

# SUPPLEMENT TO “INVERSE REGRESSION FOR LONGITUDINAL DATA”

BY CI-REN JIANG<sup>‡,\*</sup> AND WEI YU<sup>§</sup> AND JANE-LING WANG<sup>¶,†</sup>

*Academia Sinica<sup>‡</sup>, Genentech Inc,<sup>§</sup> and University of California at Davis<sup>¶</sup>*

## 1. Definition of $\Gamma^{-1/2}$ and $\Gamma^{-1/2}\Gamma_e\Gamma^{-1/2}$ when $\xi_i > 0$ , for all $i$ .

First, we show the range of  $\Gamma^{1/2}$ ,  $R_{\Gamma^{1/2}} = \{\Gamma^{1/2}g : g \in L_2(\mathcal{I})\}$ , is equal to

$$(1.1) \quad \{f \in L_2(\mathcal{I}) : \sum_i \xi_i^{-1} |\langle f, \phi_i \rangle|^2 < \infty, f \perp \ker(\Gamma)\},$$

where  $\{\xi_i, \phi_i\}$  are the eigenvalues and eigenfunctions of  $\Gamma$ .

PROOF. If  $f = \Gamma^{1/2}g$  and  $g \in L_2(\mathcal{I})$ , then  $f = \sum_k \xi_k^{1/2} \langle g, \phi_k \rangle \phi_k$ . Simple calculations show

$$\sum_i \frac{1}{\xi_i} \left| \left\langle \sum_k \xi_k^{1/2} \langle g, \phi_k \rangle \phi_k, \phi_i \right\rangle \right|^2 = \sum_i \frac{1}{\xi_i} \left| \xi_i^{1/2} \langle g, \phi_i \rangle \right|^2 < \infty,$$

and  $f \perp \ker(\Gamma)$ .

On the other hand, since  $\{\phi_k\}_{k=1}^\infty$  and  $\ker(\Gamma) = \{\psi_k\}_{k=1}^\infty$  span the  $L_2(\mathcal{I})$  space,  $f = \sum_k \langle f, \phi_k \rangle \phi_k + \sum_k \langle f, \psi_k \rangle \psi_k$  for any  $f \in L_2(\mathcal{I})$ . Under the condition  $f \perp \ker(\Gamma)$ ,  $f$  can be simplified as  $\sum_k \langle f, \phi_k \rangle \phi_k$  because  $\langle f, \psi_k \rangle = 0$  for all  $k$ . To show there exists a  $g \in L_2(\mathcal{I})$  such that  $\Gamma^{1/2}g = f$ , we specifically let  $g = \sum_k \frac{1}{\sqrt{\xi_k}} \langle f, \phi_k \rangle \phi_k$ . Simple algebra shows that  $\Gamma^{1/2}g = \sum_k \langle f, \phi_k \rangle \phi_k$  and  $\|g\|^2 = \sum_k \xi_k^{-1} |\langle f, \phi_k \rangle|^2 < \infty$ . Therefore,  $\Gamma^{1/2}g = f$  and  $g \in L_2(\mathcal{I})$ . Thus, the range of  $\Gamma^{1/2}$  can be represented by (1.1). □

1.1. *Definition of  $\Gamma^{-1/2}$ .* Similar to He et al. (2003), we consider a subset  $R_{\Gamma^{1/2}}$  of  $L_2(\mathcal{I})$ , on which the inverse of a compact operator can be defined, and for  $f \in R_{\Gamma^{1/2}}$  we define

$$\Gamma^{-1/2}f = \sum_i \xi_i^{-1/2} \langle f, \phi_i \rangle \phi_i.$$

---

<sup>\*</sup>Supported in part by NSC grant NSC 101-2118-M-001-013-MY2, Taiwan.

<sup>†</sup>Supported in part by NSF grants DMS-04-04630, DMS-09-06813, and DMS-12-28369.

Then,  $\Gamma^{-1/2}$  is in  $L_2(\mathcal{I})$  and satisfies the usual properties of an inverse in the sense that  $\Gamma^{1/2}\Gamma^{-1/2}f = f$ , for all  $f \in R_{\Gamma^{1/2}}$ , and  $\Gamma^{-1/2}\Gamma^{1/2}g = g$ , for all  $g \in R_{\Gamma^{-1/2}}$ , the range space of  $\Gamma^{-1/2}$ .

Thus  $\Gamma^{1/2}$  is a one-to-one mapping from the vector space  $R_{\Gamma^{-1/2}}$  onto the vector space  $R_{\Gamma^{1/2}}$ .

1.2. *Definition of  $\Gamma^{-1/2}\Gamma_e\Gamma^{-1/2}$ .* Since  $\Gamma(s, t) = \sum_i \xi_i \phi_i(s) \otimes \phi_i(t)$  and  $\Gamma_e(s, t) = \sum_{i,j} E\{E(A_i|y)E(A_j|y)\} \phi_i(s) \otimes \phi_j(t)$ ,

$$\begin{aligned} \Gamma^{-1/2}\Gamma_e\Gamma^{-1/2} &= \sum_{i,j} \frac{E\{E(A_i|y)E(A_j|y)\}}{(\xi_i\xi_j)^{1/2}} \phi_i(s) \otimes \phi_j(t) \\ &= \sum_{i,j} \frac{E\{A_iE(A_j|y)\}}{(\xi_i\xi_j)^{1/2}} \phi_i(s) \otimes \phi_j(t). \end{aligned}$$

In order for  $\Gamma^{-1/2}\Gamma_e\Gamma^{-1/2}(s, t) = \sum_k \lambda_k \eta_k(s) \otimes \eta_k(t)$  to be a well-defined Hilbert-Schmidt operator, we need  $\sum_k \lambda_k^2 < \infty$ . Because  $\eta_k$  is the  $k$ -th eigenfunction of  $\Gamma^{-1/2}\Gamma_e\Gamma^{-1/2}$ , we have  $\Gamma^{-1/2}\Gamma_e\Gamma^{-1/2}\eta_k = \lambda_k \eta_k$  and thus it is not difficult to show that  $\sum_k \lambda_k^2 < \infty$  if  $\sum_{i,j} \frac{E^2\{A_iE(A_j|y)\}}{\xi_i\xi_j}$  is finite. Therefore,

$$(1.2) \quad \sum_{i,j} \frac{E^2\{A_iE(A_j|y)\}}{\xi_i\xi_j} < \infty$$

is a sufficient condition for  $\Gamma^{-1/2}\Gamma_e\Gamma^{-1/2}$  to be a well-defined Hilbert-Schmidt operator.

Let  $\beta_k(t) = \Gamma^{-1/2}\eta_k$  be the e.d.r. direction. In order for  $\beta_k(t)$  to be well defined, we need  $\eta_k(t) \in R_{\Gamma^{1/2}}$ , i.e.  $\sum_i \xi_i^{-1} |\langle \eta_k, \phi_i \rangle|^2 < \infty$ . Then, for  $k \geq 1$ , we have

$$\Gamma^{-1/2}\Gamma_e\Gamma^{-1/2}\eta_k = \lambda_k \eta_k = \sum_{i,j} \frac{E\{A_iE(A_j|y)\}}{(\xi_i\xi_j)^{1/2}} \langle \phi_j, \eta_k \rangle \phi_i.$$

From the definition of  $R_{\Gamma^{1/2}}$ , we need

$$\sum_i \frac{1}{\xi_i} \left( \sum_j \frac{E\{A_iE(A_j|y)\}}{(\xi_i\xi_j)^{1/2}} \langle \phi_j, \eta_k \rangle \right)^2 < \infty$$

to show  $\eta_k \in R_{\Gamma^{1/2}}$ . By Cauchy-Schwarz inequality and that  $\sum_j \langle \phi_j, \eta_k \rangle^2 = 1$ , we can show that

$$\sum_i \frac{1}{\xi_i} \left( \sum_j \frac{E\{A_i E(A_j|y)\}}{(\xi_i \xi_j)^{1/2}} \langle \phi_j, \eta_k \rangle \right)^2 \leq \sum_i \frac{1}{\xi_i} \sum_j \frac{E^2\{A_i E(A_j|y)\}}{\xi_i \xi_j}.$$

Therefore,  $\eta_k \in R_{\Gamma^{1/2}}$  if

$$(1.3) \quad \sum_{i,j} \frac{E^2\{A_i E(A_j|y)\}}{\xi_i^2 \xi_j} < \infty.$$

**2. Simulation Studies.** Since our method is applicable to both sparse longitudinal data and functional data, we evaluate its finite sample performance for both types of data through simulations. Without loss of generality, we set the domain interval  $\mathcal{I}$  as  $[0, 1]$ . Let  $X(t)$  be a standard Brownian motion on  $[0, 1]$ , we consider the following model:

$$(2.1) \quad Y = 3 + \exp(\langle \beta(t), X(t) \rangle) + \epsilon,$$

where  $\beta(t) = \sqrt{2} \sin(3\pi t/2)$ , and the random error  $\epsilon \sim N(0, 0.1^2)$ . The standard deviation of  $\epsilon$  may look small, but the range of  $\exp(\langle \beta(t), X(t) \rangle)$  is around  $(0.5, 1.5)$  and thus the signal-noise ratio is about 10.

Two simulation studies were conducted. In simulation I, the data were generated under both sparse and complete settings to evaluate the general performance of the proposed approach. In simulation II, sparsely observed data were generated with fixed  $N_i$ , the number of observations of subject  $i$ , to evaluate the empirical convergent rate.

**2.1. Simulation I.** In each run,  $n$  sample trajectories,  $\{X_i(t), i = 1, \dots, n\}$ , are generated from Brownian motion on  $[0, 1]$ . This forms the complete data, but for practical implementation we discretized the data to equally spaced 31 time-points,  $\{t_0, t_1, \dots, t_{30}\}$ , with  $t_0 = 0$  and  $t_{30} = 1$ . Therefore, the actual dense data set is  $\{X_{ij} = X_i(t_{ij}), i = 1, \dots, n, j = 1, \dots, 30\}$  along with its response  $Y_i$ . To generate the sparse longitudinal data, we randomly selected 2 to 10 observations from  $\{t_1, t_2, \dots, t_{30}\}$ . This results in the longitudinal data  $(X_{i1}, \dots, X_{iN_i})$  for the  $i$ th subject at time points  $(t_{i1}, \dots, t_{iN_i})$ , where  $N_i$  follows a uniform distributions on  $\{2, 3, \dots, 10\}$ . The simulation consists of 100 runs and Table 1 summarizes the numerical findings when  $n$  is 100 and 200. As a comparison, we also include the results of the smoothed functional inverse regression approach in Ferré and Yao (2005), which is for complete data.

The first comparison is based on the correlation between  $\langle \beta(t), X(t) \rangle$  and  $\langle \hat{\beta}(t), X(t) \rangle$ , i.e. the correlation between the projection of  $X(t)$  on the real e.d.r. direction and that on the estimated e.d.r. direction. Averages of those correlations are reported in the third column of Table 1. The results suggest that our approach generally produces high correlations and for complete data have larger correlations than those from Ferré and Yao (2005). The rest comparisons are based on the Integrated Squared Bias (ISB), Integrated Variance (IVAR), and Integrated Mean Square Error (IMSE) (or Mean of Integrated Square Error(MISE)). The Appendix contains details of those definitions. The averages of these statistics over the 100 simulation runs are reported in Columns 4-6 of Table 1. Expectedly, the results for complete data are better than those for sparse data and the results for larger sample sizes are better. For complete data, our procedure generally led to smaller ISB, IVAR, and IMSE than Ferré and Yao's.

In addition to the above global measures, we plot in Figure 1 the mean function for each of the three  $\beta$ -estimates. The left panel of Figure 1 shows the average of  $\hat{\beta}(t)$ -functions (dashed line for complete, dotted line for sparse data and dash-dot line for Ferré and Yao (2005)) when  $n = 100$  along with the true  $\beta(t)$  (solid line), the right panel provides the same plot for  $n = 200$ . Figure 1 indicates that the biases of our approach are comparable to those of Ferré and Yao (2005) when data are observed completely. The bias of our approach is significantly reduced for sparse data when the sample size increases to 200, due to improved estimation of  $\Gamma$  and  $\Gamma_e$ .

**2.2. Simulation II.** The purpose of this simulation is to study the empirical convergent rate of the estimated standardised e.d.r.  $\hat{\eta}_j(t)$ . Thus, we applied identical bandwidths in each run even when the sample sizes are different to facilitate the comparison. All the simulation settings are identical to those in Simulation I, except for the number of observations per curve. In this study,  $N_i$ 's are not randomly selected. Instead,  $N_i = 6$  for all  $i$  so the comparison can be more transparent. For each subject, we randomly selected 6 observations from  $\{t_1, t_2, \dots, t_{30}\}$  to generate the sparse longitudinal data. This results in the longitudinal data  $(X_{i1}, \dots, X_{i6})$  for the  $i$ th subject at time points  $(t_{i1}, \dots, t_{i6})$ . The simulation consists of 100 runs and Table 2 summarizes the numerical findings when  $n$  is 100, 200 and 400. Since the order of the bias is determined by the order of the bandwidth and this is an analytical result from Taylor expansion, we focus on the rate of the variance. The empirical convergent rates of the standardized e.d.r. directions  $(\eta_k)$  can be checked through the IVAR results reported in Table 2. The ratios of the  $\sqrt{IVAR}$  for two consecutive samples in Table 2 are close

to  $\sqrt{2}$ , which is the square root of the ratio of sample sizes. This provides numerical evidence of the convergence rate.

**3. Data Analysis.** The data set we use here are the records of the lifetimes and daily reproduction of female Medflies, quantified by the number of eggs laid daily for 1000 female Mediterranean fruit flies. Details about the experimental background can be found in Carey et al. (1998). Our goal is to explore the relationship between early patterns of fecundity, quantified by the number of eggs laid per day until day 20, and mortality for each individual fly. For this reason, we exclude flies that died by day 30 and flies that did not lay any eggs. The remaining 647 flies have an average lifetime ( $Y$ ) of 43.9 days with a standard deviation of 11.9 days. It is assumed that there is an underlying stochastic predicting process  $X(t)$  which quantifies the reproduction pattern and can be characterized as a fecundity curve, represented by the daily egg counts. The numbers of eggs laid in the first 20 days are discrete observations of the function  $X(t)$ . The objective of our analysis is to find the e.d.r. directions such that the projection of the observations onto the resulting e.d.r. space will carry the key information for longevity in the regression  $E(Y|X)$ .

To test the efficiency of our method and to check the effect of sparse data, we first use the complete information of all 20 days as *complete/dense data*; and then randomly pick  $N_i$  points from each fly as our *sparse data*, where  $N_i$  is uniformly distributed in  $\{2, \dots, 10\}$ . The approach in Ferré and Yao (2005) was also applied to the complete data as a comparison. The results for the functional inverse regression model and the associated direction estimates are shown in Figures 2, 3 and 4. Figure 2 gives the plot of fraction of variance explained by the eigenvalues of the standardized covariance functions for  $E(X(t)|Y = y)$  for both sparse and complete data analysed by our approach. We find that the eigenvalues decrease very fast and the fraction of variance explained by the first eigenfunction is over 95% for sparse and 72% for complete data. Further, the first two eigenfunctions explain over 90% of the total variation for both sparse and complete data. Thus, one or at most two directions would suffice to summarize the information contained in the fecundity data in order to infer lifetime.

Figure 3 displays the directions estimated by our approach for both complete and sparse data and by the method in Ferré and Yao (2005) for the complete data only. The directions estimated by Ferré and Yao (2005) are less smooth because  $\Gamma(s, t)$  was estimated empirically without smoothing. However, the general trends of these directions are similar to ours except for the first index after 15 days ( $t > 15$ ). The global patterns of the direction

estimates by our approach are similar between the two types of data and the difference might be due to the difference in the selected bandwidths (a larger bandwidth is used for sparse data to compensate for the sparsity, and this leads to smoother directions.) The estimated  $\beta_1(t)$  indicates that daily reproduction during the period day 4 to day 10 plays an important role in mortality, while the estimated  $\beta_2(t)$  shows the effect of daily reproduction from day 10 to day 20.

One caveat is that the  $i$ -th index,  $\langle \beta_i, \mathbf{X} \rangle$ , can not be estimated well by conventional numerical integrations when covariates are observed sparsely. A remedy is to impute the indices for each subject. Specifically, for the  $j$ th index of subject  $i$ , we consider

$$E(\langle \beta_j(t), X_i(t) \rangle | \mathbf{X}_i) = \int \mu(t) \beta_j(t) dt + \sum_k E(A_{ik} | \mathbf{X}_i) \left( \int \phi_k(t) \beta_j(t) dt \right).$$

If the principal scores ( $A_{ik}$ ) are normality distributed,

$$E(A_{ik} | \mathbf{X}_i) = \lambda_k \phi_k^T(\mathbf{T}_i) \Sigma_{\mathbf{X}_i}^{-1} \{ \mathbf{X}_i - \mu(\mathbf{T}_i) \},$$

where  $\phi_k^T(\mathbf{T}_i) = (\phi_k(T_{i1}), \dots, \phi_k(T_{iN_i}))$ ,  $\Sigma_{\mathbf{X}_i}$  is the covariance of  $\mathbf{X}_i$  and  $\mu^T(\mathbf{T}_i) = (\mu(T_{i1}), \dots, \mu(T_{iN_i}))$ . When the normality assumption does not hold the above expression is no longer the true conditional expectation but is the best linear projection of the principal scores onto the linear space spanned by  $\mathbf{X}_i$ . Therefore, we propose to estimate the  $j$ -th index of subject  $i$  by

(3.1)

$$\hat{E}(\langle \beta_j(t), X_i(t) \rangle | \mathbf{X}_i) = \int \hat{\mu}(t) \hat{\beta}_j(t) dt + \sum_k \hat{E}(A_{ik} | \mathbf{X}_i) \left( \int \hat{\phi}_k(t) \hat{\beta}_j(t) dt \right),$$

where  $\hat{E}(A_{ik} | \mathbf{X}_i) = \hat{\lambda}_k \hat{\phi}_k^T(\mathbf{T}_i) \hat{\Sigma}_{\mathbf{X}_i}^{-1} \{ \mathbf{X}_i - \hat{\mu}(\mathbf{T}_i) \}$ .

Since two directions suffice to summarize the information contained in the fecundity data to infer lifetime, we further explore the relation of lifetimes with these two directions by assuming that the error  $\epsilon$  in the model is additive. The unknown bivariate regression function is estimated by a bivariate local linear smoother on the estimated bivariate indices. Specifically, the estimated link surface based on the two indices at  $(u, v)$  is  $\hat{y} = \hat{b}_0(u, v)$ , where

$$\begin{aligned} (\hat{b}_0, \hat{b}_1, \hat{b}_2)^T = \arg \min_{\mathbf{b}=(b_0, b_1, b_2)^T} \sum_{i=1}^n \{ Y_i - b_0 - b_1(u - U_i) - b_2(v - V_i) \}^2 \\ \times K_2\left(\frac{U_i - u}{h_u}, \frac{V_i - v}{h_v}\right), \end{aligned}$$

$K_2$  is the bivariate kernel function with bandwidths  $h_u$  and  $h_v$ , and  $U_i = \langle \hat{\beta}_1(t), X_i(t) \rangle$  and  $V_i = \langle \hat{\beta}_2(t), X_i(t) \rangle$  are the two estimated indices of the  $i$ -th subject. The estimated link surfaces are provided in Figure 4. The e.d.r. directions of the upper panel were obtained by our approach (left) and Ferré and Yao (2005)'s (right) from complete data, and those of the lower panel were obtained by our approach from sparse data using two different ways to calculate the index. The lower right plot uses the complete covariate  $X(t)$  but the lower left plot uses the imputed  $X(t)$  through the conditional expectation (3.1). The estimated e.d.r. directions are not identical so the ranges of the resulted indices also vary with the imputed ones most different from the rest. This is because of the shrinkage effect in the imputation, which leads to the narrowest ranges among all four methods. Despite this difference, the global pattern of all four methods are similar in the sense that lifetimes tend to increase with increasing size of the first index when the second index is held fixed. From the model fitting perspective, all four methods are also comparable since the averages of the square fitted errors shown in Table 3 differs little. Specifically, averages of the square fitted errors are similar for all three methods where the indices were obtained by plugging in the complete covariate  $X(t)$  and interestingly our approach for sparse data performed slightly better than Ferré and Yao (2005)'s approach based on complete data. Expectedly, the worse case is when both the e.d.r. directions and indices are estimated from sparse data, but the average of square fitted error is only slightly larger than the others. Overall, this supports the use of the imputed  $X(t)$  based on (3.1).

Combining Figures 3 and 4, we find that a fly laying fewer eggs from day 4 to day 10 but making it up later by reaching average number of egg production during the period day 10 to day 20 is expected to live longer. Since egg production is most intense in the early stage (day 4 to 10), this suggests a cost of early reproduction to female Medflies. One plausible explanation is that young Medflies are still fragile and reproduction depletes the needed nutrition for growth.

#### 4. Proof of Lemma 2.2.

PROOF OF LEMMA 2.2. We define

$$\tilde{\Gamma}_e(s, t) = \frac{1}{n} \sum_{i=1}^n m(s, Y_i) m(t, Y_i) - \frac{1}{n} \sum_{i=1}^n m(s, Y_i) \frac{1}{n} \sum_{i=1}^n m(t, Y_i),$$

then  $\hat{\Gamma}_e - \Gamma_e = (\hat{\Gamma}_e - \tilde{\Gamma}_e) + (\tilde{\Gamma}_e - \Gamma_e)$ .

We first consider  $\hat{\Gamma}_e - \tilde{\Gamma}_e$ :

$$\begin{aligned}\hat{\Gamma}_e - \tilde{\Gamma}_e &= \left\{ \frac{1}{n} \sum_{i=1}^n \hat{m}(s, Y_i) \hat{m}(t, Y_i) - \frac{1}{n} \sum_{i=1}^n m(s, Y_i) m(t, Y_i) \right\} \\ &\quad - \left\{ \frac{1}{n} \sum_{i=1}^n \hat{m}(s, Y_i) \frac{1}{n} \sum_{i=1}^n \hat{m}(t, Y_i) - \frac{1}{n} \sum_{i=1}^n m(s, Y_i) \frac{1}{n} \sum_{i=1}^n m(t, Y_i) \right\} \\ &= I_1 - I_2,\end{aligned}$$

where

$$\begin{aligned}I_1 &= \frac{1}{n} \sum_{i=1}^n \hat{m}(s, Y_i) \hat{m}(t, Y_i) - \frac{1}{n} \sum_{i=1}^n m(s, Y_i) m(t, Y_i), \\ I_2 &= \frac{1}{n} \sum_{i=1}^n \hat{m}(s, Y_i) \frac{1}{n} \sum_{i=1}^n \hat{m}(t, Y_i) - \frac{1}{n} \sum_{i=1}^n m(s, Y_i) \frac{1}{n} \sum_{i=1}^n m(t, Y_i).\end{aligned}$$

Rewrite  $I_1$  as

$$I_1 = S_1 + S_2 + S_3,$$

where

$$\begin{aligned}S_1 &= \frac{1}{n} \sum_{i=1}^n (\hat{m}(s, Y_i) - m(s, Y_i)) (\hat{m}(t, Y_i) - m(t, Y_i)), \\ S_2 &= \frac{1}{n} \sum_{i=1}^n m(s, Y_i) (\hat{m}(t, Y_i) - m(t, Y_i)), \\ S_3 &= \frac{1}{n} \sum_{i=1}^n (\hat{m}(s, Y_i) - m(s, Y_i)) m(t, Y_i).\end{aligned}$$

The  $L_2$  norm of  $S_1$  is

$$\begin{aligned}\|S_1\| &= \left[ \int \int \left( \frac{1}{n} \sum_{i=1}^n (\hat{m}(s, Y_i) - m(s, Y_i)) (\hat{m}(t, Y_i) - m(t, Y_i)) \right)^2 ds dt \right]^{1/2} \\ &\leq \left[ \int \int \frac{1}{n^2} \sum_{i=1}^n (\hat{m}(s, Y_i) - m(s, Y_i))^2 \sum_{i=1}^n (\hat{m}(t, Y_i) - m(t, Y_i))^2 ds dt \right]^{1/2} \\ &= \frac{1}{n} \sum_{i=1}^n \|\hat{m}(s, Y_i) - m(s, Y_i)\|^2.\end{aligned}$$

Lemma 2.1 implies that

$$E(\|\hat{m}(s, y) - m(s, y)\|^2 | \mathbf{T}_i, Y_i, i = 1, \dots, n) = O_p\left(\frac{1}{nh^2EN} + h^4\right).$$



Equivalently, for any  $\varepsilon$ , there exists  $M > 0$ , such that

$$(4.1) \quad P\left(\frac{E(\|e_n\|^2|S_n)}{r_n} \geq M\right) < \varepsilon,$$

where  $e_n = \hat{m}(s, y) - m(s, y)$ ,  $S_n = \{\mathbf{T}_i, Y_i, i = 1, \dots, n\}$  and  $r_n = \frac{1}{nh^2EN} + h^4$ .

Give  $K > 0$  such that  $\frac{M}{K} \leq \varepsilon$ , we have

$$(4.2) \quad \begin{aligned} E(\|e_n\|^2|S_n) &= \int \|e_n\|^2 dF_{e_n|S_n}(\cdot|S_n) \\ &\geq \int_{\{\frac{\|e_n\|^2}{r_n} \geq K\}} \|e_n\|^2 dF_{e_n|S_n}(\cdot|S_n) \\ &\geq r_n K P\left(\frac{\|e_n\|^2}{r_n} \geq K|S_n\right), \end{aligned}$$

where  $F_{e_n|S_n}$  is the conditional cdf of  $e_n$  given  $S_n$ .

Divided both sides of equation (4.2) by  $r_n$ , we obtain

$$\frac{E(\|e_n\|^2|S_n)}{r_n} \geq K P\left(\frac{\|e_n\|^2}{r_n} \geq K|S_n\right),$$

which implies the relationship between two sets:

$$\left\{\frac{E(\|e_n\|^2|S_n)}{r_n} \geq M\right\} \supset \left\{K P\left(\frac{\|e_n\|^2}{r_n} \geq K|S_n\right) \geq M\right\},$$

thereby

$$(4.3) \quad P\left(\frac{E(\|e_n\|^2|S_n)}{r_n} \geq M\right) \geq P\left(P\left(\frac{\|e_n\|^2}{r_n} \geq K|S_n\right) \geq \frac{M}{K}\right).$$

From equation (4.1)  $\sim$  (4.3), we obtain

$$(4.4) \quad P\left(P\left(\frac{\|e_n\|^2}{r_n} \geq K|S_n\right) \geq \frac{M}{K}\right) \leq \varepsilon.$$

We next consider  $P\left(\frac{\|e_n\|^2}{r_n} \geq K\right)$ . The simple calculation shows that

$$P\left(\frac{\|e_n\|^2}{r_n} \geq K\right) = E\left(P\left(\frac{\|e_n\|^2}{r_n} \geq K|S_n\right)\right).$$

Therefore, we have

$$\begin{aligned}
P\left(\frac{\|e_n\|^2}{r_n} \geq K\right) &= \int P\left(\frac{\|e_n\|^2}{r_n} \geq K | S_n\right) dF_{S_n} \\
&\leq \int_{\{P(\frac{\|e_n\|^2}{r_n} \geq K | S_n) \geq \frac{M}{K}\}} 1 dF_{S_n} + \frac{M}{K} \int dF_{S_n} \\
&\leq P\left(P\left(\frac{\|e_n\|^2}{r_n} \geq K | S_n\right) \geq \frac{M}{K}\right) + \frac{M}{K} \\
(4.5) \quad &\leq 2\varepsilon,
\end{aligned}$$

where  $F_{S_n}$  is the cdf of  $S_n$ , and the last inequality is from (4.4) and the definition of positive number  $K$ . The inequality (4.5) directly implies that

$$(4.6) \quad \|\hat{m}(s, y) - m(s, y)\|^2 = O_p\left(\frac{1}{nh^2EN} + h^4\right).$$

Thus, from (4.6),

$$\begin{aligned}
\|S_1\| &= O_p\left(\frac{1}{nh^2EN} + h^4\right), \\
\|S_2\| &= \left\| \frac{1}{n} \sum_{i=1}^n m(s, Y_i) (\hat{m}(t, Y_i) - m(t, Y_i)) \right\| \\
&\leq \sqrt{\frac{1}{n} \sum_{i=1}^n \|m(s, Y_i)\|^2} \|S_1\| \\
&= O_p\left(\sqrt{\frac{1}{nh^2EN} + h^4}\right).
\end{aligned}$$

$$\text{Similarly, } \|S_3\| = O_p\left(\sqrt{\frac{1}{nh^2EN} + h^4}\right).$$

Thus,  $I_1 = O_p\left(\sqrt{\frac{1}{nh^2EN} + h^4}\right)$ . Following the same arguments, we can show that  $I_2 = O_p\left(\sqrt{\frac{1}{nh^2EN} + h^4}\right)$ , and thus  $\hat{\Gamma}_e - \tilde{\Gamma}_e = O_p\left(\sqrt{\frac{1}{nh^2EN} + h^4}\right)$ .

Since  $\tilde{\Gamma}_e - \Gamma_e = O_p\left(\frac{1}{\sqrt{n}}\right) = o_p\left(\frac{1}{\sqrt{nh^2EN}}\right)$ , we have

$$\hat{\Gamma}_e - \Gamma_e = O_p\left(\sqrt{\frac{1}{nh^2EN} + h^4}\right).$$

□

## References.

- Carey, J. R., Müller, H. G., Wang, J. L., and Chiou, J. M. (1998). Relationship of age patterns of fecundity to mortality, longevity, and lifetime reproduction in a large cohort of mediterranean fruit fly females. *J. Gerontology: Biological Sciences*, 53A:B245–B251.
- Ferré, L. and Yao, A. (2005). Smoothed functional inverse regression. *Statist. Sinica*, 15:665–683.
- He, G., Müller, H. G., and Wang, J. (2003). Functional canonical analysis for square integrable stochasitc processes. *J. Multi. Anal.*, 85:54–77.

CI-REN JIANG  
INSTITUTE OF STATISTICAL SCIENCE  
ACADEMIA SINICA  
TAIPEI, 115  
TAIWAN  
E-MAIL: cirenjiang@stat.sinica.edu.tw

WEI YU  
GENENTECH, INC  
1 DNA WAY  
SOUTH SAN FRANCISCO, CA 94080  
USA  
E-MAIL: yu.wei@gene.com

JANE-LING WANG  
DEPARTMENT OF STATISTICS  
UNIVERSITY OF CALIFORNIA  
DAVIS, CA 95616  
USA  
E-MAIL: jlwang.ucdavis@gmail.com

TABLE 1

Simulation comparison of FY (Ferré and Yao (2005)) for complete data and our procedures for both complete and sparse data. The comparison is based on the averages of correlations, ISB, IVAR and IMSE in 100 simulation runs.

$n$	Data type	Correlation	ISB	IVAR	IMSE
100	FY (Complete)	0.7159	0.0114	0.2008	0.2123
	Complete	0.9912	0.0043	0.0084	0.0127
	Sparse	0.8831	0.0583	0.2823	0.3406
200	FY (Complete)	0.8218	0.0024	0.0837	0.0861
	Complete	0.9921	0.0024	0.0092	0.0116
	Sparse	0.9438	0.0274	0.1602	0.1876

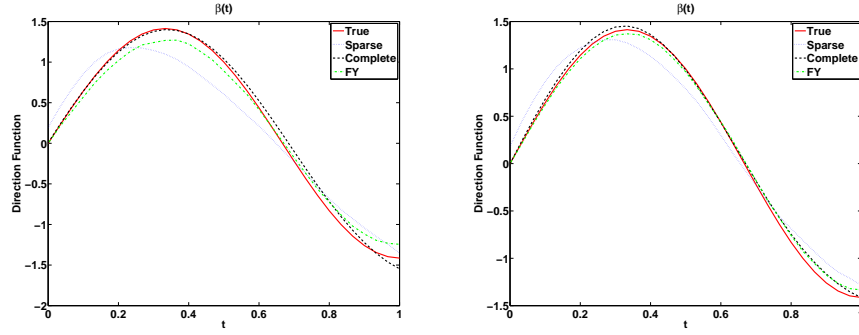


FIG 1. Simulation comparison of the average estimates of  $\beta(t)$  for the three methods in Table 1. The left panel shows the average of  $n$  estimate ( $\hat{\beta}(t)$ ) for various methods vs the target ( $\beta(t)$ ) for  $n=100$ , and the right panel for  $n=200$ .

TABLE 2

The average of  $\|\hat{\eta}(t) - \tilde{\eta}(t)\| = (\int_{\mathcal{T}} (\hat{\eta}(t) - \tilde{\eta}(t))^2 dt)^{1/2} = \sqrt{IVAR}$  for sparse data and ratios of the  $\sqrt{IVAR}$  for two consecutive samples is close to the theoretical rate  $\sqrt{2}$ .

$n$	$\ \hat{\eta}(t) - \tilde{\eta}(t)\ $
100	0.4375
200	0.3141
400	0.2138

TABLE 3

Average of the square fitted errors,  $\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$ , of the fecundity data, based on four different methods: Ferré and Yao (2005), our method with complete data, and our method with sparse data (using true  $X(t)$  to calculate the index and using imputed  $X(t)$  to calculate the index.)

Method	Complete: FY	Complete	Sparse: True $X(t)$	Sparse: Imputed $X(t)$
Fitted Error	134.28	134.05	134.13	139.38

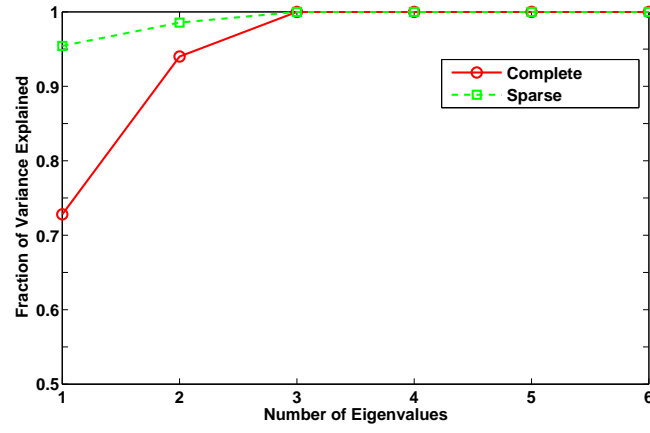


FIG 2. The fractions of variance explained from the complete data and sparse data

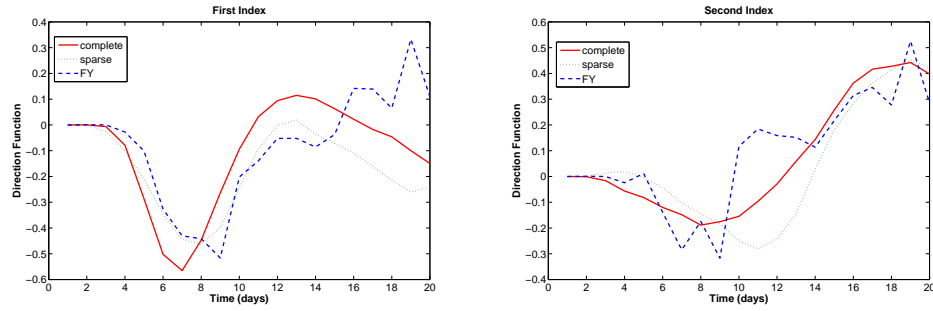


FIG 3. Estimated  $\beta_1(t)$  and  $\beta_2(t)$  from complete data and sparse data.

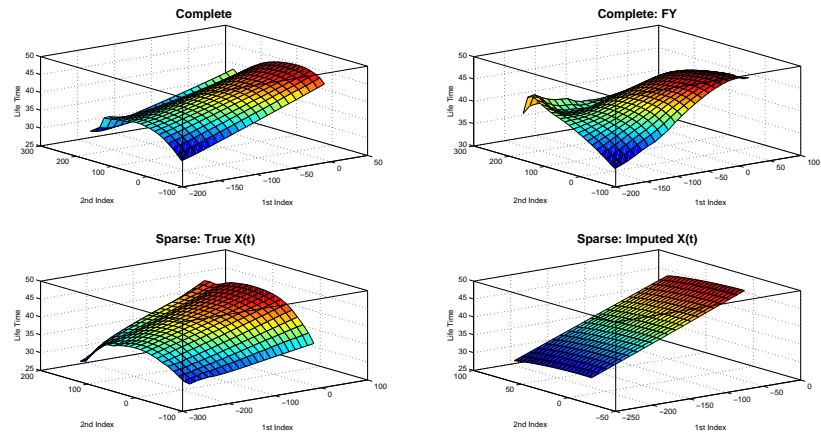


FIG 4. *Estimated link functions: the e.d.r. directions of the upper panel were obtained by our approach (left) and Ferré and Yao (2005) (right) from complete data and those of the lower panel were obtained by our approach from sparse data. The lower left shows the surface when the indices were calculated with true complete covariate and the lower right shows the surface when the indices were calculated with the imputed covariate (3.1).*