Supplement to "Pairwise Estimation of Multivariate Gaussian Process Models With Replicated Observations: Application to Multivariate Profile Monitoring"

Yongxiang Li, Qiang Zhou*, Xiaohu Huang, Li Zeng

*corresponding author email: q.zhou@arizona.edu

A. PROPOSITIONS

Proposition 1. Under the usual regularity conditions, as n → ∞, estimates of the augmented parameters in $\mathbf{\Theta}^*$ are consistent, i.e., $\hat{\mathbf{\Theta}}^* \xrightarrow{p} \mathbf{\Theta}^*$, and

$$\sqrt{n}\left(\hat{\mathbf{\Theta}}^* - \mathbf{\Theta}^*\right) \sim N\left(\mathbf{0}, \mathbf{H}^{-1}\mathbf{J}\left(\mathbf{H}^{-1}\right)^T\right) \tag{A.1}$$

where $\mathbf{H}$ is a lower triangular block matrix with blocks $\mathbf{H}_{p,q}$ and $\mathbf{J}$ is a symmetric matrix containing blocks $\mathbf{J}_{p,q}$, given by

$$\begin{cases} \mathbf{H}_{[p,q]} = E\left[-\partial U_{[p]}\left(\Theta_{[p]};\mathbf{y}_{[p]} \mid \Theta_{[1]},\cdots,\Theta_{[p-1]}\right)\big/\partial\Theta_{[q]}^T\right], \ \ for \ \ q \le p \\ \mathbf{J}_{[p,q]} = E\left[U_{[p]}\left(\Theta_{[p]};\mathbf{y}_{[p]} \mid \Theta_{[1]},\cdots,\Theta_{[p-1]}\right)U_{[q]}^T\left(\Theta_{[q]};\mathbf{y}_{[q]} \mid \Theta_{[1]},\cdots,\Theta_{[q-1]}\right)\right] \end{cases} \tag{A.2}$$

or equivalently

$$\begin{cases} \mathbf{J}_{[p,q]} \quad (p \ne q) & = \begin{bmatrix} \mathbf{F}_{[p]}^T\mathbf{C}_{[p]}^{-1}\mathbf{C}_{[p,q]}\mathbf{C}_{[q]}^{-1}\mathbf{F}_{[q]} & \mathbf{0} \\ \mathbf{0} & \dfrac{1}{2}tr\left(\mathbf{C}_{[q]}^{-1}\mathbf{C}_{[p,q]}^T\mathbf{C}_{[p]}^{-1}\dfrac{\partial\mathbf{C}_{[p]}}{\partial\mathbf{\Psi}_{[p]}}\mathbf{C}_{[p]}^{-1}\mathbf{C}_{[p,q]}\mathbf{C}_{[q]}^{-1}\dfrac{\partial\mathbf{C}_{[q]}}{\partial\mathbf{\Psi}_{[q]}^T}\right) \end{bmatrix} \\ \mathbf{H}_{[p,p]} = \mathbf{J}_{[p,p]} & = \begin{bmatrix} \mathbf{F}_{[p]}^T\mathbf{C}_{[p]}^{-1}\mathbf{F}_{[p]} & \mathbf{0} \\ \mathbf{0} & \dfrac{1}{2}tr\left(\mathbf{C}_{[p]}^{-1}\dfrac{\partial\mathbf{C}_{[p]}}{\partial\mathbf{\Psi}_{[p]}}\mathbf{C}_{[p]}^{-1}\dfrac{\partial\mathbf{C}_{[p]}}{\partial\mathbf{\Psi}_{[p]}^T}\right) \end{bmatrix} \\ \mathbf{H}_{[p,q]} \quad (p > q) & = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \dfrac{1}{2}tr\left(\mathbf{C}_{[p]}^{-1}\dfrac{\partial\mathbf{C}_{[p]}}{\partial\mathbf{\Psi}_{[p]}}\mathbf{C}_{[p]}^{-1}\dfrac{\partial\mathbf{C}_{[p]}}{\partial\omega_{[q]}}\right) \end{bmatrix} \end{cases} \tag{A.3}$$

Proof: Clearly, if $s_p = 1$ and $r_p = 2, 3,\ldots, k$ (this corresponds to $p = 1, 2, \ldots, k - 1$), $\hat{\Theta}_{[p]}$ is a consistent estimator because it falls into a regular maximum likelihood estimation framework. Next we prove that, for $s_p > 1$ (i.e., $p \ge k$), the estimator $\hat{\Theta}_{[p]}$ obtained by maximizing the conditional likelihood $L_{[p]}\left(\Theta_{[p]};\mathbf{\Upsilon}_{[p]}\middle|\hat{\Theta}_{[1]},\cdots,\hat{\Theta}_{[p-1]}\right)$ is consistent, given the consistency of all $\hat{\Theta}_{[i]}$'s with $i < p$.

Since $\hat{\Theta}_{[i]} \xrightarrow{p} \Theta_{[i]}$ with $i < p$, then $\hat{v}_{[p]} \xrightarrow{p} v_{[p]}$. Recall that $T_{[p]} = \zeta_{[p]}(u_{[p]})$ and $u_{[p]} = v_{[p]} \cup \{\omega_{[p]}\}$. Define another function $\omega_{[p]} = \tilde{\zeta}_{[p]}\left(T_{[p]}, v_{[p]}\right)$. With fixed $v_{[p]}$, $\zeta_{[p]}\left(u_{[p]}\right) = \zeta_{[p]}\left(v_{[p]}, \omega_{[p]}\right)$ is monotonically decreasing with $\omega_{[p]}$ based on relationships from the HD, i.e., $\zeta_{[p]}\left(v_{[p]}, \pi\right) < \zeta_{[p]}\left(v_{[p]}, \omega_{[p]}\right) = T_{[p]} < \zeta_{[p]}\left(v_{[p]}, 0\right)$.

Define the following events

$$A = \left\{T_{[p]} \in \left[\zeta_{[p]}\left(\hat{v}^{(n)}_{[p]}, \pi\right),\ \zeta_{[p]}\left(\hat{v}^{(n)}_{[p]}, 0\right)\right]\right\},$$

$$A^c = A_1^c \cup A_2^c = \{T_{[p]} < \zeta_{[p]}\left(\hat{v}^{(n)}_{[p]}, \pi\right)\} \cup \{T_{[p]} > \zeta_{[p]}\left(\hat{v}^{(n)}_{[p]}, 0\right)\},$$

$$B_1 = \left\{\zeta_{[p]}\left(\hat{v}^{(n)}_{[p]}, \pi\right) - \zeta_{[p]}\left(v_{[p]}, \pi\right) > a\right\},$$

$$B_2 = \left\{\zeta_{[p]}\left(v_{[p]}, 0\right) - \zeta_{[p]}\left(\hat{v}^{(n)}_{[p]}, 0\right) > a\right\},$$

$$C = \left\{\left|\hat{T}^{(n)}_{[p]} - T_{[p]}\right| > a\right\}.$$

where $\hat{v}^{(n)}_{[p]} / \hat{T}^{(n)}_{[p]}$ denotes the estimate is based on a sample of size n.

Since $\hat{v}_{[p]} \xrightarrow{p} v_{[p]}$ and $\zeta_{[p]}$ is a continuous function, then for any a satisfying $0 < a < M$, where $M = \min\left(T_{[p]} - \zeta_{[p]}\left(v_{[p]}, \pi\right),\ \zeta_{[p]}\left(v_{[p]}, 0\right) - T_{[p]}\right)$, and any $\varepsilon > 0$, there must exist a common $N_1(\varepsilon) > 0$, such that for any sample size $n_1 > N_1(\varepsilon)$,

$$\Pr\left(B_1\right) < \varepsilon / 4 \ \text{ and } \ \Pr\left(B_2\right) < \varepsilon / 4 .$$

Since $A_1^c \subset B_1$ and $A_2^c \subset B_2$, we have $A^c \subset \{B_1 \cup B_2\}$. In addition with $T_{[p]} = \zeta_{[p]}\left(v_{[p]}, \omega_{[p]}\right)$, we have $\Pr(A^c) < \varepsilon/2$ for any sample size $n_1 > N_1(\varepsilon)$. Since

$$P(C) = P(C \mid A)P(A) + P(C \mid A^c)P(A^c),$$

it immediately follows that $\Pr(C) < \Pr(C|A) + \Pr(A^c)$. It is noted that $\hat{T}_{[p]}$ can be consistently estimated from $L_{[p]}\left(\Theta_{[p]}; \Upsilon_{[p]} \middle| \hat{\Theta}_{[1]}, \cdots, \hat{\Theta}_{[p-1]}\right)$ if A is true (in this case, it is possible that $\hat{T}_{[p]} = T_{[p]}$ even if $\hat{\omega}_{[p]} \neq \omega_{[p]}$ with given $\hat{v}_{[p]}$). Hence there exists an

$N_2(\varepsilon) \geq N_1(\varepsilon)$, such that for any sample size $n_2 > N_2(\varepsilon)$, $\Pr(C|A) < \varepsilon/2$ and thus $\Pr(C) < \varepsilon$, where the choice of n_2 is independent of n_1. Therefore, for any a satisfying $0 < a < M$, and any $\varepsilon > 0$, there exists an $N = N_2(\varepsilon)$, such that for any sample size $n > N$,

$$P\left(\left|\hat{T}_{[p]} - T_{[p]}\right| > a\right) < \varepsilon \quad .$$

Thus $\hat{T}_{[p]}$ is a consistent estimator. By the continuous mapping theorem, $\hat{\omega}_{[p]} = \tilde{\zeta}_{[p]}\left(\hat{T}_{[p]}, \hat{v}_{[p]}\right)$ is also consistent. All other parameters in $\Theta_{[p]}$ are estimated consistently, which is straightforward from the regular maximum likelihood estimation. Therefore $\hat{\Theta}_{[p]}$ and hence $\hat{\mathbf{\Theta}}^*$ are consistent.

Now we prove the asymptotic normality. By applying the Taylor series expansion of the score function $U_{[p]}\left(\hat{\Theta}_{[p]}; \Upsilon_{[p]} \mid \hat{\Theta}_{[1]}, \cdots, \hat{\Theta}_{[p-1]}\right) = 0$ around the true parameter values and dropping higher order terms (Lehmann 1983; Cox and Reid 2004), we get

$$-\frac{1}{n}\begin{bmatrix} \dfrac{\partial U_{[1]}\left(\Theta_{[1]};\Upsilon_{[1]}\right)}{\partial \Theta_{[1]}^T} & & \mathbf{0} \\ \vdots & \ddots & \\ \dfrac{\partial U_{[K]}\left(\Theta_{[K]};\Upsilon_{[K]}\left|\Theta_{[1]},\cdots,\Theta_{[K-1]}\right.\right)}{\partial \Theta_{[1]}^T} & \cdots & \dfrac{\partial U_{[K]}\left(\Theta_{[K]};\Upsilon_{[K]}\left|\Theta_{[1]},\cdots,\Theta_{[K-1]}\right.\right)}{\partial \Theta_{[K]}^T} \end{bmatrix} \begin{bmatrix} \sqrt{n}\left(\hat{\Theta}_{[1]} - \Theta_{[1]}\right) \\ \vdots \\ \sqrt{n}\left(\hat{\Theta}_{[K]} - \Theta_{[K]}\right) \end{bmatrix}$$

$$\approx \begin{bmatrix} \dfrac{1}{\sqrt{n}} U_{[1]}\left(\Theta_{[1]};\Upsilon_{[1]}\right) \\ \vdots \\ \dfrac{1}{\sqrt{n}} U_{[K]}\left(\Theta_{[K]};\Upsilon_{[K]}\left|\Theta_{[1]},\cdots,\Theta_{[K-1]}\right.\right) \end{bmatrix} .$$

By the law of large numbers,

$$-\frac{1}{n}\frac{\partial U_{[p]}\left(\Theta_{[p]};\Upsilon_{[p]} \mid \Theta_{[1]},\cdots,\Theta_{[p-1]}\right)}{\partial \Theta_{[q]}^T} = \frac{1}{n}\sum_{m=1}^{n} -\frac{\partial U_{[p]}\left(\Theta_{[p]};\mathbf{y}_{[p]}^{(m)} \mid \Theta_{[1]},\cdots,\Theta_{[p-1]}\right)}{\partial \Theta_{[q]}^T} \xrightarrow{p} \mathbf{H}_{[p,q]} .$$

By the central limit theorem,

$$\begin{bmatrix} \frac{1}{\sqrt{n}} U_{[1]}\left(\Theta_{[1]};\Upsilon_{[1]}\right) \\ \vdots \\ \frac{1}{\sqrt{n}} U_{[K]}\left(\Theta_{[K]};\Upsilon_{[K]} \mid \Theta_{[1]},\cdots,\Theta_{[K-1]}\right) \end{bmatrix} = \frac{1}{\sqrt{n}} \begin{bmatrix} \sum_{m=1}^{n} U_{[1]}\left(\Theta_{[1]};\mathbf{y}_{[1]}^{(m)}\right) \\ \vdots \\ \sum_{m=1}^{n} U_{[K]}\left(\Theta_{[K]};\mathbf{y}_{[K]}^{(m)} \mid \Theta_{[1]},\cdots,\Theta_{[K-1]}\right) \end{bmatrix} \xrightarrow{d} \mathbf{U} \sim N\left(\mathbf{0},\mathbf{J}\right),$$

since $E\left[U_{[p]}\left(\Theta_{[p]};\mathbf{y}_{[p]} \mid \Theta_{[1]},\cdots,\Theta_{[p-1]}\right)\right]=0$ for all p. According to Lemma 5.2 in Chapter 6 of Lehmann (1998), the limit distribution of $\sqrt{n}\left(\hat{\mathbf{\Theta}}^{*}-\mathbf{\Theta}^{*}\right)$ is that of $\mathbf{H}^{-1}\mathbf{U}$, following which the asymptotic normality of $\hat{\mathbf{\Theta}}^{*}$ is then obtained

$$\sqrt{n}\left(\hat{\mathbf{\Theta}}^{*}-\mathbf{\Theta}^{*}\right) \sim N\left(\mathbf{0},\mathbf{H}^{-1}\mathbf{J}\left(\mathbf{H}^{-1}\right)^{T}\right).$$

To prove the equation (A.3), we first define the random column vector $\mathbf{y} = [\mathbf{y}_1; \mathbf{y}_2; \ldots; \mathbf{y}_k]$, where $\mathbf{y} - \mathbf{F\beta} \sim N(\mathbf{0}, \mathbf{C})$. The covariance matrix $\mathbf{C}$ consists of blocks $\mathbf{C}_{r,s}$ defining cross-covariance between the rth and sth profiles. Let $\mathbf{AA}^T = \mathbf{C}$ and define a random vector $\mathbf{u} \sim N(\mathbf{0}, \mathbf{I})$ with the same dimension as $\mathbf{y}$, then $\mathbf{Au} = \mathbf{y} - \mathbf{F\beta}$. Write $\mathbf{A}^{T}=\left[\mathbf{A}_1^{T},\cdots,\mathbf{A}_k^{T}\right]$ such that the block matrices satisfy $\mathbf{A}_r\mathbf{A}_s^{T}=\mathbf{C}_{r,s}$ and define $\mathbf{A}_{[p]}^{T}=[\mathbf{A}_{r_p}^{T},\mathbf{A}_{s_p}^{T}]$, then we have

$$\mathbf{A}_{[p]}\mathbf{A}_{[q]}^{T}=\mathbf{C}_{[p,q]}=\begin{bmatrix} \mathbf{C}_{r_p,r_q} & \mathbf{C}_{r_p,s_q} \\ \mathbf{C}_{s_p,r_q} & \mathbf{C}_{s_p,s_q} \end{bmatrix}.$$

The pth conditional log-likelihood component can be written as ($p = 1, 2, \ldots, K$)

$$l_{[p]}\left(\Theta_{[p]};\mathbf{y}_{[p]} \mid \Theta_{[1]},\cdots,\Theta_{[p-1]}\right)=-\frac{1}{2}\left(\log\left|\mathbf{C}_{[p]}\right|+\left(\mathbf{y}_{[p]}-\mathbf{F}_{[p]}\boldsymbol{\beta}_{[p]}\right)^{T}\mathbf{C}_{[p]}^{-1}\left(\mathbf{y}_{[p]}-\mathbf{F}_{[p]}\boldsymbol{\beta}_{[p]}\right)\right),$$

where $\mathbf{C}_{[p]} = \mathbf{C}_{[p,p]}$ and $\mathbf{F}_{[p]}=\mathrm{diag}(\mathbf{F}_{r_p},\mathbf{F}_{s_p})$. By taking its score function, plugging into (A.2) and after some derivations and simplification, the equation (A.3) can be derived. ■

Proposition 2. Under the usual regularity conditions, as n → ∞, estimates of the original parameters in $\boldsymbol{\Theta}$ are consistent, i.e., $\hat{\boldsymbol{\Theta}} \xrightarrow{p} \boldsymbol{\Theta}$, and

$$\sqrt{n}\left(\hat{\boldsymbol{\Theta}}-\boldsymbol{\Theta}\right) \sim N\left(\mathbf{0}, \mathbf{B}\mathbf{H}^{-1}\mathbf{J}\left(\mathbf{H}^{-1}\right)^{T}\mathbf{B}^{T}\right) \tag{A.4}$$

Proof: By the linear relationship $\hat{\boldsymbol{\Theta}} = \mathbf{B}\hat{\boldsymbol{\Theta}}^{*}$, the proposition 2 can be quickly obtained by the delta method. ■

Define $K \times 1$ vectors $\boldsymbol{\omega} = [\omega_{[1]}, \omega_{[2]}, \ldots, \omega_{[K]}]^{T}$, $\mathbf{t} = [T_{[1]}, T_{[2]}, \ldots, T_{[K]}]^{T}$, and $\boldsymbol{\zeta} = [\zeta_{[1]}(\cdot), \zeta_{[2]}(\cdot), \ldots, \zeta_{[K]}(\cdot)]^{T}$, whose elements are arranged according to the sequence S. The following proposition is obtained.

Proposition 3. Under the usual regularity conditions, as n → ∞, estimates of the cross-correlation factors are consistent, i.e., $\hat{\mathbf{t}} \xrightarrow{p} \mathbf{t}$, and

$$\sqrt{n}\left(\hat{\mathbf{t}}-\mathbf{t}\right) \sim N\left(\mathbf{0}, \frac{\partial \boldsymbol{\zeta}}{\partial \boldsymbol{\omega}^{T}}\mathbf{B}_{\omega}\mathbf{H}^{-1}\mathbf{J}\left(\mathbf{H}^{-1}\right)^{T}\mathbf{B}_{\omega}^{T}\frac{\partial \boldsymbol{\zeta}}{\partial \boldsymbol{\omega}}\right) \tag{A.5}$$

where matrix $\mathbf{B}_{\omega}$ is defined based on the linear transform $\hat{\boldsymbol{\omega}} = \mathbf{B}_{\omega}\hat{\boldsymbol{\Theta}}^{*}$.

Proof: The proposition 3 is obvious by the continuous mapping theorem and the delta method. ■

REFERENCES

1. Lehmann, E. L. (1998). Theory of Point Estimation, Wiley, New York.

2. Cox, D. R. and Reid, N. (2004). A note on pseudolikelihood constructed from marginal densities. Biometrika, 91, 729-737.

B. NOTATION

$[\mathbf{v}_1;\mathbf{v}_2;\cdots;\mathbf{v}_i;\cdots]$	denotes $[\mathbf{v}_1^T,\mathbf{v}_2^T,\cdots,\mathbf{v}_i^T,\cdots]^T$, a column vector
$[p]$	denotes the index to the pth bivariate GP model (a pair of profiles)
$\mathbf{y}_{[p]}$	responses of the pth model, i.e. $\mathbf{y}_{[p]}=[\mathbf{y}_{r_p};\mathbf{y}_{s_p}]$
$\boldsymbol{\beta}_{[p]}$	regression coefficient of the pth model, i.e. $\boldsymbol{\beta}_{[p]}=[\boldsymbol{\beta}_{r_p}^{(s_p)};\boldsymbol{\beta}_{s_p}^{(r_p)}]$
$\boldsymbol{\psi}_{[p]}$	all parameters in the covariance function of the pth model, i.e. $\boldsymbol{\psi}_{[p]}=[\boldsymbol{\phi}_{r_p}^{(s_p)};\boldsymbol{\phi}_{s_p}^{(r_p)};\sigma_{r_p}^{(s_p)};\sigma_{s_p}^{(r_p)};\omega_{[p]};\sigma_M^{[p]}]$
$\omega_{[p]}$	ω_{r_p,s_p} in the Hypersphere Decomposition of the pth model
$\sigma_M^{[p]}$	measurement error of the pth model, i.e. $\sigma_M^{[p]}=\sigma_M$
$\Theta_{[p]}$	all parameters of the pth model, i.e. $\Theta_{[p]}=[\boldsymbol{\beta}_{[p]};\boldsymbol{\psi}_{[p]}]$
$\boldsymbol{\Theta}$	all parameters of the MGP, i.e. $\boldsymbol{\Theta}=[\boldsymbol{\beta};\boldsymbol{\phi};\boldsymbol{\sigma};\sigma_M;\boldsymbol{\Omega}]$
$\boldsymbol{\Theta}^*$	a collection of the parameters in all bivariate models, i.e. $\boldsymbol{\Theta}^*=[\Theta_{[1]};\Theta_{[2]};...;\Theta_{[K]}]$