Functional robust regression for longitudinal data

Daniel Gervini
Department of Mathematical Sciences
University of Wisconsin–Milwaukee

September 8, 2012

Abstract

We present a robust regression estimator for longitudinal data, which is especially suited for functional data that has been observed on sparse or irregular time grids. We show by simulation that the proposed estimators possess good outlier-resistance properties compared with the traditional functional least-squares estimator. As an example of application, we study the relationship between levels of oxides of nitrogen and ozone in the city of San Francisco.

Key Words: Functional data analysis; Longitudinal data analysis; Mixed effects models; Robust statistics; Spline smoothing.

1 Introduction

In a typical longitudinal study, a number of variables are measured on a group of individuals and the goal is to analyze the relationships between the trajectories of the variables. In recent years, functional data analysis has provided efficient ways to analyze longitudinal data. In many cases the variable trajectories are discretized continuous curves that can be reconstructed by smoothing, and functional linear regression methods can be applied to study the relationship between the variables (Ramsay and Silverman, 2005). But in other situations the data is observed at sparse and irregular time points, which makes smoothing difficult or even unfeasible. Therefore, functional regression methods that can be applied directly to the raw measurements become very useful.

Methods for functional data analysis of irregularly sampled curves have been proposed by a number of authors, for the one-sample problem as well as for the functional regression problem (Chiou et al., 2004; James et al., 2000; Müller et al., 2008; Yao et al., 2005a, 2005b). Outlier-resistant techniques for the functional one-sample problem have also been proposed (Cuevas et al., 2007; Gervini, 2008, 2009; Fraiman and Muniz, 2001; Locantore et al., 1999), and two recent papers deal with robust functional regression for pre-smoothed curves (Zhu et al. 2011; Maronna and Yohai, 2012). However, outlier-resistant functional regression methods for raw functional data have not yet been proposed in the literature. In this paper we address this problem and present a computationally simple approach based on random-effect models. Our simulations show that this method attains the desired outlier resistance against atypical curves, and that the asymptotic distribution of the test statistic is approximately valid for small samples.

As an example of application, we will analyze the daily trajectories of oxides of nitrogen and ozone levels in the city of Sacramento, California, during the summer of 2005. The data is shown in Figure 1. The goal is to predict ozone concentration from oxides of nitrogen. Both types of curves follow regular patterns, but some atypical curves can be discerned in the sample. We will show in Section 4 that to a large extend it is indeed possible to predict ozone levels from oxides-of-nitrogen levels, but that the outlying curves distort the classical regression estimators and that the proposed robust method gives more reliable results.

The paper is organized as follows. Section 2 presents a brief overview of functional

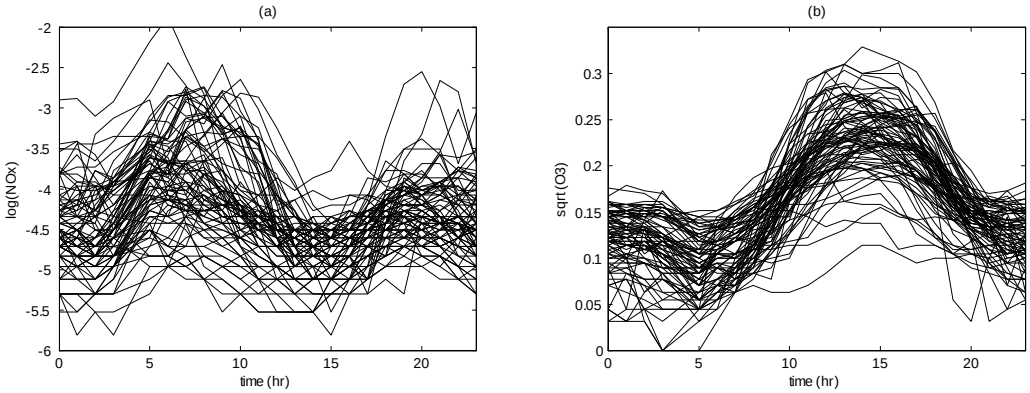

Figure 1: Ozone Example. Daily trajectories of ground-level concentrations of (a) oxides of nitrogen and (b) ozone in the city of Sacramento in the Summer of 2005.

linear regression and introduces the new method. Section 3 reports the results of a comparative simulation study, and Section 4 presents a detailed analysis of the above mentioned ozone dataset. Technical derivations and proofs are left to the Appendix. Matlab programs implementing these procedures are available on the author's webpage.

2 Method

2.1 Background: classical functional linear regression

The functional approach to longitudinal data analysis assumes that the observations $(\mathbf{x}_1, \mathbf{y}_1), \ldots, (\mathbf{x}_n, \mathbf{y}_n)$ are discrete measurements of underlying continuous curves, so

$$x_{ij} = X_i(s_{ij}) + \varepsilon_{ij}, \;\; i = 1, \ldots, n, \;\; j = 1, \ldots, m_i, \tag{1}$$

$$y_{ij} = Y_i(t_{ij}) + \varepsilon'_{ij}, \;\; i = 1, \ldots, n, \;\; j = 1, \ldots, m'_i, \tag{2}$$

where $\{X_i(s)\}$ and $\{Y_i(t)\}$ are the trajectories of interest, $\{\varepsilon_{ij}\}$ and $\{\varepsilon'_{ij}\}$ are random measurement errors, and $\{s_{ij}\}$ and $\{t_{ij}\}$ are the time points where the data is observed. The $X_i(s)$s and the $Y_i(t)$s are random functions that we assume independent and identically distributed realizations of a pair $(X(s), Y(t))$.

Suppose $X(s)$ and $Y(t)$ are square-integrable functions on an interval $[a, b]$. Define the

norm $\|f\| = \{\int_a^b f^2(s)ds\}^{1/2}$ and the inner product $\langle f, g\rangle = \int_a^b f(s)g(s)ds$. If $\mathrm{E}(\|X\|^2)$ and $\mathrm{E}(\|Y\|^2)$ are finite, then $X(s)$ and $Y(t)$ admit the decomposition

$$X(s) = \mu_X(s) + \sum_{k=1}^{p} U_k\phi_k(s), \tag{3}$$

$$Y(t) = \mu_Y(t) + \sum_{l=1}^{q} V_l\psi_l(t), \tag{4}$$

known as the Karhunen–Loève decomposition (Ash and Gardner 1975, ch. 1.4), where $\mu_X(s) = \mathrm{E}\{X(s)\}$, $\mu_Y(t) = \mathrm{E}\{Y(t)\}$, $\{\phi_k(s)\}$ and $\{\psi_l(t)\}$ are orthonormal functions (i.e. $\langle\phi_k, \phi_{k'}\rangle = \delta_{kk'}$ and $\langle\psi_l, \psi_{l'}\rangle = \delta_{ll'}$, where δ is Kronecker's delta), and $\{U_k\}$ and $\{V_l\}$ are random variables with zero mean and finite variance (without loss of generality, one can assume that $\mathrm{var}(U_1) \geq \mathrm{var}(U_2) \geq \cdots > 0$ and $\mathrm{var}(V_1) \geq \mathrm{var}(V_2) \geq \cdots > 0$.) This is the functional equivalent of the principal-component decomposition in multivariate analysis, so the $\phi_k(s)$s and $\psi_l(t)$s are called "principal components", and the U_ks and V_ls are called "component scores". In principle p and q in (3) and (4) could be infinite, but since $\mathrm{E}(\|X - \mu_X\|^2) = \sum_{k=1}^{p} \mathrm{var}(U_k)$ and $\mathrm{E}(\|Y - \mu_Y\|^2) = \sum_{l=1}^{q} \mathrm{var}(V_l)$ are finite, the sequences $\{\mathrm{var}(U_k)\}$ and $\{\mathrm{var}(V_l)\}$ usually decrease to zero fast enough that for practical purposes p and q can be assumed to be finite.

Methods for estimating the mean and the principal components of $X(s)$ and $Y(t)$ can be found in Ramsay and Silverman (2005), James et al. (2000), and Yao et al. (2005b). These methods are not resistant to outliers, though; outlier-resistant estimators of the mean and principal components have been proposed by Locantore et al. (1999), Cuevas et al. (2007), and Gervini (2008, 2009). We will use the method of Gervini (2009) to estimate the mean and the principal components in (3) and (4). This method is briefly reviewed in the Appendix.

Now suppose that there is a functional linear relationship between $X(s)$ and $Y(t)$:

$$Y(t) = \alpha_0(t) + \int_a^b \beta_0(s,t)X(s)ds + Z(t), \tag{5}$$

where $\alpha_0(t)$ is the intercept, $\beta_0(s,t)$ the slope, and $Z(t)$ the error term. We assume $\mathrm{E}\{Z(t)\} = 0$ and $\mathrm{cov}\{X(s), Z(t)\} = 0$ for all s and t. (Note that the Z is not neces-

sarily white noise; it is just the portion of Y that is not explained by X, and it is usually a smooth non-trivial process.) Since (5) implies that $\mu_Y(t) = \alpha_0(t) + \int_a^b \beta_0(s,t)\mu_X(s)ds$, we can rewrite (5) as

$$Y(t) = \mu_Y(t) + \int_a^b \beta_0(s,t)\{X(s) - \mu_X(s)\}\mathrm{ds} + Z(t). \tag{6}$$

Then the only parameter that remains to be estimated is the regression slope β_0.

Since $\{\phi_k\}$ is an orthonormal basis of the X-space and $\{\psi_l\}$ is an orthonormal basis of the Y-space, without loss of generality the regression slope can be expressed as

$$\beta_0(s,t) = \sum_{k=1}^{p}\sum_{l=1}^{q} \theta_{0kl}\phi_k(s)\psi_l(t). \tag{7}$$

In matrix form, $\beta_0(s,t) = \boldsymbol{\phi}(s)^T\boldsymbol{\Theta}_0\boldsymbol{\psi}(t)$, where $\boldsymbol{\phi}(s) = (\phi_1(s),\ldots,\phi_p(s))^T$ and $\boldsymbol{\psi}(t) = (\psi_1(t),\ldots,\psi_q(t))^T$. If we also collect the component scores $\{U_k\}$ and $\{V_l\}$ into vectors $\mathbf{U} \in \mathbb{R}^p$ and $\mathbf{V} \in \mathbb{R}^q$, from (3), (4), (6) and (7) we obtain

$$\begin{aligned} \boldsymbol{\psi}(t)^T\mathbf{V} &= \int_a^b \boldsymbol{\psi}(t)^T\boldsymbol{\Theta}_0^T\boldsymbol{\phi}(s)\boldsymbol{\phi}(s)^T\mathbf{U}\ \mathrm{ds} + Z(t) \\ &= \boldsymbol{\psi}(t)^T\boldsymbol{\Theta}_0^T\mathbf{U} + \boldsymbol{\psi}(t)^T\mathbf{W}, \end{aligned}$$

where $\mathbf{W} \in \mathbb{R}^q$ is the random vector with elements $W_l = \langle Z, \psi_l\rangle$. This reduces the functional regression model (6) to a simpler multivariate regression model,

$$\mathbf{V} = \boldsymbol{\Theta}_0^T\mathbf{U} + \mathbf{W}, \tag{8}$$

and the problem now is to estimate the regression matrix $\boldsymbol{\Theta}_0$.

2.2 Outlier-resistant functional regression

As explained above, given the data $(\mathbf{x}_1,\mathbf{y}_1),\ldots,(\mathbf{x}_n,\mathbf{y}_n)$ we use the reduced-rank t estimators of Gervini (2009) to obtain robust estimators of μ_X, μ_Y, $\{\phi_k\}$, $\{\psi_l\}$, $\{U_{ik}\}$ and

$\{V_{il}\}$. By (7) and (8), the least-squares estimator of $\beta_0(s,t)$ would be $\boldsymbol{\phi}(s)^T\hat{\boldsymbol{\Theta}}\boldsymbol{\psi}(t)$ with

$$\hat{\boldsymbol{\Theta}} = \underset{\boldsymbol{\Theta}}{\operatorname{argmin}} \sum_{i=1}^{n} \|\hat{\mathbf{V}}_i - \boldsymbol{\Theta}^T\hat{\mathbf{U}}_i\|^2 = (\sum_{i=1}^{n} \hat{\mathbf{U}}_i\hat{\mathbf{U}}_i^T)^{-1} \sum_{i=1}^{n} \hat{\mathbf{U}}_i\hat{\mathbf{V}}_i^T. \tag{9}$$

However, this estimator is not robust. Although the reduced-rank t estimators of μ_X, μ_Y, $\{\phi_k\}$ and $\{\psi_l\}$ are robust, the component scores $\hat{\mathbf{U}}_i$ and $\hat{\mathbf{V}}_i$ are individual parameters that will be outliers if the corresponding curves $X_i(s)$ and $Y_i(t)$ are outliers. Therefore, the estimator of $\boldsymbol{\Theta}_0$ has to incorporate a mechanism to downweight outlying $\hat{\mathbf{U}}_i$s and $\hat{\mathbf{V}}_i$s.

This can be accomplished, for instance, by a modification of the t-type GM-estimators of He et al. (2000), that we will call GMt for short. Let

$$(\hat{\boldsymbol{\Theta}}, \hat{\boldsymbol{\Sigma}}) = \underset{\boldsymbol{\Theta},\boldsymbol{\Sigma}}{\operatorname{argmin}} \sum_{i=1}^{n} \rho\{w(\hat{\mathbf{U}}_i)(\hat{\mathbf{V}}_i - \boldsymbol{\Theta}^T\hat{\mathbf{U}}_i)^T\boldsymbol{\Sigma}^{-1}(\hat{\mathbf{V}}_i - \boldsymbol{\Theta}^T\hat{\mathbf{U}}_i)\} + n\log|\boldsymbol{\Sigma}|, \tag{10}$$

where $\rho(x) = (\nu+q)\log(1+x/\nu)$. These are the maximum likelihood estimators of $\boldsymbol{\Theta}_0$ and $\boldsymbol{\Sigma}_0$ when $\mathbf{W}$ in (8) follows a multivariate t distribution with mean zero and scatter matrix $\boldsymbol{\Sigma}_0/w(\hat{\mathbf{U}}_i)$, although we do not actually assume that $\mathbf{W}$ follows this distribution; as in He et al. (2000), this is just the motivation behind definition (10).

It is shown in the Appendix that $\hat{\boldsymbol{\Theta}}$ and $\hat{\boldsymbol{\Sigma}}$ satisfy the fixed-point equations

$$\hat{\boldsymbol{\Theta}} = \left\{\sum_{i=1}^{n} \rho'(e_i)w(\hat{\mathbf{U}}_i)\hat{\mathbf{U}}_i\hat{\mathbf{U}}_i^T\right\}^{-1} \sum_{i=1}^{n} \rho'(e_i)w(\hat{\mathbf{U}}_i)\hat{\mathbf{U}}_i\hat{\mathbf{V}}_i^T, \tag{11}$$

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{n}\sum_{i=1}^{n} \rho'(e_i)w(\hat{\mathbf{U}}_i)\mathbf{R}_i\mathbf{R}_i^T, \tag{12}$$

where $\mathbf{R}_i = \hat{\mathbf{V}}_i - \hat{\boldsymbol{\Theta}}^T\hat{\mathbf{U}}_i$ and $e_i = w(\hat{\mathbf{U}}_i)\mathbf{R}_i^T\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{R}_i$. These equations can be solved iteratively by a reweighting algorithm.

As for the weights $w(\hat{\mathbf{U}}_i)$, they are essentially a by-product of the estimation of μ_X, $\{\phi_k\}$ and $\{\mathbf{U}_i\}$. Since $E(\mathbf{U}_i) = \mathbf{0}$ and $\operatorname{var}(\mathbf{U}_i) = \operatorname{diag}(\lambda_1,\ldots,\lambda_p)$, the $\hat{\mathbf{U}}_i$s are approximately uncorrelated with mean zero. The squared Mahalanobis distance of $\hat{\mathbf{U}}_i$ is then $D_i^2 = \sum_{k=1}^{p} \hat{U}_{ik}^2/\hat{\lambda}_k$, and large D_i^2s will correspond to X-outliers. The D_i^2s will follow an approximate χ_p^2 distribution if the data is Gaussian.

This suggests a number of weighting schemes. One possibility is to use "metric" trim-

ming,

$$w(\hat{\mathbf{U}}_i) = \begin{cases} 1, & D_i^2 \leq \chi^2_{p,1-\alpha}, \\ 0, & \text{otherwise,} \end{cases} \tag{13}$$

where $\chi^2_{p,1-\alpha}$ is the $1-\alpha$ quantile of the χ^2_p distribution. Another possibility is to use rank-based trimming,

$$w(\hat{\mathbf{U}}_i) = \begin{cases} 1, & \mathrm{rank}(D_i^2)/n \leq 1-\alpha, \\ 0, & \text{otherwise.} \end{cases} \tag{14}$$

The latter will always eliminate the αn observations with largest Mahalanobis distances, even if they are not actual outliers; so we recommend not using an unnecessarily large α for rank-based trimming. In practice, the choice of α can be based on the proportion of outliers observed in a boxplot or histogram of the D_i^2s.

The estimator $\hat{\mathbf{\Theta}}$ defined above belongs to the general class of M-estimators, which have well-known asymptotic properties (Van der Vaart, 1998, ch. 5). As shown in the Appendix, $\sqrt{n}\{\mathrm{vec}(\hat{\mathbf{\Theta}}) - \mathrm{vec}(\mathbf{\Theta}_0)\}$ follows an approximate $\mathrm{N}(\mathbf{0}, \mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1})$ distribution for large n, with

$$\mathbf{A} = 2\mathrm{E}\left\{\rho''(e)w^2(\mathbf{U})\mathbf{\Sigma}_0^{-1}\mathbf{R}\mathbf{R}^T \otimes \mathbf{U}\mathbf{U}^T\right\} + \mathbf{I}_q \otimes \mathrm{E}\left\{\rho'(e)w(\mathbf{U})\mathbf{U}\mathbf{U}^T\right\}, \tag{15}$$

$$\mathbf{B} = \mathrm{E}\left[\{\rho'(e)\}^2 w^2(\mathbf{U})\mathbf{R}\mathbf{R}^T \otimes \mathbf{U}\mathbf{U}^T\right]. \tag{16}$$

The matrices $\mathbf{A}$ and $\mathbf{B}$ can be easily estimated, replacing expectations by averages. This asymptotic distribution can be used, for instance, to test significance of the regression: if $\mathbf{\Theta}_0 = \mathbf{O}$, Wald's statistic $Q = n\mathrm{vec}(\hat{\mathbf{\Theta}})^T \hat{\mathbf{A}}\hat{\mathbf{B}}^{-1}\hat{\mathbf{A}}\mathrm{vec}(\hat{\mathbf{\Theta}})$ follows an approximate χ^2_{pq} distribution for large n, so we decide the regression is significant if $Q \geq \chi^2_{pq,1-\alpha}$ for a given level α. We can also construct marginal tests and confidence intervals for the individual coefficients θ_{kl}.

In Section 3 we will study the accuracy of this asymptotic approximation. It is our experience that the distribution of $\hat{\mathbf{\Theta}}$ approaches normality quite fast, but the above "sandwich formula" tends to underestimate the variance when the sample size n is small. In that case it is better to use bootstrap estimators of the covariance matrix of $\mathrm{vec}(\hat{\mathbf{\Theta}})$.

3 Simulations

In this section we study by simulation the finite-sample behavior of the estimators (10). To this end, we generated data from model (8) with $\mathbf{U} \sim \mathrm{N}(\mathbf{0}, \mathbf{\Lambda})$ and $\mathbf{W} \sim \mathrm{N}(\mathbf{0}, \mathbf{\Sigma})$, where $\mathbf{\Lambda} = \mathrm{diag}(1, 1/2, \ldots, 1/p)$ and $\mathbf{\Sigma} = \mathrm{diag}(1, 1/2, \ldots, 1/q)$. Two regression parameters $\mathbf{\Theta}_0$ were considered: for the first set of simulations (to study estimation error) we took $\mathbf{\Theta}_0$ with $\theta_{0,11} = 3$ and $\theta_{0,ij} = 0$ for $(i, j) \neq (1, 1)$; for the second set of simulations (to study the goodness of the asymptotic approximation of Wald's test) we took $\mathbf{\Theta}_0 = \mathbf{O}$. The curves $\{X_i(s)\}$ and $\{Y_i(t)\}$ were generated following (3) and (4), with $\mu_X(s)$ and $\mu_Y(t)$ equal to zero, $\phi_k(s) = \sqrt{2}\sin(k\pi s)$ and $\psi_l(t) = \sqrt{2}\sin(l\pi t)$, for s and t in $[0, 1]$. The raw observations were generated following (1) and (2), with random s_{ij}s uniformly distributed in $[0, 1]$, $\{\varepsilon_{ij}\}$ and $\{\varepsilon'_{ij}\}$ independent $\mathrm{N}(0, 0.01)$, and $m_i = m'_i = m$; for simplicity we took the grid $\{t_{ij}\}$ equal to $\{s_{ij}\}$.

The first series of simulations were designed to study estimation error of the $\hat{\mathbf{\Theta}}$s, both for clean and for outlier-contaminated data. We generated outliers by replacing $[\varepsilon n]$ of the pairs $(\mathbf{U}_i, \mathbf{V}_i)$ by $(\mathbf{U}^*_i, \mathbf{V}^*_i)$, with $U^*_{i1} = U_{i1} + 5$ and $U^*_{ij} = U_{ij}$ for $j \neq 1$, and $\mathbf{V}^*_i = \mathbf{W}_i$. Note that the contaminated data $(\mathbf{U}^*_i, \mathbf{V}^*_i)$ follows model (8) with $\mathbf{\Theta}_0 = \mathbf{O}$ and high-leverage $\mathbf{U}^*_i$s, so the effect of this type of contamination is an underestimation of $\theta_{0,11}$ that tends to pull $\hat{\beta}(s, t)$ towards 0.

The estimation of $\mathbf{\Theta}_0$ requires two steps: first, to estimate $\{\mathbf{U}_i\}$ and $\{\mathbf{V}_i\}$ from the raw data, and then to compute $\hat{\mathbf{\Theta}}$ from the $\hat{\mathbf{U}}_i$s and the $\hat{\mathbf{V}}_i$s. So we compared two procedures: a non-robust procedure, using reduced-rank Normal models (James et al., 2000) to estimate the component scores, followed by the ordinary least-squares regression estimator (9); and a robust procedure, using reduced-rank t-models (Gervini, 2009) to estimate the component scores, followed by the GMt regression estimator (10). For the robust procedure, we considered the two types of weights $w(\hat{\mathbf{U}}_i)$ discussed in Section 2.2, with trimming proportions $\alpha = .10$ and $\alpha = .50$; degrees of freedom $\nu = 1$ and $\nu = 5$ were used for the t-models.

Four levels of contamination ε were considered: 0 (clean data), .10, .20 and .30. We took $n = 50$ as sample size, $m = 20$ as grid size, and $p = q = 2$ as model dimensions. Each case was replicated 1000 times. As measure of the estimation error we used the expected root integrated squared error $\mathrm{E}(\|\hat{\beta} - \beta_0\|)$, where $\|\hat{\beta} - \beta_0\|^2 = \int_0^1 \int_0^1 \{\hat{\beta}(s, t) -$

$\beta_0(s,t)\}^2\, ds\, dt$.

The results are reported in Table 1, along with Monte Carlo standard errors. We see that for non-contaminated data ($\varepsilon = 0$), there is no significant difference between metric and rank trimming for a given pair (ν, α). The trimming proportion α has a larger impact on the estimator's behavior than the degrees of freedom ν. For this reason we recommend choosing α adaptively, so as not to cut off too much good data. When $\varepsilon > 0$, we see that metric trimming tends to outperform rank trimming for a given pair (ν, α). Somewhat counterintuitively, estimators with $\nu = 5$ tend to be more robust than those with $\nu = 1$ for a given α; the reason is that for this type of contamination, which affects $\hat{\mathbf{\Theta}}$ but not the $\hat{\phi}_k$s or the $\hat{\psi}_l$s, t models with $\nu = 5$ provide more accurate estimators of $\{\mathbf{U}_i\}$ and $\{\mathbf{V}_i\}$ than t models with $\nu = 1$ (for other types of contamination this is no longer true, although t models with $\nu = 5$ are still very robust; see Gervini (2009).) In general, then, the recommendation is to use t-model estimators with metrically trimmed weights and a trimming proportion chosen adaptively.

The second series of simulations were designed to assess the finite-sample adequacy of the asymptotic Wald test. To this end we generated data as before, but with $\mathbf{\Theta}_0 = \mathbf{O}$. Then $Q = n\mathrm{vec}(\hat{\mathbf{\Theta}})^T \hat{\mathbf{\Omega}}^{-1} \mathrm{vec}(\hat{\mathbf{\Theta}})$ should approximately follow a χ^2_{pq} distribution, where $\mathbf{\Omega}$ is the asymptotic covariance matrix of $\sqrt{n}\mathrm{vec}(\hat{\mathbf{\Theta}})$. For GMt estimators, $\mathbf{\Omega}$ is the "sandwich formula" given in Section 2.2; for the least-squares estimator, $\mathbf{\Omega} = \mathrm{E}(\mathbf{R}\mathbf{R}^T) \otimes \{\mathrm{E}(\mathbf{U}\mathbf{U}^T)\}^{-1}$. Table 2 reports the tail probabilities $\mathrm{P}(Q \geq \chi^2_{pq,1-\alpha})$ for the usual values of α (.10, .05 and .01) and various combinations of parameters n, m, p and q. Each combination was replicated 10,000 times. We compared only two estimators this time: the least-squares estimator and the 10% metrically trimmed GMt estimator with $\nu = 5$. We see in Table 2 that the asymptotic χ^2_{pq} approximation works reasonably well for the least-squares estimator if the ratio n/pq exceeds 15; however, for the GMt estimator a ratio n/pq of at least 35 is necessary for the asymptotic approximation to be reasonably good. Therefore, the asymptotic Wald test can be used with confidence only for large sample sizes and relatively small dimensions. In other cases, permutation tests or Wald tests with bootstrap-estimated covariances are preferable.

Estimator	Contamination proportion			
	0%	10%	20%	30%
Least squares	.293 (.004)	2.241 (.006)	2.644 (.048)	2.731 (.007)
GMt, $\nu = 1$, $\alpha = .10$				
Metric trim	.472 (.006)	.497 (.007)	1.316 (.028)	2.924 (.008)
Rank trim	.473 (.006)	.469 (.007)	2.246 (.028)	2.941 (.006)
GMt, $\nu = 1$, $\alpha = .50$				
Metric trim	.846 (.012)	.800 (.013)	1.112 (.018)	1.756 (.022)
Rank trim	.832 (.012)	.922 (.015)	1.212 (.018)	1.784 (.021)
GMt, $\nu = 5$, $\alpha = .10$				
Metric trim	.379 (.005)	.396 (.005)	1.493 (.023)	2.746 (.006)
Rank trim	.374 (.005)	.395 (.006)	2.341 (.011)	2.792 (.005)
GMt, $\nu = 5$, $\alpha = .50$				
Metric trim	.795 (.010)	.666 (.011)	.912 (.015)	1.494 (.021)
Rank trim	.783 (.010)	.829 (.013)	1.054 (.017)	1.506 (.021)

Table 1: Simulation Results. Mean root integrated squared errors of $\hat{\beta}$ under various contamination proportions (Monte Carlo standard errors in parenthesis).

Parameters	Estimator	Nominal probability		
		.10	.05	.01
$n = 50$, $m = 20$,	LS	.1426 (.0035)	.0819 (.0027)	.0219 (.0015)
$p = q = 2$	GMt	.2270 (.0042)	.1571 (.0036)	.0749 (.0026)
$n = 100$, $m = 20$,	LS	.1272 (.0033)	.0693 (.0025)	.0170 (.0013)
$p = q = 2$	GMt	.1584 (.0037)	.0952 (.0029)	.0366 (.0019)
$n = 150$, $m = 10$,	LS	.1117 (.0032)	.0561 (.0023)	.0123 (.0011)
$p = q = 2$	GMt	.1392 (.0035)	.0824 (.0027)	.0258 (.0016)
$n = 100$, $m = 20$,	LS	.1452 (.0035)	.0813 (.0027)	.0211 (.0014)
$p = q = 3$	GMt	.2750 (.0045)	.1900 (.0039)	.0875 (.0028)
$n = 150$, $m = 20$,	LS	.1316 (.0034)	.0718 (.0026)	.0144 (.0012)
$p = q = 3$	GMt	.2111 (.0041)	.1360 (.0034)	.0514 (.0022)
$n = 200$, $m = 10$,	LS	.1185 (.0032)	.0625 (.0024)	.0169 (.0013)
$p = q = 3$	GMt	.1782 (.0038)	.1122 (.0032)	.0391 (.0019)

Table 2: Simulation Results. Finite-sample tail probabilities of Wald's significance-of-regression test for nominal asymptotic probabilities .10, .05 and .01 (Monte Carlo standard errors in parenthesis).

4 Application: Ozone Pollution Data

Ground-level ozone is an air pollutant known to cause serious health problems. Unlike other pollutants, ozone is not emitted directly into the air but forms as a result of complex chemical reactions, including volatile organic compounds and oxides of nitrogen among other factors. Modeling ground-level ozone formation has been an active topic of air-quality studies for many years. The California Environmental Protection Agency database, available at http://www.arb.ca.gov/aqd/aqdcd/aqdcddld.htm, has collected data on hourly concentrations of pollutants at different locations in California for the years 1980 to 2009. Here we will focus on the trajectories of oxides of nitrogen (NOx) and ozone (O3) in the city of Sacramento (site 3011 in the database) between June 6 and August 26 of 2005, which make a total of 82 days (shown in Figure 1). There are a few days with some missing observations (9 in total), but since the method can handle unequal time grids, imputation of the missing data was not necessary.

The first step in the analysis is to fit reduced-rank models to the sample curves. We used cubic B-splines with 7 equally spaced knots every 5 years, and fitted Normal and t_1 (Cauchy) reduced-rank models with up to 10 principal components. For both the response and the explanatory curves, the leading three components explain at least 85% of the total variability, so we retained these models. The means and the principal components are plotted in Figure 2. There is no substantial difference between the estimators obtained by these models, except perhaps for the mean and the third component of log-NOx (Figures 2 (a) and (g)).

With the Normal component scores we computed the Least Squares estimator, obtaining

$$\hat{\boldsymbol{\Theta}}_{LS} = \begin{pmatrix} .0404 & -.0077 & .0083 \\ -.0537 & -.0085 & .0317 \\ -.0109 & -.0173 & -.0263 \end{pmatrix}.$$

With the Cauchy component scores we computed the GMt estimator with 1 degree of

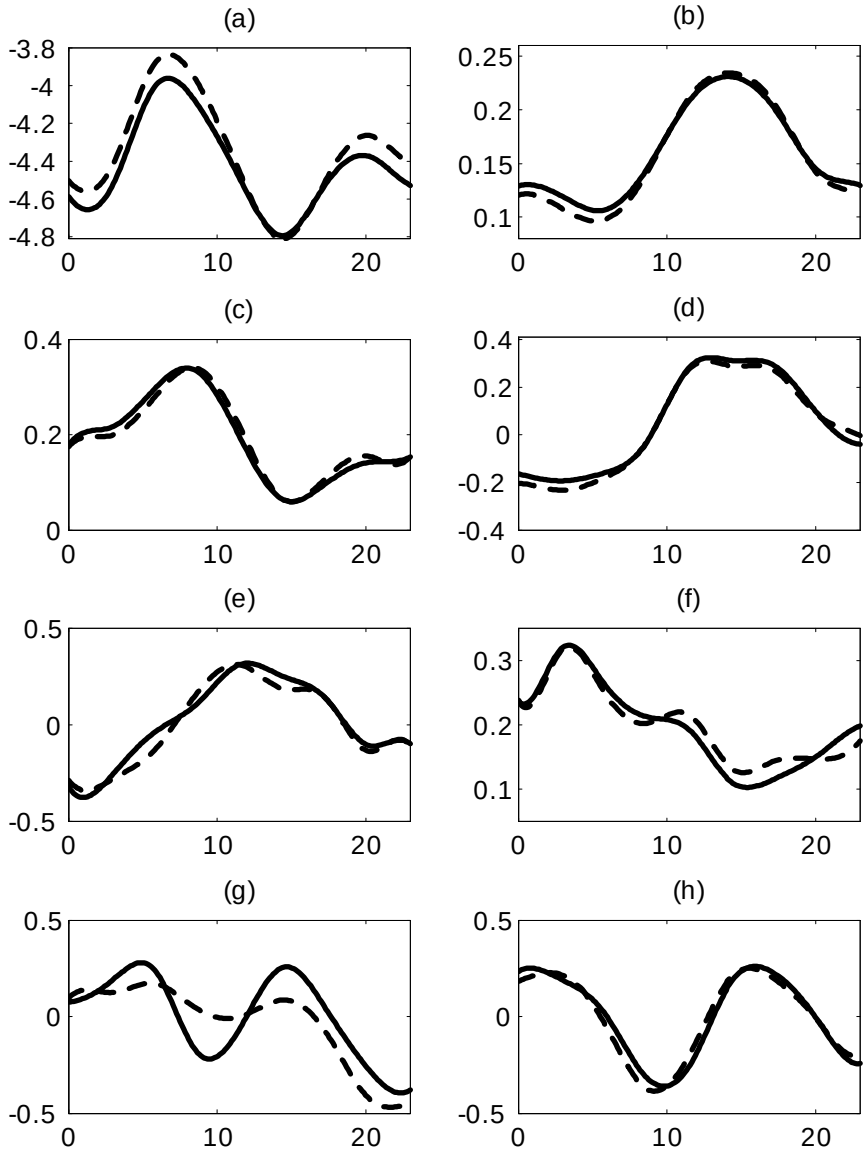

Figure 2: Ozone Example. Normal (– – –) and Cauchy (——) reduced-rank B-spline estimators of the mean [(a),(b)], the first principal component [(c),(d)], the second principal component [(e),(f)] and the third principal component [(g),(h)] of log-NOx and root-O3 trajectories.

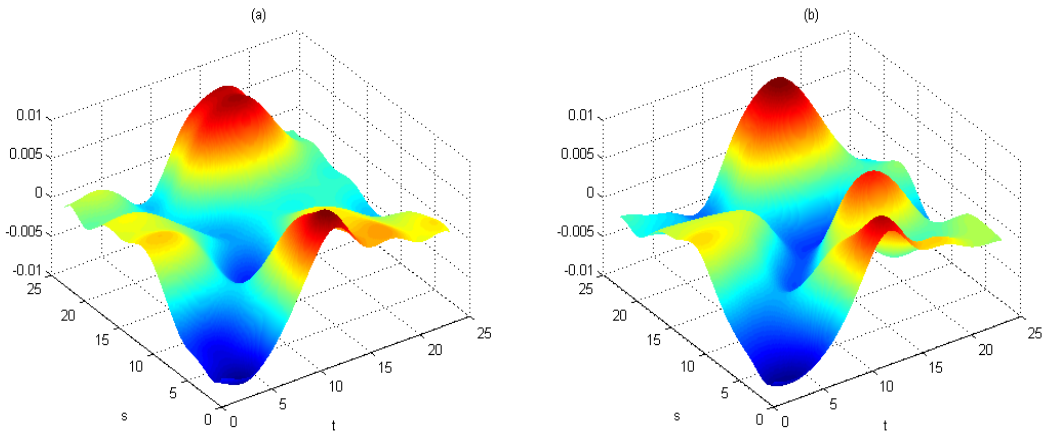

Figure 3: Ozone Example. Functional slope estimators obtained by (a) least squares using Normal scores and (b) metric-trimmed GMt using Cauchy scores.

freedom and 10% metric trimming, obtaining

$$\hat{\mathbf{\Theta}}_{GM} = \begin{pmatrix} .0406 & -.0172 & .0045 \\ -.0451 & .0029 & .0266 \\ -.0289 & -.0071 & -.0317 \end{pmatrix}.$$

The latter cut off 5 observations out of the 82. There are some noticeable differences between these two estimators, even leaving aside the third row (which are not easily comparable, since $\hat{\phi}_{LS,3}(s)$ and $\hat{\phi}_{GM,3}(s)$ are rather different). The differences are more striking in the slope estimators $\hat{\beta}_{LS}(s,t)$ and $\hat{\beta}_{GM}(s,t)$, shown in Figure 3. There is a "bump" in $\hat{\beta}_{GM}(s,t)$ around $(s,t) = (8,16)$ that does not appear in $\hat{\beta}_{LS}(s,t)$. This means that the robust slope estimator assigns positive weight to NOx values around 8am in the prediction of O3 levels around 4pm, showing that there is a persistent effect of oxides-of-nitrogen level in ozone formation.

Of course, none of this would be meaningful if the regression model was not statistically significant. But the estimated response curves, shown in Figure 4, clearly show that the model does predict the response curves to a large extent. The robust estimator provides a better fit overall, with a root median squared error of $.022$ compared to the root median squared error of $.023$ for the least squares estimator.

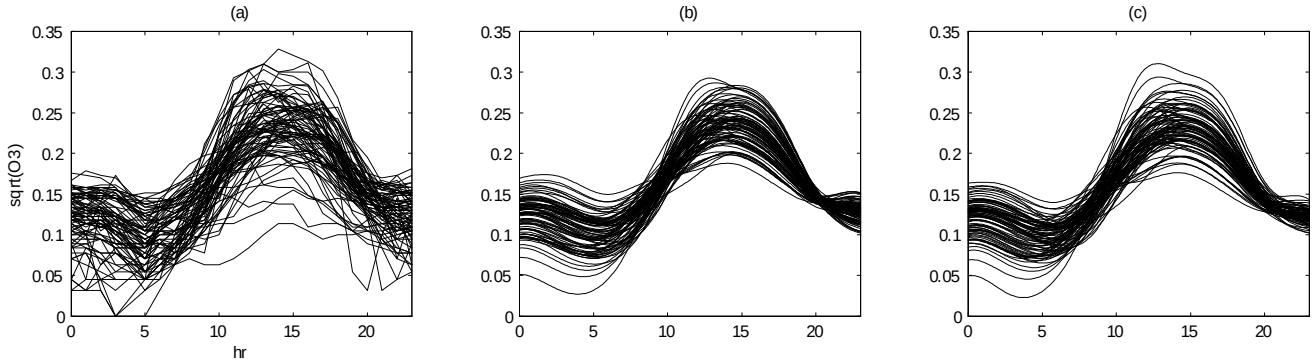

Figure 4: Ozone Example. Daily trajectories of root-O3 levels: (a) observed, (b) predicted by robust GMt estimator, and (c) predicted by least squares.

Acknowledgement

The author was partly supported by NSF grants DMS 0604396 and 1006281.

Appendix

Reduced-rank t models

The method proposed by Gervini (2009) to estimate the mean and the principal components of a stochastic process X works as follows. The mean function μ_X and the principal components $\{\phi_k\}$ are modeled as spline functions; that is, given a set of spline basis functions $b_1, \ldots, b_N$, chosen by the user, it is assumed that $\mu_X(s) = \sum_{l=1}^{N} \xi_l b_l(s)$ and $\phi_k(s) = \sum_{l=1}^{N} \eta_{kl} b_l(s)$. The observed vector $\mathbf{x}_i$ can then be expressed as

$$\mathbf{x}_i = \mathbf{B}_i\boldsymbol{\xi} + \mathbf{B}_i\mathbf{H}\boldsymbol{\Lambda}^{1/2}\mathbf{z}_i + \sigma\boldsymbol{\varepsilon}_i,$$

where $\mathbf{B}_i = [b_l(s_{ij})]_{(j,l)}$, $\mathbf{H} = [\boldsymbol{\eta}_1, \ldots, \boldsymbol{\eta}_p]$ and $\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_p)$. Note that $\mathbf{U}_i = \boldsymbol{\Lambda}^{1/2}\mathbf{z}_i$ in this notation. By assuming $(\mathbf{z}_i, \boldsymbol{\varepsilon}_i)$ has a standard multivariate t distribution, robust maximum likelihood estimators of $\boldsymbol{\xi}$, $\{\boldsymbol{\eta}_k\}$, $\{\lambda_k\}$ and σ are obtained. The estimators are computed via a standard EM algorithm. The optimal number of components p can be chosen via AIC or BIC criteria. See Gervini (2009) for details. In addition to parame-

ter estimates, the EM algorithm yields predictors of the random effects $\mathbf{z}_i$, so one obtains $\hat{\mathbf{U}}_i = \hat{\mathbf{\Lambda}}^{1/2}\hat{\mathbf{z}}_i$ as a by-product. The estimators of μ_Y, $\{\psi_k\}$, and $\{\mathbf{V}_i\}$ are obtained in a similar way from the sample $\mathbf{y}_1, \ldots, \mathbf{y}_n$.

GMt estimating equations and asymptotics

The estimators $\hat{\mathbf{\Theta}}$ and $\hat{\mathbf{\Sigma}}$ defined by (10) are M-type estimators (Van der Vaart, 1998, ch. 5), since they minimize a function of the form $M(\mathbf{\Theta}, \mathbf{\Sigma}) = \frac{1}{n}\sum_{i=1}^{n} m_{(\mathbf{\Theta},\mathbf{\Sigma})}(\hat{\mathbf{U}}_i, \hat{\mathbf{V}}_i)$. Specifically,

$$m_{(\mathbf{\Theta},\mathbf{\Sigma})}(\hat{\mathbf{U}}_i, \hat{\mathbf{V}}_i) = \rho\{w(\hat{\mathbf{U}}_i)(\hat{\mathbf{V}}_i - \mathbf{\Theta}^T\hat{\mathbf{U}}_i)^T\mathbf{\Sigma}^{-1}(\hat{\mathbf{V}}_i - \mathbf{\Theta}^T\hat{\mathbf{U}}_i)\} + \log|\mathbf{\Sigma}|\,.$$

Then $\hat{\mathbf{\Theta}}$ and $\hat{\mathbf{\Sigma}}$ solve the equations $\frac{\partial}{\partial\mathbf{\Theta}}M(\hat{\mathbf{\Theta}}, \hat{\mathbf{\Sigma}}) = \mathbf{O}$ and $\frac{\partial}{\partial\mathbf{\Sigma}}M(\hat{\mathbf{\Theta}}, \hat{\mathbf{\Sigma}}) = \mathbf{O}$. To compute matrix derivatives we use the method of differentials (Magnus and Neudecker, 1999). Differentiating with respect to $\mathbf{\Theta}$ we obtain

$$\begin{aligned}
\mathrm{d}m_{(\mathbf{\Theta},\mathbf{\Sigma})}(\hat{\mathbf{U}}_i, \hat{\mathbf{V}}_i) &= \rho'(e_i)w(\hat{\mathbf{U}}_i)2(\hat{\mathbf{V}}_i - \mathbf{\Theta}^T\hat{\mathbf{U}}_i)^T\mathbf{\Sigma}^{-1}\{-(\mathrm{d}\mathbf{\Theta})^T\hat{\mathbf{U}}_i\} \\
&= -2\rho'(e_i)w(\hat{\mathbf{U}}_i)\mathrm{tr}\{(\mathrm{d}\mathbf{\Theta})^T\hat{\mathbf{U}}_i(\hat{\mathbf{V}}_i - \mathbf{\Theta}^T\hat{\mathbf{U}}_i)^T\mathbf{\Sigma}^{-1}\} \\
&= -2\rho'(e_i)w(\hat{\mathbf{U}}_i)\mathrm{vec}(\mathrm{d}\mathbf{\Theta})^T\mathrm{vec}\{\hat{\mathbf{U}}_i(\hat{\mathbf{V}}_i - \mathbf{\Theta}^T\hat{\mathbf{U}}_i)^T\mathbf{\Sigma}^{-1}\},
\end{aligned}$$

where $e_i = w(\hat{\mathbf{U}}_i)(\hat{\mathbf{V}}_i - \mathbf{\Theta}^T\hat{\mathbf{U}}_i)^T\mathbf{\Sigma}^{-1}(\hat{\mathbf{V}}_i - \mathbf{\Theta}^T\hat{\mathbf{U}}_i)$. Then

$$\nabla_{\mathrm{vec}(\mathbf{\Theta})}m_{(\mathbf{\Theta},\mathbf{\Sigma})}(\hat{\mathbf{U}}_i, \hat{\mathbf{V}}_i) = -2\rho'(e_i)w(\hat{\mathbf{U}}_i)\mathrm{vec}\{\hat{\mathbf{U}}_i(\hat{\mathbf{V}}_i - \mathbf{\Theta}^T\hat{\mathbf{U}}_i)^T\mathbf{\Sigma}^{-1}\}, \qquad (17)$$

which can be rearranged in matrix form as

$$\frac{\partial}{\partial\mathbf{\Theta}}m_{(\mathbf{\Theta},\mathbf{\Sigma})}(\hat{\mathbf{U}}_i, \hat{\mathbf{V}}_i) = -2\rho'(e_i)w(\hat{\mathbf{U}}_i)\hat{\mathbf{U}}_i(\hat{\mathbf{V}}_i - \mathbf{\Theta}^T\hat{\mathbf{U}}_i)^T\mathbf{\Sigma}^{-1},$$

and (11) follows. Differentiating m with respect to $\mathbf{\Sigma}$ we obtain

$$\mathrm{d}m_{(\mathbf{\Theta},\mathbf{\Sigma})}(\hat{\mathbf{U}}_i, \hat{\mathbf{V}}_i) =$$

$$
\begin{aligned}
&= \rho'(e_i) w(\hat{\mathbf{U}}_i)(\hat{\mathbf{V}}_i - \boldsymbol{\Theta}^T\hat{\mathbf{U}}_i)^T\{-\boldsymbol{\Sigma}^{-1}(\mathrm{d}\boldsymbol{\Sigma})\boldsymbol{\Sigma}^{-1}\}(\hat{\mathbf{V}}_i - \boldsymbol{\Theta}^T\hat{\mathbf{U}}_i) + \mathrm{tr}\{\boldsymbol{\Sigma}^{-1}(\mathrm{d}\boldsymbol{\Sigma})\} \\
&= -\rho'(e_i) w(\hat{\mathbf{U}}_i)\mathrm{tr}\{\boldsymbol{\Sigma}^{-1}(\hat{\mathbf{V}}_i - \boldsymbol{\Theta}^T\hat{\mathbf{U}}_i)(\hat{\mathbf{V}}_i - \boldsymbol{\Theta}^T\hat{\mathbf{U}}_i)^T\boldsymbol{\Sigma}^{-1}(\mathrm{d}\boldsymbol{\Sigma})\} + \mathrm{tr}\{\boldsymbol{\Sigma}^{-1}(\mathrm{d}\boldsymbol{\Sigma})\} \\
&= -\rho'(e_i) w(\hat{\mathbf{U}}_i)\mathrm{vec}\{\boldsymbol{\Sigma}^{-1}(\hat{\mathbf{V}}_i - \boldsymbol{\Theta}^T\hat{\mathbf{U}}_i)(\hat{\mathbf{V}}_i - \boldsymbol{\Theta}^T\hat{\mathbf{U}}_i)^T\boldsymbol{\Sigma}^{-1}\}^T\mathrm{vec}(\mathrm{d}\boldsymbol{\Sigma}) \\
&\quad + \mathrm{vec}(\boldsymbol{\Sigma}^{-1})^T\mathrm{vec}(\mathrm{d}\boldsymbol{\Sigma}),
\end{aligned}
$$

so

$$
\begin{aligned}
&\nabla_{\mathrm{vec}(\boldsymbol{\Sigma})} m_{(\boldsymbol{\Theta},\boldsymbol{\Sigma})}(\hat{\mathbf{U}}_i, \hat{\mathbf{V}}_i) \\
&= -\rho'(e_i) w(\hat{\mathbf{U}}_i)\mathrm{vec}\{\boldsymbol{\Sigma}^{-1}(\hat{\mathbf{V}}_i - \boldsymbol{\Theta}^T\hat{\mathbf{U}}_i)(\hat{\mathbf{V}}_i - \boldsymbol{\Theta}^T\hat{\mathbf{U}}_i)^T\boldsymbol{\Sigma}^{-1}\} + \mathrm{vec}(\boldsymbol{\Sigma}^{-1}).
\end{aligned}
$$

Again, this can be expressed in matrix form as

$$
\begin{aligned}
&\frac{\partial}{\partial \boldsymbol{\Sigma}} m_{(\boldsymbol{\Theta},\boldsymbol{\Sigma})}(\hat{\mathbf{U}}_i, \hat{\mathbf{V}}_i) \\
&= -\rho'(e_i) w(\hat{\mathbf{U}}_i)\boldsymbol{\Sigma}^{-1}(\hat{\mathbf{V}}_i - \boldsymbol{\Theta}^T\hat{\mathbf{U}}_i)(\hat{\mathbf{V}}_i - \boldsymbol{\Theta}^T\hat{\mathbf{U}}_i)^T\boldsymbol{\Sigma}^{-1} + \boldsymbol{\Sigma}^{-1},
\end{aligned}
$$

from which (12) follows.

We will simplify the derivation of the asymptotic distribution of $\hat{\boldsymbol{\Theta}}$ by assuming that the true component scores $(\mathbf{U}_i, \mathbf{V}_i)$ are used, instead of the estimated scores $(\hat{\mathbf{U}}_i, \hat{\mathbf{V}}_i)$, and by assuming that $\boldsymbol{\Sigma}_0$ is fixed and known. In that case we can apply Theorem 5.23 of Van der Vaart (1998) directly, and obtain that $\sqrt{n}\{\mathrm{vec}(\hat{\boldsymbol{\Theta}}) - \mathrm{vec}(\boldsymbol{\Theta}_0)\}$ is asymptotically $N(\mathbf{0}, \mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1})$ with

$$
\mathbf{A} = \mathrm{E}\{\nabla_{\mathrm{vec}(\boldsymbol{\Theta})}\nabla^T_{\mathrm{vec}(\boldsymbol{\Theta})} m_{(\boldsymbol{\Theta}_0,\boldsymbol{\Sigma}_0)}(\mathbf{U}, \mathbf{V})\}
$$

and

$$
\mathbf{B} = \mathrm{E}\{\nabla_{\mathrm{vec}(\boldsymbol{\Theta})} m_{(\boldsymbol{\Theta}_0,\boldsymbol{\Sigma}_0)}(\mathbf{U}, \mathbf{V})\nabla^T_{\mathrm{vec}(\boldsymbol{\Theta})} m_{(\boldsymbol{\Theta}_0,\boldsymbol{\Sigma}_0)}(\mathbf{U}, \mathbf{V})\};
$$

these expectations are taken with respect to the true parameters $(\boldsymbol{\Theta}_0, \boldsymbol{\Sigma}_0)$. Without loss of generality we can eliminate the factor $2\boldsymbol{\Sigma}^{-1}$ in (17); then it is easy to see that (16) holds. To derive (15) we use differentials again:

$$
\mathrm{d}\{\nabla^T_{\mathrm{vec}(\boldsymbol{\Theta})} m_{(\boldsymbol{\Theta},\boldsymbol{\Sigma}_0)}(\mathbf{U}, \mathbf{V})\} =
$$

$$
\begin{aligned}
&= 2\rho''(e)w^2(\mathbf{U})(\mathbf{V}-\boldsymbol{\Theta}^T\mathbf{U})^T\boldsymbol{\Sigma}_0^{-1}(\mathrm{d}\boldsymbol{\Theta})^T\mathbf{U}\mathrm{vec}\{\mathbf{U}(\mathbf{V}-\boldsymbol{\Theta}^T\mathbf{U})^T\}^T \\
&\quad +\rho'(e)w(\mathbf{U})\mathrm{vec}\{\mathbf{U}(\mathrm{d}\boldsymbol{\Theta}^T\mathbf{U})^T\}^T \\
&= 2\rho''(e)w^2(\mathbf{U})\mathrm{tr}\{(\mathrm{d}\boldsymbol{\Theta})^T\mathbf{U}(\mathbf{V}-\boldsymbol{\Theta}^T\mathbf{U})^T\boldsymbol{\Sigma}_0^{-1}\}\mathrm{vec}\{\mathbf{U}(\mathbf{V}-\boldsymbol{\Theta}^T\mathbf{U})^T\}^T \\
&\quad +\rho'(e)w(\mathbf{U})\mathrm{vec}(\mathbf{U}\mathbf{U}^T\mathrm{d}\boldsymbol{\Theta})^T \\
&= 2\rho''(e)w^2(\mathbf{U})\mathrm{vec}(\mathrm{d}\boldsymbol{\Theta})^T\mathrm{vec}\{\mathbf{U}(\mathbf{V}-\boldsymbol{\Theta}^T\mathbf{U})^T\boldsymbol{\Sigma}_0^{-1}\}\mathrm{vec}\{\mathbf{U}(\mathbf{V}-\boldsymbol{\Theta}^T\mathbf{U})^T\}^T \\
&\quad +\rho'(e)w(\mathbf{U})\{(\mathbf{I}_q\otimes\mathbf{U}\mathbf{U}^T)\mathrm{vec}(\mathrm{d}\boldsymbol{\Theta})\}^T,
\end{aligned}
$$

so

$$
\nabla_{\mathrm{vec}(\boldsymbol{\Theta})}\nabla^T_{\mathrm{vec}(\boldsymbol{\Theta})}m_{(\boldsymbol{\Theta},\boldsymbol{\Sigma}_0)}(\mathbf{U},\mathbf{V}) =
$$

$$
\begin{aligned}
&= 2\rho''(e)w^2(\mathbf{U})\mathrm{vec}\{\mathbf{U}(\mathbf{V}-\boldsymbol{\Theta}^T\mathbf{U})^T\boldsymbol{\Sigma}_0^{-1}\}\mathrm{vec}\{\mathbf{U}(\mathbf{V}-\boldsymbol{\Theta}^T\mathbf{U})^T\}^T \\
&\quad +\rho'(e)w(\mathbf{U})(\mathbf{I}_q\otimes\mathbf{U}\mathbf{U}^T) \\
&= 2\rho''(e)w^2(\mathbf{U})(\boldsymbol{\Sigma}_0^{-1}\mathbf{R}\otimes\mathbf{U})(\mathbf{R}\otimes\mathbf{U})^T+\rho'(e)w(\mathbf{U})(\mathbf{I}_q\otimes\mathbf{U}\mathbf{U}^T),
\end{aligned}
$$

from which (15) follows.

References

Ash, R. B. and Gardner, M. F. (1975). *Topics in Stochastic Processes*. Probability and Mathematical Statistics (Vol. 27). New York: Academic Press.

Chiou, J.-M., Müller, H.-G., and Wang, J.-L. (2004). Functional response models. *Statistica Sinica* **14**, 675–693.

Cuevas, A., Febrero, M., and Fraiman, R. (2007). Robust estimation and classification for functional data via projection-based depth notions. *Computational Statistics* **22**, 481–496.

Davison, A. C. and Hinkley, D. V. (1997). *Bootstrap Methods and Their Application*. Cambridge: Cambridge University Press.

Fraiman, R., and Muniz, G. (2001). Trimmed means for functional data. *Test* **10**, 419–440.

Gervini, D. (2008). Robust functional estimation using the median and spherical principal components. *Biometrika* **95**, 587–600.

Gervini, D. (2009). Detecting and handling outlying trajectories in irregularly sampled functional datasets. *The Annals of Applied Statistics* **3**, 1758–1775.

He, X., Simpson, D. G. and Wang, G. (2000). Breakdown points of t-type regression estimators. *Biometrika* **87**, 675–687.

James, G., Hastie, T. G. and Sugar, C. A. (2000). Principal component models for sparse functional data. *Biometrika* **87**, 587–602.

Locantore, N., Marron, J. S., Simpson, D. G., Tripoli, N., Zhang, J. T. and Cohen, K. L. (1999). Robust principal components for functional data (with discussion). *Test* **8**, 1–28.

Magnus, J. R., and Neudecker, H. (1999). *Matrix Differential Calculus with Applications in Statistics and Econometrics. Revised Edition*, New York: Wiley.

Maronna, R. A., Martin, R. D. and Yohai, V. J. (2006). *Robust Statistics. Theory and Methods*. New York: Wiley.

Maronna, R. A. and Yohai, V. J. (2012). Robust functional linear regression based on splines. To appear in *Computational Statistics & Data Analysis.*

Müller, H.-G., Chiou, J.-M., and Leng, X. (2008). Inferring gene expression dynamics via functional regression analysis. *BMC Bioinformatics* **9:60**.

Ramsay, J. O. and Silverman, B. W. (2005). *Functional Data Analysis. Second Edition*. New York: Springer.

Van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge, UK: Cambridge University Press.

Yao, F., Müller, H.-G. and Wang, J.-L. (2005a). Functional linear regression analysis for longitudinal data. *The Annals of Statistics* **33**, 2873–2903.

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

Zhu, H., Brown, P.J. and Morris, J.S. (2011). Robust, adaptive functional regression in functional mixed model framework. *Journal of the American Statistical Association* **106**, 1167–1179.