Focused information criterion for copulas

Vinnie Ko*, Nils Lid Hjort, Ingrid Hobæk Haff
Department of Mathematics, University of Oslo
PB 1053, Blindern, NO-0316 Oslo, Norway

July 2018

Abstract

In this paper, we extend the focused information criterion (FIC) from Claeskens & Hjort (2008) and Jullum & 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: model selection, mean squared error, two-stage maximum likelihood, semi-parametric estimation

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, th AIC and TIC aim for the model that minimizes the Kullback–Leibler (KL) divergence from the real data-generating mechanism to the true 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.

*Corresponding author.
E-mail addresses: vinniebk@math.uio.no (V. Ko), nils@math.uio.no (N.L. Hjort), ingrihaf@math.uio.no (I. Hobæk Haff)

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 & 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 & Hjort, 2003) is that it assumes all the models to be parametric and nested. Jullum & Hjort (2017) extend the FIC to a more general setting by using a non-parametric model to estimate the bias part of the decomposed MSE. Their new FIC machinery is designed for maximum likelihood (ML) estimated parametric models and non-parametric 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, $C(\cdot, \theta)$ is unique, and can be differentiated,

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

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 non-parametric margins, Grønneberg & Hjort (2014) develop a copula information criterion (CIC) and Chen & Fan (2005) suggest a pseudo likelihood ratio test. Ko & Hjort (2018a) propose a different CIC, for the two-stage ML. The present paper involves extending the FIC theory to these different estimation procedures.

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 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 the abalone data set 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 & 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 & Hjort (2003) and later modified by Jullum & Hjort (2017), is that one wants to select a model that is good for a specific task. We quantify this specific task as that of estimating a focus parameter, say $T(M)$, which is a functional with an arbitrary distribution function M as its input. So, T can be anything as long as it can be written as a functional of a distribution function.

In our setting, the focus parameter from the data generating model is denoted as $T(G)$ or $T_{\rm true}$ and the focus parameter from our estimated candidate model is denoted as $T(\widehat{F})$ or $\widehat{T}_{\rm cnd}$. We are looking for the model $f(\cdot)$ such that $\widehat{T}_{\rm cnd}$ is as close as possible to $T_{\rm true}$. Jullum & Hjort (2017) use a quadratic loss function, which is the MSE by definition, as a measure of closeness. They define the focused information criterion (FIC) as the estimated MSE

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

We return to estimates for the variance of $\widehat{T}_{\rm cnd}$ below, with different recipes for different estimators. For estimating the bias, they use the fact that the non-parametrically estimated T, which we denote as $\widehat{T}_{\rm np}$, is asymptotically unbiased. This yields

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

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

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

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 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 & Hjort, 2017), the joint distribution of $\widehat{T}_{\text{ML}}$ and $\widehat{T}_{\text{np}}$, can in our copula setting be written as

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

where

$$\begin{aligned}
&V_{\text{np}} = \text{Var}_G\left(\text{IF}_T(y,G)\right) = \text{E}_G\left[\text{IF}(y,G)\,\text{IF}(y,G)^{\text{T}}\right], \\
&V_{\text{np,ML}} = Q_{\text{ML}}\left(\mathcal{I}_\eta^{\text{ML}}\right)^{-1}\dot{s}_\eta,\ Q_{\text{ML}} = \text{E}_G\left[\text{IF}_T(y,G)\cdot\phi_\eta^{\text{ML}}(y,\eta_0^{\text{ML}})^{\text{T}}\right],\ \dot{s}_\eta = \frac{\partial s_\eta}{\partial\eta} = \frac{\partial T(F(\eta))}{\partial\eta}, \\
&V_{\text{ML}} = \dot{s}_\eta^{\text{T}}\left(\mathcal{I}_\eta^{\text{ML}}\right)^{-1}K_\eta^{\text{ML}}\left(\mathcal{I}_\eta^{\text{ML}}\right)^{-1}\dot{s}_\eta,\ K_\eta^{\text{ML}} = \text{E}_G\left[\phi_\eta^{\text{ML}}(y,\eta_0^{\text{ML}})\,\phi_\eta^{\text{ML}}(y,\eta_0^{\text{ML}})^{\text{T}}\right], \\
&\phi_\eta^{\text{ML}}(y,\eta) = \frac{\partial\left(\sum_{j=1}^d f_j(y_j,\alpha_j) + \log c\left(F_1(y_1,\alpha_1),\cdots,F_d(y_d,\alpha_d),\theta\right)\right)}{\partial\eta}, \\
&\mathcal{I}_\eta^{\text{ML}} = -\text{E}_G\left[H_\eta(y,\eta_0^{\text{ML}})\right] = -\int g\,\frac{\partial^2\left(\sum_{j=1}^d f_j(y_j,\alpha_{0,j}^{\text{ML}}) + \log c\left(F_1(y_1,\alpha_{0,1}^{\text{ML}}),\cdots,F_d(y_d,\alpha_{0,d}^{\text{ML}}),\theta_0^{\text{ML}}\right)\right)}{\partial\eta\,\partial\eta^{\text{T}}}\,\text{d}y
\end{aligned}$$

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

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

where

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

Here $\widehat{V}_{\text{np}}$, $\widehat{Q}_{\text{ML}}$, $\widehat{\mathcal{I}}_\eta^{\text{ML}}$, $\widehat{K}_\eta^{\text{ML}}$ indicate empirical analogues of V_{np}, Q_{ML}, $\mathcal{I}_\eta^{\text{ML}}$, K_η^{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 & 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 $\widehat{\alpha}^{\text{2ML}}$ into the log-likelihood and obtains $\widehat{\theta}^{\text{2ML}}$ by maximizing the log-likelihood with respect to θ. The asymptotic properties of the two-stage ML estimator are given in Ko & Hjort (2018b) 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 & Hjort (2017) and $(A1)-(A5)$ from Ko & Hjort (2018b),*

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

where

$$\begin{aligned} V_{\text{np,2ML}} &= Q_\eta^{\text{2ML}} \left(\left(\mathcal{I}_\eta^{\text{2ML}}\right)^{-1}\right)^{\text{T}} \dot{s}_\eta, \\ V_{\text{2ML}} &= \dot{s}_\eta^{\text{T}} V_{\eta^{\text{2ML}}} \dot{s}_\eta, \\ V_{\eta^{\text{2ML}}} &= \left(\mathcal{I}_\eta^{\text{2ML}}\right)^{-1} K_\eta^{\text{2ML}} \left(\left(\mathcal{I}_\eta^{\text{2ML}}\right)^{-1}\right)^{\text{T}} \end{aligned}$$

and other quantities are defined in the proof.

The proof is given in Appendix A.

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

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

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 $\widehat{\theta}^{\text{ML}}$ and $\widehat{\theta}^{\text{2ML}}$. 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 non-parametric probability integral transforms of the data. Subsequently, one plugs these into the copula log-likelihood and maximizes it with respect to θ to obtain $\widehat{\theta}^{\text{PML}}$. For details concerning the PML estimator, see Tsukahara (2005), Chen & Fan (2005) and Kim *et al.* (2007).

Once θ is estimated, one can keep the non-parametric 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 & Fan (2005) and* $(C1)-(C4)$ *from Jullum & Hjort (2017),*

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

where

$$\begin{aligned}V_{\text{np,PMLpm}}&=Q_\eta^{\text{PMLpm}}\left(\mathcal{I}_\eta^{\text{PMLpm}}\right)^{-1}\dot{s}_\eta,\\ V_{\text{PMLpm}}&=\dot{s}_\eta^{\text{T}}V_{\eta^{\text{PMLpm}}}\dot{s}_\eta,\\ V_{\eta^{\text{PMLpm}}}&=\left(\mathcal{I}_\eta^{\text{PMLpm}}\right)^{-1}K_\eta^{\text{PMLpm}}\left(\mathcal{I}_\eta^{\text{PMLpm}}\right)^{-1}\end{aligned}$$

and other quantities are defined in the proof.

The proof is given in Appendix A.

In practice, V_{PMLpm} and $V_{\text{np,PMLpm}}$ 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 & Fan (2005). The resulting estimated variance and covariance are written as $\widehat{V}_{\text{PMLpm}}$ and $\widehat{V}_{\text{np,PMLpm}}$. Plugging these into (1) and (2) gives the FIC for PMLpm

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

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

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

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 1. *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 $2n \cdot Q(\widehat{\eta}^{\mathrm{PMLpm}})$, widely known as Akaike information criterion (AIC), is*

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

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 AFIC

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 & Hjort (2017) developed an averaged weighted FIC (AFIC) which aims for the model that obtains lowest risk for a set of focus parameters $T_{\mathrm{true}}(t)$, 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

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

which results in the following risk:

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

Then, they define the AFIC as

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

where the bias term is defined as

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

Note that when the set of estimands $T_{\mathrm{true}}(t)$ 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 & 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 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 non-parametrically, 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 & 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 & Fan (2005) based estimate for small sample sizes, and they produced comparable results.

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

The results from the simulations can be found in Tables 4 and 5 in Appendix B. The result is also visualized in Figure 1 and 3 for $n = 100$ and in Figure 2 and 4 for $n = 1000$.

The 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 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 4), the correspondence is even better.

Figure 3 shows 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 4 shows that the same holds when the sample size increases.

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

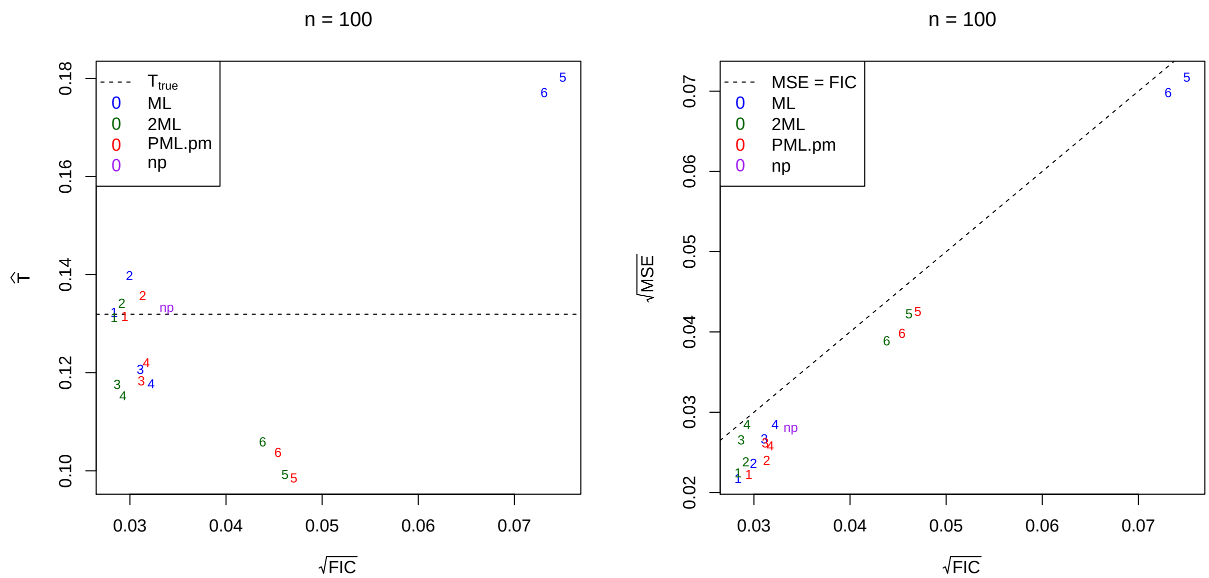

Figure 1: Results from Part 1 of the simulations for $n = 100$. The numbers refer to the model numbers in Table 1. The non-parametric model is denoted 'np'.

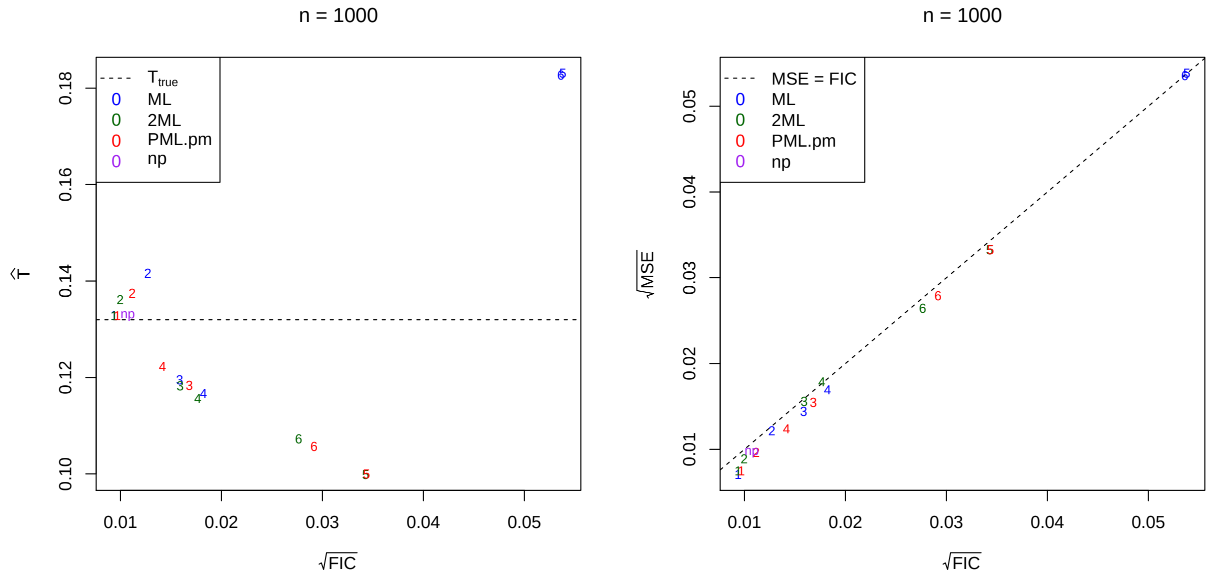

Figure 2: Results from Part 1 of the simulations for $n = 1000$. See Figure 1 for an explanation of symbols.

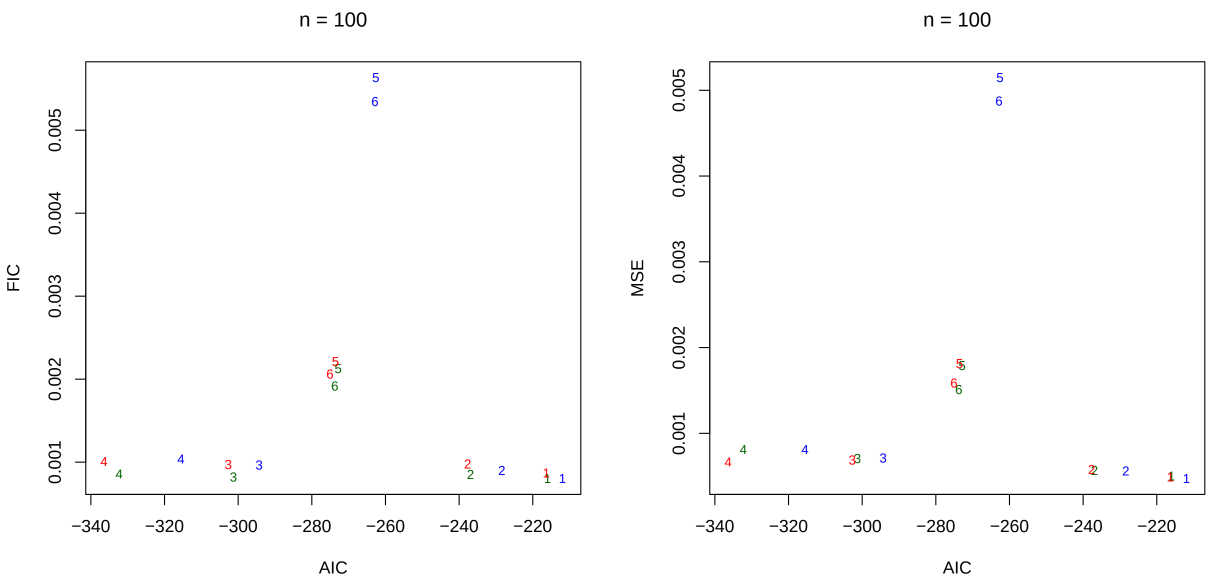

Figure 3: Results from Part 1 of the simulations for $n = 100$. See Figure 1 for an explanation of symbols.

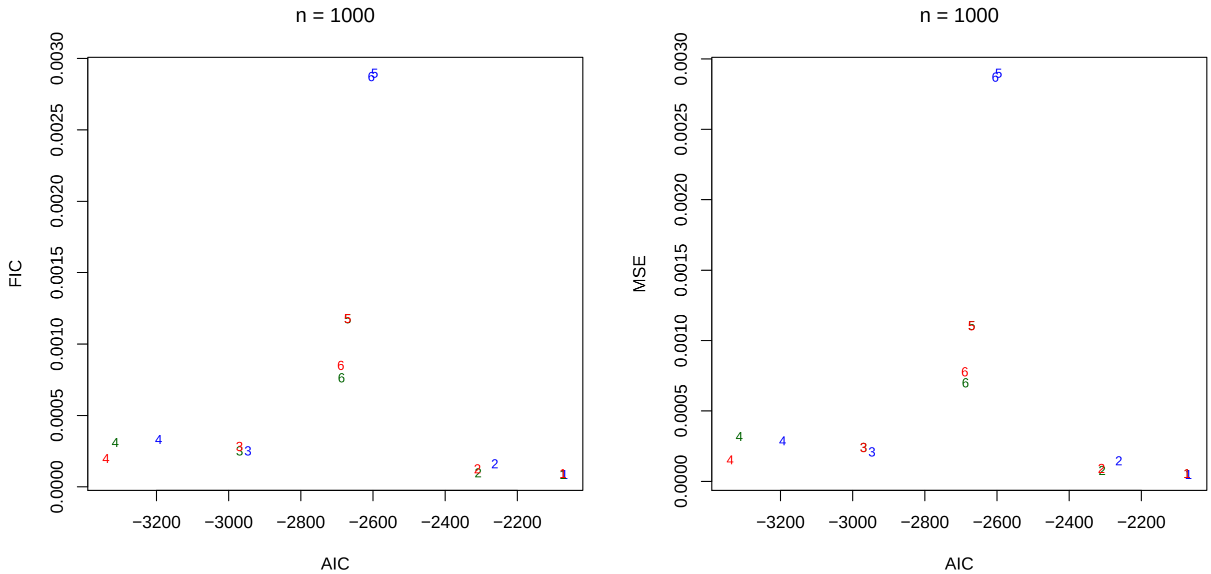

Figure 4: Results from Part 1 of the simulations for $n = 1000$. See Figure 1 for an explanation of symbols.

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 $T(F) = P(G_1^{-1}(p) < y_1 \,|\, G_2^{-1}(p) < y_2)$ for $p = 0.90, 0.91, \cdots, 0.95$. When p is close to 1, a reasonable precision of the non-parametric 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.

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

The results from the simulations with the AFIC, as well as with the FIC with single focus parameter for $p = 0.90$ and $p = 0.95$ can be found in Tables 6, 7 and 8 in Appendix B. The results are also visualized in Figures 5 to 10. The left panel of Figure 5 is a plot of $\sqrt{\mathrm{FIC}}$ against the focus parameter estimates $\widehat{T}$ for the FIC with $p = 0.9$, and the right panel shows $\sqrt{\mathrm{FIC}}$ against $\sqrt{\mathrm{MSE}}$. Figure 6 displays AIC against FIC and AIC against MSE to the left and right, respectively. Figures 7 and 8 show corresponding results for the FIC with $p = 0.95$, whereas Figures 9 and 10 are for the AFIC.

When we consider a single focus parameter with $p = 0.90$, the models with smaller $\sqrt{\mathrm{FIC}}$ 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.

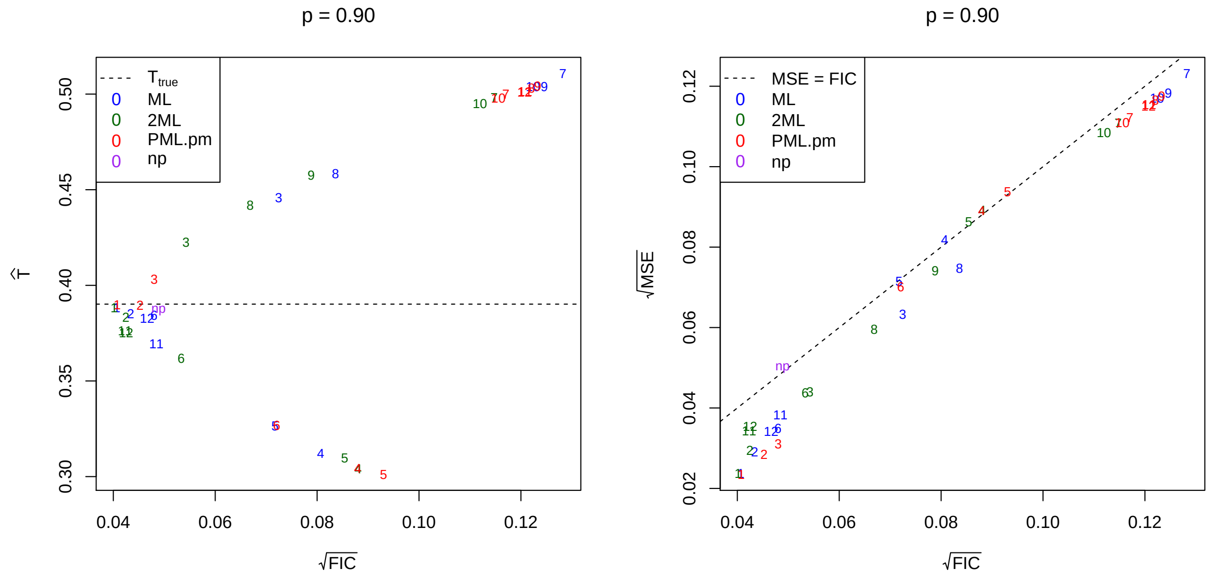

Figure 5: Results from Part 2 of the simulations for the FIC with $p = 0.90$. The numbers refer to the model numbers in Table 2.

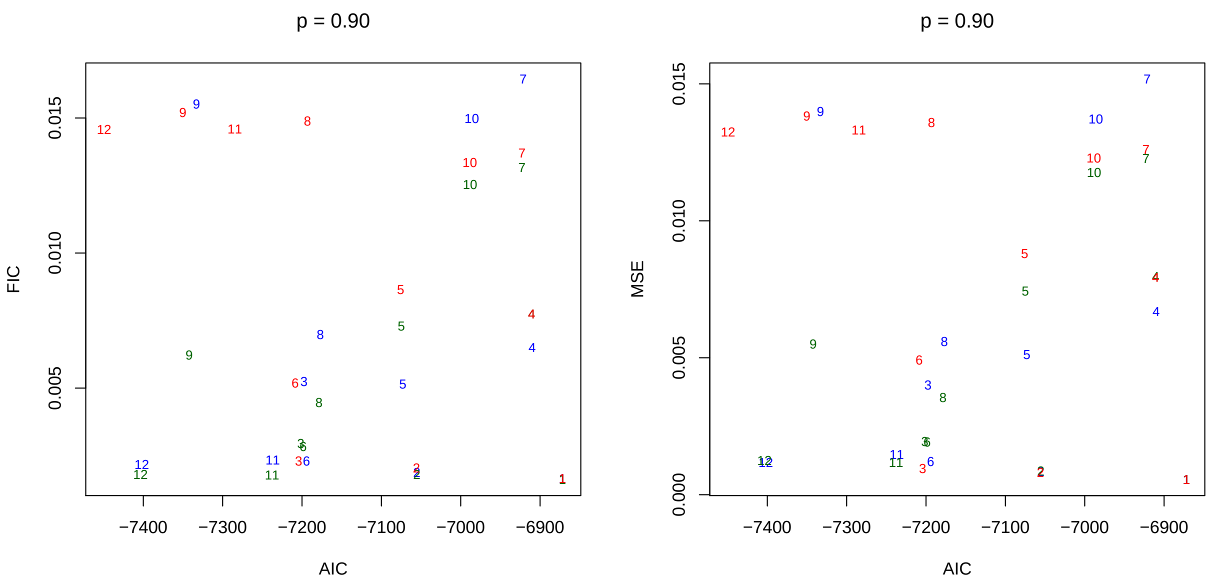

Figure 6: Results from Part 2 of the simulations for the FIC with $p = 0.90$. The numbers refer to the model numbers in Table 2.

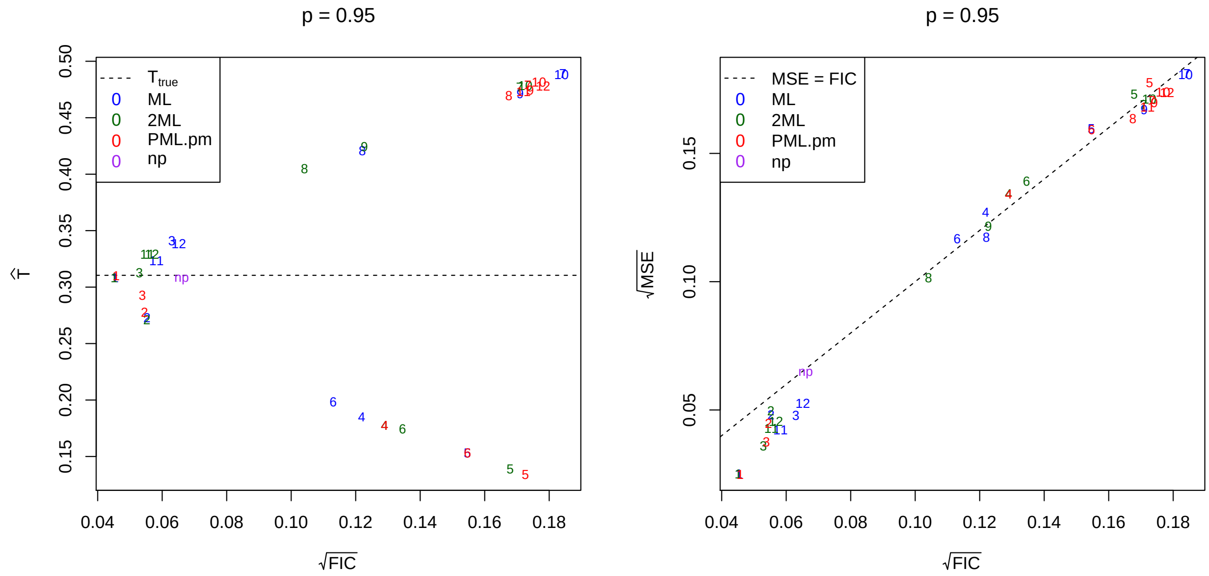

Figure 7: Results from Part 2 of the simulations for the FIC with $p = 0.95$. The numbers refer to the model numbers in Table 2.

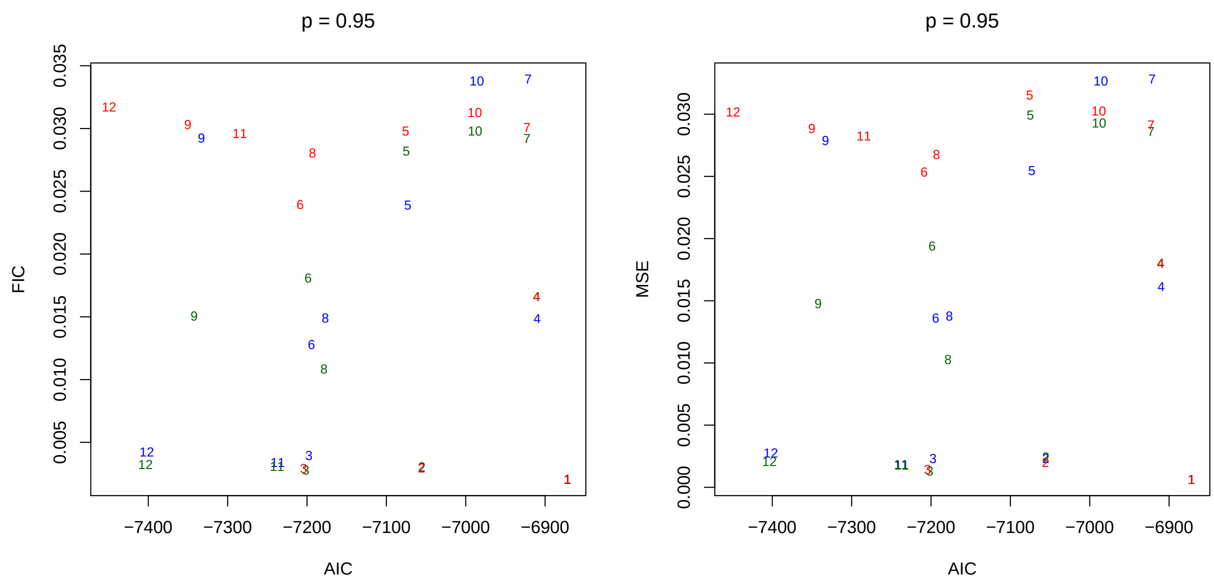

Figure 8: Results from Part 2 of the simulations for the FIC with $p = 0.95$. The numbers refer to the model numbers in Table 2.

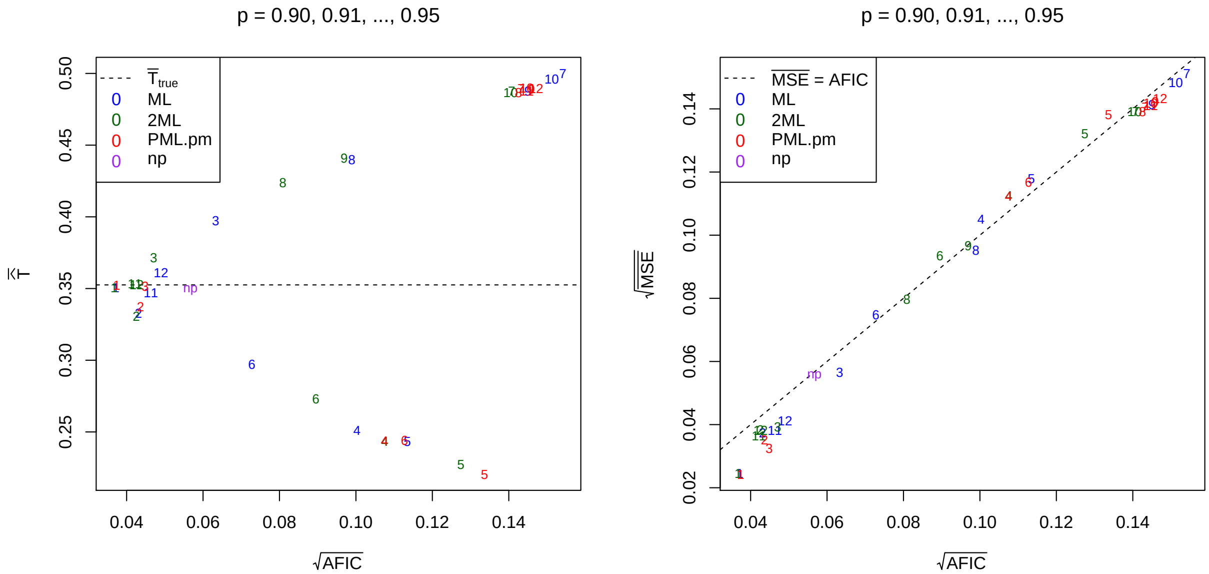

Figure 9: Results from Part 2 of the simulations for the AFIC. The numbers refer to the model numbers in Table 2. $\overline{\widehat{T}}$ indicates the average of $\widehat{T}$ across focus parameters in the set. Similarly, $\overline{\text{MSE}}$ indicates the average MSE.

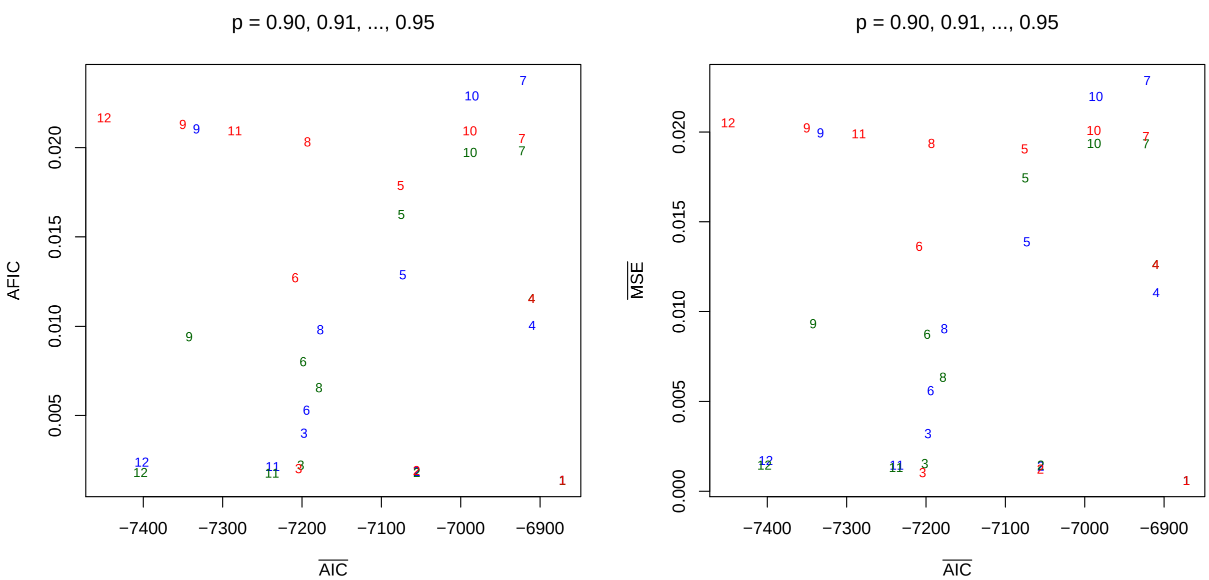

Figure 10: Results from Part 2 of the simulations for the AFIC. The numbers refer to the model numbers in Table 2. $\overline{\text{AIC}}$ indicates the average of AIC across focus parameters in the set. Similarly, $\overline{\text{MSE}}$ indicates the average MSE.

6 Example: The abalone data

As a real-life data example, we applied our FIC model selection method to the abalone data set (Asuncion & Newman, 2011), which has previously been used to illustrate different dependence modeling methods e.g. Ma *et al.* (2012) and Hobæk Haff *et al.* (2016). An 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 $N(0, 0.01^2)$.

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

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 specifically, we used the focus parameter

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

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 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 9 in Appendix B and also visualized in Figure 11. We see a similar pattern to the ones from the simulation study. One of the notable differences is that the non-parametric model is now the winning model. This seems reasonable considering that the non-parametric 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 non-parametric 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 non-parametric 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.

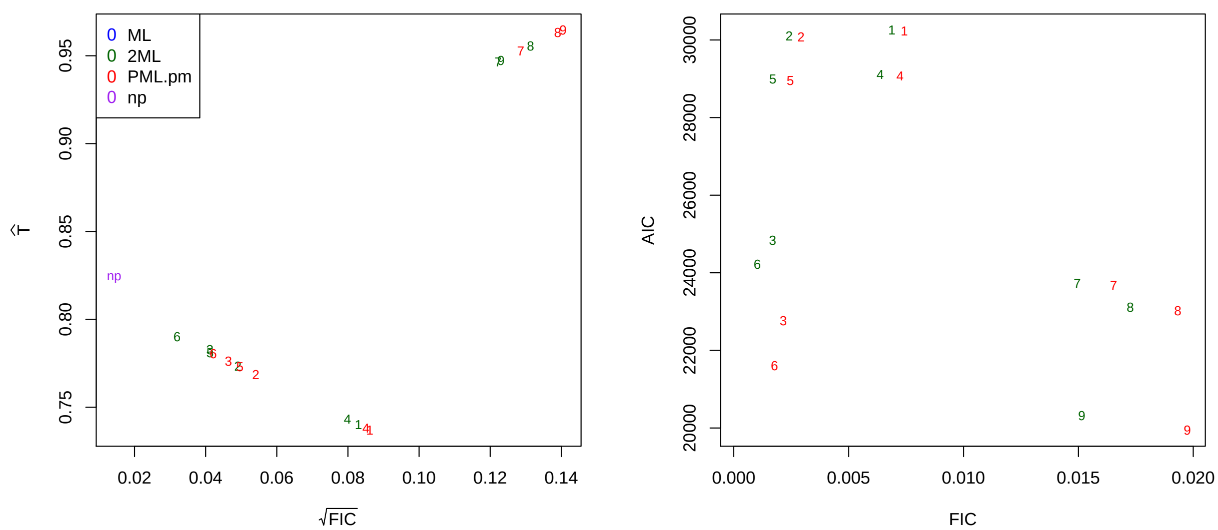

Figure 11: Results from the models fitted to the abalone data. The numbers refer to the model numbers in Table 3.

7 Conclusions and further research

In this paper, we have developed the FIC for copula models partly based on the general methods of Jullum & 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 non-parametric one. It can also easily be extended to other estimation schemes as long as the joint asymptotic distribution with the non-parametric 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 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; cf. Jullum & Hjort (2017, Section 7).

As we see from the FIC plots (Figures 1, 2, 5, 7 and 9), 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 & Hjort (2008, Chapter 7) propose and develop machinery for such model average estimators, and suggest using weights proportional to $\exp(-\lambda \cdot \mathrm{FIC})$, 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 non-parametric, 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

The authors would like to thank Steffen Grønneberg for his valuable comments and fruitful discussions. The authors also acknowledge partial funding from the Norwegian Research Council supported research group FocuStat: Focus Driven Statistical Inference With Complex Data, and from the Department of Mathematics at the University of Oslo.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE transactions on automatic control* **19**, 716–723.

Asuncion, A. & Newman, D. (2011). UCI machine learning repository, university of california, irvine, school of information and computer sciences.

Bevacqua, E., Maraun, D., Hobæk Haff, I., Widmann, M. & 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. & Fan, Y. (2005). Pseudo-likelihood ratio tests for semiparametric multivariate copula model selection. *Canadian Journal of Statistics* **33**, 389–414.

Claeskens, G. & Hjort, N. L. (2003). The focused information criterion. *Journal of the American Statistical Association* **98**, 900–916.

Claeskens, G. & 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. & Rivest, L.-P. (1995). A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika* **82**, 543–552.

Grønneberg, S. & 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*. Emerald Group Publishing Limited, pp. 45–73.

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

Hobæk Haff, I., Frigessi, A. & 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. & Hjort, N. L. (2017). Parametric or nonparametric: the fic approach. *Statistica Sinica* **27**, 951–981.

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

Ko, V. & Hjort, N. L. (2018a). Copula information criterion for model selection with two-stage maximum likelihood estimation. Submitted to a peer-reviews journal.

Ko, V. & Hjort, N. L. (2018b). Impact of model misspecification on two-stage maximum likelihood estimation for copulae. Submitted to a peer-reviews journal.

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

Ma, J., Sun, Z.-Q., Chen, S. & 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. & 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 & Hjort (2018b) that

$$\widehat{T}_{\rm np} - T_{\rm true} = \frac{1}{n}\sum_{i=1}^{n} {\rm IF}_T(y_i, G) + o_p(n^{-1/2}),$$
$$\widehat{\eta}^{\rm 2ML} - \eta_0^{\rm 2ML} = (\mathcal{I}_\eta^{\rm 2ML})^{-1}\phi_{n,\eta}^{\rm 2ML}(\eta_0^{\rm 2ML}) + o_p(n^{-1/2})$$

where

$$\phi_{n,\eta}^{\rm 2ML}(\eta) = \frac{1}{n}\sum_{i=1}^{n}\phi_\eta^{\rm 2ML}(y_i,\eta),$$
$$\phi_{n,\alpha}^{\rm 2ML}(\alpha) = \frac{1}{n}\sum_{i=1}^{n}\phi_\alpha^{\rm 2ML}(y_i,\alpha),$$
$$\phi_{n,\theta}^{\rm 2ML}(\theta) = \frac{1}{n}\sum_{i=1}^{n}\phi_\theta^{\rm 2ML}(y_i,\theta),$$
$$\phi_\eta^{\rm 2ML}(\eta) = \begin{pmatrix}\phi_\alpha^{\rm 2ML}(\alpha)\\ \phi_\theta^{\rm 2ML}(\eta)\end{pmatrix},$$
$$\phi_\alpha^{\rm 2ML}(y,\alpha) = \begin{pmatrix}\phi_{\alpha_1}^{\rm 2ML}(y_1,\alpha_1)\\ \vdots \\ \phi_{\alpha_d}^{\rm 2ML}(y_d,\alpha_d)\end{pmatrix} = \begin{pmatrix}\partial\log f_1(y_1,\alpha_1)/\partial\alpha_1\\ \vdots\\ \partial\log f_d(y_d,\alpha_d)/\partial\alpha_d\end{pmatrix},$$
$$\phi_\theta^{\rm 2ML}(y,\eta) = \frac{\partial\log c\,(F_1(y_1,\alpha_1),\cdots,F_d(y_d,\alpha_d),\theta)}{\partial\theta},$$
$$\mathcal{I}_\eta^{\rm 2ML} = \begin{pmatrix}\mathcal{I}_\alpha^{\rm 2ML} & 0\\ (\mathcal{I}_{\alpha,\theta}^{\rm 2ML})^{\rm T} & \mathcal{I}_\theta^{\rm 2ML}\end{pmatrix},$$
$$\mathcal{I}_\alpha^{\rm 2ML} = \begin{pmatrix}\mathcal{I}_{\alpha_1}^{\rm 2ML} & 0 & 0\\ 0 & \ddots & 0\\ 0 & 0 & \mathcal{I}_{\alpha_d}^{\rm 2ML}\end{pmatrix},$$
$$\mathcal{I}_{\alpha_j}^{\rm 2ML} = -{\rm E}_{G_j}\left[H_{\alpha_j}^{\rm 2ML}(y_j,\alpha_{0,j}^{\rm 2ML})\right] = -\int g_j\,\frac{\partial^2\log f_j(y_j,\alpha_{0,j}^{\rm 2ML})}{\partial\alpha_j\,\partial\alpha_j^{\rm T}}\,{\rm d}y_j,$$
$$\mathcal{I}_\theta^{\rm 2ML} = -{\rm E}_G\left[H_\theta^{\rm 2ML}(y,\eta_0^{\rm 2ML})\right] = -\int g\,\frac{\partial^2\log c\left(F_1(y_1,\alpha_{0,1}^{\rm 2ML}),\cdots,F_d(y_d,\alpha_{0,d}^{\rm 2ML}),\theta_0^{\rm 2ML}\right)}{\partial\theta\,\partial\theta^{\rm T}}\,{\rm d}y,$$
$$\mathcal{I}_{\alpha,\theta}^{\rm 2ML} = -{\rm E}_G\left[H_{\alpha,\theta}^{\rm 2ML}(y,\eta_0^{\rm 2ML})\right] = -\int g\,\frac{\partial^2\log c\left(F_1(y_1,\alpha_{0,1}^{\rm 2ML}),\cdots,F_d(y_d,\alpha_{0,d}^{\rm 2ML}),\theta_0^{\rm 2ML}\right)}{\partial\alpha\,\partial\theta^{\rm T}}\,{\rm d}y.$$

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

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

where

$$\begin{aligned}
V_{\text{np},\eta^{\text{2ML}}} &= Q_\eta^{\text{2ML}}\left(\left(\mathcal{I}_\eta^{\text{2ML}}\right)^{-1}\right)^{\text{T}},\\
V_{\eta^{\text{2ML}}} &= \left(\mathcal{I}_\eta^{\text{2ML}}\right)^{-1}K_\eta^{\text{2ML}}\left(\left(\mathcal{I}_\eta^{\text{2ML}}\right)^{-1}\right)^{\text{T}},\\
Q_\eta^{\text{2ML}} &= \text{Cov}_G\left(\text{IF}_T(y,G),\phi_\eta^{\text{2ML}}(y,\eta_0^{\text{2ML}})\right) = \text{E}_G\left[\text{IF}_T(y,G)\cdot\phi_\eta^{\text{2ML}}(y,\eta_0^{\text{2ML}})^{\text{T}}\right],\\
K_\eta^{\text{2ML}} &= \text{Var}_G\left(\phi_\eta^{\text{2ML}}(y,\eta_0^{\text{2ML}})\right) = \text{E}_G\left[\phi_\eta^{\text{2ML}}(y,\eta_0^{\text{2ML}})\,\phi_\eta^{\text{2ML}}(y,\eta_0^{\text{2ML}})^{\text{T}}\right].
\end{aligned}$$

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

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

□

Proof of Proposition 2

Proof. Under the assumed regularity conditions, we have from Theorem 5.5 in Shao (2003), Lemma 1 in Ko & Hjort (2018b) and Proposition 2 from Chen & Fan (2005) that

$$\begin{aligned}
\widehat{T}_{\text{np}}-T_{\text{true}} &= \frac{1}{n}\sum_{i=1}^n \text{IF}_T(y_i,G)+o_p(n^{-1/2}),\\
\widehat{\alpha}^{\text{2ML}}-\alpha_0^{\text{2ML}} &= (\mathcal{I}_\alpha^{\text{2ML}})^{-1}\phi_{n,\alpha}^{\text{2ML}}(\alpha_0^{\text{2ML}})+o_p(n^{-1/2}),\\
\widehat{\theta}^{\text{PML}}-\theta_0^{\text{PML}} &= (\mathcal{I}_\theta^{\text{PML}})^{-1}\phi_{n,\theta}^{\text{PML}}(\theta_0^{\text{PML}})+o_p(n^{-1/2}),
\end{aligned}$$

where

$$\begin{aligned}
\phi_{n,\theta}^{\text{PML}}(\theta) &= \frac{1}{n}\sum_{i=1}^n \phi_\theta^{\text{PML}}(G_{n,1}(y_{i,1}),\cdots,G_{n,d}(y_{i,d}),\theta),\\
G_{n,j}(y_j) &= \frac{1}{n+1}\sum_{k=1}^n I\{y_{k,j}\le y_j\},\\
\phi_\theta^{\text{PML}}(u,\theta) &= \frac{\partial \log c\,(u_1,\cdots,u_d),\theta)}{\partial\theta},\\
\mathcal{I}_\theta^{\text{PML}} &= -\text{E}_{C_0}\left[H_\theta^{\text{PML}}(u,\theta_0^{\text{PML}})\right] = -\int_{[0,1]^d} c_0\,(u)\,\frac{\partial^2\log c\,(u,\theta_0^{\text{PML}})}{\partial\theta\,\partial\theta^{\text{T}}}\,\text{d}u,\\
u &= (u_1,\cdots,u_d) = (G_1(y_1),\cdots,G_d(y_d)).
\end{aligned}$$

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

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

where

$$\begin{aligned}
\widehat{\eta}^{\text{PMLpm}} &= \left((\widehat{\alpha}^{\text{2ML}})^{\text{T}},(\widehat{\theta}^{\text{PML}})^{\text{T}}\right)^{\text{T}},\\
\eta_0^{\text{PMLpm}} &= \left((\alpha_0^{\text{2ML}})^{\text{T}},(\theta_0^{\text{PML}})^{\text{T}}\right)^{\text{T}},\\
\mathcal{I}_\eta^{\text{PMLpm}} &= \begin{pmatrix}\mathcal{I}_\alpha^{\text{2ML}} & 0\\ 0 & \mathcal{I}_\theta^{\text{PML}}\end{pmatrix},\\
V_{\text{np},\eta^{\text{PMLpm}}} &= Q_\eta^{\text{PMLpm}}\left(\mathcal{I}_\eta^{\text{PMLpm}}\right)^{-1},\\
V_{\eta^{\text{PMLpm}}} &= \left(\mathcal{I}_\eta^{\text{PMLpm}}\right)^{-1}K_\eta^{\text{PMLpm}}\left(\mathcal{I}_\eta^{\text{PMLpm}}\right)^{-1},\\
Q_\eta^{\text{PMLpm}} &= \text{Cov}_G\left(\text{IF}_T(y,G),\phi_\eta^{\text{PMLpm}}(y,\eta_0^{\text{PMLpm}})+\mathcal{W}^*(y,\theta_0^{\text{PML}})\right)=\text{E}_G\left[\text{IF}_T(y,G)\cdot\left(\phi_\eta^{\text{PMLpm}}(y,\eta_0^{\text{PMLpm}})+\mathcal{W}^*(y,\theta_0^{\text{PML}})\right)^{\text{T}}\right],\\
K_\eta^{\text{PMLpm}} &= \text{Var}_G\left(\phi_\eta^{\text{PMLpm}}(y,\eta_0^{\text{PMLpm}}+\mathcal{W}^*(y,\theta_0^{\text{PML}}))\right)\\
&= \text{E}_G\left[\left(\phi_\eta^{\text{PMLpm}}(y,\eta_0^{\text{PMLpm}})+\mathcal{W}^*(y,\theta_0^{\text{PML}})\right)\cdot\left(\phi_\eta^{\text{PMLpm}}(y,\eta_0^{\text{PMLpm}})+\mathcal{W}^*(y,\theta_0^{\text{PML}})\right)^{\text{T}}\right],\\
\phi_\eta^{\text{PMLpm}}(y,\eta) &= \begin{pmatrix}\phi_\alpha^{\text{2ML}}(y,\alpha)\\ \phi_\theta^{\text{PML}}(\mathcal{W}^*((G_1(y_1),\cdots,G_d(y_d)),\theta)\end{pmatrix},\\
\mathcal{W}^*(y,\theta) &= \begin{pmatrix}0_{\dim(\alpha)}\\ \mathcal{W}\left((G_1(y_1),\cdots,G_d(y_d)),\theta\right)\end{pmatrix},\\
\mathcal{W}(u,\theta) &= \sum_{j=1}^d \mathcal{W}_j(u_j,\theta)=\sum_{j=1}^d\int_{[0,1]^d} c_0(v)\,\frac{\partial^2\log c(v,\theta)}{\partial\theta\,\partial v_j^{\text{T}}}\left(I(u_j\le v_j)-v_j\right)\,\text{d}v.
\end{aligned}$$

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

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

□

Proof of Lemma 1

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

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

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 $\widehat{Q}(\widehat{\eta}^{\text{PMLpm}})$:

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

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

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

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 $\mathcal{I}_{\alpha_j}^{-1} = K_{\alpha_j}$ (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

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

and

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

So, we can write

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

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

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

which stems from the proof of Proposition 2, and furthermore

$$\begin{aligned}
\phi_{n,\eta}^*(\eta) &= \frac{1}{n}\sum_{i=1}^n \phi_\eta^*(y_i, \eta) = \frac{1}{n}\sum_{i=1}^n \frac{\partial \log c\left(F_1(y_{i,1}, \alpha_1), \cdots, F_d(y_{i,d}, \alpha_d), \theta\right)}{\partial \eta}, \\
H_{n,\eta}^*(\eta) &= \frac{1}{n}\sum_{i=1}^n H_\eta^*(y_i, \eta) = \frac{1}{n}\sum_{i=1}^n \frac{\partial^2 \log c\left(F_1(y_{i,1}, \alpha_1), \cdots, F_d(y_{i,d}, \alpha_d), \theta\right)}{\partial \eta \partial \eta^{\mathrm{T}}}, \\
\mathcal{I}_\eta^* &= -\mathrm{E}_G\left[H_\eta^*(y, \eta_0^{\text{PMLpm}})\right] = -\int g(y) H_\eta^*(y, \eta_0^{\text{PMLpm}})\,\mathrm{d}y.
\end{aligned}$$

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

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

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

$$
+\frac{1}{2n}\sqrt{n}(\widehat{\eta}^{\text{PMLpm}}-\eta_0^{\text{PMLpm}})^{\mathrm{T}}\int g(y)H_\eta^*(y,\eta_0^{\text{PMLpm}})\,\mathrm{d}y\cdot\sqrt{n}(\widehat{\eta}^{\text{PMLpm}}-\eta_0^{\text{PMLpm}})+o_p(n^{-1})
$$

$$
=Q_c(\eta_0^{\text{PMLpm}})-\frac{1}{2n}A_\eta^{\mathrm{T}}\mathcal{I}_\eta^*A_\eta+o_p(n^{-1})
$$

and

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

We now assume that the margins are correctly specified, i.e. $F_j(y_j,\alpha_{0,j}^{\text{2ML}})=G_j(y_j)$. This implies that $\theta_0^{\text{PML}}=\theta_0^{\text{2ML}}$ and consequently $\eta_0^{\text{PMLpm}}=\eta_0^{\text{2ML}}$. Then, the central limit theorem gives

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

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

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

where

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

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 & Hjort (2018a) gives $K_\alpha^\circ=\mathcal{I}_\alpha^{\text{2ML}}-K_\alpha^{\text{2ML}}$.

By using those results, we have

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

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

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

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

□

Appendix B: Results from the simulations study and abalone data

This Appendix contains the results from the simulation study and the abalone data.

Model no.	Estimation	$\widehat{T}$	$10^{-4} \cdot \text{MSE}(\widehat{T})$	$10^{-4} \cdot \text{FIC}$	AIC	Rank_{FIC}	$\text{Rank}_{\text{FIC}}^{-\text{np}}$	Rank_{AIC}	Copula	Margin 1	Margin 2	Margin 3	Margin 4
1	ML	0.1323	4.7453	8.0417	-211.93	1	1	1	Gumbel	Gamma	Gamma	Weibull	Log-normal
1	2ML	0.1312	5.0293	8.0484	-215.97	2	2	2	Gumbel	Gamma	Gamma	Weibull	Log-normal
3	2ML	0.1176	7.0372	8.2209	-301.24	3	3	14	Surv.Clayton	Gamma	Gamma	Weibull	Log-normal
2	2ML	0.1342	5.6744	8.5046	-236.90	4	4	5	Gumbel	Weibull	Weibull	Gamma	Gamma
4	2ML	0.1154	8.1284	8.5721	-332.28	5	5	17	Surv.Clayton	Weibull	Weibull	Gamma	Gamma
1	PML.pm	0.1315	4.9445	8.6823	-216.31	6	6	3	Gumbel	Gamma	Gamma	Weibull	Log-normal
2	ML	0.1398	5.5992	8.9757	-228.40	7	7	4	Gumbel	Weibull	Weibull	Gamma	Gamma
3	ML	0.1207	7.1126	9.6686	-294.28	8	8	13	Surv.Clayton	Gamma	Gamma	Weibull	Log-normal
3	PML.pm	0.1183	6.8651	9.7290	-302.66	9	9	15	Surv.Clayton	Gamma	Gamma	Weibull	Log-normal
2	PML.pm	0.1357	5.7744	9.8151	-237.68	10	10	6	Gumbel	Weibull	Weibull	Gamma	Gamma
4	PML.pm	0.1221	6.6745	10.0568	-336.40	11	11	18	Surv.Clayton	Weibull	Weibull	Gamma	Gamma
4	ML	0.1178	8.1122	10.3724	-315.53	12	12	16	Surv.Clayton	Weibull	Weibull	Gamma	Gamma
np	np	0.1329	7.7549	11.4463	-	13	-	-	np	np	np	np	np
6	2ML	0.1059	15.1220	19.1944	-273.76	14	13	11	Frank	Weibull	Weibull	Gamma	Gamma
6	PML.pm	0.1037	15.8880	20.6127	-275.04	15	14	12	Frank	Weibull	Weibull	Gamma	Gamma
5	2ML	0.0992	17.8678	21.2891	-272.82	16	15	9	Frank	Gamma	Gamma	Weibull	Log-normal
5	PML.pm	0.0985	18.1207	22.1375	-273.56	17	16	10	Frank	Gamma	Gamma	Weibull	Log-normal
6	ML	0.1771	48.7422	53.4333	-262.81	18	17	8	Frank	Weibull	Weibull	Gamma	Gamma
5	ML	0.1803	51.4541	56.3131	-262.59	19	18	7	Frank	Gamma	Gamma	Weibull	Log-normal

Table 4: Results from Part 1 of the simulations with $n = 100$. The results are averaged.

Model no.	Estimation	$\widehat{T}$	$10^{-4} \cdot \mathrm{MSE}(\widehat{T})$	$10^{-4} \cdot \mathrm{FIC}$	AIC	$\mathrm{Rank}_{\mathrm{FIC}}$	$\mathrm{Rank}_{\mathrm{FIC}}^{-\mathrm{np}}$	$\mathrm{Rank}_{\mathrm{AIC}}$	Copula	Margin 1	Margin 2	Margin 3	Margin 4
1	2ML	0.1328	0.5652	0.8786	-2073.55	1	1	2	Gumbel	Gamma	Gamma	Weibull	Log-normal
1	ML	0.1329	0.5018	0.8813	-2069.36	2	2	1	Gumbel	Gamma	Gamma	Weibull	Log-normal
1	PML.pm	0.1328	0.5645	0.9402	-2073.61	3	3	3	Gumbel	Gamma	Gamma	Weibull	Log-normal
2	2ML	0.1361	0.7877	0.9937	-2308.32	4	4	5	Gumbel	Weibull	Weibull	Gamma	Gamma
np	np	0.1327	0.9250	1.1499	-	5	-	-	np	np	np	np	np
2	PML.pm	0.1375	0.9330	1.2455	-2310.28	6	5	6	Gumbel	Weibull	Weibull	Gamma	Gamma
2	ML	0.1417	1.4708	1.6146	-2262.54	7	6	4	Gumbel	Weibull	Weibull	Gamma	Gamma
4	PML.pm	0.1224	1.5327	2.0082	-3339.62	8	7	18	Surv.Clayton	Weibull	Weibull	Gamma	Gamma
3	ML	0.1195	2.0818	2.5168	-2946.18	9	8	13	Surv.Clayton	Gamma	Gamma	Weibull	Log-normal
3	2ML	0.1182	2.4213	2.5317	-2969.28	10	9	14	Surv.Clayton	Gamma	Gamma	Weibull	Log-normal
3	PML.pm	0.1183	2.3987	2.8321	-2970.06	11	10	15	Surv.Clayton	Gamma	Gamma	Weibull	Log-normal
4	2ML	0.1157	3.1902	3.1229	-3314.20	12	11	17	Surv.Clayton	Weibull	Weibull	Gamma	Gamma
4	ML	0.1167	2.8793	3.3248	-3193.82	13	12	16	Surv.Clayton	Weibull	Weibull	Gamma	Gamma
6	2ML	0.1073	7.0037	7.6417	-2687.17	14	13	11	Frank	Weibull	Weibull	Gamma	Gamma
6	PML.pm	0.1057	7.7933	8.5109	-2689.18	15	14	12	Frank	Weibull	Weibull	Gamma	Gamma
5	2ML	0.0999	11.0561	11.7567	-2669.70	16	15	9	Frank	Gamma	Gamma	Weibull	Log-normal
5	PML.pm	0.1000	11.0542	11.8092	-2670.15	17	16	10	Frank	Gamma	Gamma	Weibull	Log-normal
6	ML	0.1827	28.7108	28.7463	-2604.74	18	17	8	Frank	Weibull	Weibull	Gamma	Gamma
5	ML	0.1831	28.9728	28.9566	-2595.25	19	18	7	Frank	Gamma	Gamma	Weibull	Log-normal

Table 5: Results from Part 1 of the simulations with $n = 1000$. The results are averaged.

Model no.	Estimation	$\widehat{T}$	$10^{-4} \cdot \text{MSE}(\widehat{T})$	$10^{-4} \cdot \text{FIC}$	AIC	Rank_{FIC}	$\text{Rank}_{\text{FIC}}^{-\text{np}}$	Rank_{AIC}	Copula	Margin 1	Margin 2
1	2ML	0.3882	5.6478	16.1254	-6871.78	1	1	2	Gaussian	Log-normal	Log-normal
1	ML	0.3882	5.6478	16.5589	-6871.78	2	2	1	Gaussian	Log-normal	Log-normal
1	PML.pm	0.3898	5.5264	16.6036	-6871.80	3	3	3	Gaussian	Log-normal	Log-normal
11	2ML	0.3762	11.7738	17.9116	-7237.69	4	4	29	Surv.Clayton	Gamma	Gamma
2	2ML	0.3833	8.6586	18.0289	-7055.26	5	5	14	Gaussian	Gamma	Gamma
12	2ML	0.3751	12.5833	18.0651	-7403.53	6	6	35	Surv.Clayton	Weibull	Weibull
2	ML	0.3853	8.4380	18.8414	-7055.18	7	7	13	Gaussian	Gamma	Gamma
2	PML.pm	0.3896	8.1054	20.4592	-7055.79	8	8	15	Gaussian	Gamma	Gamma
12	ML	0.3828	11.7219	21.7556	-7401.77	9	9	34	Surv.Clayton	Weibull	Weibull
6	ML	0.3844	12.1560	23.0150	-7194.21	10	10	22	Frank	Weibull	Weibull
3	PML.pm	0.4032	9.6180	23.0546	-7204.27	11	11	26	Gaussian	Weibull	Weibull
11	ML	0.3696	14.6960	23.4908	-7236.84	12	12	28	Surv.Clayton	Gamma	Gamma
np	np	0.3868	24.8964	23.8918	-	13	-	-	np	np	np
6	2ML	0.3617	19.1963	28.4171	-7198.43	14	13	24	Frank	Weibull	Weibull
3	2ML	0.4224	19.3920	29.4495	-7201.46	15	14	25	Gaussian	Weibull	Weibull
8	2ML	0.4419	35.4811	44.7175	-7178.64	16	15	20	Gumbel	Gamma	Gamma
5	ML	0.3264	51.1712	51.4565	-7073.03	17	16	16	Frank	Gamma	Gamma
6	PML.pm	0.3269	49.1468	51.9458	-7208.65	18	17	27	Frank	Weibull	Weibull
3	ML	0.4457	40.0698	52.4882	-7197.51	19	18	23	Gaussian	Weibull	Weibull
9	2ML	0.4575	54.9298	62.1409	-7342.12	20	19	32	Gumbel	Weibull	Weibull
4	ML	0.3121	67.0025	65.1113	-6909.88	21	20	4	Frank	Log-normal	Log-normal
8	ML	0.4583	55.9058	69.9129	-7176.83	22	21	19	Gumbel	Gamma	Gamma
5	2ML	0.3097	74.3431	72.9594	-7074.76	23	22	17	Frank	Gamma	Gamma
4	PML.pm	0.3044	79.2754	77.3873	-6910.60	24	23	6	Frank	Log-normal	Log-normal
4	2ML	0.3042	79.6418	77.5446	-6910.55	25	24	5	Frank	Log-normal	Log-normal
5	PML.pm	0.3012	87.9706	86.5350	-7075.58	26	25	18	Frank	Gamma	Gamma
10	2ML	0.4951	117.6930	125.3261	-6988.12	27	26	11	Surv.Clayton	Log-normal	Log-normal
7	2ML	0.4983	122.7807	131.7212	-6922.68	28	27	8	Gumbel	Log-normal	Log-normal
10	PML.pm	0.4979	122.9565	133.5784	-6988.51	29	28	12	Surv.Clayton	Log-normal	Log-normal
7	PML.pm	0.4999	126.0077	136.9588	-6922.81	30	29	9	Gumbel	Log-normal	Log-normal
12	PML.pm	0.5011	132.4155	145.6955	-7449.37	31	30	36	Surv.Clayton	Weibull	Weibull
11	PML.pm	0.5013	133.2260	145.9652	-7284.64	32	31	30	Surv.Clayton	Gamma	Gamma
8	PML.pm	0.5030	135.9110	148.9554	-7193.07	33	32	21	Gumbel	Gamma	Gamma
10	ML	0.5038	137.1085	149.7742	-6985.99	34	33	10	Surv.Clayton	Log-normal	Log-normal
9	PML.pm	0.5041	138.2748	151.9813	-7350.06	35	34	33	Gumbel	Weibull	Weibull
9	ML	0.5040	139.9657	155.2434	-7333.02	36	35	31	Gumbel	Weibull	Weibull
7	ML	0.5108	151.7796	164.4415	-6921.32	37	36	7	Gumbel	Log-normal	Log-normal

Table 6: Results from Part 2 of the simulations for the FIC with $p = 0.90$. The results are averaged.

Model no.	Estimation	$\widehat{T}$	$10^{-4} \cdot \text{MSE}(\widehat{T})$	$10^{-4} \cdot \text{FIC}$	AIC	Rank_{FIC}	$\text{Rank}_{\text{FIC}}^{-\text{np}}$	Rank_{AIC}	Copula	Margin 1	Margin 2
1	2ML	0.3086	6.3295	20.3392	-6871.78	1	1	2	Gaussian	Log-normal	Log-normal
1	ML	0.3086	6.3295	20.6200	-6871.78	2	2	1	Gaussian	Log-normal	Log-normal
1	PML.pm	0.3102	6.2141	20.9331	-6871.80	3	3	3	Gaussian	Log-normal	Log-normal
3	2ML	0.3127	12.8982	27.9917	-7201.46	4	4	25	Gaussian	Weibull	Weibull
3	PML.pm	0.2926	14.0941	28.9935	-7204.27	5	5	26	Gaussian	Weibull	Weibull
2	PML.pm	0.2776	20.1145	29.7261	-7055.79	6	6	15	Gaussian	Gamma	Gamma
2	2ML	0.2712	24.6047	30.5281	-7055.26	7	7	14	Gaussian	Gamma	Gamma
2	ML	0.2731	23.2402	30.5653	-7055.18	8	8	13	Gaussian	Gamma	Gamma
11	2ML	0.3293	18.3423	30.7280	-7237.69	9	9	29	Surv.Clayton	Gamma	Gamma
12	2ML	0.3290	20.8156	32.3012	-7403.53	10	10	35	Surv.Clayton	Weibull	Weibull
11	ML	0.3233	17.9161	33.9353	-7236.84	11	11	28	Surv.Clayton	Gamma	Gamma
3	ML	0.3412	23.0161	39.6736	-7197.51	12	12	23	Gaussian	Weibull	Weibull
12	ML	0.3386	27.7021	42.4621	-7401.77	13	13	34	Surv.Clayton	Weibull	Weibull
np	np	0.3065	41.2571	43.6136	-	14	-	-	np	np	np
8	2ML	0.4049	102.8917	108.4323	-7178.64	15	14	20	Gumbel	Gamma	Gamma
6	ML	0.1983	136.2531	127.7135	-7194.21	16	15	22	Frank	Weibull	Weibull
4	ML	0.1850	161.6738	148.5580	-6909.88	17	16	4	Frank	Log-normal	Log-normal
8	ML	0.4210	137.7982	148.9792	-7176.83	18	17	19	Gumbel	Gamma	Gamma
9	2ML	0.4243	147.6641	150.5689	-7342.12	19	18	32	Gumbel	Weibull	Weibull
4	2ML	0.1776	180.3702	166.3138	-6910.55	20	19	5	Frank	Log-normal	Log-normal
4	PML.pm	0.1778	179.9556	166.3638	-6910.60	21	20	6	Frank	Log-normal	Log-normal
6	2ML	0.1744	193.7918	180.9205	-7198.43	22	21	24	Frank	Weibull	Weibull
5	ML	0.1529	254.7066	239.1068	-7073.03	23	22	16	Frank	Gamma	Gamma
6	PML.pm	0.1532	253.6577	239.2535	-7208.65	24	23	27	Frank	Weibull	Weibull
8	PML.pm	0.4694	267.6935	280.5936	-7193.07	25	24	21	Gumbel	Gamma	Gamma
5	2ML	0.1389	299.3806	281.8677	-7074.76	26	25	17	Frank	Gamma	Gamma
7	2ML	0.4771	285.8960	291.9153	-6922.68	27	26	8	Gumbel	Log-normal	Log-normal
9	ML	0.4713	279.0023	292.4440	-7333.02	28	27	31	Gumbel	Weibull	Weibull
11	PML.pm	0.4732	282.4024	295.9925	-7284.64	29	28	30	Surv.Clayton	Gamma	Gamma
10	2ML	0.4784	292.8752	297.8630	-6988.12	30	29	11	Surv.Clayton	Log-normal	Log-normal
5	PML.pm	0.1342	315.3396	298.0236	-7075.58	31	30	18	Frank	Gamma	Gamma
7	PML.pm	0.4788	291.2749	300.6144	-6922.81	32	31	9	Gumbel	Log-normal	Log-normal
9	PML.pm	0.4746	288.3063	303.3531	-7350.06	33	32	33	Gumbel	Weibull	Weibull
10	PML.pm	0.4815	302.4989	312.7326	-6988.51	34	33	12	Surv.Clayton	Log-normal	Log-normal
12	PML.pm	0.4781	301.7405	317.2543	-7449.37	35	34	36	Surv.Clayton	Weibull	Weibull
10	ML	0.4880	326.6800	337.9244	-6985.99	36	35	10	Surv.Clayton	Log-normal	Log-normal
7	ML	0.4893	328.3196	339.4458	-6921.32	37	36	7	Gumbel	Log-normal	Log-normal

Table 7: Results from Part 2 of the simulations for the FIC with $p = 0.95$. The results are averaged.

Model no.	Estimation	$\widehat{T}$	$10^{-4} \cdot \text{AMSE}(\widehat{T})$	$10^{-4} \cdot \text{AFIC}$	AAIC	$\text{Rank}_{\text{AFIC}}$	$\text{Rank}^{-\text{np}}_{\text{AFIC}}$	$\text{Rank}_{\text{AAIC}}$	Copula	Margin 1	Margin 2
1	2ML	0.3506	6.0039	13.4946	-6871.78	1	1	2	Gaussian	Log-normal	Log-normal
1	ML	0.3506	6.0039	13.8241	-6871.78	2	2	1	Gaussian	Log-normal	Log-normal
1	PML.pm	0.3522	5.8839	13.9772	-6871.80	3	3	3	Gaussian	Log-normal	Log-normal
11	2ML	0.3535	13.2986	17.7845	-7237.69	4	4	29	Surv.Clayton	Gamma	Gamma
2	2ML	0.3309	14.7284	18.1211	-7055.26	5	5	14	Gaussian	Gamma	Gamma
12	2ML	0.3525	14.6442	18.1680	-7403.53	6	6	35	Surv.Clayton	Weibull	Weibull
2	ML	0.3328	13.9672	18.5785	-7055.18	7	7	13	Gaussian	Gamma	Gamma
2	PML.pm	0.3372	12.3515	19.0526	-7055.79	8	8	15	Gaussian	Gamma	Gamma
3	PML.pm	0.3517	10.4913	20.1153	-7204.27	9	9	26	Gaussian	Weibull	Weibull
11	ML	0.3472	14.6029	21.4949	-7236.84	10	10	28	Surv.Clayton	Gamma	Gamma
3	2ML	0.3714	15.4486	22.1663	-7201.46	11	11	25	Gaussian	Weibull	Weibull
12	ML	0.3612	17.0221	24.0154	-7401.77	12	12	34	Surv.Clayton	Weibull	Weibull
np	np	0.3490	30.6032	32.1526	-	13	-	-	np	np	np
3	ML	0.3973	31.9702	40.0586	-7197.51	14	13	23	Gaussian	Weibull	Weibull
6	ML	0.2970	55.9162	52.9555	-7194.21	15	14	22	Frank	Weibull	Weibull
8	2ML	0.4237	63.5797	65.4443	-7178.64	16	15	20	Gumbel	Gamma	Gamma
6	2ML	0.2731	87.3558	80.1738	-7198.43	17	16	24	Frank	Weibull	Weibull
9	2ML	0.4407	93.2461	93.9918	-7342.12	18	17	32	Gumbel	Weibull	Weibull
8	ML	0.4399	90.5960	97.8883	-7176.83	19	18	19	Gumbel	Gamma	Gamma
4	ML	0.2512	110.5606	100.5473	-6909.88	20	19	4	Frank	Log-normal	Log-normal
4	PML.pm	0.2436	126.1196	115.5180	-6910.60	21	20	6	Frank	Log-normal	Log-normal
4	2ML	0.2434	126.5241	115.6843	-6910.55	22	21	5	Frank	Log-normal	Log-normal
6	PML.pm	0.2441	136.3455	127.0975	-7208.65	23	22	27	Frank	Weibull	Weibull
5	ML	0.2432	138.9328	128.6879	-7073.03	24	23	16	Frank	Gamma	Gamma
5	2ML	0.2273	174.5574	162.4928	-7074.76	25	24	17	Frank	Gamma	Gamma
5	PML.pm	0.2205	190.6471	178.6701	-7075.58	26	25	18	Frank	Gamma	Gamma
10	2ML	0.4868	193.7567	197.3935	-6988.12	27	26	11	Surv.Clayton	Log-normal	Log-normal
7	2ML	0.4876	193.4383	198.1757	-6922.68	28	27	8	Gumbel	Log-normal	Log-normal
8	PML.pm	0.4866	193.6220	203.1375	-7193.07	29	28	21	Gumbel	Gamma	Gamma
7	PML.pm	0.4893	197.6532	205.2236	-6922.81	30	29	9	Gumbel	Log-normal	Log-normal
10	PML.pm	0.4897	201.0107	209.2503	-6988.51	31	30	12	Surv.Clayton	Log-normal	Log-normal
11	PML.pm	0.4878	199.1494	209.4146	-7284.64	32	31	30	Surv.Clayton	Gamma	Gamma
9	ML	0.4877	199.6092	210.3643	-7333.02	33	32	31	Gumbel	Weibull	Weibull
9	PML.pm	0.4892	202.2458	212.8958	-7350.06	34	33	33	Gumbel	Weibull	Weibull
12	PML.pm	0.4895	205.2491	216.5513	-7449.37	35	34	36	Surv.Clayton	Weibull	Weibull
10	ML	0.4959	219.7997	228.7910	-6985.99	36	35	10	Surv.Clayton	Log-normal	Log-normal
7	ML	0.5000	228.7504	237.6216	-6921.32	37	36	7	Gumbel	Log-normal	Log-normal

Table 8: Results from Part 2 of the simulations for the AFIC based on $p \in \{0.90, 0.91, \cdots, 0.95\}$. The results are averaged.

Model no.	Estimation	$\widehat{T}$	$10^{-4} \cdot$ FIC	AIC	Rank_{FIC}	$\text{Rank}_{\text{FIC}}^{-\text{np}}$	Rank_{AIC}	Copula	Margin 1	Margin 2	Margin 3	Margin 4
np	np	0.8235	2.0354	-	1	-	-	np	np	np	np	np
6	2ML	0.7902	10.2261	24219.02	2	1	10	Gaussian	Normal	Normal	Normal	Log-normal
3	2ML	0.7827	16.9604	24839.29	3	2	9	t	Normal	Normal	Normal	Log-normal
5	2ML	0.7811	17.0115	28994.80	4	3	7	Gaussian	Weibull	Weibull	Weibull	Log-normal
6	PML.pm	0.7804	17.7474	21602.45	5	4	16	Gaussian	Normal	Normal	Normal	Log-normal
3	PML.pm	0.7761	21.5499	22761.26	6	5	15	t	Normal	Normal	Normal	Log-normal
2	2ML	0.7734	24.1082	30108.31	7	6	3	t	Weibull	Weibull	Weibull	Log-normal
5	PML.pm	0.7730	24.6204	28958.97	8	7	8	Gaussian	Weibull	Weibull	Weibull	Log-normal
2	PML.pm	0.7685	29.2406	30074.78	9	8	4	t	Weibull	Weibull	Weibull	Log-normal
4	2ML	0.7432	63.7877	29114.80	10	9	5	Gaussian	Weibull	Weibull	Weibull	NIG
1	2ML	0.7401	68.8440	30257.06	11	10	1	t	Weibull	Weibull	Weibull	NIG
4	PML.pm	0.7381	72.3696	29079.53	12	11	6	Gaussian	Weibull	Weibull	Weibull	NIG
1	PML.pm	0.7369	74.2642	30230.77	13	12	2	t	Weibull	Weibull	Weibull	NIG
7	2ML	0.9465	149.5032	23729.98	14	13	11	Clayton	Weibull	Weibull	Weibull	NIG
9	2ML	0.9473	151.4692	20314.83	15	14	17	Clayton	Normal	Normal	Normal	Log-normal
7	PML.pm	0.9528	165.3779	23683.19	16	15	12	Clayton	Weibull	Weibull	Weibull	NIG
8	2ML	0.9556	172.6200	23107.05	17	16	13	Clayton	Weibull	Weibull	Weibull	Log-normal
8	PML.pm	0.9632	193.2900	23026.99	18	17	14	Clayton	Weibull	Weibull	Weibull	Log-normal
9	PML.pm	0.9647	197.4406	19942.89	19	18	18	Clayton	Normal	Normal	Normal	Log-normal

Table 9: Results for the models fitted to the abalone data.