Supplementary Material

1 POSTERIOR DISTRIBUTION AND MCMC

Let $\boldsymbol{Y_n} = (Y_1, ..., Y_n)'$ follow a multivariate normal distribution,

$$\boldsymbol{Y_n}|\boldsymbol{A}, \boldsymbol{X}, \boldsymbol{V}, \boldsymbol{\gamma} \sim MVN(\boldsymbol{Z\gamma}, \boldsymbol{\Sigma}), \tag{S1}$$

where $\boldsymbol{Z} = (1, \boldsymbol{X_i}, A_i, A_i \times \boldsymbol{X_i})_{n\times(2+2p)}$, $\boldsymbol{\gamma} = (\boldsymbol{\beta}, \boldsymbol{\alpha})'$, $\boldsymbol{\Sigma} = (\sigma_{ij})_{n\times n}$, with $\sigma_{ij} = K(\boldsymbol{v_i}, \boldsymbol{v_j}) + \sigma_0^2\delta_{ij}$. The δ_{ij} is the Kronecker function, $\delta_{ij} = 1$ if $i = j$, and 0 otherwise. The GP prior covariance function $\boldsymbol{K}$ is the squared-exponential (SE) function, where

$$K(v_i, v_j) = \sigma_f^2 exp\left(-\sum_{k=1}^{p}\frac{|v_{ki} - v_{kj}|^2}{\phi_k}\right),$$

for $i, j = 1, ..., n$. The $(\phi_1, \phi_2, ..., \phi_p)$ are the length scale parameters for each of the covariate variables. The priors of the parameters in the model are

$$\begin{aligned}
\boldsymbol{\gamma} &\sim MVN\left(\boldsymbol{0}, \sigma_f^2\boldsymbol{\Sigma_{0\gamma}}\right),\\
\sigma_0^2 &\sim IG(a_0, b_0),\\
\sigma_f^2 &\sim IG(a_f, b_f),\\
\phi_k &\sim IG(a_\phi, b_\phi).
\end{aligned}$$

We set $\boldsymbol{\Sigma_{0\gamma}} = 10^6\sigma_{lm}^2(\boldsymbol{ZZ'})^{-1}, a_\phi = b_\phi = 1, a_0 = a_f = 2, b_0 = b_f = \sigma_{lm}^2/2, \sigma_{lm}^2$ is the estimated variance from a simple linear regression model of Y on A and X for computational efficiency. For the ease of computation, we do the following transformations.

$$\begin{aligned}
\upsilon &= \frac{\sigma_0^2}{\sigma_f^2},\\
\omega &= log(\upsilon),\\
\rho_k &= -log(\phi_k).
\end{aligned}$$

Thus the priors for $(\omega, \boldsymbol{\rho}, \sigma_f^2)$ are

$$\begin{aligned}
p(\boldsymbol{\rho}) &\propto \Pi_{k=1}^{p}\left\{(exp(-\rho_k))^{-a_\phi-1}exp\left\{-\frac{b_\phi}{exp(-\rho_k)}\right\}exp(-\rho_k)\right\},\\
p(\omega, \sigma_f^2) &\propto (\sigma_f^2)^{-a_f-1}exp\left\{-\frac{b_f}{\sigma_f^2}\right\}(\sigma_f^2\upsilon)^{-a_0-1}exp\left\{-\frac{b_0}{\sigma_f^2\upsilon}\right\}\sigma_f^2 exp[\omega(-a_0-1)]exp(\omega)\\
&\propto (\sigma_f^2)^{-a_0-a_f-1}exp\left\{-\sigma_f^{-2}(b_f + \frac{b_0}{\upsilon})\right\}exp(-a_0\omega).
\end{aligned}$$

The likelihood multiplied by the priors is

$$
\begin{aligned}
L &= p(\boldsymbol{Y_n}, \boldsymbol{\gamma}, \sigma_f^2, \boldsymbol{\rho}, \omega) \\
&\propto |\boldsymbol{\Sigma}|^{-\frac{1}{2}} exp\left\{-\frac{1}{2}[(\boldsymbol{Y_n} - \boldsymbol{Z\gamma})'\boldsymbol{\Sigma^{-1}}(\boldsymbol{Y_n} - \boldsymbol{Z\gamma})]\right\} p(\boldsymbol{\gamma})p(\boldsymbol{\rho})p(\omega, \sigma_f^2).
\end{aligned}
$$

Therefore the conditional posterior distribution of $\boldsymbol{\gamma}$ is

$$
\begin{aligned}
p(\boldsymbol{\gamma}|\boldsymbol{Y_n}, \sigma_f^2, \boldsymbol{\rho}, \omega) &\propto exp\left\{-\frac{1}{2}[(\boldsymbol{Y_n} - \boldsymbol{Z\gamma})'\boldsymbol{\Sigma^{-1}}(\boldsymbol{Y_n} - \boldsymbol{Z\gamma}) + \gamma'\sigma_f^{-2}\boldsymbol{\Sigma_{0\gamma}}^{-1}\gamma]\right\} \\
&\propto exp\left\{-\frac{1}{2}[\gamma'(\boldsymbol{Z'\Sigma^{-1}Z} + \sigma_f^{-2}\boldsymbol{\Sigma_{0\gamma}}^{-1})\gamma - 2\gamma'\boldsymbol{Z'\Sigma^{-1}Y_n}]\right\} \\
&\propto exp\left\{-\frac{1}{2}(\gamma - \boldsymbol{\mu_\gamma})'\boldsymbol{\Sigma_\gamma^{-1}}(\boldsymbol{\gamma} - \boldsymbol{\mu_\gamma})\right\},
\end{aligned}
$$

where $\boldsymbol{\mu_\gamma} = \boldsymbol{\Sigma_\gamma^{-1}Z'\Sigma^{-1}Y_n}$ and $\boldsymbol{\Sigma_\gamma} = (\boldsymbol{Z'\Sigma^{-1}Z} + \sigma_f^{-2}\boldsymbol{\Sigma_{0\gamma}}^{-1})^{-1}$. Thus

$$
(\boldsymbol{\gamma}|\boldsymbol{Y_n}, \sigma_f^2, \boldsymbol{\rho}, \omega) \sim MVN(\boldsymbol{\mu_\gamma}, \boldsymbol{\Sigma_\gamma^{-1}}). \tag{S2}
$$

Let $\boldsymbol{G} = \sigma_f^{-2}\boldsymbol{\Sigma}$. Then the marginal posterior distribution of $(\sigma_0^2, \sigma_f^2, \phi_k)$ is

$$
\begin{aligned}
p(\sigma_f^2, \boldsymbol{\rho}, \omega) &\propto \iint p(\boldsymbol{Y_n}, \gamma, \sigma_f^2, \boldsymbol{\rho}, \omega)d\boldsymbol{\gamma} \\
&\propto |\boldsymbol{\Sigma}|^{-\frac{1}{2}}|\boldsymbol{\Sigma_\gamma}|^{\frac{1}{2}} exp\left\{-\frac{1}{2}\boldsymbol{Y_n}'\boldsymbol{\Sigma}^{-1}\boldsymbol{Y_n} + \frac{1}{2}\boldsymbol{\mu_\gamma}'\boldsymbol{\Sigma_\gamma^{-1}}\boldsymbol{\mu_\gamma}\right\} p(\boldsymbol{\rho})p(\omega, \sigma_f^2) \\
&\propto \sigma_f^{-\frac{1}{2}n}|\boldsymbol{G}|^{-\frac{1}{2}}\sigma_f^{\frac{1}{2}p}|\boldsymbol{Z'G^{-1}Z} + \boldsymbol{\Sigma_{0\gamma}}^{-1}|^{-\frac{1}{2}} exp\left\{-\frac{1}{2\sigma_f^2}\boldsymbol{Y_n}'\boldsymbol{G}^{-1}\boldsymbol{Y_n}\right\} \\
&\quad exp\left\{-\frac{1}{2\sigma_f^2}\left[-(\boldsymbol{Z'G^{-1}Y_n})'(\boldsymbol{Z'G^{-1}Z} + \boldsymbol{\Sigma_{0\gamma}}^{-1})^{-1}(\boldsymbol{Z'G^{-1}Y_n})\right]\right\} \\
&\quad p(\boldsymbol{\rho})(\sigma_f^2)^{-a_0-a_f-1} exp\left\{-\sigma_f^{-2}(b_f + \frac{b_0}{\upsilon})\right\} exp(-a_0\omega).
\end{aligned}
$$

Therefore we can derive

$$
p(\sigma_f^2|\boldsymbol{Y_n}, \boldsymbol{\rho}, \omega) \sim (\sigma_f^2)^{-a_{\sigma_f}-1} exp\left\{-\frac{b_{\sigma_f}}{\sigma_f^2}\right\},
$$

where $a_{\sigma_f} = \frac{1}{2}(n-p) + a_f + a_0$, $b_{\sigma_f} = b_f + \frac{b_0}{\upsilon} + \frac{1}{2}[\boldsymbol{Y_n}'\boldsymbol{G}^{-1}\boldsymbol{Y_n} - (\boldsymbol{Z'G^{-1}Y_n})'(\boldsymbol{Z'G^{-1}Z} + \boldsymbol{\Sigma_{0\gamma}}^{-1})^{-1}$ $(\boldsymbol{Z'G^{-1}Y_n})]$. Thus

$$
(\sigma_f^2|\boldsymbol{Y_n}, \boldsymbol{\rho}, \omega) \sim IG(a_{\sigma_f}, b_{\sigma_f}). \tag{S3}
$$

The posterior distribution of $(\boldsymbol{\rho}, \omega)$ is

$$p(\boldsymbol{\rho}, \omega | \boldsymbol{Y_n}) \sim b_{\sigma_f}^{-a_{\sigma_f}} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} |\boldsymbol{Z'G^{-1}Z} + \boldsymbol{\Sigma_{0\gamma}}^{-1}|^{-\frac{1}{2}} exp(-a_0\omega)$$
$$\Pi_{k=1}^{p} \left\{ (exp(-\rho_k))^{-a_\phi - 1} exp\left\{ -\frac{b_\phi}{exp(-\rho_k)} \right\} exp(-\rho_k) \right\} \tag{S4}$$

The posterior of the parameters can be obtained by implementing a Gibbs sampling algorithm: first sample the $\boldsymbol{\rho}$ and ω from its posterior distribution $p(\boldsymbol{\rho}, \omega | \boldsymbol{Y_n})$ in S4 by using the Metropolis–Hastings algorithm; then sample σ_f^2 from an inverse gamma distribution in S3; at last sample $\boldsymbol{\gamma}$ from a multivariate normal distribution in S2.

Let $\boldsymbol{Z^*}$ be a new unit and $\boldsymbol{Y^*}$ the corresponding outcomes, the joint distribution of $(\boldsymbol{Y_n}, \boldsymbol{Y^*})$ is

$$\begin{pmatrix} \boldsymbol{Y_n} \\ \boldsymbol{Y^*} \end{pmatrix} |\gamma, \sigma_f^2, \boldsymbol{\rho}, \omega \sim N \left[\begin{pmatrix} \boldsymbol{Z\gamma} \\ \boldsymbol{Z^*\gamma} \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma_{11}} & \boldsymbol{\Sigma_{12}} \\ \boldsymbol{\Sigma_{21}} & \boldsymbol{\Sigma_{22}} \end{pmatrix} \right],$$

where $\boldsymbol{\Sigma_{11}} = \Sigma$. So the predictive posterior distributions of $\boldsymbol{Y^*} | \boldsymbol{Y_n}, \gamma, \sigma_f^2, \boldsymbol{\rho}, \omega$ is

$$\boldsymbol{Y^* | Y_n, \gamma, \sigma_f^2, \rho, \omega} \sim N(\boldsymbol{\mu_{Y^*}}, \boldsymbol{\Sigma_{Y^*}})$$

where $\mu_{Y^*} = \boldsymbol{Z\gamma} + \boldsymbol{\Sigma_{21}\Sigma_{11}^{-1}}(\boldsymbol{Y_n} - \boldsymbol{Z\gamma})$ and $\boldsymbol{\Sigma_{Y^*}} = \boldsymbol{\Sigma_{22}} - \boldsymbol{\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}}$. Then we derive the posterior distribution of the potential outcomes $\boldsymbol{Y^{(0)}}$ and $\boldsymbol{Y^{(1)}}$ by setting $\boldsymbol{Z^*} = (1, \boldsymbol{X_i}, 0, 0 \times \boldsymbol{X_i})_{n \times (2+2p)}$ and $\boldsymbol{Z^*} = (1, \boldsymbol{X_i}, 1, 1 \times \boldsymbol{X_i})_{n \times (2+2p)}$. At last, we sample $\boldsymbol{Y^{(0)}}$ and $\boldsymbol{Y^{(1)}}$ from their posterior distribution and estimate the ATE from the average of $\boldsymbol{Y^{(1)}} - \boldsymbol{Y^{(0)}}$.

Figure S1: Distribution of the Estimated by Different Sample Sizes ATE from GPMatch under the Kang and Shafer Dual Misspecifcation Setting. Upper panel presents the results of GPMatch with the treatment effect only in the mean function model; lower panel presents the results of GPMatch with the treatment effect and the $X_1 - X_4$ in the mean function model. Both included $X_1 - X_4$ in the covariate function.