Focused information criteria for copulas

Vinnie Ko[1] | **Nils Lid Hjort[1]** | **Ingrid Hobæk Haff[1]**

[1]Department of Mathematics, University of Oslo, PB 1053, Blindern, NO-0316 Oslo, Norway

Correspondence

Vinnie Ko, Department of Mathematics, University of Oslo, PB 1053, Blindern, NO-0316 Oslo, Norway
Email: vinniebk@math.uio.no

Funding information

In this paper, we extend the focused information criterion (FIC) from Claeskens and Hjort (2008) and Jullum and Hjort (2017) to copula models. copulas are often used for applications where the joint tail behavior of the variables is of particular interest, and selecting a copula that captures this well is then essential. Traditional model selection methods, such as the AIC and BIC aim at finding the overall best fitting model, which is not necessarily the one best suited for the application at hand. The FIC, on the other hand, evaluates and ranks candidate models based on the precision of their point estimates of a context-given focus parameter. This could be any quantity of particular interest, e.g. the mean, a correlation, conditional probabilities, or measures of tail dependence. We derive FIC formulae for the maximum likelihood estimator, the two-stage maximum likelihood estimator and the so-called pseudo-maximum-likelihood estimator (PML) combined with parametric margins. Further, we confirm the validity of the AIC formula for the PML estimator combined with parametric margins. To study the numerical behavior of FIC, we have carried out simulation study, and we have also analyzed a multivariate abalone data set. The results from the study show that the FIC successfully ranks candidate models in terms of their performance, defined as how well they estimate the focus parameter. In terms of estimation precision, FIC clearly outperforms AIC, especially when the focus parameter relates to only a specific part of the model, like the conditional upper tail probability.

KEYWORDS
mean squared error, model selection, semi-parametric estimation,

two-stage maximum likelihood

1 | INTRODUCTION AND COPULA MODELS

Model selection is inarguably an important part of modern statistics. Most model selection criteria work by evaluating 'overall' fit of the models, in a suitable sense, as opposed to their performance related to a specific use. For example, the AIC and TIC aim for the model that minimizes the Kullback–Leibler (KL) divergence from the real data-generating mechanism to the candidate model. However, in practice, the model itself is often not the final goal. The model is to be used for some specific tasks such as estimating the mean, a quantile, correlation or the tail dependence.

Copula models are frequently used for applications within for instance finance and insurance, as well as hydrology, where the tail behavior of the joint distribution is particularly important. Measures of interest are then typically the Value-at-Risk, the Expected Shortfall or conditional upper quantiles of the distribution of sea and river levels. Selecting a copula that captures the joint tail behavior well is then essential. It is by no means certain that this is obtained with traditional model selection methods. On the other hand, the focused information criterion (FIC), proposed by Claeskens and Hjort (2003), is a model selection method that evaluates candidate models based on the precision of their estimate of a quantity of interest, measured in terms of the mean squared error (MSE). This quantity of interest is called the 'focus parameter' and hence the name 'focused information criterion'. One of the drawbacks of the original FIC (Claeskens and Hjort, 2003) is that it assumes all the models to be parametric and nested. Jullum and Hjort (2017) extend the FIC to a more general setting by using a nonparametric model to estimate the bias part of the decomposed MSE. Their new FIC machinery is designed for maximum likelihood (ML) estimated parametric models and nonparametric models. Further, the candidate models do not have to be nested. We extend this new FIC to copula models that are estimated with maximum likelihood (ML), two-stage ML or pseudo ML combined with parametric margins (PMLpm).

Our technical setting is as follows. Let $y_1, \cdots, y_n$ be independent and identically distributed d-dimensional observations from a joint density $g(y_1, \cdots, y_d)$. The data generating model g is typically unknown. Suppose $f(y_1, \cdots, y_d, \eta)$ is our parametric approximation of g, with the parameter vector η belonging to some connected subset of the appropriate Euclidean space. Let G and $F(\cdot, \eta)$ denote cumulative distribution functions corresponding to g and $f(\cdot, \eta)$, respectively. Further, $G_j(y_j)$ and $F_j(y_j, \alpha_j)$ indicate j-th marginal distribution functions corresponding to G and $F(\cdot, \eta)$, respectively, α_j being the parameter vector belonging to margin component j.

According to Sklar's theorem (Sklar, 1959), there always exists a copula $C(u_1, \ldots, u_d, \theta)$ that satisfies

$$F(y_1, \cdots, y_d, \eta) = C(F_1(y_1, \alpha_1), \cdots, F_d(y_d, \alpha_d), \theta)$$

with the full parameter vector

$$\eta = (\alpha^{\mathrm{T}}, \theta^{\mathrm{T}})^{\mathrm{T}} = (\alpha_1^{\mathrm{T}}, \cdots, \alpha_d^{\mathrm{T}}, \theta^{\mathrm{T}})^{\mathrm{T}}.$$

Under the assumption that $F_1(y_1, \alpha_1), \cdots, F_d(y_d, \alpha_d)$ are absolutely continuous and strictly increasing, $C(\cdot, \theta)$ can be differentiated. This gives

$$f(y_1, \cdots, y_d, \eta) = c\left(F_1(y_1, \alpha_1), \cdots, F_d(y_d, \alpha_d), \theta\right) \prod_{j=1}^{d} f_j(y_j, \alpha_j),$$

where $c(u_1, \ldots, u_d) = \partial^d C(u_1, \ldots, u_d, \theta)/\partial u_1 \cdots \partial u_d$ and $f_j(y_j, \alpha_j) = \partial F_j(y_j, \alpha_j)/\partial y_j$. Similarly, the true density g can also be

decomposed into marginal densities and copula density

$$g(y_1, \cdots, y_d) = c_0 \left(G_1(y_1), \cdots, G_d(y_d) \right) \prod_{j=1}^{d} g_j(y_j),$$

with $c_0(\cdot)$ the density of the true copula. For more details on copula modeling, see Joe (1997, 2014) and Nelsen (2006).

Three of the most commonly used estimation procedures for parametric copulas are full ML, two-stage ML and PML. When full ML is employed for the copula and margins, the theory of AIC, BIC and FIC holds directly. For PML estimated copula, combined with nonparametric margins, Grønneberg and Hjort (2014) develop a copula information criterion (CIC) and Chen and Fan (2005) suggest a pseudo likelihood ratio test. Ko and Hjort (2019a) propose a different CIC, for the two-stage ML. The present paper concerns extending the FIC theory to these different estimation procedures. This partly follows the line of arguments used in Jullum and Hjort (2017, 2018), but by necessity involves certain extra efforts, as we work with more complex estimation strategies than the pure nonparametric and the maximum likelihood.

The further structure of this paper is as follows. In Section 2, the focused information criterion (FIC) for copula models under ML, two-stage ML and PMLpm estimator is derived and defined. Further, we present the AIC formula for PMLpm estimator in Section 3 and cover the average weighted FIC (AFIC), designed for the situations where multiple focus parameters are of interest, in Section 4. In Section 5, the behavior of the FIC is investigated in a simulation study. Moreover, we apply the FIC to a dataset on multiple measurements on abalones in Section 6. Finally, Section 7 contains concluding remarks and suggestions for future work.

2 | FIC FOR COPULA MODELS

2.1 | General idea and derivation of the FIC

In this subsection, we briefly summarize the idea behind the focused information criterion (FIC) of Jullum and Hjort (2017), and reach a general formula. The actual details for implementing this FIC formula depend on the specific estimation scheme, which we return to later in Sections 2.2, 2.3 and 2.4.

Most model selection criteria like AIC, BIC and TIC measure how close a candidate model is to the true model in terms of a certain divergence measure. However, in many cases, the model itself is not the goal, but rather a means to estimate a specific quantity, such as the mean or the probability of certain event. The main idea behind the FIC, first introduced by Claeskens and Hjort (2003) and later modified by Jullum and Hjort (2017, 2018), is that one wants to select a model that is good for a specific task. In our setting, we quantify this specific task as that of estimating a focus parameter, say $T(G)$, a functional of the data generating distribution. The functional needs to have a clear statistical interpretation across candidate models. With F_M the distribution associated with such a candidate model M, $T(F_M)$ is the ensuing approximation to the $T_{\text{true}} = T(G)$. We also need the functional to be smooth, in the sense of admitting an influence function leading to limiting normality with a $\sqrt{n}$ rate. This involves certain regularity conditions, as (C1)–(C4) of Jullum and Hjort (2017, Section 2).

In our setting, the estimated focus parameter via the use of a candidate model is denoted as $\widehat{T}_{\text{cnd}}$ or $T(\widehat{F})$, with $\widehat{F}$ the estimated distribution via this model. We are looking for candidate models such that $\widehat{T}_{\text{cnd}}$ is close to T_{true}. As with Jullum and Hjort (2017, 2018), we use a quadratic loss function, leading by definition to the mean squared error (MSE) as the criterion of closeness. We hence define the focused information criterion (FIC) as the estimated MSE

$$\text{FIC} = \widehat{\text{MSE}}(\widehat{T}_{\text{cnd}}) = \widehat{\text{Var}}_G(\widehat{T}_{\text{cnd}}) + \widehat{\text{bias}^2} \left(\widehat{T}_{\text{cnd}}, T_{\text{true}} \right). \tag{1}$$

We return to estimates for the variance of $\widehat{T}_{\text{cnd}}$ below, with different recipes for different estimators. For estimating the bias, we use the fact that the nonparametrically estimated T, which we denote as $\widehat{T}_{\text{np}}$, is asymptotically unbiased. This suggests using

$$\widehat{\text{bias}^2}^* \left(\widehat{T}_{\text{cnd}}, T_{\text{true}} \right) = \left(\widehat{T}_{\text{cnd}} - \widehat{T}_{\text{np}} \right)^2$$

which however tends to overestimate its estimand. By correcting for the overestimation and possible negative values, we obtain the final squared bias estimator

$$\widehat{\text{bias}^2} \left(\widehat{T}_{\text{cnd}}, T_{\text{true}} \right) = \begin{cases} \max \left\{ 0, \left(\widehat{T}_{\text{cnd}} - \widehat{T}_{\text{np}} \right)^2 - \widehat{\text{Var}}_G \left(\widehat{T}_{\text{cnd}} \right) \right. & \text{if the model is (semi)parametric} \\ \left. -\widehat{\text{Var}}_G \left(\widehat{T}_{\text{np}} \right) + 2\widehat{\text{Cov}}_G \left(\widehat{T}_{\text{np}}, \widehat{T}_{\text{cnd}} \right) \right\} & \\ 0 & \text{if the model is nonparametric.} \end{cases} \tag{2}$$

Note that this derivation of the FIC is done without specifying the estimation scheme. As long as $\widehat{T}_{\text{cnd}}$, $\widehat{T}_{\text{np}}$, $\widehat{\text{Var}}_G \left(\widehat{T}_{\text{cnd}} \right)$, $\widehat{\text{Var}}_G \left(\widehat{T}_{\text{np}} \right)$ and $\widehat{\text{Cov}}_G \left(\widehat{T}_{\text{np}}, \widehat{T}_{\text{cnd}} \right)$ can be obtained, the FIC can be defined for any model and estimation scheme.

2.2 | FIC under maximum likelihood

The ML estimator $\widehat{\eta}^{\text{ML}}$ is defined as the maximizer of $\ell_n(\eta)$, the log-likelihood function of a chosen model with given observations. Its properties are covered in classic references such as White (1982) and Le Cam (1990).

Let $\widehat{T}_{\text{ML}} = T(F(\widehat{\eta}^{\text{ML}}))$ denote a focus parameter, estimated based on a given parametric model and using the ML. Further, let $T_{0,\text{ML}}$ indicate the focus parameter associated with the least false parameter value η_0^{ML}, which minimizes the KL–divergence from the true model to the candidate model. If the candidate model captures the true model, we have $T_{0,\text{ML}} = T_{\text{true}}$.

Under the regularity conditions (C1)–(C4) (Jullum and Hjort, 2017), the joint distribution of $\widehat{T}_{\text{ML}}$ and $\widehat{T}_{\text{np}}$, can in our copula setting be written as

$$\sqrt{n} \begin{pmatrix} \widehat{T}_{\text{np}} - T_{\text{true}} \\ \widehat{T}_{\text{ML}} - T_{0,\text{ML}} \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \Lambda_{\text{np}} \\ \dot{s}_\eta^{\text{T}} \left(I_\eta^{\text{ML}} \right)^{-1} \Lambda_{\eta^{\text{ML}}} \end{pmatrix} \sim N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} V_{\text{np}} & V_{\text{np,ML}} \\ V_{\text{np,ML}}^{\text{T}} & V_{\text{ML}} \end{pmatrix} \right) \tag{3}$$

where

$$V_{\text{np}} = \text{Var}_G \left(\text{IF}_T(y, G) \right) = \text{E}_G \left[\text{IF}(y, G) \, \text{IF}(y, G)^{\text{T}} \right],$$

$$V_{\text{np,ML}} = Q_{\text{ML}} \left(I_\eta^{\text{ML}} \right)^{-1} \dot{s}_\eta, \quad Q_{\text{ML}} = \text{E}_G \left[\text{IF}_T(y, G) \cdot \phi_\eta^{\text{ML}}(y, \eta_0^{\text{ML}})^{\text{T}} \right], \quad \dot{s}_\eta = \frac{\partial s_\eta}{\partial \eta} = \frac{\partial T(F(\eta))}{\partial \eta},$$

$$V_{\text{ML}} = \dot{s}_\eta^{\text{T}} \left(I_\eta^{\text{ML}} \right)^{-1} K_\eta^{\text{ML}} \left(I_\eta^{\text{ML}} \right)^{-1} \dot{s}_\eta, \quad K_\eta^{\text{ML}} = \text{E}_G \left[\phi_\eta^{\text{ML}}(y, \eta_0^{\text{ML}}) \, \phi_\eta^{\text{ML}}(y, \eta_0^{\text{ML}})^{\text{T}} \right],$$

$$\phi_\eta^{\text{ML}}(y, \eta) = \frac{\partial \left(\sum_{j=1}^d f_j(y_j, \alpha_j) + \log c \left(F_1(y_1, \alpha_1), \cdots, F_d(y_d, \alpha_d), \theta \right) \right)}{\partial \eta},$$

$$I_\eta^{\text{ML}} = -\text{E}_G \left[H_\eta(y, \eta_0^{\text{ML}}) \right] = -\int g \, \frac{\partial^2 \left(\sum_{j=1}^d f_j(y_j, \alpha_{0,j}^{\text{ML}}) + \log c \left(F_1(y_1, \alpha_{0,1}^{\text{ML}}), \cdots, F_d(y_d, \alpha_{0,d}^{\text{ML}}), \theta_0^{\text{ML}} \right) \right)}{\partial \eta \, \partial \eta^{\text{T}}} \, dy$$

and $\text{IF}_T(y, G)$ is the influence function associated with the focus parameter. Based on (3), the FIC for ML estimated parametric

models is defined as

$$\mathrm{FIC}_{\mathrm{ML}} = \frac{1}{n}\widehat{V}_{\mathrm{ML}} + \max\left\{0, \left(\widehat{T}_{\mathrm{ML}} - \widehat{T}_{\mathrm{np}}\right)^2 - \frac{1}{n}\left(\widehat{V}_{\mathrm{ML}} + \widehat{V}_{\mathrm{np}} - 2\widehat{V}_{\mathrm{np,ML}}\right)\right\}, \tag{4}$$

where

$$\widehat{V}_{\mathrm{np,ML}} = \widehat{Q}_{\mathrm{ML}}\left(\widehat{\mathcal{I}}_{\eta}^{\mathrm{ML}}\right)^{-1}\dot{s}_{\widehat{\eta}}, \quad \widehat{V}_{\mathrm{ML}} = \dot{s}_{\widehat{\eta}}^{\mathrm{T}}\widehat{V}_{\eta}^{\mathrm{ML}}\widehat{\dot{s}} = \dot{s}_{\widehat{\eta}}^{\mathrm{T}}\left(\widehat{\mathcal{I}}_{\eta}^{\mathrm{ML}}\right)^{-1}\widehat{K}_{\eta}^{\mathrm{ML}}\left(\widehat{\mathcal{I}}_{\eta}^{\mathrm{ML}}\right)^{-1}\dot{s}_{\widehat{\eta}}.$$

Here $\widehat{V}_{\mathrm{np}}, \widehat{Q}_{\mathrm{ML}}, \widehat{\mathcal{I}}_{\eta}^{\mathrm{ML}}, \widehat{K}_{\eta}^{\mathrm{ML}}$ indicate empirical analogues of $V_{\mathrm{np}}, Q_{\mathrm{ML}}, \mathcal{I}_{\eta}^{\mathrm{ML}}, K_{\eta}^{\mathrm{ML}}$ respectively.

2.3 | FIC under two-stage maximum likelihood

When the dimension d of the model grows, the length of the total parameter vector η increases rapidly and ML estimation is not always feasible. Shih and Louis (1995) proposed two-stage ML (a.k.a. inference functions for margins, IFM) as an alternative. This consists in first estimating α_j separately for each margin by ML. Then, one plugs the resulting $\widehat{\alpha}^{\mathrm{2ML}}$ into the log-likelihood and obtains $\widehat{\theta}^{\mathrm{2ML}}$ by maximizing the log-likelihood with respect to θ. The asymptotic properties of the two-stage ML estimator are given in Ko and Hjort (2019b) for the general case or Joe (2005) for the situation where one assumes that the parametric model covers the true data generating mechanism.

Proposition 1 *Under regularity conditions (C1)–(C4) from Jullum and Hjort (2017) and (A1)–(A5) from Ko and Hjort (2019b),*

$$\sqrt{n}\begin{pmatrix}\widehat{T}_{\mathrm{np}} - T_{\mathrm{true}} \\ \widehat{T}_{\mathrm{2ML}} - T_{0,\mathrm{2ML}}\end{pmatrix} \xrightarrow{d} \begin{pmatrix}\Lambda_{\mathrm{np}} \\ \dot{s}_{\eta}^{\mathrm{T}}\left(\mathcal{I}_{\eta}^{\mathrm{2ML}}\right)^{-1}\Lambda_{\eta^{2\mathrm{ML}}}\end{pmatrix} \sim N\left(\begin{pmatrix}0 \\ 0\end{pmatrix}, \begin{pmatrix}V_{\mathrm{np}} & V_{\mathrm{np,2ML}} \\ V_{\mathrm{np,2ML}}^{\mathrm{T}} & V_{\mathrm{2ML}}\end{pmatrix}\right)$$

where

$$V_{\mathrm{np,2ML}} = Q_{\eta}^{\mathrm{2ML}}\left(\left(\mathcal{I}_{\eta}^{\mathrm{2ML}}\right)^{-1}\right)^{\mathrm{T}}\dot{s}_{\eta},$$

$$V_{\mathrm{2ML}} = \dot{s}_{\eta}^{\mathrm{T}}V_{\eta^{2\mathrm{ML}}}\dot{s}_{\eta},$$

$$V_{\eta^{2\mathrm{ML}}} = \left(\mathcal{I}_{\eta}^{\mathrm{2ML}}\right)^{-1}K_{\eta}^{\mathrm{2ML}}\left(\left(\mathcal{I}_{\eta}^{\mathrm{2ML}}\right)^{-1}\right)^{\mathrm{T}}$$

and other quantities are defined in the proof.

The proof is given in Appendix A.

In practice, V_{2ML} and $V_{\mathrm{np,2ML}}$ can be estimated by taking empirical analogues and using the plug-in principle. We write those estimates as $\widehat{V}_{\mathrm{2ML}}$ and $\widehat{V}_{\mathrm{np,2ML}}$. The FIC for two-stage ML estimated parametric models is then

$$\mathrm{FIC}_{\mathrm{2ML}} = \frac{1}{n}\widehat{V}_{\mathrm{2ML}} + \max\left\{0, \left(\widehat{T}_{\mathrm{2ML}} - \widehat{T}_{\mathrm{np}}\right)^2 - \frac{1}{n}\left(\widehat{V}_{\mathrm{2ML}} + \widehat{V}_{\mathrm{np}} - 2\widehat{V}_{\mathrm{np,2ML}}\right)\right\}. \tag{5}$$

2.4 | FIC under pseudo maximum likelihood combined with parametric margins

One of the drawbacks of (two-stage) ML estimation is that the estimation of the copula parameter θ is affected by the choice of parametric distribution for the margins. Even when the copula is correctly specified, possible misspecifications of the margins can

lead to suboptimal values of $\widehat{\theta}^{\mathrm{ML}}$ and $\widehat{\theta}^{\mathrm{2ML}}$. Genest et al. (1995) propose the PML estimator that overcomes this shortcoming. In this semi-parametric estimation scheme, one first uses the pseudo-observations $G_{n,j}(y_j) = (n+1)^{-1} \sum_{k=1}^{n} I\{y_{k,j} \leq y_j\}$, which are nonparametric probability integral transforms of the data. Subsequently, one plugs these into the copula log-likelihood and maximizes it with respect to θ to obtain $\widehat{\theta}^{\mathrm{PML}}$. For details concerning the PML estimator, see Tsukahara (2005), Chen and Fan (2005) and Kim et al. (2007).

Once θ is estimated, one can keep the nonparametric margins. However, parametric margins are often used instead, estimating α_j separately for each margin by ML, just like in stage 1 of two-stage ML estimation (Hobæk Haff et al., 2015; Bevacqua et al., 2017). Below, we derive the FIC formula for this 'PML combined with parametric margins' and abbreviate the estimation scheme by 'PMLpm' throughout this paper.

Proposition 2 *Under regularity conditions (C1)–(C5) from Chen and Fan (2005) and (C1)–(C4) from Jullum and Hjort (2017),*

$$\sqrt{n} \begin{pmatrix} \widehat{T}_{\mathrm{np}} - T_{\mathrm{true}} \\ \widehat{T}_{\mathrm{PMLpm}} - T_{0,\mathrm{PMLpm}} \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \Lambda_{\mathrm{np}} \\ \dot{s}_\eta^{\mathrm{T}} \left(\mathcal{I}_\eta^{\mathrm{PMLpm}} \right)^{-1} \Lambda_\eta^{\mathrm{PMLpm}} \end{pmatrix} \sim N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} V_{\mathrm{np}} & V_{\mathrm{np,PMLpm}} \\ V_{\mathrm{np,PMLpm}}^{\mathrm{T}} & V_{\mathrm{PMLpm}} \end{pmatrix} \right)$$

where

$$V_{\mathrm{np,PMLpm}} = Q_\eta^{\mathrm{PMLpm}} \left(\mathcal{I}_\eta^{\mathrm{PMLpm}} \right)^{-1} \dot{s}_\eta,$$

$$V_{\mathrm{PMLpm}} = \dot{s}_\eta^{\mathrm{T}} V_{\eta \mathrm{PMLpm}} \dot{s}_\eta,$$

$$V_{\eta}^{\mathrm{PMLpm}} = \left(\mathcal{I}_\eta^{\mathrm{PMLpm}} \right)^{-1} K_\eta^{\mathrm{PMLpm}} \left(\mathcal{I}_\eta^{\mathrm{PMLpm}} \right)^{-1}$$

and other quantities are defined in the proof.

The proof is given in Appendix A.

In practice, V_{PMLpm} and $V_{\mathrm{np,PMLpm}}$ can be estimated by plug-in estimators. For K_η^{PMLpm} and $\mathcal{W}(u, \theta)$, one can use the consistent estimators suggested in Remark 2 of Chen and Fan (2005). The resulting estimated variance and covariance are written as $\widehat{V}_{\mathrm{PMLpm}}$ and $\widehat{V}_{\mathrm{np,PMLpm}}$. Plugging these into (1) and (2) gives the FIC for PMLpm

$$\mathrm{FIC}_{\mathrm{PMLpm}} = \frac{1}{n} \widehat{V}_{\mathrm{PMLpm}} + \max \left\{ 0, \left(\widehat{T}_{\mathrm{PMLpm}} - \widehat{T}_{\mathrm{np}} \right)^2 - \frac{1}{n} \left(\widehat{V}_{\mathrm{PMLpm}} + \widehat{V}_{\mathrm{np}} - 2\widehat{V}_{\mathrm{np,PMLpm}} \right) \right\}. \tag{6}$$

3 | AIC FOR PSEUDO MAXIMUM LIKELIHOOD WITH PARAMETRIC MARGINS

Since the eventual model form of PMLpm is fully parametric, we can use the Kullback–Leibler (KL) divergence to evaluate the relative closeness of the candidate model to the true model. Minimizing KL divergence is equal to maximizing

$$Q(\eta) = \int g(y) \left\{ \log f_1(y_1, \alpha_1) + \cdots + \log f_d(y_d, \alpha_d) + \log c \left(F_1(y_1, \alpha_1), \cdots, F_d(y_d, \alpha_d), \theta \right) \right\} dy.$$

The empirical equivalent is

$$\widehat{Q}(\eta) = \frac{1}{n} \sum_{i=1}^{n} \left[\log f_1(y_{i,1}, \alpha_1) + \cdots + \log f_d(y_{i,d}, \alpha_d) + \log c \left(F_1(y_{i,1}, \alpha_1), \cdots, F_d(y_{i,d}, \alpha_d), \theta \right) \right]$$

which is biased. AIC is defined as bias corrected and rescaled estimator of $Q(\eta)$ under the true model assumption. We derive AIC for PMLpm estimated parametric model.

Lemma 3 *Assuming that the parametric model captures the model that generated the data, i.e. $f(\cdot, \eta_0) = g(\cdot)$ for an appropriate η_0, an approximately unbiased empirical estimator of $2n \cdot Q(\widehat{\eta}^{\text{PMLpm}})$, widely known as Akaike information criterion (AIC), is*

$$\text{AIC} = 2\ell_n(\widehat{\eta}^{\text{PMLpm}}) - 2\dim(\eta).$$

The proof is given in Appendix A.

Since the AIC for the ML, two-stage ML and PMLpm estimators are all aiming for the same quantity under the model conditions, they are compatible. This means that AIC can be used to compare closeness to the true model for different estimation schemes.

4 | AVERAGE WEIGHTED FOCUSED INFORMATION CRITERION

The FIC apparatus is designed to select the best model for estimating a chosen focus parameter precisely. However, there are cases where one is interested in multiple focus parameters. Jullum and Hjort (2017) developed an averaged weighted FIC (AFIC) which aims for the model that obtains lowest risk for a set of focus parameters $T_{\text{true}}(t)$, with t in some index set. This set of focus parameters is specified with a cumulative weight function $W(t)$. Under this setting, they use the weighted quadratic loss function

$$L = \int \left(\widehat{T}_{\text{cnd}}(t) - T_{\text{true}}(t)\right)^2 dW(t),$$

which results in the following risk:

$$\text{E}_G[L] = \int \text{E}_G\left[\left(\widehat{T}_{\text{cnd}}(t) - T_{\text{true}}(t)\right)^2\right] dW(t).$$

Then, they define the AFIC as

$$\text{AFIC} = \int \widehat{\text{Var}}_G\left(\widehat{T}_{\text{cnd}}(t)\right) dW(t) + \widehat{\text{bias}}^2_{\text{cnd,true}}. \tag{7}$$

where the bias term is defined as

$$\widehat{\text{bias}}^2_{\text{cnd,true}} = \begin{cases} \max\left\{0, \int \left(\widehat{T}_{\text{cnd}}(t) - \widehat{T}_{\text{np}}(t)\right)^2 - \widehat{\text{Var}}_G\left(\widehat{T}_{\text{cnd}}(t)\right) & \text{if (semi)parametric} \\ \qquad -\widehat{\text{Var}}_G\left(\widehat{T}_{\text{np}}(t)\right) + 2\widehat{\text{Cov}}_G\left(\widehat{T}_{\text{np}}(t), \widehat{T}_{\text{cnd}}(t)\right)\right\} dW(t) & \\ 0 & \text{if nonparametric.} \end{cases} \tag{8}$$

Note that when the set of estimands $T_{\text{true}}(t)$ is finite and the weight function is not stochastic, the integral with respect to $W(t)$ simply becomes a weighted sum. For a practical illustration of AFIC, see the supplementary material of Jullum and Hjort (2017).

5 | SIMULATION STUDY

To study the behavior of FIC, we have carried out a simulation study in two parts. The performance of the FIC with a joint box probability as focus parameter is assessed in the Part 1 simulations. The purpose of the Part 2 simulations is to evaluate the FIC

TABLE 1 Description of the models used in Part 1 of the simulations.

	Copula	Margin 1	Margin 2	Margin 3	Margin 4
Data generating model	Gumbel $\theta = 3$	Gamma $\alpha_1 = (1, 2)^{\mathrm{T}}$ (shape, rate)	Gamma $\alpha_2 = (3, 1)^{\mathrm{T}}$ (shape, rate)	Weibull $\alpha_3 = (1, 2)^{\mathrm{T}}$ (shape, scale)	Log-normal $\alpha_4 = (-1, 0.5)^{\mathrm{T}}$ (mean, sd)
Model 1	Gumbel	Gamma	Gamma	Weibull	Log-normal
Model 2	Gumbel	Weibull	Weibull	Gamma	Gamma
Model 3	Survival Clayton	Gamma	Gamma	Weibull	Log-normal
Model 4	Survival Clayton	Weibull	Weibull	Gamma	Gamma
Model 5	Frank	Gamma	Gamma	Weibull	Log-normal
Model 6	Frank	Weibull	Weibull	Gamma	Gamma

with other focus parameters, more specifically conditional probabilities, as well as the AFIC.

5.1 | Part 1: Evaluation of the FIC

We have generated datasets from a model consisting of a Gumbel copula, two Gamma, one Weibull and one log-normal margins and fitted 6 candidate models for each dataset using the ML, two-stage ML and PMLpm estimators. Further, the sample size is either $n = 100$ or $n = 1000$. The data generating and candidate models are described in Table 1. With each fitted candidate model, we have computed the focus parameter $T(G) = P(q_{0.8} < Y)$ where $q_{0.8}$ is a vector that contains the 0.8-quantile value of each margin according to the true model. i.e. $q_{0.8} = \left(G_1^{-1}(0.8), \cdots, G_4^{-1}(0.8)\right)$. In addition, we have estimated the focus parameter nonparametrically, which is necessary for the FIC machinery. We have repeated this process 100 times and the results are averaged. Since the natural empirical estimator for $\mathcal{W}$ from Chen and Fan (2005), needed for PMLpm, is computationally very expensive for a large sample size, we have estimated the asymptotic variance from Proposition 2 numerically with jackknife (Efron, 1982). We have compared the jackknife estimated variance to the Chen and Fan (2005) based estimate for small sample sizes, and they produced comparable results.

The results from the simulations can be found in Tables 1 and 2 from the supplementary material (Ko et al., 2019). The result is also visualized in Figure 1 for $n = 100$ and in Figure 2 for $n = 1000$.

The top left panel of Figure 1 is a plot of the estimated focus parameter $\widehat{T}$ against $\sqrt{\mathrm{FIC}}$, which is on the same scale as $\widehat{T}$. We see that the models with smaller $\sqrt{\mathrm{FIC}}$ value tend to give focus parameter estimates that are closer to the true focus parameter value. This tendency becomes clearer as the sample size increases (see Figure 2).

In the top right panel of Figure 1, we have plotted $\sqrt{\mathrm{MSE}}$ against $\sqrt{\mathrm{FIC}}$. We can observe that the FIC provides a good estimate of the MSE. As the sample size increases (Figure 2), the correspondence is even better.

The two panels at the bottom of Figure 1 show the AIC plotted against the FIC (left panel) and the MSE (right panel) for $n = 100$. Note that models with large scores are preferred when using the AIC. We see that both the AIC and the FIC pick the correctly specified model (model 1) as the best model. The AIC rank for models 5 and 6 indicate implies that those are overall rather close to the true model (relatively to other candidates). However, the FIC suggests that they do not estimate $P(q_{0.8} < Y)$ well (again in comparison to the other candidate models). The estimated focus parameter values and MSEs indicate that this is indeed the case. This illustrates that the FIC does choose models that are good at a specific task, as intended. Figure 2 shows that the same holds when the sample size increases.
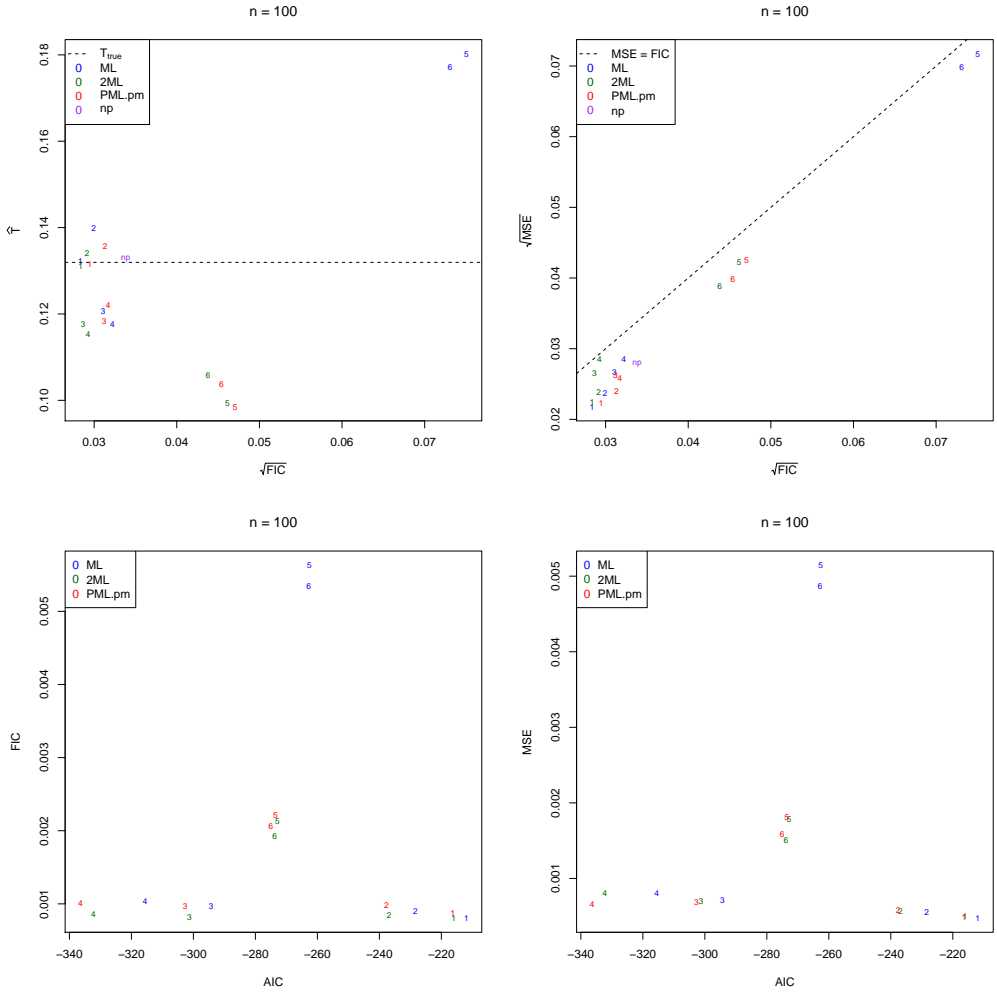
FIGURE 1   Results from Part 1 of the simulations for $n = 100$. The colored numbers refer to the model numbers in Table 1. The nonparametric model is denoted "np".

One of the noticeable changes as the sample size increases from $n = 100$ to $n = 1000$ is that the rank of the nonparametric model decreases from 13 to 5. The reason is that the (co)variance terms shrink toward 0, such that the asymptotic unbiasedness of nonparametric model becomes stronger advantage. When n is large enough, the (co)variance terms will be negligible and the nonparametric model will be the overall 'winning' model.

Further, the focus parameter values from models estimated with two-stage ML and PMLpm tend to be better than the ones obtained with ML when the model is misspecified, with correspondingly lower FIC scores. The differences are less systematic when we compare two-stage ML and PMLpm.
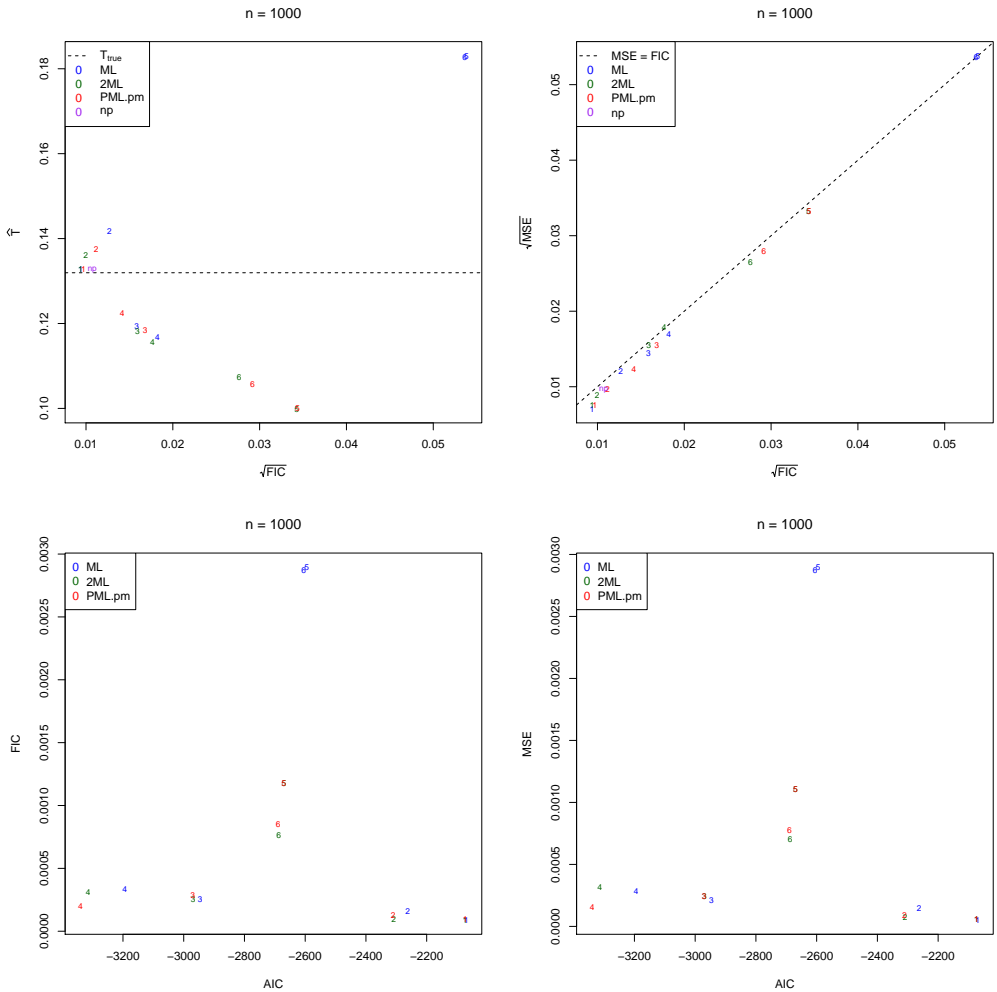
FIGURE 2 Results from Part 1 of the simulations for $n = 1000$. See Figure 1 for an explanation of symbols.

5.2 | Part 2: Evaluation of the AFIC

In Part 2 of the simulations, we have generated data sets from a two-dimensional model consisting of a Gaussian copula and log-normal margins and fitted 12 candidate models by using ML, two-stage ML and PMLpm estimators. The data generating and candidate models are described in Table 2 and the sample size is $n = 1000$. With each fitted candidate model, we have computed a set of focus parameters $T(F) = P(G_1^{-1}(p) < Y_1 \mid G_2^{-1}(p) < Y_2)$ for $p = 0.90, 0.91, \cdots, 0.95$. When p is close to 1, a reasonable precision of the nonparametric estimate of this conditional probability requires a very large sample size to ensure that there are enough observations that satisfy the inequality conditions. Therefore, we did not include probabilities that are even closer to 1. In addition, we have computed the AFIC by assigning equal weights to each p. The whole process was repeated 100 times and the results are averaged. As for the Part 1 simulations, the asymptotic variance of PMLpm is estimated with jackknife.

The results from the simulations with the AFIC, as well as with the FIC with single focus parameter for $p = 0.90$ and

**TABLE 2** Description of the models used in Part 2 of the simulations.

	Copula	Margin 1	Margin 2
Data generating model	Gaussian $\theta = 0.6$	Log-normal $\alpha_1 = (1, 0.8)^{\mathrm{T}}$ (mean, sd)	Log-normal $\alpha_2 = (0.4, 0.7)^{\mathrm{T}}$ (mean, sd)
Model 1	Gaussian	Log-normal	Log-normal
Model 2	Gaussian	Gamma	Gamma
Model 3	Gaussian	Weibull	Weibull
Model 4	Frank	Log-normal	Log-normal
Model 5	Frank	Gamma	Gamma
Model 6	Frank	Weibull	Weibull
Model 7	Gumbel	Log-normal	Log-normal
Model 8	Gumbel	Gamma	Gamma
Model 9	Gumbel	Weibull	Weibull
Model 10	Survival Clayton	Log-normal	Log-normal
Model 11	Survival Clayton	Gamma	Gamma
Model 12	Survival Clayton	Weibull	Weibull

$p = 0.95$ can be found in Tables 3, 4 and 5 from the supplementary material (Ko et al., 2019). The results are also visualized in Figures 3 and 4, together with Figure 1 from the supplementary material (Ko et al., 2019). The top left panel of Figure 3 is a plot of $\sqrt{\text{FIC}}$ against the focus parameter estimates $\widehat{T}$ for the FIC with $p = 0.9$, and the top right panel shows $\sqrt{\text{FIC}}$ against $\sqrt{\text{MSE}}$. The bottom left and right panels display AIC against FIC and AIC against MSE, respectively. Figures 4 show corresponding results for the FIC with $p = 0.95$. The corresponding result for AFIC is displayed in Figure 1 from the supplementary material (Ko et al., 2019).

When we consider a single focus parameter with $p = 0.90$, the models with smaller $\sqrt{\text{FIC}}$ score tend to have $\widehat{T}$ values closer to the true focus parameter value, as anticipated. Moreover, the FIC is overall a good estimate of the MSE. When $p = 0.95$, the patterns are the same. However, the wrongly specified models now have larger FIC and MSE values. This is logical since the upper conditional probability, which becomes upper tail dependence when $p \to 1$, differs highly from model to model as p increases. Further, there is little correspondence between the models preferred by the AIC and the FIC, except that they both pick the true model as the best model. Again, the models that are favored by the AIC are not necessarily good at estimating the upper tail conditional probability, though they may be overall closer to the true model. This confirms that there is a clear gain in using the FIC when the model is to be used for estimating a specific quantity. The results for the AFIC are in line with the previous results with a single focus parameter.

6 | EXAMPLE: THE ABALONE DATA

As a real-life data example, we applied our FIC model selection method to the abalone data set (Asuncion and Newman, 2011), which has previously been used to illustrate different dependence modeling methods e.g. Ma et al. (2012) and Hobæk Haff et al.
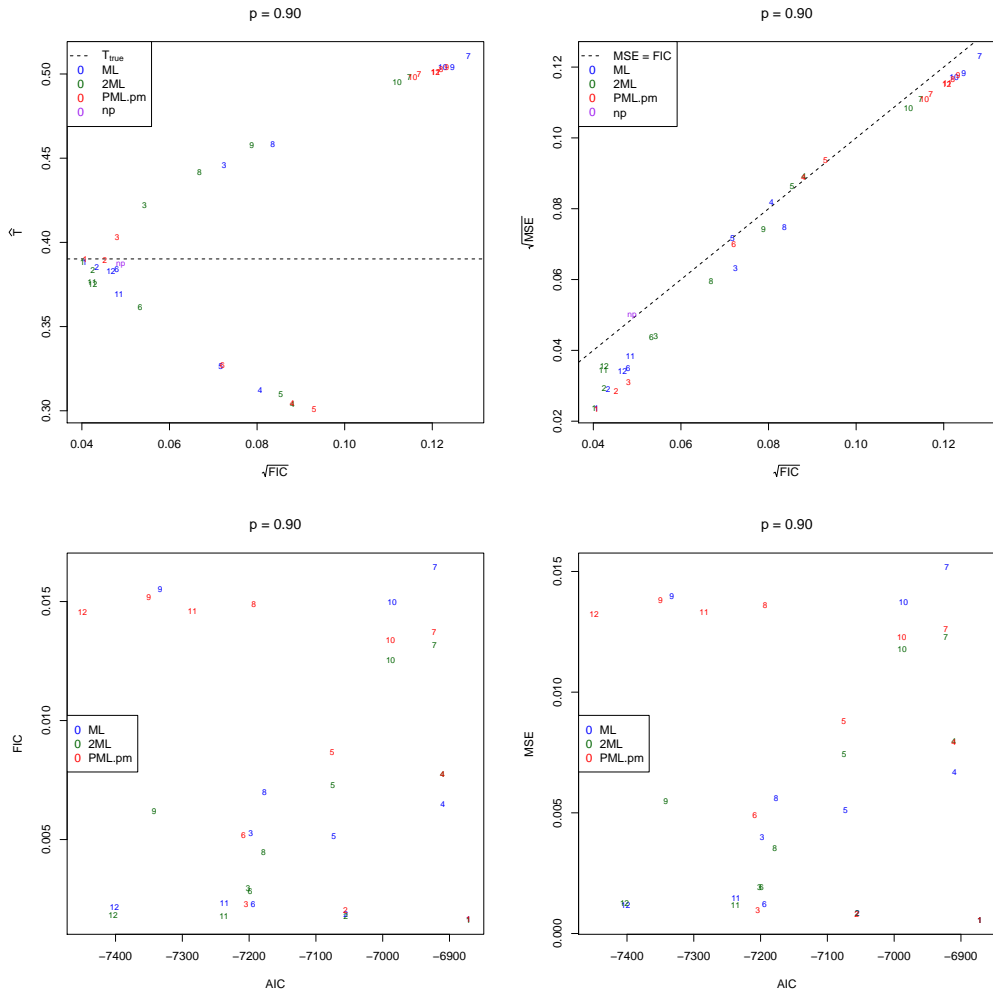
FIGURE 3   Results from Part 2 of the simulations for the FIC with $p = 0.90$. The colored numbers refer to the model numbers in Table 2.

(2016). Abalone is an edible sea snail, the harvest of which is subject to quotas. The quotas should reflect the age distribution of abalones and is based on among others their size. As the age of an abalone is cumbersome to determine, one would like to estimate the age based on some physical measurements such as weight and height. The dataset consists of 4177 samples of 9 variables. We focus on the dependence relations among the 4 variables 'diameter', 'height', 'shell weight' and 'rings'. According to the description of the dataset, the age in years is given by the number of rings $+ 1.5$. So we convert the variable 'ring' into 'age' by adding this constant. Further, we make the 'age' variable continuous by adding Gaussian noise $N(0, 0.01^2)$.

Since the harvest quotas are related to the age of abalones as well as their size, an interesting focus parameter to consider is the probability that an abalone is under certain age given that physical measurements are smaller than certain bounds. More
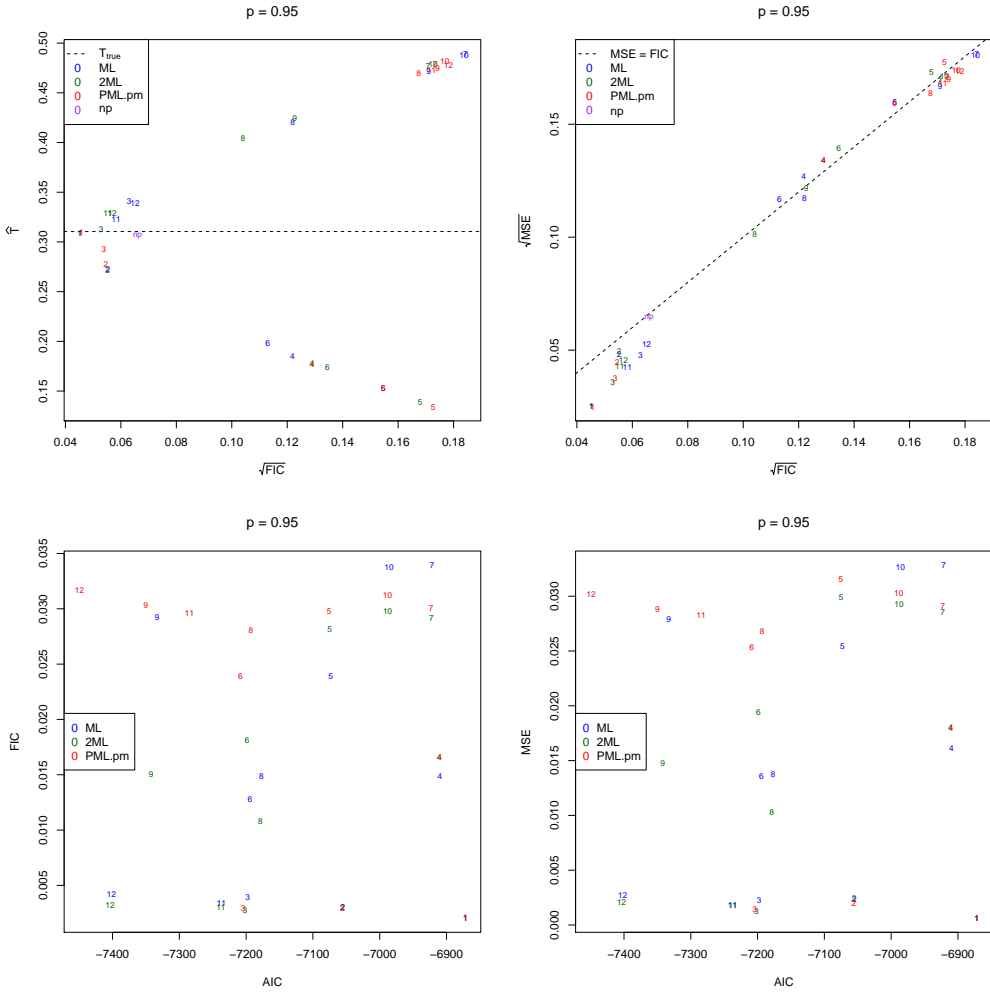
FIGURE 4 Results from Part 2 of the simulations for the FIC with $p = 0.95$. The colored numbers refer to the model numbers in Table 2.

specifically, we used the focus parameter

$$T(F) = P(\text{age} < 8 \mid \text{diameter} < 0.325, \ \text{height} < 0.105, \ \text{shellweight} < 0.109).$$

Here, 0.325, 0.105 and 0.109 were sample 20%-quantiles values of 'diameter', 'height' and 'shell weight', respectively, whereas 8 is the 27%-quantile of 'age'.

We have tried the 9 candidate models described in Table 3. To choose adequate candidate margins, we fitted a set of well-known univariate distributions with ML estimator. We then evaluated them by AIC. For copula, we fitted different copula models with the PML estimator and looked at maximized likelihood values. We composed Table 3 by choosing certain combinations of

TABLE 3   Description of the models used to fit the abalone data set.

|	Copula	Margin 1	Margin 2	Margin 3	Margin 4
Model 1	t	Weibull	Weibull	Weibull	NIG
Model 2	t	Weibull	Weibull	Weibull	Log-normal
Model 3	t	Normal	Normal	Normal	Log-normal
Model 4	Gaussian	Weibull	Weibull	Weibull	NIG
Model 5	Gaussian	Weibull	Weibull	Weibull	Log-normal
Model 6	Gaussian	Normal	Normal	Normal	Log-normal
Model 7	Clayton	Weibull	Weibull	Weibull	NIG
Model 8	Clayton	Weibull	Weibull	Weibull	Log-normal
Model 9	Clayton	Normal	Normal	Normal	Log-normal

copula and margins. Each model in Table 3 is estimated with the two-stage ML and the PMLpm estimators. The ML estimator encountered numeric problems for some models.

The results are shown in Table 6 from the supplementary material (Ko et al., 2019) and also visualized in Figure 5. We see a similar pattern to the ones from the simulation study. One of the notable differences is that the nonparametric model is now the winning model. This seems reasonable considering that the nonparametric model becomes quite precise when the sample size is large enough. This is also in line with results from Part 1 of the simulations, where the FIC rank of the nonparametric model decreased from 13 to 5 when the sample size increased from 100 to 1000. Further, scatter plots of the pseudo-observations (not shown here) indicate that the copula is rather different from well-known 4-dimensional parametric copulas, with both asymmetric dependence and large differences between pairs. This also explains the preference for the nonparametric model. The winning models among the parametric ones, according the the FIC, are the ones based on the Gaussian or the t copula, that, as opposed to the Clayton, allow for different dependence between different pairs. Finally, the right panel shows a moderate agreement between the AIC and the FIC.

7 | CONCLUSIONS AND FURTHER RESEARCH

In this paper, we have developed the FIC for copula models partly based on the general methods of Jullum and Hjort (2017) for three different estimators, namely the ML, two-stage ML and PMLpm. This is a model selection criterion that aims for the model that minimizes the estimated MSE of a chosen focus parameter, and has the advantage that it can be used to compare models that are estimated under different estimation schemes, including the nonparametric one. It can also easily be extended to other estimation schemes as long as the joint asymptotic distribution with the nonparametric focus parameter estimate can be derived. In addition, it is a model-robust model selection criterion since it does not assume that any of the candidate model captures the data generating model.

We performed a series of simulations to study the behavior of the FIC. We also applied our method to the abalone data set. The results show that models with lower FIC values give focus parameter estimates that are closer to the true parameter value, which is the aim of the method. Further, it turns out that FIC also fulfills its role as an estimator of the MSE. Moreover, when the focus parameter is heavily based on a specific part of the distribution, in particular the tails, the models favored by the AIC does not necessarily give good estimates of the focus parameter. This seems natural since AIC aims for the model that minimizes
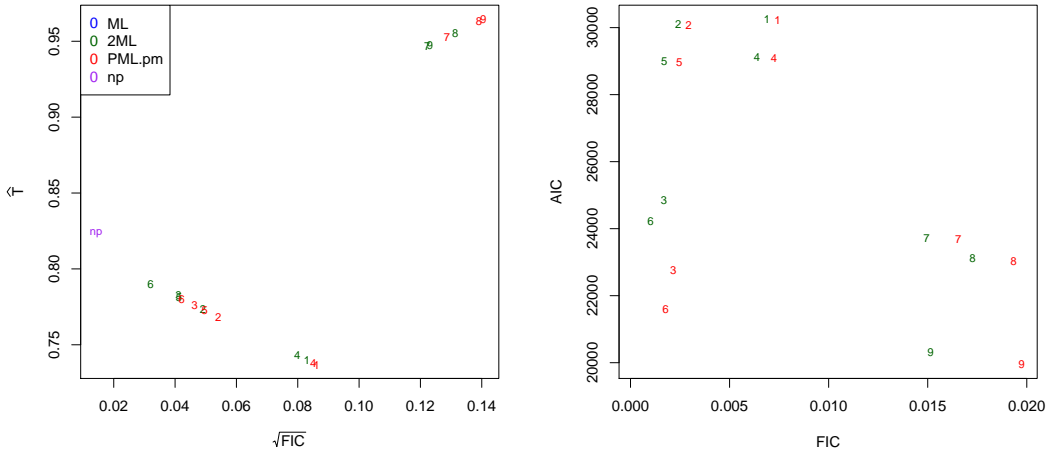
FIGURE 5 Results from the models fitted to the abalone data. The colored numbers refer to the model numbers in Table 3.

overall closeness to the true model in terms of KL–divergence, and also demonstrates that there is a clear advantage in using the FIC instead of the AIC in such situations.

Even though the FIC machinery can be applied to many different kinds of focus parameters, a limitation of the method is that it needs the focus parameter to be a smooth functional of the full distribution G, so that an influence function can be derived. For some functionals, like the density itself, or the copula function c_0 in a given location, there is no influence function, and more elaborate estimators are called for, involving smoothing parameters. Versions of FIC may still be put up, but requires further elaboration; see Jullum and Hjort (2017, Section 7).

As we see from the FIC plots (Figures 1, 2, 3 and 4), often the estimates from the best models are in reasonable agreement, and a final estimate can be based on a suitable average over these. Claeskens and Hjort (2008, Chapter 7) propose and develop machinery for such model average estimators, and suggest using weights proportional to $\exp(-\lambda \cdot \text{FIC})$, with λ a tuning parameter. We do not pursue that theme in this paper, however.

In this paper, we have only considered parametric copula models with parametric margins and i.i.d. variables. One can for example extend the FIC framework to the PML estimator where margins are nonparametric, as well as to time series and vine models. Furthermore, it is possible to derive another variant of FIC by using another loss function than the quadratic loss function (e.g. absolute loss).

ACKNOWLEDGMENTS

REFERENCES

Akaike, H. (1974) A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, **19**, 716–723.

Asuncion, A. and Newman, D. (2011) UCI Machine Learning Repository, University of California, Irvine, School of Information and Computer Sciences. URL: `https://archive.ics.uci.edu/ml/datasets.html`.

Bevacqua, E., Maraun, D., Hobæk Haff, I., Widmann, M. and Vrac, M. (2017) Multivariate statistical modelling of compound events via pair-copula constructions: analysis of floods in Ravenna (Italy). *Hydrology and Earth System Sciences*, **21**, 2701–2723.

Chen, X. and Fan, Y. (2005) Pseudo-likelihood ratio tests for semiparametric multivariate copula model selection. *Canadian Journal of Statistics*, **33**, 389–414.

Claeskens, G. and Hjort, N. L. (2003) The focused information criterion [with discussion contributions and a rejoinder]. *Journal of the American Statistical Association*, **98**, 900–916.

Claeskens, G. and Hjort, N. L. (2008) *Model Selection and Model Averaging*. Cambridge University Press.

Efron, B. (1982) *The Jackknife, the Bootstrap, and Other Resampling Plans*, vol. 38. Siam.

Genest, C., Ghoudi, K. and Rivest, L.-P. (1995) A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika*, **82**, 543–552.

Grønneberg, S. and Hjort, N. L. (2014) The copula information criteria. *Scandinavian Journal of Statistics*, **41**, 436–459.

Hardin, J. W. (2003) The sandwich estimate of variance. In *Maximum likelihood estimation of misspecified models: Twenty years later*, 45–73. Emerald Group Publishing Limited.

Hobæk Haff, I., Aas, K., Frigessi, A. and Lacal, V. (2016) Structure learning in Bayesian networks using regular vines. *Computational Statistics & Data Analysis*, **101**, 186–208.

Hobæk Haff, I., Frigessi, A. and Maraun, D. (2015) How well do regional climate models simulate the spatial dependence of precipitation? an application of pair-copula constructions. *Journal of Geophysical Research: Atmospheres*, **120**, 2624–2646.

Joe, H. (1997) *Multivariate Models and Multivariate Dependence Concepts*. CRC Press.

Joe, H. (2005) Asymptotic efficiency of the two-stage estimation method for copula-based models. *Journal of Multivariate Analysis*, **94**, 401–419.

Joe, H. (2014) *Dependence Modeling with Copulas*. CRC Press.

Jullum, M. and Hjort, N. L. (2017) Parametric or nonparametric: the FIC approach. *Statistica Sinica*, **27**, 951–981.

Jullum, M. and Hjort, N. L. (2018) What price semiparametric Cox regression? *Lifetime data analysis*, 1–33. URL: `https://doi.org/10.1007/s10985-018-9450-7`.

Kim, G., Silvapulle, M. J. and Silvapulle, P. (2007) Comparison of semiparametric and parametric methods for estimating copulas. *Computational Statistics & Data Analysis*, **51**, 2836–2850.

Ko, V. and Hjort, N. L. (2019a) Copula information criterion for model selection with two-stage maximum likelihood estimation. *Econometrics and Statistics*. (in press).

Ko, V. and Hjort, N. L. (2019b) Model robust inference with two-stage maximum likelihood estimation for copulas. *Journal of Multivariate Analysis*. (in press).

Ko, V., Hjort, N. L. and Hobæk Haff, I. (2019) Supplement to "Focused information criteria for copulas". *Scandinavian Journal of Statistics*.

Le Cam, L. (1990) Maximum likelihood: an introduction. *International Statistical Review*, **58**, 153–171.

Ma, J., Sun, Z.-Q., Chen, S. and Liu, H.-H. (2012) Dependence tree structure estimation via copula. *International Journal of Automation and Computing*, **9**, 113–121.

Nelsen, R. B. (2006) *An Introduction to Copulas*. Springer Science & Business Media.

Shao, J. (2003) *Mathematical Statistics*. Springer Texts in Statistics. Springer.

Shih, J. H. and Louis, T. A. (1995) Inferences on the association parameter in copula models for bivariate survival data. *Biometrics*, **51**, 1384–1399.

Sklar, A. (1959) Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris*, **8**, 229–231.

Tsukahara, H. (2005) Semiparametric estimation in copula models. *Canadian Journal of Statistics*, **33**, 357–375.

White, H. (1982) Maximum likelihood estimation of misspecified models. *Econometrica: Journal of the Econometric Society*, 1–25.

APPENDIX A: PROOFS

| Proof of Proposition 1

Proof Under the assumed regularity conditions, we have from Theorem 5.5 in Shao (2003) and Proposition 1 in Ko and Hjort (2019b) that

$$\widehat{T}_{\text{np}} - T_{\text{true}} = \frac{1}{n} \sum_{i=1}^{n} \text{IF}_T(y_i, G) + o_p(n^{-1/2}),$$

$$\widehat{\eta}^{\text{2ML}} - \eta_0^{\text{2ML}} = (\mathcal{I}_\eta^{\text{2ML}})^{-1} \phi_{n,\eta}^{\text{2ML}}(\eta_0^{\text{2ML}}) + o_p(n^{-1/2})$$

where

$$\phi_{n,\eta}^{\text{2ML}}(\eta) = \frac{1}{n} \sum_{i=1}^{n} \phi_\eta^{\text{2ML}}(y_i, \eta),$$

$$\phi_{n,\alpha}^{\text{2ML}}(\alpha) = \frac{1}{n} \sum_{i=1}^{n} \phi_\alpha^{\text{2ML}}(y_i, \alpha),$$

$$\phi_{n,\theta}^{\text{2ML}}(\theta) = \frac{1}{n} \sum_{i=1}^{n} \phi_\theta^{\text{2ML}}(y_i, \theta),$$

$$\phi_\eta^{\text{2ML}}(\eta) = \begin{pmatrix} \phi_\alpha^{\text{2ML}}(\alpha) \\ \phi_\theta^{\text{2ML}}(\eta) \end{pmatrix},$$

$$\phi_\alpha^{\text{2ML}}(y, \alpha) = \begin{pmatrix} \phi_{\alpha_1}^{\text{2ML}}(y_1, \alpha_1) \\ \vdots \\ \phi_{\alpha_d}^{\text{2ML}}(y_d, \alpha_d) \end{pmatrix} = \begin{pmatrix} \partial \log f_1(y_1, \alpha_1)/\partial \alpha_1 \\ \vdots \\ \partial \log f_d(y_d, \alpha_d)/\partial \alpha_d \end{pmatrix},$$

$$\phi_\theta^{\text{2ML}}(y, \eta) = \frac{\partial \log c \left(F_1(y_1, \alpha_1), \cdots, F_d(y_d, \alpha_d), \theta \right)}{\partial \theta},$$

$$\mathcal{I}_\eta^{\text{2ML}} = \begin{pmatrix} \mathcal{I}_\alpha^{\text{2ML}} & 0 \\ (\mathcal{I}_{\alpha,\theta}^{\text{2ML}})^{\text{T}} & \mathcal{I}_\theta^{\text{2ML}} \end{pmatrix},$$

$$\mathcal{I}_\alpha^{\text{2ML}} = \begin{pmatrix} \mathcal{I}_{\alpha_1}^{\text{2ML}} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \mathcal{I}_{\alpha_d}^{\text{2ML}} \end{pmatrix},$$

$$\mathcal{I}_{\alpha_j}^{\text{2ML}} = -\text{E}_{G_j} \left[H_{\alpha_j}^{\text{2ML}}(y_j, \alpha_{0,j}^{\text{2ML}}) \right] = -\int g_j \frac{\partial^2 \log f_j(y_j, \alpha_{0,j}^{\text{2ML}})}{\partial \alpha_j \, \partial \alpha_j^{\text{T}}} \, dy_j,$$

$$\mathcal{I}_\theta^{\text{2ML}} = -\text{E}_G \left[H_\theta^{\text{2ML}}(y, \eta_0^{\text{2ML}}) \right] = -\int g \frac{\partial^2 \log c \left(F_1(y_1, \alpha_{0,1}^{\text{2ML}}), \cdots, F_d(y_d, \alpha_{0,d}^{\text{2ML}}), \theta_0^{\text{2ML}} \right)}{\partial \theta \, \partial \theta^{\text{T}}} \, dy,$$

$$\mathcal{I}_{\alpha,\theta}^{\text{2ML}} = -\text{E}_G \left[H_{\alpha,\theta}^{\text{2ML}}(y, \eta_0^{\text{2ML}}) \right] = -\int g \frac{\partial^2 \log c \left(F_1(y_1, \alpha_{0,1}^{\text{2ML}}), \cdots, F_d(y_d, \alpha_{0,d}^{\text{2ML}}), \theta_0^{\text{2ML}} \right)}{\partial \alpha \, \partial \theta^{\text{T}}} \, dy.$$

Applying the multivariate central limit theorem to the summand yields the joint distribution

$$\sqrt{n} \begin{pmatrix} \widehat{T}_{\text{np}} - T_{\text{true}} \\ \widehat{\eta}^{\text{2ML}} - \eta_0^{\text{2ML}} \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \Lambda_{\text{np}} \\ \left(\mathcal{I}_\eta^{\text{2ML}} \right)^{-1} \Lambda_{\eta^{\text{2ML}}} \end{pmatrix} \sim N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} V_{\text{np}} & V_{\text{np},\eta^{\text{2ML}}} \\ \left(V_{\text{np},\eta^{\text{2ML}}} \right)^{\text{T}} & V_{\eta^{\text{2ML}}} \end{pmatrix} \right)$$

where

$$V_{\mathrm{np},\eta^{\mathrm{2ML}}} = Q_\eta^{\mathrm{2ML}} \left(\left(\mathcal{I}_\eta^{\mathrm{2ML}} \right)^{-1} \right)^{\mathrm{T}},$$

$$V_{\eta^{\mathrm{2ML}}} = \left(\mathcal{I}_\eta^{\mathrm{2ML}} \right)^{-1} K_\eta^{\mathrm{2ML}} \left(\left(\mathcal{I}_\eta^{\mathrm{2ML}} \right)^{-1} \right)^{\mathrm{T}},$$

$$Q_\eta^{\mathrm{2ML}} = \mathrm{Cov}_G \left(\mathrm{IF}_T(y, G), \phi_\eta^{\mathrm{2ML}}(y, \eta_0^{\mathrm{2ML}}) \right) = \mathrm{E}_G \left[\mathrm{IF}_T(y, G) \cdot \phi_\eta^{\mathrm{2ML}}(y, \eta_0^{\mathrm{2ML}})^{\mathrm{T}} \right],$$

$$K_\eta^{\mathrm{2ML}} = \mathrm{Var}_G \left(\phi_\eta^{\mathrm{2ML}}(y, \eta_0^{\mathrm{2ML}}) \right) = \mathrm{E}_G \left[\phi_\eta^{\mathrm{2ML}}(y, \eta_0^{\mathrm{2ML}}) \phi_\eta^{\mathrm{2ML}}(y, \eta_0^{\mathrm{2ML}})^{\mathrm{T}} \right].$$

Now by applying the delta method with transformation function $s_{x,\eta}(x, \eta) = (x, s_\eta)^{\mathrm{T}} = (x, T(F(\eta)))^{\mathrm{T}}$, we obtain

$$\sqrt{n} \begin{pmatrix} \widehat{T}_{\mathrm{np}} - T_{\mathrm{true}} \\ \widehat{T}_{\mathrm{2ML}} - T_{0,\mathrm{2ML}} \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \Lambda_{\mathrm{np}} \\ \dot{s}_\eta^{\mathrm{T}} \left(\mathcal{I}_\eta^{\mathrm{2ML}} \right)^{-1} \Lambda_{\eta^{\mathrm{2ML}}} \end{pmatrix} \sim N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} V_{\mathrm{np}} & V_{\mathrm{np},\mathrm{2ML}} \\ V_{\mathrm{np},\mathrm{2ML}}^{\mathrm{T}} & V_{\mathrm{2ML}} \end{pmatrix} \right).$$

| Proof of Proposition 2

Proof Under the assumed regularity conditions, we have from Theorem 5.5 in Shao (2003), Lemma 1 in Ko and Hjort (2019b) and Proposition 2 from Chen and Fan (2005) that

$$\widehat{T}_{\mathrm{np}} - T_{\mathrm{true}} = \frac{1}{n} \sum_{i=1}^{n} \mathrm{IF}_T(y_i, G) + o_p(n^{-1/2}),$$

$$\widehat{\alpha}^{\mathrm{2ML}} - \alpha_0^{\mathrm{2ML}} = (\mathcal{I}_\alpha^{\mathrm{2ML}})^{-1} \phi_{n,\alpha}^{\mathrm{2ML}}(\alpha_0^{\mathrm{2ML}}) + o_p(n^{-1/2}),$$

$$\widehat{\theta}^{\mathrm{PML}} - \theta_0^{\mathrm{PML}} = (\mathcal{I}_\theta^{\mathrm{PML}})^{-1} \phi_{n,\theta}^{\mathrm{PML}}(\theta_0^{\mathrm{PML}}) + o_p(n^{-1/2}),$$

where

$$\phi_{n,\theta}^{\mathrm{PML}}(\theta) = \frac{1}{n} \sum_{i=1}^{n} \phi_\theta^{\mathrm{PML}}(G_{n,1}(y_{i,1}), \cdots, G_{n,d}(y_{i,d}), \theta),$$

$$G_{n,j}(y_j) = \frac{1}{n+1} \sum_{k=1}^{n} I\{y_{k,j} \le y_j\},$$

$$\phi_\theta^{\mathrm{PML}}(u, \theta) = \frac{\partial \log c \left(u_1, \cdots, u_d \right), \theta)}{\partial \theta},$$

$$\mathcal{I}_\theta^{\mathrm{PML}} = -\mathrm{E}_{C_0} \left[H_\theta^{\mathrm{PML}}(u, \theta_0^{\mathrm{PML}}) \right] = - \int_{[0,1]^d} c_0(u) \frac{\partial^2 \log c \left(u, \theta_0^{\mathrm{PML}} \right)}{\partial \theta \, \partial \theta^{\mathrm{T}}} \, du,$$

$$u = \left(u_1, \cdots, u_d \right) = \left(G_1(y_1), \cdots, G_d(y_d) \right).$$

Applying Lemma 2 from Chen and Fan (2005), which becomes the central limit theorem when one holds the empirical process in the marginal parts constant, yields the joint distribution

$$\sqrt{n} \begin{pmatrix} \widehat{T}_{\mathrm{np}} - T_{\mathrm{true}} \\ \widehat{\eta}^{\mathrm{PMLpm}} - \eta_0^{\mathrm{PMLpm}} \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \Lambda_{\mathrm{np}} \\ (\mathcal{I}_\eta^{\mathrm{PMLpm}})^{-1} \Lambda_{\eta^{\mathrm{PMLpm}}} \end{pmatrix} \sim N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} V_{\mathrm{np}} & V_{\mathrm{np},\eta^{\mathrm{PMLpm}}} \\ \left(V_{\mathrm{np},\eta^{\mathrm{PMLpm}}} \right)^{\mathrm{T}} & V_{\eta^{\mathrm{PMLpm}}} \end{pmatrix} \right)$$

where

$$\widehat{\eta}^{\text{PMLpm}} = \left((\widehat{\alpha}^{\text{2ML}})^{\text{T}}, (\widehat{\theta}^{\text{PML}})^{\text{T}} \right)^{\text{T}},$$

$$\eta_0^{\text{PMLpm}} = \left((\alpha_0^{\text{2ML}})^{\text{T}}, (\theta_0^{\text{PML}})^{\text{T}} \right)^{\text{T}},$$

$$\mathcal{I}_{\eta}^{\text{PMLpm}} = \begin{pmatrix} \mathcal{I}_{\alpha}^{\text{2ML}} & 0 \\ 0 & \mathcal{I}_{\theta}^{\text{PML}} \end{pmatrix},$$

$$V_{\text{np},\eta^{\text{PMLpm}}} = Q_{\eta}^{\text{PMLpm}} \left(\mathcal{I}_{\eta}^{\text{PMLpm}} \right)^{-1},$$

$$V_{\eta^{\text{PMLpm}}} = \left(\mathcal{I}_{\eta}^{\text{PMLpm}} \right)^{-1} K_{\eta}^{\text{PMLpm}} \left(\mathcal{I}_{\eta}^{\text{PMLpm}} \right)^{-1},$$

$$Q_{\eta}^{\text{PMLpm}} = \text{Cov}_G \left(\text{IF}_T(y, G), \phi_{\eta}^{\text{PMLpm}}(y, \eta_0^{\text{PMLpm}}) + \mathcal{W}^*(y, \theta_0^{\text{PML}}) \right) = \text{E}_G \left[\text{IF}_T(y, G) \cdot \left(\phi_{\eta}^{\text{PMLpm}}(y, \eta_0^{\text{PMLpm}}) + \mathcal{W}^*(y, \theta_0^{\text{PML}}) \right)^{\text{T}} \right],$$

$$K_{\eta}^{\text{PMLpm}} = \text{Var}_G \left(\phi_{\eta}^{\text{PMLpm}}(y, \eta_0^{\text{PMLpm}} + \mathcal{W}^*(y, \theta_0^{\text{PML}})) \right)$$

$$= \text{E}_G \left[\left(\phi_{\eta}^{\text{PMLpm}}(y, \eta_0^{\text{PMLpm}}) + \mathcal{W}^*(y, \theta_0^{\text{PML}}) \right) \cdot \left(\phi_{\eta}^{\text{PMLpm}}(y, \eta_0^{\text{PMLpm}}) + \mathcal{W}^*(y, \theta_0^{\text{PML}}) \right)^{\text{T}} \right],$$

$$\phi_{\eta}^{\text{PMLpm}}(y, \eta) = \begin{pmatrix} \phi_{\alpha}^{\text{2ML}}(y, \alpha) \\ \phi_{\theta}^{\text{PML}}(\mathcal{W}^*((G_1(y_1), \cdots, G_d(y_d)), \theta)) \end{pmatrix},$$

$$\mathcal{W}^*(y, \theta) = \begin{pmatrix} 0_{\dim(\alpha)} \\ \mathcal{W}\left((G_1(y_1), \cdots, G_d(y_d)), \theta \right) \end{pmatrix},$$

$$\mathcal{W}(u, \theta) = \sum_{j=1}^d \mathcal{W}_j(u_j, \theta) = \sum_{j=1}^d \int_{[0,1]^d} c_0(v) \frac{\partial^2 \log c(v, \theta)}{\partial \theta \, \partial v_j^{\text{T}}} \left(I(u_j \le v_j) - v_j \right) \, dv.$$

Now by applying the delta method with transformation function $s_{x,\eta}(x, \eta)$, we obtain

$$\sqrt{n} \begin{pmatrix} \widehat{T}_{\text{np}} - T_{\text{true}} \\ \widehat{T}_{\text{PMLpm}} - T_{0,\text{PMLpm}} \end{pmatrix} \overset{d}{\to} \begin{pmatrix} \Lambda_{\text{np}} \\ \dot{s}_{\eta}^{\text{T}} \left(\mathcal{I}_{\eta}^{\text{PMLpm}} \right)^{-1} \Lambda_{\eta^{\text{PMLpm}}} \end{pmatrix} \sim N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} V_{\text{np}} & V_{\text{np,PMLpm}} \\ V_{\text{np,PMLpm}}^{\text{T}} & V_{\text{PMLpm}} \end{pmatrix} \right).$$

| Proof of Lemma 3

Proof When the copula model is estimated with the pseudo maximum likelihood with parametric margins, the non-constant part of the KL divergence is

$$Q(\widehat{\eta}^{\text{PMLpm}}) = \int g(y) \left\{ \log f_1(y_1, \widehat{\alpha}_1^{\text{2ML}}) + \cdots + \log f_d(y_d, \widehat{\alpha}_d^{\text{2ML}}) + \log c \left(F_1(y_1, \widehat{\alpha}_1^{\text{2ML}}), \cdots, F_d(y_d, \widehat{\alpha}_d^{\text{2ML}}), \widehat{\theta}^{\text{PML}} \right) \right\} \, dy.$$

The empirical equivalent is

$$\widehat{Q}(\widehat{\eta}^{\text{PMLpm}}) = \frac{1}{n} \sum_{i=1}^n \left[\log f_1(y_{i,1}, \widehat{\alpha}_1^{\text{2ML}}) + \cdots + \log f_d(y_{i,d}, \widehat{\alpha}_d^{\text{2ML}}) + \log c \left(F_1(y_{i,1}, \widehat{\alpha}_1^{\text{2ML}}), \cdots, F_d(y_{i,d}, \widehat{\alpha}_d^{\text{2ML}}), \widehat{\theta}^{\text{PML}} \right) \right].$$

Now we check the bias of $\hat{Q}(\hat{\eta}^{\mathrm{PMLpm}})$:

$$
\begin{aligned}
\mathrm{E}_G\left[\hat{Q}(\hat{\eta}^{\mathrm{PMLpm}})\right] - Q(\hat{\eta}^{\mathrm{PMLpm}}) = {} & \mathrm{E}_{G_1}\left[\frac{1}{n}\sum_{i=1}^n \log f_1(y_{i,1},\hat{\alpha}_1^{2\mathrm{ML}})\right] - \int g(y_1)\log f_1(y_1,\hat{\alpha}_1^{2\mathrm{ML}})\mathrm{d}y_1 \\
& + \cdots \\
& + \mathrm{E}_{G_d}\left[\frac{1}{n}\sum_{i=1}^n \log f_d(y_{i,d},\hat{\alpha}_d^{2\mathrm{ML}})\right] - \int g(y_d)\log f_d(y_d,\hat{\alpha}_d^{2\mathrm{ML}})\mathrm{d}y_d \\
& + \mathrm{E}_G\left[\frac{1}{n}\sum_{i=1}^n \log c\left(F_1(y_{i,1},\hat{\alpha}_1^{2\mathrm{ML}}),\cdots,F_d(y_{i,d},\hat{\alpha}_d^{2\mathrm{ML}}),\hat{\theta}^{\mathrm{PML}}\right)\right] \\
& - \int g(y)\log c\left(F_1(y_1,\hat{\alpha}_1^{2\mathrm{ML}}),\cdots,F_d(y_d,\hat{\alpha}_d^{2\mathrm{ML}}),\hat{\theta}^{\mathrm{PML}}\right)\mathrm{d}y.
\end{aligned}
$$

Since the parameter estimates $\hat{\alpha}_j^{2\mathrm{ML}}$ for the margins are obtained with ML estimation, we can use the results from the derivation of the TIC (Claeskens and Hjort, 2008) directly and obtain

$$
\mathrm{E}_{G_j}\left[\frac{1}{n}\sum_{i=1}^n \log f_j(y_{i,j},\hat{\alpha}_j^{2\mathrm{ML}})\right] - \int g(y_j)\log f(y_j,\hat{\alpha}_j^{2\mathrm{ML}})\mathrm{d}y_j = \frac{1}{n}\mathrm{tr}\left(\mathcal{I}_{\alpha_j}^{-1}K_{\alpha_j}\right) + o_p(n^{-1}).
$$

In a nutshell, $\mathcal{I}_{\alpha_j}$ is the Fisher information of the j-th margin and K_{α_j} is the covariance matrix of the score vector that belongs to the j-th margin. If one assumes that the model is correctly specified (i.e. $f = g$), we have $\mathcal{I}_{\alpha_j} = K_{\alpha_j}$ (White, 1982; Le Cam, 1990; Hardin, 2003). So, $\mathrm{tr}\left(\mathcal{I}_{\alpha_j}^{-1}K_{\alpha_j}\right) = \dim(\alpha_j)$.

Further, let

$$
Q_c(\hat{\eta}^{\mathrm{PMLpm}}) = \int g(y)\log c\left(F_1(y_1,\hat{\alpha}_1^{2\mathrm{ML}}),\cdots,F_d(y_d,\hat{\alpha}_d^{2\mathrm{ML}}),\hat{\theta}^{\mathrm{PML}}\right)\mathrm{d}y
$$

and

$$
\hat{Q}_c(\hat{\eta}^{\mathrm{PMLpm}}) = \mathrm{E}\left[\frac{1}{n}\sum_{i=1}^n \log c\left(F_1(y_{i,1},\hat{\alpha}_1^{2\mathrm{ML}}),\cdots,F_d(y_{i,d},\hat{\alpha}_d^{2\mathrm{ML}}),\hat{\theta}^{\mathrm{PML}}\right)\right].
$$

So, we can write

$$
\mathrm{E}_G\left[\hat{Q}(\hat{\eta}^{\mathrm{PMLpm}})\right] - Q(\hat{\eta}^{\mathrm{PMLpm}}) = \frac{1}{n}\sum_{j=1}^d \mathrm{tr}\left(\mathcal{I}_{\alpha_j}^{-1}K_{\alpha_j}\right) + \mathrm{E}_G\left[\hat{Q}_c(\hat{\eta}^{\mathrm{PMLpm}})\right] - Q_c(\hat{\eta}^{\mathrm{PMLpm}}) + o(n^{-1}).
$$

Now, $\mathrm{E}_G\left[\hat{Q}_c(\hat{\eta}^{\mathrm{PMLpm}})\right] - Q_c(\hat{\eta}^{\mathrm{PMLpm}})$ is the only element that has to be evaluated. Let

$$
\begin{aligned}
Q_c(\eta_0^{\mathrm{PMLpm}}) &= \int g(y)\log c\left(F_1(y_1,\alpha_{0,1}^{2\mathrm{ML}}),\cdots,F_d(y_d,\alpha_{0,d}^{2\mathrm{ML}}),\theta_0^{\mathrm{PML}}\right)\mathrm{d}y, \\
Z_i &= \log c\left(F_1(y_{i,1},\alpha_{0,1}^{2\mathrm{ML}}),\cdots,F_d(y_{i,d},\alpha_{0,d}^{2\mathrm{ML}}),\theta_0^{\mathrm{PML}}\right) - Q_c(\eta_0^{\mathrm{PMLpm}}), \\
A_\eta &= \sqrt{n}\left(\hat{\eta}^{\mathrm{PMLpm}} - \eta_0^{\mathrm{PMLpm}}\right) = \sqrt{n}\left(\mathcal{I}_\eta^{\mathrm{PMLpm}}\right)^{-1}\phi_\eta^{\mathrm{PMLpm}}(y,\eta_0^{\mathrm{PMLpm}}) + \begin{pmatrix} o_p(1) \\ o_p(1) \end{pmatrix},
\end{aligned}
$$

which stems from the proof of Proposition 2, and furthermore

$$\phi_{n,\eta}^*(\eta) = \frac{1}{n}\sum_{i=1}^n \phi_\eta^*(y_i,\eta) = \frac{1}{n}\sum_{i=1}^n \frac{\partial \log c\left(F_1(y_{i,1},\alpha_1),\cdots,F_d(y_{i,d},\alpha_d),\theta\right)}{\partial \eta},$$

$$H_{n,\eta}^*(\eta) = \frac{1}{n}\sum_{i=1}^n H_\eta^*(y_i,\eta) = \frac{1}{n}\sum_{i=1}^n \frac{\partial^2 \log c\left(F_1(y_{i,1},\alpha_1),\cdots,F_d(y_{i,d},\alpha_d),\theta\right)}{\partial \eta \partial \eta^{\mathrm{T}}},$$

$$\mathcal{I}_\eta^* = -\mathrm{E}_G\left[H_\eta^*(y,\eta_0^{\mathrm{PMLpm}})\right] = -\int g(y) H_\eta^*(y,\eta_0^{\mathrm{PMLpm}})\,\mathrm{d}y.$$

Then we have $\mathrm{E}[\overline{Z}_n] = \frac{1}{n}\sum_{i=1}^n \mathrm{E}[Z_i] = 0$, along with

$$\begin{aligned}
\widehat{Q}_c(\widehat{\eta}^{\mathrm{PMLpm}}) &= \frac{1}{n}\sum_{i=1}^n \log c(y_i,\widehat{\eta}^{\mathrm{PMLpm}}) \\
&= \frac{1}{n}\sum_{i=1}^n \Bigg[\log c(y_i,\eta_0^{\mathrm{PMLpm}}) - Q_c(\eta_0^{\mathrm{PMLpm}}) + Q_c(\eta_0^{\mathrm{PMLpm}}) + (\widehat{\eta}^{\mathrm{PMLpm}} - \eta_0^{\mathrm{PMLpm}})^{\mathrm{T}}\phi_\eta^*(y_i,\eta_0^{\mathrm{PMLpm}}) \\
&\quad + \frac{1}{2}(\widehat{\eta}^{\mathrm{PMLpm}} - \eta_0^{\mathrm{PMLpm}})^{\mathrm{T}} H_\eta^*(y_i,\eta_0^{\mathrm{PMLpm}})(\widehat{\eta}^{\mathrm{PMLpm}} - \eta_0^{\mathrm{PMLpm}})\Bigg] + o_p(n^{-1}) \\
&= \frac{1}{n}\sum_{i=1}^n \left[Z_i\right] + Q_c(\eta_0^{\mathrm{PMLpm}}) + (\widehat{\eta}^{\mathrm{PMLpm}} - \eta_0^{\mathrm{PMLpm}})^{\mathrm{T}}\phi_{n,\eta}^*(\eta_0^{\mathrm{PMLpm}}) \\
&\quad + \frac{1}{2}(\widehat{\eta}^{\mathrm{PMLpm}} - \eta_0^{\mathrm{PMLpm}})^{\mathrm{T}} H_{n,\eta}^*(\eta_0^{\mathrm{PMLpm}})(\widehat{\eta}^{\mathrm{PMLpm}} - \eta_0^{\mathrm{PMLpm}}) + o_p(n^{-1}) \\
&= Q_c(\eta_0^{\mathrm{PMLpm}}) + \overline{Z}_n + \frac{1}{\sqrt{n}} A_\eta^{\mathrm{T}}\phi_{n,\eta}^*(\eta_0^{\mathrm{PMLpm}}) + \frac{1}{2n} A_\eta^{\mathrm{T}} H_{n,\eta}^*(\eta_0^{\mathrm{PMLpm}}) A_\eta + o_p(n^{-1}),
\end{aligned}$$

$$\begin{aligned}
Q_c(\widehat{\eta}^{\mathrm{PMLpm}}) &= \int g(y)\log c\left(F_1(y_1,\widehat{\alpha}_1^{\mathrm{2ML}}),\cdots,F_d(y_d,\widehat{\alpha}_d^{\mathrm{2ML}}),\widehat{\theta}^{\mathrm{PML}}\right)\mathrm{d}y \\
&= \int g(y)\Bigg[\log c(y,\eta_0^{\mathrm{PMLpm}}) + (\widehat{\eta}^{\mathrm{PMLpm}} - \eta_0^{\mathrm{PMLpm}})^{\mathrm{T}}\phi_\eta^*(y,\eta_0^{\mathrm{PMLpm}}) \\
&\quad + \frac{1}{2}(\widehat{\eta}^{\mathrm{PMLpm}} - \eta_0^{\mathrm{PMLpm}})^{\mathrm{T}} H_\eta^*(y,\eta_0^{\mathrm{PMLpm}})(\widehat{\eta}^{\mathrm{PMLpm}} - \eta_0^{\mathrm{PMLpm}})\Bigg]\mathrm{d}y + o_p(n^{-1}) \\
&= \int g(y)\log c(y,\eta_0^{\mathrm{PMLpm}})\,\mathrm{d}y + (\widehat{\eta}^{\mathrm{PMLpm}} - \eta_0^{\mathrm{PMLpm}})^{\mathrm{T}}\int g(y)\phi_\eta^*(y,\eta_0^{\mathrm{PMLpm}})\,\mathrm{d}y \\
&\quad + \frac{1}{2}(\widehat{\eta}^{\mathrm{PMLpm}} - \eta_0^{\mathrm{PMLpm}})^{\mathrm{T}}\int g(y) H_\eta^*(y,\eta_0^{\mathrm{PMLpm}})\,\mathrm{d}y \cdot (\widehat{\eta}^{\mathrm{PMLpm}} - \eta_0^{\mathrm{PMLpm}}) + o_p(n^{-1}) \\
&= Q_c(\eta_0^{\mathrm{PMLpm}}) + (\widehat{\eta}^{\mathrm{PMLpm}} - \eta_0^{\mathrm{PMLpm}})^{\mathrm{T}}\cdot 0 \\
&\quad + \frac{1}{2n}\sqrt{n}(\widehat{\eta}^{\mathrm{PMLpm}} - \eta_0^{\mathrm{PMLpm}})^{\mathrm{T}}\int g(y) H_\eta^*(y,\eta_0^{\mathrm{PMLpm}})\,\mathrm{d}y \cdot \sqrt{n}(\widehat{\eta}^{\mathrm{PMLpm}} - \eta_0^{\mathrm{PMLpm}}) + o_p(n^{-1}) \\
&= Q_c(\eta_0^{\mathrm{PMLpm}}) - \frac{1}{2n} A_\eta^{\mathrm{T}}\mathcal{I}_\eta^* A_\eta + o_p(n^{-1}),
\end{aligned}$$

and

$$n\{\widehat{Q}_c(\widehat{\eta}^{\mathrm{PMLpm}}) - Q_c(\widehat{\eta}^{\mathrm{PMLpm}})\} = n\overline{Z}_n + \sqrt{n} A_\eta^{\mathrm{T}}\phi_{n,\eta}^*(\eta_0^{\mathrm{PMLpm}}) + \frac{1}{2} A_\eta^{\mathrm{T}} H_{n,\eta}^*(\eta_0^{\mathrm{PMLpm}}) A_\eta + \frac{1}{2} A_\eta^{\mathrm{T}}\mathcal{I}_\eta^* A_\eta + o_p(1).$$

We now assume that the margins are correctly specified, i.e. $F_j(y_j,\alpha_{0,j}^{\mathrm{2ML}}) = G_j(y_j)$. This implies that $\theta_0^{\mathrm{PML}} = \theta_0^{\mathrm{2ML}}$ and

consequently $\eta_0^{\text{PMLpm}} = \eta_0^{\text{2ML}}$. Then, the central limit theorem gives

$$\sqrt{n}\phi_{n,\eta}^*(\eta_0^{\text{PMLpm}}) = \sqrt{n}\phi_{n,\eta}^*(\eta_0^{\text{2ML}}) = \sqrt{n}\left\{\phi_{n,\eta}^*(\eta_0^{\text{2ML}}) - \mathrm{E}\left[\phi_\eta^*(Y, \eta_0^{\text{2ML}})\right]\right\} \xrightarrow{d} \Lambda_\eta^* \sim N\left(0, K_\eta^*\right),$$

where $K_\eta^* = \mathrm{Var}\left(\phi_\eta^*(y, \eta_0^{\text{2ML}})\right) = \mathrm{E}\left[\phi_\eta(y, \eta_0^{\text{2ML}})\phi_\eta(y, \eta_0^{\text{2ML}})^\mathrm{T}\right]$.

Now we evaluate $\mathrm{E}_G[n\{\hat{Q}_c(\hat{\eta}^{\text{PMLpm}}) - Q_c(\hat{\eta}^{\text{PMLpm}})\}]$:

$$
\begin{aligned}
\mathrm{E}_G\left[n\left(\hat{Q}_c(\hat{\eta}^{\text{PMLpm}}) - Q_c(\hat{\eta}^{\text{PMLpm}})\right)\right] &= \mathrm{E}_G\left[n\overline{Z}_n + \sqrt{n}A_\eta^\mathrm{T}\phi_{n,\eta}^*(\eta_0^{\text{PMLpm}}) + \frac{1}{2}A_\eta^\mathrm{T}H_{n,\eta}^*(\eta_0^{\text{PMLpm}})A_\eta + \frac{1}{2}A_\eta^\mathrm{T}\mathcal{I}_\eta^*A_\eta\right] + o_p(1) \\
&\xrightarrow{p} \mathrm{E}_G\left[\left(\left(\mathcal{I}_\eta^{\text{PMLpm}}\right)^{-1}\Lambda_{\eta^{\text{2ML}}}\right)^\mathrm{T}\Lambda_\eta^*\right] \\
&= \mathrm{E}_G\left[\mathrm{tr}\left(\left(\mathcal{I}_\eta^{\text{PMLpm}}\right)^{-1}\Lambda_{\eta^{\text{PMLpm}}}\left(\Lambda_\eta^*\right)^\mathrm{T}\right)\right] \\
&= \mathrm{tr}\left(\left(\mathcal{I}_\eta^{\text{PMLpm}}\right)^{-1}\mathrm{E}_G\left[\Lambda_{\eta^{\text{PMLpm}}}\left(\Lambda_\eta^*\right)^\mathrm{T}\right]\right) \\
&= \mathrm{tr}\left(\left(\mathcal{I}_\eta^{\text{PMLpm}}\right)^{-1}K_\eta^\circ\right),
\end{aligned}
$$

where

$$K_\eta^\circ = \mathrm{E}_G\left[\Lambda_{\eta^{\text{PMLpm}}}(\Lambda_\eta^*)^\mathrm{T}\right] = \mathrm{E}_G\left[\begin{pmatrix} \Lambda_{\alpha^{\text{2ML}}}\left(\Lambda_\alpha^*\right)^\mathrm{T} & \Lambda_{\alpha^{\text{2ML}}}\left(\Lambda_\theta^*\right)^\mathrm{T} \\ \Lambda_{\theta^{\text{PML}}}\left(\Lambda_\alpha^*\right)^\mathrm{T} & \Lambda_{\theta^{\text{PML}}}\left(\Lambda_\alpha^*\right)^\mathrm{T} \end{pmatrix}\right] = \begin{pmatrix} K_\alpha^\circ & K_{\alpha,\theta}^\circ \\ K_{\theta,\alpha}^\circ & K_\theta^\circ \end{pmatrix}.$$

Further, we assume that $c = c_0$ such that the whole model is correctly specified. According to Genest et al. (1995), the true model assumption yields $\mathrm{Cov}_{C_0}\left(\phi_\theta^{\text{PML}}(u, \theta_0^{\text{PML}}), \mathcal{W}(u, \theta_0^{\text{PML}}))\right) = 0$ and $\mathcal{I}_\theta^{\text{PML}} = \Sigma_\theta^{\text{PML}}$ where $\Sigma_\theta^{\text{PML}} = \mathrm{E}_{C_0}\left[\left(\phi_\eta^{\text{PMLpm}}(u, \eta_0^{\text{PMLpm}})\right) \cdot \left(\phi_\eta^{\text{PMLpm}}(u, \eta_0^{\text{PMLpm}})\right)^\mathrm{T}\right]$. Since we also assumed true margins, we have

$$\log c\left(F_1(y_1, \alpha_{0,1}^{\text{2ML}}), \cdots, F_d(y_d, \alpha_{0,d}^{\text{2ML}}), \theta\right) = \log c\left(G_1(y_1), \cdots, G_d(y_d), \theta\right).$$

This leads to $K_\theta^\circ = \mathcal{I}_\theta^{\text{PML}}$. In addition, Lemma 1 of Ko and Hjort (2019a) gives $K_\alpha^\circ = \mathcal{I}_\alpha^{\text{2ML}} - K_\alpha^{\text{2ML}}$.

By using those results, we have

$$
\begin{aligned}
\mathrm{tr}\left(\left(\mathcal{I}_\eta^{\text{PMLpm}}\right)^{-1}K_\eta^\circ\right) &= \mathrm{tr}\left(\begin{pmatrix} \mathcal{I}_\alpha^{\text{2ML}} & 0 \\ 0 & \mathcal{I}_\theta^{\text{PML}} \end{pmatrix}^{-1}\begin{pmatrix} K_\alpha^\circ & K_{\alpha,\theta}^\circ \\ K_{\theta,\alpha}^\circ & K_\theta^\circ \end{pmatrix}\right) \\
&= \mathrm{tr}\left(\begin{pmatrix} \left(\mathcal{I}_\alpha^{\text{2ML}}\right)^{-1}K_\alpha^\circ & \left(\mathcal{I}_\alpha^{\text{2ML}}\right)^{-1}K_{\alpha,\theta}^\circ \\ \left(\mathcal{I}_\theta^{\text{PML}}\right)^{-1}K_{\theta,\alpha}^\circ & \left(\mathcal{I}_\theta^{\text{PML}}\right)^{-1}K_\theta^\circ \end{pmatrix}\right) \\
&= \mathrm{tr}\left(\begin{pmatrix} \left(\mathcal{I}_\alpha^{\text{2ML}}\right)^{-1}\left(\mathcal{I}_\alpha^{\text{2ML}} - K_\alpha^{\text{2ML}}\right) & 0 \\ 0 & \left(\mathcal{I}_\theta^{\text{PML}}\right)^{-1}K_\theta^\circ \end{pmatrix}\right) \\
&= \mathrm{tr}\left(\begin{pmatrix} 0 & 0 \\ 0 & I_{\dim(\theta)} \end{pmatrix}\right) \\
&= \dim(\theta).
\end{aligned}
$$

Thus, the unbiased estimator of $Q(\widehat{\eta}^{\text{PMLpm}})$ is

$$\widehat{Q}(\widehat{\eta}^{\text{PMLpm}}) = \frac{1}{n}\ell_n(\widehat{\eta}^{\text{PMLpm}}) - \frac{1}{n}\dim(\eta)$$

By scaling this unbiased estimator with $2n$ such that it is one the same scale as the AIC for maximum likelihood (Akaike, 1974), we define the AIC for the copula model estimated with PML with parametric margins

$$\text{AIC} = 2\ell_n(\widehat{\eta}^{\text{PMLpm}}) - 2\dim(\eta).$$