Appendix F: Estimation Errors of Fourier Coefficients for Random Balance Design Sampling

Based on the definition of Fourier coefficients $C^{(\theta)}{}_{00\cdots r_i\cdots\cdots 00}$ in eq. (13) and eq. (14), we have

$$
\begin{aligned}
C^{(\theta)}{}_{00\cdots r_i\cdots 00} &= E_\theta(g(G(\theta_1),\ldots,G(\theta_n))e^{-\mathbf{i}(r_i\theta_i)}) \\
&= (\frac{1}{2\pi})^n \int_0^{2\pi}\cdots\int_0^{2\pi} g(G(\theta_1),\ldots,G(\theta_n))e^{-\mathbf{i}(r_i\theta_i)}d\theta_1\cdots d\theta_n \\
&= \frac{1}{2\pi}\int_0^{2\pi} E_\theta(g(G(\theta_1),\ldots,G(\theta_n))\,|\,\theta_i)e^{-\mathbf{i}(r_i\theta_i)}d\theta_i \\
&= \frac{1}{2\pi}\int_0^{2\pi} u(\theta_i)e^{-\mathbf{i}(r_i\theta_i)}d\theta_i \\
&= E_{\theta_i}(u(\theta_i)e^{-\mathbf{i}(r_i\theta_i)})
\end{aligned}
\qquad \text{(F1)}
$$

where

$$u(\theta_i) = E_\theta(g(G(\theta_1),\ldots,G(\theta_n))\,|\,\theta_i)\,. \qquad \text{(F2)}$$

Eq. (D2) suggests that, if we know $u(\theta_i)$, then the estimation of $C^{(\theta)}{}_{00\cdots r_i\cdots\cdots 00}$ can be reduced to a one-dimension estimation. For a sample with size *N*, $C^{(\theta)}{}_{00\cdots r_i\cdots\cdots 00}$ can be estimated by

$$\hat{C}^{(\theta,u)}{}_{00\cdots r_i\cdots\cdots 00} = \frac{1}{N}\sum_{j=1}^{N}[(u(\theta_i^{(j)}))e^{-\mathbf{i}(r_i\theta_i^{(j)})}]\,. \qquad \text{(F3)}$$

For a grid sample with the same size *N* in eq. (39), the estimation of $C^{(\theta)}{}_{00\cdots r_i\cdots 00}$ based on eq. (F3) is essentially a numerical integral using the rectangle rule, which has a relatively high accuracy since the estimate error will decay at a rate of squared sample size (see Appendix C for details). However, we are not able to know $u(\theta_i)$ exactly. Using eq. (F2), the model $y = g(G(\theta_1),\ldots,G(\theta_n))$ can be decomposed as follows,

$$g(G(\theta_1),\ldots,G(\theta_n)) = u(\theta_i) + \xi_{-x_i} \qquad \text{(F4)}$$

where ξ_{-x_i} is the random noise independent from $u(\theta_i)$, which is resulted from the uncertainty in other parameters. In view that

$$\begin{aligned} E(u(\theta_i)) &= E(E_\theta(g(G(\theta_1),...,G(\theta_n)) \mid \theta_i)) \\ &= E(g(G(\theta_1),...,G(\theta_n))) \end{aligned}$$

based on eq. (F4), we get

$$\mathrm{E}(\xi_{-x_i})=0. \quad \text{(F5)}$$

Using the well known fact about the variance decomposition (Bickel and Doksum 2001)

$$\mathrm{V}(g(G(\theta_1),...,G(\theta_n)))=\mathrm{V}(E_\theta(g(G(\theta_1),...,G(\theta_n)) \mid \theta_i))+\mathrm{E}(V(g(G(\theta_1),...,G(\theta_n)) \mid \theta_i),$$

we have

$$\begin{aligned} V(\xi_i) &= E(V(g(G(\theta_1),...,G(\theta_n)) \mid \theta_i)) \\ &= V(g(G(\theta_1),...,G(\theta_n))) - V(E_\theta(g(G(\theta_1),...,G(\theta_n)) \mid \theta_i). \end{aligned}$$

Using the relationship of the variances between parameter space and θ-space [see eq. (6)], we get

$$V(\xi_{-x_i}) = V^{(x_1,...,x_n)} - V^{(x_i)}. \quad \text{(F6)}$$

For the estimation of Fourier coefficient $C^{(\theta)}{}_{00\cdot\cdot r_i\cdot\cdot 00}$ in eq. (15), we have

$$\begin{aligned} \hat{C}^{(\theta)}_{00\cdot\cdot r_i\cdots\cdots 00} &= \frac{1}{N}\sum_{j=1}^{N}(g(G(\theta_1^{(j)}),...,G(\theta_n^{(j)}))e^{-\mathbf{i}(r_i\theta_i^{(j)})}) \\ &= \frac{1}{N}\sum_{j=1}^{N}[(u(\theta_i^{(j)})+\xi_i^{(j)})e^{-\mathbf{i}(r_i\theta_i^{(j)})}] \\ &= \frac{1}{N}\sum_{j=1}^{N}[(u(\theta_i^{(j)})e^{-\mathbf{i}(r_i\theta_i^{(j)})}]+\frac{1}{N}\sum_{j=1}^{N}\xi_i^{(j)}e^{-\mathbf{i}(r_i\theta_i^{(j)})} \\ &= \hat{C}^{(\theta,u)}_{00\cdot\cdot r_i\cdots\cdots 00}+\hat{C}^{(\theta,\xi)}_{00\cdot\cdot r_i\cdots\cdots 00}, \end{aligned} \quad \text{(F7)}$$

where

$$\hat{C}^{(\theta,\xi)}_{00\cdot\cdot r_i\cdot\cdot 00} = \frac{1}{N}\sum_{j=1}^{N}\xi_i^{(j)}e^{-\mathbf{i}(r_i\theta_i^{(j)})}. \quad \text{(F8)}$$

For $\hat{C}^{(\theta,\xi)}_{00\cdots r_i\cdots 00} = \hat{a}^{(\theta,\xi)}_{00\cdots r_i\cdots 00} - \mathbf{i}\hat{b}^{(\theta,\xi)}_{00\cdots r_i\cdots 00}$, with

$$\begin{aligned} \hat{a}^{(\theta,\xi)}_{00\cdots r_i\cdots 00} &= \frac{1}{N}\sum_{j=1}^{N}(\xi^{(j)})\cos(r_i\theta_i^{(j)}), \\ \hat{b}^{(\theta,\xi)}_{00\cdots r_i\cdots 00} &= \frac{1}{N}\sum_{j=1}^{N}(\xi^{(j)})\sin(r_i\theta_i^{(j)}), \end{aligned}$$

based on the Central Limit theorem, the coefficient $\hat{a}^{(\theta,\xi)}{}_{00\cdots r_i\cdots\cdots 00}$ and $\hat{b}^{(\theta,\xi)}{}_{00\cdots r_i\cdots\cdots 00}$ will approximately follow a normal distribution. Specifically,

$$\begin{aligned} \hat{a}^{(\theta,\xi)}_{00\cdots r_i\cdots\cdots 00} &\sim \ddot{N}(0, \frac{V(\xi_{-x_i})}{2N}), \\ \hat{b}^{(\theta,\xi)}_{00\cdots r_i\cdots\cdots 00} &\sim \ddot{N}(0, \frac{V(\xi_{-x_i})}{2N}), \end{aligned} \qquad \text{(F9)}$$

where $\ddot{N}(u,v)$ represents a normal distribution with mean *u* and variance *v*. Eq. (F9) indicates that the estimation error decays at a rate of $\frac{1}{\sqrt{N}}$ with an increasing sample size *N*. Therefore, the estimation error $\hat{C}^{(\theta,\xi)}{}_{00\cdots r_i\cdots\cdots 00}$ is much higher than the estimation error for $\hat{C}^{(\theta,u)}{}_{00\cdots r_i\cdots\cdots 00}$, which decays at a rate of $\frac{1}{N^2}$. For a relatively large sample size *N*, if we assume that the estimation error for $\hat{C}^{(\theta,u)}{}_{00\cdots r_i\cdots\cdots 00}$ is negligible compared to $\hat{C}^{(\theta,\xi)}{}_{00\cdots r_i\cdots\cdots 00}$ so that

$$\hat{C}^{(\theta,u)}{}_{00\cdots r_i\cdots\cdots 00} \approx C^{(\theta)}{}_{00\cdots r_i\cdots\cdots 00}, \qquad \text{(F10)}$$

then

$$\begin{aligned} E(\left|\hat{C}^{(\theta)}_{00\cdots r_i\cdots\cdots 00}\right|^2) &\approx E(\left|C^{(\theta)}{}_{00\cdots r_i\cdots\cdots 00}\right|^2) + E(\left|\hat{C}^{(\theta,\xi)}_{00\cdots r_i\cdots\cdots 00}\right|^2) \\ &= \left|C^{(\theta)}{}_{00\cdots r_i\cdots\cdots 00}\right|^2 + \frac{1}{N^2}\sum_{j=1}^{N} E[\xi_{-x_i}{}^{(j)} e^{-\mathbf{i}(r_i\theta_i^{(j)})} \xi_{-x_i}{}^{(j)} e^{\mathbf{i}(r_i\theta_i^{(j)})}] \\ &= \left|C^{(\theta)}{}_{00\cdots r_i\cdots\cdots 00}\right|^2 + \frac{1}{N^2}\sum_{j=1}^{N} E[\xi_{-x_i}{}^{(j)}]^2 \\ &= \left|C^{(\theta)}{}_{00\cdots r_i\cdots\cdots 00}\right|^2 + \frac{V(\xi_{-x_i})}{N}. \end{aligned} \qquad \text{(F11)}$$

For $\hat{C}^{(\theta)}_{00\cdots r_i\cdots 00} = \hat{a}^{(\theta)}_{00\cdots r_i\cdots 00} - \mathbf{i}\hat{b}^{(\theta)}_{00\cdots r_i\cdots 00}$, if we assume that the estimation error for $\hat{C}^{(\theta,u)}_{00\cdots r_i\cdots 00}$

is negligible compared to $\hat{C}^{(\theta,\xi)}_{00\cdots r_i\cdots 00}$, based on eq. (F9), eq. (F7) and eq. (F1), we have

$$\begin{aligned} \hat{a}^{(\theta)}_{00\cdots r_i\cdots 00} &\sim \ddot{N}\left(a^{(\theta)}{}_{00\cdots r_i\cdots 00}, \frac{V(\xi_{-x_i})}{2N}\right), \\ \hat{b}^{(\theta)}_{00\cdots r_i\cdots 00} &\sim \ddot{N}\left(b^{(\theta)}{}_{00\cdots r_i\cdots 00}, \frac{V(\xi_{-x_i})}{2N}\right), \end{aligned} \qquad \text{(F12)}$$

where

$$\begin{aligned} a^{(\theta)}{}_{00\cdots r_i\cdots 00} &= E[g(G(\theta_1),\ldots,G(\theta_n))\cos(r_i\theta_i)], \\ b^{(\theta)}{}_{00\cdots r_i\cdots 00} &= E[g(G(\theta_1),\ldots,G(\theta_n))\sin(r_i\theta_i)]. \end{aligned}$$

This also suggests that the estimation error for Fourier coefficient $C^{(\theta)}{}_{00\cdots r_i\cdots\cdots 00}$ obtained with random balance design sampling is much smaller than that with simple random sampling, in view that estimation error for $\hat{C}^{(\theta,u)}_{00\cdots r_i\cdots 00}$ is negligible compared to $\hat{C}^{(\theta,\xi)}_{00\cdots r_i\cdots 00}$ for random balance design sampling but not for simple random sampling.

Reference

Bickel, P. J., and K. A. Doksum. 2001. Mathematical statistics : basic ideas and selected topics, 2nd edition. Prentice Hall, Upper Saddle River, N.J.