Web-based Supplementary Materials for "Robust Estimation for Ordinary Differential Equation Models"

by J. Cao, L. Wang and J. Xu.

Web Appendix A: Technical Details for Subsection 2.1: Robust Penalized Smoothing with the ODE-defined Penalty

The integration in (3) can be approximated by numerical quadrature. The composite Simpson's rule provides an adequate approximation to the exact integral (Burden and Douglas 2000), which is used here. For an arbitrary function $g(t)$, the composite Simpson's rule is given by

$$\int_{t_1}^{t_n} g(t)dt \approx \frac{\delta}{3}\left\{g(s_0) + 2\sum_{q=1}^{Q/2-1} g(s_{2q}) + 4\sum_{q=1}^{Q/2} g(s_{2q-1}) + g(s_Q)\right\},$$

where the quadrature points $s_q = t_1 + q\delta$, $q = 0, \cdots, Q$, and $\delta = (t_n - t_1)/Q$.

The estimate $\widehat{\mathbf{c}}$ is obtained by minimizing $J(\mathbf{c}|\boldsymbol{\theta})$. This optimization can be implemented quickly via the Matlab optimization function `lsqnonlin` with analytic gradients provided. The Matlab function `lsqnonlin` uses the trust-region-reflective algorithm by default. This algorithm is a subspace trust-region method and is based on the interior-reflective Newton method described in Coleman and Li (1996). `lsqnonlin` requires the initial value of $\mathbf{c}$, which can be given by the penalized spline smoothing.

Web Appendix B: Technical Details for Section 2.3: Variances of the ODE Parameters

The analytic derivative for $d\Psi_{ij}/d\boldsymbol{\theta}$ is

$$\begin{aligned}\frac{d}{d\boldsymbol{\theta}}\Psi_{ij}(\boldsymbol{\theta}) &= \sum_{k=1}^{K_j} \frac{d^2\hat{c}_{jk}}{d\boldsymbol{\theta} d\boldsymbol{\theta}^T}\phi_{jk}(t_{ij})\rho'\{y_j(t_{ij}) - \boldsymbol{\phi}_j^T(t_{ij})\widehat{\mathbf{c}}_j(\boldsymbol{\theta})\} \\ &\quad -\left(\frac{d\widehat{\mathbf{c}}_j}{d\boldsymbol{\theta}}\right)^T \boldsymbol{\phi}_j(t_{ij})\rho''\{y_j(t_{ij}) - \boldsymbol{\phi}_j^T(t_{ij})\widehat{\mathbf{c}}_j(\boldsymbol{\theta})\}\boldsymbol{\phi}_j^T(t_{ij})\left(\frac{d\widehat{\mathbf{c}}_j}{d\boldsymbol{\theta}}\right),\end{aligned}$$

where $d^2\hat{c}_{jk}/d\boldsymbol{\theta}d\boldsymbol{\theta}^T$ is obtained using the implicit function theorem as follows.

Taking the second-order $\boldsymbol{\theta}$-derivative on both sides of the identity $\partial J/\partial c_{jk}|_{\hat{c}_{jk}} =$

0:

$$\frac{d^2}{d\boldsymbol{\theta} d\boldsymbol{\theta}^T}\left(\frac{\partial J}{\partial c_{jk}|_{\hat{c}_{jk}}}\right) = \left.\frac{\partial^3 J}{\partial c_{jk}\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\right|_{\hat{c}_{jk}} + \left.\frac{\partial^3 J}{\partial c_{jk}^2\partial\boldsymbol{\theta}}\right|_{\hat{c}_{jk}}\frac{d\hat{c}_{jk}}{d\boldsymbol{\theta}^T} + \left.\frac{\partial^3 J}{\partial c_{jk}^3}\right|_{\hat{c}_{jk}}\frac{d\hat{c}_{jk}}{d\boldsymbol{\theta}}\frac{d\hat{c}_{jk}}{d\boldsymbol{\theta}^T}$$
$$+ \left.\frac{\partial^2 J}{\partial c_{jk}^2}\right|_{\hat{c}_{jk}}\frac{d^2\hat{c}_{jk}}{d\boldsymbol{\theta} d\boldsymbol{\theta}^T} = 0\,.$$

Assuming that $\left.\frac{\partial^2 J}{\partial c_{jk}^2}\right|_{\hat{c}_{jk}} \neq 0$, the analytic expression for the second-order derivative of $\hat{c}_{jk}$ with respect to $\boldsymbol{\theta}$ is obtained:

$$\frac{d^2\hat{c}_{jk}}{d\boldsymbol{\theta} d\boldsymbol{\theta}^T} = -\left[\left.\frac{\partial^2 J}{\partial c_{jk}^2}\right|_{\hat{c}_{jk}}\right]^{-1}\left[\left.\frac{\partial^3 J}{\partial c_{jk}\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\right|_{\hat{c}_{jk}} + \left.\frac{\partial^3 J}{\partial c_{jk}^2\partial\boldsymbol{\theta}}\right|_{\hat{c}_{jk}}\frac{d\hat{c}_{jk}}{d\boldsymbol{\theta}^T} + \left.\frac{\partial^3 J}{\partial c_{jk}^3}\right|_{\hat{c}_{jk}}\frac{d\hat{c}_{jk}}{d\boldsymbol{\theta}}\frac{d\hat{c}_{jk}}{d\boldsymbol{\theta}^T}\right].$$

Web Appendix C: A simulation study on a Linear ODE

When a temperature probe is firmly held between our thumb and forefinger, the temperature of the probe may be modeled by a linear ODE (Lomen and Lovelock 1996):

$$\frac{dX(t)}{dt} = -\alpha_1 X(t) + \alpha_2\,. \tag{1}$$

Two parameters, α_1 and α_2 in the above ODE, are to be estimated from the simulated noisy data.

The simulated data are generated as follows. The above ODE is numerically solved at 101 equally-spaced points on [0,1] with the initial value $X(t_0) = 1$ and the true parameter values $\alpha_1 = 3$ and $\alpha_2 = 10$. The noisy data are generated by adding noise distributed as $\texttt{Normal}(0, 0.5^2)$ to the ODE solution. We then randomly select m observations using the discrete Uniform distribution on [1,101]. These selected m observations are replaced with randomly generated outliers that are added or subtracted from the original observations with equal probability. Each outlier is generated from the Pareto distribution, Pareto($\upsilon = 3, \xi = 1.5$), where the density function of Pareto(υ, ξ) is $f(x) = \upsilon\xi^{\upsilon}/x^{\upsilon+1}$, $x > \xi$.

The ODE parameters, α_1 and α_2, are estimated from the simulated data using the robust method and the generalized profiling method. Both methods represent the dynamic process $X(t)$ with a cubic B-spline using 101 equally-spaced knots on [0,1]. The effect of outliers is studied by varying the number of outliers $m = 0$, 10, 20, 30.

Table 1 displays the bias, standard deviation (SD) and root mean squared error (RMSE) for parameter estimates on 100 simulation replicates in the four scenarios. When no outliers are added to the simulated data (m = 0), the robust method has 2% larger RMSE than the generalized profiling method. When outliers are present in the simulated data, the robust method has much smaller bias, SD and RMSE than the generalized profiling method. For example, when the simulated data have 30% outliers, the RMSE of the parameter estimates using the robust method is only around 50% of that from the generalized profiling method.

The standard errors (SEs) for parameter estimates are estimated using the sandwich method. Table 2 shows the mean and standard deviation (SD)

Table 1: The biases, standard deviations (SDs), and root mean squared errors (RMSEs) of parameter estimates on 100 simulation replicates using the robust method and the generalized profiling (GP) method for the linear ODE (1). The true values of parameters are $\alpha_1 = 3$, and $\alpha_2 = 10$.

	Parameters	α_1		α_2	
Scenario	Methods	Robust	GP	Robust	GP
No Outliers	BIAS	0.10	0.07	0.35	0.28
	SD	0.82	0.81	2.28	2.26
	RMSE	0.83	0.81	2.30	2.26
10 Outliers	BIAS	0.19	0.37	0.63	1.19
	SD	0.97	1.60	2.68	4.51
	RMSE	0.98	1.64	2.74	4.64
20 Outliers	BIAS	0.13	0.25	0.42	0.82
	SD	1.26	1.97	3.44	5.31
	RMSE	1.26	1.97	3.45	5.35
30 Outliers	BIAS	0.28	0.41	0.87	1.26
	SD	1.60	3.17	4.53	8.44
	RMSE	1.62	3.18	4.59	8.49

Table 2: The means and standard deviations (SDs) for the standard error estimates using the sandwich method over 100 simulation replicates for the linear ODE (1). "Sample" represents the sample SDs of the parameter estimates. "CP" stands for the coverage probabilities of the 95% confidence intervals for the parameters.

Scenario	Parameter	Sample	Mean	SD	CP
10%	α_1	0.97	0.92	0.26	95%
Outliers	α_2	2.68	2.51	0.73	94%
20%	α_1	1.26	1.16	0.42	96%
Outliers	α_2	3.44	3.13	1.17	95%
30%	α_1	1.60	1.43	0.74	94%
Outliers	α_2	4.53	3.87	2.18	94%

of the standard error estimates over 100 simulation replicates. We also calculate the sample standard deviation of the parameter estimates in the same 100 simulation replicates, which can be treated as another estimate for the standard error using the parametric bootstrap method (Efron and Tibshirani 1993). The mean of the sandwich estimates is slightly smaller than the sample standard deviation. We also calculate the 95% confidence intervals for the parameters as $[\hat{\alpha}_j - 1.96. * \widehat{SE}(\hat{\alpha}_j), \hat{\alpha}_j + 1.96. * \widehat{SE}(\hat{\alpha}_j)]$, $j = 1, 2$. The coverage probabilities of the 95% confidence intervals are given in Table 2, and are very close to 95%.