Robust Mean Signal Estimation and Inference for Imaging Data

Yang Long[1], Guanqun Cao[2], David Kepplinger[1], and Lily Wang[1]

[1] *George Mason University, Fairfax, VA 22030, USA.*

[2]*Michigan State University, East Lansing, MI 48824, USA.*

Supplementary Material

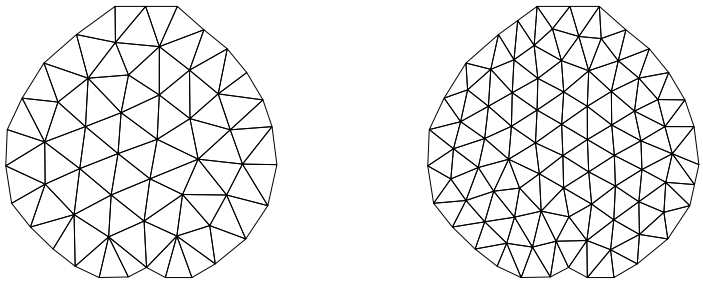
In this supplement, Section S1 presents additional simulation studies, and Section S2 includes detailed proofs of the theoretical results in the main article.

S1 More Results from Simulation Studies

S1.1 Sensitivity Analysis of Tuning Parameters

In this section, we provide additional simulation results to examine the sensitivity of the proposed method to critical tuning parameters. Specifically, we explore the effects of the spline degree d, triangulation choices, and Huber constants k on the performance of the proposed estimators and SCCs. Figure S.1 shows the triangulation options considered: $\triangle_1$ (80 triangles and 54 vertices) and $\triangle_2$ (144 triangles and 87 vertices). Compared to $\triangle_1$ used in the main text, $\triangle_2$ provides a finer triangulation, resulting in a larger number of basis functions.

For the spline degree d, we consider $d = 4$ and $d = 6$ in addition to $d = 5$ used in Section 6 of the main paper, with larger values of d providing a higher degree of Bernstein basis polynomials. We further explore the Huber constants $k = 0.01,\ 0.1,\ 1.0$. These results aim to provide a deeper understanding of how each parameter influences estimation and inference and help guide practitioners in selecting appropriate parameter values for different datasets.



(a) Triangulation $\triangle_1$ (b) Triangulation $\triangle_2$

Figure S.1: Triangulation settings $\triangle_1$ and $\triangle_2$ constructed using the `ImageSCC` R package.

The sensitivity analysis for mean function estimation is presented in Tables S.1, S.2, and S.3. The simulation datasets feature the exponential mean function (as defined in Section 6 of the main paper), considering uncontaminated datasets as the baseline, datasets with 10% surface-type contamination, and datasets with heavy-tailed error distributions with 15% mixing weights. Fixing k, the results across all three tables demonstrate similar performance across different choices of $\triangle$ and d for the M-BPST estimators, further indicating that the proposed estimator is relatively insensitive to these hyperparameters for the chosen simulation

datasets. Regarding the choice of k, all three values exhibit sufficient robustness against the contamination scenarios considered in the simulations. The mean integrated squared error (MISE) remains consistent across different contamination scenarios and decreases as the sample size increases. The M-BPST estimator with $k = 1.0$ shows marginally smaller MISE compared to the other two choices, suggesting a subtle improvement in estimation efficiency while maintaining robust performance.

Table S.1: Mean (SD) of integrated squared errors of different estimators across different sample sizes (n) and hyperparameters for the exponential mean function when there is no contamination.

Hyperpar.	n	BPST	M-BPST		
			$k = 0.01$	$k = 0.1$	$k = 1.0$
$\triangle_1, d = 4$	100	0.012 (0.014)	0.018 (0.018)	0.018 (0.018)	0.014 (0.016)
	200	0.006 (0.007)	0.008 (0.008)	0.008 (0.008)	0.007 (0.007)
	400	0.003 (0.003)	0.004 (0.003)	0.004 (0.003)	0.003 (0.003)
$\triangle_1, d = 5$	100	0.012 (0.014)	0.018 (0.018)	0.018 (0.018)	0.014 (0.016)
	200	0.006 (0.007)	0.008 (0.008)	0.008 (0.008)	0.007 (0.007)
	400	0.003 (0.003)	0.004 (0.003)	0.004 (0.003)	0.003 (0.003)
$\triangle_1, d = 6$	100	0.012 (0.014)	0.018 (0.018)	0.018 (0.018)	0.014 (0.016)
	200	0.006 (0.007)	0.008 (0.008)	0.008 (0.008)	0.007 (0.007)
	400	0.003 (0.003)	0.004 (0.003)	0.004 (0.003)	0.003 (0.003)
$\triangle_2, d = 4$	100	0.012 (0.014)	0.018 (0.018)	0.018 (0.018)	0.014 (0.016)
	200	0.006 (0.007)	0.008 (0.008)	0.008 (0.008)	0.007 (0.007)
	400	0.003 (0.003)	0.004 (0.003)	0.004 (0.003)	0.003 (0.003)
$\triangle_2, d = 5$	100	0.012 (0.014)	0.018 (0.018)	0.018 (0.018)	0.014 (0.016)
	200	0.006 (0.007)	0.008 (0.008)	0.008 (0.008)	0.007 (0.007)
	400	0.003 (0.003)	0.004 (0.003)	0.004 (0.003)	0.003 (0.003)

Table S.2: Mean (SD) of integrated squared errors of different estimators across different sample sizes (n) and hyperparameters for the exponential mean function when there is surface-type contamination with 15% contamination rate.

	Cont. Type = Stripe				Cont. Type = Square			
	BPST	M-BPST			BPST	M-BPST		
		$k = 0.01$	$k = 0.1$	$k = 1.0$		$k = 0.01$	$k = 0.1$	$k = 1.0$
$\triangle_1, d = 4$								
$n = 100$	0.101 (0.032)	0.018 (0.019)	0.018 (0.019)	0.015 (0.016)	0.142 (0.019)	0.018 (0.019)	0.018 (0.019)	0.015 (0.016)
$n = 200$	0.095 (0.025)	0.009 (0.008)	0.009 (0.008)	0.008 (0.007)	0.135 (0.014)	0.009 (0.008)	0.009 (0.008)	0.008 (0.007)
$n = 400$	0.092 (0.017)	0.005 (0.004)	0.005 (0.004)	0.004 (0.004)	0.132 (0.009)	0.005 (0.004)	0.005 (0.004)	0.004 (0.004)
$\triangle_1, d = 5$								
$n = 100$	0.164 (0.032)	0.018 (0.019)	0.018 (0.019)	0.015 (0.016)	0.144 (0.019)	0.018 (0.019)	0.018 (0.019)	0.015 (0.016)
$n = 200$	0.158 (0.025)	0.009 (0.008)	0.009 (0.008)	0.008 (0.007)	0.138 (0.014)	0.009 (0.008)	0.009 (0.008)	0.008 (0.007)
$n = 400$	0.155 (0.017)	0.005 (0.004)	0.005 (0.004)	0.004 (0.004)	0.134 (0.009)	0.005 (0.004)	0.005 (0.004)	0.004 (0.004)
$\triangle_1, d = 6$								
$n = 100$	0.260 (0.032)	0.018 (0.019)	0.018 (0.019)	0.015 (0.016)	0.169 (0.019)	0.018 (0.019)	0.018 (0.019)	0.015 (0.016)
$n = 200$	0.254 (0.025)	0.009 (0.008)	0.009 (0.008)	0.008 (0.007)	0.162 (0.014)	0.009 (0.008)	0.009 (0.008)	0.008 (0.007)
$n = 400$	0.250 (0.017)	0.005 (0.004)	0.005 (0.004)	0.004 (0.004)	0.159 (0.009)	0.005 (0.004)	0.005 (0.004)	0.004 (0.004)
$\triangle_2, d = 4$								
$n = 100$	0.162 (0.032)	0.018 (0.019)	0.018 (0.019)	0.015 (0.016)	0.144 (0.019)	0.018 (0.019)	0.018 (0.019)	0.015 (0.016)
$n = 200$	0.156 (0.025)	0.009 (0.008)	0.009 (0.008)	0.008 (0.007)	0.137 (0.014)	0.009 (0.008)	0.009 (0.008)	0.008 (0.007)
$n = 400$	0.152 (0.017)	0.005 (0.004)	0.005 (0.004)	0.004 (0.004)	0.134 (0.009)	0.005 (0.004)	0.005 (0.004)	0.004 (0.004)
$\triangle_2, d = 5$								
$n = 100$	0.265 (0.032)	0.018 (0.019)	0.018 (0.019)	0.015 (0.016)	0.169 (0.019)	0.018 (0.019)	0.018 (0.019)	0.015 (0.016)
$n = 200$	0.259 (0.025)	0.009 (0.008)	0.009 (0.008)	0.008 (0.007)	0.163 (0.014)	0.009 (0.008)	0.009 (0.008)	0.008 (0.007)
$n = 400$	0.256 (0.017)	0.005 (0.004)	0.005 (0.004)	0.004 (0.004)	0.159 (0.009)	0.005 (0.004)	0.005 (0.004)	0.004 (0.004)

Table S.3: Mean (SD) of integrated squared errors of different estimators across different sample size (n) and hyperparameters for the exponential mean function when errors follow a heavy-tailed distribution with 15% mixing weights.

	Cont. Type = Cauchy				Cont. Type = Slash			
	BPST	M-BPST			BPST	M-BPST		
		$k = 0.01$	$k = 0.1$	$k = 1.0$		$k = 0.01$	$k = 0.1$	$k = 1.0$
$\triangle_1, d = 4$								
$n = 100$	7.2e2 (6.9e3)	0.018 (0.018)	0.017 (0.018)	0.015 (0.016)	5.982 (34.24)	0.017 (0.018)	0.017 (0.018)	0.014 (0.016)
$n = 200$	18.00 (91.18)	0.008 (0.007)	0.008 (0.007)	0.007 (0.007)	1.1e2 (9.9e2)	0.008 (0.008)	0.008 (0.008)	0.007 (0.007)
$n = 400$	83.63 (8.0e2)	0.004 (0.003)	0.004 (0.003)	0.003 (0.003)	4.2e2 (2.9e3)	0.004 (0.003)	0.004 (0.003)	0.003 (0.003)
$\triangle_1, d = 5$								
$n = 100$	7.2e2 (7.0e3)	0.018 (0.018)	0.018 (0.018)	0.015 (0.016)	9.845 (66.79)	0.017 (0.018)	0.017 (0.018)	0.014 (0.016)
$n = 200$	19.17 (94.60)	0.008 (0.007)	0.008 (0.007)	0.007 (0.007)	1.2e2 (9.9e2)	0.008 (0.008)	0.008 (0.007)	0.007 (0.007)
$n = 400$	1.0e2 (9.9e2)	0.004 (0.004)	0.004 (0.004)	0.003 (0.003)	4.9e2 (3.4e3)	0.005 (0.004)	0.004 (0.004)	0.003 (0.003)
$\triangle_1, d = 6$								
$n = 100$	3.2e2 (3.0e3)	0.018 (0.018)	0.018 (0.018)	0.015 (0.016)	5.328 (30.88)	0.017 (0.018)	0.017 (0.018)	0.014 (0.016)
$n = 200$	28.20 (1.9e2)	0.008 (0.007)	0.008 (0.007)	0.007 (0.007)	96.83 (8.1e2)	0.008 (0.008)	0.008 (0.008)	0.007 (0.007)
$n = 400$	1.1e2 (1.0e3)	0.006 (0.003)	0.004 (0.004)	0.003 (0.003)	5.2e2 (3.7e3)	0.006 (0.003)	0.005 (0.004)	0.003 (0.003)
$\triangle_2, d = 4$								
$n = 100$	7.1e2 (7.0e3)	0.018 (0.018)	0.018 (0.018)	0.015 (0.016)	9.335 (66.57)	0.017 (0.018)	0.017 (0.018)	0.014 (0.016)
$n = 200$	14.85 (82.98)	0.008 (0.007)	0.008 (0.007)	0.007 (0.007)	1.0e2 (9.9e2)	0.008 (0.008)	0.008 (0.007)	0.007 (0.007)
$n = 400$	96.57 (9.3e2)	0.004 (0.004)	0.004 (0.003)	0.003 (0.003)	4.7e2 (3.3e3)	0.005 (0.004)	0.004 (0.004)	0.003 (0.003)
$\triangle_2, d = 5$								
$n = 100$	3.2e2 (3.0e3)	0.018 (0.018)	0.018 (0.018)	0.015 (0.016)	5.351 (30.92)	0.017 (0.018)	0.017 (0.018)	0.014 (0.016)
$n = 200$	15.56 (79.84)	0.008 (0.007)	0.008 (0.007)	0.007 (0.007)	88.82 (8.0e2)	0.009 (0.008)	0.008 (0.008)	0.007 (0.007)
$n = 400$	1.0e2 (1.0e3)	0.006 (0.004)	0.004 (0.004)	0.003 (0.003)	5.1e2 (3.6e3)	0.006 (0.003)	0.005 (0.004)	0.003 (0.003)

Next, we evaluate the effect of the Huber constant k on our proposed robust SCC (RSCC) procedure. Tables S.4 and S.5 compare the performance under

different values of k, benchmarking against the non-robust SCC (NRSCC) introduced by Wang et al. (2020). The Huber constant k controls the robustness of the M-BPST estimator, with smaller k providing greater robustness. The additional simulation results focus on the datasets with an exponential true mean function, considering all contamination scenarios discussed in Section 6.2: uncontaminated datasets as the baseline, 5% or 10% surface-type contamination, and heavy-tailed error distributions with 5% or 15% mixing weights. We evaluate the performance for $k = 0.01,\ 0.1$ and 1.0, representing decreasing levels of robustness. The trade-off between robustness and efficiency of the M-BPST estimator is evident in both tables. For smaller k, the empirical coverage rate improves for both uncontaminated and contaminated datasets, particularly in challenging scenarios such as high contamination rates with square-type contamination. However, this increased robustness results in wider corridors, leading to overly conservative SCCs when the contamination rates are low. In cases of a heavy-tailed error distribution, where contamination is symmetric, all three values of k exhibit strong empirical coverage rates. However, $k = 0.01$ results in the widest and most conservative SCC, while $k = 1.0$ provides the narrowest corridor but slightly lower coverage rates compared to the other two. These results support the choice of $k = 0.1$ in Section 6 of the main paper, which strikes a balance between robustness and efficiency.

Table S.4: Coverage rates (average corridor width) of the robust and non-robust 95% SCC (RSCC and NRSCC) for the exponential true mean function based on non-contaminated samples with different Huber constants k.

n	NRSCC	RSCC		
		$k = 0.01$	$k = 0.1$	$k = 1.0$
100	0.94 (0.541)	0.95 (1.390)	0.90 (0.761)	0.90 (0.570)
200	0.97 (0.386)	0.99 (0.851)	0.95 (0.513)	0.95 (0.406)
400	0.97 (0.275)	0.99 (0.520)	0.98 (0.354)	0.95 (0.290)

Table S.5: Coverage rates (average corridor width) of the robust and non-robust 95% SCCs (RSCC and NRSCC) for the exponential true mean function based on contaminated samples with different Huber constants k.

	Cont. Rate/Mix Weight = 0.05				Cont. Rate=0.1/Mix Weight = 0.15			
	NRSCC	RSCC			NRSCC	RSCC		
		$k = 0.01$	$k = 0.1$	$k = 1.0$		$k = 0.01$	$k = 0.1$	$k = 1.0$
Stripe								
$n = 100$	0.50 (0.716)	0.95 (1.410)	0.90 (0.762)	0.88 (0.573)	0.00 (0.782)	0.94 (1.389)	0.89 (0.764)	0.88 (0.576)
$n = 200$	0.09 (0.509)	0.98 (0.830)	0.94 (0.517)	0.94 (0.409)	0.00 (0.556)	0.98 (0.831)	0.91 (0.517)	0.89 (0.412)
$n = 400$	0.00 (0.360)	0.97 (0.520)	0.94 (0.356)	0.93 (0.292)	0.00 (0.394)	0.94 (0.525)	0.87 (0.357)	0.87 (0.295)
Square								
$n = 100$	0.00 (0.662)	0.95 (1.398)	0.90 (0.760)	0.89 (0.573)	0.00 (0.709)	0.94 (1.396)	0.88 (0.763)	0.88 (0.575)
$n = 200$	0.00 (0.471)	0.98 (0.848)	0.92 (0.517)	0.91 (0.409)	0.00 (0.504)	0.97 (0.858)	0.90 (0.519)	0.78 (0.413)
$n = 400$	0.00 (0.335)	0.99 (0.521)	0.93 (0.355)	0.88 (0.292)	0.00 (0.358)	0.90 (0.523)	0.73 (0.356)	0.62 (0.295)
Cauchy								
$n = 100$	0.77 (4.219)	0.94 (1.391)	0.90 (0.763)	0.90 (0.579)	0.78 (11.74)	0.94 (1.346)	0.89 (0.780)	0.88 (0.594)
$n = 200$	0.70 (1.741)	0.98 (0.835)	0.94 (0.514)	0.93 (0.413)	0.73 (4.625)	0.95 (0.833)	0.94 (0.524)	0.94 (0.426)
$n = 400$	0.58 (1.787)	0.99 (0.520)	0.97 (0.363)	0.96 (0.299)	0.68 (4.935)	0.97 (0.547)	0.94 (0.385)	0.95 (0.313)
Slash								
$n = 100$	0.78 (1.799)	0.98 (1.396)	0.90 (0.756)	0.89 (0.576)	0.71 (4.491)	0.95 (1.384)	0.91 (0.758)	0.91 (0.597)
$n = 200$	0.72 (3.690)	0.98 (0.828)	0.94 (0.515)	0.94 (0.417)	0.70 (10.55)	0.98 (0.846)	0.94 (0.539)	0.94 (0.441)
$n = 400$	0.54 (3.324)	0.98 (0.519)	0.97 (0.361)	0.96 (0.303)	0.68 (9.571)	0.96 (0.538)	0.94 (0.388)	0.97 (0.320)

Table S.6 evaluates the performance of the proposed RSCC procedure vs the classic NRSCC procedure under different triangulation schemes ($\triangle$) and polynomial degrees (d). The simulations focus on datasets with the exponential mean function, considering uncontaminated datasets as the baseline, datasets with 10% surface-type contamination, and datasets with heavy-tailed error distributions with 15% mixing weights. The results in Table S.6 demonstrate consistent performance across different $\triangle$ and d, indicating that the RSCCs are relatively insensitive to these hyperparameters for the chosen simulation datasets

and resolution settings.

Table S.6: Coverage rates (average corridor width) of the robust and non-robust 95% SCCs (RSCC and NRSCC) for the exponential true mean function with $k = 0.1$, different $\triangle$, d, and contamination types (10% contamination rate for the surface-type cases, and 15% mixture weight for heavy-tailed error distribution scenarios). Bold values indicate the better-performing method in each scenario.

Hyperpar.	Cont. Type	$n = 100$		$n = 200$		$n = 400$	
		RSCC	NRSCC	RSCC	NRSCC	RSCC	NRSCC
$\triangle_1, d = 4$	Stripe	**0.89 (0.762)**	0.06 (0.674)	**0.91 (0.515)**	0.00 (0.478)	**0.87 (0.355)**	0.00 (0.340)
	Square	**0.91 (0.761)**	0.00 (0.694)	**0.90 (0.517)**	0.00 (0.494)	**0.73 (0.355)**	0.00 (0.351)
	Cauchy	**0.90 (0.775)**	0.69 (11.77)	**0.95 (0.518)**	0.63 (4.559)	**0.96 (0.368)**	0.64 (4.867)
	Slash	**0.93 (0.755)**	0.60 (4.419)	**0.94 (0.533)**	0.65 (10.35)	**0.97 (0.370)**	0.58 (9.532)
	Clean	0.90 (0.759)	**0.94 (0.538)**	0.95 (0.511)	**0.97 (0.384)**	0.98 (0.352)	**0.97 (0.273)**
$\triangle_1, d = 5$	Stripe	**0.89 (0.764)**	0.00 (0.782)	**0.91 (0.517)**	0.00 (0.556)	**0.87 (0.357)**	0.00 (0.394)
	Square	**0.88 (0.763)**	0.00 (0.709)	**0.90 (0.519)**	0.00 (0.504)	**0.73 (0.356)**	0.00 (0.358)
	Cauchy	**0.89 (0.780)**	0.78 (11.74)	**0.94 (0.524)**	0.73 (4.625)	**0.94 (0.385)**	0.68 (4.935)
	Slash	**0.91 (0.758)**	0.71 (4.491)	**0.94 (0.539)**	0.70 (10.55)	**0.94 (0.388)**	0.68 (9.571)
	Clean	0.90 (0.761)	**0.94 (0.541)**	0.95 (0.513)	**0.97 (0.386)**	0.98 (0.354)	**0.97 (0.275)**
$\triangle_1, d = 6$	Stripe	**0.89 (0.765)**	0.00 (0.894)	**0.91 (0.518)**	0.00 (0.635)	**0.87 (0.358)**	0.00 (0.451)
	Square	**0.88 (0.764)**	0.00 (0.791)	**0.90 (0.519)**	0.00 (0.561)	**0.73 (0.357)**	0.00 (0.399)
	Cauchy	**0.89 (0.783)**	0.52 (10.27)	**0.94 (0.527)**	0.49 (4.599)	**0.96 (0.405)**	0.40 (4.511)
	Slash	**0.91 (0.760)**	0.39 (4.212)	**0.93 (0.542)**	0.46 (10.19)	**0.94 (0.414)**	0.42 (9.369)
	Clean	0.90 (0.762)	**0.94 (0.543)**	0.96 (0.513)	**0.97 (0.387)**	0.98 (0.355)	**0.97 (0.275)**
$\triangle_2, d = 4$	Stripe	**0.89 (0.764)**	0.00 (0.776)	**0.91 (0.517)**	0.00 (0.550)	**0.87 (0.357)**	0.00 (0.390)
	Square	**0.88 (0.762)**	0.00 (0.706)	**0.89 (0.518)**	0.00 (0.502)	**0.73 (0.356)**	0.00 (0.356)
	Cauchy	**0.89 (0.779)**	0.81 (11.73)	**0.93 (0.523)**	0.76 (4.471)	**0.94 (0.381)**	0.74 (5.036)
	Slash	**0.92 (0.757)**	0.73 (4.622)	**0.94 (0.538)**	0.78 (9.996)	**0.95 (0.383)**	0.69 (9.756)
	Clean	0.90 (0.761)	**0.94 (0.540)**	0.95 (0.512)	**0.97 (0.386)**	0.98 (0.354)	**0.96 (0.274)**
$\triangle_2, d = 5$	Stripe	**0.89 (0.765)**	0.00 (0.890)	**0.91 (0.518)**	0.00 (0.631)	**0.86 (0.358)**	0.00 (0.448)
	Square	**0.88 (0.764)**	0.00 (0.792)	**0.90 (0.519)**	0.00 (0.563)	**0.72 (0.357)**	0.00 (0.399)
	Cauchy	**0.89 (0.784)**	0.53 (10.69)	**0.94 (0.527)**	0.49 (4.342)	**0.96 (0.407)**	0.40 (4.565)
	Slash	**0.91 (0.760)**	0.39 (4.337)	**0.93 (0.542)**	0.44 (9.744)	**0.94 (0.421)**	0.43 (9.587)
	Clean	0.90 (0.762)	**0.94 (0.543)**	0.96 (0.514)	**0.97 (0.387)**	0.98 (0.355)	**0.97 (0.275)**

S1.2 Additional Simulation Results

In this section, we present additional comparison results for the performance of five mean function estimators: the non-robust BPST mean function estimator (BPST) introduced by Wang et al. (2020), our proposed BPST-based M-mean function estimator (M-BPST), the robust deep neural network mean function estimator (RDNN) proposed by Wang and Cao (2022), and two additional 2D

smoothing methods, thin-plate (TPS) and tensor-product spline with B-splines (TPB), as implemented in `mgcv`. The comparative performance of these estimators under different contamination scenarios is shown in Figures S.2 through S.4, which illustrate the behavior of the estimators using simulated data generated from a quadratic, exponential and cubic mean function, respectively, on a brain-shaped domain.

Tables S.7 through S.9 summarize the performance of the mean function estimators, reporting the mean and standard deviation of the integrated squared errors under both uncontaminated and contamination scenarios. As seen in the results presented in Tables 1 and 2 in Section 6 of the main paper, the non-robust estimators are highly susceptible to outliers. Its convergence properties deteriorate markedly as sample size increases in contaminated datasets. In contrast, both the M-BPST and RDNN estimators exhibit consistent MISE across different contamination scenarios. Between the M-BPST and RDNN estimators, M-BPST demonstrates a slight advantage in terms of MISE. In addition, the challenge of tuning hyperparameters for the RDNN persists, as its MISE does not necessarily improve with an increase in sample size n to 400.

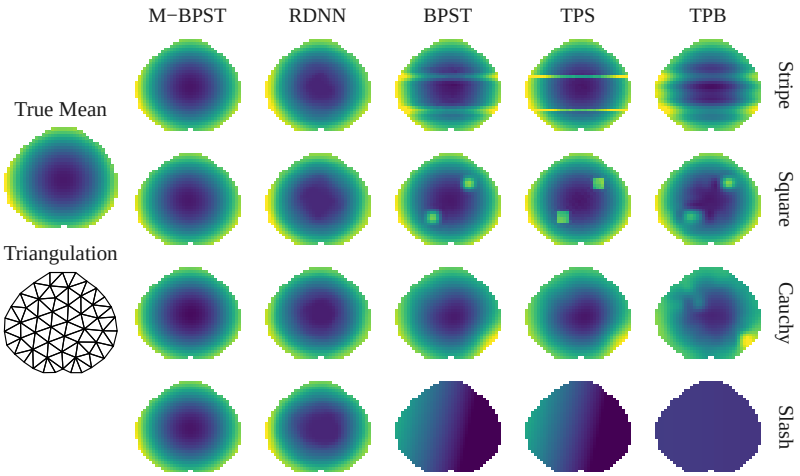


Figure S.2: Simulation run demonstrating the performance of mean function estimation methods on a **quadratic** mean function. The leftmost column displays the true mean function alongside the triangulation setting used for the BPST and M-BPST estimators. The subsequent four columns demonstrate the performance of the estimators on a dataset with $n = 200$ under four contamination scenarios: stripe and square contamination at a 15% contamination rate, 15% Normal-Cauchy mixture, and 15% Normal-Slash mixture.

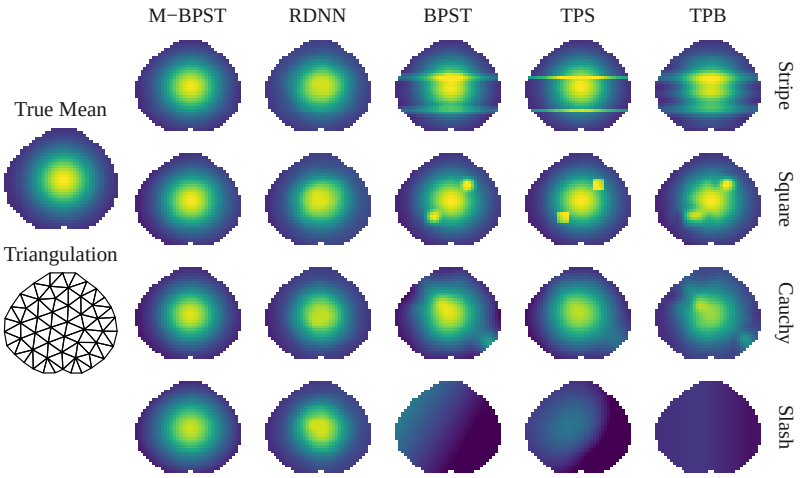


Figure S.3: Simulation run demonstrating the performance of mean function estimation methods on an **exponential** mean function. The leftmost column displays the true mean function alongside the triangulation setting used for the BPST and M-BPST estimators. The subsequent five columns demonstrate the performance of the estimators on a dataset with $n = 200$ under four contamination scenarios: stripe and square contamination at a 15% contamination rate, 15% Normal-Cauchy mixture, and 15% Normal-Slash mixture.

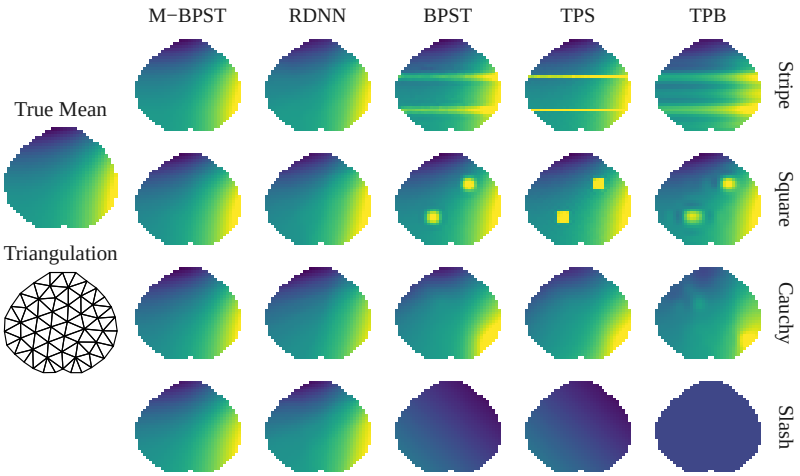


Figure S.4: Simulation run demonstrating the performance of mean function estimation methods on a **cubic** mean function. The leftmost column displays the true mean function alongside the triangulation setting used for the BPST and M-BPST estimators. The subsequent four columns demonstrate the performance of the estimators on a dataset with $n = 200$ under four contamination scenarios: stripe and square contamination at a 15% contamination rate, 15% Normal-Cauchy mixture, and 15% Normal-Slash mixture.

Table S.7: Mean (and SD) of integrated squared errors ($\times 100$) of different estimators for the **quadratic** mean function. Bold values indicate the best-performing estimator in each scenario.

Cont.	$n = 100$					$n = 200$					$n = 400$				
	M-BPST	RDNN	BPST	TPS	TPB	M-BPST	RDNN	BPST	TPS	TPB	M-BPST	RDNN	BPST	TPS	TPB
Stripe															
0.05	**1.77** (**1.87**)	2.71 (2.29)	2.90 (1.63)	5.25 (1.64)	2.83 (1.64)	**0.94** (**0.77**)	1.84 (1.21)	2.29 (1.10)	4.65 (1.10)	2.22 (1.11)	**0.42** (**0.35**)	1.31 (0.62)	1.95 (0.64)	4.32 (0.64)	1.88 (0.64)
0.10	**1.81** (**1.88**)	2.76 (2.16)	7.96 (2.34)	17.5 (2.36)	7.48 (2.34)	**0.96** (**0.78**)	1.75 (1.17)	7.36 (1.77)	16.9 (1.77)	6.89 (1.77)	**0.44** (**0.37**)	1.62 (0.83)	7.01 (1.14)	16.5 (1.15)	6.55 (1.14)
0.15	**1.83** (**1.92**)	2.75 (2.23)	16.4 (3.21)	37.8 (3.23)	15.5 (3.21)	**0.98** (**0.80**)	1.99 (1.35)	15.8 (2.52)	37.3 (2.53)	14.9 (2.51)	**0.48** (**0.41**)	1.45 (0.66)	15.5 (1.66)	37.0 (1.69)	14.5 (1.66)
Square															
0.05	**1.74** (**1.85**)	2.54 (1.78)	2.69 (1.43)	3.07 (1.43)	2.68 (1.43)	**0.94** (**0.76**)	1.80 (0.96)	2.07 (0.83)	2.46 (0.83)	2.07 (0.83)	**0.43** (**0.34**)	1.53 (0.77)	1.74 (0.42)	2.13 (0.42)	1.74 (0.42)
0.10	**1.75** (**1.86**)	2.76 (1.91)	7.07 (1.63)	8.64 (1.64)	6.85 (1.63)	**0.96** (**0.77**)	1.83 (1.12)	6.46 (1.08)	8.04 (1.09)	6.24 (1.09)	**0.44** (**0.35**)	1.55 (0.75)	6.11 (0.62)	7.71 (0.62)	5.89 (0.62)
0.15	**1.77** (**1.87**)	2.91 (2.16)	14.4 (1.92)	18.0 (1.94)	13.5 (1.93)	**0.98** (**0.78**)	1.86 (0.98)	13.8 (1.40)	17.3 (1.40)	12.9 (1.40)	**0.47** (**0.36**)	1.54 (0.82)	13.4 (0.86)	17.0 (0.87)	12.6 (0.86)
Cauchy															
0.05	**1.92** (**1.83**)	2.56 (1.90)	8.2e3 (7.7e4)	1.5e3 (1.4e4)	3.9e3 (3.8e4)	**0.92** (**0.75**)	1.81 (1.09)	261 (1.2e3)	98.8 (497)	243 (1.4e3)	**0.42** (**0.33**)	1.44 (0.76)	1.5e3 (1.5e4)	1.5e3 (1.5e4)	911 (8.8e3)
0.10	**1.99** (**1.80**)	2.77 (2.02)	3.3e4 (3.1e5)	6.2e3 (5.7e4)	1.6e4 (1.5e5)	**0.92** (**0.75**)	1.85 (1.04)	1.0e3 (4.9e3)	386 (2.0e3)	949 (5.8e3)	**0.41** (**0.33**)	1.53 (0.74)	6.1e3 (5.9e4)	5.9e3 (5.8e4)	3.6e3 (3.5e4)
0.15	**1.98** (**1.77**)	2.68 (1.80)	7.4e4 (7.0e5)	1.4e4 (1.3e5)	3.5e4 (3.4e5)	**0.91** (**0.74**)	1.93 (1.21)	2.3e3 (1.1e4)	860 (4.5e3)	2.1e3 (1.3e4)	**0.41** (**0.33**)	1.60 (0.92)	1.4e4 (1.3e5)	1.3e4 (1.3e5)	8.1e3 (7.9e4)
Slash															
0.05	**1.95** (**1.81**)	2.72 (1.95)	116 (743)	42.7 (213)	102 (576)	**0.92** (**0.76**)	1.82 (1.36)	1.8e3 (1.1e4)	918 (8.9e3)	483 (4.5e3)	**0.42** (**0.33**)	1.44 (0.76)	7.6e3 (5.5e4)	3.2e3 (2.6e4)	2.2e3 (1.6e4)
0.10	**1.98** (**1.77**)	2.60 (1.85)	451 (3.0e3)	150 (812)	370 (2.3e3)	**0.94** (**0.80**)	1.76 (0.90)	7.0e3 (4.5e4)	3.7e3 (3.6e4)	1.9e3 (1.8e4)	**0.41** (**0.33**)	1.38 (0.60)	3.1e4 (2.2e5)	1.3e4 (1.0e5)	8.9e3 (6.4e4)
0.15	**1.96** (**1.75**)	2.65 (2.05)	1.0e3 (6.7e3)	321 (1.8e3)	802 (5.1e3)	**0.94** (**0.79**)	2.02 (1.36)	1.8e4 (1.2e5)	8.2e3 (8.0e4)	4.2e3 (4.0e4)	**0.41** (**0.33**)	1.45 (0.73)	6.9e4 (5.0e5)	2.9e4 (2.4e5)	2.0e4 (1.4e5)
Clean	1.73 (1.84)	2.70 (2.02)	1.23 (1.39)	1.23 (1.39)	**1.23** (**1.39**)	0.92 (0.76)	1.85 (1.17)	0.62 (0.71)	0.62 (0.71)	**0.62** (**0.71**)	0.42 (0.33)	1.45 (0.79)	**0.28** (**0.31**)	0.29 (0.31)	0.28 (0.31)

Table S.8: Mean (and SD) of integrated squared errors ($\times 100$) of different estimators for the **exponential** mean function. Bold values indicate the best-performing estimator in each scenario.

Cont.	$n = 100$					$n = 200$					$n = 400$				
	M-BPST	RDNN	BPST	TPS	TPB	M-BPST	RDNN	BPST	TPS	TPB	M-BPST	RDNN	BPST	TPS	TPB
Stripe															
0.05	**1.75**	2.90	2.90	5.25	2.75	**0.82**	2.57	2.29	4.65	2.15	**0.38**	4.00	1.95	4.32	1.81
	(1.86)	(2.23)	(1.63)	(1.64)	(1.62)	**(0.76)**	(1.27)	(1.10)	(1.10)	(1.09)	**(0.34)**	(1.41)	(0.64)	(0.64)	(0.63)
0.10	**1.77**	2.89	7.97	17.5	7.46	**0.85**	2.51	7.36	16.9	6.86	**0.41**	4.32	7.01	16.5	6.53
	(1.89)	(2.38)	(2.34)	(2.36)	(2.33)	**(0.77)**	(1.22)	(1.77)	(1.77)	(1.76)	**(0.37)**	(1.64)	(1.14)	(1.15)	(1.13)
0.15	**1.80**	2.67	16.4	37.8	15.5	**0.90**	2.55	15.8	37.3	14.9	**0.47**	4.00	15.5	37.0	14.6
	(1.92)	(1.69)	(3.21)	(3.23)	(3.20)	**(0.79)**	(1.25)	(2.52)	(2.53)	(2.51)	**(0.41)**	(1.64)	(1.66)	(1.69)	(1.66)
Square															
0.05	**1.76**	2.94	2.69	3.07	2.57	**0.82**	2.46	2.08	2.46	1.96	**0.38**	4.10	1.74	2.13	1.63
	(1.85)	(1.92)	(1.43)	(1.43)	(1.43)	**(0.76)**	(1.19)	(0.83)	(0.83)	(0.83)	**(0.34)**	(1.74)	(0.42)	(0.42)	(0.42)
0.10	**1.78**	2.80	7.07	8.64	6.45	**0.85**	2.40	6.46	8.04	5.85	**0.41**	3.99	6.11	7.71	5.51
	(1.86)	(1.76)	(1.63)	(1.64)	(1.62)	**(0.76)**	(1.20)	(1.08)	(1.09)	(1.08)	**(0.35)**	(1.47)	(0.62)	(0.62)	(0.62)
0.15	**1.81**	2.83	14.4	18.0	12.9	**0.89**	2.49	13.8	17.3	12.3	**0.47**	4.03	13.4	17.0	11.9
	(1.87)	(1.85)	(1.92)	(1.94)	(1.92)	**(0.76)**	(1.02)	(1.40)	(1.40)	(1.39)	**(0.36)**	(1.71)	(0.86)	(0.87)	(0.85)
Cauchy															
0.05	**1.76**	2.66	8.2e3	1.6e3	3.9e3	**0.81**	2.48	261	102	244	**0.38**	3.82	1.5e3	1.5e3	910
	(1.83)	(1.84)	(7.7e4)	(1.5e4)	(3.8e4)	**(0.75)**	(1.15)	(1.2e3)	(509)	(1.4e3)	**(0.34)**	(1.60)	(1.5e4)	(1.5e4)	(8.8e3)
0.10	**1.74**	2.93	3.3e4	6.2e3	1.6e4	**0.80**	2.31	1.0e3	394	951	**0.38**	3.96	6.1e3	5.9e3	3.6e3
	(1.81)	(2.04)	(3.1e5)	(5.8e4)	(1.5e5)	**(0.74)**	(1.16)	(4.9e3)	(2.0e3)	(5.8e3)	**(0.34)**	(1.34)	(5.9e4)	(5.8e4)	(3.5e4)
0.15	**1.76**	2.72	7.4e4	1.4e4	3.5e4	**0.80**	2.42	2.3e3	874	2.1e3	**0.40**	4.01	1.4e4	1.3e4	8.1e3
	(1.79)	(2.00)	(7.0e5)	(1.3e5)	(3.4e5)	**(0.74)**	(1.29)	(1.1e4)	(4.5e3)	(1.3e4)	**(0.35)**	(1.52)	(1.3e5)	(1.3e5)	(7.9e4)
Slash															
0.05	**1.72**	2.57	119	42.4	96.9	**0.80**	2.36	1.8e3	924	485	**0.38**	3.93	7.6e3	3.2e3	2.2e3
	(1.82)	(1.91)	(745)	(198)	(550)	**(0.75)**	(1.08)	(1.1e4)	(8.9e3)	(4.5e3)	**(0.33)**	(1.30)	(5.5e4)	(2.6e4)	(1.6e4)
0.10	**1.70**	2.62	453	145	356	**0.81**	2.51	7.0e3	3.7e3	1.9e3	**0.39**	4.04	3.1e4	1.3e4	8.9e3
	(1.79)	(1.79)	(3.0e3)	(764)	(2.2e3)	**(0.75)**	(1.43)	(4.5e4)	(3.6e4)	(1.8e4)	**(0.34)**	(1.50)	(2.2e5)	(1.0e5)	(6.4e4)
0.15	**1.71**	2.72	1.0e3	312	780	**0.82**	2.39	1.8e4	8.2e3	4.2e3	**0.41**	3.77	6.9e4	2.9e4	2.0e4
	(1.76)	(1.61)	(6.7e3)	(1.7e3)	(5.0e3)	**(0.74)**	(1.28)	(1.2e5)	(8.0e4)	(4.0e4)	**(0.36)**	(1.42)	(5.0e5)	(2.4e5)	(1.4e5)
Clean	1.75	2.77	1.23	**1.23**	1.23	0.81	2.44	0.62	**0.62**	0.62	0.37	3.87	**0.28**	0.28	0.28
	(1.84)	(1.89)	(1.39)	**(1.39)**	(1.39)	(0.76)	(1.23)	(0.71)	**(0.71)**	(0.71)	(0.33)	(1.55)	**(0.31)**	(0.31)	(0.31)

Table S.9: Mean (and SD) of integrated squared errors ($\times 100$) of different estimators for the **cubic** mean function. Bold values indicate the best-performing estimator in each scenario.

Cont.	$\boldsymbol{n = 100}$					$\boldsymbol{n = 200}$					$\boldsymbol{n = 400}$				
	M-BPST	RDNN	BPST	TPS	TPB	M-BPST	RDNN	BPST	TPS	TPB	M-BPST	RDNN	BPST	TPS	TPB
Stripe															
0.05	**1.69**	2.18	2.90	5.25	2.78	**0.79**	1.33	2.29	4.65	2.18	**0.37**	1.46	1.95	4.32	1.84
	(1.85)	(2.12)	(1.63)	(1.64)	(1.64)	**(0.76)**	(1.06)	(1.10)	(1.10)	(1.11)	**(0.34)**	(1.01)	(0.64)	(0.64)	(0.64)
0.10	**1.69**	2.10	7.96	17.5	7.70	**0.81**	1.29	7.36	16.9	7.10	**0.38**	1.35	7.01	16.5	6.77
	(1.87)	(2.02)	(2.34)	(2.36)	(2.37)	**(0.77)**	(0.98)	(1.77)	(1.77)	(1.79)	**(0.37)**	(0.79)	(1.14)	(1.15)	(1.15)
0.15	**1.70**	2.18	16.4	37.8	16.0	**0.83**	1.22	15.8	37.3	15.5	**0.41**	1.32	15.5	37.0	15.1
	(1.90)	(2.16)	(3.21)	(3.23)	(3.27)	**(0.79)**	(0.87)	(2.52)	(2.53)	(2.55)	**(0.40)**	(0.84)	(1.66)	(1.69)	(1.69)
Square															
0.05	**1.69**	2.19	2.69	3.07	2.57	**0.79**	1.30	2.07	2.46	1.96	**0.37**	1.56	1.74	2.13	1.63
	(1.84)	(2.07)	(1.43)	(1.43)	(1.43)	**(0.76)**	(1.17)	(0.83)	(0.83)	(0.83)	**(0.33)**	(1.01)	(0.42)	(0.42)	(0.42)
0.10	**1.70**	2.09	7.07	8.64	6.45	**0.80**	1.26	6.46	8.04	5.84	**0.38**	1.79	6.11	7.71	5.50
	(1.84)	(1.86)	(1.63)	(1.64)	(1.62)	**(0.76)**	(0.94)	(1.08)	(1.09)	(1.08)	**(0.34)**	(1.16)	(0.62)	(0.62)	(0.62)
0.15	**1.71**	2.08	14.4	18.0	12.9	**0.82**	1.31	13.8	17.3	12.3	**0.40**	1.97	13.4	17.0	12.0
	(1.85)	(1.85)	(1.92)	(1.94)	(1.93)	**(0.76)**	(0.98)	(1.40)	(1.40)	(1.40)	**(0.36)**	(1.30)	(0.86)	(0.87)	(0.86)
Cauchy															
0.05	**1.69**	2.16	8.2e3	1.5e3	3.9e3	**0.54**	1.22	2.4e4	1.8e4	3.6e3	**0.26**	0.92	140	47.0	194
	(1.82)	(2.29)	(7.7e4)	(1.4e4)	(3.8e4)	**(0.36)**	(0.68)	(2.3e5)	(1.8e5)	(3.5e4)	**(0.16)**	(0.43)	(1.0e3)	(99.9)	(1.2e3)
0.10	**1.67**	2.12	3.3e4	6.2e3	1.6e4	**0.55**	1.18	9.4e4	7.4e4	1.4e4	**0.26**	0.90	550	134	745
	(1.80)	(1.87)	(3.1e5)	(5.7e4)	(1.5e5)	**(0.36)**	(0.55)	(9.3e5)	(7.3e5)	(1.4e5)	**(0.16)**	(0.43)	(4.1e3)	(303)	(4.7e3)
0.15	**1.66**	2.20	7.4e4	1.4e4	3.5e4	**0.55**	1.14	2.1e5	1.7e5	3.2e4	**0.26**	0.90	1.2e3	266	1.7e3
	(1.77)	(2.11)	(7.0e5)	(1.3e5)	(3.4e5)	**(0.36)**	(0.54)	(2.1e6)	(1.6e6)	(3.2e5)	**(0.16)**	(0.42)	(9.2e3)	(649)	(1.0e4)
Slash															
0.05	**1.67**	2.04	111	37.4	91.5	**0.78**	1.31	1.8e3	915	476	**0.39**	1.42	7.6e3	3.2e3	2.2e3
	(1.81)	(1.87)	(744)	(202)	(560)	**(0.75)**	(0.88)	(1.1e4)	(8.9e3)	(4.5e3)	**(0.33)**	(1.03)	(5.5e4)	(2.6e4)	(1.6e4)
0.10	**1.65**	2.01	438	138	347	**0.78**	1.30	7.0e3	3.7e3	1.9e3	**0.41**	1.42	3.1e4	1.3e4	8.9e3
	(1.79)	(1.84)	(3.0e3)	(792)	(2.2e3)	**(0.74)**	(1.02)	(4.5e4)	(3.6e4)	(1.8e4)	**(0.32)**	(0.86)	(2.2e5)	(1.0e5)	(6.4e4)
0.15	**1.62**	2.12	1.0e3	304	768	**0.80**	1.30	1.6e4	8.2e3	4.2e3	**0.41**	1.58	6.9e4	2.9e4	2.0e4
	(1.77)	(1.98)	(6.7e3)	(1.8e3)	(5.0e3)	**(0.75)**	(1.05)	(1.0e5)	(8.0e4)	(4.0e4)	**(0.32)**	(0.97)	(5.0e5)	(2.4e5)	(1.4e5)
Clean	1.69	2.08	**1.23**	1.23	1.23	0.79	1.24	**0.62**	0.62	0.62	0.36	1.49	**0.28**	0.28	0.28
	(1.83)	(1.90)	**(1.39)**	(1.39)	(1.39)	(0.76)	(0.91)	**(0.71)**	(0.71)	(0.71)	(0.33)	(0.83)	**(0.31)**	(0.31)	(0.31)

S2 Technical Proofs

In the following, we give the proofs to Theorems 1 and 2 described in the main paper. We start by citing two important results from Lai and Schumaker (2007).

Lemma S1. **(Theorem 2.7, Lai and Schumaker (2007))** Let $\{B_m\}_{m\in\mathcal{M}}$ be the Bernstein polynomial basis for spline basis $\mathbb{S}_d^r(\triangle)$ defined over a ϖ-quasi-uniform triangulation $\triangle$. Then there exist positive constants $c,\ C$ depending on the smoothness $r,\ d$, and the shape parameter ϖ such that $c|\triangle|^2\sum_{m\in\mathcal{M}}\gamma_m^2 \leq \left\|\sum_{m\in\mathcal{M}}\gamma_m B_m\right\|_2^2 \leq C|\triangle|^2\sum_{m\in\mathcal{M}}\gamma_m^2$.

Lemma S2. **(Theorem 10.10, Lai and Schumaker (2007))** Let $d\geq 3r+2$, and suppose that $\triangle$ is a ϖ-quasi-uniform triangulation of a polygonal domain Ω, and $g\in\mathcal{W}^{d+1,\infty}(\Omega)$. Then, for bi-integer (a_1,a_2) with $0\leq a_1+a_2\leq d$, there exists a spline $g^*\in\mathbb{S}_d^r(\triangle)$ such that $\|\nabla_{z_1}^{a_1}\nabla_{z_2}^{a_2}(g-g^*)\|_\infty \leq C|\triangle|^{d+1-a_1-a_2}|g|_{d+1}$, where C is a constant depending on $d,\ r$, and the shape parameter ϖ.

Lemma S2 shows that $\mathcal{S}_d^r(\triangle)$ has full approximation power if $d\geq 3r+2$. Define $R_n(\boldsymbol{z})=\widetilde{\mathbf{B}}^\top(\boldsymbol{z})\boldsymbol{\theta}-\mu(\boldsymbol{z})$ for $\boldsymbol{z}\in\Omega$ and some $\boldsymbol{\theta}\in\mathbb{R}^{|\widetilde{\mathcal{M}}|}$. Then, letting $a_1=a_2=0$ in Lemma S2 implies that there exists $\boldsymbol{\theta}^*\in\mathbb{R}^{|\widetilde{\mathcal{M}}|}$ such that

$$\sup_{\boldsymbol{z}\in\Omega}|R_n(\boldsymbol{z})| = \sup_{\boldsymbol{z}\in\Omega}\left|\widetilde{\mathbf{B}}^\top(\boldsymbol{z})\boldsymbol{\theta}^*-\mu(\boldsymbol{z})\right| = O\left(|\triangle|^{d+1}\right). \tag{S2.1}$$

Therefore, to prove Theorem 1, we can replace $\mu(\cdot)$ by $\mu^*(\cdot)=\widetilde{\mathbf{B}}^\top(\cdot)\boldsymbol{\theta}^*$ up to an order of $|\triangle|^{d+1}$, which implies what is left to show is $\|\widehat{\mu}-\mu^*\|_2 =$

$O_p(n^{-1/2}|\triangle|^{-1})$. To that end, we first define

$$\mathbf{S}_{n,h} = \frac{n}{N}\sum_{j=1}^{N}\widetilde{\mathbf{B}}(\boldsymbol{z}_j)\widetilde{\mathbf{B}}^\top(\boldsymbol{z}_j) + \frac{h}{N}\mathbf{Q}_2^\top\left[\langle B_m, B_{m'}\rangle_{\mathcal{E}}\right]_{m,m'\in\mathcal{M}}\mathbf{Q}_2,$$

$$\mathbf{V}(\boldsymbol{z}) = \mathbf{S}_{n,h}^{-1/2}\widetilde{\mathbf{B}}(\boldsymbol{z}),\ \boldsymbol{\alpha} = \mathbf{S}_{n,h}^{1/2}(\boldsymbol{\theta}-\boldsymbol{\theta}^*).$$

The following lemma from Wang et al. (2020) guarantees that the maximum and minimum eigenvalues of the matrix $n^{-1}\mathbf{S}_{n,h}$ are bounded by certain orders, and hence $\mathbf{S}_{n,h}$ is invertible.

Lemma S3. **(Lemma B.6, Wang et al. (2020))** Under Assumption (A1), if $N^{1/2}|\triangle| \to \infty$ as $N \to \infty$, then there exist constants $0 < c_h < C_h < \infty$ such that with probability approaching 1 as $N \to \infty$, $n \to \infty$,

$$c_h|\triangle|^2 \le \lambda_{\min}\left(n^{-1}\mathbf{S}_{n,h}\right) \le \lambda_{\max}\left(n^{-1}\mathbf{S}_{n,h}\right) \le C_h\left(|\triangle|^2 + \frac{h}{nN|\triangle|^2}\right).$$

Let $R_{nj} = R_n(\boldsymbol{z}_j)$. By definition of $\mathbf{V}(\cdot)$ and $\boldsymbol{\alpha}$, we can rewrite $Y_{ij} - \widetilde{\mathbf{B}}^\top(\boldsymbol{z}_j)\boldsymbol{\theta} = e_{ij} - R_{nj} - \mathbf{V}^\top(\boldsymbol{z}_j)\boldsymbol{\alpha}$, which further implies that

$$\min_{\boldsymbol{\theta}\in\mathbb{R}^{|\widetilde{\mathcal{M}}|}}\sum_{i=1}^{n}\sum_{j=1}^{N}\rho\left\{Y_{ij} - \widetilde{\mathbf{B}}^\top(\boldsymbol{z}_j)\boldsymbol{\theta}\right\} + \frac{h}{2}\boldsymbol{\theta}^\top\mathbf{D}\boldsymbol{\theta} \tag{S2.2}$$

$$= \min_{\boldsymbol{\alpha}\in\mathbb{R}^{|\widetilde{\mathcal{M}}|}}\sum_{i=1}^{n}\sum_{j=1}^{N}\rho\left\{e_{ij} - R_{nj} - \mathbf{V}^\top(\boldsymbol{z}_j)\boldsymbol{\alpha}\right\} + \frac{h}{2}\left(\mathbf{S}_{n,h}^{-1/2}\boldsymbol{\alpha} + \boldsymbol{\theta}^*\right)^\top\mathbf{D}\left(\mathbf{S}_{n,h}^{-1/2}\boldsymbol{\alpha} + \boldsymbol{\theta}^*\right).$$

Then, the minimizer $\widehat{\boldsymbol{\alpha}}$ of (S2.2) is also the minimizer of $\boldsymbol{\Gamma}_n(\boldsymbol{\alpha}) + \mathbf{P}_h(\boldsymbol{\alpha})$, where

$$\boldsymbol{\Gamma}_n(\boldsymbol{\alpha}) = \sum_{i=1}^{n} \frac{1}{N} \sum_{j=1}^{N} \left[\rho \left\{ e_{ij} - R_{nj} - \mathbf{V}^\top(\boldsymbol{z}_j)\boldsymbol{\alpha} \right\} - \rho \left(e_{ij} - R_{nj} \right) \right],$$

$$\text{and } \mathbf{P}_h(\boldsymbol{\alpha}) = \frac{h}{2N} \left(\mathbf{S}_{n,h}^{-1/2}\boldsymbol{\alpha} + \boldsymbol{\theta}^* \right)^\top \mathbf{D} \left(\mathbf{S}_{n,h}^{-1/2}\boldsymbol{\alpha} + \boldsymbol{\theta}^* \right) - \frac{h}{2N} \boldsymbol{\theta}^{*\top} \mathbf{D} \boldsymbol{\theta}^*.$$

This is true because the additional terms do not include the parameter $\boldsymbol{\alpha}$. Define further

$$\boldsymbol{\Xi}_n(\boldsymbol{\alpha}) = \boldsymbol{\Gamma}_n(\boldsymbol{\alpha}) - \mathbb{E}\left\{ \boldsymbol{\Gamma}_n(\boldsymbol{\alpha}) \right\} + \sum_{i=1}^{n} \frac{1}{N} \sum_{j=1}^{N} \left\{ \psi(e_{ij}) \mathbf{V}^\top(\boldsymbol{z}_j)\boldsymbol{\alpha} \right\}. \tag{S2.3}$$

To prove Theorem 1, We first obtain the asymptotic bounds of the terms $\boldsymbol{\Xi}_n(\boldsymbol{\alpha})$, $\boldsymbol{\Gamma}_n(\boldsymbol{\alpha})$, and $\mathbb{E}\{\boldsymbol{\Gamma}_n(\boldsymbol{\alpha})\}$ in the following three lemmas.

Lemma S4. Under Assumptions (A1)–(A5) and (A7), for a fixed constant $L > 1$,

$$\sup_{\|\boldsymbol{\alpha}\|_2 \leq L} \left| |\triangle|^2 \boldsymbol{\Xi}_n \left(|\triangle|^{-1} \boldsymbol{\alpha} \right) \right| = o_P(1).$$

Proof. We will prove that $\mathbb{P}(\sup_{\|\boldsymbol{\alpha}\|_2 \leq L} ||\triangle|^2 \boldsymbol{\Xi}_n(|\triangle|^{-1}\boldsymbol{\alpha})| \geq \epsilon) \to 0$ for any $\epsilon > 0$ using Bernstein's Inequality. First, we define $\Pi = \{\boldsymbol{\alpha} \in \mathbb{R}^{|\widetilde{M}|}; \|\boldsymbol{\alpha}\|_2 \leq L\}$, and find a decomposition $\Pi = \Pi_1 \cup \cdots \cup \Pi_{K_n}$, where $\{\Pi_k\}_{k=1}^{K_n}$ are pairwise disjoint sets and for any $1 \leq k \leq K_n$, $\text{diam}(\Pi_k) = \max_{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2 \in \Pi_k} \{\|\boldsymbol{\alpha}_1 - \boldsymbol{\alpha}_2\|_2\} \leq q_0 = \epsilon C^{-1} |\triangle|^{-2} n^{-1}$. Notice that we can find such decomposition with $K_n \leq (2\sqrt{|\widetilde{M}|}/q_0 + 1)^{|\widetilde{M}|}$ (Lima et al., 2019). For each $1 \leq k \leq K_n$, select

$\boldsymbol{\alpha}_k \in \Pi_k$. Then, by (S2.3) we have

$$\begin{aligned}
&\min_{1\le k\le K_n} |\triangle|^2 \left|\boldsymbol{\Xi}_n\left(|\triangle|^{-1}\boldsymbol{\alpha}\right) - \boldsymbol{\Xi}_n\left(|\triangle|^{-1}\boldsymbol{\alpha}_k\right)\right| \\
&\quad \le \min_{1\le k\le K_n} |\triangle|^2 \left|\boldsymbol{\Gamma}_n\left(|\triangle|^{-1}\boldsymbol{\alpha}\right) - \boldsymbol{\Gamma}_n\left(|\triangle|^{-1}\boldsymbol{\alpha}_k\right)\right| \\
&\qquad + \min_{1\le k\le K_n} |\triangle|^2 \left|\mathbb{E}\left\{\boldsymbol{\Gamma}_n\left(|\triangle|^{-1}\boldsymbol{\alpha}\right)\right\} - \mathbb{E}\left\{\boldsymbol{\Gamma}_n\left(|\triangle|^{-1}\boldsymbol{\alpha}_k\right)\right\}\right| \\
&\qquad + \min_{1\le k\le K_n} |\triangle|^2 \left|\,|\triangle|^{-1}\sum_{i=1}^{n}\frac{1}{N}\sum_{j=1}^{N}\left\{\psi(e_{ij})\mathbf{V}^\top(\boldsymbol{z}_j)\boldsymbol{\alpha} - \psi(e_{ij})\mathbf{V}^\top(\boldsymbol{z}_j)\boldsymbol{\alpha}_k\right\}\right| \\
&\quad := I + II + III.
\end{aligned}$$

We then proceed to find asymptotic upper bounds for $I,\ II,$ and III. By definition of $\boldsymbol{\Gamma}_n(|\triangle|^{-1}\boldsymbol{\alpha})$ and applying the mean value theorem on ρ, we obtain

$$\begin{aligned}
I = \min_{1\le k\le K_n} |\triangle|^2 &\left|\sum_{i=1}^{n}\frac{1}{N}\sum_{j=1}^{N}\rho\left\{e_{ij} - R_{nj} - |\triangle|^{-1}\mathbf{V}^\top(\boldsymbol{z}_j)\boldsymbol{\alpha}\right\}\right. \qquad \text{(S2.4)} \\
&\left. -\rho\left\{e_{ij} - R_{nj} - |\triangle|^{-1}\mathbf{V}^\top(\boldsymbol{z}_j)\boldsymbol{\alpha}_k\right\}\right| \\
\le C|\triangle|^2 n \max_{\boldsymbol{z}\in\Omega}\left\|\mathbf{V}(\boldsymbol{z})\right\|_2 &|\triangle|^{-1} \min_{1\le k\le K_n}\left\|\boldsymbol{\alpha} - \boldsymbol{\alpha}_k\right\|_2. \qquad \text{(S2.5)}
\end{aligned}$$

Similarly, we can show that the upper bounds of II and III have the same order as I. Therefore,

$$\begin{aligned}
&\min_{1\le k\le K_n} |\triangle|^2 \left|\boldsymbol{\Xi}_n\left(|\triangle|^{-1}\boldsymbol{\alpha}\right) - \boldsymbol{\Xi}_n\left(|\triangle|^{-1}\boldsymbol{\alpha}_k\right)\right| \\
&\qquad \le 3C|\triangle|^2 n \max_{\boldsymbol{z}\in\Omega}\left\|\mathbf{V}(\boldsymbol{z})\right\|_2 |\triangle|^{-1} \min_{1\le k\le K_n}\left\|\boldsymbol{\alpha} - \boldsymbol{\alpha}_k\right\|_2 \\
&\qquad \le 3C|\triangle|^2 q_0 d_n = 3\epsilon d_n,
\end{aligned}$$

where q_0 is defined at the beginning of the proof, and

$$d_n = |\triangle|^{-2} \max_{\boldsymbol{z}\in\Omega} \|\mathbf{V}(\boldsymbol{z})\|_2 |\triangle|^{-1} = O(|\triangle|^{-3} n^{-1/2}) = o(1),$$

and the last equality is determined by Assumption (A3). To compute the order of $\max_{\boldsymbol{z}\in\Omega} \|\mathbf{V}(\boldsymbol{z})\|_2$, note that $\mathbf{V}(\boldsymbol{z}) = \mathbf{S}_{n,h}^{-1/2} \widetilde{\mathbf{B}}(\boldsymbol{z})$, and $\max_{\boldsymbol{z}\in\Omega} \|\mathbf{V}(\boldsymbol{z})\|_2 \asymp \|\mathbf{S}_{n,h}^{-1/2}\|_2 \max_{\boldsymbol{z}\in\Omega} \|\widetilde{\mathbf{B}}(\boldsymbol{z})\|_2$. By Lemma S3, $\|\mathbf{S}_{n,h}^{-1/2}\|_2 = O(n^{-1/2}|\triangle|^{-1})$. Since $\|\widetilde{\mathbf{B}}(\boldsymbol{z})\|_2 = O(|\triangle|)$ for all $\boldsymbol{z} \in \Omega$, we have $\max_{\boldsymbol{z}\in\Omega} \|\mathbf{V}(\boldsymbol{z})\|_2 = O(n^{-1/2})$. We can choose $d_n < 1/12$ to have

$$\min_{1\le k\le K_n} |\triangle|^2 \left| \boldsymbol{\Xi}_n \left(|\triangle|^{-1} \boldsymbol{\alpha} \right) - \boldsymbol{\Xi}_n \left(|\triangle|^{-1} \boldsymbol{\alpha}_k \right) \right| < \epsilon/4.$$

Next, for any $1 \le i \le n$, $1 \le j \le N$, and $\boldsymbol{\alpha} \in \mathbb{R}^{|\widetilde{\mathcal{M}}|}$, define

$$\omega_{ij}(\boldsymbol{\alpha}) = \rho\left\{ e_{ij} - R_{nj} - |\triangle|^{-1} \mathbf{V}^\top(\boldsymbol{z}_j)\boldsymbol{\alpha} \right\} - \rho\left(e_{ij} - R_{nj}\right) + \psi(e_{ij}) |\triangle|^{-1} V(\boldsymbol{z}_j)^\top \boldsymbol{\alpha}.$$

By a similar argument to (S2.4), we can show $\sup_{\|\boldsymbol{\alpha}\|_2\le L} |\omega_{ij}(\boldsymbol{\alpha})| = O(|\triangle|^{-1} n^{-1/2})$. Consequently,

$$\sup_{\|\boldsymbol{\alpha}\|_2\le L} |\omega_{ij}(\boldsymbol{\alpha}) - \mathbb{E}\left\{\omega_{ij}(\boldsymbol{\alpha})\right\}| = O\left(|\triangle|^{-1} n^{-1/2} \right), \tag{S2.6}$$

which then implies that

$$\sup_{\|\boldsymbol{\alpha}\|_2\le L} \left| \sum_{i=1}^{n} \mathrm{Var}\left\{ N^{-1} \sum_{j=1}^{N} \omega_{ij}(\boldsymbol{\alpha}) \right\} \right| = O\left(|\triangle|^{-2} N^{-1} \right). \tag{S2.7}$$

Using these upper bounds, by Bernstein's Inequality and Assumption (A3), we

have

$$\begin{aligned}
\mathbb{P}\left\{\sup_{\|\boldsymbol{\alpha}\|_2\le L}\left||\triangle|^2\boldsymbol{\Xi}_n\left(|\triangle|^{-1}\boldsymbol{\alpha}\right)\right|\ge\epsilon\right\} &\le \sum_{k=1}^{K_n}\mathbb{P}\left\{\left|\boldsymbol{\Xi}_n\left(|\triangle|^{-1}\boldsymbol{\alpha}_k\right)\right|\ge\frac{\epsilon}{2|\triangle|^2}\right\}\\
&\le \sum_{k=1}^{K_n}\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n}\left[\frac{1}{N}\sum_{j=1}^{N}\left|\omega_{ij}\left(\boldsymbol{\alpha}_k\right)-\mathbb{E}\left\{\omega_{ij}\left(\boldsymbol{\alpha}_k\right)\right\}\right|\right]\ge\frac{\epsilon}{2n|\triangle|^2}\right)\\
&\le K_n\exp\left\{-\frac{Cn\left(\frac{\epsilon}{2n|\triangle|^2}\right)^2}{\frac{1}{nN|\triangle|^2}+\left(\frac{\epsilon}{2n|\triangle|^2}\right)\frac{1}{|\triangle|n^{1/2}}}\right\}\\
&\le \exp\left\{-O\left(\frac{Nn^{1/2}}{n^{1/2}|\triangle|^2+|\triangle|N}\right)+O(|\triangle|^{-2}\log|\triangle|)+O(|\triangle|^{-2}\log n)\right\}=o(1).
\end{aligned}$$

This concludes the proof of Lemma S4. ■

Lemma S5. Under Assumptions (A1), (A4), and (A5), for a fixed constant $L>1$,

$$\sup_{\|\boldsymbol{\alpha}\|_2\le L}\left||\triangle|\sum_{i=1}^{n}\frac{1}{N}\sum_{j=1}^{N}\left\{\psi(e_{ij})\mathbf{V}^\top(\boldsymbol{z}_j)\boldsymbol{\alpha}\right\}\right|=o_P(1).$$

Proof. First, by Assumption (A4), $\mathbb{E}\{\psi(e_{ij})\}^2$ is bounded. Then, by Assumption (A5) and Lemma S3, for $\|\boldsymbol{\alpha}\|_2\le L$ one has

$$\begin{aligned}
\text{Var}\left[|\triangle|\sum_{i=1}^{n}\frac{1}{N}\sum_{j=1}^{N}\left\{\psi(e_{ij})\mathbf{V}^\top(\boldsymbol{z}_j)\boldsymbol{\alpha}\right\}\right] &= |\triangle|^2\frac{1}{N^2}\sum_{i=1}^{n}\sum_{j=1}^{N}\mathbb{E}\left\{\psi(e_{ij})\right\}^2\left\{\mathbf{V}^\top(\boldsymbol{z}_j)\boldsymbol{\alpha}\right\}^2\\
&= O\left(|\triangle|^2N^{-1}\right).
\end{aligned}$$

Finally, by Chebyshev's Inequality, for any $\epsilon > 0$,

$$\mathbb{P}\left[\sup_{\|\boldsymbol{\alpha}\|_2 \le L}\left||\triangle|\sum_{i=1}^{n}\frac{1}{N}\sum_{j=1}^{N}\left\{\psi(e_{ij})\mathbf{V}^\top(\boldsymbol{z}_j)\boldsymbol{\alpha}\right\}\right| > \epsilon\right] \le \frac{C|\triangle|^2}{\epsilon^2 N} \to 0.$$

This concludes the proof of Lemma S5. ■

Lemma S6. Under Assumptions (A1)–(A6), for a fixed constant $L > 1$,

$$\mathbb{P}\left[\inf_{\|\boldsymbol{\alpha}\|_2 = L}|\triangle|^2\mathbb{E}\left\{\boldsymbol{\Gamma}_n(|\triangle|^{-1}\boldsymbol{\alpha})\right\} > 0\right] \to 1.$$

Proof. By Lemma S3, we can assume that $\max_{\|\boldsymbol{\alpha}\|_2=L}(\||\triangle|^{-1}\mathbf{V}^\top(\boldsymbol{z}_j)\boldsymbol{\alpha}\|_2 + |R_{nj}|)$ following Wei and He (2006). Then, by Assumption (A5), we have

$$\begin{aligned}
|\triangle|^2\mathbb{E}\left\{\boldsymbol{\Gamma}_n\left(|\triangle|^{-1}\boldsymbol{\alpha}\right)\right\} &= |\triangle|^2\sum_{i=1}^{n}\frac{1}{N}\sum_{j=1}^{N}\int_{-R_{nj}}^{-R_{nj}-|\triangle|^{-1}\mathbf{V}^\top(\boldsymbol{z}_j)\boldsymbol{\alpha}}\mathbb{E}\left\{\psi(e_{ij}+u)\right\}du \\
&= |\triangle|^2\sum_{i=1}^{n}\frac{1}{N}\sum_{j=1}^{N}\int_{-R_{nj}}^{-R_{nj}-|\triangle|^{-1}\mathbf{V}^\top(\boldsymbol{z}_j)\boldsymbol{\alpha}}\delta(\boldsymbol{z}_j)u + O(u^2)du \\
&= |\triangle|^2\sum_{i=1}^{n}\frac{1}{N}\sum_{j=1}^{N}\delta(\boldsymbol{z}_j)\left[\frac{1}{2}\left\{R_{nj}+|\triangle|^{-1}\mathbf{V}^\top(\boldsymbol{z}_j)\boldsymbol{\alpha}\right\}^2 - \frac{1}{2}R_{nj}^2\right] \\
&\quad + |\triangle|^2\sum_{i=1}^{n}\frac{1}{N}\sum_{j=1}^{N}C\left[-\frac{1}{3}\left\{R_{nj}+|\triangle|^{-1}\mathbf{V}^\top(\boldsymbol{z}_j)\boldsymbol{\alpha}\right\}^3 + \frac{1}{3}R_{nj}^3\right] \\
&\quad - |\triangle|^2\sum_{i=1}^{n}\frac{1}{N}\sum_{j=1}^{N}C|\triangle|^{-1}\left\|\mathbf{V}^\top(\boldsymbol{z}_j)\boldsymbol{\alpha}\right\|_2\max_{\|\boldsymbol{\alpha}\|_2=L}\left(\left\||\triangle|^{-1}\mathbf{V}^\top(\boldsymbol{z}_j)\boldsymbol{\alpha}\right\|_2 + |R_{nj}|\right)^2.
\end{aligned}$$

Thus,

$$
\begin{aligned}
&|\triangle|^2\mathbb{E}\left\{\mathbf{\Gamma}_n\left(|\triangle|^{-1}\boldsymbol{\alpha}\right)\right\} \\
&\geq |\triangle|^2\sum_{i=1}^{n}\frac{1}{N}\sum_{j=1}^{N}\delta(\boldsymbol{z}_j)\left[\frac{1}{2}|\triangle|^{-2}\left\|\mathbf{V}^\top(\boldsymbol{z}_j)\boldsymbol{\alpha}\right\|_2^2+\left\{R_{nj}|\triangle|^{-1}\mathbf{V}^\top(\boldsymbol{z}_j)\boldsymbol{\alpha}\right\}\right] \\
&\quad\geq Cn\inf_{\boldsymbol{z}\in\Omega}\delta(\boldsymbol{z})\left\|\boldsymbol{\alpha}\right\|_2^2-\sum_{i=1}^{n}\frac{1}{N}\sum_{j=1}^{N}R_{nj}|\triangle|^{-1}\mathbf{V}^\top(\boldsymbol{z}_j)\boldsymbol{\alpha}+o(1) \\
&\quad= CnL^2-CnL+o(1),
\end{aligned}
$$

which is positive for large L. This finishes the proof of the lemma. ■

Proof of Theorem 1: By Theorem 2.19 in Lai and Schumaker (2007), the (m,m')-th element of the penalty matrix $\mathbf{P}$ is $O(|\triangle|^{-2})$ for any $m,\ m'\in\mathcal{M}$. It follows that, with $\|\boldsymbol{\alpha}\|_2=L$,

$$
\begin{aligned}
|\triangle|^2\mathbf{P}_h\left(|\triangle|^{-1}\boldsymbol{\alpha}\right) &= |\triangle|^2\left\{\frac{h}{2N}\left(\mathbf{S}_{n,h}^{-1/2}|\triangle|^{-1}\boldsymbol{\alpha}+\boldsymbol{\theta}^*\right)^\top\mathbf{D}\left(\mathbf{S}_{n,h}^{-1/2}|\triangle|^{-1}\boldsymbol{\alpha}+\boldsymbol{\theta}^*\right)\right. \\
&\qquad\left.-\frac{h}{2N}\boldsymbol{\theta}^{*\top}\mathbf{D}\boldsymbol{\theta}^*\right\} \\
&= \frac{h}{2N}\left(2|\triangle|\boldsymbol{\alpha}^\top\mathbf{S}_{n,h}^{-1/2}\mathbf{Q}_2^\top\mathbf{P}\mathbf{Q}_2\boldsymbol{\theta}^*+\boldsymbol{\alpha}^\top\mathbf{S}_{n,h}^{-1/2}\mathbf{Q}_2^\top\mathbf{P}\mathbf{Q}_2\mathbf{S}_{n,h}^{-1/2}\boldsymbol{\alpha}\right) \\
&= O\left(h|\triangle|^{-3}n^{-1/2}N^{-1}\right),
\end{aligned}
$$

which, by Assumption (A3), is $o(1)$. Then, combining the above result with Lemmas S4, S5, and S6 we obtain that, for any $\epsilon>0$,

$$
\mathbb{P}\left[\inf_{\|\boldsymbol{\alpha}\|_2=L}|\triangle|^2\left\{\mathbf{\Gamma}_n\left(|\triangle|^{-1}\boldsymbol{\alpha}\right)+\mathbf{P}_h\left(|\triangle|^{-1}\boldsymbol{\alpha}\right)\right\}\right]>1-\epsilon. \tag{S2.8}
$$

By convexity and Corollary 25 of Eggleston (1958), (S2.8) implies that

$$\mathbb{P}\left[\inf_{\|\boldsymbol{\alpha}\|_2 \geq L|\triangle|^{-1}} |\triangle|^2 \left\{\boldsymbol{\Gamma}_n(\boldsymbol{\alpha}) + \mathbf{P}_h(\boldsymbol{\alpha})\right\}\right] > 1 - \epsilon. \tag{S2.9}$$

Let $\widehat{\boldsymbol{\alpha}} = \mathbf{S}_{n,h}^{1/2}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*) = \arg\min_{\boldsymbol{\alpha}\in\mathbb{R}^{|\widetilde{\mathcal{M}}|}} \boldsymbol{\Gamma}_n(\boldsymbol{\alpha}) + \mathbf{P}_h(\boldsymbol{\alpha})$, then (S2.9) implies $\|\widehat{\boldsymbol{\alpha}}\|_2 = O_P(|\triangle|^{-1})$, which further implies that $\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\|_2 = O_P(n^{-1/2}|\triangle|^{-2})$. Next, by Lemma S1,

$$\|\widehat{\mu} - \mu^*\|_2 = \left\|\widetilde{\mathbf{B}}^\top \left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\right)\right\|_2 = O_p\left(|\triangle| \times n^{-1/2}|\triangle|^{-2}\right) = O_p\left(n^{-1/2}|\triangle|^{-1}\right).$$

Finally, by the triangle inequality,

$$\|\widehat{\mu} - \mu\|_2 = O_p\left(n^{-1/2}|\triangle|^{-1}\right) + O_p\left(|\triangle|^{d+1}\right),$$

which completes the proof of Theorem 1. ■

Next, we prove the asymptotic normality of the estimator $\widehat{\mu}(\boldsymbol{z}_j)$ for $\boldsymbol{z}_j \in \Omega$. Define

$$\mathbf{L}_{n,0} = \frac{n}{N}\sum_{j=1}^{N} \delta(\boldsymbol{z}_j)\widetilde{\mathbf{B}}(\boldsymbol{z}_j)\widetilde{\mathbf{B}}^\top(\boldsymbol{z}_j), \text{ and } \widetilde{\mathbf{L}}_{n,h^*} = \mathbf{S}_{n,h}^{-1/2}\mathbf{L}_{n,h^*}\mathbf{S}_{n,h}^{-1/2},$$

where h^* indicates the dependency on $\mathbf{L}_{n,h^*}$. In particular,

$$\widetilde{\mathbf{L}}_{n,0} = \mathbf{S}_{n,h}^{-1/2}\mathbf{L}_{n,0}\mathbf{S}_{n,h}^{-1/2}, \text{ and } \widetilde{\mathbf{L}}_{n,h} = \mathbf{S}_{n,h}^{-1/2}\mathbf{L}_{n,h}\mathbf{S}_{n,h}^{-1/2}.$$

Define also

$$\widetilde{\boldsymbol{\alpha}} = \widetilde{\mathbf{L}}_{n,h}^{-1}\sum_{i=1}^{n}\frac{1}{N}\sum_{j=1}^{N}\mathbf{V}(\boldsymbol{z}_j)\psi\left(e_{ij}\right).$$

The following lemma gives the asymptotic upper bound on the difference between $\widehat{\boldsymbol{\alpha}}$ and $\widetilde{\boldsymbol{\alpha}}$.

Lemma S7. Under Assumptions (A1) – (A8), assuming additionally $|\triangle| \ll n^{-1/\{2(d+2)\}}$, if $N^{1/2}|\triangle| \to \infty$ as $N \to \infty$, we have $|\triangle| \left\|\widehat{\boldsymbol{\alpha}} - \widetilde{\boldsymbol{\alpha}}\right\|_2 = o_P(1)$.

Proof. By Assumption (A6) and Lemma S3, the matrix $\widetilde{\mathbf{L}}_{n,h}$ is invertible, and for all n, $\lambda_{\min}(\widetilde{\mathbf{L}}_{n,h}) > \widetilde{\lambda}_0 > 0$ for some constant $\widetilde{\lambda}_0$. We will use an analogy to the proof of Theorem 1 and show that, for any $\epsilon > 0$,

$$\mathbb{P}\left[\inf_{\|\boldsymbol{\alpha}-\widetilde{\boldsymbol{\alpha}}\|_2 \geq \epsilon|\triangle|^{-1}} |\triangle|^2 \left\{\boldsymbol{\Gamma}_n(\boldsymbol{\alpha}) + \mathbf{P}_h(\boldsymbol{\alpha}) - \boldsymbol{\Gamma}_n(\widetilde{\boldsymbol{\alpha}}) - \mathbf{P}_h(\widetilde{\boldsymbol{\alpha}})\right\} > 0\right] \to 1.$$

In fact, by convexity of $\boldsymbol{\Gamma}_n + \mathbf{P}_h$, we only need to show

$$\mathbb{P}\left[\inf_{\|\boldsymbol{\alpha}-\widetilde{\boldsymbol{\alpha}}\|_2 = \epsilon|\triangle|^{-1}} |\triangle|^2 \left\{\boldsymbol{\Gamma}_n(\boldsymbol{\alpha}) + \mathbf{P}_h(\boldsymbol{\alpha}) - \boldsymbol{\Gamma}_n(\widetilde{\boldsymbol{\alpha}}) - \mathbf{P}_h(\widetilde{\boldsymbol{\alpha}})\right\} > 0; \|\widetilde{\boldsymbol{\alpha}}\|_2 \leq L|\triangle|^{-1}\right] \to 1. \tag{S2.10}$$

By equation (S2.3) and using a similar argument as in Lemma S6, we have

$$\begin{aligned}
|\triangle|^2 \mathbb{E}\left\{\boldsymbol{\Gamma}_n(\boldsymbol{\alpha})\right\} &= |\triangle|^2 \sum_{i=1}^{n} \frac{1}{N} \sum_{j=1}^{N} \int_{-R_{nj}}^{-R_{nj}-\mathbf{V}^\top(\boldsymbol{z}_j)\boldsymbol{\alpha}} \mathbb{E}\left\{\psi(e_{ij}+u)\right\} du \\
&= |\triangle|^2 \sum_{i=1}^{n} \frac{1}{N} \sum_{j=1}^{N} \int_{-R_{nj}}^{-R_{nj}-\mathbf{V}^\top(\boldsymbol{z}_j)\boldsymbol{\alpha}} \delta(\boldsymbol{z}_j)u + O(u^2) du \\
&= |\triangle|^2 \sum_{i=1}^{n} \frac{1}{N} \sum_{j=1}^{N} \delta(\boldsymbol{z}_j) \left[\frac{1}{2}\left\|\mathbf{V}^\top(\boldsymbol{z}_j)\boldsymbol{\alpha}\right\|_2^2 + \left\{R_{nj}\mathbf{V}^\top(\boldsymbol{z}_j)\boldsymbol{\alpha}\right\}\right] + o(1) \\
&= \frac{|\triangle|^2}{2} \boldsymbol{\alpha}^\top \widetilde{\mathbf{L}}_{n,0} \boldsymbol{\alpha} + o(1),
\end{aligned}$$

where the last $o(1)$ comes from the additional condition $|\triangle| \ll n^{-1/\{2(d+2)\}}$.

Note also $nN^{-1}\sum_{j=1}^{N}\psi(e_{ij})\mathbf{V}^{\top}(\boldsymbol{z}_j)\boldsymbol{\alpha}=\widetilde{\boldsymbol{\alpha}}^{\top}\widetilde{\mathbf{L}}_{n,h}\boldsymbol{\alpha}$. Therefore,

$$\begin{aligned}|\triangle|^2\boldsymbol{\Gamma}_n(\boldsymbol{\alpha}) &= |\triangle|^2\left[\boldsymbol{\Xi}_n(\boldsymbol{\alpha})+\mathbb{E}\left\{\boldsymbol{\Gamma}_n(\boldsymbol{\alpha})\right\}-\sum_{i=1}^{n}\frac{1}{N}\sum_{j=1}^{N}\left\{\psi(e_{ij})\mathbf{V}^{\top}(\boldsymbol{z}_j)\boldsymbol{\alpha}\right\}\right]\\ &= |\triangle|^2\left(\boldsymbol{\Xi}_n(\boldsymbol{\alpha})+\frac{1}{2}\boldsymbol{\alpha}^{\top}\widetilde{\mathbf{L}}_{n,0}\boldsymbol{\alpha}-\widetilde{\boldsymbol{\alpha}}^{\top}\widetilde{\mathbf{L}}_{n,h}\boldsymbol{\alpha}\right)+o(1).\end{aligned}$$

For the penalty part, from the proof of Theorem 1 we have

$$|\triangle|^2\mathbf{P}_h\left(\boldsymbol{\alpha}\right)=|\triangle|^2\frac{h}{2N}\left\{2\boldsymbol{\alpha}^{\top}\mathbf{S}_{n,h}^{-1/2}\mathbf{Q}_2^{\top}\mathbf{P}\mathbf{Q}_2\boldsymbol{\theta}^*+\boldsymbol{\alpha}^{\top}\mathbf{S}_{n,h}^{-1/2}\mathbf{Q}_2^{\top}\mathbf{P}\mathbf{Q}_2\mathbf{S}_{n,h}^{-1/2}\boldsymbol{\alpha}\right\}.$$

Combining the two yields

$$\begin{aligned}|\triangle|^2\left\{\boldsymbol{\Gamma}_n(\boldsymbol{\alpha})+\mathbf{P}_h\left(\boldsymbol{\alpha}\right)\right\} &= |\triangle|^2\left\{\boldsymbol{\Xi}_n(\boldsymbol{\alpha})+\frac{1}{2}\boldsymbol{\alpha}^{\top}\mathbf{S}_{n,h}^{-1/2}\left(\mathbf{L}_{n,0}+\frac{h}{N}\mathbf{Q}_2^{\top}\mathbf{P}\mathbf{Q}_2\right)\mathbf{S}_{n,h}^{-1/2}\boldsymbol{\alpha}\right.\\ &\qquad\left.+\frac{h}{N}\boldsymbol{\alpha}^{\top}\mathbf{S}_{n,h}^{-1/2}\mathbf{Q}_2^{\top}\mathbf{P}\mathbf{Q}_2\boldsymbol{\theta}^*-\widetilde{\boldsymbol{\alpha}}^{\top}\widetilde{\mathbf{L}}_{n,h}\boldsymbol{\alpha}\right\}+o(1)\\ &= |\triangle|^2\left\{\boldsymbol{\Xi}_n(\boldsymbol{\alpha})+\frac{1}{2}\boldsymbol{\alpha}^{\top}\widetilde{\mathbf{L}}_{n,h}\boldsymbol{\alpha}+\frac{h}{N}\boldsymbol{\alpha}^{\top}\mathbf{S}_{n,h}^{-1/2}\mathbf{Q}_2^{\top}\mathbf{P}\mathbf{Q}_2\boldsymbol{\theta}^*\right.\\ &\qquad\left.-\widetilde{\boldsymbol{\alpha}}^{\top}\widetilde{\mathbf{L}}_{n,h}\boldsymbol{\alpha}\right\}+o(1).\end{aligned}$$

Then, using the fact that

$$2\widetilde{\boldsymbol{\alpha}}^{\top}\widetilde{\mathbf{L}}_{n,h}\boldsymbol{\alpha}=\boldsymbol{\alpha}^{\top}\widetilde{\mathbf{L}}_{n,h}\boldsymbol{\alpha}+\widetilde{\boldsymbol{\alpha}}^{\top}\widetilde{\mathbf{L}}_{n,h}\widetilde{\boldsymbol{\alpha}}-(\boldsymbol{\alpha}-\widetilde{\boldsymbol{\alpha}})^{\top}\widetilde{\mathbf{L}}_{n,h}(\boldsymbol{\alpha}-\widetilde{\boldsymbol{\alpha}}),$$

we obtain

$$\begin{aligned}&|\triangle|^2\left\{\boldsymbol{\Gamma}_n(\boldsymbol{\alpha})+\mathbf{P}_h(\boldsymbol{\alpha})-\boldsymbol{\Gamma}_n(\widetilde{\boldsymbol{\alpha}})-\mathbf{P}_h(\widetilde{\boldsymbol{\alpha}})\right\}\\ &\quad= |\triangle|^2\left\{\boldsymbol{\Xi}_n(\boldsymbol{\alpha})-\boldsymbol{\Xi}_n(\widetilde{\boldsymbol{\alpha}})+(\boldsymbol{\alpha}-\widetilde{\boldsymbol{\alpha}})^{\top}\widetilde{\mathbf{L}}_{n,h}(\boldsymbol{\alpha}-\widetilde{\boldsymbol{\alpha}})\right.\\ &\qquad\left.+\frac{h}{N}(\boldsymbol{\alpha}-\widetilde{\boldsymbol{\alpha}})^{\top}\mathbf{S}_{n,h}^{-1/2}\mathbf{Q}_2^{\top}\mathbf{P}\mathbf{Q}_2\boldsymbol{\theta}^*\right\}+o(1). \qquad \text{(S2.11)}\end{aligned}$$

Next, we find the order of $\|\widetilde{\boldsymbol{\alpha}}\|_2$. Note that

$$\begin{aligned}\|\widetilde{\boldsymbol{\alpha}}\|_2^2 &\asymp \frac{1}{N^2}\sum_{i,i'=1}^{n}\sum_{j,j'=1}^{N}\mathbf{V}^\top(\boldsymbol{z}_j)\mathbf{V}(\boldsymbol{z}_{j'})\psi\left(e_{ij}\right)\psi\left(e_{i'j'}\right)\\ &\asymp |\triangle|^{-2}n^{-1}\frac{1}{N^2}\sum_{i,i'=1}^{n}\sum_{j,j'=1}^{N}\widetilde{\mathbf{B}}^\top(\boldsymbol{z}_j)\widetilde{\mathbf{B}}(\boldsymbol{z}_{j'})\psi\left(e_{ij}\right)\psi\left(e_{i'j'}\right).\end{aligned}$$

By the independence of errors, for $i \neq i',\ j,\ j'$, we have

$$\mathbb{E}\left\{\widetilde{\mathbf{B}}^\top(\boldsymbol{z}_j)\widetilde{\mathbf{B}}(\boldsymbol{z}_{j'})\psi\left(e_{ij}\right)\psi\left(e_{i'j'}\right)\right\} = \sum_{m\in\widetilde{\mathcal{M}}}\widetilde{B}_m(\boldsymbol{z}_j)\widetilde{B}_m(\boldsymbol{z}_{j'})\mathbb{E}\left\{\psi\left(e_{ij}\right)\psi\left(e_{i'j'}\right)\right\} = 0.$$

Next, because $\widetilde{\mathbf{B}}^\top(\boldsymbol{z}_j)\widetilde{\mathbf{B}}(\boldsymbol{z}_{j'}) = \mathbf{B}^\top(\boldsymbol{z}_j)\mathbf{Q}_2\mathbf{Q}_2^\top\mathbf{B}(\boldsymbol{z}_{j'})$ and the eigenvalues of $\mathbf{Q}_2\mathbf{Q}_2^\top$ are either 0 or 1,

$$\begin{aligned}&|\triangle|^{-2}\frac{1}{N^2}\sum_{j,j'=1}^{N}\mathbb{E}\left\{\widetilde{\mathbf{B}}^\top(\boldsymbol{z}_j)\widetilde{\mathbf{B}}(\boldsymbol{z}_{j'})\psi\left(e_{ij}\right)\psi\left(e_{i'j'}\right)\right\}\\ &\leq |\triangle|^{-2}\frac{1}{N^2}\sum_{j,j'=1}^{N}\mathbb{E}\left\{\mathbf{B}^\top(\boldsymbol{z}_j)\mathbf{B}(\boldsymbol{z}_{j'})\psi\left(e_{ij}\right)\psi\left(e_{i'j'}\right)\right\}\\ &= \sum_{m\in\mathcal{M}}|\triangle|^{-2}\frac{1}{N^2}\sum_{j,j'=1}^{N}\mathbb{E}\left\{B_m(\boldsymbol{z}_j)B_m(\boldsymbol{z}_{j'})\psi\left(e_{ij}\right)\psi\left(e_{i'j'}\right)\right\}.\end{aligned}$$

Assumption (A4) implies that, for each $m \in \mathcal{M}$

$$\frac{1}{N^2}\sum_{j\neq j'}^{N}\mathbb{E}\left\{B_m(\boldsymbol{z}_j)B_m(\boldsymbol{z}_{j'})\psi\left(e_{ij}\right)\psi\left(e_{i'j'}\right)\right\} = O\left(|\triangle|^2\right).$$

Therefore,

$$|\triangle|^{-2}\frac{1}{N^2}\sum_{j,j'=1}^{N}\mathbb{E}\left\{\widetilde{\mathbf{B}}^\top(\boldsymbol{z}_j)\widetilde{\mathbf{B}}(\boldsymbol{z}_{j'})\psi\left(e_{ij}\right)\psi\left(e_{i'j'}\right)\right\} = |\triangle|^{-2}O(|\triangle|^{-2})O\left(|\triangle|^2\right) = O(|\triangle|^{-2}),$$

where the term $O(|\triangle|^{-2})$ in the first equality comes from the number of spline basis functions $|\mathcal{M}|$. This implies that $\mathbb{E}\|\widetilde{\boldsymbol{\alpha}}\|_2^2 = O(|\triangle|^{-2})$ and $\|\widetilde{\boldsymbol{\alpha}}\|_2 = O_p(|\triangle|^{-1})$,

which implies that for a large constant $L > 0$, we can assume $|\triangle|\|\widetilde{\boldsymbol{\alpha}}\|_2 < L$. With the condition that $|\triangle|\|\boldsymbol{\alpha} - \widetilde{\boldsymbol{\alpha}}\|_2 = \epsilon$, we have $|\triangle|\|\boldsymbol{\alpha}\|_2 \leq L + \epsilon$. Then, going back to equation (S2.11), we have

$$
\begin{aligned}
&|\triangle|^2 \left\{\boldsymbol{\Gamma}_n(\boldsymbol{\alpha}) + \mathbf{P}_h(\boldsymbol{\alpha}) - \boldsymbol{\Gamma}_n(\widetilde{\boldsymbol{\alpha}}) - \mathbf{P}_h(\widetilde{\boldsymbol{\alpha}})\right\} \\
&\quad = |\triangle|^2 \left\{\boldsymbol{\Xi}_n(\boldsymbol{\alpha}) - \boldsymbol{\Xi}_n(\widetilde{\boldsymbol{\alpha}}) + (\boldsymbol{\alpha} - \widetilde{\boldsymbol{\alpha}})^\top \widetilde{\mathbf{L}}_{n,h}(\boldsymbol{\alpha} - \widetilde{\boldsymbol{\alpha}})\right\} \\
&\qquad + |\triangle|^2 \left\{\frac{h}{N}(\boldsymbol{\alpha} - \widetilde{\boldsymbol{\alpha}})^\top \mathbf{S}_{n,h}^{-1/2} \mathbf{Q}_2^\top \mathbf{P} \mathbf{Q}_2 \boldsymbol{\theta}^*\right\} + o(1) \\
&\quad \geq \frac{\widetilde{\lambda}_0 \epsilon^2}{2} - 2 \sup_{|\triangle|\|\boldsymbol{\alpha}\|_2 \leq L+\epsilon} \left||\triangle|^2 |\boldsymbol{\Xi}_n(\boldsymbol{\alpha})\right| + O\left(h|\triangle|^{-3} n^{-1/2} N^{-1}\right) + o(1),
\end{aligned}
$$

where $\widetilde{\lambda}_0$ is defined at the beginning of the proof, the inequality comes from Lemma S4, and the penalty term $O(h|\triangle|^{-3}n^{-1/2}N^{-1})$ is $o(1)$ by Assumption (A3), as shown in the proof of Theorem 1. This proves (S2.10), which then implies that $|\triangle|\|\widehat{\boldsymbol{\alpha}} - \widetilde{\boldsymbol{\alpha}}\|_2 = o_P(1)$ by a similar argument as in the proof for Theorem 1. ■

Proof of Theorem 2 By Lemma S3 and S7, we have

$$
\begin{aligned}
\left\|\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\right) - \mathbf{S}_{n,h}^{-1/2} \widetilde{\boldsymbol{\alpha}}\right\|_2 &= \left\|\mathbf{S}_{n,h}^{-1/2}\left\{\mathbf{S}_{n,h}^{1/2}\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\right) - \widetilde{\boldsymbol{\alpha}}\right\}\right\|_2 \\
&= \left\|\mathbf{S}_{n,h}^{-1/2}\left(\widehat{\boldsymbol{\alpha}} - \widetilde{\boldsymbol{\alpha}}\right)\right\|_2 = o_P\left(n^{-1/2}|\triangle|^{-2}\right).
\end{aligned}
$$

This implies that, for any vector $\boldsymbol{\beta} \in \mathbb{R}^{|\widetilde{\mathcal{M}}|}$ with $\|\boldsymbol{\beta}\|_2 \leq L$,

$$\begin{aligned}\boldsymbol{\beta}^\top \left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\right) &= \boldsymbol{\beta}^\top \mathbf{S}_{n,h}^{-1/2} \widetilde{\boldsymbol{\alpha}} + o_P\left(n^{-1/2}|\triangle|^{-2}\right) \\ &= \boldsymbol{\beta}^\top \mathbf{L}_{n,h}^{-1} \sum_{i=1}^{n} \frac{1}{N} \sum_{j=1}^{N} \widetilde{\mathbf{B}}(\boldsymbol{z}_j) \psi\left(e_{ij}\right) + o_P\left(n^{-1/2}|\triangle|^{-2}\right).\end{aligned} \tag{S2.12}$$

Letting $\boldsymbol{\nu} = \widetilde{\mathbf{B}} \mathbf{L}_{n,h}^{-1} \boldsymbol{\beta} = (\boldsymbol{\beta}^\top \mathbf{L}_{n,h}^{-1} \widetilde{\mathbf{B}}(\boldsymbol{z}_1), \cdots, \boldsymbol{\beta}^\top \mathbf{L}_{n,h}^{-1} \widetilde{\mathbf{B}}(\boldsymbol{z}_N))^\top$, we have

$$\boldsymbol{\beta}^\top \mathbf{L}_{n,h}^{-1} \sum_{i=1}^{n} \frac{1}{N} \sum_{j=1}^{N} \mathbf{V}(\boldsymbol{z}_j) \psi\left(e_{ij}\right) = \sum_{i=1}^{n} \frac{1}{N} \boldsymbol{\nu}^\top \boldsymbol{\psi}\left(\boldsymbol{e}_i\right). \tag{S2.13}$$

Note that

$$\text{Var}\left\{\sum_{i=1}^{n} \frac{1}{N} \boldsymbol{\nu}^\top \boldsymbol{\psi}\left(\boldsymbol{e}_i\right)\right\} = \sum_{i=1}^{n} \frac{1}{N^2} \boldsymbol{\nu}^\top \mathbf{K}_i \boldsymbol{\nu}.$$

Next, we rewrite (S2.13) as $\sum_i^n a_i \varsigma_i$, where $a_i^2 = N^{-2} \boldsymbol{\nu}^\top \mathbf{K}_i \boldsymbol{\nu}$, $1 \leq i \leq n$, and $\{\varsigma_i\}_{i=1}^n$ are independent with zero mean and unit variance. In the following, we are going to show the asymptotic normality of $\sum_{i=1}^n a_i \varsigma_i$ using Lindeberg's Central Limit Theorem. To this end, we need to check its condition first. We have that

$$\mathbb{E}\left(\frac{1}{N^2} \boldsymbol{\nu}^\top \mathbf{K}_i \boldsymbol{\nu}\right) \leq \frac{\|\boldsymbol{\nu}\|_2^2}{N^2} \sum_{j=1}^{N} \mathbb{E}\left\{\psi\left(e_{ij}\right)\right\}^2 \leq \frac{C}{N} \sum_{j=1}^{N} \left\{\boldsymbol{\beta}^\top \mathbf{L}_{n,h}^{-1} \widetilde{\mathbf{B}}(\boldsymbol{z}_j)\right\}^2 = O\left(\frac{1}{n^2|\triangle|^4}\right).$$

Thus, $a_i^2 \leq C n^{-2}|\triangle|^{-4}$ for all $1 \leq i \leq n$. Also, by Assumption (A8) we have

$$\sum_{i=1}^{n} a_i^2 = \sum_{i=1}^{n} \frac{1}{N^2} \boldsymbol{\nu}^\top \mathbf{K}_i \boldsymbol{\nu} \geq \frac{1}{N} \lambda_0 \boldsymbol{\beta}^\top \mathbf{L}_{n,h}^{-1} \widetilde{\mathbf{B}}^\top \widetilde{\mathbf{B}} \mathbf{L}_{n,h}^{-1} \boldsymbol{\beta} = O\left(n^{-1} N^{-1} |\triangle|^{-2}\right).$$

Therefore,

$$\max_{1\le i\le n}\frac{a_i^2}{\sum_{i=1}^n a_i^2}=O\left(n^{-1}N|\triangle|^{-2}\right)=o(N^{-1}),$$

where the last equality comes from Assumption (A3). This shows that the condition of Lindeberg's central limit theorem is satisfied. Setting $\boldsymbol{\beta}=\widetilde{\mathbf{B}}(\boldsymbol{z})$ for $\boldsymbol{z}\in\Omega$, we obtain

$$\sum_{i=1}^n\frac{1}{N^2}\boldsymbol{\nu}^\top\mathbf{K}_i\boldsymbol{\nu}=\widetilde{\mathbf{B}}^\top(\boldsymbol{z})\mathbf{L}_{n,h}^{-1}\left(\sum_{i=1}^n N^{-2}\widetilde{\mathbf{B}}^\top\mathbf{K}_i\widetilde{\mathbf{B}}\right)\mathbf{L}_{n,h}^{-1}\widetilde{\mathbf{B}}(\boldsymbol{z})=\Sigma_n(\boldsymbol{z}).$$

This helps us establish the asymptotic normality of $\widetilde{\mathbf{B}}(\boldsymbol{z})^\top(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}^*)$ for $\boldsymbol{z}\in\Omega$ based on (S2.12). Due to the condition $|\triangle|\ll n^{-1/\{2(d+2)\}}$, the bias term $\widetilde{\mathbf{B}}(\boldsymbol{z})^\top(\boldsymbol{\theta}^*-\boldsymbol{\theta})$ is asymptotically negligible, and hence the asymptotic normality of $\widetilde{\mathbf{B}}(\boldsymbol{z})^\top(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}^*)$ is established. This concludes the proof of Theorem 2. ■

Bibliography

Eggleston, H. G. (1958), *Convexity*, Cambridge Tracts in Mathematics, Cambridge: Cambridge University Press.

Lai, M.-J. and Schumaker, L. L. (2007), *Spline Functions on Triangulations*, Encyclopedia of Mathematics and Its Applications, Cambridge: Cambridge University Press.

Lima, I. R., Cao, G., and Billor, N. (2019), "M-based simultaneous inference for the mean function of functional data," *Annals of the Institute of Statistical Mathematics*, 71, 577–598.

Wang, S. and Cao, G. (2022), "Robust deep neural network estimation for multidimensional functional data," *Electronic Journal of Statistics*, 16, 6461–6488.

Wang, Y., Wang, G., Wang, L., and Ogden, R. T. (2020), "Simultaneous confidence corridors for mean functions in functional data analysis of imaging data," *Biometrics*, 76, 427–437.

Wei, Y. and He, X. (2006), "Conditional growth charts," *The Annals of Statistics*, 34, 2069–2097.