# Improvement of the estimation of FPC scores

Bitao Liu & Hans-Georg Müller

March 3, 2008

Let $Y_{ij}$ be the $j$th observation of the random function $X_i(\cdot)$, made at a random time $T_{ij}$ and $\varepsilon_{ij}$ the additional measurement errors that are assumed to be i.i.d. and independent of the random coefficients (FPC scores) $\xi_{ik}$, where $i = 1, \ldots, n$, $j = 1, \ldots, n_i$, $k = 1, 2, \ldots$. Then the model we consider is

$$Y_{ij} = X_i(T_{ij}) + \varepsilon_{ij} = \mu(T_{ij}) + \sum_{k=1}^{\infty} \xi_{ik}\phi_k(T_{ij}) + \varepsilon_{ij}, \ \ T_{ij} \in \mathcal{T},$$

where $\mathrm{E}\varepsilon_{ij} = 0$, $\mathrm{var}(\varepsilon_{ij}) = \sigma^2$.

Write $\widetilde{\mathbf{X}}_i = (X_i(T_{i1}), \ldots, X_i(T_{in_i}))^T$, $\widetilde{\mathbf{Y}}_i = (Y_{i1}, \ldots, Y_{in_i})^T$, $\boldsymbol{\mu}_i = (\mu(T_{i1}), \ldots, \mu(T_{in_i}))^T$, $\boldsymbol{\phi}_{ik} = (\phi_k(T_{i1}), \ldots, \phi_k(T_{in_i}))^T$. Yao *et al.* (2005) proposed to estimate the $\xi_{ik}$ through conditional expectation by assuming $\xi_{ik}$ and $\varepsilon_{ij}$ are jointly Gaussian, that is,

$$\hat{\xi}_{ik} = \hat{\mathrm{E}}[\xi_{ik}|\widetilde{\mathbf{Y}}_i] = \hat{\lambda}_k \hat{\boldsymbol{\phi}}_{ik}^T \hat{\boldsymbol{\Sigma}}_{Y_i}^{-1} (\widetilde{\mathbf{Y}}_i - \hat{\boldsymbol{\mu}}_i) \tag{1}$$

where the $(j, l)$ element of $(\hat{\boldsymbol{\Sigma}}_{Y_i})_{j,l} = \hat{G}(T_{ij}, T_{il}) + \hat{\sigma}^2\delta_{jl}$ and $\hat{G}(T_{ij}, T_{il}) = \sum_{k=1}^{K} \hat{\lambda}_k \hat{\phi}_k(T_{ij})\hat{\phi}_k(T_{il})$.

When $\hat{\boldsymbol{\Sigma}}_{Y_i}$ is close to numerically singular, the resulting $\hat{\xi}_{ik}$ can be highly unstable due to the difficulty in the inversion. In PACE 2.5, we apply truncation related to ridge regression to stabilize the inversion of $\hat{\boldsymbol{\Sigma}}_{Y_i}$, which leads to stable estimates of $\hat{\xi}_{ik}$. The idea of this approach is to truncate small values of $\hat{\sigma}^2$ used in (1) implicitly and set them equal to a positive threshold value.

The algorithm for the implementation of this approach is as follows:

**Step 1:** Compute residual sum of squares:

$$\hat{\sigma}^2_{new,1} = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{n_i}\sum_{j=1}^{n_i}(Y_{ij} - \hat{Y}_{ij})^2,$$

where $\hat{Y}_{ij} = \hat{\mu}(T_{ij}) + \sum_{k=1}^{K} \hat{\xi}_{ik}\hat{\phi}_k(T_{ij})$. Here, the estimation of $\xi_{ik}$ is based on (1) and $\hat{\sigma}^2$ in $(\hat{\boldsymbol{\Sigma}}_{Y_i})_{j,l}$ is based on (2) in Yao *et al.* (2005).

**Step 2:** Repeat Step 1 and compute a new residual sum of squares:

$$\hat{\sigma}^2_{new,2} = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{n_i}\sum_{j=1}^{n_i}(Y_{ij} - \hat{Y}_{ijnew})^2,$$

where $\hat{Y}_{ijnew}$ differs from $\hat{Y}_{ij}$ in the estimation of $\xi_{ik}$, as it involves $\hat{\sigma}_{new,1}$ in $(\hat{\boldsymbol{\Sigma}}_{Y_i})_{j,l}$ instead of $\hat{\sigma}^2$.

**Step 3:** Reset the value of $\hat{\sigma}^2_{new,2}$ to some $\rho$, if $\hat{\sigma}^2_{new,2} < \rho$. Three choices for this truncation step are avaiable in PACE 2.5:

**Choice** $\rho = -1$**:** compute unadjusted FPC scores the same way as previous PACE versions;

**Choice** $\rho > 0$**:** user-defined choice of $\rho$;

**Choice** $\rho = 0$**:** do not restrict the $\hat{\sigma}^2_{new,2}$. Since $\hat{\sigma}^2_{new,2}$ is always non-negative, when $\rho$ is set to be zero, this corresponds to omitting the truncation step;

**Choice** $\rho =$**'cv':** use randomized leave-one-measurement-out CV approach to find the optimal value of $\rho$ (default choice). The grid of truncation threshold $\rho$ is tied to a measure of the overal signal size $\gamma$, given by

$$\begin{aligned}\gamma &= \left[\mathrm{E}\frac{1}{|\mathcal{T}|}\int_{\mathcal{T}}\{\mu(t)+\sum_{k=1}^{K}\xi_k\phi_k(t)\}^2 dt\right]^{\frac{1}{2}}\\ &= \frac{1}{|\mathcal{T}|^{\frac{1}{2}}}\left[\int_{\mathcal{T}}\{\mu^2(t)+\sum_{k=1}^{K}\lambda_k\phi_k^2(t)\}dt\right]^{\frac{1}{2}}\\ &= \frac{1}{|\mathcal{T}|^{\frac{1}{2}}}\left[\{\int_{\mathcal{T}}\mu^2(t)dt+\sum_{k=1}^{K}\lambda_k\}\right]^{\frac{1}{2}}\end{aligned}$$

Then define $\rho_l = \alpha_l\gamma$ for $l = 1, \ldots, r$, where $\alpha_l$ is some positive constant that is used to create an array of data-dependent candidate choices of $\rho$. In the program, we set $\alpha \in [0.01, 0.225]$, for $r = 50$ equidistant grid points. Define the set $J = \{i : n_i \geq 2\}$ and

$$\rho_{opt} = \operatorname*{argmin}_{\rho_1,\ldots,\rho_r} CV(\rho_l) = \sum_{i\in J}\left(Y_{ij} - \hat{\mu}(T_{ij}) - \sum_{k=1}^{K}\hat{\xi}_{ik}^{(-j)}\hat{\phi}_k(T_{ij})\right)^2.$$

Here, for subject $i$, we randomly select one observation $(T_{ij}, Y_{ij})$ which is then left out from the sample for this subject, and then re-estimate the FPC scores using (1) with the constraint that one resets $\hat{\sigma}^2_{new,2} = \rho_l$ if $\hat{\sigma}^2_{new,2} < \rho_l$.

**Step 4:** Estimate $\xi_{ik}$ using (1) with the updated $\hat{\sigma}^2_{new,2}$ for $(\hat{\mathbf{\Sigma}}_{Y_i})_{j,l}$.

# References

Yao, F., Müller, H.-G. & Wang, J.-L. (2005). Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association* **100**, 577–590.