Supplementary Material for "Semi-parametric goodness-of-fit test for clustered point processes with a shape-constrained pair correlation function"

Ganggang Xu, Chen Liang, Rasmus Waagepetersen, and Yongtao Guan

Abstract

This supplementary material includes additional simulation studies, details on the numerical implementation of the test statistics, and technical conditions and proofs.

January 10, 2022

Keywords: Clustered point process, Goodness-of-fit, Pair correlation function, Shape constraints

---

Ganggang Xu (Email: gangxu@bus.miami.edu) is Assistant Professor and Yongtao Guan is Professor, Department of Management Science, University of Miami, Coral Gables, FL 33124. Chen Liang was a Ph.D. candidate, Department of Mathematical Sciences, Binghamton University, Binghamton, NY 13901, and is now a research scientist at Amazon.com, Inc. Rasmus Waagepetersen is Professor, Department of Mathematical Sciences, Aalborg University, Aalborg, Denmark. Xu's research is supported by NSF grant SES-1902195 and Yongtao Guan is supported by NSF grant DMS-1810591.

S.1 Additional Simulation Results

In this section, we provide additional simulation results, including simulation results of the log-Gaussian Cox processes, the impact of M-spline degrees, and a comparison between the infill asymptotics and the increasing domain asymptotics.

S.1.1 Log-Gaussian Cox Process

The log-Gaussian Cox processes are generated using latent isotropic Gaussian random fields equipped with the following covariance functions

1. Circular Covariance Model (LGCP-C), with $g_0(r) = \exp\left[\sigma^2 C(r/\phi)\right]$, where $C(r) = 1 - \frac{2}{\pi}\left[r \arcsin(r) + r\sqrt{1-r^2}\right] I(0 \leq r \leq 1)$ for some $\sigma^2 > 0, \phi > 0$.
2. Exponential Covariance Model (LGCP-E), with $g_0(r) = \exp\left[\sigma^2 C(r/\phi)\right]$, where $C(r) = \exp(-r)$ for some $\sigma^2 > 0, \phi > 0$.
3. Gaussian Covariance Model (LGCP-G), with $g_0(r) = \exp\left[\sigma^2 C(r/\phi)\right]$, where $C(r) = \exp(-r^2)$ for some $\sigma^2 > 0, \phi > 0$.

The parameters for the LGCPs are chosen so that the true PCFs of LGCP-G, LGCP-E, and LGCP-C are matched closely with those of the Thomas, VarGamma, and MatClust processes, respectively. See Figure S1(b)-(c) for a comparison.

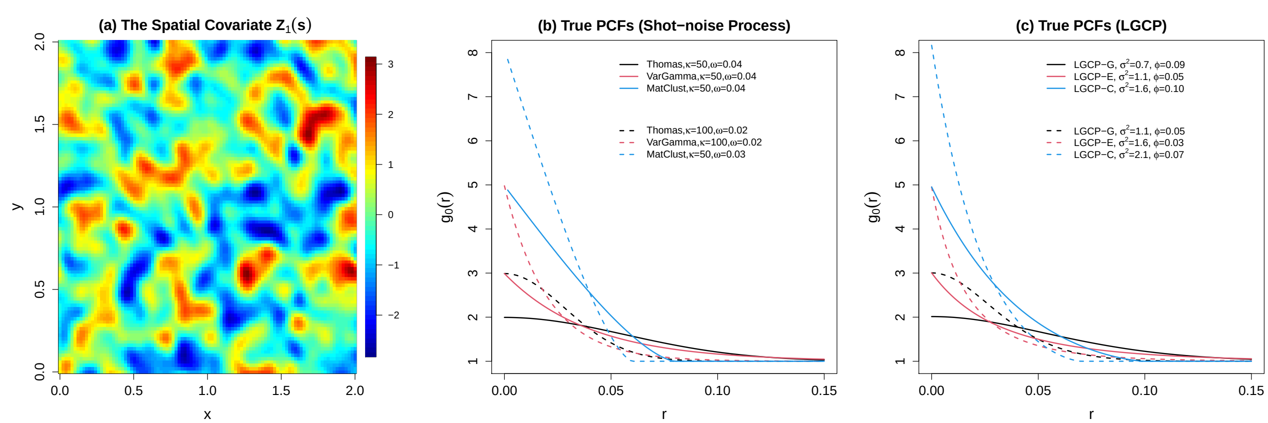

Figure S1: Panel (a): spatial covariates; Panels (b)-(c) the true PCFs of Cox processes.

S.1.1.1 Estimation Accuracies of PCFs

The same first-order intensity model is used for generation as in Section 6.1. Summary statistics based on $B = 3,000$ simulation runs are summarized in Figures S2. As illustrated in Figure S1 (b)-(c), the true PCFs between corresponding panels in Figure 3 and Figure S2, e.g., Figure3(a) v.s. Figure S2(a), are almost the same. We can observe that with the same level of second-order spatial correlation, estimated PCFs of the log-Gaussian Cox process have much greater estimation errors than those of the shot-noise Cox process. For log-Gaussian Cox processes, the proposed shape-constrained PCF estimators appear to have smaller advantages over the kernel PCF estimators, compared to the cases for shot-noise processes.

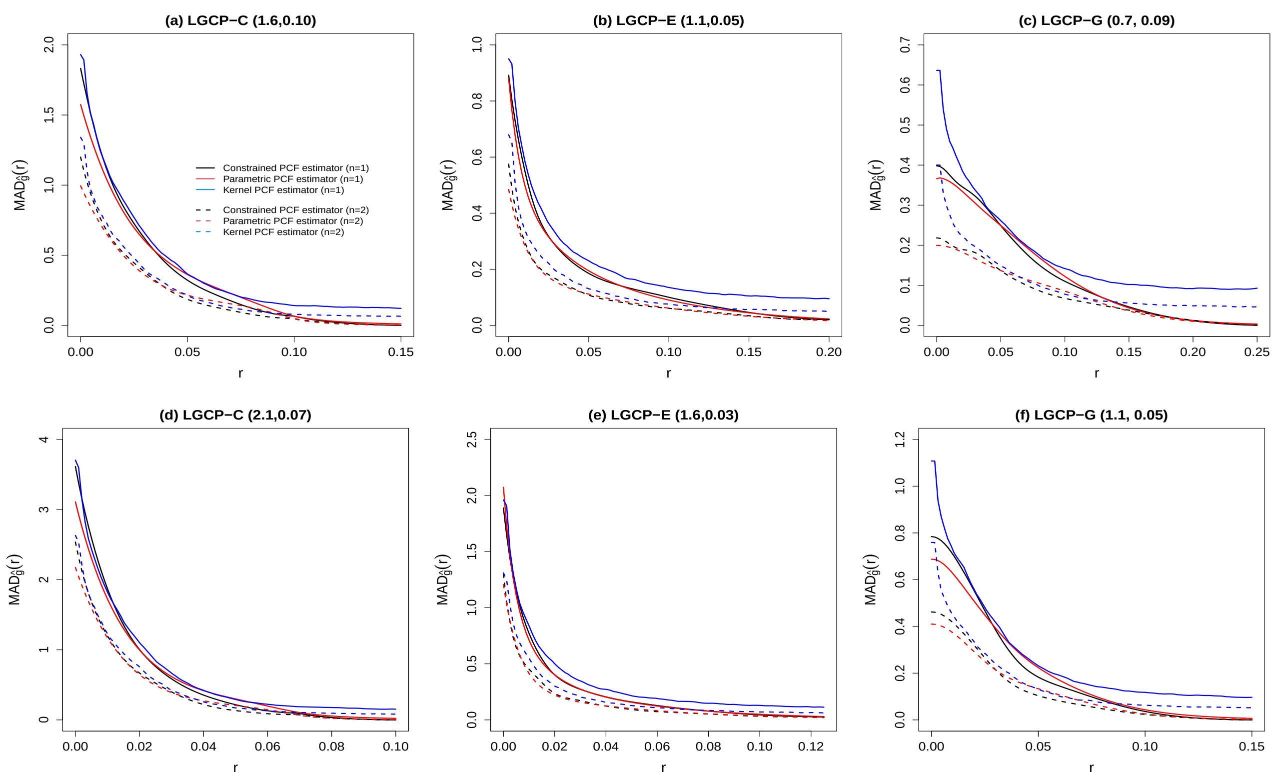

Figure S2: Estimation accuracy of three types of PCF estimators for log-Gaussian Cox processes.

S.1.1.2 Empirical Sizes of Goodness-of-fit Tests

The same first-order intensity model is used for generation as in Section 6.2. The rejection rates of H_0 at the $\alpha = 0.05$ and 0.10 levels based on $B = 3,000$ simulation runs are summarized in Table S1, where we can see that the empirical sizes for the log-Gaussian Cox processes are overall

Table S1: Empirical sizes of proposed goodness-of-fit tests at 0.05 and 0.10 levels

			Semi-G test		Semi-L test		Semi-C test	
Process	(σ^2, ϕ)	n	$\alpha=0.05$	$\alpha=0.10$	$\alpha=0.05$	$\alpha=0.10$	$\alpha=0.05$	$\alpha=0.10$
LGCP-G	(0.7,0.09)	1	0.063	0.118	0.067	0.122	0.074	0.134
		1.5	0.058	0.101	0.065	0.118	0.071	0.121
		2	0.052	0.106	0.055	0.101	0.050	0.096
	(1.1,0.05)	1	0.060	0.122	0.062	0.118	0.060	0.114
		1.5	0.066	0.119	0.064	0.113	0.065	0.119
		2	0.063	0.116	0.053	0.102	0.061	0.114
LGCP-E	(1.1,0.05)	1	0.067	0.116	0.066	0.116	0.063	0.120
		1.5	0.054	0.110	0.061	0.116	0.055	0.110
		2	0.060	0.114	0.064	0.121	0.065	0.122
	(1.6,0.03)	1	0.061	0.115	0.064	0.115	0.064	0.119
		1.5	0.048	0.101	0.056	0.109	0.056	0.111
		2	0.052	0.102	0.061	0.118	0.064	0.113
LGCP-C	(1.6,0.10)	1	0.059	0.116	0.058	0.117	0.064	0.116
		1.5	0.043	0.092	0.058	0.107	0.049	0.097
		2	0.059	0.108	0.062	0.116	0.060	0.110
	(2.1,0.07)	1	0.054	0.103	0.056	0.104	0.058	0.101
		1.5	0.047	0.094	0.056	0.112	0.061	0.105
		2	0.054	0.109	0.056	0.106	0.060	0.110

close to the nominal levels as well.

S.1.1.3 Empirical Powers of Goodness-of-fit Tests

The same settings are used as those in Section 6.3, except that data are generated from three types of log-Gaussian Cox processes, namely, LGCP-G $(0.7, 0.09)$, LGCP-E $(1.1, 0.05)$ and LGCP-C $(1.6, 0.10)$, which have similar spatial dependences as corresponding shot-noise Cox processes used in Section 6.3. The empirical rejection rates at the $\alpha = 0.05$ level based on $B = 1,000$ simulation runs are summarized in Figures S3. Similar to simulation results in Section 6.3, we can see that the Semi-G test has the greatest power against the alternative H_1^a while practically no power against H_1^b. On the contrary, the Semi-L test displays the opposite trend. This is consistent with our discussion in Section 4.4. The Semi-C test using the test function (28) is able to inherit strengths of both the Semi-G and the Semi-L test, and achieves much greater overall power than all Monte Carlo tests.

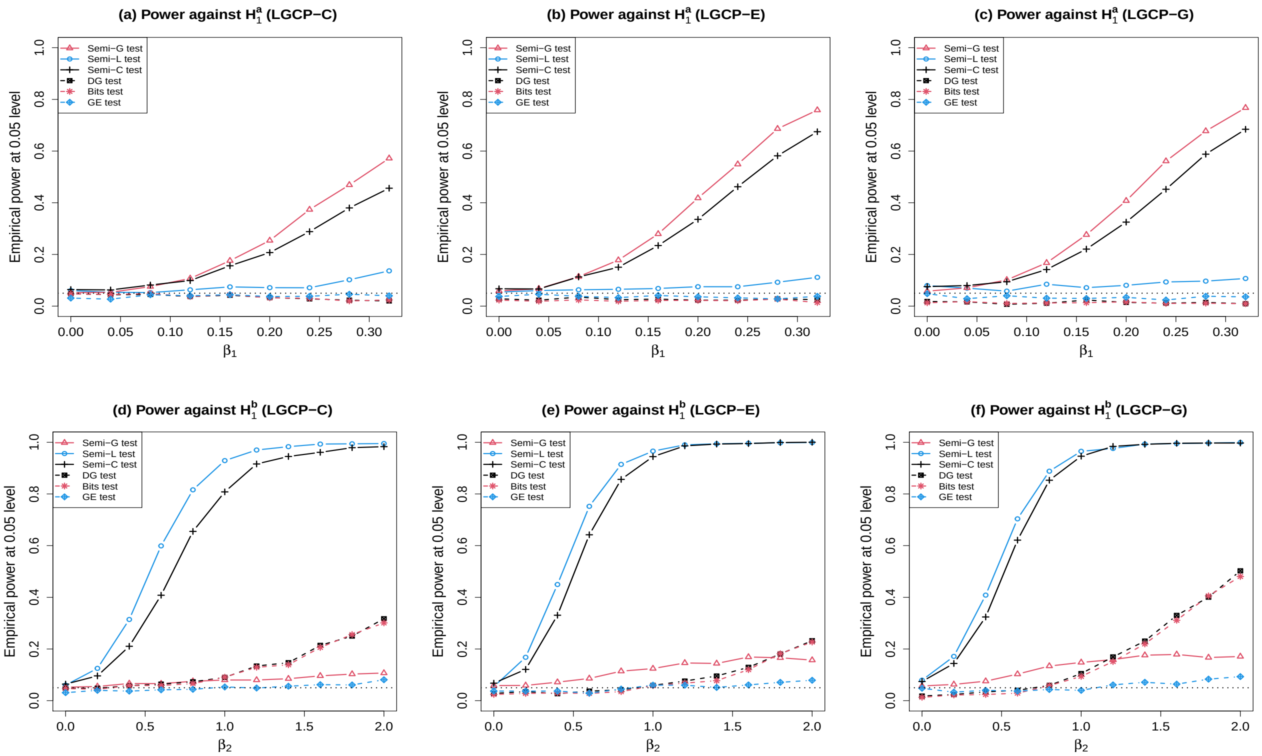

Figure S3: Empirical powers of goodness-of-fit tests for log-Gaussian Cox processes.

S.1.2 Estimation Accuracies of PCFs when Shape Constraints Are Violated

In this subsection, we compare performances of the shape-constrained PCF estimator and the kernel PCF estimator when the shape constraints S1 and S2 are violated. To do so, we simulate data from the log-Gaussian Cox process with the Bessel family covariance function (LGCP-B), in which case $g_0(r) = \exp\left[\sigma^2 C(r/\phi)\right]$ with

$$C(r) = 2^{\nu}\Gamma(\nu+1)r^{-\nu}J_{\nu}(r),$$

where $\nu \geq 0$, $\Gamma(\cdot)$ is the Gamma function and $J_{\nu}(\cdot)$ is a Bessel function of first kind. We fix $\sigma^2 = 1$ and $\phi = 0.01$, and consider multiple $\nu = 0.5, 0.75, 2$. From Figure S4, the true $g_0(\cdot)$'s are no longer monotone functions as required by the shape constraint S1. Therefore, the shape constraints on $g_0(\cdot)$ are violated. For simulation, the same first-order intensity model is used as in Section 6.1. Summary statistics based on $B = 1{,}000$ simulation runs are summarized in Figures S4(b)-(d).

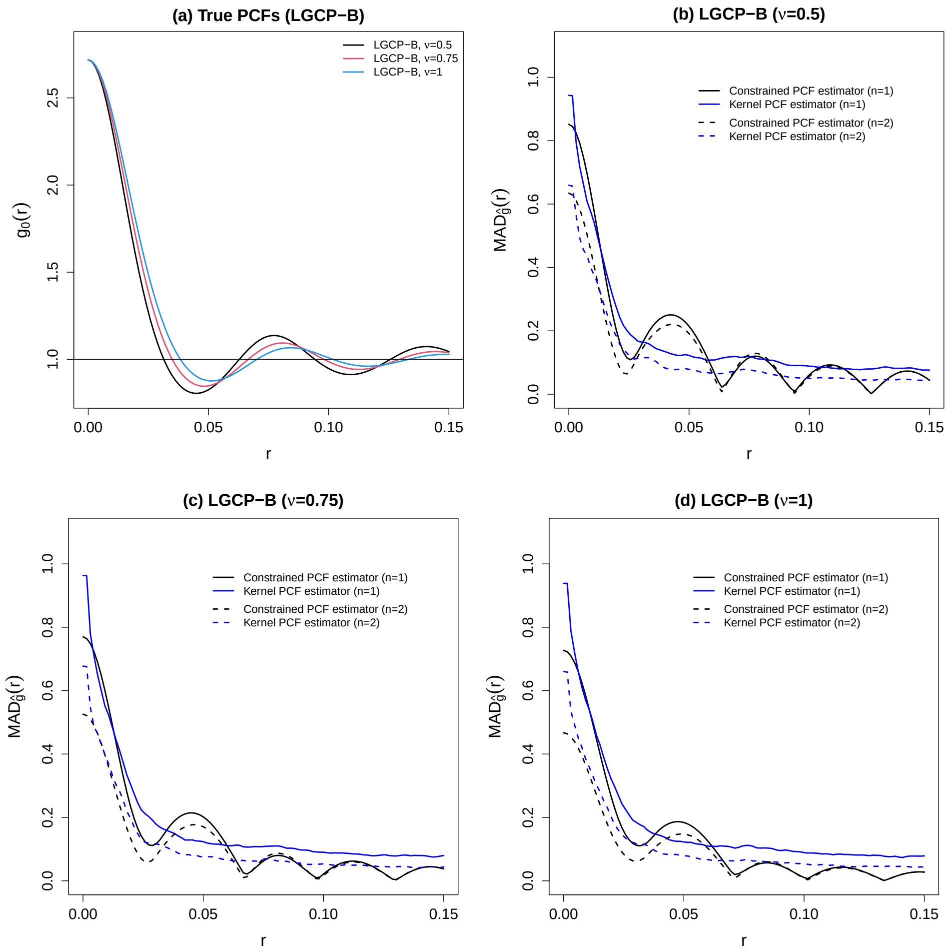

Figure S4: (a) True PCFs of LGCP-B with various ν's; (b)-(d) Estimation accuracies of shape-constrained PCF estimator v.s. Kernel PCF estimator;

From Figure S4(b), we can see that, as expected, the estimation accuracies of the shape-constrained PCF estimator does not change much when n increases from 1 to 2 for $r \geq 0.03$, indicating large estimation biases. The kernel PCF estimator outperforms the shape-constrained PCF estimator, especially with $n = 2$. This is not surprising because the kernel estimator is universal for isotropic PCFs with various shapes. As ν increases, the true PCFs become less oscillating and the shape constraint violations become less severe, and consequently, the advantages of the kernel PCF estimator over the shape constrained PCF estimator gradually fade away.

S.1.3 Impacts of M-spline Degrees

In this section, we investigate the impacts of the degrees of the M-spline on the shape-constrained PCF estimator. We simulate data in the same way as in Section 6.2 and estimate the PCF with degrees of the M-spline increasing from $m = 0$ to $m = 5$. The averaged $\mathrm{MAD}(\hat{g})=\int_0^R \mathrm{MAD}_{\hat{g}}(r)dr$ based on $1,000$ simulation runs are summarized in Figure S5. We can see that the choice of m has little impact on the estimation accuracy as long as it is not too small (e.g., $m = 0$). On one hand, when $m = 0$, piecewise constant functions are used to approximate the spectral density as in (6), which may result in significant approximation error, leading to larger $\mathrm{MAD}(\hat{g})$ values. On the other hand, although higher order spline basis functions typically lead to more variability of function estimators in the regression setting, the imposed shape constraints in our work serve as a strong regularizer for the smoothness of the estimated function and may reduce/offset impacts of higher degrees of the M-spline.

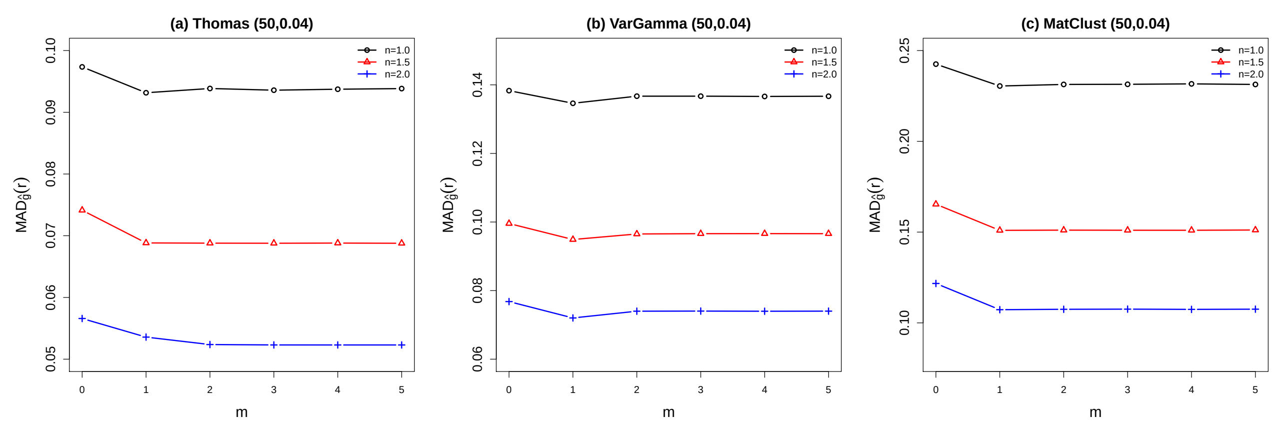

Figure S5: $\mathrm{MAD}(\hat{g})$ of the shape-constrained PCF estimator using M-splines with different degrees.

S.1.4 Infill Asymptotics v.s. Increasing Domain Asymptotics

Our theoretical results in Section 5 depend on the key assumption that the observation window W_n is increasing as n grows, which is commonly referred to as the increasing domain asymptotic framework. In the literature, there exists another type of asymptotic framework where the observation window is fixed while the number of points increases. In this section, we conduct a simulation to illustrate the differences between these two asymptotic frameworks when applied to the proposed shaped constrained PCF estimator. The data are simulated in the same way as in Section 6.2. For a given window $W_n = [0, n]^2$, we choose the intercept β_0 under H_0 appropriately so that the average number of points per unit square is around $\rho_{unit} = 200, 400, \cdots, 1600$ in increasing order. The averaged MAD($\hat{g}$)=$\int_0^R \text{MAD}_{\hat{g}}(r)dr$ based on $1,000$ simulation runs are summarized in Figure S6.

Figure S6(a)-(c) reflects the increasing domain asymptotics considered in this paper, for which ρ_{unit} is fixed when W_n expands. There is a clear linear relationship between MAD($\hat{g}$) and $\log(n)$, indicating that the convergence rate of the shape constrained PCF estimator is of the order $O(|W_n|^{-\delta})$ for some $\delta > 0$, which is as expected. On the contrary, under the infill asymptotic framework when W_n is fixed, the estimation error first decreases at a fast rate when ρ_{unit} is small and then slower as ρ_{unit} gets larger. The nonlinear relationship between MAD($\hat{g}$) and $\log(\rho_{unit})$ suggests that the limiting behavior of the shape constrained PCF estimator under the infill asymptotic framework is rather different from that under the increasing domain asymptotic framework considered in this paper, and will be an interesting future research topic.

S.2 Numerical Implementation of Test Statistic

Recall the definition of the test statistic

$$T_{\boldsymbol{f}_{\hat{g}}} = \left[\mathbf{Q}_{\boldsymbol{f}_{\hat{g}}}(\widehat{\boldsymbol{\beta}}_{\boldsymbol{\phi}_{\hat{g}}})\right]^T \left[\boldsymbol{\Sigma}_{\boldsymbol{f}}(\widehat{\boldsymbol{\beta}}_{\boldsymbol{\phi}_{\hat{g}}}, \hat{g})\right]^- \left[\mathbf{Q}_{\boldsymbol{f}_{\hat{g}}}(\widehat{\boldsymbol{\beta}}_{\boldsymbol{\phi}_{\hat{g}}})\right],$$

where

$$\mathbf{Q}_{\boldsymbol{f}_{\hat{g}}}(\widehat{\boldsymbol{\beta}}_{\boldsymbol{\phi}_{\hat{g}}}) = \frac{1}{\sqrt{|W|}}\left[\sum_{\mathbf{s}\in N\cap W} \boldsymbol{f}_{\hat{g}}(\mathbf{s}; \widehat{\boldsymbol{\beta}}_{\boldsymbol{\phi}_{\hat{g}}}) - \int_W \psi(\mathbf{s}; \widehat{\boldsymbol{\beta}}_{\boldsymbol{\phi}_{\hat{g}}})\boldsymbol{f}_{\hat{g}}(\mathbf{s}; \widehat{\boldsymbol{\beta}}_{\boldsymbol{\phi}_{\hat{g}}})\mathrm{d}\mathbf{s}\right], \tag{S.1}$$

and based Theorem 2, it can be shown that

$$\begin{aligned}\boldsymbol{\Sigma}_{\boldsymbol{f}}(\widehat{\boldsymbol{\beta}}_{\boldsymbol{\phi}_{\hat{g}}}, \hat{g}) = |W|^{-1} &\iint_{W^2} \widetilde{\boldsymbol{f}}_{\hat{g}}(\mathbf{s}; \widehat{\boldsymbol{\beta}}_{\boldsymbol{\phi}_{\hat{g}}})\widetilde{\boldsymbol{f}}_{\hat{g}}^T(\mathbf{t}; \widehat{\boldsymbol{\beta}}_{\boldsymbol{\phi}_{\hat{g}}})\left[\hat{g}(\|\mathbf{s}-\mathbf{t}\|) - 1\right]\psi(\mathbf{s}; \widehat{\boldsymbol{\beta}}_{\boldsymbol{\phi}_{\hat{g}}})\psi(\mathbf{t}; \widehat{\boldsymbol{\beta}}_{\boldsymbol{\phi}_{\hat{g}}})\mathrm{d}\mathbf{s}\mathrm{d}\mathbf{t} \\ &+ |W|^{-1}\int_W \widetilde{\boldsymbol{f}}_{\widetilde{g}}(\mathbf{s}; \widehat{\boldsymbol{\beta}}_{\boldsymbol{\phi}_{\hat{g}}})\widetilde{\boldsymbol{f}}_{\widetilde{g}}^T(\mathbf{s}; \widehat{\boldsymbol{\beta}}_{\boldsymbol{\phi}_{\hat{g}}})\psi(\mathbf{s}; \widehat{\boldsymbol{\beta}}_{\boldsymbol{\phi}_{\hat{g}}})\mathrm{d}\mathbf{s},\end{aligned} \tag{S.2}$$

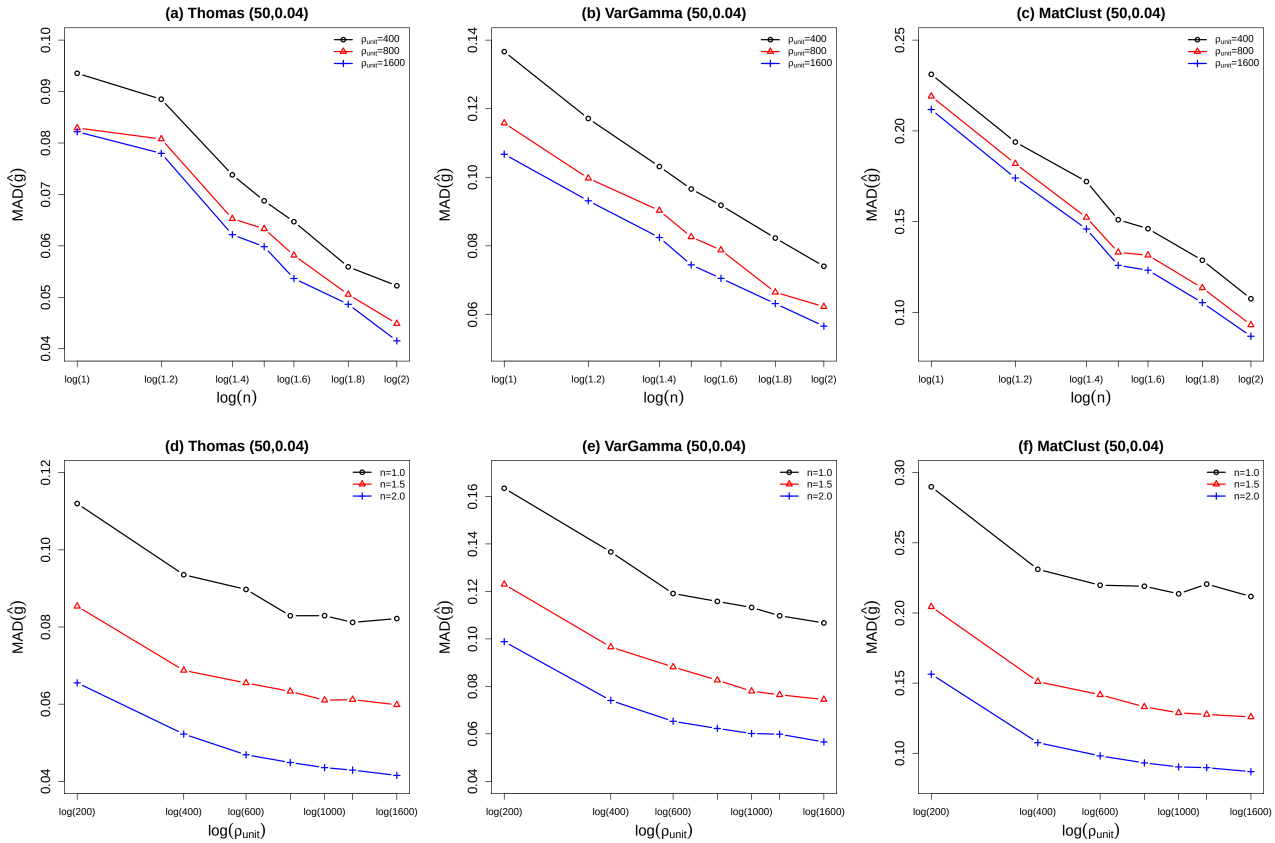

Figure S6: Increasing domain asymptotics (top panels) v.s. Infill asymptotics (bottom panels).

with $\widetilde{\boldsymbol{f}}_{\hat{g}}(\mathbf{s};\widehat{\boldsymbol{\beta}}_{\boldsymbol{\phi}_{\hat{g}}}) = \boldsymbol{f}_{\hat{g}}(\mathbf{s};\widehat{\boldsymbol{\beta}}_{\boldsymbol{\phi}_{\hat{g}}}) - [\mathbf{S}_{\boldsymbol{f}}(\widehat{\boldsymbol{\beta}}_{\boldsymbol{\phi}_{\hat{g}}},\hat{g})]^T[\mathbf{S}_{\boldsymbol{\phi}}(\widehat{\boldsymbol{\beta}}_{\boldsymbol{\phi}_{\hat{g}}},\hat{g})]^{-1}\boldsymbol{\phi}_{\hat{g}}(\mathbf{s};\widehat{\boldsymbol{\beta}}_{\boldsymbol{\phi}_{\hat{g}}})$, where then sensitivity matrix is defined as $\mathbf{S}_{\boldsymbol{f}}(\boldsymbol{\beta}, g) = |W|^{-1}\int_W \psi(\mathbf{s};\boldsymbol{\beta})\boldsymbol{\eta}(\mathbf{s};\boldsymbol{\beta})\boldsymbol{f}_g^T(\mathbf{s};\boldsymbol{\beta})\mathrm{d}\mathbf{s}$.

Remark 1. *The covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{f}}(\widehat{\boldsymbol{\beta}}_{\boldsymbol{\phi}_{\hat{g}}},\hat{g})$ given in* (S.2) *appears to be different from the one given in* (15)*. However, it is straightforward to show that if $\boldsymbol{\phi}_g(\cdot;\boldsymbol{\beta})$ is the optimal estimating function obtained by solving the Fredholm integral equation* (3)*, equations* (S.2) *and* (15) *give the same estimated covariance matrices. However, covariance matrix* (S.2) *is more general in the sense that it holds even if $\boldsymbol{\phi}_g(\cdot;\boldsymbol{\beta})$ is not the optimal estimating function.*

Suppose that the spatial domain W can be partitioned into m small sub-domains centered at quadrature points $\mathbf{t}_1,\cdots,\mathbf{t}_m$ with associated areas as $w_1,\cdots,w_m$. Then the integrals in (S.1) and (S.2) can be approximated by

$$\mathbf{Q}_{\boldsymbol{f}_{\hat{g}}}(\widehat{\boldsymbol{\beta}}_{\boldsymbol{\phi}_{\hat{g}}}) = \frac{1}{\sqrt{|W|}}\left[\sum_{\mathbf{s}\in N\cap W}\boldsymbol{f}_{\hat{g}}(\mathbf{s};\widehat{\boldsymbol{\beta}}_{\boldsymbol{\phi}_{\hat{g}}}) - \sum_{i=1}^{m}\psi(\mathbf{t}_i;\widehat{\boldsymbol{\beta}}_{\boldsymbol{\phi}_{\hat{g}}})\boldsymbol{f}_{\hat{g}}(\mathbf{t}_i;\widehat{\boldsymbol{\beta}}_{\boldsymbol{\phi}_{\hat{g}}})w_i\right], \tag{S.3}$$

$$\begin{aligned}\boldsymbol{\Sigma}_{\boldsymbol{f}}(\widehat{\boldsymbol{\beta}}_{\boldsymbol{\phi}_{\hat{g}}},\hat{g}) &= |W|^{-1}\sum_{i=1}^{m}\sum_{j=1}^{m}\widetilde{\boldsymbol{f}}_{\hat{g}}(\mathbf{t}_i;\widehat{\boldsymbol{\beta}}_{\boldsymbol{\phi}_{\hat{g}}})\widetilde{\boldsymbol{f}}_{\hat{g}}^{T}(\mathbf{t}_j;\widehat{\boldsymbol{\beta}}_{\boldsymbol{\phi}_{\hat{g}}})\left[\hat{g}(\|\mathbf{t}_i-\mathbf{t}_j\|)-1\right]\psi(\mathbf{t}_i;\widehat{\boldsymbol{\beta}}_{\boldsymbol{\phi}_{\hat{g}}})\psi(\mathbf{t}_j;\widehat{\boldsymbol{\beta}}_{\boldsymbol{\phi}_{\hat{g}}})w_iw_j\\ &\quad + |W|^{-1}\sum_{i=1}^{m}\widetilde{\boldsymbol{f}}_{\widetilde{g}}(\mathbf{t}_i;\widehat{\boldsymbol{\beta}}_{\boldsymbol{\phi}_{\hat{g}}})\widetilde{\boldsymbol{f}}_{\widetilde{g}}^{T}(\mathbf{t}_i;\widehat{\boldsymbol{\beta}}_{\boldsymbol{\phi}_{\hat{g}}})\psi(\mathbf{t}_i;\widehat{\boldsymbol{\beta}}_{\boldsymbol{\phi}_{\hat{g}}})w_i,\end{aligned} \tag{S.4}$$

where $\mathbf{S}_{\boldsymbol{f}}(\widehat{\boldsymbol{\beta}}_{\boldsymbol{\phi}_{\hat{g}}},\hat{g})$ and $\mathbf{S}_{\boldsymbol{\phi}}(\widehat{\boldsymbol{\beta}}_{\boldsymbol{\phi}_{\hat{g}}},\hat{g})$ are similarly obtained by quadrature approximation. Now to approximate $\mathbf{Q}_{\boldsymbol{f}_{\hat{g}}}(\widehat{\boldsymbol{\beta}}_{\boldsymbol{\phi}_{\hat{g}}})$ and $\boldsymbol{\Sigma}_{\boldsymbol{f}}(\widehat{\boldsymbol{\beta}}_{\boldsymbol{\phi}_{\hat{g}}},\hat{g})$, it remains to find the test function values at quadrature points, i.e., $\boldsymbol{f}_{\hat{g}}(\mathbf{t}_i;\widehat{\boldsymbol{\beta}}_{\boldsymbol{\phi}_{\hat{g}}})$, $i=1,\cdots,m$ and for any $\mathbf{s}\in W$, $\boldsymbol{f}_{\hat{g}}(\mathbf{s};\widehat{\boldsymbol{\beta}}_{\boldsymbol{\phi}_{\hat{g}}})$ takes the same value as the nearest quadrature point to $\mathbf{s}$.

S.2.1 Global Semi-parametric (Semi-G) Test

The definition of Semi-G test involves the following test function

$$\boldsymbol{f}_g^{\mathrm{G}}(\mathbf{s};\boldsymbol{\beta}) = \mathrm{Vec}^{\mathrm{sub}}\left[\boldsymbol{\eta}(\mathbf{s};\boldsymbol{\beta})\boldsymbol{\phi}_g^{T}(\mathbf{s};\boldsymbol{\beta})\right].$$

Then it immediately follows that

$$\boldsymbol{f}_{\hat{g}}(\mathbf{t}_i;\widehat{\boldsymbol{\beta}}_{\boldsymbol{\phi}_{\hat{g}}}) = \mathrm{Vec}^{\mathrm{sub}}\left[\boldsymbol{\eta}(\mathbf{t}_i;\widehat{\boldsymbol{\beta}}_{\boldsymbol{\phi}_{\hat{g}}})\boldsymbol{\phi}_{\hat{g}}^{T}(\mathbf{t}_i;\widehat{\boldsymbol{\beta}}_{\boldsymbol{\phi}_{\hat{g}}})\right], i=1,\cdots m,$$

which can then be plugged back in (S.3) and (S.4).

S.2.2 Local Semi-parametric (Semi-L) Test

The definition of Semi-L test involves the following test function

$$\boldsymbol{f}_g^{\mathrm{L}}(\mathbf{s};\boldsymbol{\beta}) = \int_W \mathrm{diag}\left\{\left[\mathbf{S}_{\boldsymbol{\phi}_{\mathbf{x}}}(\boldsymbol{\beta},g)\right]_{11}^{-1/2},\cdots,\left[\mathbf{S}_{\boldsymbol{\phi}_{\mathbf{x}}}(\boldsymbol{\beta},g)\right]_{pp}^{-1/2}\right\}\boldsymbol{\phi}_{\mathbf{x},g}(\mathbf{s};\boldsymbol{\beta})I(\mathbf{s}\in B_{\mathbf{x}}(d))\mathrm{d}\mathbf{x},$$

where $\mathbf{S}_{\boldsymbol{\phi}_{\mathbf{x}}}(\boldsymbol{\beta},g) = |W|^{-1}\int_{W\cap B_{\mathbf{x}}(d)}\psi(\mathbf{s};\boldsymbol{\beta})\boldsymbol{\eta}(\mathbf{s};\boldsymbol{\beta})\boldsymbol{\phi}_{\mathbf{x},g}^{T}(\mathbf{s};\boldsymbol{\beta})\mathrm{d}\mathbf{s}$ and $\boldsymbol{\phi}_{\mathbf{x},g}(\mathbf{s};\boldsymbol{\beta})$ is the solution to the localized version of equation (3) as follows

$$\boldsymbol{\phi}_{\mathbf{x},g}(\mathbf{s};\boldsymbol{\beta}) + \int_{W\cap B_{\mathbf{x}}(d)}\boldsymbol{\phi}_{\mathbf{x},g}(\mathbf{u};\boldsymbol{\beta})\psi(\mathbf{u};\boldsymbol{\beta})\left[g(\mathbf{s},\mathbf{t})-1\right]\mathrm{d}\mathbf{u} = \boldsymbol{\eta}(\mathbf{s};\boldsymbol{\beta}),$$

for $\mathbf{s} \in W \cap B_{\mathbf{x}}(d)$.

Using the quadrature approximation for the integral, one has that, for any $\mathbf{t}_i$,

$$\boldsymbol{f}_g^{\mathrm{L}}(\mathbf{t}_i; \widehat{\boldsymbol{\beta}}_{\boldsymbol{\phi}_{\hat{g}}}) = \sum_{j=1}^{m} \operatorname{diag}\left\{ \left[\mathbf{S}_{\boldsymbol{\phi}_{\mathbf{t}_j}}(\widehat{\boldsymbol{\beta}}_{\boldsymbol{\phi}_{\hat{g}}}, \hat{g})\right]_{11}^{-1/2}, \cdots, \left[\mathbf{S}_{\boldsymbol{\phi}_{\mathbf{t}_j}}(\widehat{\boldsymbol{\beta}}_{\boldsymbol{\phi}_{\hat{g}}}, \hat{g})\right]_{pp}^{-1/2} \right\} \boldsymbol{\phi}_{\mathbf{t}_j,\hat{g}}(\mathbf{t}_i; \widehat{\boldsymbol{\beta}}_{\boldsymbol{\phi}_{\hat{g}}}) I(\mathbf{t}_i \in B_{\mathbf{t}_j}(d)) w_j. \tag{S.5}$$

To find $\boldsymbol{\phi}_{\mathbf{t}_j,\hat{g}}(\mathbf{t}_i; \widehat{\boldsymbol{\beta}}_{\boldsymbol{\phi}_{\hat{g}}})$ for all $\mathbf{t}_i \in B_{\mathbf{t}_j}(d)$, we approximate the spatial domain $B_{\mathbf{t}_j}(d)$ by a subset of the quadrature points, say $\{\mathbf{t}_{j_1}, \cdots, \mathbf{t}_{j_{n_j}}\} \subset \{\mathbf{t}_1, \mathbf{t}_2, \cdots, \mathbf{t}_m\}$. Then the integral equation can be approximated by

$$\boldsymbol{\phi}_{\mathbf{t}_j,\hat{g}}(\mathbf{t}_{j_k}; \widehat{\boldsymbol{\beta}}_{\boldsymbol{\phi}_{\hat{g}}}) + \sum_{l=1}^{n_j} \boldsymbol{\phi}_{\mathbf{t}_j,\hat{g}}(\mathbf{t}_{j_l}; \widehat{\boldsymbol{\beta}}_{\boldsymbol{\phi}_{\hat{g}}}) \psi(\mathbf{t}_{j_l}; \widehat{\boldsymbol{\beta}}_{\boldsymbol{\phi}_{\hat{g}}}) \left[\hat{g}(\mathbf{t}_{j_k}, \mathbf{t}_{j_l}) - 1\right] w_{j_l} = \boldsymbol{\eta}(\mathbf{t}_{j_k}; \widehat{\boldsymbol{\beta}}_{\boldsymbol{\phi}_{\hat{g}}}), k = 1, \cdots, n_j,$$

which has a closed form solution

$$\begin{bmatrix} \boldsymbol{\phi}^T_{\mathbf{t}_j,\hat{g}}(\mathbf{t}_{j_1}; \widehat{\boldsymbol{\beta}}_{\boldsymbol{\phi}_{\hat{g}}}) \\ \boldsymbol{\phi}^T_{\mathbf{t}_j,\hat{g}}(\mathbf{t}_{j_2}; \widehat{\boldsymbol{\beta}}_{\boldsymbol{\phi}_{\hat{g}}}) \\ \vdots \\ \boldsymbol{\phi}^T_{\mathbf{t}_j,\hat{g}}(\mathbf{t}_{j_{n_j}}; \widehat{\boldsymbol{\beta}}_{\boldsymbol{\phi}_{\hat{g}}}) \end{bmatrix} = \begin{bmatrix} 1 + a_{11} & a_{12} & \cdots & a_{1n_j} \\ a_{21} & 1 + a_{22} & \cdots & a_{2n_j} \\ \vdots & \vdots & \cdots & \vdots \\ a_{n_j1} & 1 + a_{n_j2} & \cdots & 1 + a_{n_jn_j} \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{\eta}^T(\mathbf{t}_{j_1}; \widehat{\boldsymbol{\beta}}_{\boldsymbol{\phi}_{\hat{g}}}) \\ \boldsymbol{\eta}^T(\mathbf{t}_{j_2}; \widehat{\boldsymbol{\beta}}_{\boldsymbol{\phi}_{\hat{g}}}) \\ \vdots^T \\ \boldsymbol{\eta}^T(\mathbf{t}_{j_{n_j}}; \widehat{\boldsymbol{\beta}}_{\boldsymbol{\phi}_{\hat{g}}}) \end{bmatrix}, \tag{S.6}$$

where $a_{kl} = \psi(\mathbf{t}_{j_l}; \widehat{\boldsymbol{\beta}}_{\boldsymbol{\phi}_{\hat{g}}}) \left[\hat{g}(\mathbf{t}_{j_k}, \mathbf{t}_{j_l}) - 1\right] w_{j_l}$ for $k, l = 1, \cdots, n_j$ and $j = 1, \cdots, m$. Using (S.6), $\mathbf{S}_{\boldsymbol{\phi}_{\mathbf{t}_j}}(\widehat{\boldsymbol{\beta}}_{\boldsymbol{\phi}_{\hat{g}}}, \hat{g})$ can be approximated by

$$\mathbf{S}_{\boldsymbol{\phi}_{\mathbf{t}_j}}(\widehat{\boldsymbol{\beta}}_{\boldsymbol{\phi}_{\hat{g}}}, \hat{g}) = |W|^{-1} \sum_{l=1}^{n_j} \psi(\mathbf{t}_{j_l}; \widehat{\boldsymbol{\beta}}_{\boldsymbol{\phi}_{\hat{g}}}) \boldsymbol{\eta}(\mathbf{t}_{j_l}; \widehat{\boldsymbol{\beta}}_{\boldsymbol{\phi}_{\hat{g}}}) \boldsymbol{\phi}^T_{\mathbf{t}_j,\hat{g}}(\mathbf{t}_{j_l}; \widehat{\boldsymbol{\beta}}_{\boldsymbol{\phi}_{\hat{g}}}) w_{j_l}, \quad j = 1, \cdots m. \tag{S.7}$$

By plugging in (S.6) and (S.7) back into (S.5), we can obtain all $\boldsymbol{f}_g^{\mathrm{L}}(\mathbf{t}_1; \widehat{\boldsymbol{\beta}}_{\boldsymbol{\phi}_{\hat{g}}}), \cdots, \boldsymbol{f}_g^{\mathrm{L}}(\mathbf{t}_m; \widehat{\boldsymbol{\beta}}_{\boldsymbol{\phi}_{\hat{g}}})$.

S.3 Proof of Proposition 1

It can be readily seen that $\beta_0^\dagger = \beta_0 + \sum_{j=1}^p a_j \beta_j$ and $\beta_{1,j}^\dagger = b_j \beta_{1,j}$, $j = 1, \cdots, p$. Then the true first-order intensity can be written as $\psi(\mathbf{s}; \boldsymbol{\beta}) = \exp\left(\beta_0 + \boldsymbol{\beta}_1^T \mathbf{Z}(\mathbf{s})\right) = \exp\left(\beta_0^\dagger + \boldsymbol{\beta}_1^{\dagger T} \widetilde{\mathbf{Z}}(\mathbf{s})\right) = \psi\left(\mathbf{s}; \boldsymbol{\beta}^\dagger\right)$, where $\boldsymbol{\beta}_1 = (\beta_1, \cdots, \beta_p)^T$, $\boldsymbol{\beta}_1^\dagger = (\beta_1^\dagger, \cdots, \beta_p^\dagger)^T$, $\boldsymbol{\beta} = (\beta_0, \boldsymbol{\beta}_1^T)^T$ and $\boldsymbol{\beta}^\dagger = (\beta_0^\dagger, \boldsymbol{\beta}_1^{\dagger T})^T$. Some straightforward calculus yields that

$$\boldsymbol{\eta}(\mathbf{s}; \boldsymbol{\beta}) = \frac{1}{\psi(\mathbf{s}; \boldsymbol{\beta})} \frac{\partial \psi(\mathbf{s}; \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \frac{1}{\psi\left(\mathbf{s}; \boldsymbol{\beta}^\dagger\right)} \left(\frac{\partial \boldsymbol{\beta}^\dagger}{\partial \boldsymbol{\beta}^T}\right)^T \frac{\partial \psi\left(\mathbf{s}; \boldsymbol{\beta}^\dagger\right)}{\partial \boldsymbol{\beta}^\dagger} = \mathbf{T} \boldsymbol{\eta}^\dagger(\mathbf{s}; \boldsymbol{\beta}^\dagger),$$

where the matrix

$$\mathbf{T} = \left(\frac{\partial \boldsymbol{\beta}^{\dagger}}{\partial \boldsymbol{\beta}^{T}}\right) = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ a_1 & b_1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ a_p & 0 & \cdots & b_p \end{pmatrix}. \tag{S.8}$$

Under the model $\psi(\mathbf{s};\boldsymbol{\beta})$, the optimal weight function $\boldsymbol{\phi}_g(\mathbf{s};\boldsymbol{\beta})$ solves the integral equation

$$\boldsymbol{\phi}_g(\mathbf{s};\boldsymbol{\beta}) + \int_W \boldsymbol{\phi}_g(\mathbf{u};\boldsymbol{\beta})\psi(\mathbf{u};\boldsymbol{\beta})\left[g(\mathbf{s},\mathbf{u})-1\right]\mathrm{d}\mathbf{u} = \boldsymbol{\eta}(\mathbf{s};\boldsymbol{\beta}).$$

Multiple both sides by the matrix $\mathbf{T}^{-1}$ and denote $\boldsymbol{\phi}_g^{\dagger}(\mathbf{s};\boldsymbol{\beta}^{\dagger}) = \mathbf{T}^{-1}\boldsymbol{\phi}_g(\mathbf{s};\boldsymbol{\beta})$, then one has that

$$\boldsymbol{\phi}_g^{\dagger}(\mathbf{s};\boldsymbol{\beta}^{\dagger}) + \int_W \boldsymbol{\phi}_g^{\dagger}(\mathbf{u};\boldsymbol{\beta}^{\dagger})\psi(\mathbf{u};\boldsymbol{\beta}^{\dagger})\left[g(\mathbf{s},\mathbf{u})-1\right]\mathrm{d}\mathbf{u} = \boldsymbol{\eta}^{\dagger}(\mathbf{s};\boldsymbol{\beta}),$$

suggesting that $\boldsymbol{\phi}_g^{\dagger}(\mathbf{s};\boldsymbol{\beta}^{\dagger}) = \mathbf{T}^{-1}\boldsymbol{\phi}_g(\mathbf{s};\boldsymbol{\beta})$ is the optimal weight function under the re-parameterized model $\psi(\mathbf{u};\boldsymbol{\beta}^{\dagger})$.

When $\psi(\mathbf{s};\boldsymbol{\beta}) = \exp\left(\beta_0 + \boldsymbol{\beta}_1^T\mathbf{Z}(\mathbf{s})\right)$, one has that $\boldsymbol{\eta}(\mathbf{s};\boldsymbol{\beta}) = (1,\mathbf{Z}^T(\mathbf{s}))^T$ and similarly $\boldsymbol{\eta}^{\dagger}(\mathbf{s};\boldsymbol{\beta}) = (1,\widetilde{\mathbf{Z}}^T(\mathbf{s}))^T$. Then one has that $\eta_0(\mathbf{s};\boldsymbol{\beta})\phi_{0,g}(\mathbf{s};\boldsymbol{\beta}) = \phi_{0,g}(\mathbf{s};\boldsymbol{\beta})$, which means that the $(1,1)$ entry in the matrix is of the form

$$\left[\widehat{\mathbf{S}}_{\boldsymbol{\phi}}(\boldsymbol{\beta}_0,g_0) - \widehat{\boldsymbol{\Sigma}}^*_{\boldsymbol{\phi}}(\boldsymbol{\beta}_0,g_0)\right]_{(1,1)} = \frac{1}{|W|}\left[\sum_{\mathbf{s}\in N\cap W}\phi_{0,\hat{g}}(\mathbf{s};\widehat{\boldsymbol{\beta}}_{\boldsymbol{\phi}_{\hat{g}}}) - \int_W \psi(\mathbf{s};\widehat{\boldsymbol{\beta}}_{\boldsymbol{\phi}_{\hat{g}}})\phi_{0,\hat{g}}(\mathbf{s};\widehat{\boldsymbol{\beta}}_{\boldsymbol{\phi}_{\hat{g}}})\mathrm{d}\mathbf{s}\right] = 0, \tag{S.9}$$

where the last equality follows from the definition of $\widehat{\boldsymbol{\beta}}_{\boldsymbol{\phi}_{\hat{g}}}$ and $\phi^T_{0,\hat{g}}(\mathbf{s};\widehat{\boldsymbol{\beta}}_{\boldsymbol{\phi}_{\hat{g}}})$.

Similarly, one has that under the re-parameterized model $\psi(\mathbf{u};\boldsymbol{\beta}^{\dagger})$

$$\left[\widehat{\mathbf{S}}_{\boldsymbol{\phi}^{\dagger}}(\boldsymbol{\beta}_0^{\dagger},g_0) - \widehat{\boldsymbol{\Sigma}}^*_{\boldsymbol{\phi}^{\dagger}}(\boldsymbol{\beta}_0^{\dagger},g_0)\right]_{(1,1)} = \frac{1}{|W|}\left[\sum_{\mathbf{s}\in N\cap W}\phi^{\dagger}_{0,\hat{g}}(\mathbf{s};\widehat{\boldsymbol{\beta}}^{\dagger}_{\boldsymbol{\phi}^{\dagger}_{\hat{g}}}) - \int_W \psi(\mathbf{s};\widehat{\boldsymbol{\beta}}^{\dagger}_{\boldsymbol{\phi}^{\dagger}_{\hat{g}}})\phi^{\dagger}_{0,\hat{g}}(\mathbf{s};\widehat{\boldsymbol{\beta}}^{\dagger}_{\boldsymbol{\phi}^{\dagger}_{\hat{g}}})\mathrm{d}\mathbf{s}\right] = 0. \tag{S.10}$$

Consequently, by the definition of the Semi-G test, under the model $\psi(\mathbf{u};\boldsymbol{\beta})$, the test function is of the form.

$$\boldsymbol{f}_g^G(\mathbf{s};\boldsymbol{\beta}) = \left[Z_1(\mathbf{s})\phi_{0,g}(\mathbf{s};\boldsymbol{\beta}),\cdots,Z_p(\mathbf{s})\phi_{0,g}(\mathbf{s};\boldsymbol{\beta}),Z_1(\mathbf{s})\phi_{1,g}(\mathbf{s};\boldsymbol{\beta}),\cdots,Z_p(\mathbf{s})\phi_{p,g}(\mathbf{s};\boldsymbol{\beta})\right]^T.$$

Similarly, under the re-parameterized model $\psi(\mathbf{u};\boldsymbol{\beta}^\dagger)$

$$\boldsymbol{f}_g^{G\dagger}(\mathbf{s};\boldsymbol{\beta}) = \left[\widetilde{Z}_1(\mathbf{s})\phi_{0,g}^\dagger(\mathbf{s};\boldsymbol{\beta}^\dagger),\cdots,\widetilde{Z}_p(\mathbf{s})\phi_{0,g}^\dagger(\mathbf{s};\boldsymbol{\beta}^\dagger),\widetilde{Z}_1(\mathbf{s})\phi_{1,g}^\dagger(\mathbf{s};\boldsymbol{\beta}^\dagger),\cdots,\widetilde{Z}_p(\mathbf{s})\phi_{p,g}^\dagger(\mathbf{s};\boldsymbol{\beta})\right]^T.$$

By the definition of matrix $\mathbf{T}$ in (S.8), and the relationship $\boldsymbol{\eta}(\mathbf{s};\boldsymbol{\beta}) = \mathbf{T}\boldsymbol{\eta}^\dagger(\mathbf{s};\boldsymbol{\beta}^\dagger)$ and $\boldsymbol{\phi}_g(\mathbf{s};\boldsymbol{\beta}) = \mathbf{T}\boldsymbol{\phi}_g^\dagger(\mathbf{s};\boldsymbol{\beta}^\dagger)$, we have that $\phi_{0,g}(\mathbf{s};\boldsymbol{\beta}) = \phi_{0,g}^\dagger(\mathbf{s};\boldsymbol{\beta}^\dagger)$ and that $Z(\mathbf{s}) = a_j + b_j\widetilde{Z}_j(\mathbf{s})$, $\phi_{j,g}(\mathbf{s};\boldsymbol{\beta}) = a_j\phi_{0,g}^\dagger(\mathbf{s};\boldsymbol{\beta}^\dagger) + b_j\phi_{j,g}^\dagger(\mathbf{s};\boldsymbol{\beta}^\dagger)$, $j = 1,\cdots,p$. Plugging these inequality back to the definition

$$\mathbf{Q}^{\mathrm{G}}(\widehat{\boldsymbol{\beta}}_{\boldsymbol{\phi}_{\hat g}}) = \frac{1}{\sqrt{|W|}}\left[\sum_{\mathbf{s}\in N\cap W}\boldsymbol{f}_g^G(\mathbf{s};\widehat{\boldsymbol{\beta}}_{\boldsymbol{\phi}_{\hat g}}) - \int_W \psi(\mathbf{s};\widehat{\boldsymbol{\beta}}_{\boldsymbol{\phi}_{\hat g}})\boldsymbol{f}_g^G(\mathbf{s};\widehat{\boldsymbol{\beta}}_{\boldsymbol{\phi}_{\hat g}})\mathrm{d}\mathbf{s}\right], \text{ and}$$

$$\mathbf{Q}^{\mathrm{G}}(\widehat{\boldsymbol{\beta}}^\dagger_{\boldsymbol{\phi}^\dagger_{\hat g}}) = \frac{1}{\sqrt{|W|}}\left[\sum_{\mathbf{s}\in N\cap W}\boldsymbol{f}_g^{G\dagger}(\mathbf{s};\widehat{\boldsymbol{\beta}}^\dagger_{\boldsymbol{\phi}^\dagger_{\hat g}}) - \int_W \psi(\mathbf{s};\widehat{\boldsymbol{\beta}}^\dagger_{\boldsymbol{\phi}^\dagger_{\hat g}})\boldsymbol{f}_g^{G\dagger}(\mathbf{s};\widehat{\boldsymbol{\beta}}^\dagger_{\boldsymbol{\phi}^\dagger_{\hat g}})\mathrm{d}\mathbf{s}\right],$$

and making use of the equality (S.10) together with the fact $\psi(\cdot;\widehat{\boldsymbol{\beta}}_{\boldsymbol{\phi}_{\hat g}}) = \psi(\cdot;\widehat{\boldsymbol{\beta}}^\dagger_{\boldsymbol{\phi}^\dagger_{\hat g}})$, one has the relationship that

$$\mathbf{Q}^{\mathrm{G}}(\widehat{\boldsymbol{\beta}}_{\boldsymbol{\phi}_{\hat g}}) = \mathbf{T}_1\mathbf{Q}^{\mathrm{G}}(\widehat{\boldsymbol{\beta}}^\dagger_{\boldsymbol{\phi}^\dagger_{\hat g}}),$$

where the matrix

$$\mathbf{T}_1 = \begin{pmatrix} \mathrm{diag}\{b_1,\cdots,b_p\} & \mathbf{0}_{p\times p} \\ \mathrm{diag}\{a_1b_1,\cdots,a_pb_p\} & \mathrm{diag}\{b_1^2,\cdots,b_p^2\}\end{pmatrix}$$

Then by the definition of the test statistic, some straightforward but tedious algebra gives

$$\begin{aligned} T_{\boldsymbol{f}_g^G,n} - \widetilde{T}_{\boldsymbol{f}_g^G,n} &= \left[\mathbf{Q}^{\mathrm{G}}(\widehat{\boldsymbol{\beta}}_{\boldsymbol{\phi}_{\hat g}})\right]^T \mathrm{Var}\left[\mathbf{Q}^{\mathrm{G}}(\widehat{\boldsymbol{\beta}}_{\boldsymbol{\phi}_{\hat g}})\right]^{-1}\mathbf{Q}^{\mathrm{G}}(\widehat{\boldsymbol{\beta}}_{\boldsymbol{\phi}_{\hat g}}) - \left[\mathbf{Q}^{\mathrm{G}}(\widehat{\boldsymbol{\beta}}^\dagger_{\boldsymbol{\phi}^\dagger_{\hat g}})\right]^T \mathrm{Var}\left[\mathbf{Q}^{\mathrm{G}}(\widehat{\boldsymbol{\beta}}^\dagger_{\boldsymbol{\phi}^\dagger_{\hat g}})\right]^{-1}\mathbf{Q}^{\mathrm{G}}(\widehat{\boldsymbol{\beta}}^\dagger_{\boldsymbol{\phi}^\dagger_{\hat g}}) + o_p(1) \\ &= \left[\mathbf{Q}^{\mathrm{G}}(\widehat{\boldsymbol{\beta}}_{\boldsymbol{\phi}_{\hat g}})\right]^T \mathrm{Var}\left[\mathbf{Q}^{\mathrm{G}}(\widehat{\boldsymbol{\beta}}_{\boldsymbol{\phi}_{\hat g}})\right]^{-1}\mathbf{Q}^{\mathrm{G}}(\widehat{\boldsymbol{\beta}}_{\boldsymbol{\phi}_{\hat g}}) - \left[\mathbf{Q}^{\mathrm{G}}(\widehat{\boldsymbol{\beta}}_{\boldsymbol{\phi}_{\hat g}})\right]^T \mathbf{T}_1^T\left\{\mathbf{T}_1\mathrm{Var}\left[\mathbf{Q}^{\mathrm{G}}(\widehat{\boldsymbol{\beta}}_{\boldsymbol{\phi}_{\hat g}})\right]\mathbf{T}_1^T\right\}^{-1}\mathbf{T}_1\mathbf{Q}^{\mathrm{G}}(\widehat{\boldsymbol{\beta}}_{\boldsymbol{\phi}_{\hat g}}) + o_p(1). \end{aligned}$$

Finally, since $\boldsymbol{\Sigma}_{\boldsymbol{f}_g^G}(\boldsymbol{\beta}_0, g_0)$ is full rank and $\mathrm{Var}\left[\mathbf{Q}^{\mathrm{G}}(\widehat{\boldsymbol{\beta}}_{\boldsymbol{\phi}_{\hat g}})\right] = \boldsymbol{\Sigma}_{\boldsymbol{f}_g^G}(\boldsymbol{\beta}_0, g_0) + o(1)$, we have that $\mathrm{Var}\left[\mathbf{Q}^{\mathrm{G}}(\widehat{\boldsymbol{\beta}}_{\boldsymbol{\phi}_{\hat g}})\right]$ is invertable, which further gives that

$$T_{\boldsymbol{f}_g^G,n} - \widetilde{T}_{\boldsymbol{f}_g^G,n} = o_p(1).$$

The proof of Proposition 1 is completed.

S.4 Proof of Theorem 1

S.4.1 Notations and Conditions

Let $\mathbf{A}_n$, $\mathbf{b}_n$, $g_{\mathrm{s},n}(\cdot;\boldsymbol{\theta})$ be the corresponding quantities defined in Section 3.2. With slight abuse of notation, we denote all linear constraints in the quadratic programming problem outline in (8)-(11) as $\mathbf{C}_n\boldsymbol{\theta} \geq 0$, although we set the first constraint as an equality constraint. For two sequences $\{a_n\}$ and $\{b_n\}$, we write $a_n = O(b_n)$ or $a_n = o(b_n)$ if a_n/b_n is uniformly bounded for any n or $a_n/b_n \to 0$ as $n \to \infty$, respectively. Let $\xrightarrow{p}$ denote convergence in probability and $\xrightarrow{d}$ denote convergence in distribution. For a sequence of random variables $\{X_n\}$, $X_n = O_p(a_n)$ or $X_n = o_p(a_n)$ if X_n/a_n is bounded in probability or $X_n/a_n \xrightarrow{p} 0$, respectively. Finally, let $\|\mathbf{D}\|_{\max} = \max_{i,j}(|e_{ij}|)$ be the max norm of a matrix or vector $\mathbf{D}$, $\|\mathbf{e}\|_1 = \sum_i |e_i|$ be the ℓ_1 norm of the vector $\mathbf{e}$, and $\lambda_{\max}(\mathbf{D})$ and $\lambda_{\min}(\mathbf{D})$ be the largest and the smallest eigenvalues of the matrix $\mathbf{D}$, respectively.

Define the k'th order joint intensity $\lambda^{(k)}(\cdot)$ by the identity

$$\mathrm{E}\left[\sum_{\mathbf{s}_1\cdots,\mathbf{s}_k\in N}^{\neq} I(\mathbf{s}_1 \in B_1, \cdots, \mathbf{s}_k \in B_k)\right] = \int_{B_1\times\cdots\times B_k} \lambda^{(k)}(\mathbf{t}_1, \cdots, \mathbf{t}_k)\mathrm{d}\mathbf{t}_1\cdots\mathrm{d}\mathbf{t}_k$$

for bounded subsets $B_j \subset \mathbb{R}^2$, $j = 1, \cdots, k$, where the sum is over distinct $\mathbf{s}_1 \cdots, \mathbf{s}_k$. Then, $g^{(k)}(\mathbf{s}_1, \cdots, \mathbf{s}_k) = \lambda^{(k)}(\mathbf{s}_1, \cdots, \mathbf{s}_k)/\left[\lambda(\mathbf{s}_1)\cdots\lambda(\mathbf{s}_k)\right]$ is called the kth-order normalized joint intensities, which are assumed to be translation invariant, i.e. $g^{(k)}(\mathbf{s}_1, \cdots, \mathbf{s}_k) = g_0^{(k)}(\mathbf{s}_2-\mathbf{s}_1, \cdots, \mathbf{s}_k-\mathbf{s}_1)$ for some function $g_0^{(k)}(\cdot)$, for $k = 2, 3, 4$. In particular, we write $g_0(\cdot)$ for $g_0^{(2)}(\cdot)$.

We first establish uniform consistency of the shape constrained PCF estimator $g_{\mathrm{s},n}(\cdot;\widehat{\boldsymbol{\theta}}_n)$ defined in (12), for which the following conditions are sufficient.

[C1] There exist $0 < c_\lambda < C_\lambda < \infty$ such that $c_\lambda \leq \psi(\mathbf{s};\boldsymbol{\beta}) \leq C_\lambda$ for any $\boldsymbol{\beta}$ in a compact set and $\mathbf{s} \in W_n$. Furthermore, $\psi(\mathbf{s};\boldsymbol{\beta})$ is twice differentiable with respect to $\boldsymbol{\beta}$ and $\|\boldsymbol{\psi}^{(1)}(\mathbf{s};\boldsymbol{\beta})\|_{\max} \leq C_\lambda$, $\|\partial^2\psi(\mathbf{s};\boldsymbol{\beta})/(\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T)\|_{\max} \leq C_\lambda$ for any $\boldsymbol{\beta}$ and $\mathbf{s} \in W_n$.

[C2] Assume that there exist a $C_g > 0$ such that (a) $\int_{\mathbb{R}^2} |g_0(\|\mathbf{s}\|) - 1|\mathrm{d}\mathbf{s} \leq C_g$; (b) $\left|g^{(k)}(\mathbf{s}_1, \mathbf{s}_2, \cdots, \mathbf{s}_k)\right| \leq C_g$ for $\mathbf{s}_j \in W_n$, $j = 1, \cdots, k$ and $k = 2, 3$; and (c) $\int_{\mathbb{R}^2} |g_0^{(4)}(\mathbf{s}, \mathbf{t} +$

$\mathbf{w}, \mathbf{w}) - g_0(\|\mathbf{s}\|)g_0(\|\mathbf{t}\|)|\mathrm{d}\mathbf{w} \le C_g$.

[C3] $\lim_{n\to\infty} \lambda_{\min}\left[\mathbf{S}_n^{\mathrm{p}}(\boldsymbol{\beta}_0)\right] > 0$, with the sensitivity matrix of the Poisson likelihood $\mathbf{S}_n^{\mathrm{p}}(\boldsymbol{\beta}_0) = \frac{1}{|W_n|}\int_{W_n}\left[\psi(\mathbf{s};\boldsymbol{\beta}_0)\right]^{-1}\boldsymbol{\psi}^{(1)}(\mathbf{s};\boldsymbol{\beta}_0)\boldsymbol{\psi}^{(1)T}(\mathbf{s};\boldsymbol{\beta}_0)\mathrm{d}\mathbf{s}$.

[C4] Denoting (7) as $\mathbf{b}_n = (b_{1,n}, \cdots, b_{J_n,n})^T$, there exist $C_0 > 0$, $t_0 > 0$ and n_0, such that for any $n \ge n_0$, it holds that $P\left(\sqrt{|W_n|}\,|b_{j,n} - \mathrm{E}(b_{j,n})| > t\right) \le \exp(-C_0 t)$, for any $t > t_0$, $j = 1, \cdots, J_n$.

[C5] With $\boldsymbol{\theta}_{0,n} = \arg\min_{\mathbf{C}_n\boldsymbol{\theta}\ge\mathbf{0}} \int_0^R w(r)\left[g_{\mathrm{s},n}(r;\boldsymbol{\theta}) - g_0(r)\right]^2 \mathrm{d}r$, assume the approximation error $e_n = \sup_{r\in[0,R]}|g_{\mathrm{s},n}(r;\boldsymbol{\theta}_{0,n}) - g_0(r)| \to 0$ as $|W_n| \to \infty$.

[C6] Define the partition of the constraint matrix $\mathbf{C}_n = \left[\mathbf{C}_{n,\mathcal{A}}^T, \mathbf{C}_{n,\mathcal{A}^c}^T\right]^T$ such that $\mathbf{C}_{n,\mathcal{A}}\boldsymbol{\theta}_{0,n} = \mathbf{0}$ and $\mathbf{C}_{n,\mathcal{A}^c}\boldsymbol{\theta}_{0,n} > \mathbf{0}$, and the quantity $\eta_n = \min_{\|\boldsymbol{\delta}\|_1=1, \mathbf{C}_{n,\mathcal{A}}\boldsymbol{\delta}\ge 0} \boldsymbol{\delta}^T\mathbf{A}_n\boldsymbol{\delta} > 0$. We assume that $\eta_n = O(1)$ and that $\left(e_n + |W_n|^{-1/2}\log J_n\right)/\eta_n = o(1)$ as $|W_n| \to \infty$ (with e_n defined in C5).

Conditions C1-C3 are standard conditions that ensure the consistency of $\widehat{\boldsymbol{\beta}}^{\mathrm{p}}$ and have been widely used in the literature, see, e.g., Schoenberg (2005); Prokešová et al. (2017). Condition C4 requires that each component of $\mathbf{b}_n$ has a distribution whose tail decays exponentially fast when W_n is sufficiently large. This is a mild condition due to the definition of $\mathbf{b}_n$ in (7), each component of which can be shown to be asymptotically normal as $|W_n| \to \infty$ under some additional mild conditions such as N1 and N2 in Section S.6.1. The tail of a normal distribution decays even faster than the exponential rate. Condition C5 asserts that $g_{\mathrm{s},n}(\cdot;\boldsymbol{\theta}_{0,n})$ can approximate the true $g_0(\cdot)$ sufficiently well, which applies to a large class of existing point process models provided that J_n and L_n are sufficiently large.

Condition C6 is similar to the "self-regularizing property" (Slawski et al., 2013) and the "minimal positive compatible eigenvalue" (Meinshausen et al., 2013) that are used to introduce sparsity in the solution $\widehat{\boldsymbol{\theta}}_n$ to the sign constrained optimization problem defined by (8)-(11). In other words, a large η_n means that more components of the solution $\widehat{\boldsymbol{\theta}}_n$ are 0's. The magnitude of η_n depends on three quantities: J_n, L_n and $\boldsymbol{\theta}_{0,n}$. Recall that $\mathbf{A}_n$ is a $J_n \times J_n$ matrix, and thus increasing J_n will potentially decrease η_n. However, the effect of a large J_n will be greatly offset by the number of equality constraints in $\mathbf{C}_{n,\mathcal{A}}$, which is controlled by the value of $\boldsymbol{\theta}_{0,n}$. The more constraints in $\mathbf{C}_{n,\mathcal{A}}$, the larger η_n is by its definition. This explains the observation that increasing the number of knots J_n in the M-spline approximation (6) has rather limited impact on $g_{\mathrm{s},n}(\cdot;\widehat{\boldsymbol{\theta}}_n)$ because most of the components in $\widehat{\boldsymbol{\theta}}_n$ are forced to be 0's. This is consistent with the self-regularizing property of the nonnegative least square estimators studied in Slawski et al. (2013) and Meinshausen et al. (2013). Finally, the impact of L_n on η_n is that increasing L_n tend to increase the singularity of the matrix $\mathbf{A}_n$, leading to a smaller η_n. Therefore, condition C6 essentially imposes some implicit restrictions on J_n, L_n and the underlying truth $\boldsymbol{\theta}_{0,n}$.

S.4.2 Proof

Lemma 1. *For a sequence of random variables $X_1, \cdots, X_{J_n}$ such that $\mathrm{E}X_j = 0$ and $P(|X_j| > t) \leq \exp(-C_0 t)$ for some constant $C_0 > 0$ and $t > t_0$ with $t_0 > 0$ being some constant, and $J = 1, \cdots, J_n$, then we have that*

$$\max\left(|X_1|, \cdots, |X_{J_n}|\right) = O_p\left(\log J_n\right).$$

Proof. For any constant $C > 0$, it is straightforward to show that

$$\begin{aligned} P\left(\max\left(|X_1|, \cdots, |X_{J_n}|\right) > C\log J_n\right) &= P\left(\cup_{j=1}^{J_n}\{|X_j| > C\log J_n\}\right) \\ &\leq \sum_{j=1}^{J_n} P\left(|X_j| > C\log J_n\right) \\ &\leq J_n \exp(-C_0 C\log J_n) = J_n^{1-CC_0}, \end{aligned}$$

where the right-hand side can be arbitrarily small when C increases. Therefore, by the definition of convergence in probability, we have that

$$\max\left(|X_1|, \cdots, |X_{J_n}|\right) = O_p\left(\log J_n\right).$$

□

Proof of Theorem 1. Under conditions C1-C3, Theorem 1 of Waagepetersen and Guan (2009) asserts that $\sqrt{|W_n|}\left(\widehat{\boldsymbol{\beta}}^{\mathrm{p}} - \boldsymbol{\beta}_0\right) = O_p(1)$. Therefore, to simplify the presentation, we assume that

$\widehat{\boldsymbol{\beta}}^{\mathrm{p}}$ can be replaced safely with $\boldsymbol{\beta}_0$ without altering the asymptotic results.

Define the constant $c_{\max} = \|\mathbf{C}_n^T \boldsymbol{\theta}_{0,n}\|_{\max}$, then for any $0 < \alpha_n \leq c_{\max}$ and a vector $\boldsymbol{\delta}$ such that $\|\boldsymbol{\delta}\|_1 \leq 1$ and $\mathbf{C}_{n,\mathcal{A}}\boldsymbol{\delta} \geq \mathbf{0}$, it is straightforward to see that $\tilde{\boldsymbol{\theta}}_n = \boldsymbol{\theta}_{0,n} + \alpha_n \boldsymbol{\delta}$ is a feasible solution for the optimization problem in equation (9), that is,

$$\text{minimize } Q_n(\boldsymbol{\theta}) = \frac{1}{2}\boldsymbol{\theta}^T \mathbf{A}_n \boldsymbol{\theta} + \boldsymbol{\theta}^T \mathbf{b}_n \text{ s.t. } \mathbf{C}_n \boldsymbol{\theta} \geq 0. \tag{S.11}$$

Let $\alpha_n = C\eta_n^{-1}(e_n + |W_n|^{-1/2}\log J_n)$ for some large constant C and it suffices to show that for any given $\epsilon > 0$, for large enough W_n, we have

$$P\left[\inf_{\|\boldsymbol{\delta}\|_1 = 1, \mathbf{C}_{n,\mathcal{A}}\boldsymbol{\delta} \geq \mathbf{0}} Q_n(\boldsymbol{\theta}_{0,n} + \alpha_n \boldsymbol{\delta}) > Q_n(\boldsymbol{\theta}_{0,n})\right] \geq 1 - \varepsilon. \tag{S.12}$$

Note that any feasible solution to (S.11) that is close to $\boldsymbol{\theta}_{0,n}$, denoted as $\tilde{\boldsymbol{\theta}}_n$ such that $\|\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_{0,n}\|_1 \leq \alpha_n$, must satisfy $\mathbf{C}_{n,\mathcal{A}}(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_{0,n}) \geq \mathbf{0}$ by the definition of $\mathbf{C}_{n,\mathcal{A}}$. Therefore, combining the fact that $Q_n(\boldsymbol{\theta})$ is a convex function of $\boldsymbol{\theta}$, the inequality (S.12) implies that, with a probability tending to 1, there exists a local minimizer $\widehat{\boldsymbol{\theta}}_n$ in the feasible solution region $\{\boldsymbol{\theta}_{0,n} + \alpha_n \boldsymbol{\delta} : \|\boldsymbol{\delta}\|_1 \leq 1, \mathbf{C}_{n,\mathcal{A}}\boldsymbol{\delta} \geq \mathbf{0}\}$ such that $\|\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_{0,n}\|_1 = O_p(\alpha_n)$.

It is straightforward to show that

$$Q_n(\boldsymbol{\theta}_{0,n} + \alpha_n \boldsymbol{\delta}) - Q_n(\boldsymbol{\theta}_{0,n}) = \frac{\alpha_n^2}{2}\boldsymbol{\delta}^T \mathbf{A}_n \boldsymbol{\delta} + \alpha_n \boldsymbol{\delta}^T\left(\mathbf{b}_n + \mathbf{A}_n \boldsymbol{\theta}_{0,n}\right).$$

Define the random variable $Z_n = \sup_{\|\boldsymbol{\delta}\|_1 = 1} |\boldsymbol{\delta}^T\left(\mathbf{b}_n + \mathbf{A}_n \boldsymbol{\theta}_{0,n}\right)|$, and it is easy to see that

$$\inf_{\|\boldsymbol{\delta}\| = 1, \mathbf{C}_{n,\mathcal{A}}\boldsymbol{\delta} \geq \mathbf{0}} Q_n(\boldsymbol{\theta}_{0,n} + \alpha_n \boldsymbol{\delta}) - Q_n(\boldsymbol{\theta}_{0,n}) \geq \frac{\alpha_n^2}{2}\eta_n - \alpha_n Z_n = \frac{\alpha_n^2}{2}\eta_n - \frac{1}{C}\frac{\alpha_n^2 \eta_n Z_n}{e_n + |W_n|^{-1/2}\log J_n}, \tag{S.13}$$

which implies that to show (S.12), it suffices to show that $Z_n = O_p(e_n + |W_n|^{-1/2}\log J_n)$ with a sufficiently large constant C.

Note that $\boldsymbol{\delta}^T\left(\mathbf{b}_n + \mathbf{A}_n \boldsymbol{\theta}_{0,n}\right) = \int_0^R w(r)\left[\boldsymbol{\theta}_{0,n}^T \mathbf{x}_g(r) + 1 - g_0(r)\right]\left[\boldsymbol{\delta}^T \mathbf{x}_{g,n}(r)\right] \mathrm{d}r + \boldsymbol{\delta}^T(\mathbf{b}_n - \mathrm{E}\mathbf{b}_n)$, then using the triangle inequality, we have that

$$
\begin{aligned}
Z_n &\le \sup_{\|\boldsymbol{\delta}\|_1=1} \left| \int_0^R w(r) \left[\boldsymbol{\theta}_{0,n}^T \mathbf{x}_g(r) + 1 - g_0(r)\right] \left[\boldsymbol{\delta}^T \mathbf{x}_{g,n}(r)\right] \mathrm{d}r \right| + \sup_{\|\boldsymbol{\delta}\|_1=1} \left|\boldsymbol{\delta}^T (\mathbf{b}_n - \mathrm{E}\mathbf{b}_n)\right| \\
&\le \sup_{r\in[0,R]} \left|\boldsymbol{\theta}_{0,n}^T \mathbf{x}_g(r) + 1 - g_0(r)\right| \times \int_0^R w(r) \|\mathbf{x}_{g,n}(r)\|_{\max} \mathrm{d}r + \|\mathbf{b}_n - \mathrm{E}\mathbf{b}_n\|_{\max} \\
&= O(1) e_n + \|\mathbf{b}_n - \mathrm{E}\mathbf{b}_n\|_{\max},
\end{aligned}
$$

where the last equality follows from condition C5 and the fact that $\|\mathbf{x}_{g,n}(r)\|_{\max}$ is uniformly bounded by its definition for any $r \in [0, R]$. Using condition C4 and Lemma 1, it readily follows that $\|\mathbf{b}_n - \mathrm{E}\mathbf{b}_n\|_{\max} = O_p(|W_n|^{-1/2} \log J_n)$ and as a result $Z_n = O_p(e_n + |W_n|^{-1/2} \log J_n)$, which completes the proof for inequality (S.12). Therefore, we have that

$$
\|\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_{0,n}\|_1 = O_p\left(\frac{e_n + |W_n|^{-1/2}\log J_n}{\eta_n}\right).
$$

Furthermore, using the triangular inequality, we have that

$$
\begin{aligned}
\sup_{0\le r\le R} |g_{\mathrm{s},n}(r; \widehat{\boldsymbol{\theta}}_n) - g_0(r)| &\le \sup_{0\le r\le R} |g_{\mathrm{s},n}(r; \widehat{\boldsymbol{\theta}}_n) - g_{\mathrm{s},n}(r; \boldsymbol{\theta}_{0,n})| + \sup_{0\le r\le R} |g_{\mathrm{s},n}(r; \boldsymbol{\theta}_{0,n}) - g_0(r)| \\
&\le \sup_{0\le r\le R} \left| \mathbf{x}_{g,n}^T(r) \left(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_{0,n}\right)\right| + \sup_{0\le r\le R} |g_{\mathrm{s},n}(r; \boldsymbol{\theta}_{0,n}) - g_0(r)| \\
&\le \left[\sup_{0\le r\le R} \|\mathbf{x}_{g,n}^T(r)\|_{\max}\right] \|\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_{0,n}\|_1 + \sup_{0\le r\le R} |g_{\mathrm{s},n}(r; \boldsymbol{\theta}_{0,n}) - g_0(r)| \\
&= O_p\left(\frac{e_n + |W_n|^{-1/2}\log J_n}{\eta_n}\right) + O_p(e_n) \\
&= O_p\left(\frac{e_n + |W_n|^{-1/2}\log J_n}{\eta_n}\right),
\end{aligned}
$$

which completes the proof of Theorem 1. □

S.5 Proof of Theorem 2

S.5.1 Conditions and Lemmas

[L1] Assume that $\boldsymbol{\phi}_{n,\boldsymbol{\theta}}(\mathbf{s}; \boldsymbol{\beta})$ is differentiable with respect to $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$, and that there exists some constant $K_1 > 0$ such that $\sup_{\mathbf{s}\in W_n} \|\boldsymbol{\phi}_{n,\boldsymbol{\theta}}(\mathbf{s}; \boldsymbol{\beta})\|_{\max}$, $\sup_{\mathbf{s}\in W_n} \|\frac{\partial \boldsymbol{\phi}_{n,\boldsymbol{\theta}}(\mathbf{s};\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\|_{\max}$, $\sup_{\mathbf{s}\in W_n} \|\frac{\partial \boldsymbol{\phi}_{n,\boldsymbol{\theta}}(\mathbf{s};\boldsymbol{\beta})}{\partial \boldsymbol{\theta}}\|_{\max}$ are uniformly bounded for $(\boldsymbol{\beta}, \boldsymbol{\theta})$ satisfying $\|\boldsymbol{\beta} - \boldsymbol{\beta}_0\|_1 \le K_1$, $\|\boldsymbol{\theta} - \boldsymbol{\theta}_{0,n}\|_1 \le K_1$.

[L2] Denote matrix $\mathbf{H}_{\boldsymbol{\phi},n}(\boldsymbol{\theta}) = \frac{1}{\sqrt{|W_n|}}\left[\sum_{\mathbf{s}\in N\cap W_n} \frac{\partial \boldsymbol{\phi}_{n,\boldsymbol{\theta}}(\mathbf{s};\boldsymbol{\beta}_0)}{\partial \boldsymbol{\theta}} - \int_{W_n} \psi(\mathbf{s};\boldsymbol{\beta}_0)\frac{\partial \boldsymbol{\phi}_{n,\boldsymbol{\theta}}(\mathbf{s};\boldsymbol{\beta}_0)}{\partial \boldsymbol{\theta}}\mathrm{d}\mathbf{s}\right]$ and its ijth element as $h_{ij,\boldsymbol{\phi},n}(\boldsymbol{\theta})$. Assume that there exists a constant $K_2 > 0$ and n_0 such that for any $n \geq n_0$, it holds that for any $\boldsymbol{\theta}$ satisfying $\|\boldsymbol{\theta} - \boldsymbol{\theta}_{0,n}\|_1 \leq K_1$, $P\left(|h_{ij,\boldsymbol{\phi},n}(\boldsymbol{\theta})| > t\right) \leq \exp\left(-K_2 t\right)$, for any $t > t_0$ with $t_0 > 0$ being some constant and $i = 1, \cdots, p$, $j = 1, \cdots, J_n$.

[L3] Let $\mathbf{S}_{\boldsymbol{\phi},n}(\boldsymbol{\beta}_0, \boldsymbol{\theta}_{0,n}) = \frac{1}{|W_n|}\int_{W_n} \boldsymbol{\psi}^{(1)}(\mathbf{s};\boldsymbol{\beta}_0)\boldsymbol{\phi}^T_{n,\boldsymbol{\theta}_{0,n}}(\mathbf{s};\boldsymbol{\beta}_0)\mathrm{d}\mathbf{s}$, where $\boldsymbol{\psi}^{(1)}(\mathbf{s};\boldsymbol{\beta}) = \frac{\partial \psi(\mathbf{s};\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}$. We assume that $\liminf_{n\to\infty} \lambda_{\min}\left[\mathbf{S}_n(\boldsymbol{\beta}_0, \boldsymbol{\theta}_{0,n})\right] > 0$.

Conditions L1 and L3 are taken from Guan et al. (2015) and are rather mild conditions. Condition L2 is the same as the condition C4 and can be justified by the fact that under some mild conditions, each component in $\mathbf{H}_{\boldsymbol{\phi},n}(\boldsymbol{\theta})$ is asymptotically normal as $|W_n| \to \infty$, whose tail decays even faster than the exponential rate.

Lemma 2. *For any bounded function $h : \mathbb{R}^2 \to \mathbb{R}$, under conditions C1-C2, we have that*

$$\mathrm{Var}\left[\sum_{\mathbf{s}\in N\cap W_n} h(\mathbf{s})\right] = O(|W_n|).$$

Proof. Using the Campbell's formula, it is straightforward to show that, under conditions C1-C2,

$$\begin{aligned}\mathrm{Var}\left[\sum_{\mathbf{s}\in N\cap W_n} h(\mathbf{s})\right] &= \int_{W_n} \psi(\mathbf{s};\boldsymbol{\beta}_0)h^2(\mathbf{s}) + \int_{W_n}\int_{W_n} \psi(\mathbf{s};\boldsymbol{\beta}_0)\psi(\mathbf{t};\boldsymbol{\beta}_0)\left[g_0(\|\mathbf{s}-\mathbf{t}\|) - 1\right]h(\mathbf{s})h(\mathbf{t})\mathrm{d}\mathbf{s}\mathrm{d}\mathbf{t} \\ &\leq C_\lambda |W_n| \sup_{\mathbf{s}\in W_n} h^2(\mathbf{s}) + C_\lambda^2 |W_n| C_g \left[\sup_{\mathbf{s}\in W_n} h(\mathbf{s})\right]^2 = O(|W_n|).\end{aligned}$$

□

Lemma 3. *Under conditions C1, C2 and L1, if $\|\boldsymbol{\beta} - \boldsymbol{\beta}_0\|_1 \to 0$ and $\|\boldsymbol{\theta} - \boldsymbol{\theta}_{0,n}\|_1 \to 0$, then*

$$\sup_{\mathbf{s}\in W_n} \left\|\boldsymbol{\phi}_{n,\boldsymbol{\theta}}(\mathbf{s};\boldsymbol{\beta}) - \boldsymbol{\phi}_{n,\boldsymbol{\theta}_{0,n}}(\mathbf{s};\boldsymbol{\beta}_0)\right\|_{\max} = O(\|\boldsymbol{\beta} - \boldsymbol{\beta}_0\|_1 + \|\boldsymbol{\theta} - \boldsymbol{\theta}_{0,n}\|_1).$$

Proof. For any given $\boldsymbol{\beta}, \boldsymbol{\theta}$, recall the Fredholm integral equation of the second kind using the shape constrained PCF estimator $g_{\mathrm{r}}(r;\boldsymbol{\theta})$ that gives the optimal weight function $\boldsymbol{\phi}_{\boldsymbol{\theta}}(\mathbf{s};\boldsymbol{\beta})$

$$\boldsymbol{\phi}_{n,\boldsymbol{\theta}}(\mathbf{s};\boldsymbol{\beta}) + \int_{W_n} \boldsymbol{\phi}_{n,\boldsymbol{\theta}}(\mathbf{u};\boldsymbol{\beta})\psi(\mathbf{u};\boldsymbol{\beta})\left[g_{\mathrm{s},n}(\|\mathbf{s}-\mathbf{u}\|;\boldsymbol{\theta}) - 1\right]\mathrm{d}\mathbf{u} = \boldsymbol{\psi}^{(1)}(\mathbf{s};\boldsymbol{\beta})/\psi(\mathbf{s};\boldsymbol{\beta}). \tag{S.14}$$

Define function $\mathbf{h}_n(\mathbf{s}) = \boldsymbol{\phi}_{n,\boldsymbol{\theta}}(\mathbf{s};\boldsymbol{\beta}) - \boldsymbol{\phi}_{n,\boldsymbol{\theta}_{0,n}}(\mathbf{s};\boldsymbol{\beta}_0)$, which must satisfy the integral equation

$$\mathbf{h}_n(\mathbf{s}) + \int_{W_n} \mathbf{h}_n(\mathbf{u})\psi(\mathbf{u};\boldsymbol{\beta}_0)\left[g_{\mathrm{s},n}(\|\mathbf{s}-\mathbf{u}\|;\boldsymbol{\theta}_{0,n}) - 1\right] \mathrm{d}\mathbf{u} = \Delta_n(\mathbf{s}), \tag{S.15}$$

where

$$\begin{aligned}
\Delta_n(\mathbf{s}) =& \frac{\boldsymbol{\psi}^{(1)}(\mathbf{s};\boldsymbol{\beta})}{\psi(\mathbf{s};\boldsymbol{\beta})} - \frac{\boldsymbol{\psi}^{(1)}(\mathbf{s};\boldsymbol{\beta}_0)}{\psi(\mathbf{s};\boldsymbol{\beta}_0)} - \int_{W_n} \boldsymbol{\phi}_{n,\boldsymbol{\theta}}(\mathbf{u};\boldsymbol{\beta})\psi(\mathbf{u};\boldsymbol{\beta})\left[g_{\mathrm{s},n}(\|\mathbf{s}-\mathbf{u}\|;\boldsymbol{\theta}) - 1\right] \mathrm{d}\mathbf{u} \\
&+ \int_{W_n} \boldsymbol{\phi}_{n,\boldsymbol{\theta}}(\mathbf{u};\boldsymbol{\beta})\psi(\mathbf{u};\boldsymbol{\beta}_0)\left[g_{\mathrm{s},n}(\|\mathbf{s}-\mathbf{u}\|;\boldsymbol{\theta}_{0,n}) - 1\right] \mathrm{d}\mathbf{u} \\
=& \frac{\boldsymbol{\psi}^{(1)}(\mathbf{s};\boldsymbol{\beta})}{\psi(\mathbf{s};\boldsymbol{\beta})} - \frac{\boldsymbol{\psi}^{(1)}(\mathbf{s};\boldsymbol{\beta}_0)}{\psi(\mathbf{s};\boldsymbol{\beta}_0)} - \int_{W_n} \boldsymbol{\phi}_{n,\boldsymbol{\theta}}(\mathbf{u};\boldsymbol{\beta})\left[\psi(\mathbf{u};\boldsymbol{\beta}) - \psi(\mathbf{u};\boldsymbol{\beta}_0)\right]\left[g_{\mathrm{s},n}(\|\mathbf{s}-\mathbf{u}\|;\boldsymbol{\theta}) - 1\right] \mathrm{d}\mathbf{u} \\
&+ \int_{W_n} \boldsymbol{\phi}_{n,\boldsymbol{\theta}}(\mathbf{u};\boldsymbol{\beta})\psi(\mathbf{u};\boldsymbol{\beta}_0)\left[g_{\mathrm{s},n}(\|\mathbf{s}-\mathbf{u}\|;\boldsymbol{\theta}_{0,n}) - g_{\mathrm{s},n}(\|\mathbf{s}-\mathbf{u}\|;\boldsymbol{\theta})\right] \mathrm{d}\mathbf{u}.
\end{aligned}$$

Under conditions C1-C2 and L1, and the fact that $g_{\mathrm{s},n}(r;\boldsymbol{\theta})$ is bounded and $g_{\mathrm{s},n}(r;\boldsymbol{\theta}) = 1$ for any $r > R$ by design, a straightforward application of Taylor expansion yields that

$$\sup_{\mathbf{s}\in W_n} \|\Delta_n(\mathbf{s})\|_{\max} = O(1)\|\boldsymbol{\beta}-\boldsymbol{\beta}_0\|_1 + O(1)\|\boldsymbol{\theta}-\boldsymbol{\theta}_{0,n}\|_1. \tag{S.16}$$

Define the identity functional operator $\mathbf{I}(f) = f(\mathbf{s})$ for $\mathbf{s} \in W_n$ and

$$\boldsymbol{K}_n(f) = \int_{W_n} f(\mathbf{u})\psi(\mathbf{u};\boldsymbol{\beta}_0)\left[g_{\mathrm{s},n}(\|\mathbf{s}-\mathbf{u}\|;\boldsymbol{\theta}_{0,n}) - 1\right] \mathrm{d}\mathbf{u}.$$

Denote by $(\mathbf{I}+\boldsymbol{K}_n)^{-1}$ the inverse operator of the linear operator $\mathbf{I}+\boldsymbol{K}_n$. Since the solution to the integral equation (S.15) must satisfy $\mathbf{h}_n(\mathbf{s}) = (\mathbf{I}+\boldsymbol{K}_n)^{-1}\Delta_n(\mathbf{s})$, we have that

$$\sup_{\mathbf{s}\in W_n} \|\mathbf{h}_n(\mathbf{s})\|_{\max} = \sup_{\mathbf{s}\in W_n} \|(\mathbf{I}+\boldsymbol{K}_n)^{-1}\Delta(\mathbf{s})\|_{\max} \le \|(\mathbf{I}+\boldsymbol{K}_n)^{-1}\|_{op} \sup_{\mathbf{s}\in W_n} \|\Delta(\mathbf{s})\|_{\max}, \tag{S.17}$$

where $\|(\mathbf{I}+\boldsymbol{K}_n)^{-1}\|_{op}$ is the operator norm of the linear operator $(\mathbf{I}+\mathbf{K}_n)^{-1}$. Since under the constraint S2, $g_{\mathrm{s},n}(\|\mathbf{s}-\mathbf{t}\|;\boldsymbol{\theta}_{0,n})-1$ is a positive semi-definite function of $(\mathbf{s},\mathbf{t}) \in W_n^2$, hence the linear operator $(\mathbf{I}+\boldsymbol{K}_n)$ does not have an eigen-value 0, and hence $\|(\mathbf{I}+\boldsymbol{K}_n)^{-1}\|_{op} < \infty$. Therefore, making use of equation (S.16) and (S.17), we conclude that $\sup_{\mathbf{s}\in W_n} \left\|\boldsymbol{\phi}_{n,\boldsymbol{\theta}}(\mathbf{s};\boldsymbol{\beta}) - \boldsymbol{\phi}_{n,\boldsymbol{\theta}_{0,n}}(\mathbf{s};\boldsymbol{\beta}_0)\right\|_{\max} = O(\|\boldsymbol{\beta}-\boldsymbol{\beta}_0\|_1 + \|\boldsymbol{\theta}-\boldsymbol{\theta}_{0,n}\|_1)$.

□

S.5.2 Proof of Theorem 2

Some straightforward calculus yields that the negative partial derivatives of the optimal estimating function $\mathbf{Q}_{\boldsymbol{\phi}_g}(\boldsymbol{\beta})$ defined in equation (3) is of the form

$$\begin{aligned}\mathbf{J}_{\boldsymbol{\phi},n}(\boldsymbol{\beta},\boldsymbol{\theta}) &= -\frac{1}{\sqrt{|W_n|}}\frac{\partial \mathbf{Q}_{\boldsymbol{\phi},n}(\boldsymbol{\beta},\boldsymbol{\theta})}{\partial \boldsymbol{\beta}} \\ &= -\frac{1}{|W_n|}\left[\sum_{\mathbf{s}\in N\cap W_n}\frac{\partial \boldsymbol{\phi}_{n,\boldsymbol{\theta}}(\mathbf{s};\boldsymbol{\beta})}{\partial\boldsymbol{\beta}} - \int_{W_n}\psi(\mathbf{s};\boldsymbol{\beta})\frac{\partial \boldsymbol{\phi}_{n,\boldsymbol{\theta}}(\mathbf{s};\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}\mathrm{d}\mathbf{s}\right] + \mathbf{S}_{\boldsymbol{\phi},n}(\boldsymbol{\beta},\boldsymbol{\theta}),\end{aligned} \tag{S.18}$$

where $\mathbf{S}_{\boldsymbol{\phi},n}(\boldsymbol{\beta},\boldsymbol{\theta})$ is as defined in conditions L3. And the variance of the optimal estimating function is of the form

$$\begin{aligned}\boldsymbol{\Sigma}^{\dagger}_{\boldsymbol{\phi},n}(\boldsymbol{\beta},\boldsymbol{\theta}) &= \mathrm{Var}\left[\mathbf{Q}_{\boldsymbol{\phi}}(\boldsymbol{\beta},\boldsymbol{\theta})\right] = \frac{1}{|W_n|}\int_{W_n}\psi(\mathbf{s};\boldsymbol{\beta}_0)\boldsymbol{\phi}_{n,\boldsymbol{\theta}}(\mathbf{s};\boldsymbol{\beta})\boldsymbol{\phi}^T_{n,\boldsymbol{\theta}}(\mathbf{s};\boldsymbol{\beta}) \\ &+ \frac{1}{|W_n|}\int_{W_n}\int_{W_n}\psi(\mathbf{s};\boldsymbol{\beta}_0)\psi(\mathbf{t};\boldsymbol{\beta}_0)\left[g_0(\|\mathbf{s}-\mathbf{t}\|)-1\right]\boldsymbol{\phi}_{n,\boldsymbol{\theta}}(\mathbf{s};\boldsymbol{\beta})\boldsymbol{\phi}^T_{n,\boldsymbol{\theta}}(\mathbf{s};\boldsymbol{\beta})\mathrm{d}\mathbf{s}\mathrm{d}\mathbf{t}.\end{aligned} \tag{S.19}$$

By the design of the optimal weight function $\boldsymbol{\phi}(\cdot)$ in equation (4), it is straightforward to show that

$$\begin{aligned}\boldsymbol{\Sigma}^{\dagger}_{\boldsymbol{\phi},n}(\boldsymbol{\beta}_0,\boldsymbol{\theta}_{0,n}) - \mathbf{S}_{\boldsymbol{\phi},n}(\boldsymbol{\beta}_0,\boldsymbol{\theta}_{0,n}) &= \frac{1}{|W_n|}\int_{W_n}\int_{W_n}\psi(\mathbf{s};\boldsymbol{\beta}_0)\psi(\mathbf{t};\boldsymbol{\beta}_0)\left[g_0(\|\mathbf{s}-\mathbf{t}\|)-g_{\mathrm{s},n}(\|\mathbf{s}-\mathbf{t}\|;\boldsymbol{\theta}_{0,n})\right] \\ &\qquad\times \boldsymbol{\phi}_{n,\boldsymbol{\theta}}(\mathbf{s};\boldsymbol{\beta})\boldsymbol{\phi}^T_{n,\boldsymbol{\theta}}(\mathbf{s};\boldsymbol{\beta})\mathrm{d}\mathbf{s}\mathrm{d}\mathbf{t} \\ &\le \sup_{r\in[0,R]}|g_0(r)-g_{\mathrm{s},n}(r;\boldsymbol{\theta}_{0,n})|\times \\ &\quad \frac{1}{|W_n|}\int_{W_n}\int_{W_n}\psi(\mathbf{s};\boldsymbol{\beta}_0)\psi(\mathbf{t};\boldsymbol{\beta}_0)I(\|\mathbf{s}-\mathbf{t}\|\le R)\boldsymbol{\phi}_{n,\boldsymbol{\theta}}(\mathbf{s};\boldsymbol{\beta})\boldsymbol{\phi}^T_{n,\boldsymbol{\theta}}(\mathbf{s};\boldsymbol{\beta})\mathrm{d}\mathbf{s}\mathrm{d}\mathbf{t}\end{aligned}$$

Under the conditions C1, C5 and L1, we have that

$$\left\|\boldsymbol{\Sigma}^{\dagger}_{\boldsymbol{\phi},n}(\boldsymbol{\beta}_0,\boldsymbol{\theta}_{0,n}) - \mathbf{S}_{\boldsymbol{\phi},n}(\boldsymbol{\beta}_0,\boldsymbol{\theta}_{0,n})\right\|_{\max} = O(e_n) = o(1). \tag{S.20}$$

We first show that $\sqrt{|W_n|}(\widehat{\boldsymbol{\beta}}_{\widehat{\boldsymbol{\phi}},n}-\boldsymbol{\beta}_0) = O_p(1)$ following the approach in Guan et al. (2015), which amounts to show

R1 $\||W_n|^{-1}\boldsymbol{\Sigma}^{\dagger-1}_{\boldsymbol{\phi},n}(\boldsymbol{\beta}_0,\boldsymbol{\theta}_{0,n})\|_{\max} = o(1)$.

R2 Let $\alpha_n = \eta_n^{-1}(e_n + |W_n|^{-1/2}\log J_n)$, show that for some constant $d > 0$,

$$\sup_{\|\boldsymbol{\beta}-\boldsymbol{\beta}_0\|_1 \le \frac{d}{\sqrt{|W_n|}}, \|\boldsymbol{\theta}-\boldsymbol{\theta}_{0,n}\|_1 \le d\alpha_n} \|\mathbf{J}_{\boldsymbol{\phi},n}(\boldsymbol{\beta},\boldsymbol{\theta}) - \mathbf{J}_{\boldsymbol{\phi},n}(\boldsymbol{\beta}_0,\boldsymbol{\theta}_{0,n})\|_{\max} = o_p(1).$$

R3 $\left\|\mathbf{J}_{\boldsymbol{\phi},n}(\boldsymbol{\beta}_0,\widehat{\boldsymbol{\theta}}_n) - \mathbf{S}_{\boldsymbol{\phi},n}(\boldsymbol{\beta}_0,\boldsymbol{\theta}_{0,n})\right\|_{\max} = o_p(1)$.

R4 $\left\|\mathbf{Q}_{\boldsymbol{\phi},n}(\boldsymbol{\beta}_0,\widehat{\boldsymbol{\theta}}_n) - \mathbf{Q}_{\boldsymbol{\phi},n}(\boldsymbol{\beta}_0,\boldsymbol{\theta}_{0,n})\right\|_{\max} = o_p(1)$.

R5 $\lim_{n\to\infty} \lambda_{\min}\left[\boldsymbol{\Sigma}^{\dagger-1}_{\boldsymbol{\phi},n}(\boldsymbol{\beta}_0,\boldsymbol{\theta}_{0,n})\mathbf{S}_{\boldsymbol{\phi},n}(\boldsymbol{\beta}_0,\boldsymbol{\theta}_{0,n})\boldsymbol{\Sigma}^{\dagger-1}_{\boldsymbol{\phi},n}(\boldsymbol{\beta}_0,\boldsymbol{\theta}_{0,n})\right] > 0$.

Proof of R1: As $|W_n| \to \infty$, R1 follows directly from condition L3 and equation (S.20).

Proof of R2: Using equation (S.18) and Lemma 2, it readily follows that under condition L1, for any $\|\boldsymbol{\beta}-\boldsymbol{\beta}_0\|_1 \le \frac{d}{\sqrt{|W_n|}}, \|\boldsymbol{\theta}-\boldsymbol{\theta}_{0,n}\|_1 \le d\alpha_n$,

$$\mathrm{Var}\left[J_{ij,\boldsymbol{\phi},n}(\boldsymbol{\beta},\boldsymbol{\theta}) - S_{ij,\boldsymbol{\phi},n}(\boldsymbol{\beta},\boldsymbol{\theta})\right] = O(|W_n|^{-1}), \quad i,j = 1,\cdots,p,$$

and

$$\begin{aligned}
\|\mathrm{E}\left[\mathbf{J}_{\boldsymbol{\phi},n}(\boldsymbol{\beta},\boldsymbol{\theta}) - \mathbf{S}_{\boldsymbol{\phi},n}(\boldsymbol{\beta},\boldsymbol{\theta})\right]\|_{\max} &= \frac{1}{|W_n|}\left\|\int_{W_n}\left[\psi(\mathbf{s};\boldsymbol{\beta}) - \psi(\mathbf{s};\boldsymbol{\beta}_0)\right]\frac{\partial \boldsymbol{\phi}_{n,\boldsymbol{\theta}}(\mathbf{s};\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\mathrm{d}\mathbf{s}\right\|_{\max} \\
&\le \frac{1}{|W_n|}\int_{W_n}|\psi(\mathbf{s};\boldsymbol{\beta}) - \psi(\mathbf{s};\boldsymbol{\beta}_0)|\left\|\frac{\partial \boldsymbol{\phi}_{n,\boldsymbol{\theta}}(\mathbf{s};\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\right\|_{\max}\mathrm{d}\mathbf{s} \\
&= \frac{1}{|W_n|}\int_{W_n}\left|(\boldsymbol{\beta}-\beta_0)^T\boldsymbol{\psi}^{(1)}(\mathbf{s};\boldsymbol{\beta}^*)\right|\left\|\frac{\partial \boldsymbol{\phi}_{n,\boldsymbol{\theta}}(\mathbf{s};\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\right\|_{\max}\mathrm{d}\mathbf{s} \\
&= \frac{\|\boldsymbol{\beta}-\beta_0\|_1}{|W_n|}\int_{W_n}\left\|\boldsymbol{\psi}^{(1)}(\mathbf{s};\boldsymbol{\beta}^*)\right\|_{\max}\left\|\frac{\partial \boldsymbol{\phi}_{n,\boldsymbol{\theta}}(\mathbf{s};\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\right\|_{\max}\mathrm{d}\mathbf{s} \\
&= O(1)\|\boldsymbol{\beta}-\boldsymbol{\beta}_0\|_1,
\end{aligned}$$

where the last equality follows from conditions C1 and L1. Therefore, we have that for any $\|\boldsymbol{\beta}-\boldsymbol{\beta}_0\|_1 \le \frac{d}{\sqrt{|W_n|}}, \|\boldsymbol{\theta}-\boldsymbol{\theta}_{0,n}\|_1 \le d\alpha_n$,

$$\mathbf{J}_{\boldsymbol{\phi},n}(\boldsymbol{\beta},\boldsymbol{\theta}) = \mathbf{S}_{\boldsymbol{\phi},n}(\boldsymbol{\beta},\boldsymbol{\theta}) + O_p(|W_n|^{-1/2}). \tag{S.21}$$

Based on (S.21), we have that

$$\mathbf{J}_{\boldsymbol{\phi},n}(\boldsymbol{\beta},\boldsymbol{\theta}) - \mathbf{J}_{\boldsymbol{\phi},n}(\boldsymbol{\beta}_0,\boldsymbol{\theta}_{0,n}) = \mathbf{S}_{\boldsymbol{\phi},n}(\boldsymbol{\beta},\boldsymbol{\theta}) - \mathbf{S}_{\boldsymbol{\phi},n}(\boldsymbol{\beta}_0,\boldsymbol{\theta}_{0,n}) + O_p(|W_n|^{-1/2}),$$

which implies that to show R2, it suffices to show that

$$\sup_{\|\boldsymbol{\beta}-\boldsymbol{\beta}_0\|_1\le \frac{d}{\sqrt{|W_n|}}, \|\boldsymbol{\theta}-\boldsymbol{\theta}_{0,n}\|_1\le d\alpha_n} \|\mathbf{S}_{\boldsymbol{\phi},n}(\boldsymbol{\beta},\boldsymbol{\theta}) - \mathbf{S}_{\boldsymbol{\phi},n}(\boldsymbol{\beta}_0,\boldsymbol{\theta}_{0,n})\|_{\max} = o(1). \tag{S.22}$$

To show (S.22), using the Taylor expansion, we have that

$$\begin{aligned}
\|\mathbf{S}_{\boldsymbol{\phi},n}(\boldsymbol{\beta},\boldsymbol{\theta}) - \mathbf{S}_{\boldsymbol{\phi},n}(\boldsymbol{\beta}_0,\boldsymbol{\theta}_{0,n})\|_{\max} \le\ & \frac{1}{|W_n|}\int_{W_n} \left\| \left[\boldsymbol{\psi}^{(1)}(\mathrm{s};\boldsymbol{\beta}) - \boldsymbol{\psi}^{(1)}(\mathrm{s};\boldsymbol{\beta}_0)\right] \boldsymbol{\phi}_{n,\boldsymbol{\theta}}^T(\mathrm{s};\boldsymbol{\beta})\right\|_{\max} \mathrm{ds} \\
& + \frac{1}{|W_n|}\int_{W_n} \left\| \boldsymbol{\psi}^{(1)}(\mathrm{s};\boldsymbol{\beta}_0)\left[\boldsymbol{\phi}_{n,\boldsymbol{\theta}}(\mathrm{s};\boldsymbol{\beta}) - \boldsymbol{\phi}_{n,\boldsymbol{\theta}_{0,n}}(\mathrm{s};\boldsymbol{\beta}_0)\right]\right\|_{\max} \mathrm{ds} \\
\le\ & \frac{\|\boldsymbol{\beta}-\boldsymbol{\beta}_0\|_1}{|W_n|}\int_{W_n} \left\|\boldsymbol{\psi}^{(2)}(\mathrm{s};\boldsymbol{\beta}^*)\right\|_{\max} \left\|\boldsymbol{\phi}_{n,\boldsymbol{\theta}}(\mathrm{s};\boldsymbol{\beta})\right\|_{\max} \mathrm{ds} \\
& + \frac{1}{|W_n|}\int_{W_n} \left\|\boldsymbol{\psi}^{(1)}(\mathrm{s};\boldsymbol{\beta}_0)\right\|_{\max} \left\|\boldsymbol{\phi}_{n,\boldsymbol{\theta}}(\mathrm{s};\boldsymbol{\beta}) - \boldsymbol{\phi}_{n,\boldsymbol{\theta}_{0,n}}(\mathrm{s};\boldsymbol{\beta}_0)\right\|_{\max} \mathrm{ds}.
\end{aligned}$$

Therefore, under conditions C1, L1 and Lemma 3 (recall that under condition C6, $\alpha_n \to 0$ and thus conditions of Lemma 3 are satisfied), we have that

$$\|\mathbf{S}_{\boldsymbol{\phi},n}(\boldsymbol{\beta},\boldsymbol{\theta}) - \mathbf{S}_{\boldsymbol{\phi},n}(\boldsymbol{\beta}_0,\boldsymbol{\theta}_{0,n})\|_{\max} = O(\|\boldsymbol{\beta}-\boldsymbol{\beta}_0\|_1 + \|\boldsymbol{\theta}-\boldsymbol{\theta}_{0,n}\|_1), \tag{S.23}$$

which completes the proof of equation (S.22) and thus the proof of R2.

Proof of R3: Using the triangular inequality, we have that

$$\begin{aligned}
& \left\|\mathbf{J}_{\boldsymbol{\phi},n}(\boldsymbol{\beta}_0,\widehat{\boldsymbol{\theta}}_n) - \mathbf{S}_{\boldsymbol{\phi},n}(\boldsymbol{\beta}_0,\boldsymbol{\theta}_{0,n})\right\|_{\max} \\
& \le \left\|\mathbf{J}_{\boldsymbol{\phi},n}(\boldsymbol{\beta}_0,\widehat{\boldsymbol{\theta}}_n) - \mathbf{J}_{\boldsymbol{\phi},n}(\boldsymbol{\beta}_0,\boldsymbol{\theta}_{0,n})\right\|_{\max} + \|\mathbf{J}_{\boldsymbol{\phi},n}(\boldsymbol{\beta}_0,\boldsymbol{\theta}_{0,n}) - \mathbf{S}_{\boldsymbol{\phi},n}(\boldsymbol{\beta}_0,\boldsymbol{\theta}_{0,n})\|_{\max} \\
& = o_p(1) + O_p(|W_n|^{-1/2}) = o_p(1),
\end{aligned}$$

where the second last equality follows from R2 (by Theorem 1, $\|\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_{0,n}\|_1 = O_p(\alpha_n)$) and equation (S.21).

Proof of R4: To prove R4, it suffices to show that

$$\sup_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_{0,n}\|_1\le d\alpha_n} \|\mathbf{Q}_{\boldsymbol{\phi},n}(\boldsymbol{\beta}_0,\boldsymbol{\theta}) - \mathbf{Q}_{\boldsymbol{\phi},n}(\boldsymbol{\beta}_0,\boldsymbol{\theta}_{0,n})\|_{\max} = o_p(1). \tag{S.24}$$

To show (S.24), using the first order Taylor expansion of $\mathbf{Q}_{\boldsymbol{\phi},n}(\boldsymbol{\beta}_0,\boldsymbol{\theta})$ around $\boldsymbol{\theta} = \boldsymbol{\theta}_{0,n}$, we have that

$$\|\mathbf{Q}_{\boldsymbol{\phi},n}(\boldsymbol{\beta}_0,\boldsymbol{\theta}) - \mathbf{Q}_{\boldsymbol{\phi},n}(\boldsymbol{\beta}_0,\boldsymbol{\theta}_{0,n})\|_{\max} = \|\mathbf{H}_{\boldsymbol{\phi},n}(\boldsymbol{\theta}^*)(\boldsymbol{\theta}-\boldsymbol{\theta}_{0,n})\|_{\max} \le \|\mathbf{H}_{\boldsymbol{\phi},n}(\boldsymbol{\theta}^*)\|_{\max}\|\boldsymbol{\theta}-\boldsymbol{\theta}_{0,n}\|_1,$$

where $\mathbf{H}_{\boldsymbol{\phi},n}(\boldsymbol{\theta})$ is as defined in condition L2 and $\|\boldsymbol{\theta}^* - \boldsymbol{\theta}_{0,n}\|_1 \leq \|\boldsymbol{\theta} - \boldsymbol{\theta}_{0,n}\|_1$. Under condition L2 and using Lemma 1, we have that $\|\mathbf{H}_{\boldsymbol{\phi},n}(\boldsymbol{\theta}^*)\|_{\max} = O_p(\log J_n)$, which further improves that

$$\sup_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_{0,n}\|_1 \leq d\alpha_n} \|\mathbf{Q}_{\boldsymbol{\phi},n}(\boldsymbol{\beta}_0, \boldsymbol{\theta}) - \mathbf{Q}_{\boldsymbol{\phi},n}(\boldsymbol{\beta}_0, \boldsymbol{\theta}_{0,n})\|_{\max} = O_p(\alpha_n \log J_n) = o_p(1),$$

which completes the proof of (S.24).

Proof of R5: As $|W_n| \to \infty$, R5 follows directly from condition L3 and equation (S.20).

Therefore, we have shown that $\sqrt{|W_n|}(\widehat{\boldsymbol{\beta}}_{\widehat{\boldsymbol{\phi}},n} - \boldsymbol{\beta}_0) = O_p(1)$. Then using the Taylor expansion, we have that

$$\mathbf{Q}_{\widehat{\boldsymbol{\phi}},n}\left(\widehat{\boldsymbol{\beta}}_{\widehat{\boldsymbol{\phi}},n}, \widehat{\boldsymbol{\theta}}_n\right) - \mathbf{Q}_{\boldsymbol{\phi},n}\left(\boldsymbol{\beta}_0, \widehat{\boldsymbol{\theta}}_n\right) = -\mathbf{J}_{\boldsymbol{\phi},n}\left(\boldsymbol{\beta}^*, \widehat{\boldsymbol{\theta}}_n\right)\sqrt{|W_n|}\left(\widehat{\boldsymbol{\beta}}_{\widehat{\boldsymbol{\phi}},n} - \boldsymbol{\beta}_0\right),$$

where $\|\boldsymbol{\beta}^* - \boldsymbol{\beta}_0\|_1 \leq \|\widehat{\boldsymbol{\beta}}_{\widehat{\boldsymbol{\phi}},n} - \boldsymbol{\beta}_0\|_1$, and $\mathbf{J}_{\boldsymbol{\phi},n}(\cdot,\cdot)$ is as defined in (S.18). Consequently, we have that

$$\sqrt{|W_n|}\left(\widehat{\boldsymbol{\beta}}_{\widehat{\boldsymbol{\phi}},n} - \boldsymbol{\beta}_0\right) = \mathbf{J}^{-1}_{\boldsymbol{\phi},n}\left(\boldsymbol{\beta}^*, \widehat{\boldsymbol{\theta}}_n\right)\mathbf{Q}_{\boldsymbol{\phi},n}\left(\boldsymbol{\beta}_0, \widehat{\boldsymbol{\theta}}_n\right) = \left[\mathbf{S}_{\boldsymbol{\phi},n}(\boldsymbol{\beta}_0, \boldsymbol{\theta}_{0,n})\right]^{-1}\mathbf{Q}_{\boldsymbol{\phi},n}\left(\boldsymbol{\beta}_0, \boldsymbol{\theta}_{0,n}\right) + o_p(1),$$

where the last equality follows from R2, R3 and R4. The proof of Theorem 2 is complete.

S.6 Proof of Theorem 3

S.6.1 Notations and Conditions

For the asymptotic distribution of the test statistic, we introduce the definition of α-mixing coefficients for point processes to quantify the strength of spatial dependence. For two σ-algebras $\mathcal{F}$ and $\mathcal{G}$ define $\alpha\left[\mathcal{F}, \mathcal{G}\right] = \sup\left\{|P(F \cap G) - P(F)P(G)| : F \in \mathcal{F}, G \in \mathcal{G}\right\}$. For any $s, c_1, c_2 \geq 0$, the α-mixing coefficients of a point process N are defined as $\alpha_N(s; c_1, c_2) = \sup\{\alpha\left[\sigma(N \cap E_1), \sigma(N \cap E_2)\right] : B_k \subset \mathbb{R}^2, |B_k| \leq c_k, k = 1, 2,\ \mathrm{d}(B_1, B_2) \geq s\}$, where $\mathrm{d}(B_1, B_2) = \inf\{\max_{1 \leq i \leq 2} |s_i - t_i| : \mathbf{s} = (s_1, s_2)^T \in B_1, \mathbf{t} = (t_1, t_2)^T \in B_2\}$.

The following additional conditions are needed for asymptotic distributions.

[N1] The mixing coefficient of N, $\alpha_N(r; 2, \infty) = O(r^{-2-\epsilon})$ for some $\epsilon > 0$.

[N2] There exist constants C_g and $\delta > 4/\epsilon$ such that $\left|g^{(k)}(\mathbf{s}_1, \mathbf{s}_2, \cdots, \mathbf{s}_k)\right| \leq C_g$ for any $\mathbf{s}_j \in W_n$, $j = 1, \cdots, k$ and $k = 2, \cdots, 2(2 + \lceil\delta\rceil)$, with $\lceil\delta\rceil$ being the smallest integer greater than δ.

These two conditions are standard conditions, see, e.g., Prokešová and Jensen (2013); Prokešová et al. (2017), and impose no rigid restrictions. We also need the following technical conditions.

[L4] Assume that $\boldsymbol{f}_{n,\boldsymbol{\theta}}(\mathbf{s};\boldsymbol{\beta})$ is differentiable with respect to $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$, and that there exists some constant $K_1 > 0$ such that (a) $\sup_{\mathbf{s}\in W_n} \|\boldsymbol{f}_{n,\boldsymbol{\theta}}(\mathbf{s};\boldsymbol{\beta})\|_{\max}$, $\sup_{\mathbf{s}\in W_n} \|\frac{\partial \boldsymbol{f}_{n,\boldsymbol{\theta}}(\mathbf{s};\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\|_{\max}$, $\sup_{\mathbf{s}\in W_n} \|\frac{\partial \boldsymbol{f}_{n,\boldsymbol{\theta}}(\mathbf{s};\boldsymbol{\beta})}{\partial \boldsymbol{\theta}}\|_{\max}$ are uniformly bounded for $(\boldsymbol{\beta},\boldsymbol{\theta})$ satisfying $\|\boldsymbol{\beta}-\boldsymbol{\beta}_0\|_1 \leq K_1$, $\|\boldsymbol{\theta}-\boldsymbol{\theta}_{0,n}\|_1 \leq K_1$; (b) $\sup_{\mathbf{s}\in W_n} \left\|\boldsymbol{f}_{n,\boldsymbol{\theta}}(\mathbf{s};\boldsymbol{\beta}) - \boldsymbol{f}_{n,\boldsymbol{\theta}_{0,n}}(\mathbf{s};\boldsymbol{\beta}_0)\right\|_{\max} = O(\|\boldsymbol{\beta}-\boldsymbol{\beta}_0\|_1 + \|\boldsymbol{\theta}-\boldsymbol{\theta}_{0,n}\|_1)$ if $\|\boldsymbol{\beta}-\boldsymbol{\beta}_0\|_1 \to 0$ and $\|\boldsymbol{\theta}-\boldsymbol{\theta}_{0,n}\|_1 \to 0$ as $|W_n| \to \infty$.

[L5] Denote matrix $\mathbf{H}_{\boldsymbol{f},n}(\boldsymbol{\theta}) = \frac{1}{\sqrt{|W_n|}}\left[\sum_{\mathbf{s}\in N\cap W_n} \frac{\partial \boldsymbol{f}_{n,\boldsymbol{\theta}}(\mathbf{s};\boldsymbol{\beta}_0)}{\partial \boldsymbol{\theta}} - \int_{W_n} \psi(\mathbf{s};\boldsymbol{\beta}_0)\frac{\partial \boldsymbol{f}_{n,\boldsymbol{\theta}}(\mathbf{s};\boldsymbol{\beta}_0)}{\partial \boldsymbol{\theta}}\mathrm{d}\mathbf{s}\right]$ and its ijth element as $h_{ij,\boldsymbol{f},n}(\boldsymbol{\theta})$. Assume that there exists a constant $K_2 > 0$ and n_0 such that for any $n \geq n_0$, it holds that for any $\boldsymbol{\theta}$ satisfying $\|\boldsymbol{\theta}-\boldsymbol{\theta}_{0,n}\|_1 \leq K_1$, $P\left(|h_{ij,\boldsymbol{f},n}(\boldsymbol{\theta})| > t\right) \leq \exp\left(-K_2 t\right)$, for any $t > t_0$ with $t_0 > 0$ being some constant and $i = 1, \cdots, p$, $j = 1, \cdots, J_n$.

[L6] Let $df = \mathrm{rank}\left[\boldsymbol{\Sigma}_{\boldsymbol{f},n}(\boldsymbol{\beta}_0, g_0)\right]$ and there exists a $df \times q$ matrix $\mathbf{B}(\boldsymbol{\beta}_0, g_0)$ such that $\|\mathbf{B}(\boldsymbol{\beta}_0, g_0)\|_{\max} < \infty$ and $\mathbf{B}(\boldsymbol{\beta}_0, g_0)\boldsymbol{\Sigma}_{\boldsymbol{f},n}(\boldsymbol{\beta}_0, g_0)\mathbf{B}^T(\boldsymbol{\beta}_0, g_0) = \mathbf{I}_{df}$.

Conditions L4(a) and L5 are identical to conditions L1-L2, with $\boldsymbol{\phi}_{n,\boldsymbol{\theta}}(\cdot;\boldsymbol{\beta})$ replaced by $\boldsymbol{f}_{n,\boldsymbol{\theta}}(\cdot;\boldsymbol{\beta})$. Condition L4(b) is similar to the Lemma 3, and is rather mild based on the definition of $\boldsymbol{f}^{\mathrm{L}}_{n,\boldsymbol{\theta}}(\cdot;\boldsymbol{\beta})$ and $\boldsymbol{f}^{\mathrm{G}}_{n,\boldsymbol{\theta}}(\cdot;\boldsymbol{\beta})$.

S.6.2 Proof of Theorem 3

Using the first-order stochastic Taylor expansion, one has that

$$\mathbf{Q}_{\boldsymbol{f},n}(\widehat{\boldsymbol{\beta}}_{\widehat{\boldsymbol{\phi}},n}, \widehat{\boldsymbol{\theta}}_n) - \mathbf{Q}_{\boldsymbol{f},n}(\boldsymbol{\beta}_0, \widehat{\boldsymbol{\theta}}_n) = -\mathbf{J}_{\boldsymbol{f},n}(\boldsymbol{\beta}^*, \widehat{\boldsymbol{\theta}}_n)\sqrt{|W_n|}\left(\widehat{\boldsymbol{\beta}}_{\widehat{\boldsymbol{\phi}},n} - \boldsymbol{\beta}_0\right),$$

where $\mathbf{J}_{\boldsymbol{f},n}(\boldsymbol{\beta},\widehat{\boldsymbol{\theta}})$ is as defined in (S.18). Using the same arguments in the proof of R2 and R3 of Theorem 2, one can show that under the condition L4,

$$\sup_{\|\boldsymbol{\beta}-\boldsymbol{\beta}_0\|_1\le\frac{d}{\sqrt{|W_n|}},\|\boldsymbol{\theta}-\boldsymbol{\theta}_{0,n}\|_1\le d\alpha_n}\|\mathbf{J}_{\boldsymbol{f},n}(\boldsymbol{\beta},\boldsymbol{\theta})-\mathbf{S}_{\boldsymbol{f},n}(\boldsymbol{\beta}_0,\boldsymbol{\theta}_{0,n})\|_{\max}=o_p(1),$$

where $\alpha_n=\eta_n^{-1}(e_n+|W_n|^{-1/2}\log J_n)$ and some constant $d>0$. Using Theorem 1 and the above equality, it immediately follows that

$$\mathbf{Q}_{\boldsymbol{f},n}(\widehat{\boldsymbol{\beta}}_{\widehat{\boldsymbol{\phi}},n},\widehat{\boldsymbol{\theta}}_n)=\mathbf{Q}_{\boldsymbol{f},n}(\boldsymbol{\beta}_0,\widehat{\boldsymbol{\theta}}_n)-\mathbf{S}_{\boldsymbol{f},n}(\boldsymbol{\beta}_0,\boldsymbol{\theta}_{0,n})\sqrt{|W_n|}\left(\widehat{\boldsymbol{\beta}}_{\widehat{\boldsymbol{\phi}},n}-\boldsymbol{\beta}_0\right)+o_p(1).$$

Then using the same arguments in the proof of R4 of Theorem 2, under conditions L4-L5, one has that $\left\|\mathbf{Q}_{\boldsymbol{f},n}(\boldsymbol{\beta}_0,\widehat{\boldsymbol{\theta}}_n)-\mathbf{Q}_{\boldsymbol{f},n}(\boldsymbol{\beta}_0,\boldsymbol{\theta}_{0,n})\right\|_{\max}=o_p(1)$, which further implies that

$$\mathbf{Q}_{\boldsymbol{f},n}(\widehat{\boldsymbol{\beta}}_{\widehat{\boldsymbol{\phi}},n},\widehat{\boldsymbol{\theta}}_n)=\mathbf{Q}_{\boldsymbol{f},n}(\boldsymbol{\beta}_0,\boldsymbol{\theta}_{0,n})-\mathbf{S}_{\boldsymbol{f},n}(\boldsymbol{\beta}_0,\boldsymbol{\theta}_{0,n})\sqrt{|W_n|}\left(\widehat{\boldsymbol{\beta}}_{\widehat{\boldsymbol{\phi}},n}-\boldsymbol{\beta}_0\right)+o_p(1).$$

Then using Theorem 2, it immediately follows that

$$\mathbf{Q}_{\boldsymbol{f},n}(\widehat{\boldsymbol{\beta}}_{\widehat{\boldsymbol{\phi}},n},\widehat{\boldsymbol{\theta}}_n)=\mathbf{Q}_{\boldsymbol{f},n}(\boldsymbol{\beta}_0,\boldsymbol{\theta}_{0,n})-\mathbf{S}^T_{\boldsymbol{f},n}(\boldsymbol{\beta}_0,\boldsymbol{\theta}_{0,n})\left[\mathbf{S}_{\boldsymbol{\phi},n}(\boldsymbol{\beta}_0,\boldsymbol{\theta}_{0,n})\right]^{-1}\mathbf{Q}_{\boldsymbol{\phi},n}\left(\boldsymbol{\beta}_0,\boldsymbol{\theta}_{0,n}\right)+o_p(1). \quad \text{(S.25)}$$

Using (S.25), some straightforward algebra yields that

$$\begin{aligned}
&\mathrm{Var}\left[\mathbf{Q}_{\boldsymbol{f},n}(\widehat{\boldsymbol{\beta}}_{\widehat{\boldsymbol{\phi}},n},\widehat{\boldsymbol{\theta}}_n)\right]\\
&=\mathrm{Var}\left[\mathbf{Q}_{\boldsymbol{f},n}(\boldsymbol{\beta}_0,\boldsymbol{\theta}_{0,n})\right]-\mathbf{S}^T_{\boldsymbol{f},n}(\boldsymbol{\beta}_0,\boldsymbol{\theta}_{0,n})\left[\mathbf{S}_{\boldsymbol{\phi},n}(\boldsymbol{\beta}_0,\boldsymbol{\theta}_{0,n})\right]^{-1}\mathrm{Cov}\left[\mathbf{Q}_{\boldsymbol{\phi},n}(\boldsymbol{\beta}_0,\boldsymbol{\theta}_{0,n}),\mathbf{Q}_{\boldsymbol{f},n}(\boldsymbol{\beta}_0,\boldsymbol{\theta}_{0,n})\right]\\
&\quad-\mathrm{Cov}\left[\mathbf{Q}_{\boldsymbol{f},n}(\boldsymbol{\beta}_0,\boldsymbol{\theta}_{0,n}),\mathbf{Q}_{\boldsymbol{\phi},n}(\boldsymbol{\beta}_0,\boldsymbol{\theta}_{0,n})\right]\left[\mathbf{S}_{\boldsymbol{\phi},n}(\boldsymbol{\beta}_0,\boldsymbol{\theta}_{0,n})\right]^{-1}\mathbf{S}_{\boldsymbol{f},n}(\boldsymbol{\beta}_0,\boldsymbol{\theta}_{0,n})\\
&\quad+\mathbf{S}^T_{\boldsymbol{f},n}(\boldsymbol{\beta}_0,\boldsymbol{\theta}_{0,n})\left[\mathbf{S}_{\boldsymbol{\phi},n}(\boldsymbol{\beta}_0,\boldsymbol{\theta}_{0,n})\right]^{-1}\mathrm{Var}\left[\mathbf{Q}_{\boldsymbol{\phi},n}\left(\boldsymbol{\beta}_0,\boldsymbol{\theta}_{0,n}\right)\right]\left[\mathbf{S}_{\boldsymbol{\phi},n}(\boldsymbol{\beta}_0,\boldsymbol{\theta}_{0,n})\right]^{-1}\mathbf{S}_{\boldsymbol{f},n}(\boldsymbol{\beta}_0,\boldsymbol{\theta}_{0,n}).
\end{aligned}$$

$$\begin{aligned}
&\mathrm{Cov}\left[\mathbf{Q}_{\boldsymbol{\phi},n}(\boldsymbol{\beta}_0,\boldsymbol{\theta}_{0,n}),\mathbf{Q}_{\boldsymbol{f},n}(\boldsymbol{\beta}_0,\boldsymbol{\theta}_{0,n})\right]=\frac{1}{|W_n|}\int_{W_n}\psi(\mathbf{s};\boldsymbol{\beta}_0)\boldsymbol{\phi}_{n,\boldsymbol{\theta}_{0,n}}(\mathbf{s};\boldsymbol{\beta}_0)\boldsymbol{f}^T_{n,\boldsymbol{\theta}_{0,n}}(\mathbf{s};\boldsymbol{\beta}_0)\\
&\quad+\frac{1}{|W_n|}\int_{W_n}\int_{W_n}\psi(\mathbf{s};\boldsymbol{\beta}_0)\psi(\mathbf{t};\boldsymbol{\beta}_0)\left[g_0(\|\mathbf{s}-\mathbf{t}\|)-1\right]\boldsymbol{\phi}_{n,\boldsymbol{\theta}_{0,n}}(\mathbf{s};\boldsymbol{\beta}_0)\boldsymbol{f}^T_{n,\boldsymbol{\theta}_{0,n}}(\mathbf{s};\boldsymbol{\beta}_0)\mathrm{d}\mathbf{s}\mathrm{d}\mathbf{t}\\
&\quad=\frac{1}{|W_n|}\int_{W_n}\psi(\mathbf{s};\boldsymbol{\beta}_0)\boldsymbol{\phi}_{n,\boldsymbol{\theta}_{0,n}}(\mathbf{s};\boldsymbol{\beta}_0)\boldsymbol{f}^T_{n,\boldsymbol{\theta}_{0,n}}(\mathbf{s};\boldsymbol{\beta}_0)\\
&\quad+\frac{1}{|W_n|}\int_{W_n}\int_{W_n}\psi(\mathbf{s};\boldsymbol{\beta}_0)\psi(\mathbf{t};\boldsymbol{\beta}_0)\left[g_{\mathrm{s},n}(\|\mathbf{s}-\mathbf{t}\|;\boldsymbol{\theta}_{0,n})-1\right]\boldsymbol{\phi}_{n,\boldsymbol{\theta}_{0,n}}(\mathbf{s};\boldsymbol{\beta}_0)\boldsymbol{f}^T_{n,\boldsymbol{\theta}_{0,n}}(\mathbf{s};\boldsymbol{\beta}_0)\mathrm{d}\mathbf{s}\mathrm{d}\mathbf{t}+o(1)\\
&\quad=\mathbf{S}_{\boldsymbol{f},n}(\boldsymbol{\beta}_0,\boldsymbol{\theta}_{0,n})+o(1),
\end{aligned}$$

(S.26)

where the last equality follows from the definition of $\boldsymbol{\phi}_{n,\boldsymbol{\theta}_{0,n}}(\mathbf{s};\boldsymbol{\beta}_0)$ in the integral equation (3). Note that (S.26) also indicates that $\mathrm{Var}\left[\mathbf{Q}_{\boldsymbol{\phi},n}\left(\boldsymbol{\beta}_0,\boldsymbol{\theta}_{0,n}\right)\right]=\mathbf{S}_{\boldsymbol{\phi},n}(\boldsymbol{\beta}_0,\boldsymbol{\theta}_{0,n})+o(1)$. Consequently, using (S.26), the $\mathrm{Var}\left[\mathbf{Q}_{\boldsymbol{f},n}(\widehat{\boldsymbol{\beta}}_{\widehat{\boldsymbol{\phi}},n},\widehat{\boldsymbol{\theta}}_n)\right]$ can be simplified as

$$\mathrm{Var}\left[\mathbf{Q}_{\boldsymbol{f},n}(\widehat{\boldsymbol{\beta}}_{\widehat{\boldsymbol{\phi}},n},\widehat{\boldsymbol{\theta}}_n)\right]=\mathrm{Var}\left[\mathbf{Q}_{\boldsymbol{f},n}(\boldsymbol{\beta}_0,\boldsymbol{\theta}_{0,n})\right]-\mathbf{S}^T_{\boldsymbol{f},n}(\boldsymbol{\beta}_0,\boldsymbol{\theta}_{0,n})\left[\mathbf{S}_{\boldsymbol{\phi},n}(\boldsymbol{\beta}_0,\boldsymbol{\theta}_{0,n})\right]^{-1}\mathbf{S}_{\boldsymbol{f},n}(\boldsymbol{\beta}_0,\boldsymbol{\theta}_{0,n})+o(1).$$

Similarly, since $\sup_{r\in[0,R]}|g_0(r)-g_{\mathrm{s},n}(r;\boldsymbol{\theta}_{0,n})|$=o(1) by Theorem 2, under condition L4 and Lemma 3, it is straightforward to show that,

$$\mathrm{Var}\left[\mathbf{Q}_{\boldsymbol{f},n}(\boldsymbol{\beta}_0,\boldsymbol{\theta}_{0,n})\right]=\mathrm{Var}\left[\mathbf{Q}_{\boldsymbol{f}_{g_0}}(\boldsymbol{\beta}_0)\right]+o(1),\quad \mathbf{S}_{\boldsymbol{f},n}(\boldsymbol{\beta}_0,\boldsymbol{\theta}_{0,n})=\mathbf{S}_{\boldsymbol{f},n}(\boldsymbol{\beta}_0,g_0)+o(1),$$

$$\text{and }\mathbf{S}_{\boldsymbol{\phi},n}(\boldsymbol{\beta}_0,\boldsymbol{\theta}_{0,n})=\mathbf{S}_{\boldsymbol{\phi},n}(\boldsymbol{\beta}_0,g_0)+o(1),y$$

which immediately suggests that

$$\mathrm{Var}\left[\mathbf{Q}_{\boldsymbol{f},n}(\widehat{\boldsymbol{\beta}}_{\widehat{\boldsymbol{\phi}},n},\widehat{\boldsymbol{\theta}}_n)\right]=\boldsymbol{\Sigma}_{\boldsymbol{f},n}(\boldsymbol{\beta}_0,g_0)+o(1), \tag{S.27}$$

where $\boldsymbol{\Sigma}_{\boldsymbol{f},n}(\boldsymbol{\beta}_0,g_0)=\mathrm{Var}\left[\mathbf{Q}_{\boldsymbol{f}_{g_0}}(\boldsymbol{\beta}_0)\right]-\mathbf{S}^T_{\boldsymbol{f},n}(\boldsymbol{\beta}_0,g_0)\mathbf{S}_{\boldsymbol{\phi},n}(\boldsymbol{\beta}_0,g_0)\mathbf{S}_{\boldsymbol{f},n}(\boldsymbol{\beta}_0,g_0)$.

To prove Theorem 3, it suffices to show that

$$\mathbf{B}(\boldsymbol{\beta}_0,g_0)\mathbf{Q}_{\boldsymbol{f},n}(\widehat{\boldsymbol{\beta}}_{\widehat{\boldsymbol{\phi}},n},\widehat{\boldsymbol{\theta}}_n)\xrightarrow{d}N(\mathbf{0},\mathbf{I}_{df}),\ \text{as } |W_n|\to\infty, \tag{S.28}$$

where $\mathbf{B}(\boldsymbol{\beta}_0,g_0)$ is defined in condition L6.

Using approximation (S.25), it is straightforward to show that

$$\mathbf{B}(\boldsymbol{\beta}_0,g_0)\mathbf{Q}_{\boldsymbol{f},n}(\widehat{\boldsymbol{\beta}}_{\widehat{\boldsymbol{\phi}},n},\widehat{\boldsymbol{\theta}}_n)=\mathbf{Q}_{\boldsymbol{f}^*,n}(\boldsymbol{\beta}_0,\boldsymbol{\theta}_{0,n})+o_p(1),$$

where $f^*_{n,\boldsymbol{\theta}_{0,n}}(\mathbf{s};\boldsymbol{\beta}_0)=\mathbf{B}(\boldsymbol{\beta}_0,g_0)\boldsymbol{f}_{n,\boldsymbol{\theta}_{0,n}}(\mathbf{s};\boldsymbol{\beta}_0)-\mathbf{B}(\boldsymbol{\beta}_0,g_0)\mathbf{S}^T_{\boldsymbol{f},n}(\boldsymbol{\beta}_0,\boldsymbol{\theta}_{0,n})\left[\mathbf{S}_{\boldsymbol{\phi},n}(\boldsymbol{\beta}_0,\boldsymbol{\theta}_{0,n})\right]^{-1}\boldsymbol{\phi}_{n,\boldsymbol{\theta}_{0,n}}(\mathbf{s};\boldsymbol{\beta}_0)$.

By the definition of $\mathbf{B}(\boldsymbol{\beta}_0,g_0)$ in condition L6, we have that $\mathrm{Var}\left[\mathbf{Q}_{\boldsymbol{f}^*,n}(\boldsymbol{\beta}_0,\boldsymbol{\theta}_{0,n})\right]=\mathbf{I}_{df}$. Combining this fact with conditions N1-N2, (S.28) follows from Theorem 1 of Biscio and Waagepetersen (2019). Consequently, one has that

$$\mathbf{Q}^T_{\boldsymbol{f},n}(\widehat{\boldsymbol{\beta}}_{\widehat{\boldsymbol{\phi}},n},\widehat{\boldsymbol{\theta}}_n)\mathbf{B}^T(\boldsymbol{\beta}_0,g_0)\mathbf{B}(\boldsymbol{\beta}_0,g_0)\mathbf{Q}_{\boldsymbol{f},n}(\widehat{\boldsymbol{\beta}}_{\widehat{\boldsymbol{\phi}},n},\widehat{\boldsymbol{\theta}}_n)\to\chi^2(df).$$

Note that $\mathbf{B}^T(\boldsymbol{\beta}_0, g_0)\mathbf{B}(\boldsymbol{\beta}_0, g_0)$ is a version of generalized inverse of $\boldsymbol{\Sigma}(\boldsymbol{\beta}_0, g_0)$ and consequently Theorem 3 follows from the Slutsky's Theorem.

References

Biscio, C. A. N. and Waagepetersen, R. (2019), "A general central limit theorem and a subsampling variance estimator for α-mixing point processes," *Scandinavian Journal of Statistics*, 46, 1168–1190.

Guan, Y., Jalilian, A., and Waagepetersen, R. (2015), "Quasi-likelihood for spatial point processes," *Journal of the Royal Statistical Society: Series B*, 77, 677–697.

Meinshausen, N. et al. (2013), "Sign-constrained least squares estimation for high-dimensional regression," *Electronic Journal of Statistics*, 7, 1607–1631.

Prokešová, M., Dvořák, J., and Jensen, E. B. V. (2017), "Two-step estimation procedures for inhomogeneous shot-noise Cox processes," *Annals of the Institute of Statistical Mathematics*, 69, 513–542.

Prokešová, M. and Jensen, E. B. V. (2013), "Asymptotic Palm likelihood theory for stationary point processes," *Annals of the Institute of Statistical Mathematics*, 65, 387–412.

Schoenberg, F. P. (2005), "Consistent parametric estimation of the intensity of a spatial–temporal point process," *Journal of Statistical Planning and Inference*, 128, 79–93.

Slawski, M., Hein, M., et al. (2013), "Non-negative least squares for high-dimensional linear models: Consistency and sparse recovery without regularization," *Electronic Journal of Statistics*, 7, 3004–3056.

Waagepetersen, R. and Guan, Y. (2009), "Two-step estimation for inhomogeneous spatial point processes," *Journal of the Royal Statistical Society: Series B*, 71, 685–702.