Supplementary Material

Proof of lemma 4.2.
Let $\mathbf{y} = (y_1 \dots y_n)$, and $F_{\mathbf{Y}}(\mathbf{y}) = (F_{y_1}(y_{y_1}) \dots, F_{y_n}(y_n))$ (for readability, we omit m that indexes the instantiation). By definition of the expectation and conditional density we have:

$$\mathbb{E}_{f(X|\mathbf{y},y^*)}(X) = \int x \frac{f(x,\mathbf{y},y^*)}{f(\mathbf{y},y^*)} dx. \tag{12}$$

Next, recall that the joint density $f(x,\mathbf{y},y^*)$ can be written as:

$$f(x,\mathbf{y},y^*) = c(F_1(x_1), F_{\mathbf{Y}}(\mathbf{y}), F_{y^*}(y^*)) \prod_i f_i(x_i),$$

where $f_i(x_i)$ are the univariate densities and $c(\cdot)$ is the copula density. Therefore:

$$f_{\mathbf{Y},Y^*}(\mathbf{y},y^*) = \left(\prod_i f_i(y_i)\right) f_{Y^*}(y^*) \int c(F_X(x), \mathbf{F_Y}(\mathbf{y}), F_{Y^*}(y^*)) f_X(x) dx,$$

Plugging this into Equation (12), we then have:

$$\mathbb{E}_{f(X|\mathbf{y},y^*)}(X) = \int x \frac{c(F_X(x), \mathbf{F_Y}(\mathbf{y}), F_{Y^*}(y^*))}{c(\mathbf{F_Y}(\mathbf{y}), F_{Y^*}(y^*))} f_X(x) dx,$$

and, using a change of variable $U = F_X(x)$, we get:

$$\mathbb{E}_{f(\mathbf{X}|\mathbf{y},y^*)}(\mathbf{X}) = \int F_X^{-1}(u) c(u|\mathbf{F_Y}(\mathbf{y}), F_{Y^*}(y^*)) du,$$

where $c(u|\mathbf{F_Y}(\mathbf{y}), F_{Y^*}(y^*)) = \frac{c(u,\mathbf{F_Y}(\mathbf{y}),F_{Y^*}(y^*))}{c(\mathbf{F_Y}(\mathbf{y}),F_{Y^*}(y^*))}$. □

Proof of Lemma 4.4.
Using the notations introduced in Section 3, recall that in the case of a Gaussian copula $c(U_i|\mathbf{u}_{-i}) = \frac{\phi\left(Z_i|\mathbf{z}_{-i};\boldsymbol{\beta}_{-i},\sigma\right)}{\phi(z_i)}$, where $\phi\left(Z_i|\mathbf{z}_{-i};\boldsymbol{\beta}_{-i},\sigma\right)$ is the conditional density induced from the joint density $\phi_\Sigma(\mathbf{z})$ by conditioning Z_i on $\mathbf{z}_{-i}$, and ϕ is the standrad univariate normal p.d.f. Therefore:

$$\begin{aligned}\mathbb{E}_{C_\Sigma}(U_i|\mathbf{u}_{-i}) &= \int_0^1 u_i c(u_i|\mathbf{u}_{-i}) du_i \\ &= \int_0^1 u_i \phi\left(z_i|\mathbf{z}_{-i};\boldsymbol{\beta}_{-i},\sigma\right)/\phi(z_i) du_i.\end{aligned}$$

Using a change of variables $z_i = \Phi^{-1}(u_i)$, we obtain

$$\begin{aligned}\mathbb{E}_{C_\Sigma}(U_i|\mathbf{u}_{-i}) &= \int_{-\infty}^{\infty} \Phi(z_i)\phi\left(z_i|\mathbf{z}_{-i};\boldsymbol{\beta}_{-i},\sigma\right) du_i \\ &= \mathbb{E}_{\phi\left(Z_i|\mathbf{z}_{-i};\boldsymbol{\beta}_{-i},\sigma\right)}[\Phi(z_i)].\end{aligned}$$

□

Proof of Lemma 5.1.
Recall that when a new variable W is added as parent of X_i, we only estimate the scale parameter associated with $Z_w = \Phi^{-1}(F_W(w))$, while all other parameters are held fixed (see Section 5). That is, $\beta_j^+ = \hat{\beta}_j$, $\forall j \neq i$, and similarly for the variance parameter $\sigma^+ = \hat{\sigma}$.

Now, by the definition of QIP

$$\left(z_i[m] - \sum_{j:j\neq i} \hat{\beta}_j z_j[m]\right)^2 = z^*[m]^2, \quad \forall m.$$

Similarly,

$$\left(z_i[m] - \sum_{j:j\neq i} \hat{\beta}_j z_j[m] - \beta_w^+ z_w[m]\right)^2 = \left(z^*[m] - \beta_w^+ z_w[m]\right)^2, \quad \forall m.$$

Therefore, the change in the likelihood score is given by the difference

$$\begin{aligned}\Delta_{X_i|\mathrm{Par}_i(W)} &= -\frac{M}{2}\left[\ln(\sigma^+)^2 - \ln\hat{\sigma}^2\right] \\ &- \frac{1}{2(\sigma^+)^2}\sum_m \left(z^*[m] - \beta_w^+ z_w[m]\right)^2 + \frac{1}{2\hat{\sigma}^2}\sum_m z^*[m]^2.\end{aligned}$$

Since $\sigma^+ = \hat{\sigma}$, using standard algebraic manipulations this reduces to

$$\Delta_{X_i|\mathrm{Par}_i(W)} = -\frac{1}{2(\hat{\sigma})^2}\left(-2\beta_w^+(\mathbf{z}_w \cdot \mathbf{z}^*) + (\beta_w^+)^2\|\mathbf{z}_w\|^2\right).$$

Taking the derivative w.r.t. β_w^+ and setting it to zero we obtain

$$\hat{\beta}_w^+ = \frac{(\mathbf{z}^* \cdot \mathbf{z}_w)}{\|\mathbf{z}^*\|^2},$$

where $(\cdot)$ is the standard inner product. Plugging this into Equation (7), we get

$$\widetilde{\Delta}_{X_i|\mathrm{Par}_i}(W) = \frac{1}{2\hat{\sigma}^2}\frac{(\mathbf{z}^* \cdot \mathbf{z}_w)^2}{\|\mathbf{z}_w\|^2}. \tag{13}$$

Finally, we note that the ML estimator for σ^2 is given by:

$$\hat{\sigma}^2 = \frac{1}{M}\sum_m \left(z_i[m] - \sum_{j:j\neq i}\hat{\beta}_i z_j[m]\right)^2 = \frac{1}{M}\|\mathbf{z}^*\|^2. \tag{14}$$

Denote the angle between $\mathbf{z}^*$ and $\mathbf{z}_w$ by $\angle(\mathbf{z}^*, \mathbf{z}_w)$. Plugging Equation (14) into Equation (13), and using the identity $\cos^2(\angle(\mathbf{z}^*, \mathbf{z}_w)) = \frac{(\mathbf{z}^* \cdot \mathbf{z}_w)^2}{\|\mathbf{z}^*\|^2\|\mathbf{z}_w\|^2}$, we obtain

$$\widetilde{\Delta}_{X_i|\mathrm{Par}_i}(W) = \frac{1}{2\hat{\sigma}^2}\frac{(\mathbf{z}^* \cdot \mathbf{z}_w)^2}{\|\mathbf{z}_w\|^2} = \frac{M}{2}\cos^2(\angle(\mathbf{z}^*, \mathbf{z}_w)).$$

□

Proof of Corollary 5.2.
Recall that

$$\begin{aligned}&\Delta_{X_i|\mathrm{Par}_i}(W) \\ &\equiv \max_{\theta^+} l_X(\mathbf{D} : \mathrm{Par}_i \cup \{W\}, \theta^+) - l_X(\mathbf{D} : \mathrm{Par}_i, \hat{\boldsymbol{\theta}}) \\ &= \max_{\boldsymbol{\theta}^+} \textstyle\sum_m \log f(x_i[m]|\mathrm{Par}_i[m], w[m]; \boldsymbol{\theta}^+) \\ &\quad - \textstyle\sum_m \log f(x_i[m]|\mathrm{Par}_i[m]; \hat{\boldsymbol{\theta}}) \\ &= \max_{\boldsymbol{\theta}^+} \textstyle\sum_m \log c(F_i[m]|\{F_j[m]\}_{j\in\mathrm{Par}_i}, F_W[m]; \boldsymbol{\theta}^+) \\ &\quad - \textstyle\sum_m \log c(F_i[m]|\{F_j[m]\}_{j\in\mathrm{Par}_i}; \hat{\boldsymbol{\theta}}) \\ &= \max_{\boldsymbol{\beta}^+,\sigma^+} \textstyle\sum_m \phi\left(z_i[m]|\mathbf{z}_{\mathrm{Par}_i}[m], z_w[m]; \boldsymbol{\beta}^+, \sigma^+\right) \\ &\quad - \textstyle\sum_m \phi\left(z_i[m]|\mathbf{z}_{\mathrm{Par}_i}[m]; \hat{\boldsymbol{\beta}}, \hat{\sigma}\right),\end{aligned} \tag{15}$$

Denote by Ω the parameter space of $\boldsymbol{\beta}^+$ and σ^+. When introducing a new parent variable, W, as a parent of X_i, ideally, we should estimate θ by $\hat{\boldsymbol{\theta}}_c$, such that

$$\hat{\boldsymbol{\theta}}_c \in \operatorname*{argmax}_{(\boldsymbol{\beta}^+,\sigma^+)\in\Omega} l_X(\mathbf{D} : \mathrm{Par}_i \cup \{W\}, \theta^+).$$

Denote the corresponding change in the likelihood function by:

$$\Delta^c_{X_i|\mathrm{Par}_i}(W) \equiv l_X(\mathbf{D} : \mathrm{Par}_i \cup \{W\}, \hat{\boldsymbol{\theta}}_c) - l_X(\mathbf{D} : \mathrm{Par}_i, \hat{\boldsymbol{\theta}}).$$

However, this can be prohibitive since we need to estimate β^+, σ^+, for each candidate parent W, as well as due to the constrain $(\boldsymbol{\beta}^+, \sigma^+) \in \Omega$. Instead, as described in Section 5, we estimate $\Delta_{X_i|\mathrm{Par}_i}(W)$ by $\widetilde{\Delta}_{X_i|\mathrm{Par}_i}(W)$, that is, by maximizing only over the scale parameter associated with Z_w, β_w, while keeping all other parameters fixed to their value before W was added. Note that by doing so, we implicitly remove the constrain over β_w and estimate it by solving the following unconstrained optimization problem:

$$\hat{\beta}_w \in \operatorname*{argmax}_{\beta_w} l_X(\mathbf{D} : \mathrm{Par}_i \cup \{W\}, \theta^+),$$

Next, let $\hat{\boldsymbol{\theta}}_{uc}$ be an estimator of θ such that

$$\hat{\boldsymbol{\theta}}_{uc} \in \operatorname*{argmax}_{(\boldsymbol{\beta},\sigma)} l_X(\mathbf{D} : \mathrm{Par}_i \cup \{W\}, \theta^+),$$

that is, $\hat{\boldsymbol{\theta}}_{uc}$ is a solution to an optimization problem which is a relaxation of the original problem. Denote the corresponding change in the likelihood function by:

$$\Delta^{uc}_{X_i|\mathrm{Par}_i}(W) \equiv l_X(\mathbf{D} : \mathrm{Par}_i \cup \{W\}, \hat{\boldsymbol{\theta}}_{uc}) - l_X(\mathbf{D} : \mathrm{Par}_i, \hat{\boldsymbol{\theta}}).$$

Finally, let $\boldsymbol{\theta}_0$ be the true underlying parameters. Note that by assumption $\boldsymbol{\theta}_0 \in \Omega$.

Denote the true change in the likelihood function by:

$$\Delta^0_{X_i|\mathrm{Par}_i}(W) \equiv l_X(\mathbf{D} : \mathrm{Par}_i \cup \{W\}, \hat{\boldsymbol{\theta}}_0) - l_X(\mathbf{D} : \mathrm{Par}_i, \hat{\boldsymbol{\theta}}).$$

By definition, for all M the following holds:

$$\widetilde{\Delta}_{X_i|\mathrm{Par}_i}(W) \leq \Delta^{uc}_{X_i|\mathrm{Par}_i}(W). \tag{16}$$

Due to consistency of ML parameters, as $M \to \infty$, $\hat{\boldsymbol{\theta}}_{uc} \to \boldsymbol{\theta}_0$, therefore

$$\Delta^{uc}_{X_i|\mathrm{Par}_i}(W) \to \Delta^0_{X_i|\mathrm{Par}_i}(W), \text{ a.s.}$$

Similarly, since we have assumed that $\boldsymbol{\theta}_0 \in \Omega$, $\lim_{M\to\infty} \hat{\boldsymbol{\theta}}_c = \boldsymbol{\theta}_0$, *a.s.*, and therefore

$$\Delta^c_{X_i|\mathrm{Par}_i}(W) \to \Delta^0_{X_i|\mathrm{Par}_i}(W), \text{ a.s.}$$

To summarize, we have that:

$$\lim_{M\to\infty} \hat{\boldsymbol{\theta}}_c = \lim_{M\to\infty} \hat{\boldsymbol{\theta}}_{uc} = \boldsymbol{\theta}_0, \text{ a.s.},$$

and,

$$\begin{aligned}\lim_{M\to\infty} \Delta^c_{X_i|\mathrm{Par}_i}(W) &= \lim_{M\to\infty} \Delta^{uc}_{X_i|\mathrm{Par}_i}(W) \\ &= \Delta^0_{X_i|\mathrm{Par}_i}(W), \quad M \to \infty, \text{ a.s.}\end{aligned} \tag{17}$$

Combining now 17 and 16, by continuity of the limit we get:

$$\begin{aligned}\widetilde{\Delta}_{X_i|\mathrm{Par}_i}(W) \leq \Delta^{uc}_{X_i|\mathrm{Par}_i}(W) &= \lim_{M\to\infty} \Delta^c_{X_i|\mathrm{Par}_i}(W) \\ &= \Delta^0_{X_i|\mathrm{Par}_i}(W), \quad M \to \infty, \text{ a.s.}\end{aligned} \tag{18}$$

$\square$