Online Supplement for:
A Penalized Spline Estimator For Fixed Effects Panel Data Models

Peter Pütz[*], Thomas Kneib[†]

Further Simulation Results

As discussed in Section 4 of the article, we conduct further simulation studies to investigate whether the confidence bands for the derivatives, in particular for $f_2'(x_{2it})$, approach the nominal coverage rate of 95%. Therefore, we use the setting in the article with $N = (150)$ and $n = (1050)$ and vary several parameters: the error standard deviation σ_u, the number of knots k and the difference order q of the penalty. The results in Table 1 indicate that the coverate rates do not universally improve for lower error variance or higher difference order of the penalty. The width of the bands decreases, but the coverage rates for $f_2'(x_{2it})$ generally rather impair. A larger number of knots results in improved coverage rates but wider bands. Similar problems are observable if we substitute the second function for $f_2(x_{2it}) = (0.5 - x_{2it})^3$ but use the same setting apart from that. As Table 2 shows, even an increasing amount of knots does not prevent bad coverage rates for the derivatives.

Furthermore, we check the sensitivity of the validity of our approach to deviations from the model assumptions. Thus, we generate the data with autocorrelated and non-Gaussian errors and check the coverage rates for the confidence bands around the true functions (the derivatives are not considered). The autocorrelation is introduced by an AR(1)-process in the error term for each individual in a balanced panel data setting ($T = 7$) with the AR(1)-coefficient set to 0.2. As non-Gaussian error distributions the uniform distribution and the poisson distribution with parameter $\lambda = 10$ are chosen. For the sake of comparability, after generating the error terms, their standard deviation is set to 0.5 in all settings. The results in Table 3 indicate an accurate performance of the estimated confidence bands despite the deviations from the standard model assumptions, even for moderate sample sizes.

[*]Georg-August-Universität Göttingen, Centre for Statistics, e-mail: ppuetz@uni-goettingen.de.
[†]Georg-August-Universität Göttingen, Faculty of Economic Sciences, Chair of Statistics.

Table 1: Coverage rates in simulations (500 replicates) for the derivatives, average areas between confidence bands in parantheses. Columns (i) denote estimation with using GLS, columns (ii) without using GLS.

Parameters	f_1'		f_2'		f_3'	
	(i)	(ii)	(i)	(ii)	(i)	(ii)
$k = 40, \sigma_u = 0.5, q = 2$	0.90 (23.51)	0.77 (24.42)	0.85 (25.19)	0.66 (25.85)	0.94 (14.06)	0.84 (14.92)
$k = 40, \sigma_u = 0.1, q = 2$	0.92 (8.23)	0.76 (8.36)	0.77 (8.88)	0.60 (8.92)	0.95 (4.82)	0.92 (4.93)
$k = 40, \sigma_u = 0.5, q = 3$	0.93 (20.13)	0.83 (20.45)	0.81 (22.84)	0.66 (23.08)	0.99 (10.80)	0.87 (11.05)
$k = 40, \sigma_u = 0.5, q = 4$	0.94 (19.07)	0.85 (19.09)	0.74 (22.62)	0.62 (22.70)	0.97 (12.91)	0.87 (12.87)
$k = 80, \sigma_u = 0.5, q = 3$	0.93 (20.27)	0.83 (20.62)	0.82 (23.05)	0.67 (23.33)	0.99 (10.81)	0.87 (11.08)
$k = 80, \sigma_u = 0.5, q = 2$	0.92 (24.79)	0.80 (25.99)	0.87 (26.63)	0.68 (27.65)	0.96 (14.83)	0.88 (15.85)
$k = 120, \sigma_u = 0.5, q = 2$	0.93 (25.35)	0.81 (26.65)	0.89 (27.25)	0.71 (28.39)	0.97 (15.12)	0.89 (16.20)

Table 2: Coverage rates in simulations (500 replicates) for the derivatives, average areas between confidence bands in parantheses, now with a cubic function $f_2(x_{2it})$. Columns (i) denote estimation with using GLS, columns (ii) without using GLS.

Parameters	f_1'		f_2'		f_3'	
	(i)	(ii)	(i)	(ii)	(i)	(ii)
$k = 40, \sigma_u = 0.5, q = 2$	0.90 (23.45)	0,76 (24.36)	0.7 (13.81)	0.62 (14.81)	0.94 (14.23)	0.84 (15.03)
$k = 80, \sigma_u = 0.5, q = 2$	0.91 (24.75)	0.79 (25.95)	0.72 (14.36)	0.65 (15.54)	0.96 (14.80)	0.88 (15.78)
$k = 120, \sigma_u = 0.5, q = 2$	0.92 (25.30)	0.80 (26.61)	0.72 (14.65)	0.67 (15.90)	0.97 (15.10)	0.89 (16.13)

Table 3: Coverage rates in simulations (500 replicates) for different error distributions, average areas between confidence bands in parantheses. Columns (i) denote estimation with using GLS, columns (ii) without using GLS.

	f_1		f_2		f_3	
$u_{it} \sim \text{unif}(0,1)$	(i)	(ii)	(i)	(ii)	(i)	(ii)
$n = 525$	0.94 (3.43)	0.89 (3.48)	0.92 (3.66)	0.86 (3.62)	0.95 (3.07)	0.90 (3.24)
$n = 1050$	0.95 (2.45)	0.89 (2.51)	0.96 (2.44)	0.86 (2.44)	0.97 (2.12)	0.91 (2.14)
$n = 2100$	0.95 (1.81)	0.89 (1.80)	0.97 (1.90)	0.89 (1.89)	0.97 (1.54)	0.89 (1.55)
$u_{it} \sim \text{Pois}(10)$						
$n = 525$	0.94 (3.42)	0.87 (3.47)	0.92 (3.65)	0.85 (3.61)	0.96 (3.07)	0.90 (3.24)
$n = 1050$	0.95 (2.45)	0.86 (2.51)	0.95 (2.44)	0.86 (2.44)	0.95 (2.12)	0.87 (2.14)
$n = 2100$	0.96 (1.81)	0.83 (1.80)	0.98 (1.89)	0.88 (1.88)	0.98 (1.54)	0.88 (1.55)
$u_{it} = 0.2u_{i,t-1} + w_{it},\ w_{it} \sim N(0, 0.09)$						
$n = 525$	0.86 (3.62)	0.81 (3.34)	0.95 (3.85)	0.89 (3.69)	0.96 (3.13)	0.88 (3.04)
$n = 1050$	0.94 (2.54)	0.87 (2.37)	0.96 (2.70)	0.88 (2.53)	0.97 (2.16)	0.91 (2.08)
$n = 2100$	0.95 (1.89)	0.91 (1.74)	0.95 (1.88)	0.86 (1.75)	0.95 (1.67)	0.89 (1.57)