of the following penalized criterion

$$\sum_{i=1}^{N} \left\{y_i - \sum_{k=1}^{K} \int_S X_{k,i}(s) \gamma_k(s, z_i) ds\right\}^2 + \lambda \sum_{k=1}^{K} \left(\|\gamma_k\|^2 + \phi_1\|\gamma''_k\|^2 + \phi_2\|\gamma'''_k\|^2\right)^{1/2}; \quad (4.4)$$

where $\lambda > 0$ controls the model sparsity, and $\phi_1, \phi_2 > 0$ control the smoothness of $\gamma_k(\cdot, \cdot)$ in the s and z dimension, respectively. Let $\phi = \{\phi_1, \phi_2\}$ and $P_\phi(\gamma_k) = (\|\gamma_k\|^2 + \phi_1\|\gamma''_k\|^2 + \phi_2\|\gamma'''_k\|^2)^{1/2}$. The proposed penalty function $P_\phi(\gamma_k)$ was first introduced by Meier et al. (2009) for variable selection in high dimensional additive model. As $\lambda$ increases, more penalty weight is placed on both the magnitude and smoothness of the surface which results in $\|\gamma_k\|^2 = 0$, excluding the corresponding functional predictor from the model. Although $\gamma_k(\cdot, \cdot)$ is assumed to be a smooth function, the fitted regression surfaces may be rough if the penalty on the second order derivatives $\|\gamma''_k\|^2$ and $\|\gamma'''_k\|^2$ are not included. The extent of smoothing of the non-zero regression surfaces is controlled by the vector of smoothing parameters $\phi$. For example, as $\phi$ increases, the model more heavily penalizes the departure from linearity, producing linear effects in both directions. The proposed penalty $P_\phi(\gamma_k)$ ensures smoothness and sparseness in two dimensions; it can be easily extended for multiple dimensions without loss of generality.

We now turn our attention to efficient calculation of $P_\phi(\gamma_k)$ for a given function $\gamma_k$. For simplicity, we further assume that the bases $\{\omega_l(\cdot)\}_{l=1}^{L}$ and $\{\tau_m(\cdot)\}_{m=1}^{M}$ are orthogonal B-splines. Using the orthogonal property of the bases, define by $\Omega_\lambda = I_L$ the $L \times L$ identity
matrix with the \((l, l')\)th element as \(\int_{S} \omega_l(s) \omega_{l'}(s) ds = 1\) if \(l = l'\) and 0 otherwise, and define \(\Omega_z = I_M\) similarly in terms of the \(\tau_m(z)\)'s. General bases will not have \(\Omega_s\) and \(\Omega_z\) equal to the identity. Let \(\omega''(\cdot)\) and \(\tau''(\cdot)\) be the second derivatives of \(\omega_l(\cdot)\) and \(\tau_m(\cdot)\), respectively; let \(\omega''(s)\) and \(\tau''(z)\) be the vectors of these \(L\) and \(M\) functions. Define \(\Omega_{ss} = \int \omega''(s) \omega''(s)^T ds\) and \(\Omega_{zz} = \int \tau''(z) \tau''(z)^T dz\).

It follows that

\[
||\gamma_k||^2 = \sum_{l,l'} \sum_{m,m'} \left\{ \int_{S} \omega_l(s) \omega_{l'}(s) ds \right\} \left\{ \int_{S} \tau_m(z) \tau_{m'}(z) dz \right\} \beta_{klm}^2 = \beta_k^T \beta_k,
\]

where the last equality follows from the orthogonality of the bases. Similarly, we obtain

\[
||\gamma''_{k,s}||^2 = \beta_k^T (\Omega_{ss} \otimes I_M) \beta_k \quad \text{and} \quad ||\gamma''_{k,z}||^2 = \beta_k^T (I_L \otimes \Omega_{zz}) \beta_k.
\]

The penalty \(P_\phi(\gamma_k)\) may now be expressed as a sparse-smooth regularization (Meier et al., 2008; Meier, 2009)

\[
P_\phi(\gamma_k) = (\beta_k^T Q_\phi \beta_k)^{1/2},
\]

where \(Q_\phi = I_{LM} + \phi_1(\Omega_{ss} \otimes I_M) + \phi_2(I_L \otimes \Omega_{zz})\) is a \(LM \times LM\) symmetric positive-definite matrix. Cholesky decomposition on \(Q_\phi\) gives \(Q_\phi = R_\phi R_\phi^T\), where \(R_\phi\) is a lower triangular non-singular matrix. Using (4.3) and \(R_\phi\), reparametrize the model coefficients \(\beta_k\) as \(\tilde{\beta}_k = R_\phi^{-1} \beta_k\) and transform each \(\tilde{X}_{ki}\) to \(W_{ki} = R_\phi^{-1} \tilde{X}_{ki}\). It follows that (4.4) can be written equivalently as

\[
\sum_{i=1}^{N} (y_i - \sum_{k=1}^{K} W_{ki}^T \tilde{\beta}_k)^2 + \lambda \sum_{k=1}^{K} ||\tilde{\beta}_k||, \tag{4.5}
\]

which is similar to the group LASSO penalized criterion described in Yuan & Lin (2006); Yang & Zou (2013, 2015).

We use the groupwise-majorization-descent (GMD) algorithm described by Yang & Zou (2013) to solve (4.5), which is implemented by the \texttt{R} package \texttt{gglasso} (Yang & Zou, 2013). A benefit of the GMD algorithm is that it does not require group-wise orthonormality and is applicable for a general design matrix. In addition, for given values of the tuning parameters \(\lambda\) and \(\phi\), the minimizer of (4.5) with respect to \(\tilde{\beta}_k\) has a close-form solution; denote the solution by \(\tilde{\beta}_k\) by omitting its dependence on \(\lambda\) and \(\phi\).

It is straightforward to restructure the estimated basis coefficients in \(\tilde{\beta}_k\) to give \(\tilde{B}_k\), and we have

\[
\tilde{\gamma}_k(s, z) \approx \omega(s)^T \tilde{B}_k \tau(z).
\]

Let \(K_{\lambda,\phi}\) be the estimated index set after solving (4.5) under \(\lambda\) and \(\phi\); and denote the set
of their corresponding estimates by $\mathcal{B}_{\lambda, \phi} = \{\tilde{\gamma}_k(\cdot, \cdot); k \in \mathcal{K}_{\lambda, \phi}\}$. In reality, the sparseness and smoothing parameters, $\lambda$ and $\phi$ are unknown and need to be chosen empirically; for instance by cross-validation. We discuss this aspect in more detail in Section 4.4.4.

The penalized criterion (4.4) does not allow for different shrinkage and smoothness for different functional predictors, and assumes equal weight. This may inflate the number of false positives in variable selection and to mitigate this problem, adaptive estimation is recommended for variable selection in high-dimensional data (Meier, 2009; Tutz & Gertheiss, 2010; Gertheiss et al., 2013b). We too adopt this approach and discuss it next.

### 4.4.3 Adaptive penalized criterion

We generalize the criterion (4.4) by adaptive penalized criterion. Specifically, we use initial weights $w = \{f_k, g_k, h_k\}_{k=1}^K$ to introduce prior information on the relative importance and smoothness of functional predictors, where $f_k$ is related to the sparse penalty factor and $g_k$ and $h_k$ are related to the smooth penalty factors in the $s$ and $z$ dimension, respectively.

Define the adaptive penalty function by

$$P_{\phi}(\gamma_k) = \left[ \beta_k^T \{ f_k I_{LM} + g_k \phi_1(\Omega_{ss} \otimes I_M) + h_k \phi_2(I_L \otimes \Omega_{zz}) \} \beta_k \right]^{1/2} \quad (4.6)$$

where the weights $w$ are strictly positive and calculated based on the parameter estimates $\{\tilde{\gamma}_{k0}(\cdot, \cdot)\}_{k=1}^K$ associated with a functional additive model (McLean et al., 2014; Scheipl et al., 2015) fit without a sparseness penalty. The subscript 0 is used to distinguish the initial estimates from the model estimates of (4.5). Define the weights by $f_k = \frac{1}{\|\tilde{\gamma}_{k0}\|_d}$, $g_k = \frac{1}{\|\tilde{\gamma}_{k,s0}\|_d}$, and $h_k = \frac{1}{\|\tilde{\gamma}_{k,z0}\|_d}$, $d > 0$; for details we refer to Zou (2006); Gertheiss et al. (2013b); Guo et al. (2015); Ciuperca (2016); Ivanoff et al. (2016). Alternatively, one can adopt a semi-adaptive approach using only $f_k$ while keeping equal weights for smooth penalty factors such as $g_k = 1$ and $h_k = 1$ for all $k$. Like Yang & Zou (2015), the weights based on the number of parameters within each group can also be adopted alternatively. Here, as $\|\tilde{\gamma}_{k0}\|$ decreases to 0, $f_k$ increases to $\infty$ which yields sparser solution for $\gamma_k$ for a given $\lambda$. Similarly, as $g_k$ and $h_k$ increase to $\infty$, the regression surface yields linear pattern in both directions for a given $\phi$. We denote the minimizer of the criteria with penalty (4.6) by $\mathcal{B}_{\lambda, \phi}$ associated with $\mathcal{K}_{\lambda, \phi}$ for a given $\lambda$ and $\phi$. 

87
4.4.4 Selection of the tuning parameters, $\lambda$ and $\phi$

A widely used method to select tuning parameters is $K$-fold cross-validation (CV). We use 5-fold block CV which has been found appealing for data with temporal correlations (see Roberts et al. (2017)). Specifically, we partition the data into 5 equally-sized sections; let $n = 1, \ldots, 5$ index the folds and denote by $y_{i_{n,m}}$ be the $m$th ordered response of the $n$th fold corresponding to the instance $i_{n,m}$, and $N_n$ be the total number of observations in the $n$th subsample. We aim for the $N_n$ to be as equal as possible across $n$. The test set is formed by one of the five folds; the remaining four folds form the training set. For fixed values of $\lambda$, $\phi$, and $n$, the model parameters are estimated based on the training set (e.g., all data with the $n$th fold removed), and the prediction accuracy is evaluated based on the performance of the test set (e.g., the $n$th fold); i.e., $PE_{\text{out},n} = \sqrt{\sum_{m=1}^{N_n} (y_{i_{n,m}} - \hat{y}_{i_{n,m}})^2 / N_n}$, where $\hat{y}_{i_{n,m}}$ is the predicted value for $y_{i_{n,m}}$.

For each value of the tuning parameters, we compute the average $PE_{\text{out},n}$'s and select the values corresponding to the minimum prediction error. Furthermore, CV with the one standard error rule can be also adopted; the main idea is to select the simplest model whose numerical performance is comparable with the optimal model that is chosen by minimum prediction error (Friedman et al., 2001; Krstajic et al., 2014; Yang & Zou, 2015).

4.4.5 Sequential Adaptive Functional Empirical (SAFE) Selection

Most variable selection algorithms cannot guarantee the exclusion of noise variables (Meinshausen et al., 2009). In particular, the group LASSO approach has been shown to select a larger number of groups than necessary (Meier et al., 2008; Gertheiss et al., 2013b); this may happen due to the uncertainty in selecting the optimum combination of tuning parameters. As a remedy, we propose to perform the selection step twice where the second stage considers only the covariates $k \in K_{\lambda^1,\phi^1}^{(1)}$; where the superscript (1) denotes the first stage, and $\lambda^1$ and $\phi^1$ are the optimal tuning parameters determined by CV in the first stage. The adaptive weights for the second stage are then calculated based on $B_{\lambda^1,\phi^1}^{(1)}$.

The proposed idea is motivated by the relaxed LASSO (Meinshausen, 2007) approach, however, the fitting procedure is different in terms of calculating the adaptive weights for the second stage. In contrast to Meinshausen (2007), our solution path of the second stage is driven by both $B_{\lambda^1,\phi^1}^{(1)}$ and the selection of second-stage tuning parameters $\lambda$ and $\phi$. In Meinshausen (2007), the estimates of the second stage are not driven by weights but rather
based on the exhaustive search of the sparse tuning parameter. This idea was also promoted by Wei & Huang (2010) and Guo et al. (2015) who adopted a two-stage variable selection strategy for the scalar-on-scalar regression.

With weights re-calculated, we minimize the objective function (4.6) using only those $\mathcal{K}^{(1)}_{\lambda^*,\phi}$ with respect to $\lambda$ and $\phi$. Let $\mathcal{K}^{(2)}_{\lambda^*,\phi^*}$ be the estimated index set resulting from the second stage of the regularization fit under the optimal combination of tuning parameters $\lambda^*$ and $\phi^*$ found by CV.

### 4.4.6 Post-selection inference and prediction

Shrinkage penalties cause the estimates of the non-zero coefficients to be biased towards zero (Zhao et al., 2017; Friedman et al., 2001). In the similar spirit to Leeb et al. (2015); Zhao et al. (2017), once $\mathcal{K}^{(2)}_{\lambda^*,\phi^*}$ is determined, we refit the model using the second order smooth regularization to reduce the prediction bias. Specifically, we solve the following penalized criterion

$$
\sum_{i=1}^{N} \left\{ y_i - \sum_{k \in \mathcal{K}^{(2)}_{\lambda^*,\phi^*}} \int_{S} X_{k,i}(s) \gamma_k(s,z_i) ds \right\}^2 + \sum_{k \in \mathcal{K}^{(2)}_{\lambda^*,\phi^*}} P_{\phi}(\gamma_k); \tag{4.7}
$$

where $\phi = \{\phi_1, \phi_2\}$ are the smoothing parameters that control the curvature of the fit and $P_{\phi}(\gamma_k) = \phi_1||\gamma''_{k,s}||^2 + \phi_2||\gamma''_{k,z}||^2$ is the smoothing penalty function as described in Wood (2006a) and Eilers & Marx (2003). We choose the optimal tuning parameters using CV as discussed in Section 4.4.4.

The model (4.7) is conditional on the event that the same data is being used twice: once in the variable selection stage and again in the post-selection fit on the selected subset. Wu et al. (2009); Zhao et al. (2017) argued that the use of standard prediction (confidence) intervals or $p$-values without adjustment are invalid as they neglect the complex selection procedure to define the reduced model in the first place. Thus an adjustment is required to construct valid prediction intervals or obtain valid $p$-values (Tibshirani & Johnstone, 2014; Fithian et al., 2014; Lee et al., 2016). To the best of our knowledge, post-selection inference for the group LASSO in the context of functional data is still an open problem. Alternatively, inference after selection based on data splitting is commonly adopted for high-dimensional data; see Wasserman & Roeder (2009); Meinshausen et al. (2009). We extend these ideas to the case of functional covariate and focus on post-selection predictive inference based on data splitting and construct split conformal prediction bands following Lei et al. (2017); for
4.5 Data Application: EMG selection for finger/wrist movement

In this section, we present the variable selection and prediction results for our data collected from an AB subject across multiple postures and movement patterns as described in Section 4.3. Recall there are 16 EMG signals: 14 coming from different forearm muscles that could potentially contribute to finger or wrist movements and 2 that are randomly generated noise. The ideal selection scheme would pick one EMG signal for extension and another EMG for flexion for both finger and wrist movements, but we do not enforce this restriction in the estimation. This selection should be consistent across postures and movement patterns. In addition to identifying these EMG signals, we are also interested in a model with clear interpretability of the regression surfaces and high predictive ability.

Here we consider fitting the procedure described in Section 4.4 to our data application. Let \( i \in \{1, \ldots, N\} \) denote the current time point and \( y_i \) be the velocity at time point \( i \). Let \( z_i \) be the finger or wrist position at time point \( i \). Define \( S = [-\delta, 0] \) for some integer \( \delta > 0 \) as the time window for the recent past EMG signal relative to the current time \( i \), so the \( \delta + 1 \) s’s make up the sequence \{\( -\delta, -\delta + 1, \ldots, 0 \}\}. For \( i \geq \delta + 1 \), let \( \{X_{k,-\delta}, X_{k,-\delta+1}, \ldots, X_{k,0}\} \) be the realizations of the \( k \)th EMG signal at the original time points \{\( i-\delta, i-\delta+1, \ldots, i \)\}. Recall from Section 4.3.2 that we proposed and justified a recent past time window that spanned roughly 1/3 second, making \( \delta = 40 \) since data were collected at 120Hz. The following results were found to be relatively insensitive to other \( \delta \) within a reasonable neighborhood of 40.

The underlying muscle-movement mechanism for an AB person is known to clinicians; this prior information about the underlying true signals allows us to define the correct model size in the data application. To evaluate the model selection performance, we partition \( \mathcal{K} \) according to important movement classes based on expert knowledge. Let \( \mathcal{K}_F \) and \( \mathcal{K}_E \) denote the partitioning subsets contributing to flexion and extension movement, respectively, which depend on finger or wrist movements. Figure 4.4 shows in circled labels the clinically relevant muscles for different movement classes for an AB person. Specifically for finger movements, \( \mathcal{K}_F = \{12\} \) and \( \mathcal{K}_E = \{5, 7\} \); these muscles are known as flexor digitorum and extensor digitorum, respectively. Similarly for wrist movements, \( \mathcal{K}_F = \{8, 10, 11, 14\} \) which are known as flexor carpi ulnaris, flexor digitorum superficialis, flexor carpi ulnaris, and flexor carpi...
radials, respectively, and \( \mathcal{K}_E = \{2, 7, 13, 15\} \) which are coined as \textit{extensor carpi radialis longus, extensor digitorum, extensor carpi radialis brevis, and extensor carpi ulnaris}, respectively. All muscles associated with a specific movement class produce highly correlated signals. The expert consensus is that one muscle in \( \mathcal{K}_F \) and another in \( \mathcal{K}_E \) are sufficient to define the relationship between the movements and EMG signals. The ideal index set identifies one muscle in each movement class or equivalently \(|\mathcal{K} \cap \mathcal{K}_F| = 1\) and \(|\mathcal{K} \cap \mathcal{K}_E| = 1\), where \(|\cdot|\) denotes the cardinality of a set.

Figure 4.4: Reference forearm muscles for finger (top) and wrist (bottom) movements.

We use the method described in Section 4.4 to study which EMG signals \( X_{k,i} \)'s are related to the velocity and determine representative members of the sets \( \mathcal{K}_F \) and \( \mathcal{K}_E \). We also consider three alternative approaches; the first competitor is the method proposed by Gertheiss \textit{et al.} (2013b) in which the authors propose the functional variable selection technique using adaptive group LASSO (agLASSO) type penalized criterion inducing both sparsity and smoothness. The second competitor is the method proposed by Pannu & Billor (2017) extending the idea of Gertheiss \textit{et al.} (2013b); here the authors use the objective function based on the least absolute deviation and select the functional variables using group lasso penalty. This approach is explicitly designed to account for potential outliers in the functional predictors, reducing overfitting of the regression surfaces. We refer this approach to LAD-gLASSO. The third competitor (Fan \textit{et al.}, 2015) uses functional additive regression (FAR) with a groupwise smoothly clipped absolute deviation (gSCAD) penalty, denoted by FAR-gSCAD. Unlike other approaches, FAR-gSCAD imposes penalty on the integral.
\[ \int_S X_{k,i}(s) \gamma_k(s) ds \] assuming the integral is well-defined. Note that all these methods are designed to assess the relationship between the scalar responses and functional predictors. In stark contrast to SAFE-gLASSO, all three competitors assume non-varying smooth coefficients, \( \gamma_k(\cdot) \)'s defined on \( S \), which accounts for the passive forces but assumes the relationship is not position dependent.

The performance of all the methods is evaluated in terms of how close they yield an ideal selection of the EMG signals as well as their prediction ability. Specifically if \( \hat{K} \) denotes an estimator produced by one of the methods, we evaluate its performance by

- **Size** = \( |\hat{K}| \), ideally size = 2 with \( |\hat{K} \cap K_F| = 1 \) and \( |\hat{K} \cap K_E| = 1 \);
- **Sparsity (SP)** = 1 - \( |\hat{K}|/K \). In our application, \( K \) is 16 and the optimal sparsity is 0.88 which is associated with the ideal model size 2. Define the relative sparsity (RSP) with respect to the optimal sparsity by \( \text{RSP} = \text{SP}/0.88 \);
- **False positive rate (FPR)** = \( FP / |K^c| \); where we define the number of false positives (\( FP \)) or falsely identified EMG signals in \( \hat{K} \) by \( FP = |\hat{K} \cap K^c| \), and \( K^c \) is the complement of \( K \);
- **True positive rate (TPR)** = \( TP/2 \); where the number of true positives (\( TP \)) is defined by \( TP = 1(|\hat{K} \cap K_F| \geq 1) + 1(|\hat{K} \cap K_E| \geq 1) \) and focuses only on whether we capture the two index sets \( K_F \) and \( K_E \);
- **Mean squared error (MSE)** = \( \sum_{i=1}^{N}(y_i - \hat{y}_i)^2/N \).

Note that in general one cannot calculate FPRs and TPRs in a data application as the underlying truth is unknown. Our definitions for the above matrices are somewhat unconventional focusing on group identification of \( K_F \) and \( K_E \) rather than identification of all \( K \).

### 4.5.1 Computational Details

We briefly describe the computational details of our implementation as performed in R, starting with the necessary steps after data processing including construction of the recent past EMG curves using \( \delta = 40 \). We approximate the \( \gamma_k(\cdot, \cdot) \) using a tensor product of two orthogonal B-splines with \( L = 12 \) basis functions in the \( s \) direction and \( M = 22 \) basis functions in the \( z \) direction. Knots are placed uniformly throughout \( S \) and \( Z \). The tuning parameters to
produce the final estimates at each stage are chosen using 5-fold CV following the procedure described in Roberts et al. (2017).

For the competitors, the smooth coefficient functions \( \{ \gamma_k(\cdot) \}_{k=1}^{16} \) are modeled using B-splines with 12 basis functions and the sparsity-smoothness tuning parameters are estimated by 5-fold CV (Roberts et al., 2017) for all the competing methods. We used the R package grplasso (Meier, 2009) to fit the agLASSO (Gertheiss et al., 2013b) and rqPen (Sherwood et al., 2017) to fit the robust LAD-gLASSO (Pannu & Billor, 2017). We used the code provided by the corresponding author Fan et al. (2015) for FAR-gSCAD.

### 4.5.2 Variable selection, prediction, and implication

Table 4.1 shows the results of the model selection performance for the finger movements. In most cases all the competing methods select one EMG signal that contributes for finger extension and another for finger flexion. All methods perform fairly well in terms of TPR. In particular, the proposed SAFE-gLASSO attains desirable values of TPRs, FPRs, and RSPs across all settings. We found that the first stage of SAFE-gLASSO tended to select more EMG signals than needed but the second stage corrected the surplus. In contrast, agLASSO exhibits good numerical performance reflected by the optimal levels of FPRs and RSPs but did have poor TPR in the last two settings. LAD-gLASSO exhibited large FPRs and RSPs, indicating a high Type I error rate. FAR-gSCAD demonstrates good model selection performance but selected more EMG signals than necessary, as indicated by its RSP values. The numerical performances are similar for the wrist movements and due to the interest of the space, we include the results in Table C.1 of the Supplementary Materials, Section C.2.3.

Figure 4.5 illustrates graphically the MSEs of the competitive methods. For both finger and wrist movements, SAFE-gLASSO outperforms the competitors across all patterns and postures. The proposed approach was able to consistently fit the data better with a smaller model size, specifically relative to FAR-gSCAD and LAD-gLASSO.

Figure 4.6 shows a segment of the fitted velocities based on SAFE-gLASSO and agLASSO where the shaded region corresponds to the split conformal prediction bands following Lei et al. (2017). SAFE-gLASSO yields superior prediction trajectory than agLASSO in the events with a larger absolute velocity. In the highlighted event of Figure 4.6, fingers return to the neutral position around 1.45 to 1.88 seconds due to passive forces and while both methods predict movement, agLASSO greatly underestimates the observed velocity. The improvement is most directly attributable to the inclusion of position information. The agLASSO relies
on the assumption that the effect of the EMG signals on finger/wrist movements is invariant across $z_i$, which appears to limit the fit across a variety of movement events.

Since our methodology offers low-dimensional modeling with negligible false positives, biomedical engineers can use our method as a screening approach and focus primarily on the selected subsets of forearm muscles in collecting data from TRAs. This approach significantly reduces the burden of data collection with respect to time and cost. Furthermore, since our model and that by Crouch & Huang (2016) account for similar biomechanical phenomenon, our variable selection results may be directly applied to their method.

Table 4.1: Finger movements EMG signal selection for consistent (top three rows) and random (bottom three row) patterns at different postures and results with superscript † correspond to the first stage of SAFE-gLASSO. RSPs, number of variables (square brackets), and the percentages of TPRs and FPRs are presented.

<table>
<thead>
<tr>
<th>Pattern</th>
<th>agLASSO</th>
<th>LAD-gLASSO</th>
<th>FAR-gSCAD</th>
<th>SAFE-gLASSO</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RSP</td>
<td>TPR</td>
<td>FPR</td>
<td>RSP</td>
</tr>
<tr>
<td>1.00</td>
<td>100</td>
<td>0</td>
<td>0.57</td>
<td>100</td>
</tr>
<tr>
<td>[2]</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1.00</td>
<td>100</td>
<td>0</td>
<td>0.72</td>
<td>100</td>
</tr>
<tr>
<td>[2]</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1.00</td>
<td>100</td>
<td>0</td>
<td>0.57</td>
<td>100</td>
</tr>
<tr>
<td>[2]</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1.00</td>
<td>100</td>
<td>0</td>
<td>0.78</td>
<td>100</td>
</tr>
<tr>
<td>[2]</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1.06</td>
<td>50</td>
<td>0</td>
<td>0.85</td>
<td>100</td>
</tr>
<tr>
<td>[1]</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1.06</td>
<td>50</td>
<td>0</td>
<td>0.85</td>
<td>100</td>
</tr>
<tr>
<td>[1]</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Figure 4.5: Comparison of MSEs based on alternative approaches. Results correspond to finger (left) and wrist (right) movements with different patterns and postures.

Figure 4.6: Predicted velocity corresponding to consistent finger movements based on SAFE-gLASSO (red dashed line) and agLASSO (blue dashed line). Vertical reference lines are drawn at 1.45 and 1.88 second at which movements due to passive forces occur in the absence of muscle contractions. Shaded regions correspond to pointwise 95% split conformal prediction bands for SAFE-gLASSO.

4.5.3 Estimation of smooth coefficients

Figure 4.7 illustrates the estimated coefficient functions for the finger (top) and wrist (bottom) movements. For the purpose of illustration, we focus on interpreting the second posture shown in Table 4.1. In particular the top and bottom panels of Figure 4.7 correspond to con-
sistent finger and wrist movement, respectively, for the two identified EMG signals from SAFE-gLASSO. For finger movement, there is a clear distinction in the movement contributions: concurrent contraction of *extensor digitorum* is responsible for finger extension (positive velocity) while *flexor digitorum* leads to flexion (negative velocity). This matches with the intuition of the biomechanical system. As expected, these associations also depend on the positions. Indeed, the impact of the left surface on the velocity is most important for angles between 20 and 55 radians which correspond to finger flexion. Similarly, the impact of the right surface is most relevant between angles $-40$ and 10 radians which are primarily associated with finger extension. The *flexor digitorum* only leads to finger flexion when the hand is in a neutral position (0 to 15 radians).

Finger flexion occurs at other positions occur due to past behavior of the *extensor digitorum*. The observed concurrent relationship described above has the opposite historical relationship for the positions when the signals are presently active. This implies that past activation of one of these EMG signals can lead to the opposite type of concurrent movement they produce. In particular, this corresponds to two difficult cases where the model tries to establish the systematic relationship between: (1) passive force movements in the absence of muscle contraction, and (2) lack of movements in the presence of consistent muscle contraction due to physical constraints. Our model borrows information from the past, where the muscles were active, to predict such passive movements. Failing to appropriately account for passive forces results in poor prediction performance in those cases as shown in Figure 4.6.

The bottom panels of Figure 4.7 plot the estimated regression coefficients for the selected EMG channels for wrist movements. The interpretation of the regression surfaces for the wrist flexion/extension follows the same intuition as of hand movements. Concurrently, *flexor carpi ulnaris* leads to wrist flexion around 0 to 50 radians while *extensor carpi ulnaris* yields wrist extension around -60 to -15 radians. When the wrist is in a neutral position, *flexor carpi ulnaris* has to be flexed to keep the wrist upright around -5 to -15 radians. As before, the observed concurrent relationship exhibits the opposite historical association for wrist flexion and extension.

The regression surfaces for other patterns of finger/wrist movements can be interpreted in the similar manner; we report the regression surfaces corresponding to other patterns in Section C.2.5 of the Supplementary Material.
Figure 4.7: Regression surfaces for finger flexion (top-left), finger extension (top-right), wrist flexion (bottom-left), and wrist extension (bottom-right). The selected forearm muscles are pointed out by red circles.

4.6 Simulation study

4.6.1 Simulation mimicking the data application

In this section, we consider a simulation study that mimicks the EMG data from Section 4.5. Specifically, we focus on finger movements with consistent movement. Let the observed data be \([y_i, z_i, \{X_{k,i}(s_r); r = 1, \ldots, 41\}; k = 1, \ldots, 16, i = 1, \ldots, N]\); where \(y_i\) is now the simulated velocity at time \(i\), and all other data match the descriptions from section 4.5. In particular,
consider the generating model $y_i = \alpha + \sum_{k=1}^{16} \int_S X_{k,i}(s) \gamma_k(s, z_i) ds + \epsilon_i$ where $\gamma_k(\cdot, \cdot) = 0$ for all $k$ except $k = 5$ and $k = 12$, and $\gamma_5(\cdot, \cdot)$ and $\gamma_{12}(\cdot, \cdot)$ are the estimated effects from one of the scenarios. In addition, $\epsilon_i$ is a zero-mean error process with isotropic covariance function described by
\[
cov(y_i, y_{i'}) = \sigma_h^2 \left[ \mathbb{I}(i = i') + \theta \exp \left\{ -\frac{|i - i'|/\eta}{\sqrt{\log(\kappa)}} \right\} \right].
\]
Here $\theta$ is related to the dominant sources of dependence; $\theta = 0$ means that the responses are uncorrelated while large $\theta$ reflects higher degree of autocorrelation between the responses. In addition, $\eta$ controls the strength of correlation between any two measurements. The choice of $\eta$ is driven by the parameter $\kappa$, the correlation between any two consecutive measurements, using the relationship $\eta = |i - i'|/\sqrt{-\log(\kappa)}$.

The simulation study varies over three factors: The first factor we examine is the dominant sources of dependence determined by $\theta$.

- **A1. Dominant white noise.** $\theta << 1$.
- **A2. Equal process.** $\theta = 1$.
- **A3. Dominant dependent process.** $\theta >> 1$.

The second factor is the strength of the correlation between successive measurements.

- **B1. Low correlation.** $\kappa = 0.2$.
- **B2. High correlation.** $\kappa = 0.9$.

The third factor is the magnitude of noise variance $\sigma^2_h$.

- **C1. Small noise.** $\sigma^2_h = 0.01\sigma^2$ corresponds to the situation of having smaller variance than that of the original data.
- **C2. Large noise.** $\sigma^2_h = \sigma^2$ corresponds to the case of equal variance.

Let $\mu_i = \alpha + \sum_{k=1}^{16} \int_S X_{k,i}(s) \gamma_k(s, z_i) ds$. Calculate signal-to-noise ratio (SNR) as $\text{SNR} = \frac{\text{var}(\mu_i)}{\text{var}(\epsilon_i)}$. In particular, we consider $\text{SNR} = \{875, 445, 80, 8, 4, 0.8\}$. The results are based on 100 independent samples for each combination of the simulation settings.

Table 4.2 illustrates the numerical performance corresponding to the simulated kinematic data for different SNRs. The results are consistent to the findings of finger movements in Section 4.5. As before, we observe that the mean model size of the second stage of the
proposed is nearer to the truth (i.e. 2) than that of the first stage; see the “SP” column under SAFE-gLASSO in Table 4.2. This is expected due to the fact that by adopting the two-stage variable selection scheme, we shrink the surplus variables in the second stage of the procedure. The proposed approach exhibits lower FPRs than that of the agLASSO, LAD-gLASSO, and FAR-gSCAD across all simulation settings. Similar to the other approaches, SAFE-gLASSO also attains high TPRs.

SAFE-gLASSO exhibited superior prediction performance relative to the competing methods across all simulation settings; see Figure 4.8. These findings are in agreement with the results of Section 4.5. As expected the prediction accuracy improves with the SNR; compare the box-plots corresponding to the SAFE-gLASSO for different SNRs in Figure 4.8. Unlike the proposed approach, the numerical performance of the alternative methods suffers due to not considering positing-varying smooth coefficients as the methods fail to quantify the systematic relationship between the velocity and muscle activities of a biomechanical model. We report the additional simulation results in the Supplementary Materials, Section C.3.

Table 4.2: Analysis of finger movements with fixed motion. Data is generated assuming noise variance $C1$ and $C2$ with different dominant processes ($A1$-$A3$) for high correlation coefficient ($B2$). Reported are the SPs (%), model size (in square brackets), TPRs (%), and FPRs (%) averaged over 100 simulations. Results with superscript $\dagger$ correspond to the first stage of SAFE-gLASSO.
4.6.2 Numerical experiment

Next we consider another simulation study where data are generated from a purely mathematical perspective. Let \([y_i, z_i, \{X_{k,i}(s_r); r = 1, \ldots, 100\}, k = 1, \ldots, 10, i = 1, \ldots, 500]\) be the simulated data; where \(y_i\) is a scalar response at an instance \(i\), \(z_i\) is a scalar covariate observed at \(i\), \(X_{k,i}(s_r)\)’s are the realizations of the \(k\)th functional predictor such that \(s_r \in [0, 1]\). The 10 functional predictors are generated similar to Gertheiss et al. (2013b) and Pannu & Billor (2017). The nonzero functional coefficients \(\{\gamma_k(\cdot, \cdot); k = 1, 2\}\)’s are varying over \(z\), \(z_i \in [-1, 1]\), and defined as \(\gamma_1(s, z) = 1 + \sqrt{2}Cz + \sqrt{2}k\cos(\pi s)\) and \(\gamma_2(s, z) = 1 + C\exp(-0.5z) + s + 0.5s^2\). Note that \(C\) controls the strength of functional coefficients over \(z\). The error variance \(\sigma^2\) is calculated based on different SNRs; i.e., \(\text{SNR} = \{0.5, 1, 5\}\). More details about the functional predictors, functional coefficients, and simulation results are summarized in the Supplementary Materials, Section C.4.

As expected, when \(C = 0\), meaning the functional coefficient should have only one dimension, \(s\), all competitors showed comparable performance across all metrics, having high TPR and similar MSE. All methods except LAD-gLASSO had low FPR, which consistently selected 5 of the 10 possible variables. As the difficulty of the model increases, say for \(C = 5\) the competitors perform much worse than SAFE-gLASSO. FAR-gSCAD and agLASSO have low FPR, but tend to select only 1 of the 2 important variables, while LAD-gLASSO has high TPR but also high FPR, similar to the \(C = 0\) scenario. As expected, the MSE performance
also deteriorates for the competitors as $C$ increases.

We also investigated the coverage probabilities and length of the prediction intervals for the SAFE-gLASSO models. Denote the lower and upper bound of the in-sample prediction interval by $C_{\text{roo}}(X_i) = (C_{\text{roo}}^l(X_i), C_{\text{roo}}^u(X_i))$, which is constructed following rank-one-out (ROO) split conformal prediction inference (Lei et al., 2017). Define the pointwise average coverage probability by $CP_{\text{roo}} = \sum_{i=1}^{N} \mathbb{1}\{y_i \in C_{\text{roo}}(X_i)\}/N$ where $N$ is the total number of in-sample observations. Calculate the expected length of the interval by $\sum_{i=1}^{N} (C_{\text{roo}}^u(X_i) - C_{\text{roo}}^l(X_i))/N$. Let $i^*$ index the instance at which new information of the predictor $X_{i^*}$ is recorded; where $i^* = 1, \ldots, M$, and $M$ is the total number of future observations. Define the pointwise average coverage probability for the future observations by $CP_{\text{split}} = \sum_{i^*=1}^{M} \mathbb{1}\{y_{i^*} \in C_{\text{split}}(X_{i^*})\}/M$; where $C_{\text{split}}(X_{i^*})$ is the prediction intervals computed using the split conformal prediction inference (Lei et al., 2017). In our simulation, we consider $M = 250$.

Table 4.3 demonstrates the predictive inference of the proposed approach in terms of the actual coverage and expected length of the interval for both the in-sample ($y_i$) and future ($y_{i^*}$) observations. In general, the average coverage stays around the nominal levels irrespective of the complexity of the model as defined by $C$; see the results for $C = 0$ and $C = 10$. As expected, the width of the interval increases as $C$ departs from 0 but reduces for larger SNR; see the average length of the interval for $C = 10$ at miscoverage level $\alpha = 0.20$ for $y_i$. We also observe the similar phenomenon with the prediction intervals for the future observations.
Table 4.3: Average pointwise coverage of 80%, 90%, 95%, and 99% prediction bands for the in-sample ($y_i$) and future ($y_i^*$) observations using ROO and split conformal prediction inference, respectively; standard errors (in parenthesis) and average lengths (in square bracket) of prediction bands are reported. Results are based on 100 simulations.

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>SNR = 0.5</th>
<th>SNR = 5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$C = 0$</td>
<td>$C = 5$</td>
</tr>
<tr>
<td>$y_i$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.01</td>
<td>0.990 (0.010) [1.169]</td>
<td>0.990 (0.010) [5.202]</td>
</tr>
<tr>
<td>0.05</td>
<td>0.949 (0.022) [1.877]</td>
<td>0.950 (0.022) [8.870]</td>
</tr>
<tr>
<td>0.10</td>
<td>0.900 (0.030) [2.745]</td>
<td>0.900 (0.030) [5.14]</td>
</tr>
<tr>
<td>0.20</td>
<td>0.800 (0.040) [0.572]</td>
<td>0.799 (0.040) [2.504]</td>
</tr>
<tr>
<td>$y_i^*$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.01</td>
<td>0.992 (0.009) [1.168]</td>
<td>0.990 (0.010) [5.209]</td>
</tr>
<tr>
<td>0.05</td>
<td>0.949 (0.022) [1.877]</td>
<td>0.948 (0.022) [8.860]</td>
</tr>
<tr>
<td>0.10</td>
<td>0.898 (0.030) [2.745]</td>
<td>0.900 (0.030) [5.228]</td>
</tr>
<tr>
<td>0.20</td>
<td>0.862 (0.040) [0.571]</td>
<td>0.802 (0.040) [2.817]</td>
</tr>
</tbody>
</table>
4.7 Discussion

In this paper, we proposed a covariate-dependent, scalar-on-function regression model that appropriately accounts for the biomechanical processes involved in hand movement, such as passive forces and physical constraints. The functional predictors were the recent past behavior of EMG signals measured across multiple muscles in the subject’s forearm and the responses were finger and wrist velocity. The functional coefficients for each EMG signal were allowed to vary based on the current finger or wrist position. The bivariate coefficients were then approximated using a tensor product of rich basis expansions that were then estimated with a combined multi-dimensional smoothing and sparseness penalty, which is an extension of Gertheiss et al. (2013b). We developed a two-step variable selection procedure, called Sequential Adaptive Functional Empirical group LASSO (SAFE-gLASSO), that was shown through numerical investigations to have superior performance over standard selection approaches (Gertheiss et al., 2013b; Pannu & Billor, 2017; Fan et al., 2015) by reducing the number of false positives irrespective of model complexity.

The results of the data application showed SAFE-gLASSO was able to identify the important EMG signals for finger and wrist movement for an AB subject. Furthermore, the estimated varying functional coefficients were relatively sparse, easy to interpret, and had exceptional predictive performance compared to standard selection approaches. Our model and fitting algorithm have great potential to outperform current state-of-the-art data driven methods for prosthesis control such as pattern recognition because they ignore biomechanical constraints, do not perform variable selection, and are prone to overfitting. The variable selection results from SAFE-gLASSO could also be incorporated in the method by Crouch & Huang (2016), which currently does not perform variable selection and uses a planar link-segment dynamic model. Although our model mathematically differs from theirs, both models account for the biomechanical system with enough similarity that our variable selection results still apply. Our model does have the advantage in that, after training our model, there is minimal data processing required to produce predictions in the prosthetic limb, reducing the burden of real-time data collection.

There are many opportunities for extensions and new applications of the approaches taken in this paper. The two-step fitting approach used in SAFE-gLASSO easily applies to the more common univariate functional coefficients. We also discussed an approach to assess the predictive ability based on data splitting, which as far as the authors are aware has not been applied to functional regression. Additionally, our model, which uses a tensor product
basis, can also be applied in a more general setting with many functional predictors having functional coefficients varying over multiple covariates. This would be of interest for this data application if the coefficients were considered to vary significantly across postures.

In our current developments, the functional covariates were assumed to be observed without error on a fine grid of points. Extensions to a situation where the functional measurements are perturbed by error or when the grid points $s_r$'s are sparse would require a preliminary smoothing of the functional covariates using existing approaches Yao et al. (2005); Xiao et al. (2016) and then SAFE-gLASSO may be employed on the smoothed functional covariates. In a preliminary investigation, not reported here, we observed that the variable selection performance of SAFE-gLASSO is unaffected due to the use of estimated smooth profiles in the place of true functional predictors. However prediction error of the responses is affected by large noise variance of the functional predictors, as would any method.

As in non-parametric regressions, our method relies on the tensor products of basis functions; as a result, the number of parameters to estimate can explode very quickly. Even though our method can tackle the dimensionality issue, it becomes computationally intensive with large number of basis functions. Note that we adopt CV based approach in selecting the tuning parameters for the data application. We also acknowledge that there are other methods to select optimum tuning parameters and CV based approach may not be theoretically the best approach (Gertheiss et al., 2013b). While the method works well for our data application, future research is needed to investigate the optimality in selecting the tuning parameters. In addition, we focused on the development of a subject-specific modeling procedure and were not concerned with estimating subject-specific variability. By accounting for this source of variability, we can develop highly functional, user-specific robotic prosthetic limb that perform well across multiple subjects through estimation of population parameters.
BIBLIOGRAPHY


Benavides, Mendoza, Maria, Santa, et al. 2016. Dietary supplementation of the osmolyte betaine to swine exposed to heat stress.


Crainiceanu, Ciprian, Reiss, Philip, Goldsmith, Jeff, Huang, Lei, Huo, Lan, Scheipl, Fabian, Greven, S, Harezlak, J, Kundu, MG, & Zhao, Y. 2012a. refund: Regression with functional data. *R package version 0.1-6*.


110


Meier, Lukas. 2009. grplasso: Fitting user specified models with group lasso penalty. *R package version 0.4-2*.


Park, So Young, Li, Cai, Han, Zhen, Mendoza, Santa-Maria, Heugten, Eric van, & Ana-María, Staicu. 2016. Conditional analysis for mixed covariates, with application to feed intake of lactating sows. *preprint*.


Scheipl, Fabian, & Bolker, Ben. 2016. Package rlrsim.

Scheipl, Fabian, Greven, Sonja, & Küchenhoff, Helmut. 2008. Size and power of tests for a zero random effect variance or polynomial regression in additive and linear mixed models. Computational statistics & data analysis, 52(7), 3283–3299.


Scheme, Erik, Fougner, A., Stavdahl, Ø, Chan, A. C., & Englehart, Kevin. 2010. Examining the adverse effects of limb position on pattern recognition based myoelectric control. Pages 6337–6340 of: 2010 annual international conference of the ieee engineering in medicine and biology. IEEE.


117


APPENDICES
Appendix A

Additional details for Chapter 2

This Supplementary Material contains three sections. Section A.1 provides additional simulation results for the settings described in the main article, for both Gaussian and binary cases. Section A.2 presents additional analysis results for the sows data application. Section C provides technical details for implementation.

A.1 Additional simulation results

A.1.1 Predictor functions

The choice of the predictor curves $W_{ij}(\cdot)$ in the simulation study was purely arbitrarily. Figure ??, left and middle panels, depict the shape of $W_{ij}(\cdot)$ in two scenarios, one corresponding to SNR = 0.5 (left) and the other corresponding to SNR = 2.5 (middle). In the current re-submission we also considered a simulation scenario mimicking the data application. Figure ?? right panel, illustrates the covariates $W_{ij}(\cdot)$ in this case with SNR = 1.3.
Table A.1 describes the signal-to-noise ratio for different simulation settings. Define $\mu_i(t) = \alpha(t) + \int X_i(s,t) \gamma(s,t) ds$. Signal-to-noise-ratio for response is defined as

$$SNR_y = \frac{\int_T \text{var}\{\mu_i(t)\} dt}{\sigma^2 + \sigma^2_2 + \sigma^2_3 + \int_T \text{var}\{\epsilon_i(t)\} dt}.$$ 

Notice, this SNR is a sort of effective SNR and accounts for the noise level in the functional covariate as well as the noise level in the response.

![Gaussian response](image)

Figure A.1: Functional observations for a randomly chosen subject for the initial simulation design with SNR = 0.5 (left) and with SNR = 2.5 (middle) and for the simulation mimicking the data application, SNR = 1.3 (right). The solid lines correspond to the smoothed profiles.

<table>
<thead>
<tr>
<th>SNR = 0.5 (B1)</th>
<th>$\delta = 1$</th>
<th>$\delta = 2$</th>
<th>$\delta = 5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>No dependence (IID)</td>
<td>0.1</td>
<td>0.3</td>
<td>1.0</td>
</tr>
<tr>
<td>Compound symmetry (D1ii)</td>
<td>0.1</td>
<td>0.3</td>
<td>1.0</td>
</tr>
<tr>
<td>Random effects model (D1iii)</td>
<td>0.1</td>
<td>0.3</td>
<td>1.0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SNR = 2.5 (B2)</th>
<th>$\delta = 1$</th>
<th>$\delta = 2$</th>
<th>$\delta = 5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>No dependence (IID)</td>
<td>0.4</td>
<td>0.8</td>
<td>3.0</td>
</tr>
<tr>
<td>Compound symmetry (D1ii)</td>
<td>0.5</td>
<td>1.0</td>
<td>3.5</td>
</tr>
<tr>
<td>Random effects model (D1iii)</td>
<td>0.5</td>
<td>1.0</td>
<td>3.5</td>
</tr>
</tbody>
</table>

Table A.1: $SNR_y$ levels for different simulation settings; values are rounded to one decimal place.

### A.1.2 Gaussian response

In this section, we present additional results from the settings described in Section 5.1 of the main manuscript, for Gaussian responses. They confirm high prediction accuracy of the proposed LDFR compared with LPFR (Goldsmith et al. (2012a)), irrespective of the complexity of the dependence among the repeated measurements, or the number of repeated...
instances per subject, or the level of noise in measuring the functional covariate. In most cases the model is correctly specified although we looked at mild misspecification of the model covariance structure.

Tables A.2-A.3 consider Gaussian responses generated with independence covariance structure \((D_{1i})\) observed in sparse \((C_1)\) and moderately sparse \((C_2)\) longitudinal design, and functional covariates observed with high \((B_1)\) and low \((B_2)\) noise variance, and with effect \(E_1\). The model is fitted using CS type covariance structure. We observe that both IN and OUT prediction errors are similar for LDFR and LPFR models when \(\delta\) is equal to zero. As \(\delta\) value departs from zero, LDFR outperforms LPFR in all settings. As expected, prediction accuracy improves with the number of repeated measurements per subject. Also the prediction error does not seem to be affected by the magnitude of the noise in the functional predictor at least for the magnitudes of errors that we studied (compare the the two tables).

Tables A.4-A.6 show the prediction accuracy results for Gaussian responses and various dependence structures \((D_{1ii} and D_{1iii})\) using correct model specification for the covariance of the responses and with effect \(E_1\). The findings are similar as before: LDFR outweighs LPFR method as \(\delta\) departs from zero, while the two models perform similarly for \(\delta = 0\).

Table A.7 illustrates the numerical performance of the proposed and LPEER method; we generate the functional coefficient as \(E_2\) and Gaussian responses with CS type covariance structure \((D_{1ii})\) for sparse \((C_1)\) longitudinal design while functional covariates are observed with high \((B_1)\) and low \((B_2)\) SNR. As previously, we notice that both methods perform competitively for \(\delta = 1\), however as \(\delta\) departs from 1, LDFR outperforms LPEER in terms of prediction accuracy and computing time.
Table A.2: Gaussian responses with independent covariance structure ($D_{1i}$), when the longitudinal design is sparse ($C_1$) and moderately sparse (mod sparse, $C_2$); the functional covariates are observed with high noise variance ($B_{1i}$) with effect $E_1$. Model is fitted assuming CS type dependence structure. Median prediction errors and IQR in parenthesis are reported for 1000 simulations.

<table>
<thead>
<tr>
<th></th>
<th>$\delta = 0$</th>
<th>$\delta = 2$</th>
<th>$\delta = 5$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\text{IN}_{PF}$</td>
<td>$\text{OUT}_{PF}$</td>
<td>$\text{IN}_{PF}$</td>
</tr>
<tr>
<td>sparse $I = 100$</td>
<td>LDFR</td>
<td>LDFR</td>
<td>LPFR</td>
</tr>
<tr>
<td>sparse</td>
<td>2.03</td>
<td>2.08</td>
<td>2.11</td>
</tr>
<tr>
<td></td>
<td>(0.08)</td>
<td>(0.08)</td>
<td>(0.09)</td>
</tr>
<tr>
<td>mod sparse</td>
<td>2.01</td>
<td>2.08</td>
<td>2.04</td>
</tr>
<tr>
<td></td>
<td>(0.05)</td>
<td>(0.05)</td>
<td>(0.09)</td>
</tr>
<tr>
<td>sparse $I = 200$</td>
<td>LDFR</td>
<td>LDFR</td>
<td>LPFR</td>
</tr>
<tr>
<td>sparse</td>
<td>2.03</td>
<td>2.09</td>
<td>2.09</td>
</tr>
<tr>
<td></td>
<td>(0.06)</td>
<td>(0.05)</td>
<td>(0.07)</td>
</tr>
<tr>
<td>mod sparse</td>
<td>2.01</td>
<td>2.08</td>
<td>2.03</td>
</tr>
<tr>
<td></td>
<td>(0.03)</td>
<td>(0.04)</td>
<td>(0.06)</td>
</tr>
<tr>
<td>sparse $I = 300$</td>
<td>LDFR</td>
<td>LDFR</td>
<td>LPFR</td>
</tr>
<tr>
<td>sparse</td>
<td>2.04</td>
<td>2.08</td>
<td>2.08</td>
</tr>
<tr>
<td></td>
<td>(0.05)</td>
<td>(0.05)</td>
<td>(0.05)</td>
</tr>
<tr>
<td>mod sparse</td>
<td>2.02</td>
<td>2.08</td>
<td>2.03</td>
</tr>
<tr>
<td></td>
<td>(0.03)</td>
<td>(0.03)</td>
<td>(0.05)</td>
</tr>
</tbody>
</table>
Table A.3: Gaussian responses with independent covariance structure ($D_{1i}$), when the longitudinal design is sparse ($C_1$) and moderately sparse (mod sparse, $C_2$); the functional covariates are observed with low noise variance ($B_2$) with effect $E_1$. Model is fitted assuming CS type dependence structure. Median prediction errors and IQR in parenthesis are reported for 1000 simulations.

<table>
<thead>
<tr>
<th></th>
<th>$\delta = 0$</th>
<th></th>
<th>$\delta = 2$</th>
<th></th>
<th>$\delta = 5$</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>INPE</td>
<td>OUTPE</td>
<td>INPE</td>
<td>OUTPE</td>
<td>INPE</td>
<td>OUTPE</td>
</tr>
<tr>
<td></td>
<td>LDFR LPFR</td>
<td>LDFR LPFR</td>
<td>LDFR LPFR</td>
<td>LDFR LPFR</td>
<td>LDFR LPFR</td>
<td>LDFR LPFR</td>
</tr>
<tr>
<td>sparse</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I = 100</td>
<td>2.02 (0.08)</td>
<td>2.06 (0.08)</td>
<td>2.10 (0.09)</td>
<td>2.09 (0.09)</td>
<td>2.07 (0.09)</td>
<td>2.33 (0.11)</td>
</tr>
<tr>
<td>mod sparse</td>
<td>2.01 (0.05)</td>
<td>2.06 (0.05)</td>
<td>2.07 (0.09)</td>
<td>2.07 (0.09)</td>
<td>2.07 (0.09)</td>
<td>2.37 (0.11)</td>
</tr>
<tr>
<td>sparse</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I = 200</td>
<td>2.03 (0.05)</td>
<td>2.07 (0.06)</td>
<td>2.09 (0.07)</td>
<td>2.07 (0.07)</td>
<td>2.34 (0.08)</td>
<td>2.14 (0.08)</td>
</tr>
<tr>
<td>mod sparse</td>
<td>2.01 (0.04)</td>
<td>2.07 (0.04)</td>
<td>2.08 (0.06)</td>
<td>2.03 (0.05)</td>
<td>2.35 (0.05)</td>
<td>2.04 (0.07)</td>
</tr>
<tr>
<td>sparse</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I = 300</td>
<td>2.03 (0.04)</td>
<td>2.07 (0.05)</td>
<td>2.10 (0.05)</td>
<td>2.07 (0.05)</td>
<td>2.35 (0.05)</td>
<td>2.12 (0.05)</td>
</tr>
<tr>
<td>mod sparse</td>
<td>2.01 (0.03)</td>
<td>2.07 (0.03)</td>
<td>2.08 (0.05)</td>
<td>2.03 (0.05)</td>
<td>2.35 (0.05)</td>
<td>2.04 (0.07)</td>
</tr>
</tbody>
</table>
Table A.4: Gaussian responses with dependent covariance structure ($D_{1ii}$), when the longitudinal design is sparse ($C_1$) and moderately sparse (mod sparse, $C_2$); the functional covariates are observed with low noise variance ($B_2$) with effect $E_1$. Model is fitted assuming CS type dependence structure. Median prediction errors and IQR in parenthesis are reported for 1000 simulations.

<table>
<thead>
<tr>
<th></th>
<th>$\delta = 0$</th>
<th>$\delta = 2$</th>
<th>$\delta = 5$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>INPE</td>
<td>OUTPE</td>
<td>INPE</td>
</tr>
<tr>
<td></td>
<td>LDFR</td>
<td>LDFR</td>
<td>LPFR</td>
</tr>
<tr>
<td>L = 100</td>
<td>sparse</td>
<td>0.76 (0.04)</td>
<td>0.83 (0.05)</td>
</tr>
<tr>
<td></td>
<td>mod sparse</td>
<td>0.73 (0.02)</td>
<td>0.87 (0.04)</td>
</tr>
<tr>
<td>L = 200</td>
<td>sparse</td>
<td>0.75 (0.03)</td>
<td>0.85 (0.03)</td>
</tr>
<tr>
<td></td>
<td>mod sparse</td>
<td>0.73 (0.02)</td>
<td>0.87 (0.04)</td>
</tr>
<tr>
<td>L = 300</td>
<td>sparse</td>
<td>0.74 (0.03)</td>
<td>0.85 (0.03)</td>
</tr>
<tr>
<td></td>
<td>mod sparse</td>
<td>0.72 (0.02)</td>
<td>0.87 (0.03)</td>
</tr>
</tbody>
</table>
Table A.5: Gaussian responses with dependent covariance structure (D1iii), when the longitudinal design is sparse (C1) and moderately sparse (mod sparse, C2); the functional covariates are observed with high noise variance (B1) with effect E1. Model is fitted assuming both subject-specific random intercept and slope. Median prediction errors and IQR in parenthesis are reported for 1000 simulations.

<table>
<thead>
<tr>
<th></th>
<th>$\delta = 0$</th>
<th>$\delta = 2$</th>
<th>$\delta = 5$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>INPE</td>
<td>OUTPPE</td>
<td>INPE</td>
</tr>
<tr>
<td></td>
<td>LDFR LPFR</td>
<td>LDFR LPFR</td>
<td>LDFR LPFR</td>
</tr>
<tr>
<td>sparse</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I = 100</td>
<td>0.60 (0.04)</td>
<td>0.80 (0.13)</td>
<td>0.92 (0.05)</td>
</tr>
<tr>
<td></td>
<td>0.57 (0.03)</td>
<td>0.82 (0.05)</td>
<td>0.87 (0.04)</td>
</tr>
<tr>
<td>mod sparse</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I = 200</td>
<td>0.59 (0.04)</td>
<td>0.81 (0.11)</td>
<td>0.91 (0.04)</td>
</tr>
<tr>
<td></td>
<td>0.57 (0.03)</td>
<td>0.82 (0.05)</td>
<td>0.86 (0.03)</td>
</tr>
<tr>
<td>sparse</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I = 300</td>
<td>0.59 (0.04)</td>
<td>0.80 (0.10)</td>
<td>0.90 (0.03)</td>
</tr>
<tr>
<td></td>
<td>0.56 (0.03)</td>
<td>0.82 (0.04)</td>
<td>0.86 (0.03)</td>
</tr>
</tbody>
</table>
Table A.6: Gaussian responses with dependent covariance structure ($D_{1iii}$), when the longitudinal design is sparse ($C_1$) and moderately sparse (mod sparse, $C_2$); the functional covariates are observed with low noise variance ($B_2$) with effect $E_1$. Model is fitted assuming both subject-specific random intercept and slope. Median prediction errors and IQR in parenthesis are reported for 1000 simulations.

<table>
<thead>
<tr>
<th></th>
<th>$\delta = 0$</th>
<th>$\delta = 2$</th>
<th>$\delta = 5$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>INPE</td>
<td>OUTPE</td>
<td>INPE</td>
</tr>
<tr>
<td></td>
<td>LDFR</td>
<td>LPFR</td>
<td>LDFR</td>
</tr>
<tr>
<td>sparse</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$I = 100$</td>
<td>0.59 (0.04)</td>
<td>0.77 (0.04)</td>
<td>0.85 (0.05)</td>
</tr>
<tr>
<td>mod sparse</td>
<td>0.57 (0.02)</td>
<td>0.78 (0.03)</td>
<td>0.65 (0.04)</td>
</tr>
<tr>
<td>$I = 200$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sparse</td>
<td>0.58 (0.03)</td>
<td>0.78 (0.03)</td>
<td>0.78 (0.03)</td>
</tr>
<tr>
<td>mod sparse</td>
<td>0.56 (0.02)</td>
<td>0.78 (0.03)</td>
<td>0.63 (0.03)</td>
</tr>
<tr>
<td>$I = 300$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sparse</td>
<td>0.57 (0.03)</td>
<td>0.76 (0.03)</td>
<td>0.75 (0.03)</td>
</tr>
<tr>
<td>mod sparse</td>
<td>0.56 (0.02)</td>
<td>0.78 (0.02)</td>
<td>0.62 (0.02)</td>
</tr>
</tbody>
</table>
Table A.7: Gaussian responses with CS dependence structure (D1ii), when the longitudinal design is sparse (C1) for B1 and B2 with effect E2. Model is fitted assuming CS type dependence structure. Median prediction errors and IQR in parenthesis are reported.

<table>
<thead>
<tr>
<th>SNR 0.5</th>
<th>SNR 2.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>I = 100</td>
<td>I = 300</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>$\delta = 1$</th>
<th>$\delta = 5$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$IN_{PE}$</td>
<td>$OUT_{PE}$</td>
</tr>
<tr>
<td></td>
<td>LDFR</td>
<td>LPEER</td>
</tr>
<tr>
<td>0.99</td>
<td>1.26</td>
<td>1.34</td>
</tr>
<tr>
<td>0.96</td>
<td>1.19</td>
<td>1.29</td>
</tr>
<tr>
<td>0.95</td>
<td>1.25</td>
<td>1.31</td>
</tr>
<tr>
<td>0.92</td>
<td>1.18</td>
<td>1.18</td>
</tr>
</tbody>
</table>
Figure A.2 illustrates the prediction accuracy of the response trajectories using LDFR. Data are generated using setting $D1ii$ for the responses using moderately sparse design $C2$. The setting for the functional covariates is $B2$ for Figure A.2. The LDFR model is fitted with subject-specific random intercept. As expected, the prediction accuracy improves substantially as the number of observations per subject increases. Also, prediction accuracy seems to be affected slightly by the size: compare the box-plots of the residual mean prediction error for a particular $\delta$ depicted in Figure A.2. As expected, the prediction error is larger for the new subjects compared to the existing ones; this is due to the fact of not accounting the subject-specific random effects in constructing the response trajectory of new subjects.

![Figure A.2: Gaussian responses with CS dependence structure ($D1ii$), when the longitudinal design is sparse (left) and moderately sparse (right); the functional covariates are observed with low noise variance ($B2$) and effect $E1$. Fitted model assumes CS covariance structure. Reported is $RMPE_{trj}$ for observed (white boxplot) and unobserved (gray boxplot) subjects based on 1000 simulations. Reference lines are drawn at RMPE values 1, 2, and 3 for convenience.]

A.1.3 Binary response

Next we present additional results for the case when the responses are binary, as described by setting $D2$ under sparse ($C1$) and moderately sparse ($C2$) longitudinal design, and when the functional measurements are observed with low noise ($B2$). The model is fitted using a misspecified model for the dependence among the repeated measurements, by assuming a subject-specific random intercept. Table A.8 shows the prediction error for using our model
LDFR and the LPFR, which does not account for a time-varying effect of the functional covariate. As noted in the paper, the two approaches lead to similar prediction performance for $\delta = 0$; however as $\delta$ departs from zero, the accuracy improves for LDFR with respect to the estimation of linear predictors and prediction of successes (true prediction error, TPR). We notice that the magnitude of noise in the functional predictors seems to slightly affect the prediction accuracy.

Table A.8: Binary responses ($D2$), when the longitudinal design is sparse ($C1$) and moderately sparse ($C2$); the functional covariates are observed with low noise variance ($B2$) and with effect $E1$. Model is fitted assuming subject-specific random intercept. We report the median of prediction errors of the linear predictor trajectories and, in parenthesis, the median of true-positive-rates.

<table>
<thead>
<tr>
<th>$\delta = 0$</th>
<th>$\delta = 5$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>IN$P_{FR}$</strong></td>
<td><strong>OUT$P_{FR}$</strong></td>
</tr>
<tr>
<td>LDFR</td>
<td>LPFR</td>
</tr>
<tr>
<td>I = 100 sparse</td>
<td>1.26</td>
</tr>
<tr>
<td></td>
<td>(0.96)</td>
</tr>
<tr>
<td>mod sparse</td>
<td>1.19</td>
</tr>
<tr>
<td></td>
<td>(0.96)</td>
</tr>
<tr>
<td>I = 200 sparse</td>
<td>1.22</td>
</tr>
<tr>
<td></td>
<td>(0.96)</td>
</tr>
<tr>
<td>mod sparse</td>
<td>1.17</td>
</tr>
<tr>
<td></td>
<td>(0.96)</td>
</tr>
<tr>
<td>I = 300 sparse</td>
<td>1.21</td>
</tr>
<tr>
<td></td>
<td>(0.96)</td>
</tr>
<tr>
<td>mod sparse</td>
<td>1.16</td>
</tr>
<tr>
<td></td>
<td>(0.96)</td>
</tr>
</tbody>
</table>

A.1.4 Comparison of results using B-splines and truncated linear splines

In this section, we compare the numerical results when the non-linear effects $\beta_k(\cdot)$'s are approximated using B-splines and truncated linear splines with 15 basis functions. For the case with B-splines, we penalize the roughness of the functions using the second order difference penalty; see Eilers & Marx (1996, 2010). REML is used to select the smoothing parameters in both cases. Table A.9 illustrates the numerical performance in the case when the data are generated from CS type covariance structure with low noise variance ($B2$) and with effect $E1$. We observe similar in-sample and out-of-sample prediction accuracy irrespective of the
basis functions used across different sample size and sparsity level.

Table A.9: Modeling the unknown smooth parameters by B-splines and truncated linear splines. Reported are the median and IQR of in-sample and out-of-sample prediction errors. Results are based on 100 independent samples.

<table>
<thead>
<tr>
<th></th>
<th>( I = 100 )</th>
<th>( I = 300 )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( \delta = 2 )</td>
<td>( \delta = 5 )</td>
</tr>
<tr>
<td>sparse</td>
<td>( 0.86 ) (0.05)</td>
<td>( 1.28 ) (0.14)</td>
</tr>
<tr>
<td>mod sparse</td>
<td>( 0.86 ) (0.05)</td>
<td>( 1.11 ) (0.11)</td>
</tr>
</tbody>
</table>

**A.2 Additional results for the sows data application**

In this section we provide more results of the analysis of the sows data.

**A.2.1 Autoregressive model**

Model (10) does not account for the previous day feed intake, which may be viewed as an important predictor for the current day feed intake. Here, we further explore such direction by examining the residuals of the model (10) for evidence of autoregressive dependence as well as by comparing the prediction error to the counterpart obtained by accounting for the lagged dependence on feed intake.

For the first direction, we consider the mean model (10) in the current submission (model 8 in the initial submission), and examine the residuals for their dependence pattern. The residuals are obtained as the difference between the observed and the fitted values and are
shown in Figure A.3, left panel. We use the auto-correlation (ACF) and the partial ACF (PACF) measures to examine the autoregressive pattern of the residuals. The results displayed in Figure A.4 show no significant "spikes" across different lags, indicating no evidence for such autoregressive dependence in the residual term. This brief exploratory analysis shows that there is no need to use an autoregressive covariance model for the residuals $\epsilon_{ij}$ and that the dependence among the repeated measures is well captured by the random term $b_i$'s.

![Figure A.3: Residuals versus predicted values for different models: model (10) in the current submission (left panel) and lagged model (A.1) in the right.](image-url)
For the second direction, we explicitly account for the feed intake on the previous day by adding an extra term that accounts for a lagged covariate effect to the mean model described by (10). Specifically the lagged mean model we use is

$$FI_{ij} = \beta_p(t_{ij}) + \eta FI_{i,j-1} + \int Temp_{ij}(s)\gamma(s, t_{ij})ds + b_{g_i} + b_{0i(g_i)} + b_{1i(g_i)}t_{ij} + \varepsilon_{ij}, \quad (A.1)$$

where $\eta$ is an unknown parameter that quantifies the effect of the previous day feed-intake on the mean feed intake of the current day. Figure A.3 right panel, plots the residuals versus the predicted. Visually, there seems to be none or negligible difference between the residuals for model (10) and the ones for the lagged model detailed in (A.1), indicating a negligible effect of the previous day feed intake. To further examine this point we compare the prediction error for each of the two models. Define the root mean prediction error (RMPE) by

$$RMPE = \sqrt{\sum_{i=1}^{I} \left( \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_{ij})^2 / n_i \right) / I}.$$  

The two models yield: RMPE = 1.335 for model (10) in the main manuscript and RMPE = 1.343 for the lagged model (A.1). The very close prediction error confirms the point above, that the current mean model does not seem to benefit for accounting for the previous day feed intake. Furthermore, we considered more complex lagged mean models, by accounting for the feed intake at previous $d$ days. Specifically, we also examined the following lagged models

$$FI_{ij} = \beta_p(t_{ij}) + \sum_{d'=1}^{d} \eta_{d'} FI_{i,j-d'} + \int Temp_{ij}(s)\gamma(s, t_{ij})ds + b_{g_i} + b_{0i(g_i)} + b_{1i(g_i)}t_{ij} + \varepsilon_{ij}, \quad (A.2)$$

Figure A.4: ACF (left) and PACF (right) function for residuals.
for \( j \geq d + 1 \), where \( \eta_{d'} \) is the effect of the \( d' \)th day feed intake prior to the current day, for \( d = 2, 3, 4, 5 \). The prediction error, as assessed by RMPE, for each of these models is shown in Table A.10; notice that all the errors are larger than the RMPE corresponding to mean model (10), which is discussed in the manuscript and uses IID errors.

Table A.10: Prediction accuracy of the lagged models as assessed by RMPE.

<table>
<thead>
<tr>
<th></th>
<th>( d = 2 )</th>
<th>( d = 3 )</th>
<th>( d = 4 )</th>
<th>( d = 5 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>LFDR model (2) with lagged covariates</td>
<td>1.371</td>
<td>1.357</td>
<td>1.337</td>
<td>1.335</td>
</tr>
</tbody>
</table>

A.2.2 Validation of residual assumption and model selection

The left panel of Figure A.5 validates the homoscedasticity of residual variance for model (10); the residuals are obtained as the difference between the observed and the predicted values. Notice residuals vary randomly around zero and there is no systematic pattern. The middle and right panels of Figure A.5 investigate the normality assumption of the residuals via Q-Q plot.
Figure A.5: Diagnostics of the residual assumption of model (10); scatter plot of the residuals against the fitted values (left), histogram of the residuals (middle), and normal Q-Q plot (right).
Next we investigate the covariance assumption between random effects. We assumed independent random effects in our data application. The idea was to use a covariance model that accounts for the dependence between the repeated measurements of the response, and this choice seemed sufficiently simple, while illustrating the point; plus we were able to adjust the competitive approaches to fit such model. However such choice may be restrictive in some situations and one should consider correlated random effects induced covariance models.

We fitted both a model with independent group intercept and slope, and the one with dependent group intercept and slope. The results are displayed in Table A.11 below. They do not provide convincing evidence that a model with dependent group intercept and slope would be better.

<table>
<thead>
<tr>
<th>Assumption</th>
<th>INPE (a)</th>
<th>INPE (b)</th>
<th>OUTPE (a)</th>
<th>OUTPE (b)</th>
<th>AIC (a)</th>
<th>AIC (b)</th>
<th>BIC (a)</th>
<th>BIC (b)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \text{cov}(b_{0i}, b_{1i}) = 0 )</td>
<td>1.22 (0.01)</td>
<td>1.32 (0.01)</td>
<td>1.69 (0.02)</td>
<td>1.43 (0.03)</td>
<td>19344.42 (56.11)</td>
<td>23942.34 (90.71)</td>
<td>19506.73 (56.11)</td>
<td>24105.27 (90.71)</td>
</tr>
<tr>
<td>( \text{cov}(b_{0i}, b_{1i}) \neq 0 )</td>
<td>1.22 (0.01)</td>
<td>1.31 (0.01)</td>
<td>1.69 (0.02)</td>
<td>1.43 (0.03)</td>
<td>19349.57 (59.76)</td>
<td>23938.21 (85.02)</td>
<td>19514.59 (59.76)</td>
<td>24107.93 (85.02)</td>
</tr>
</tbody>
</table>

Table A.11: Model comparison. Reported are the median prediction accuracy, marginal AIC, and BIC for model 10 with independent and dependent covariance assumption of random intercept and slope; IQR are reported in parenthesis.

A.2.3 Additional results for the model presented in the main article

In this section, we briefly describe the analysis of the noisy temperature profiles. We estimate the bivariate mean function of temperature using the approach described in Xiao et al. (2013); see Figure A.6. Next we model the profiles using the approach introduced by Park & Staicu (2015) using the marginal covariance induced by these data. Using a percentage of explained variance equal to 95%, we estimate \( K = 7 \); the estimated directions are shown in Figure A.7 along with the corresponding explained percentage of variance.
Figure A.6: Daily mean temperature $\tilde{\tau}(\cdot, \cdot)$ during the lactations days.
Figure A.7: Estimated basis functions $\hat{\phi}_k(\cdot)$ corresponding to functional PCs.

Figure A.8 exhibits the observed temperature and the corresponding smoothed profiles for randomly chosen two sows. Figure A.9 illustrates the estimated full trajectory of time-varying basis coefficients $\hat{\xi}_{ik}(\cdot)$ for $k = 1, \ldots, 5$. 
Figure A.8: Observed (left) and smooth (right) temperature profiles for a randomly chosen sow.

Figure A.9: Full trajectory of time-varying basis coefficients corresponding to first (left of first row), second (right of first row), third (left of second row), fourth (right of second row), fifth (left of third row), sixth (right of third row), and seventh (bottom row) functional PCs.
We fit the longitudinal dynamic functional regression model (LDFR) as described in the main paper. We evaluate the percentage of variation that is explained by the various components. Recall that $i$ indexes the sow and $j$ the repeated measurements per sow. Let $n_j$ be the number of sows that are observed at time $t_j$. Let $\int \hat{Temp}_{ij}(s)\hat{\gamma}(s,t_j)ds = \sum_{k=1}^{7} \hat{\xi}_{ik}(t_j)\hat{\beta}_k(t_j)$ be the functional effect at $t_j$ and $\overline{T_M}(t_j) = \sum_{i=1}^{n_j} (\sum_{k=1}^{7} \hat{\xi}_{ik}(t_j)\hat{\beta}_k(t_j))/n_j$ be the mean effect at $t_j$. Write $\overline{Y}_j = 1/n_j \sum_{i=1}^{n_j} Y_{ij}$ and $\hat{\overline{Y}}_j = 1/n_j \sum_{i=1}^{n_j} \hat{Y}_{ij}$. Define the mean parity effect at $t_j$ by $\hat{\beta}(t_j) = (\hat{\beta}_{p_i=0}(t_j) + \hat{\beta}_{p_i=1}(t_j))/2$, where $p_i = 0$ (young sows) and $p_i = 1$ (old sows) respectively. Find at each $t_j$,

\[
\hat{\text{Var}}(Y_j) = \sum_{i=1}^{n_j} (Y_{ij} - \overline{Y}_j)^2/(n_j - 1),
\]
\[
\hat{\text{Var}}(E_j) = \sum_{i=1}^{n_j} (Y_{ij} - \hat{\overline{Y}}_{ij})^2/(n_j - 1),
\]
\[
\hat{\text{Var}}(M_j) = \sum_{i=1}^{n_j} (\overline{Y}_{ij} - \hat{\overline{Y}}_j)^2/(n_j - 1),
\]
\[
\hat{\text{Var}}(P_j) = (\hat{\beta}_{old}(t_j) - \hat{\beta}(t_j))^2 + (\hat{\beta}_{yng}(t_j) - \hat{\beta}(t_j))^2,
\]
\[
\hat{\text{Var}}(T_j) = \sum_{i=1}^{n_j} \left( \int \hat{Temp}_{ij}(s)\hat{\gamma}(s,t_j)ds - \overline{T_M}(t_j) \right)^2/(n_j - 1).
\]

Find the corresponding sum-of-squares as follows

\[
SS_{\text{resp}} = (1/m) \sum_{j=1}^{m} \hat{\text{Var}}(Y_j),
\]
\[
SS_{\text{error}} = (1/m) \sum_{j=1}^{m} \hat{\text{Var}}(E_j),
\]
\[
SS_{\text{model}} = (1/m) \sum_{j=1}^{m} \hat{\text{Var}}(M_j),
\]
\[
SS_{\text{par}} = (1/m) \sum_{j=1}^{m} \hat{\text{Var}}(P_j),
\]
\[
SS_{\text{cov}} = (1/m) \sum_{j=1}^{m} \hat{\text{Var}}(T_j).
\]

Table A.12 illustrates the contribution of model components in explaining the variation.
Table A.12: Relative importance of model components.

<table>
<thead>
<tr>
<th>SS_{resp}</th>
<th>SS_{error}</th>
<th>SS_{model}</th>
<th>SS_{par}</th>
<th>SS_{cov}</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.92</td>
<td>1.73</td>
<td>0.97</td>
<td>0.63</td>
<td>0.18</td>
</tr>
</tbody>
</table>

In particular, the percentage of variation in the data that is accounted for by the model is calculated as $0.97/2.92 \approx 33\%$ and the percentage of additive model that is accounted for by the functional par is $0.18/0.97 \approx 19\%$, while the percentage of the additive part that is accounted by for the parity factor is $0.63/0.97 \approx 65\%$. Thus it indicates that the group most important part of the variation in the additive model is due to the parity.

Figure A.10 shows the boxplots for the random group specific term, the random subject-specific intercept and the subject-specific slope.

Figure A.10: Prediction of sow-group specific random coefficients.
Figure A.11 illustrates the comparison of the predicted responses using our method with the existing approaches, LPEER and LPFR, for the sows of Figure 4 in the manuscript. Note that LDFR predicts the full response trajectory, while LPEER, as LPFR, predicts the response only for the times at which the sow was observed.
Figure A.11: Predicted responses with LDFR (red), LPEER (blue), and LPFR (purple) for two young (left and middle) and one old (right) sows; the actual observations are given in black. Shaded regions correspond to 95% pointwise prediction bands based on LDFR.
A.2.4 Fitting simpler model

In the main paper the alternative approach, LPEER is based on a random intercept-based covariance model, as the available R function `lpeer()` only accommodates random intercepts and it is not clear how it can be modified to account for a random slope. For a fair comparison, we also fit LPFR and LDFR using the same covariance structure, using random subject intercept and random group intercept.

Table A.13: Comparison of prediction accuracy and run time for the model with subject-specific random intercept and random block effect.

<table>
<thead>
<tr>
<th></th>
<th>IN_{PE}</th>
<th></th>
<th>OUT_{PE}</th>
<th></th>
<th>Computing time</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>LDPR</td>
<td>LPEER</td>
<td>LPFR</td>
<td></td>
<td>LDPR</td>
</tr>
<tr>
<td>(a)</td>
<td>1.26 (0.01)</td>
<td>1.32 (0.01)</td>
<td>1.40 (0.01)</td>
<td></td>
<td>1.61 (0.01)</td>
</tr>
<tr>
<td>(b)</td>
<td>1.35 (0.01)</td>
<td>1.39 (0.01)</td>
<td>1.45 (0.01)</td>
<td></td>
<td>1.44 (0.03)</td>
</tr>
</tbody>
</table>

A.2.5 Simulated data

Figure A.12 exhibits the simulated response for a randomly chosen subject while data are generated mimicking data application.

![Figure A.12](image)

Figure A.12: Illustration of simulated response for a randomly chosen subject.
A.3 Technical details

In this section, we detail the procedure of finding the variance of $\hat{Y}_i$. Let $\hat{\theta} = \{\hat{\phi}_k(\cdot), K, \hat{\psi}_{kl}(\cdot), L, L_k\}$ be the collection of functional principal decomposition objects. Define $Y_i(t) = \alpha(t) + \sum_{k=1}^2 \beta_k(t)\xi_{ik}(t) + Z_{b,it}b_i + \epsilon_{it}$. Let $t$ be some generic time in $T$.

Recall from section 3.2 that the model for Gaussian response $Y$ can be written, in matrix form, as $\mu = \hat{\nu} + \hat{\theta} + \epsilon$, where all terms are defined appropriately as before. Define the overall design matrix by stacking up the design matrices corresponding to fixed, $\beta$ and random effects, $u$ and $b$. Let $C$ be a such $\sum_i m_i \times ((p+1)(K+1)+L(K+1)+I_q)$ design matrix; where $C = (\hat{V}|\hat{Z}|Z_b)$. Define by $\beta^* = (\beta^T, u^T, b^T)^T$ the vector of unknown coefficients. Let $D_{\lambda,b}$ be a block diagonal matrix such that $D_{\lambda,b} = \text{diag}(\sigma^2 e \lambda_0, \sigma^2 e \lambda, \ldots, \sigma^2 e \lambda) \otimes I_L, I \otimes \sigma^2 D^{-1}$. Following Ruppert et al. (2003) and Wood (2017, 2006b), we minimize (6) and estimate

$$\hat{\beta}^* = (C^T C + D_{\lambda,b})^{-1} C^T Y,$$

and compute the bias adjusted variability for $\hat{\beta}^*$ as

$$\text{Var}(\hat{\beta}^*) = (C^T C + D_{\lambda,b})^{-1} \sigma_e^2,$$

which is equivalent to the corresponding Bayesian posterior covariance matrix developed by Wahba (1983). Let $C_{it}$ be the covariates information for $i$ at some time $t$; it yields that

$$\text{Var}\{\hat{Y}_i(t)\} = \text{Var}\{C_{it}\hat{\beta}^*\} = C_{it}\text{Var}(\hat{\beta}^*)C_{it}^T.$$

In particular, $\text{gam}$ function in mgcv (Wood, 2001) package computes the bias adjusted variability for the parameter estimates $\hat{\beta}^*$. Alternatively, one can use $\text{lme}$ function in nlme (Pinheiro et al., 2009) or $\text{lme4}$ function in lme4 (Bates et al., 2014) package to obtain $\text{Var}(\hat{\beta}^*)$. However, for large number of random effects, $\text{gam}$ performs slower relative to the other alternatives due to the fact of not exploiting the sparsity feature of parametric random effects (Wood, 2012a).
Appendix B

Additional details for Chapter 3

This Supplementary Material contains three sections. Section B.1 validates the underlying covariance assumption for the response-adjusted latent process. Section B.2 provides additional simulation results for the settings \( G2 \) and \( G4 \). Section B.3 provides details for the mixed model representation using natural cubic splines.

B.1 Validation of covariance assumptions

Recall that \( S \) and \( T \) are the domains in the \( s \)-direction and \( t \)-direction respectively. We assume that the response-adjusted functional measurements \( V_i(s,t) \) is a realization of a true random process \( V(s,t) \) with mean \( \tau(s,t) \) and smooth covariance function \( c((s,t),(s',t')) = E[(V(s,t) - \tau(s,t))(V(s',t') - \tau(s',t'))^T] \) that satisfies the following regularity conditions:

(A1) \( V = \{V(s,t) : (s,t) \in S \times T\} \) is a square integrable element defined in \( L^2(S \times T) \) such that \( E[\int \int V^2(s,t)dsdt] < \infty \), where \( S = [0,1] \) and \( T = [0,1] \) are closed compact sets.

(A2) The sampling density of \( g(t) \) is continuous and \( \sup_{t \in T} |g(t)| < \infty \).

Corollary 1

Under (A1) and (A2), the covariance function \( \Sigma_\theta(s,s') = \int c((s,t),(s',t'))dt \) is (i) symmetric, (ii) positive-definite, and (iii) has eigenvalues satisfying that \( \sum_{k \geq 1} \eta_k < \infty \).
Proof:

(i) $\Sigma_v(s, s')$ is symmetric.

\[
\Sigma_v(s, s') = \int c((s, t), (s', t))g(t)dt = \int E\left[\{V(s, t) - \tau(s, t)\}\{V(s', t) - \tau(s', t)\}^T\right]g(t)dt \\
= \int E\left[\{V(s', t) - \tau(s', t)\}\{V(s, t) - \tau(s, t)\}\right]g(t)dt = \int c((s', t), (s, t))g(t)dt \\
= \Sigma_v(s', s).
\]

(ii) $\Sigma_v(s, s')$ is positive definite.

If $\Sigma_v(s, s')$ is positive definite, then the following must hold.

\[
\int \int \Sigma_v(s, s')z(s)z(s')dsds' = \int \int \left\{\int c((s, t), (s', t))g(t)dt\right\}z(s)z(s')dsds' \\
= \int g(t) \left\{\int \int c((s, t), (s', t))z(s)z(s')dsds'\right\}dt \\
= (By \text{ Fubini’s Theorem}) \\
\geq 0.
\]

Therefore it is sufficient to show that $\int \int c((s, t), (s', t))z(s)z(s')dsds' \geq 0$ for any $z(\cdot) \in L^2(\mathcal{S})$; this further implies that $V(s, t^*)$ is square integrable for any $t^* \in \mathcal{T}$. We prove this by contradiction. It is to say that $\int \int V^2(s, t)dsdt = \infty$. Let $Z(t) = \int V^2(s, t)ds$ and assume $E[Z(t)] = \infty$ for any $t \in \mathcal{T}_0 \subset \mathcal{T}$ such that $\int_{\mathcal{T}_0} g(t)dt > 0$; this further yields $\int_{\mathcal{T}} Z(t)dt > \int_{\mathcal{T}_0} Z(t)dt$. Since expectation is a linear operator, we argue that

\[
E\left[\int_{\mathcal{T}} Z(t)dt\right] > E\left[\int_{\mathcal{T}_0} Z(t)dt\right].
\]

This implies $E\left[\int_{\mathcal{T}} Z(t)dt\right] = \infty$ since we assume that $E[Z(t)] = \infty$ for any $t \in \mathcal{T}_0$. But under assumption (A1), this cannot be true as it says that $E\left[\int_{\mathcal{T}} \int_S V^2(s, t)dsdt\right] < \infty$. Thus by contradiction, we show that $E\left[\int_{\mathcal{T}} Z(t)dt\right]$ is finite and $V(s, t^*)$ is square integrable for fixed $t^* \in \mathcal{T}$. This further implies that $\Sigma_v(s, s')$ is positive definite.

(iii) $\int \Sigma_v(s, s')ds < \infty$. 

148
If $\int \Sigma_v(s, s') ds < \infty$ is true then the sum of the eigenvalues, $\eta_k$'s of $\Sigma_v(s, s')$ satisfy $\sum_k \eta_k < \infty$.

$$\int \Sigma_v(s, s') ds = \int \left\{ \int c((s, t), (s', t)) g(t) dt \right\} ds$$

$$= \int g(t) \left\{ \int c((s, t), (s', t)) ds \right\} dt \quad \text{(By Fubini’s Theorem)}$$

$$\leq \sup_{t \in T} g(t) \left\{ \int \int c((s, t), (s', t)) ds dt \right\}$$

$$< \infty$$

Therefore, by Horváth & Kokoszka (2012a), we deduce that $\Sigma_v(s, s')$ is a proper covariance function.

### B.2 Additional simulation results

Table B.1 presents the empirical type-I error rates for testing the nullity and time-invariant effect simultaneously ($F3$). As before, the method performs satisfactorily in maintaining the correct nominal levels across different simulation settings. In the situation, when the scores are estimated directly based on the marginal covariance function of the functional predictor ($\xi_W$), we observe moderate undercoverage. This is due to the fact of assuming implicitly that the leading $K$ components are the most relevant for predicting the responses and thus fails to account for the true correlation between the functional predictor and responses. In contrast, the model based on the scores $\xi_V$ that are calculated using the response-adjusted functional measurements performs satisfactorily in attaining the nominal levels across different simulation settings; compare the columns at any $\alpha$ in different blocks in Table B.1. In addition, the numerical performances are comparable with the oracle testing procedure based on $\xi_T$.

Figure B.1 plots the power rates for simultaneous testing of time-invariant and null effect ($F3$). As before, the test with $\xi_V$ produces good power properties while the testing procedure with $\xi_W$ remains inferior. Indeed the difference between the power curves associated with the oracle scores $\xi_T$ and estimated scores $\xi_V$ decreases for larger $\delta$; compare the power probabilities at $\delta = 0.20$ and $\delta = 2$ for any setting in Figure B.1. In addition, the numerical performance improves with larger size and more information per subject.
Table B.1: Testing for nullity and time-invariant effect ($F3$) simultaneously for the case $H1$. The leading $K$ FPCs explain at least 90% variation of marginal covariance function. Reported are the observed type-I errors and the corresponding standard errors (in parenthesis) based on 5000 Monte Carlo simulations. The functional covariates are observed as in $A1$ and $A2$ with low ($B1$) SNR$_w$; the smooth functional coefficients are represented as in $G3$. Responses are generated assuming normality ($C$) and with CS type covariance structure ($D$). Both moderate ($E1$) and low ($E2$) sparse sampling designs are considered.

<table>
<thead>
<tr>
<th>$n$</th>
<th>sparsity $m_i$</th>
<th>SNR$_y$</th>
<th>$\xi_T$ level of significance $\alpha$</th>
<th>$\xi_W$ level of significance $\alpha$</th>
<th>$\xi_V$ level of significance $\alpha$</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>mod 2</td>
<td>0.008 0.017 0.055 0.108</td>
<td>(0.001) (0.002) (0.003) (0.004)</td>
<td>0.007 0.019 0.045 0.094</td>
<td>(0.001) (0.002) (0.003) (0.004)</td>
</tr>
<tr>
<td></td>
<td>low 2</td>
<td>0.009 0.014 0.052 0.159</td>
<td>(0.001) (0.002) (0.003) (0.004)</td>
<td>0.008 0.019 0.040 0.080</td>
<td>(0.001) (0.002) (0.003) (0.004)</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>0.010 0.016 0.048 0.104</td>
<td>(0.001) (0.002) (0.003) (0.004)</td>
<td>0.008 0.019 0.044 0.086</td>
<td>(0.001) (0.002) (0.003) (0.004)</td>
</tr>
<tr>
<td>200</td>
<td>mod 2</td>
<td>0.011 0.020 0.051 0.110</td>
<td>(0.002) (0.002) (0.003) (0.004)</td>
<td>0.012 0.021 0.047 0.095</td>
<td>(0.002) (0.002) (0.003) (0.004)</td>
</tr>
<tr>
<td></td>
<td>low 2</td>
<td>0.009 0.018 0.054 0.108</td>
<td>(0.001) (0.002) (0.003) (0.004)</td>
<td>0.008 0.021 0.050 0.103</td>
<td>(0.001) (0.002) (0.003) (0.004)</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>0.012 0.020 0.054 0.108</td>
<td>(0.002) (0.002) (0.003) (0.004)</td>
<td>0.008 0.015 0.034 0.081</td>
<td>(0.001) (0.002) (0.003) (0.004)</td>
</tr>
</tbody>
</table>

150
Figure B.1: Observed power function for simultaneously testing the adequacy of time-varying functional effect and nullity ($F_4$) for low (top-left-to-right) and high (bottom-left-to-right) SNR$_y$. Reported are the proportion of rejections based on 1000 simulations for $\alpha = 0.05$ at $\delta = \{0.05, 0.20, 0.50, 1.00, 2.00\}$. The functional covariates are observed as in A1 and A2 with low SNR$_w$ (B1) while the smooth effects are represented by $G_3$.

B.3 Mixed model representation

In general, truncated polynomial splines has the sub-optimal computational advantage to model smooth effects due to disregarding the underlying correlation between the basis functions. A more popular approach to model unknown smooth functions is using natural cubic splines or B-splines basis functions. With some cumbersome transformation, B-splines with second-order penalty can also be used to cast the model into a mixed model framework as discussed in the main manuscript Section 2.4. The main idea is to exploit the relationship between natural cubic splines/B-splines and truncated polynomial functions (TPF) following De Boor et al. (1978); Dierckx (1995). In our implementation, we consider correlation between the random effects associated with the spline terms and transform them to independence via Cholesky decomposition. We refer to Ruppert et al. (2003); Wood (2017); Verbyla (1999); Pinheiro & Bates (2000) for details. For instance, we represent each smooth coefficient by

$$\beta_k(t) \approx \beta_{0k} + \beta_{1kt} + \sum_{r=1}^{R_k} u_{rk} P_j(t).$$ (B.1)
Let $\beta_k = (\beta_{0k}, \beta_{1k})^T$ be a vector of fixed and $u_k = (u_{1k}, \ldots, u_{R_k})^T$ be a vector of random parameters such that $u_k \sim \mathcal{N}(0, \sigma_k^2 G_s)$, where, $G_s$ is a symmetric positive-definite matrix and $P_j(\cdot)$ is the set of basis functions which can be obtained by scaling the truncated power functions following Verbyla (1999); Pinheiro & Bates (2000). Let $X_s = (1, t)$ and $P_s = (P_1(t), \ldots, P_{R_k}(t))$. In matrix form, write (B.1) as

$$\beta_k(t) \approx X_s \beta_k + P_s u_k.$$

Apply Cholesky decomposition on $G_s$ such that $G_s = LL^T$ where $L$ is the lower triangular matrix. Transform the set of random effects $u_k$ by $v_k = Lu_k$ where $v_k \sim \mathcal{N}(0_{R_k}, \sigma_k^2 I_{R_k})$. The design matrix is transformed appropriately by $P = P_s L$ and define $\beta_k(t) \approx X_s \beta_k + P v_k$ which is similar to the approximating mixed model representation defined in Section 3.2.2 in the main manuscript for $p = 1$. 

152
Appendix C

Additional details for Chapter 4

This Supplementary Material consists of five sections. Section C.1 provides details for the fit without sparse penalty. Section C.2 presents additional analysis results for the kinematic data. Section C.3 and Section C.4 provide additional simulation results. Section C.5 details the algorithm of the post-selection predictive inference.

C.1 Generalized additive model

Let the observed data, in general, be \([y_i, z_i, \{X_{k,i}(s_r); r = 1, \ldots, R\}, k = 1, \ldots, K, i = 1, \ldots, N]\), where all terms bear the usual meaning as before and are described in the manuscript. Write the penalized criterion for the model (4.3) with the smoothing penalty only in matrix form as

\[
(Y - \tilde{X}\beta)^T (Y - \tilde{X}\beta) + \phi_1\beta^T \Omega_s \beta + \phi_2\beta^T \Omega_z \beta;
\]

where \(Y\) is a \(N \times 1\) vector with elements \(y_i\)'s, \(\tilde{X} = (\tilde{X}_1|\cdots|\tilde{X}_K)\), \(\tilde{X}_k\) is an \(N \times LM\) matrix having the \(i^{th}\) row \(\tilde{X}_{k,i}^T\), \(\beta = (\beta_1^T, \ldots, \beta_K^T)^T\) is a vector of \(LM\) parameters, \(\Omega_s = \text{diag}\{\Omega_{s1} \otimes I_M, \ldots, \Omega_{sK} \otimes I_M\}\), and \(\Omega_z = \text{diag}\{I_L \otimes \Omega_{z1}, \ldots, I_L \otimes \Omega_{zK}\}\). Following Ruppert et al. (2003); Wood (2017), taking the derivative with respect to \(\beta\) yields \(\tilde{\beta} = (\tilde{X}^T \tilde{X} + \phi_1 \Omega_s + \phi_2 \Omega_z)^{-1} \tilde{X}^T Y\); we obtain the initial estimates of the regression coefficients as \(\tilde{\beta}_{k0}(\cdot, \cdot)\)'s for a given \(\phi_{sz}\). The corresponding Bayesian posterior covariance matrix is \(\text{Var}(\tilde{\beta}) = (\tilde{X}^T \tilde{X} + \phi_1 \Omega_s + \phi_2 \Omega_z)^{-1}\tilde{\sigma}^2\), where \(\tilde{\sigma}^2\) is estimated from the residual sum of squares. Predict the responses as \(\tilde{Y} = \tilde{X} \tilde{\beta}\).

The \(\phi_1\) and \(\phi_2\) are unknown in practice; one approach to select the optimal tuning parameters is block CV. The post-selection model (4.7) is fitted using the above intuition but with the reduced subset \(K_{\lambda^*, \phi^*}^{(2)}\).

153
C.2 Additional results of EMG selection for movement data

In this section, we present more results of the analysis of the EMG and movement data.

C.2.1 Data restructuring of EMG signals

Our idea is to characterize the velocity at an instance \( i \) as a function of the recent past EMG signals. Figure C.1 displays visually the data reconstruction of functional predictors using the recent past EMG signals.

![Figure C.1: Visualization of data restructuring of EMG signals and joint finger angles at 0.9 second and 2.3 second. Time at current position (blue dot on black line) used to extract concurrent and previous \( \delta \) values of the two EMG signals, shown in red and green in (a). The time domain for each set of past \( \delta + 1 \) EMG signal measurements is rescaled to \([-\delta, 0]\). Reconstructed EMG curves are plotted in gray lines on the top-right and bottom-right panels, and two curves corresponding to 0.9 second and 2.3 second are highlighted in (b) and (c) for EMG-7 (green solid line) and EMG-12 (red solid line) respectively.](image)

C.2.2 Variation in EMG curves

We use functional principal component analysis technique to examine the main sources of variability in the curves. Figure C.2 illustrates the EMG curves \( X_{k,i} (\cdot) \)’s associated with a
muscle (*flexor digitorum*) in the forearm that contributes to finger flexion. The first three functional principal components (FPCs) of the estimated marginal covariance of $X_{k,i}(\cdot)$’s are also plotted. We observe one key feature that solely explains the majority of variation in the curves associated with the EMG signal; see the first FPC in Figure C.2.

![Figure C.2](image)

Figure C.2: Restructured EMG curves corresponding to the *extensor digitorum* muscle and the first three eigenfunctions from left to right.

### C.2.3 Variable selection for wrist movements

Table C.1 shows the results of the model selection performance for the wrist movements with consistent and varying patterns. In most cases, the competing methods select one muscle for wrist extension and another for flexion. As described in the manuscript, there are many potential muscles that contribute to similar wrist movements. Therefore selection between the alike muscles is desired to reduce the redundancy of EMG information. Notice both agLASSO and SAFE-gLASSO attain optimal RSPs. In contrast FAR-gSCAD shows suboptimal RSPs and LAD-gLASSO performs poorly in selecting desired variables; follow the column “RSP” for competing methods in Table C.1. While both these methods have tendency to select noise variables in addition to the true positives, the problem in the selection by LAD-gLASSO is severe; follow the column “FPR” in Table C.1.
Table C.1: Wrist movements EMG signal selection. RSPs, model size (square brackets), and the percentages of TPRs and FPRs are presented for consistent (top three rows) and random (bottom three row) patterns at different postures. Results with superscript † correspond to the first stage of SAFE-gLASSO.

<table>
<thead>
<tr>
<th>Pattern</th>
<th>agLASSO</th>
<th>LAD-gLASSO</th>
<th>FAR-gSCAD</th>
<th>SAFE-gLASSO</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RSP</td>
<td>TPR</td>
<td>FPR</td>
<td>RSP</td>
</tr>
<tr>
<td>1.00</td>
<td>100</td>
<td>0</td>
<td>0.57</td>
<td>100</td>
</tr>
<tr>
<td>[2]</td>
<td>[2]</td>
<td>[0]</td>
<td>[6]</td>
<td>[2]</td>
</tr>
<tr>
<td>1.00</td>
<td>100</td>
<td>0</td>
<td>0.50</td>
<td>100</td>
</tr>
<tr>
<td>[2]</td>
<td>[2]</td>
<td>[0]</td>
<td>[9]</td>
<td>[5]</td>
</tr>
<tr>
<td>1.00</td>
<td>100</td>
<td>0</td>
<td>0.57</td>
<td>100</td>
</tr>
<tr>
<td>[2]</td>
<td>[2]</td>
<td>[0]</td>
<td>[8]</td>
<td>[3]</td>
</tr>
<tr>
<td>1.00</td>
<td>100</td>
<td>0</td>
<td>0.72</td>
<td>100</td>
</tr>
<tr>
<td>[2]</td>
<td>[2]</td>
<td>[0]</td>
<td>[6]</td>
<td>[4]</td>
</tr>
<tr>
<td>1.00</td>
<td>100</td>
<td>0</td>
<td>0.50</td>
<td>100</td>
</tr>
<tr>
<td>[2]</td>
<td>[2]</td>
<td>[0]</td>
<td>[9]</td>
<td>[6]</td>
</tr>
</tbody>
</table>

C.2.4 Residual diagnostics

Reported are the results based on a varying finger movement data. We also carry out the residual diagnostics to check the statistical assumptions for other data sets in a similar fashion, and hence we omit it here.

The left panel of Figure C.3 checks the constant error variance assumption for the model (4.1). Let the residual at an instance \( i \) be \( r_i = y_i - \hat{y}_i \); where \( y_i \) and \( \hat{y}_i \) are the observed and predicted responses, respectively. Notice the residuals “bounce randomly” around zero and there is no systematic structure. In addition, no one residual “stands out” from the underlying random pattern implying the absence of outliers. The middle and right panels of Figure C.3 assess the normality assumption of the model via the density curve and Q-Q plot based on the residuals.

Figure C.4 illustrates the underlying dependence structure in the data. We use the auto-correlation function (ACF) and the partial ACF (PACF) to measure the strength of correlation between the successive measurements. We consider the mean model (4.7) and examine the auto-regressive patterns of the residuals. We do not observe any significant autocorrelation in Figure C.4.
C.2.5 Regression surfaces for finger movements

In this section, we interpret the regression surfaces corresponding to the finger movements with consistent and varying patterns; Figure C.5 illustrates the corresponding postures. Reported are the estimates based on the second stage of SAFE-gLASSO.
Figure C.5: Postures for finger/wrist extension and flexion.

- Figure C.6 corresponds to the neutral posture demonstrated in the left panel of Figure C.5. SAFE-gLASSO selects two muscles: extensor digitorum (muscle 5) and flexor digitorum (muscle 12). The interpretation of the extensor digitorum’s (right panel) activity on finger extension is as same as before; the impact is most relevant between $-30$ and $20$ radians which are primarily associated with finger opening. On the other hand, the flexor digitorum is active for the positions $20$ and $40$ radians which correspond to making a fist. As before, the muscle mechanism at the past time window depicts the opposite features to that of the concurrent relationship and specifically describes the involuntary movements due to the passive forces when finger is fully flexed (left) and extended (right).

- Figure C.7 corresponds to the posture demonstrated in the right panel of Figure C.5. Both stages of the SAFE-gLASSO select two muscles: one is extensor digitorum (muscle 5) and the other is flexor digitorum (muscle 12). The impact of the left surface on velocity is most important for positions between $10$ and $60$ radians which correspond to finger closing. Similarly, the impact of the right surface is most relevant between $-30$ and $20$ radians which are primarily associated with finger opening. The corresponding historical relationship follows the same intuition as before.

- Figure C.8, Figure C.9, and Figure C.10 correspond to the postures demonstrated in the left, middle, and right panel of Figure C.5, respectively. The interpretation of the corresponding regression surfaces follows the same idea as before. Notice that the smooth estimates in the left surfaces for Figure C.8 and Figure C.9 are not zero. Here one important specific is that the magnitude of the effect of flexor digitorum is considerably smaller than than that of the extensor digitorum. This happens due to the fact that user focuses primarily on opening the finger at varying pattern and does not make the fist properly. This is not entirely surprising since the pattern at these postures is not directive by the experimenter but solely determined by the user.
Figure C.6: Regression surface for finger flexion (left) and extension (right). Reported are the estimates corresponding to consistent finger movements with neutral posture depicted in the left panel of Figure C.5.

Figure C.7: Regression surfaces for finger flexion (left) and extension (right). Reported are the estimates corresponding to consistent finger movements with posture depicted in the right panel of Figure C.5.
Figure C.8: Regression surfaces for finger flexion (left) and extension (right). Reported are the estimates corresponding to varying finger movements with neutral posture depicted in the left panel of Figure C.5.

Figure C.9: Regression surface for finger flexion (left) and extension (right). Reported are the estimates corresponding to varying finger movements with posture depicted in the middle panel of Figure C.5.
C.2.6 Regression surfaces for wrist movements

In this section, we plot the regression surfaces corresponding to wrist movements based on the stage of SAFE-gLASSO.

- Figure C.11 corresponds to the neutral posture demonstrated in the left panel of Figure C.5. The second stage of SAFE-gLASSO select *flexor carpi ulnaris*, *flexor digitorum superficialis*, and *extensor carpi radialis longus*. While the first two muscles contribute for wrist flexion, the last one leads to extension. Since the regression surfaces associated with both *flexor carpi ulnaris* and *flexor digitorum superficialis* are similar, we report and interpret the effect of *flexor carpi ulnaris* only. The impact of this muscle on velocity is most important for positions between 5 and 50 radians. We also notice some activity of this muscle when the wrist is in the neutral state, say around −20 and 0 radians. This says that the *flexor carpi ulnaris* needs to be flexed to hold the wrist in the upright position. In contrast, the impact of the *extensor carpi radialis longus* is most relevant between positions −50 and −10 radians which are primarily associated with wrist extension. The corresponding muscles have opposite historical relationship which in particular quantifies the movements due to the physical constraints and passive forces.
Figure C.12 corresponds to the posture demonstrated in the right panel of Figure C.5. Unlike previous case, the impact of the flexor carpi ulnaris on velocity is most important for angles between 10 and 50 radians which correspond to wrist flexion. In contrast, the impact of the extensor carpi ulnaris (right surface) is most relevant between positions $-50$ and $-20$ radians which are primarily associated with wrist extension.

Figure C.13, Figure C.14, and Figure C.15 correspond to the postures demonstrated in the left, middle, and right panel of Figure C.5. The interpretation of the muscle mechanism follows the same intuition as before and hence we omit it here.

Figure C.11: Regression surfaces for wrist flexion (left) and extension (right). Reported are the estimates corresponding to consistent wrist movements with neutral posture depicted in the left panel of Figure C.5.
Figure C.12: Regression surfaces for wrist flexion (left) and extension (right). Reported are the estimates corresponding to consistent wrist movements with posture depicted in the right panel of Figure C.5.

Figure C.13: Regression surfaces for wrist flexion (left) and extension (right). Reported are the estimates corresponding to varying wrist movements with neutral posture depicted in the left panel of Figure C.5.
Figure C.14: Regression surfaces for wrist flexion (left) and extension (right). Reported are the estimates corresponding to varying wrist movements with posture depicted in the middle panel of Figure C.5.

Figure C.15: Regression surfaces for wrist flexion (left) and extension (right). Reported are the estimates corresponding to varying wrist movements with posture depicted in the right panel of Figure C.5.
C.2.7 Regression surfaces based on competing approaches

Figure C.16 plots the regression surfaces based on the competing approaches for the finger movements with fixed pattern (i.e., second row in Table 1). We notice that the effect of the selected EMG signals on velocity is linear in both agLASSO and LAD-gLASSO. Concurrently, say at \( s = 0 \), the effect of the *flexor digitorum* is negative to velocity while the effect of the *extensor digitorum* is positive. Historically, say at \( s = -40 \), they imply the opposite effects which matches with the surface interpreted by SAFE-gLASSO. While agLASSO and SAFE-gLASSO account for passive forces, they fail to acknowledge the changes in the regression surfaces across different finger positions. In particular, they implicitly assume that the muscle mechanism is same irrespective of the finger opening/closing positions which is indeed unrealistic. As pointed out in James *et al.* (2009), the smooth estimates based on FAR-gSCAD is difficult to interpret due to the issue of identifiability; and hence we omit it here.

![Figure C.16: Estimates of smooth coefficients for the selected muscles: flexor digitorum (left) and extensor digitorum (right) corresponding to agLASSO (top panels) and LAD-gLASSO (bottom panels).](image)

Figure C.16: Estimates of smooth coefficients for the selected muscles: *flexor digitorum* (left) and *extensor digitorum* (right) corresponding to agLASSO (top panels) and LAD-gLASSO (bottom panels).
Figure C.17: Simulated velocity across different simulation settings at SNR = \{875, 80, 8, 4, 0.8\}. Data correspond to non-varying finger movements.

C.3 Additional simulation results for the data mimicking application

Figure C.17 depicts the simulated velocity for finger movements at different scenario. As we increase the level of noise, we depart more from the truth; compare the top-left and bottom-right panels.

Table C.2 illustrates the numerical performance corresponding to the simulated kinematic data for different SNRs at low correlation coefficient. The results are consistent to the findings of finger movements in Section 5 and Section 6. The proposed approach has lower or equal FPRs than that of the competitors across different SNRs. However, at SNR = 0.8, the numerical performance of the competitors deteriorates. As before, the model size for the second stage of SAFE-gLASSO is smaller than that of the first stage; see the column “SP” for the proposed at SNR = 0.8. In addition, the proposed method outperforms the competitors in terms of prediction accuracy at all SNRs; see Figure C.18.
Table C.2: Analysis of finger movements with fixed motion. Data is generated assuming noise variance $C1$ and $C2$ with different dominant processes ($A1$-$A3$) for low correlation coefficient ($B1$). Reported are the SPs (%), model size (in square brackets), TPRs (%), and FPRs (%) averaged over 100 simulations. Results with superscript † correspond to the first stage of SAFE-gLASSO.

<table>
<thead>
<tr>
<th>Setting</th>
<th>SNR</th>
<th>agLASSO SP</th>
<th>agLASSO TPR</th>
<th>agLASSO FPR</th>
<th>LAD-gLASSO SP</th>
<th>LAD-gLASSO TPR</th>
<th>LAD-gLASSO FPR</th>
<th>FAR-gSCAD SP</th>
<th>FAR-gSCAD TPR</th>
<th>FAR-gSCAD FPR</th>
<th>SAFE-gLASSO SP</th>
<th>SAFE-gLASSO TPR</th>
<th>SAFE-gLASSO FPR</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1 + C1</td>
<td>875</td>
<td>87 (2.13)</td>
<td>100 (5.32)</td>
<td>100 (3.00)</td>
<td>87 (2.00)</td>
<td>100 (2.00)</td>
<td>100 (2.00)</td>
<td>87 (2.00)</td>
<td>100 (2.00)</td>
<td>100 (2.00)</td>
<td>88 (100)</td>
<td>100 (100)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>A2 + C1</td>
<td>445</td>
<td>86 (2.13)</td>
<td>100 (5.35)</td>
<td>100 (3.00)</td>
<td>87 (2.00)</td>
<td>100 (2.00)</td>
<td>100 (2.00)</td>
<td>88 (100)</td>
<td>100 (100)</td>
<td>100 (100)</td>
<td>88 (100)</td>
<td>100 (100)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>A3 + C1</td>
<td>80</td>
<td>87 (2.05)</td>
<td>0 (5.01)</td>
<td>100 (3.00)</td>
<td>81 (2.00)</td>
<td>100 (2.00)</td>
<td>100 (2.00)</td>
<td>88 (100)</td>
<td>100 (100)</td>
<td>100 (100)</td>
<td>88 (100)</td>
<td>100 (100)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>A1 + C2</td>
<td>8</td>
<td>88 (2.00)</td>
<td>100 (4.15)</td>
<td>100 (2.99)</td>
<td>81 (2.00)</td>
<td>100 (2.00)</td>
<td>100 (2.00)</td>
<td>88 (100)</td>
<td>100 (100)</td>
<td>100 (100)</td>
<td>88 (100)</td>
<td>100 (100)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>A2 + C2</td>
<td>4</td>
<td>87 (2.00)</td>
<td>100 (4.56)</td>
<td>100 (2.90)</td>
<td>81 (2.00)</td>
<td>100 (2.00)</td>
<td>100 (2.00)</td>
<td>88 (100)</td>
<td>100 (100)</td>
<td>100 (100)</td>
<td>88 (100)</td>
<td>100 (100)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>A3 + C2</td>
<td>0.8</td>
<td>87 (2.01)</td>
<td>100 (4.91)</td>
<td>100 (2.86)</td>
<td>82 (2.00)</td>
<td>100 (2.00)</td>
<td>100 (2.00)</td>
<td>88 (100)</td>
<td>100 (100)</td>
<td>100 (100)</td>
<td>88 (100)</td>
<td>100 (100)</td>
<td>0 (0)</td>
</tr>
</tbody>
</table>

Figure C.18: Comparison of MSEs based on competing approaches for high (left) and low (right) SNRs with high correlation coefficient ($B2$). Simulated data corresponds to the non-varying finger movements. Reported are the results based on 100 simulations.
C.4 Numerical experiment

As described in Gertheiss et al. (2013b), the 10 functional predictors are generated by

\[ X_{k,i}(s) = \{\sigma(s)\}^{-1} \sum_{r=1}^{5} [a_{ikr}\sin\{2\pi s(5-a_{ikr})\} - m_{ikr}] \].

Figure C.19 illustrates the first two functional predictors. Similar to Gertheiss et al. (2013b) and Pannu & Billor (2017), we assume \( a_{ikr} \sim U(0,5) \), \( m_{ikr} \sim U(0,2\pi) \) and \( \sigma(s) \) is defined such that \( \text{var}\{X_{k,i}(s)\} = 0.01 \), for all \( s \). The response is generated from

\[ y_i = \alpha + \sum_{k=1}^{2} \int_{s} X_{k,i}(s)\gamma_k(s,z_i)ds + \epsilon_i \], where \( \epsilon_i \)’s are assumed to be independent and identically distributed (IID) as \( \epsilon_i \sim \mathcal{N}(0,\sigma^2) \). The nonzero functional coefficients \( \{\gamma_k(\cdot,\cdot); k=1,2\} \)’s are varying over \( z_i \in [-1,1] \) and defined as \( \gamma_1(s,z) = 1 + \sqrt{2}Czk + \sqrt{2}k\cos(\pi s) \) and \( \gamma_2(s,z) = 1+C\exp(-0.5z)+s+0.5s^2 \). For example, when \( C = 0 \), we have that \( \gamma_1(s,z) = 1+\sqrt{2} k\cos(\pi s) \) and \( \gamma_2(s,z) = 1+s+0.5s^2 \) which are non-varying functional coefficients; for \( C \neq 0 \), \( \gamma_k(\cdot,\cdot) \)’s vary over \( z \). We investigate the cases for \( C = \{0,2,5,10\} \). Figure C.20 depicts the effect of functional predictors at different values of \( C \).

The steps for fitting the procedure is similar to the one described in section 5 of the main document. One difference is that we use 15 basis functions in the \( s \) direction and 7 basis functions in the \( z \) direction in modeling \( \gamma_k(\cdot,\cdot) \)’s. The invariant functional coefficients \( \gamma_k(\cdot) \)’s
are modeled using 15 basis functions for agLASSO, LAD-gLASSO, and FAR-gSCAD approach. Additionally, the tuning parameters are selected using random CV for all competing methods.

Table C.3 presents the results using the performance metrics as described in section 5 of the main document, but focus only on recovery of $K = \{1, 2\}$ (there are no $K_F$ and $K_E$ here). As expected when $C = 0$, the numerical performance of all the methods is competitive. However, as $C$ departs from 0, say at $C = 5$, the variable selection and prediction accuracy of the competitors deteriorates.

Figure C.20: $\beta_1$ (top) and $\beta_2$ (bottom) for $C = \{0, 2, 5, 10\}$ from left to right respectively.
Table C.3: Variable selection for the toy example. Reported are the SPs (%), model size (in square brackets), TPRs (%), and FPRs (%) averaged across 100 simulations. Results with superscript † correspond to the first stage of SAFE-gLASSO.

<table>
<thead>
<tr>
<th>Combination</th>
<th>agLASSO</th>
<th>LAD-gLASSO</th>
<th>FAR-gSCAD</th>
<th>SAFE-gLASSO</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SP</td>
<td>TPR</td>
<td>FPR</td>
<td>SP</td>
</tr>
<tr>
<td>SNR = 0.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C = 0</td>
<td>77</td>
<td>100</td>
<td>4</td>
<td>80</td>
</tr>
<tr>
<td></td>
<td>[2.33]</td>
<td>[5.93]</td>
<td></td>
<td>[2.01]</td>
</tr>
<tr>
<td>C = 2</td>
<td>87</td>
<td>94</td>
<td>5</td>
<td>89</td>
</tr>
<tr>
<td></td>
<td>[2.26]</td>
<td>[4.94]</td>
<td></td>
<td>[1.09]</td>
</tr>
<tr>
<td>C = 5</td>
<td>81</td>
<td>69</td>
<td>7</td>
<td>90</td>
</tr>
<tr>
<td></td>
<td>[1.90]</td>
<td>[4.37]</td>
<td></td>
<td>[1.03]</td>
</tr>
<tr>
<td>SNR = 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C = 0</td>
<td>79</td>
<td>100</td>
<td>2</td>
<td>80</td>
</tr>
<tr>
<td></td>
<td>[2.11]</td>
<td>[6.53]</td>
<td></td>
<td>[2.02]</td>
</tr>
<tr>
<td>C = 2</td>
<td>79</td>
<td>100</td>
<td>2</td>
<td>86</td>
</tr>
<tr>
<td></td>
<td>[2.09]</td>
<td>[5.63]</td>
<td></td>
<td>[1.44]</td>
</tr>
<tr>
<td>C = 5</td>
<td>84</td>
<td>74</td>
<td>2</td>
<td>90</td>
</tr>
<tr>
<td></td>
<td>[1.64]</td>
<td>[4.31]</td>
<td></td>
<td>[1.03]</td>
</tr>
<tr>
<td>SNR = 5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C = 0</td>
<td>80</td>
<td>100</td>
<td>0</td>
<td>80</td>
</tr>
<tr>
<td></td>
<td>[2.00]</td>
<td>[5.66]</td>
<td></td>
<td>[2.00]</td>
</tr>
<tr>
<td>C = 2</td>
<td>80</td>
<td>100</td>
<td>0</td>
<td>80</td>
</tr>
<tr>
<td></td>
<td>[2.00]</td>
<td>[5.50]</td>
<td></td>
<td>[2.00]</td>
</tr>
<tr>
<td>C = 5</td>
<td>84</td>
<td>81</td>
<td>0</td>
<td>88</td>
</tr>
<tr>
<td></td>
<td>[1.62]</td>
<td>[4.54]</td>
<td></td>
<td>[1.22]</td>
</tr>
</tbody>
</table>

170
Figure C.21: Reported is MSEs based on competing approaches for different SNRs at $C = 0$ (left), $C = 2$ (middle), and $C = 5$ (right). Results are based on 100 simulations. Reference lines are drawn at MSE = 0.04 (left), MSE = 0.20 (middle), and MSE = 0.75 (right) for convenience.
C.5 Algorithm of split conformal prediction band

For the simplicity of exposition, denote by $\mathcal{A}$ the proposed algorithm which includes the steps of (i) variable selection and (ii) predictive modeling. In particular, we construct the split conformal prediction bands (Lei et al., 2017, 2015) for in-sample and new observations. We use the subscript $in$ and $new$ to refer to the in-sample and new data, respectively. Indeed the conformal inference offers distribution-free predictive inference in high-dimensional regression; while the method preserves the consistency properties of the estimators, it also provides the valid prediction coverage (Lei et al., 2017). Let $y_i$ be the scalar response, $N$ be the total number of observations, and $X_{in,i}$ be the corresponding covariates information. We adopt split conformal inference to construct prediction bands for future observations that are indexed by $i^*$ and use rank-one-out (ROO) split conformal inference for in-sample observations which are indexed by $i$. For completeness, we present both algorithms as below.
Algorithm Split Conformal Prediction

**Input:** Data \((y_i, X_{in,i})_i = 1, \ldots, N\); where \(X_{in,i} = [z_i, \{X_{k,i}(s_r); r = 1, \ldots, R\}; k = 1, \ldots, K]\). Let miscoverage level be \(\alpha \in (0, 1)\) and denote the proposed SAFE-gLASSO algorithm by \(A\)

**Output:** Prediction intervals at future observations \((i^*)\) with covariates \(X_{new,i^*}\); \(i^* = N + 1, N + 2, \ldots\)

Randomly split \(\{1, \ldots, N\}\) into two equal-sized subsets \(I_1, I_2\)

\[
\hat{\mu} = A\left(\{(y_i, X_{in,i}) : i \in I_1\}\right)
\]

\[
R_i = |y_i - \hat{\mu}(X_{in,i})| \text{ where } i \in I_2;
\]

\(d = \text{the } p\text{th smallest value in } \{R_i : i \in I_2\}, \text{ where } p = \left\lfloor \frac{N/2 + 1}{1 - \alpha} \right\rfloor\)

Return \(C_{split}(X_{new,i^*}) = [\hat{\mu}(X_{new,i^*}) - d, \hat{\mu}(X_{new,i^*}) + d]\), for all \(i^*\)
**Algorithm** Rank-One-Out (ROO) Split Conformal Prediction

**Input:** Data \((y_i, X_i), i = 1, \ldots, N; \) where \(X_i = [z_i, \{X_{k,i}(s_r); r = 1, \ldots, R\}; k = 1, \ldots, K]\).

Let miscoverage level be \(\alpha \in (0, 1)\) and denote the proposed SAFE-gLASSO algorithm by \(\mathcal{A}\).

**Output:** In-sample prediction intervals at \(X_i\)

Randomly split \(\{1, \ldots, N\}\) into two equal-sized subsets \(\mathcal{I}_n, n = 1, 2\)

for \(n \in \{1, 2\}\) do

\[\bar{\mu}_n = \mathcal{A}\left\{\left( (y_i, X_i) : i \in \mathcal{I}_n \right) \right\}\]

for \(i \notin \mathcal{I}_n\) do

\[R_i = |y_i - \bar{\mu}_n(X_i)|\]

end for

for \(i \notin \mathcal{I}_n\) do

\[d_i = \text{the } m\text{th smallest value in } \{R_j : j \notin \mathcal{I}_n, j \neq i\}, \text{ where } m = \left\lfloor \frac{N}{2}(1 - \alpha) \right\rfloor\]

\[C_{\text{roo}}(X_i) = [\bar{\mu}_n(X_i) - d_i, \bar{\mu}_n(X_i) + d_i]\]

end for

end for

Return intervals \(C_{\text{roo}}(X_i); i = 1, \ldots, N\)