cemmap

centre for microdata methods and practice

Efficient estimation of copula-based semiparametric Markov models

Xiaohong Chen
Wei Biao Wu
Yanping Yi

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP06/09

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

An ESRC Research Centre

Efficient Estimation of Copula-Based Semiparametric Markov Models[1]

Xiaohong Chen[a,2] , Wei Biao Wu[b], Yanping Yi[c]

[a] *Cowles Foundation for Research in Economics, Yale University, Box 208281, New Haven, CT 06520, USA*

[b] *Department of Statistics, University of Chicago, 5734 S. University Ave., Chicago, IL 60637, USA*

[c] *Department of Economics, New York University, 19 West 4th street, New York, NY 10012, USA*

First version: October 2007; This version: January 2009

Abstract

This paper considers efficient estimation of copula-based semiparametric strictly stationary Markov models. These models are characterized by nonparametric invariant distributions and parametric copula functions; where the copulas capture all scale-free temporal dependence and tail dependence of the processes. The Markov models generated via tail dependent copulas may look highly persistent and are useful for financial and economic applications. We first show that Markov processes generated via Clayton, Gumbel and Student's t copulas (with tail dependence) are all geometric ergodic. We then propose a sieve maximum likelihood estimation (MLE) for the copula parameter, the invariant distribution and the conditional quantiles. We show that the sieve MLEs of any smooth functionals are root-n consistent, asymptotically normal and efficient; and that the sieve likelihood ratio statistics is chi-square distributed. We present Monte Carlo studies to compare the finite sample performance of the sieve MLE, the two-step estimator of Chen and Fan (2006), the correctly specified parametric MLE and the incorrectly specified parametric MLE. The simulation results indicate that our sieve MLEs perform very well; having much smaller biases and smaller variances than the two-step estimator for Markov models generated by Clayton, Gumbel and other copulas having strong tail dependence.

JEL classification: C14; C22
AMS 2000 subject classifications: Primary 62M05; secondary 62F07

Keywords: Copula; Tail dependence; Nonlinear Markov models; Geometric ergodicity; Sieve MLE; Semiparametric efficiency; Sieve likelihood ratio statistics; Value-at-Risk

[1] This paper is dedicated to Professor Peter C. B. Phillips on the occasion of his 60th birthday.
We thank Don Andrews, Brendan Beare, Xu Cheng, Rohit Deo, Rustam Ibragimov, Andrew Patton, Peter Phillips, Demian Pouzo, Zhijie Xiao for very useful discussions, and Zhentao Shi for excellent research assistance on Monte Carlo studies. We also acknowledge helpful comments from conference participants at 2007 Stochastics and Dependence in Finance, Risk Management and Insurance at the Radcliffe Institute for Advanced Study at Harvard University, 2008 International Symposium on Financial Engineering and Risk Management (FERM), China, 2008 Conference in honor of Peter Phillips, SMU, and 2008 Oxford-Man Institute Symposium on Modelling Multivariate Dependence and Extremes in Finance.

[2] Corresponding author. Tel.: (203) 432-5852; fax: (212) 995-4186. Supported by the National Science Foundation SES-0631613. E-mail: xiaohong.chen@yale.edu

1 Introduction

A copula function is a multivariate probability distribution function with uniform marginals. Copula-based method has become one popular tool in modeling nonlinear, asymmetric and tail dependence in financial and insurance risk managements. See Embrechts, et al. (2002), Embrechts (2008), Genest et al. (2008), Patton (2002, 2006, 2008) and the references therein for reviews of various theoretical properties and financial applications of the copula approach.

While the majority of the previous work using copulas have focused on modeling the contemporaneous dependence between multiple univariate series, there are also a growing number of papers using copulas to model the temporal dependence of a univariate nonlinear time series. Granger (2003) suggests to define persistence (such as 'long memory' or 'short memory') for general nonlinear time series models via copulas. Darsow, et al. (1992), de la Pena et al. (2006) and Ibragimov (2009) provide characterizations of a copula-based time series to be a Markov process. Joe (1997) proposes a class of parametric (strictly) stationary Markov models based on parametric copulas and parametric invariant (or marginal) distributions. Chen and Fan (2006) study a class of semiparametric stationary Markov models based on parametric copulas and nonparametric invariant distributions.

Let $\{Y_t\}$ be a stationary Markov process of order one with a continuous invariant distribution G. Then its probabilistic properties are completely determined by the bivariate joint distribution function of Y_{t-1} and Y_t, $H(y_1, y_2)$ (say). By Sklar's theorem, one can uniquely express $H(\cdot, \cdot)$ in terms of the invariant distribution G and the copula function $C(\cdot, \cdot)$ of Y_{t-1} and Y_t:

$$H(y_1, y_2) \equiv C(G(y_1), G(y_2)).$$

Thus one can always specify a stationary first order Markov model with continuous state space by directly specifying the marginal distribution of Y_t and the copula function of Y_{t-1} and Y_t. The advantage of the copula approach is that one can freely choose the marginal distribution and the copula function separately; the former characterizes the marginal behavior such as the fat-tails of the time series $\{Y_t\}_{t=1}^n$, while the latter characterizes all the scale-free temporal dependence and tail dependence properties of the time series. Although being strictly stationary first-order Markov, a model generated via a copula (especially a tail dependent copula) is very flexible. This model can generate a rich array of nonlinear time series patterns, including persistent clustering of extreme values via tail dependent copulas evaluated at fat-tailed marginals, asymmetric dependence, and other "look alike" behaviors present in many popular nonlinear models such as Arch, Garch, stochastic volatility, near-unit root, long-memory, structural break, Markov switching, etc. From the point of view of financial applications, one attractive property of the copula-based Markov model is that the implied (nonlinear) conditional quantiles (value-at-risks) are automatically monotonic across quantiles. This nice feature has been exploited by Chen et al. (2008) and Bouye and Salmon (2008) in their study of copula-based nonlinear quantile autoregression and Value at risk.

In this paper, we shall focus on the class of copula-based, strictly stationary, semiparametric first order Markov models, in which the true copula density function has a parametric form ($c(\cdot, \cdot; \alpha_0)$), and the true invariant distribution is of an unknown form ($G_0(\cdot)$) but is absolutely continuous with respect to the Lebesgue measure on the real line. Any model of this class is completely characterized by two unknown parameters: the copula dependence parameter α_0 and the invariant distribution $G_0(\cdot)$. To establish the asymptotic properties of any semiparametric estimators of (α_0, G_0), one needs to know temporal dependence properties of the copula-based Markov models. For this class of models, Chen and Fan (2006) show that the beta-mixing temporal dependence measure is purely determined by the properties of copulas (and does not depend on the invariant distributions); and Beare (2008) provides sufficient conditions for geometric beta-mixing in terms of copulas without any tail dependence (such as Gaussian and Frank copulas). Neither paper is able to verify whether or not a Markov process generated via a tail dependent copula (such as Clayton, survival Clayton, Gumbel, survival Gumbel, Student's t) is geometric beta-mixing. Ibragimov and Lentzas (2008) demonstrate via simulation that Clayton copula-based first order strictly stationary Markov models could behave as 'long memory' in copula levels. In this paper, we show that Clayton, survival Clayton, Gumbel, survival Gumbel, Student's t copula based Markov models are actually geometric ergodic (hence geometric beta-mixing). Therefore, according to our this theorem, although a time series plot of a Clayton copula (or survival Clayton, Gumbel, survival Gumbel, other tail dependent copula) generated Markov model may look highly persistent and 'long memory alike', it is in fact weakly dependent and 'short memory'.

In this paper, we propose a sieve maximum likelihood estimation (MLE) procedure for the copula parameter α_0, the invariant distribution G_0 and the conditional quantiles of a copula-based semiparametric Markov model. We show that the sieve MLEs of any smooth functionals are root-n consistent, asymptotically normal and efficient; and that the sieve likelihood ratio statistics is chi-square distributed. It is interesting to note that although the conditional distribution of a copula-based semiparametric stationary Markov model depends on the unknown invariant distribution, the plug-in sieve MLE estimators of the nonlinear conditional quantiles (VaR) are still $\sqrt{n}$-consistent, asymptotically normal and efficient.

To the best of our knowledge, Atlason (2008) is the only other paper that also consider the semiparametric efficient estimation of a copula parameter α_0 for a copula-based first-order strictly stationary Markov model. His work and our work have been carried through independently but are around the same time. While we propose sieve likelihood joint estimation of G_0 and α_0, Atlason (2008) proposes rank likelihood estimation of the copula parameter α_0, and relies on simulation method to evaluate his rank likelihood. However, Atlason (2008) does not investigate semiparametric efficient estimation of the invariant distribution G_0 nor the conditional quantiles.

Previously, Chen and Fan (2006) propose a simple two-step estimation procedure, in which one first estimates the invariant cdf $G_0(\cdot)$ by a re-scaled empirical cdf G_n of the data $\{Y_t\}_{t=1}^n$, and then estimate the copula parameter α_0 by maximizing the pseudo log-likelihood corresponding to copula density evaluated at pseudo observations $\{G_n(Y_t)\}_{t=1}^n$. Their procedure can be viewed as an extension of the one proposed by Genest et al. (1995) for a bivariate copula-based joint distribution model of a random sample $\{(X_i, Y_i)\}_{i=1}^n$ to a univariate first-order Markov model of a time series data $\{Y_i\}_{i=1}^n$ (with $X_i = Y_{i-1}$). Just as the two-step estimator of Genest et al. (1995) is generally inefficient for a bivariate random sample (see, e.g., Genest and Werker (2001)), the two-step estimator of Chen and Fan (2006) is inefficient for a univariate Markov model.

We present Monte Carlo studies to compare the finite sample performance of our sieve MLE, the two-step estimator of Chen and Fan (2006), the correctly specified parametric MLE and the incorrectly specified parametric MLE for Clayton and Gumbel copula-based Markov models. Numerous simulation studies demonstrate that the two-step estimator of Chen and Fan (2006) is not only inefficient but also severely biased (in finite sample) when the time series has strong tail dependence, and it leads to a biased and inefficient plug-in estimator of conditional quantiles (or VaR). The simulation results indicate that our sieve MLEs perform very well; when the copula-based Markov process has strong tail dependence, the sieve MLEs have much smaller biases and smaller variances than the two-step estimators.

The rest of this paper is organized as follows. In Section 2, we present the class of copula-based semiparametric strictly stationary Markov models, and show that several widely used tailed dependent copula (Clayton, Gumbel, Student's t) based Markov models are geometric beta-mixing. In Section 3, we introduce the sieve MLE, and obtain its consistency and rate of convergence. Section 4 establishes the asymptotic normality and semiparametric efficiency of the sieve MLE. Section 5 shows that the sieve maximum likelihood ratio statistics is asymptotically chi-square distributed, which suggests a simple way to construct confidence region for copula parameter and other smooth functionals. In Section 6, we first review some popular existing estimators (the two-step estimator, the correctly specified parametric MLE, the misspecified parametric MLE, the infeasible MLE). We then conduct some simulation studies to compare the finite sample performance of our sieve MLE vs these alternative estimators. Section 7 briefly concludes. All the proofs are relegated to the Appendix.

2 Copula-Based Markov Models

In this section we first present the model, and then some implied temporal dependence properties.

2.1 The model

Throughout this paper, we assume the true data generating process (DGP) satisfies the following assumption:

Assumption M (DGP): (1) $\{Y_t : t = 1, \cdots, n\}$ is a sample of a strictly stationary first order Markov process generated from $(G_0(\cdot), C(\cdot, \cdot; \alpha_0))$, where $G_0(\cdot)$ is the true invariant distribution which is absolutely continuous with respect to Lebesgue measure on the real line (with its support $\mathcal{Y}$ a nonempty interval of $\mathcal{R}$); $C(\cdot, \cdot; \alpha_0)$ is the true parametric copula for (Y_{t-1}, Y_t) up to unknown value α_0, is absolutely continuous with respect to Lebesgue measure on $[0,1]^2$, and is neither the Fréchet-Hoeffding upper ($C(u_1, u_2) = \min(u_1, u_2)$) nor the lower ($C(u_1, u_2) = \max(u_1 + u_2 - 1, 0)$) bound. (2) the true marginal density $g_0(\cdot)$ of $G_0(\cdot)$ is positive on the interior of its support $\mathcal{Y}$; and the true copula density $c(\cdot, \cdot; \alpha_0)$ of $C(\cdot, \cdot; \alpha_0)$ is positive on $(0,1)^2$.

Under Assumption M (1), the true conditional probability density function, $p^0(Y_t|Y^{t-1})$ of Y_t given $Y^{t-1} \equiv (Y_{t-1}, ..., Y_1)$ is given by:

$$p^0(Y_t|Y^{t-1}) = h_0(Y_t|Y_{t-1}) \equiv g_0(Y_t)c(G_0(Y_{t-1}), G_0(Y_t); \alpha_0),$$

where $h_0(\cdot|Y_{t-1})$ denotes the true conditional density of Y_t given Y_{t-1}. We note that the conditional density is a function of both copula and marginal; hence the $q-$th conditional quantile of Y_t given Y^{t-1} is also a function of both copula and marginal:

$$Q_q^Y(y) = G_0^{-1}\left(C_{2|1}^{-1}\left[q|G_0(y); \alpha_0\right]\right)$$

where $C_{2|1}[\cdot|u; \alpha_0] \equiv \frac{\partial}{\partial u_1} C(u, \cdot; \alpha_0) \equiv C_1(u, \cdot; \alpha_0)$ is the conditional distribution of $U_t \equiv G_0(Y_t)$ given $U_{t-1} = u$; and $C_{2|1}^{-1}[q|u; \alpha_0]$ is the $q-$th conditional quantile of U_t given $U_{t-1} = u$.

Under assumption M (1), we have that the transformed process $\{U_t : U_t \equiv G_0(Y_t)\}$ is also a strictly stationary first order Markov process with uniform marginals and $C(\cdot, \cdot; \alpha_0)$ the joint distribution of U_{t-1} and U_t. Chen and Fan (2006) express any copula-based first-order strictly stationary Markov model for $\{Y_t\}$ in terms of the following semiparametric transformation autoregression model for the transformed process $\{U_t\}$:

$$\Lambda_1(U_t) = \Lambda_2(U_{t-1}) + \varepsilon_t, \quad E\{\varepsilon_t|U_{t-1}, ..., U_1\} = 0,$$

where $\Lambda_1(\cdot)$ is an increasing function, $\Lambda_2(U_{t-1}) \equiv E\{\Lambda_1(U_t)|U_{t-1}, ..., U_1\}$, and the conditional density of ε_t given $U^{t-1} \equiv (U_{t-1}, ..., U_1)$ satisfies:

$$f_{\varepsilon_t|U^{t-1}}(\varepsilon) = c(U_{t-1}, \Lambda_1^{-1}(\varepsilon + \Lambda_2(U_{t-1})); \alpha_0) \div \frac{d\Lambda_1(\varepsilon + \Lambda_2(U_{t-1}))}{d\varepsilon}.$$

2.2 Temporal dependence properties

All the scale-free dependence measures can be expressed in terms of copulas (see Nelson (2006), Joe (1997)). For example, Kendall's tau is $\tau \equiv 4\int\int_{[0,1]^2} C(u_1,u_2)dC(u_1,u_2) - 1$. The lower tail ($\lambda_L$) and upper tail ($\lambda_U$) dependence in terms of copulas are respectively

$$\begin{aligned}\lambda_L &\equiv \lim_{u\to 0^+} \Pr(U_2 \le u | U_1 \le u) = \lim_{u\to 0^+} \frac{C(u,u)}{u}, \quad \text{and} \\ \lambda_U &\equiv \lim_{u\to 1^-} \Pr(U_2 \ge u | U_1 \ge u) = \lim_{u\to 1^-} \frac{1-2u+C(u,u)}{1-u}\end{aligned}$$

provided the limits exist.

For analyzing asymptotic properties of any semiparametric estimators of (α_0, G_0), it is convenient to apply empirical processes results for strictly stationary geometric ergodic (or geometric beta mixing) Markov processes. In the following we recall the equivalent definitions of beta-mixing and ergodicity for strictly stationary Markov process:

Definition 2.1. *(1) (Davydov, 1973) For a strictly stationary Markov process $\{Y_t\}_{t=1}^{\infty}$, the $\beta-$mixing coefficients are given by:*

$$\beta_t = \int \sup_{0\le\phi\le 1} \left| E[\phi(Y_{t+1})|Y_1 = y] - E[\phi(Y_{t+1})] \right| dG_0(y).$$

The process $\{Y_t\}$ is $\beta-$mixing if $\lim_{t\to\infty}\beta_t = 0$; is $\beta-$mixing with exponential decay rate if $\beta_t \le \gamma\exp(-\delta t)$ for some $\delta, \gamma > 0$; and is $\beta-$mixing with sub-exponential decay rate if $\lim_{t\to\infty}\xi_t\beta_t = 0$ for some positive non-decreasing rate function ξ satisfying $\xi_t \to \infty$, $t^{-1}\ln\xi_t \to 0$ as $t \to \infty$.

(2) (Chan and Tong, 2001) A strictly stationary Markov process $\{Y_t\}$ is (Harris) ergodic if

$$\lim_{t\to\infty} \sup_{0\le\phi\le 1} \left| E[\phi(Y_{t+1})|Y_1 = y] - E[\phi(Y_{t+1})] \right| = 0 \textit{ for almost all } y;$$

is geometrically ergodic if there exist a measurable function W with $\int W(y)dG_0(y) < \infty$ and a constant $\kappa \in [0,1)$ such that for all $t \ge 1$,

$$\sup_{0\le\phi\le 1} \left| E[\phi(Y_{t+1})|Y_1 = y] - E[\phi(Y_{t+1})] \right| \le \kappa^t W(y)$$

Remark 2.1: (1) Under assumption M, the time series $\{Y_t\}_{t=1}^{n}$ is strictly stationary ergodic and is also beta-mixing. See, e.g., Bradley (2005, corollary 3.6) and Chen and Fan (2006).

(2) Proposition 2.1 of Chen and Fan (2006) presents some high-level sufficient conditions in terms of a copula to ensure beta-mixing decaying either exponentially fast or polynomially fast.

(3) Beare (2008, theorems 3.1 and 4.2) shows that all symmetric copulas whose copula densities are bounded away from zero and square integrable are geometric beta-mixing. However, Beare

(2008, theorem 3.2) also shows that all copulas with square integrable densities do not have any tail dependence.

Both Chen and Fan (2006) and Beare (2008) point out that Gaussian copula (which has no tail dependence) generated Markov model is geometric beta mixing, but neither is able to verify whether any tail dependent copulas (such as Clayton copula) generated Markov models are still geometric beta-mixing.

For financial risk management, the Markov models generated via tail-dependent copulas are much more relevant than models without tail dependence. In particular, the following three examples have been widely used in financial applications:

Example 2.1 (Clayton copula-based Markov model): The bivariate Clayton copula is

$$C(u_1, u_2, \alpha) = \left[u_1^{-\alpha} + u_2^{-\alpha} - 1\right]^{-1/\alpha}, \ 0 \leq \alpha < \infty.$$

Clayton copula has Kendall's tau $\tau = \frac{\alpha}{2+\alpha}$, and lower tail dependence $\lambda_L = 2^{-1/\alpha}$ that is increasing in α, but no upper tail dependence. Clayton copula becomes the independence copula $C_I(u_1, u_2) = u_1 u_2$ when $\alpha = 0$.

Example 2.2 (Gumbel copula-based Markov model): The bivariate Gumbel copula is

$$C(u_1, u_2; \alpha) = \exp(-[(-\ln u_1)^\alpha + (-\ln u_2)^\alpha]^{1/\alpha}), \quad 1 \leq \alpha < \infty.$$

Gumbel copula has Kendall's tau $\tau = 1 - \frac{1}{\alpha}$, and upper tail dependence $\lambda_U = 2 - 2^{1/\alpha}$ that is increasing in α, but no lower tail dependence. Gumbel copula becomes the independence copula $C_I(u_1, u_2) = u_1 u_2$ when $\alpha = 1$.

Example 2.3 (Student t copula-based Markov model): The bivariate Student $t-$ copula is

$$C(u_1, u_2; \alpha) = \mathbf{t}_{\nu,\rho}(t_\nu^{-1}(u_1), t_\nu^{-1}(u_2)), \quad \alpha = (\nu, \rho), \ |\rho| < 1, \ \nu \in (1, \infty],$$

where $\mathbf{t}_{\nu,\rho}(\cdot,\cdot)$ is the bivariate Student-t distribution with mean zeros, correlation matrix having off-diagonal element ρ, and degrees of freedom ν, and $t_\nu(\cdot)$ is the cdf of a univariate Student-t distribution with mean zero, and degrees of freedom ν. Student t copula has Kendall's tau $\tau = \frac{2}{\pi} \arcsin \rho$, and symmetric tail dependence: $\lambda_L = \lambda_U = 2t_{\nu+1}(-\sqrt{(\nu+1)(1-\rho)/(1+\rho)})$ that is decreasing in ν. Student t copula becomes Gaussian copula when $\nu = \infty$.

Ibragimov and Lentzas (2008) demonstrate via simulation that Clayton copula generated first order strictly stationary Markov models behave as 'long memory' in copula levels when Clayton copula parameter α is big. The time series plots (see Figure 1) of such Markov processes do look 'long memory alike'. Nevertheless, our next theorem shows that they are in fact geometric ergodic hence 'short memory' processes.

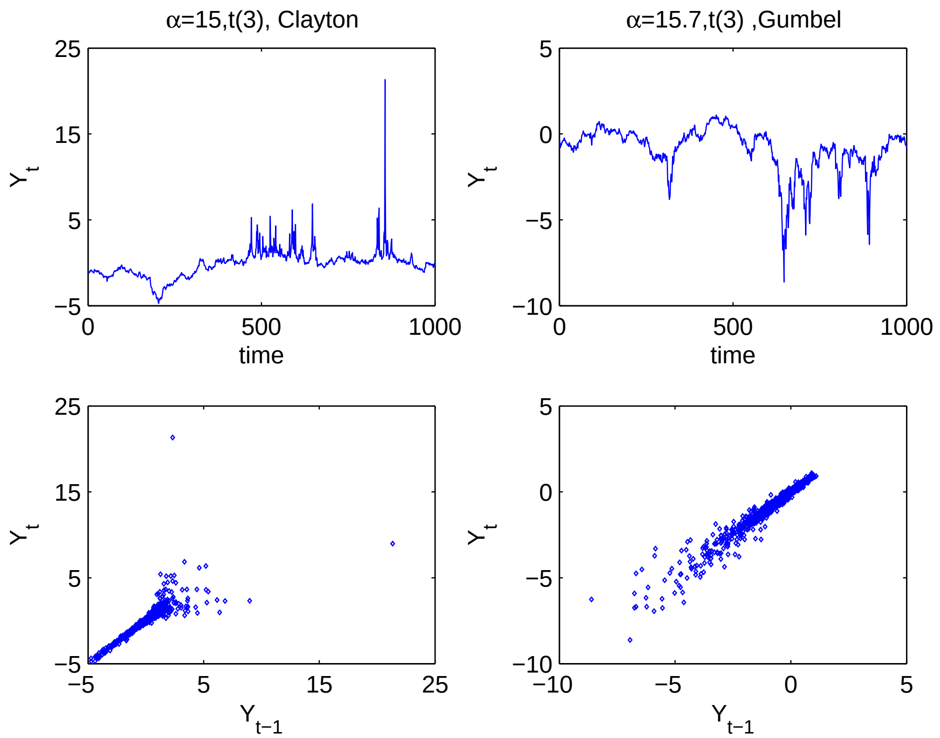


Figure 1: Markov time series: tail dependence index = 0.9548, student t_3 marginal distribution

Theorem 2.1 (geometric ergodicity): Under Assumption M, the Markov time series $\{Y_t\}_{t=1}^n$ generated via Clayton copula with $0 \leq \alpha < \infty$, Gumbel copula with $1 \leq \alpha < \infty$, Student's t copula with $|\rho| < 1$ and $\nu \in [2, \infty]$, are all geometric ergodic (hence geometric beta-mixing).

Remark 2.2: If $\{U_t\}_{t=1}^n$ is a $C_U(\cdot, \cdot)$ copula generated strictly stationary first order Markov model with uniform marginals, then $\{V_t \equiv 1 - U_t\}_{t=1}^n$ is also a copula based strictly stationary first order Markov model with uniform marginals and copula function:

$$
\begin{aligned}
C_V(v_1, v_2) &\equiv \Pr\left(V_{t-1} \leq v_1, V_t \leq v_2\right) = \Pr\left(U_{t-1} \geq 1 - v_1, U_t \geq 1 - v_2\right) \\
&= v_1 + v_2 - 1 + C_U(1 - v_1, 1 - v_2) \equiv C_U^s(v_1, v_2)
\end{aligned}
$$

which is the survival copula of $C_U^s(u_1, u_2)$ (see Nelson, 2006). Therefore, a copula $C_U(\cdot, \cdot)$ generated strictly stationary first order Markov process is (geometric ergodic) or beta-mixing with certain decay speed $\beta(j) = o(1)$ if and only if its survival copula $C_U^s(\cdot, \cdot)$ generated Markov process is (geometric ergodic) or beta-mixing with the same decay speed $\beta(j) = o(1)$.

By Theorem 2.1 and Remark 2.2, we immediately have that survival Clayton and survival Gumbel generated first order stationary Markov processes are also geometric ergodic.

3 Sieve MLE, Consistency with Rate

Under assumption M, we have that the true conditional density of Y_t given $Y^{t-1} \equiv (Y_{t-1}, ..., Y_1)$ is given by: $p^0(\cdot|Y^{t-1}) = h_0(\cdot|Y_{t-1}) \equiv g_0(\cdot)c(G_0(Y_{t-1}), G_0(\cdot); \alpha_0)$. Let

$$p(\cdot|Y^{t-1}) = h(\cdot|Y_{t-1}; \alpha, g) \equiv g(\cdot)c(G(Y_{t-1}), G(\cdot); \alpha)$$

denote any candidate conditional density of Y_t given Y^{t-1}. Let $Z_t = (Y_{t-1}, Y_t)$, and denote

$$\begin{aligned} \ell(\alpha, g, Z_t) &\equiv \log p(Y_t|Y^{t-1}) = \log \{h(Y_t|Y_{t-1}; \alpha, g)\} \equiv \log g(Y_t) + \log c\,(G(Y_{t-1}), G(Y_t); \alpha) \\ &\equiv \log g(Y_t) + \log c\left(\int 1(y \leq Y_{t-1})g(y)dy, \int 1(y \leq Y_t)g(y)dy; \alpha\right) \end{aligned}$$

as the log-likelihood associated with the conditional density $p(Y_t|Y^{t-1})$. Then the joint log-likelihood function of the data $\{Y_t\}_{t=1}^n$ is given by

$$L_n(\alpha, g) \equiv \frac{1}{n}\sum_{t=2}^{n} \ell(\alpha, g, Z_t) + \frac{1}{n}\log g(Y_1).$$

The approximate sieve MLE $\widehat{\gamma}_n \equiv (\widehat{\alpha}_n, \widehat{g}_n)$ is defined as

$$L_n(\widehat{\alpha}_n, \widehat{g}_n) \geq \max_{\alpha \in \mathcal{A}, g \in \mathcal{G}_n} L_n(\alpha, g) - O_p\left(\delta_n^2\right), \tag{3.1}$$

where $\delta_n = o(1)$, and $\mathcal{G}_n$ denotes the sieve space (i.e., a sequence of finite dimensional parameter spaces that becomes dense (as $n \to \infty$) in the entire parameter space $\mathcal{G}$ for g_0).

There exist many sieves for approximating a univariate probability density function. In this paper, we will focus on using linear sieves to directly approximate either a square root density:

$$\mathcal{G}_n = \left\{ g_{K_n} \in \mathcal{G} : g_{K_n}(y) = [\sum_{k=1}^{K_n} a_k A_k(y)]^2, \quad \int g_{K_n}(y)dy = 1 \right\}, \quad K_n \to \infty, \frac{K_n}{n} \to 0, \tag{3.2}$$

or a log density:

$$\mathcal{G}_n = \left\{ g_{K_n} \in \mathcal{G} : g_{K_n}(y) = \exp\{\sum_{k=1}^{K_n} a_k A_k(y)\}, \quad \int g_{K_n}(y)dy = 1 \right\}, \quad K_n \to \infty, \frac{K_n}{n} \to 0, \tag{3.3}$$

where $\{A_k(\cdot) : k \geq 1\}$ consists of known basis functions, and $\{a_k : k \geq 1\}$ is the collection of unknown sieve coefficients.

Suppose the support $\mathcal{Y}$ (of the true g_0) is either a compact interval (say $[0, 1]$) or the whole real line $\mathcal{R}$. A real-valued function g on $\mathcal{Y}$ is said to be *r-smooth* if it is J times continuously differentiable on $\mathcal{Y}$ and its J-th derivative satisfies a Hölder condition with exponent $r - J \in (0, 1]$ (i.e., there is a positive number K such that $|D^J g(y) - D^J g(y')| \leq K|y - y'|^{r-J}$ for all $y, y' \in \mathcal{Y}$).

We denote $\Lambda^r(\mathcal{Y})$ as the class of all real-valued functions on $\mathcal{Y}$ which are r-*smooth*; it is called a Hölder space.

Let the true marginal density function g_0 satisfy either $\sqrt{g_0} \in \Lambda^r(\mathcal{Y})$ or $\log g_0 \in \Lambda^r(\mathcal{Y})$. Then any function in $\Lambda^r(\mathcal{Y})$ can be approximated by some appropriate sieve spaces. For example, if $\mathcal{Y}$ is a bounded interval and $r > 1/2$, it can be approximated by the spline sieve $Spl(s, K_n)$ with $s > [r]$, the polynomial sieve, the trigonometric sieve, the cosine series and etc. When the support or $\mathcal{Y}$ is unbounded, thin-tailed density can be approximated by Hermite polynomial sieve, while polynomial fat-tailed density can be approximated by spline wavelet sieve. See Chen (2007) for detailed descriptions of various sieve spaces $\mathcal{G}_n$. In our simulation study, we chose the sieve number of terms using modified AIC and BIC, although one could also use cross-validation (see, e.g., Fan and Yao (2003), Gao (2007), Li and Racine (2007)) and other computationally more intensive model selection methods (see, e.g., Shen et al. (2004)) to choose the sieve number of terms K_n. See Chen et al. (2006) for further discussions.

3.1 Consistency

In the following we denote $Q_n(\alpha, g) \equiv \frac{n-1}{n} E_0[\ell(\alpha, g, Z_2)] + \frac{1}{n} E_0[\log g(Y_1)]$, where E_0 is the expectation under the true DGP (i.e., Assumption M). Denote $\gamma \equiv (\alpha, g)$ and $\gamma_0 \equiv (\alpha_0, g_0) \in \Gamma \equiv \mathcal{A} \times \mathcal{G}$.
Assumption 3.1: *(1) $\alpha_0 \in \mathcal{A}$, where $\mathcal{A}$ is a compact set of $\mathcal{R}^d$ with nonempty interior; (2) $g_0 \in \mathcal{G}$, let $r > 1/2$, either $\mathcal{G} = \{g = f^2 : f \in \Lambda^r, \int g(y)dy = 1\}$ and $\mathcal{G}_n$ is given in (3.2), or $\mathcal{G} = \{g = \exp(f) : f \in \Lambda^r, \int g(y)dy = 1\}$ and $\mathcal{G}_n$* is given in (3.3) *; (3) $Q_n(\alpha_0, g_0) > -\infty$, there are a metric $||\gamma||_c \equiv \sqrt{\alpha'\alpha} + ||g||_c$ on $\Gamma \equiv \mathcal{A} \times \mathcal{G}$* and a positive measurable function $\eta()$ such that for all $\varepsilon > 0$ and for all $k \geq 1$,

$$Q_n(\alpha_0, g_0) - \sup_{\alpha \in \mathcal{A}, g \in \mathcal{G}_k : ||\gamma_0 - \gamma||_c \geq \varepsilon} Q_n(\alpha, g) \geq \eta(\varepsilon) > 0.$$

(4) *the sieve spaces $\mathcal{G}_n$* are compact under the metric $||g||_c$; *(5) there is $\Pi_n \gamma_0 \in \Gamma_n \equiv \mathcal{A} \times \mathcal{G}_n$ such that* $||\Pi_n \gamma_0 - \gamma_0||_c = o(1)$; and $|Q_n(\Pi_n \gamma_0) - Q_n(\gamma_0)| = o(1)$.

For the norm $||\gamma||_c \equiv \sqrt{\alpha'\alpha} + ||g||_c$ *on* $\Gamma \equiv \mathcal{A} \times \mathcal{G}$, one can use either sup norm $||g||_\infty$ (or a weighted sup norm) or even lower order Hölder norm $||g||_{\Lambda^{r'}}$ for $r' \in [0, r)$ (or its weighted version).
Assumption 3.2: (1) $E_0\left[\sup_{\gamma \in \Gamma_n} |\ell(\gamma, Z_t)|\right]$ is bounded; (2) there are a finite constant $\kappa > 0$ and a measurable function $M(\cdot)$ with $E_0[M(Z_t)] \leq const. < \infty$, such that for all $\delta > 0$,

$$\sup_{\{\gamma, \gamma_1 \in \Gamma_n : ||\gamma - \gamma_1||_c \leq \delta\}} |\ell(\gamma, Z_t) - \ell(\gamma_1, Z_t)| \leq \delta^\kappa M(Z_t) \quad a.s. - Z_t$$

We note that under assumption 3.1(1)(4), assumption 3.2(1) is implied by assumption 3.2(2).

Proposition 3.1: *Under Assumptions M, 3.1 - 3.2,* $\delta_n = o(1)$ and $\frac{K_n}{n} \to 0$, *we have:*

$$||\widehat{\gamma}_n - \gamma_0||_c = o_p(1).$$

3.2 Convergence rate

Denote $\mathcal{N} = \{\gamma \in \Gamma : ||\gamma_0 - \gamma||_c = o(1)\}$ and $\mathcal{N}_n = \{\gamma \in \Gamma_n : ||\gamma_0 - \gamma||_c = o(1)\}$. Denote Var_0 as the variance under the true DGP (i.e., Assumption M).

Assumption 3.3: *(1) there are a metric* $||\gamma||_s \equiv \sqrt{\alpha'\alpha} + ||g||_s$ *on* $\mathcal{N}$ such that $||\gamma||_s \leq ||\gamma||_c$, and a *constant* $J_0 > 0$ such that for all $\varepsilon > 0$ and for all $n \geq 1$,

$$Q_n(\alpha_0, g_0) - \sup_{\gamma \in \mathcal{N}_n : ||\gamma_0 - \gamma||_s \geq \varepsilon} Q_n(\alpha, g) \geq J_0 \varepsilon^2 > 0.$$

(2) $\sup_{\{\gamma \in \mathcal{N}_n : ||\gamma_0 - \gamma||_s \leq \epsilon\}} Var_0(\ell(\gamma, Z_t) - \ell(\gamma_0, Z_t)) \leq const. \times \epsilon^2$ for all small $\epsilon > 0$.

Assumption 3.3 suggests that a natural choice of $||\gamma||_s$ could be $\sqrt{Q_n(\gamma_0) - Q_n(\gamma)}$.

Assumption 3.4: (1) $\{Y_t\}_{t=1}^n$ is geometric ergodic (hence geometric beta mixing); (2) there are a constant $\kappa \in (0, 2)$ and a measurable function $M(\cdot)$ with $E_0[M(Z_t)^2 \log(1 + M(Z_t))] \leq const. < \infty$, such that for any $\delta > 0$,

$$\sup_{\{\gamma \in \mathcal{N}_n : ||\gamma_0 - \gamma||_s \leq \delta\}} |\ell(\gamma, Z_t) - \ell(\gamma_0, Z_t)| \leq \delta^\kappa M(Z_t) \quad a.s. - Z_t.$$

Although we do not need any beta-mixing decay rate to establish consistency in Proposition 3.1, we need some beta-mixing decay rate for rate of convergence.[3] Given the results in subsection 2.2, Assumption 3.4(1) is typically satisfied by copula-based Markov models. Note that in assumption 3.4(2), the moment restriction on the envelop function $M(Z_t)$ is weaker than the one ($E_0[M(Z_t)^\varsigma] \leq$ *const.* $< \infty$ for some $\zeta > 2$) imposed in Chen and Shen (1998). This is because Chen and Shen (1998) only assumed beta mixing with polynomial decay speed while our assumption 3.4(1) assumes geometric beta mixing. It is well known that there are trade-off between speed of mixing decay rate and finiteness of moments. See Doukhan, et al (1995).

The next proposition is a direct application of theorem 1 of Chen and Shen (1998) hence we omit its proof.

Proposition 3.2: *Under Assumptions M, 3.1 - 3.4, we have*

$$||\widehat{\gamma}_n - \gamma_0||_s = O_p(\delta_n), \quad \delta_n = \max\left\{\sqrt{\frac{K_n}{n}}, ||\gamma_0 - \Pi_n \gamma_0||_s\right\} = o(1).$$

[3] It is common to assume some beta mixing or strong mixing decay rates in semi/nonparametric estimation and testing; see, e.g., Robinson (1983), Fan and Yao (2003), Gao (2007), Li and Racine (2007), Kosorok (2008).

4 Normality and Efficiency of Sieve MLE of Smooth Functionals

Let $\rho : \mathcal{A} \times \mathcal{G} \to \mathcal{R}$ be a smooth functional and $\rho(\widehat{\gamma}_n)$ be the plug-in sieve MLE of $\rho(\gamma_0)$. In this section, we extend the results of Chen et al (2006) on root-n normality and efficiency of their sieve MLE for copula based multivariate joint distribution model using i.i.d. data to our scalar strictly stationary first order Markov setting.

4.1 Asymptotic Normality and Efficiency of $\rho(\widehat{\gamma}_n)$

In the following we denote $(U_1, U_2) = (G_0(Y_1), G_0(Y_2))$, $u = (u_1, u_2) \in [0,1]^2$ and $c(G_0(Y_{t-1}), G_0(Y_t); \alpha_0) = c(U; \alpha_0) = c(\gamma_0, Z_t)$ (with the danger of slightly abusing notations). We also denote $\mathcal{N}_0 = \{\gamma \in \mathcal{N} : ||\gamma_0 - \gamma||_s = O(\delta_n)\}$ and $\mathcal{N}_{0n} = \{\gamma \in \mathcal{N}_n : ||\gamma_0 - \gamma||_s = O(\delta_n)\}$.

Assumption 4.1: $\alpha_0 \in int(\mathcal{A})$.

Assumption 4.2: the second order partial derivatives $\frac{\partial^2 \log c(u;\alpha)}{\partial\alpha\alpha'}$, $\frac{\partial^2 \log c(u;\alpha)}{\partial u_j \partial\alpha}$, $\frac{\partial^2 \log c(u;\alpha)}{\partial u_j \partial u_k}$ for $k, j = 1, 2$, are all well-defined and continuous in $\gamma \in \mathcal{N}_0$.

Denote $\mathbf{V}$ as the linear span of $\Gamma - \gamma_0$. Under Assumption 4.2, for any $v = (v_\alpha, v_g)' \in \mathbf{V}$, we have that $\ell(\gamma_0 + sv, Z)$ is continuously differentiable in small $s \in [0,1]$. For any $\gamma \in \mathcal{N}_0$, define the first order directional derivative of $\ell(\gamma, Z_t)$ at the direction $v \in \mathbf{V}$ as:

$$
\begin{aligned}
& \frac{d\ell(\gamma + sv, Z_t)}{ds}|_{s=0} \equiv \frac{\partial \ell(\gamma, Z_t)}{\partial \gamma'}[v] \\
= \quad & \frac{\partial \log c(\gamma, Z_t)}{\partial \alpha'}[v_\alpha] + \frac{v_g(Y_t)}{g(Y_t)} + \sum_{j=1}^{2} \frac{\partial \log c(\gamma, Z_t)}{\partial u_j} \int 1\{y \le Y_{t-2+j}\} v_g(y) dy,
\end{aligned}
$$

and the second order directional derivative as:

$$
\frac{d^2 \ell(\gamma + sv + \widetilde{s}\widetilde{v}, Z_t)}{d\widetilde{s} ds}|_{s=0}|_{\widetilde{s}=0} = \frac{d}{d\widetilde{s}} \left\{ \frac{\partial \ell(\gamma + \widetilde{s}\widetilde{v}, Z_t)}{\partial \gamma'}[v] \right\} |_{\widetilde{s}=0} \equiv \frac{\partial^2 \ell(\gamma, Z_t)}{\partial \gamma \partial \gamma'}[v, \widetilde{v}].
$$

Assumption 4.3: (1) $0 < E_0 \left[\left(\frac{\partial \ell(\gamma_0, Z_t)}{\partial \gamma'}[v] \right)^2 \right] < \infty$ for $v \neq 0, v \in \mathbf{V}$;

(2) $h(Y_t|Y_{t-1}; \alpha, g) \equiv g(Y_t) c(G(Y_{t-1}), G(Y_t); \alpha) > 0$ in the neighborhood $\mathcal{N}_0$ of γ_0;

(3) Let $\mathcal{S}_v = \{s \in [0,1] : \gamma_0 + sv \in \mathcal{N}_0\}$. $\int \sup_{s \in \mathcal{S}_v} |\frac{dh(y|Y_{t-1};\gamma_0+sv)}{ds}| dy < \infty$ and $\int \sup_{s \in \mathcal{S}_v} |\frac{d^2 h(y|Y_{t-1};\gamma_0+sv)}{ds^2}| dy < \infty$ almost surely, for $v \neq 0, v \in \mathbf{V}$.

Lemma 4.1: Under assumptions M, 4.1, 4.2 and 4.3, we have: (1) $E_0 \left(\left(\frac{\partial \ell(\gamma_0, Z_t)}{\partial \gamma'}[v] \right) \left(\frac{\partial \ell(\gamma_0, Z_s)}{\partial \gamma'}[\tilde{v}] \right) \right) = 0$ for $v, \tilde{v} \in \mathbf{V}$ and all $s < t$. (2) $\{\frac{\partial \ell(\gamma_0, Z_t)}{\partial \gamma'}[v] : v \in \mathbf{V}\}_{t=1}^n$ is a martingale difference sequence. (3) $E_0 \left(\left(\frac{\partial \ell(\gamma_0, Z_t)}{\partial \gamma'}[v] \right)^2 \right) = -E_0 \left(\frac{\partial^2 \ell(\gamma_0, Z_t)}{\partial \gamma \partial \gamma'}[v, v] \right)$.

Lemma 4.1 suggests that we can define the Fisher inner product on the space $\mathbf{V}$ as

$$
\langle v, \tilde{v} \rangle \equiv E_0 \left[\left(\frac{\partial \ell(\gamma_0, Z_t)}{\partial \gamma'}[v] \right) \left(\frac{\partial \ell(\gamma_0, Z_t)}{\partial \gamma'}[\tilde{v}] \right) \right]
$$

and the Fisher norm for $v \in \mathbf{V}$ as $\|v\|^2 \equiv \langle v, v\rangle$. Let $\overline{\mathbf{V}}$ be the closed linear span of $\mathbf{V}$ under the Fisher norm. Then $(\overline{\mathbf{V}}, \|\cdot\|)$ is a Hilbert space.

The asymptotic properties of $\rho(\widehat{\gamma}_n)$ depend on the smoothness of the functional ρ and the rate of convergence of $\widehat{\gamma}_n$. For any $v \in \overline{\mathbf{V}}$, we denote

$$\frac{d\rho(\gamma_0 + sv)}{ds}|_{s=0} \equiv \frac{\partial \rho(\gamma_0)}{\partial \gamma'}[v],$$

whenever the limit is well defined.

Assumption 4.4: (1) for any $v \in \overline{\mathbf{V}}$, $\rho(\gamma_0 + sv)$ is continuously differentiable in $s \in [0,1]$ near $s = 0$, and

$$\|\frac{\partial \rho(\gamma_0)}{\partial \gamma'}\| \equiv \sup_{v \in \overline{\mathbf{V}}:\|v\|>0} \frac{|\frac{\partial \rho(\gamma_0)}{\partial \gamma'}[v]|}{\|v\|} < \infty;$$

(2) there exist constants $c > 0$, $\omega > 0$, and a small $\epsilon > 0$ such that for any $v \in \overline{\mathbf{V}}$ with $\|v\| < \epsilon$, we have

$$|\rho(\gamma_0 + v) - \rho(\gamma_0) - \frac{\partial \rho(\gamma_0)}{\partial \gamma'}[v]| \le c\|v\|^{\omega}$$

Under this assumption, by the Riesz representation theorem, there exists a $v^* \in \overline{\mathbf{V}}$ such that

$$\frac{\partial \rho(\gamma_0)}{\partial \gamma'}[v] \equiv \langle v^*, v\rangle, \ \ for\ all\ v \in \overline{\mathbf{V}} \tag{4.1}$$

and

$$\|v^*\|^2 = \|\frac{\partial \rho(\gamma_0)}{\partial \gamma'}\|^2 = \sup_{v \in \overline{\mathbf{V}}:\|v\|>0} \frac{|\frac{\partial \rho(\gamma_0)}{\partial \gamma'}[v]|^2}{\|v\|^2} < \infty$$

Assumption 4.5: (1) $\|\widehat{\gamma}_n - \gamma_0\| = O_p(\delta_n)$ for a decreasing sequence δ_n satisfying $(\delta_n)^{\omega} = o(n^{-1/2})$; (2) there exists $\Pi_n v^* \in \Gamma_n - \{\gamma_0\}$ such that $\delta_n \times \|\Pi_n v^* - v^*\| = o(n^{-1/2})$.

Assumption 4.6: for all $\tilde{\gamma} \in \mathcal{N}_{0n}$ with $\|\tilde{\gamma} - \gamma_0\| \le \delta_n$ and all $v = (v_\alpha, v_g)' \in \overline{\mathbf{V}}$ with $\|v\| \le \delta_n$ we have:

$$E_0\left(\frac{\partial^2 \ell(\tilde{\gamma}, Z_t)}{\partial \gamma \partial \gamma'}[v,v] - \frac{\partial^2 \ell(\gamma_0, Z_t)}{\partial \gamma \partial \gamma'}[v,v]\right) = o(n^{-1}).$$

Assumption 4.7: $\left\{\frac{\partial \ell(\gamma, Z_t)}{\partial \gamma'}[\Pi_n v^*] : \gamma \in \mathcal{N}_0, \|\gamma - \gamma_0\| = O(\delta_n)\right\}$ is a Donsker class.

Under assumption 3.4(1), Assumption 4.7 is satisfied by applying the results of Doukhan, et al (1995) on Donsker theorems for strictly stationary beta mixing processes.

Theorem 4.1 (**Normality**): Suppose that Assumptions M, 3.1-3.4 and 4.1-4.7 hold. Then: $\sqrt{n}(\rho(\widehat{\gamma}_n) - \rho(\gamma_0)) \Rightarrow N(0, \|\frac{\partial \rho(\gamma_0)}{\partial \gamma'}\|^2)$.

We follow the approach of Wong (1992) to establish semiparametric efficiency. Related work can be found in Shen (1997), Bickel et al. (1993), Bickel and Kwon (2001) and the references therein. Recall that a probability family $\{P_\gamma : \gamma \in \Gamma\}$ for the sample $\{Y_t\}_{t=1}^n$ is *locally asymptotically normal*

(LAN) at γ_0, if (1) for any v in the linear span of $\Gamma - \gamma_0$, $\gamma_0 + sn^{-1/2}v \in \Gamma$ for all small $s \geq 0$, and (2)

$$\frac{dP_{\gamma_0+n^{-1/2}v}}{dP_{\gamma_0}}(Y_1, \cdots, Y_n) = \exp\left\{n[L_n(\gamma_0 + \frac{1}{\sqrt{n}}v) - L_n(\gamma_0)]\right\} = \exp\left\{\Sigma_n(v) - \frac{1}{2}\|v\|^2 + R_n(\gamma_0, v)\right\},$$

where $\Sigma_n(v)$ is linear in v, $\Sigma_n(v) \xrightarrow{d} \mathcal{N}(0, \|v\|^2)$ and $\text{plim}_{n\to\infty} R_n(\gamma_0, v) = 0$ (both limits are under the true probability measure P_{γ_0}). To avoid the "super-efficiency" phenomenon, certain regularity conditions on the estimates are required. In estimating a smooth functional in the infinite-dimensional case, Wong (1992, p.58) defines the class of *pathwise regular* estimates. An estimate $T_n(Y_1, \cdots, Y_n)$ of $\rho(\gamma_0)$ is *pathwise regular* if for any real number $s > 0$ and any v in the linear span of $\Gamma - \gamma_0$, we have

$$\limsup_{n\to\infty} P_{\gamma_{n,s}}(T_n < \rho(\gamma_{n,s})) \leq \liminf_{n\to\infty} P_{\gamma_{n,-s}}(T_n < \rho(\gamma_{n,-s})),$$

where $\gamma_{n,s} = \gamma_0 + sn^{-1/2}v$. See Wong (1992) and Shen (1997) for details.

Theorem 4.2 (**Efficiency**): Under conditions in Theorem 4.1, if LAN holds, then the plug in sieve MLE $\rho(\widehat{\gamma}_n)$ achieves the efficiency lower bound for pathwise regular estimates.

4.2 $\sqrt{n}$ Normality and Efficiency of $\widehat{\alpha}_n$

We take $\rho(\gamma) = \lambda'\alpha$ for any arbitrarily fixed $\lambda \in \mathcal{R}^d$ with $0 < |\lambda| < \infty$. It satisfies Assumption 4.4(2) with $\frac{\partial\rho(\gamma_0)}{\partial\gamma'}[v] = \lambda' v_\alpha$ and $\omega = \infty$. Assumption 4.4(1) is equivalent to finding a Riesz representer $v^* \in \overline{\mathbf{V}}$ satisfying (4.2) and (4.3):

$$\lambda'(\alpha - \alpha_0) = \langle \gamma - \gamma_0, v^* \rangle \quad \text{for any } \gamma - \gamma^* \in \overline{\mathbf{V}} \tag{4.2}$$

and

$$\|\frac{\partial\rho(\gamma_0)}{\partial\gamma'}\|^2 = ||v^*||^2 = \langle v^*, v^* \rangle = \sup_{v\neq 0, v\in\overline{\mathbf{V}}} \frac{|\lambda' v_\alpha|^2}{||v||^2} < \infty. \tag{4.3}$$

Let us change the variables before making statements on (4.3). Denote:

$$\mathcal{L}_2^0([0,1]) \equiv \left\{e : [0,1] \to \mathcal{R} : \int_0^1 e(v)dv = 0, \int_0^1 [e(v)]^2 dv < \infty\right\}$$

By change of variables, for any $v_g \in \overline{\mathbf{V}}_g$, there is a unique function $b_g \in \mathcal{L}_2^0([0,1])$ with $b_g(u) = \frac{v_g(G_0^{-1}(u))}{g_0(G_0^{-1}(u))}$, and vice versa. So we can express $\frac{\partial\ell(\gamma_0, Z_t)}{\partial\gamma'}[v]$ as:

$$\begin{aligned} &\frac{\partial\ell(\gamma_0, Z_t)}{\partial\gamma'}[v] = \frac{\partial\ell(\gamma_0, U_t, U_{t-1})}{\partial\gamma'}[(v_\alpha', b_g)'] \\ = \;& \frac{\partial \log c(U_{t-1}, U_t; \alpha_0)}{\partial\alpha'}[v_\alpha] + b_g(U_t) + \sum_{j=1}^{2} \frac{\partial \log c(U_{t-1}, U_t; \alpha_0)}{\partial u_j} \int_0^{U_{t-2+j}} b_g(u)du \end{aligned}$$

and

$$
\begin{aligned}
\|v\|^2 &= E_0\left[\left(\frac{\partial \ell(\gamma_0, U_t, U_{t-1})}{\partial \gamma'}[(v_\alpha', b_g)']\right)^2\right] \\
&= E_0\left[\left(\frac{\partial \log c(U_{t-1}, U_t; \alpha_0)}{\partial \alpha'}[v_\alpha] + b_g(U_t) + \sum_{j=1}^{2} \frac{\partial \log c(U_{t-1}, U_t; \alpha_0)}{\partial u_j} \int_0^{U_{t-2+j}} b_g(u)du\right)^2\right].
\end{aligned}
$$

Define:

$$
\overline{\mathbf{B}} = \left\{ b = (v_\alpha', b_g)' \in (\mathcal{A} - \alpha_0) \times \mathcal{L}_2^0([0,1]) : ||b||^2 \equiv E_0\left[\left(\frac{\partial \ell(\gamma_0, U_t, U_{t-1})}{\partial \gamma'}[b]\right)^2\right] < \infty \right\}.
$$

Then there is a one-to-one onto mapping between the two Hilbert spaces $(\overline{\mathbf{B}}, ||\cdot||)$ and $(\overline{\mathbf{V}}, ||\cdot||)$. So the Riesz representer $v^* = (v_\alpha^{*\prime}, v_g^*)' \in \overline{\mathbf{V}}$ is uniquely determined by $b^* = (v_\alpha^{*\prime}, b_g^*)' \in \overline{\mathbf{B}}$ (and vice versa) via the relation: $v_g^*(y) = b_g^*(G_0(y))g_0(y)$ for all $y \in \mathcal{Y}$. Notice that

$$
\begin{aligned}
& \sup_{v \neq 0, v \in \overline{\mathbf{V}}} \frac{|\lambda' v_\alpha|^2}{||v||^2} \\
= & \sup_{b \neq 0, b \in \overline{\mathbf{B}}} \frac{|\lambda' v_\alpha|^2}{E_0\left[\left(\frac{\partial \log c(U_{t-1}, U_t; \alpha_0)}{\partial \alpha'}[v_\alpha] + b_g(U_t) + \sum_{j=1}^{2} \frac{\partial \log c(U_{t-1}, U_t; \alpha_0)}{\partial u_j} \int_0^{U_{t-2+j}} b_g(u)du\right)^2\right]} \\
= & \lambda' \mathcal{I}_*(\alpha_0)^{-1} \lambda = \lambda' \left(E_0[\mathcal{S}_{\alpha_0} \mathcal{S}_{\alpha_0}']\right)^{-1} \lambda,
\end{aligned}
$$

where $\mathcal{S}_{\alpha_0}$ is the efficient score function for α_0,

$$
\mathcal{S}_{\alpha_0}' = \frac{\partial \log c(\alpha_0, U_t, U_{t-1})}{\partial \alpha'} - \mathbf{e}^*(U_t) - \sum_{j=1}^{2} \frac{\partial \log c(\alpha_0, U_t, U_{t-1})}{\partial u_j} \int_0^{U_{t-2+j}} \mathbf{e}^*(u)du \tag{4.4}
$$

and $\mathbf{e}^* = (e_1^*, \cdots, e_d^*) \in (\mathcal{L}_2^0([0,1]))^d$ solves the following infinite-dimensional optimization problems for $k = 1, \cdots, d$,

$$
\inf_{e_k \in \mathcal{L}_2^0([0,1])} E_0\left\{\left(\frac{\partial \log c(U_{t-1}, U_t; \alpha_0)}{\partial \alpha_k} - e_k(U_t) - \sum_{j=1}^{2} \frac{\partial \log c(U_{t-1}, U_t; \alpha_0)}{\partial u_j} \int_0^{U_{t-2+j}} e_k(u)du\right)^2\right\}.
$$

Therefore $b^* = (v_\alpha^{*\prime}, b_g^*)'$ with $v_\alpha^* = \mathcal{I}_*(\alpha_0)^{-1}\lambda$ and $b_g^*(u) = -e^*(u) \times v_\alpha^*$, and $v^* = [I_d, -e^*(G_0(\cdot))g_0(\cdot)] \times \mathcal{I}_*(\alpha_0)^{-1}\lambda$. Hence (4.3) is satisfied if and only if $\mathcal{I}_*(\alpha_0) = E_0[\mathcal{S}_{\alpha_0} \mathcal{S}_{\alpha_0}']$ is *non-singular,* which in turn is satisfied under the following Assumption:

Assumption 4.4': (1) $\int \frac{\partial c(u;\alpha_0)}{\partial u_j} du_{-j} = \frac{\partial}{\partial u_j} \int c(u; \alpha_0) du_{-j} = 0$ for $(j, -j) = (1, 2)$ with $j \neq -j$; (2) $\Sigma_{ideal} \equiv E_0\left(\frac{\partial \log c(U_{t-1}, U_t; \alpha_0)}{\partial \alpha} \{\frac{\partial \log c(U_{t-1}, U_t; \alpha_0)}{\partial \alpha}\}'\right)$ is finite and positive definite; (3) $\int \frac{\partial^2 c(u;\alpha_0)}{\partial u_j \partial \alpha} du_{-j} = \frac{\partial^2}{\partial u_j \partial \alpha} \int c(u; \alpha_0) du_{-j} = 0$ for $(j, -j) = (1, 2)$ with $j \neq -j$; (4) there exists a constant K such that $\max_{j=1,2} \sup_{0 < u_j < 1} E\left[\left(u_j(1-u_j)\frac{\partial \log c(U_1, U_2; \alpha_0)}{\partial u_j}\right)^2 |U_j = u_j\right] \leq K$.

We can now apply Theorems 4.1 and 4.2 to obtain the following result:

Proposition 4.1: Suppose that assumptions M, 3.1-3.4 and 4.1-4.3, 4.4', 4.5-4.7 hold. Then: $\sqrt{n}(\widehat{\alpha}_n - \alpha_0) \Rightarrow N\left(0, \mathcal{I}_*(\alpha_0)^{-1}\right)$, and $\widehat{\alpha}_n$ is semiparametrically efficient.

In general, there is no closed-form solution of $\mathcal{I}_*(\alpha_0)$. Nevertheless it can be consistently estimated by a sieve least square method using its characterization in (4.4). Let $\widehat{U}_t = \widehat{G}_n(Y_t)$ for $t = 1, \cdots, n$. Let $\mathbf{B}_n$ be some sieve space such as:

$$\mathbf{B}_n = \{e(u) = \sum_{k=1}^{K_{n\alpha}} a_k \sqrt{2}\cos(k\pi u),\ u \in [0,1], \sum_{k=1}^{K_{n\alpha}} a_k^2 < \infty\}, \tag{4.5}$$

where $K_{n\alpha} \to \infty, (K_{n\alpha})^d/n \to 0$. For $k = 1, \cdots, d$, we compute $\widehat{e}_k$ as the solution to

$$\min_{e_k \in \mathbf{B}_n} \frac{1}{n-1} \sum_{t=2}^{n} \left(\frac{\partial \log c(\widehat{U}_{t-1}, \widehat{U}_t; \widehat{\alpha})}{\partial \alpha_k} - e_k(\widehat{U}_t) - \sum_{j=1}^{2} \frac{\partial \log c(\widehat{U}_{t-1}, \widehat{U}_t; \widehat{\alpha})}{\partial u_j} \int_0^{\widehat{U}_{t-2+j}} e_k(u)du \right)^2.$$

Denote $\widehat{\mathbf{e}} = (\widehat{e}_1, \cdots, \widehat{e}_d)$ and

$$\widehat{\mathcal{I}_*} = \frac{1}{n-1} \sum_{t=2}^{n} \left\{ \begin{array}{c} \left(\frac{\partial \log c(\widehat{U}_{t-1}, \widehat{U}_t; \widehat{\alpha})}{\partial \alpha'} - \widehat{\mathbf{e}}(\widehat{U}_t) - \sum_{j=1}^{2} \frac{\partial \log c(\widehat{U}_{t-1}, \widehat{U}_t; \widehat{\alpha})}{\partial u_j} \int_0^{\widehat{U}_{t-2+j}} \widehat{\mathbf{e}}(u)du \right)' \times \\ \left(\frac{\partial \log c(\widehat{U}_{t-1}, \widehat{U}_t; \widehat{\alpha})}{\partial \alpha'} - \widehat{\mathbf{e}}(\widehat{U}_t) - \sum_{j=1}^{2} \frac{\partial \log c(\widehat{U}_{t-1}, \widehat{U}_t; \widehat{\alpha})}{\partial u_j} \int_0^{\widehat{U}_{t-2+j}} \widehat{\mathbf{e}}(u)du \right) \end{array} \right\}.$$

Following the proof of theorem 5.1 in Ai and Chen (2003) we immediately obtain:

Proposition 4.2: Under all the assumptions of Proposition 4.1, $\widehat{\mathcal{I}_*} = \mathcal{I}_*(\alpha_0) + o_p(1)$.

4.3 Sieve MLE of the marginal distribution

Let us consider the estimation of $\rho(\gamma_0) = G_0(y)$ for some fixed $y \in \mathcal{Y}$ by the plug-in sieve MLE: $\rho(\widehat{\gamma}_n) = \widehat{G}_n(y) = \int 1(x \le y)\widehat{g}_n(x)dx$, where $\widehat{g}_n$ is the sieve MLE for g_0.

Clearly $\frac{\partial \rho(\gamma_0)}{\partial \gamma'}[v] = \int_{\mathcal{Y}} 1(x \le y) v_g(x)dx$ for any $v = (v_\alpha', v_g)' \in \overline{\mathbf{V}}$. It is easy to see that $\omega = \infty$ in Assumption 4.4, and

$$\|\frac{\partial \rho(\gamma_0)}{\partial \gamma'}\|^2 = \sup_{v \in \overline{\mathbf{V}}: ||v||>0} \frac{\left|\int_{\mathcal{Y}} 1(x \le y) v_g(x)dx\right|^2}{||v||^2} < \infty.$$

Hence the representer $v^* \in \overline{\mathbf{V}}$ should satisfy (4.6) and (4.7):

$$\langle v^*, v \rangle = \frac{\partial \rho(\gamma_0)}{\partial \gamma'}[v] = E_0\left(1(Y_t \le y)\frac{v_g(Y_t)}{g_0(Y_t)}\right) \quad \text{for all } v \in \overline{\mathbf{V}} \tag{4.6}$$

$$\|\frac{\partial \rho(\gamma_0)}{\partial \gamma'}\|^2 = ||v^*||^2 = ||b^*||^2 = \sup_{b \in \overline{\mathbf{B}}: ||b||>0} \frac{|E_0\left[1(U_t \le G_0(y))b_g(U_t)\right]|^2}{||b||^2}. \tag{4.7}$$

Proposition 4.3: Let $v^* \in \overline{\mathbf{V}}$ solve (4.6) and (4.7). Suppose that assumptions M, 3.1-3.4 and 4.1-4.3, 4.5-4.7 hold. Then for any fixed $y \in \mathcal{Y}$, $\sqrt{n}(\widehat{G}_n(y) - G_0(y)) \Rightarrow N\left(0, ||v^*||^2\right)$. Moreover, $\widehat{G}_n$ is semiparametrically efficient.

Again, there are currently no closed-form expressions for the asymptotic variance $||v^*||^2$. Nevertheless, it can also be consistently estimated by the sieve method. Let $\widehat{\sigma}_G^2 \equiv$

$$\max_{v_\alpha \neq 0, b_g \in \mathbf{B}_n} \frac{\left|\frac{1}{n}\sum_{t=1}^n 1\{\widehat{U}_t \leq \widehat{G}_n(y)\} b_g(\widehat{U}_t)\right|^2}{\frac{1}{n-1}\sum_{t=2}^n \left[\frac{\partial \log c(\widehat{U}_{t-1}, \widehat{U}_t; \widehat{\alpha})}{\partial \alpha'} v_\alpha + b_g(\widehat{U}_t) + \sum_{j=1}^2 \frac{\partial \log c(\widehat{U}_{t-1}, \widehat{U}_t; \widehat{\alpha})}{\partial u_j} \int_0^{\widehat{U}_{t-2+j}} b_g(u) du\right]^2}$$

where $\widehat{U}_t = \widehat{G}_n(Y_t)$, and $\mathbf{B}_n$ is given in (4.5).

Proposition 4.4: Under all the assumptions of Proposition 4.3, we have: for any fixed $y \in \mathcal{Y}$, $\widehat{\sigma}_G^2 = ||v^*||^2 + o_p(1)$.

4.4 Plug-in estimates of conditional quantiles

Under assumption M, the $q-$th conditional quantile of Y_t given $Y_{t-1} = y$ is given by $Q_q^Y(y) = G_0^{-1}\left(C_{2|1}^{-1}[q|G_0(y); \alpha_0]\right)$. Its plug-in sieve MLE estimate is given by:

$$\hat{Q}_q^Y(y) = \hat{G}_n^{-1}\left(C_{2|1}^{-1}\left[q|\hat{G}_n(y); \hat{\alpha}_n\right]\right)$$

Let $\rho(\gamma_0) = Q_q^Y(y)$, then by some calculation,

$$\frac{\partial \rho(\gamma_0)}{\partial \gamma'}[v] = \frac{\frac{-C_{11}\int 1(x \leq y) v_g(x) dx - C_{1\alpha} v_\alpha}{c(U_{t-1}, C_1^{-1}(U_{t-1}, q; \alpha_0), \alpha_0)} - \int 1(x \leq Q_q^Y(y)) v_g(x) dx}{g_0(Q_q^Y(y))}$$

for any $v = (v_\alpha, v_g)' \in \overline{\mathbf{V}}$, $C_{11} = \frac{\partial^2 C(U_{t-1}, C_1^{-1}(U_{t-1}, q; \alpha_0), \alpha_0)}{\partial u_1^2}$, $C_{1\alpha} = \frac{\partial^2 C(U_{t-1}, C_1^{-1}(U_{t-1}, q; \alpha_0), \alpha_0)}{\partial u_1 \partial \alpha}$.

We can see $\omega = 2$ in Assumption 4.4, as long as $g_0(Q_q^Y(y)) \neq 0$ and $c(U_{t-1}, C_1^{-1}(U_{t-1}, q; \alpha_0), \alpha_0) \neq 0$, which are satisfied under assumption M (2). Thus we have:

$$||\frac{\partial \rho(\gamma_0)}{\partial \gamma'}||^2 = \sup_{v \in \overline{\mathbf{V}}: ||v|| > 0} \frac{\left|\{g_0(Q_q^Y(y))\}^{-1}\left[\frac{-C_{11}\int 1(x \leq y) v_g(x) dx - C_{1\alpha} v_\alpha}{c(U_{t-1}, C_1^{-1}(U_{t-1}, q; \alpha_0), \alpha_0)} - \int 1(x \leq Q_q^Y(y)) v_g(x) dx\right]\right|^2}{||v||^2} < \infty.$$

Hence the Riesz representer $v^* \in \overline{\mathbf{V}}$ should satisfy: $\langle v^*, v\rangle = \frac{\partial \rho(\gamma_0)}{\partial \gamma'}[v]$ for all $v \in \overline{\mathbf{V}}$, and $||v^*||^2 = ||\frac{\partial \rho(\gamma_0)}{\partial \gamma'}||^2$. Applying Theorems 4.1 and 4.2 we immediately obtain:

Proposition 4.5: Let $v^* \in \overline{\mathbf{V}}$ be the Riesz representer for $Q_q^Y(y)$. Suppose that assumptions M, 3.1-3.4, 4.1-4.3, 4.5-4.7 hold. Then: for a fixed $y \in \mathcal{Y}$, $\sqrt{n}(\hat{Q}_q^Y(y) - Q_q^Y(y)) \Rightarrow N\left(0, ||v^*||^2\right)$. Moreover, $\hat{Q}_q^Y(y)$ is semiparametrically efficient.

5 Sieve Likelihood Ratio Inference for Smooth Functionals

In this section, we are interested in sieve likelihood ratio inference for smooth functional $\rho(\gamma) = (\rho_1(\gamma), \cdots, \rho_k(\gamma))' : \Gamma \to \mathcal{R}^k$:

$$H_0 : \rho(\gamma_0) = 0,$$

where ρ is a vector of known functionals. For instance, $\rho(\gamma) = \alpha - \alpha_0 \in \mathcal{R}^d$ or $\rho(\gamma) = G(y) - G_0(y) \in \mathcal{R}$ for fixed y.

Shen and Shi (2005) provide a theory on sieve likelihood ratio inference for i.i.d. data. We now extend their result to strictly stationary Markov time series data,[4] and derive the following

$$2n\left(\max_{\alpha\in\mathcal{A},g\in\mathcal{G}_n} L_n(\alpha,g) - \max_{\alpha\in\mathcal{A},g\in\mathcal{G}_n:\rho(\gamma)=0} L_n(\alpha,g)\right) \to^d \mathcal{X}^2_{(m)}, \tag{5.1}$$

where m is the maximum number of linearly independent $\frac{\partial\rho_1(\gamma_0)}{\partial\gamma'}, \cdots, \frac{\partial\rho_k(\gamma_0)}{\partial\gamma'}$. Without loss of generality, we assume $k = m$, (i.e., $\frac{\partial\rho_1(\gamma_0)}{\partial\gamma'}, \cdots, \frac{\partial\rho_k(\gamma_0)}{\partial\gamma'}$ are linearly independent), otherwise a linear transformation can be conducted for the hypothesis.

Suppose that ρ_i satisfies Assumption 4.4 for $i = 1, \cdots, k$. Then by the Riesz representation theorem, there exists a $v_i^* \in \overline{\mathbf{V}}$ such that

$$\frac{\partial\rho_i(\gamma_0)}{\partial\gamma'}[v] \equiv \langle v_i^*, v\rangle, \text{ for all } v \in \overline{\mathbf{V}}.$$

Denote $v^* = (v_1^*, \cdots, v_k^*)'$. By the Gram-Schmidt orthogonalization, without loss of generality, we assume $\langle v_i^*, v_j^*\rangle = 0$ for any $i \neq j$. In addition, assume that v_i^* satisfies Assumption 4.5(2) for $i = 1, \cdots, k$.

Assumption 5.1: *For some positive sequence* $\{\delta_n, n \geq 1\}$, $\delta_n \to 0$, $\liminf_{n\to\infty} n^{1/2}\delta_n > 0$,

$$\lim\sup_{K\to\infty} \lim\sup_{n\to\infty} \Pr\left(\sup_{\{\gamma\in\Gamma_n, ||\gamma-\Pi_n\gamma_0||\geq K\delta_n\}} |L_n(\gamma) - L_n(\Pi_n\gamma_0)| \geq 0\right) = 0$$

In addition, $||\gamma_0 - \Pi_n\gamma_0|| = O(\delta_n)$.

This assumption is the same as that in Shen and Shi (2005). In this section, we require δ_n in Assumption 5.1 is the same as those in Assumptions 4.5, 4.6, and 4.7.

Denote

$$\widehat{\gamma}_n = \arg\max_{\alpha\in\mathcal{A},g\in\mathcal{G}_n} L_n(\alpha,g); \quad \overline{\gamma}_n = \arg\max_{\alpha\in\mathcal{A},g\in\mathcal{G}_n,\rho(\gamma)=0} L_n(\alpha,g).$$

Theorem 5.1: Suppose that assumptions M, 3.1-3.4, 4.1-4.3, 4.5-4.7 and 5.1 hold, also that assumption 4.4 holds with ρ_i for $i = 1, \cdots, k$ and v_i^* satisfies assumption 4.5(2) for $i = 1, \cdots, k$. Then:

$$2n(L_n(\widehat{\gamma}_n) - L_n(\overline{\gamma}_n)) \to^d \mathcal{X}^2_{(k)},$$

[4] If we only care about estimation of copula dependence parameter, we could extends the results of Murphy and van der Vaart (2000) on profile likelihood to our copula based semiparametric Markov models.

where $\frac{\partial \rho_1(\gamma_0)}{\partial \gamma'}, \cdots, \frac{\partial \rho_k(\gamma_0)}{\partial \gamma'}$ are assumed to be linearly independent.

We can apply Theorem 5.1 to construct confidence regions of any smooth functionals. For example, we can compute confidence region for sieve MLE of copula parameter α. Define the following profiled sieve likelihood:

$$\tilde{L}_n(\alpha) \equiv L_n(\alpha, \tilde{g}_n(\alpha)), \quad \text{with } \tilde{g}_n(\alpha) = \arg\max_{g \in \mathcal{G}_n} L_n(\alpha, g).$$

By Theorem 5.1, $2n(L_n(\widehat{\alpha}_n, \tilde{g}_n(\widehat{\alpha}_n)) - L_n(\alpha_0, \tilde{g}_n(\alpha_0))) \to^d \mathcal{X}^2_{(d)}$, where $(\widehat{\alpha}_n, \tilde{g}_n(\widehat{\alpha}_n)) = \widehat{\gamma}_n$ is the original sieve MLE.

6 Monte Carlo Comparison of Several Estimators

In this section we address the finite sample performance of sieve MLE by comparing it to several existing popular estimators: the two-step semiparametric estimator proposed in Chen and Fan (2006), the ideal (or infeasible) MLE, the correctly specified parametric MLE and the misspecified parametric MLE.

6.1 Existing Estimators

For comparison, we review several existing estimators that have been used in applied work.

6.1.1 Two-step semiparametric estimator

Chen and Fan (2006) propose the following two-step semiparametric procedure:

Step 1, estimate the unknown true marginal distribution $G_0(y)$ by the empirical distribution function: $\frac{n+1}{n} G_n(y)$, where $G_n(y) \equiv \frac{1}{n+1} \sum_{t=1}^n 1\{Y_t \leq y\}$.

Step 2, estimate the copula dependence parameter α_0 by:

$$\hat{\alpha}_n^{2sp} \equiv \arg\max_{\alpha \in \mathcal{A}} \frac{1}{n} \sum_{t=2}^n \log c(G_n(Y_{t-1}), G_n(Y_t); \alpha).$$

Assuming that the process $\{Y_t\}_{t=1}^n$ is beta-mixing with certain decay rate, under Assumption M and some other mild regularity conditions, Chen and Fan (2006) show that

$$\sqrt{n}(\hat{\alpha}_n^{2sp} - \alpha_0) \to_d N\left(0, \sigma_{2sp}^2\right), \quad \text{with } \sigma_{2sp}^2 \equiv B_0^{-1} \Sigma_{2sp} B_0^{-1}$$

where $B_0 \equiv -E_0\left(\frac{\partial^2 \log c(U_{t-1}, U_t; \alpha_0)}{\partial \alpha \partial \alpha'}\right) = \Sigma_{ideal}$ (under assumption 4.4'), and

$$\begin{aligned}
\Sigma_{2sp} &\equiv \lim_{n \to \infty} Var_0 \left\{ \frac{1}{\sqrt{n}} \sum_{t=2}^n \left[\frac{\partial \log c(U_{t-1}, U_t; \alpha_0)}{\partial \alpha} + W_1(U_{t-1}) + W_2(U_t) \right] \right\} < \infty, \\
W_1(U_{t-1}) &\equiv \int_0^1 \int_0^1 [1\{U_{t-1} \leq v_1\} - v_1] \frac{\partial^2 \log c(v_1, v_2; \alpha_0)}{\partial \alpha \partial u_1} c(v_1, v_2; \alpha_0) dv_1 dv_2, \\
W_2(U_t) &\equiv \int_0^1 \int_0^1 [1\{U_t \leq v_2\} - v_2] \frac{\partial^2 \log c(v_1, v_2; \alpha_0)}{\partial \alpha \partial u_2} c(v_1, v_2; \alpha_0) dv_1 dv_2.
\end{aligned}$$

Example 6.1 (Two-step semiparametric estimator of Gaussian copula parameter): The bivariate Gaussian copula is

$$C(u_1, u_2; \alpha) = \Phi_\alpha(\Phi^{-1}(u_1), \Phi^{-1}(u_2)), \quad |\alpha| < 1,$$

where Φ_α is the bivariate standard normal distribution with correlation α, and Φ is the scalar standard normal distribution. Chen and Fan (2006) show that:

$$\sqrt{n}(\hat{\alpha}_n^{2sp} - \alpha_0) \to_d N\left(0, 1 - \alpha_0^2\right).$$

Klaassen and Wellner (1997) establish that the semiparametric efficient variance bound for estimating a Gaussian copula parameter α is $1 - \alpha_0^2$; hence $\hat{\alpha}_n^{2sp}$ is semiparametrically efficient for Gaussian copula. However, as pointed out by Genest and Werker (2002), Gaussian copula and the independence copula are the only two copulas for which the two-step semiparametric estimator is efficient for α_0. Moreover, the empirical cdf estimator is still inefficient for $G_0()$ even in this Gaussian copula-based Markov model.

6.1.2 Possibly misspecified parametric MLE

Parametric MLE is the estimator that assuming a parametric functional form for the marginal probability density. Denote $G(y, \theta)$ ($g(y, \theta)$) as the marginal distribution (marginal density) whose functional form is known up to the unknown finite dimensional parameter θ. Then the observed joint parametric log-likelihood is:

$$L_n(\alpha, \theta) = \frac{1}{n}\sum_{t=1}^{n} \log g(Y_t, \theta) + \frac{1}{n}\sum_{t=2}^{n} \log c\left(G(Y_{t-1}, \theta), G(Y_t, \theta); \alpha\right),$$

and the parametric MLE is: $(\hat{\alpha}_n^p, \hat{\theta}_n^p) = \arg\max_{(\alpha,\theta)\in\mathcal{A}\times\Theta} L_n(\alpha, \theta)$, where $\mathcal{A} \times \Theta$ is the parameter space.

Denote $\ell(\alpha, \theta, Z_t) \equiv \log g(Y_t, \theta) + \log c\left(G(Y_{t-1}, \theta), G(Y_t, \theta); \alpha\right)$ as the parametric log-likelihood for one data point $Z_t \equiv (Y_{t-1}, Y_t)$.

Assumption 6.1 (1) $\mathcal{A}\times\Theta$ is a compact set of $\mathcal{R}^p$ with nonempty interior. $(\alpha^*, \theta^*) \in \mathcal{A}\times\Theta$ is the unique maximizer of $E_0(\ell(\alpha, \theta, Z_t))$ over $\mathcal{A}\times\Theta$; (2) $\ell(\alpha, \theta, Z_t)$ is continuous in (α, θ) for any data Z_t, and is a measurable function of Z_t for all $(\alpha, \theta) \in \mathcal{A}\times\Theta$; (3) $E_0[\sup_{(\alpha,\theta)\in\mathcal{A}\times\Theta} |\ell(\alpha, \theta, Z_t)|] < \infty$.

Assumption 6.2 (1) $(\alpha^*, \theta^*) \in int(\mathcal{A}\times\Theta)$; (2) the second order partial derivatives $\frac{\partial^2 \log g(y,\theta)}{\partial\theta\partial\theta'}$, $\frac{\partial^2 \log c(u_1,u_2,\alpha)}{\partial\alpha\alpha'}$, $\frac{\partial^2 \log c(u_1,u_2,\alpha)}{\partial u_j \partial\alpha}$, $\frac{\partial^2 \log c(u_1,u_2,\alpha)}{\partial u_j \partial u_k}$ for $k, j = 1, 2$ are all well-defined and continuous in a neighborhood $\mathcal{N}$ of (α^*, θ^*), and for all $y \in \mathcal{Y}$, $(u_1, u_2) \in (0,1)^2$; (3) $E_0\left(\sup_{(\alpha,\theta)\in\mathcal{N}} ||\frac{\partial^2 \ell(\alpha,\theta,Z_t)}{\partial(\alpha,\theta)\partial(\alpha,\theta)'}||\right) < \infty$; (4) $B_{*p} \equiv -E_0\left(\frac{\partial^2 \ell(\alpha^*,\theta^*,Z_t)}{\partial(\alpha,\theta)\partial(\alpha,\theta)'}\right)$ is nonsingular.

Assumption 6.3 $\frac{1}{\sqrt{n}}\sum_{t=2}^{n} \frac{\partial \ell(\alpha^*,\theta^*,Z_t)}{\partial(\alpha,\theta)} \to_d N(0, \Sigma_{*p})$ with $\Sigma_{*p} \equiv \lim_{n\to\infty} Var\{\frac{1}{\sqrt{n}}\sum_{t=2}^{n} \frac{\partial \ell(\alpha^*,\theta^*,Z_t)}{\partial(\alpha,\theta)}\} < \infty$.

Assumption 6.3 is satisfied by many well-known CLTs, such as Gordin's CLT for zero-mean ergodic stationary processes, which holds under assumptions M, 3.4(1) and $E_0\left(\frac{\partial\ell(\alpha^*,\theta^*,Z_t)}{\partial(\alpha,\theta)}\left[\frac{\partial\ell(\alpha^*,\theta^*,Z_t)}{\partial(\alpha,\theta)}\right]'\right) < \infty$. The next Proposition 6.1 follows trivially from propositions 7.3 and 7.8 of Hayashi (2000); hence we omit its proof.

Proposition 6.1 (possibly misspecified case): Let $(\hat{\alpha}_n^p, \hat{\theta}_n^p) = \arg\max_{(\alpha,\theta)\in\mathcal{A}\times\Theta} L_n(\alpha,\theta)$. Under Assumptions M and 6.1 - 6.3, we have:

$$\sqrt{n}\left((\hat{\alpha}_n^p, \hat{\theta}_n^p) - (\alpha^*, \theta^*)\right) \to_d N\left(0, B_{*p}^{-1}\Sigma_{*p}B_{*p}^{-1}\right).$$

6.1.3 Efficiency of correctly specified parametric MLE

Under assumption M and the correct specification of marginal $G(Y_t, \theta^*) = G_0(Y_t)$, we have: $\alpha^* = \alpha_0$.

Assumption 6.3' (1) the range of Y_t given Y_{t-1} does not depend of (α,θ); the 1st and 2nd order differentiations of $\ell(\alpha,\theta,Z_t)$ wrt $(\alpha,\theta)\in\mathcal{N}$ may be carried out under the integral sign, integration being wrt Y_t; (2) $\Sigma_{0p} \equiv E_0\left(\frac{\partial\ell(\alpha_0,\theta^*,Z_t)}{\partial(\alpha,\theta)}\{\frac{\partial\ell(\alpha_0,\theta^*,Z_t)}{\partial(\alpha,\theta)}\}'\right) < \infty$.

Under assumptions M and 6.3'(1), we have: $E_0\left[\frac{\partial\ell(\alpha_0,\theta^*,Z_t)}{\partial(\alpha,\theta)}|Y_{t-1},...,Y_1\right] = 0$ (see Lemma 4.1). Proposition 6.1 and Billingsley's (1961) ergodic stationary martingale difference CLT together imply the following result:

Proposition 6.2 (correctly specified case): Let $(\hat{\alpha}_n^p, \hat{\theta}_n^p) = \arg\max_{(\alpha,\theta)\in\mathcal{A}\times\Theta} L_n(\alpha,\theta)$. Under Assumptions M with $G(Y_t,\theta^*) = G_0(Y_t)$, 6.1, 6.2 and 6.3', we have: $\alpha^* = \alpha_0$, $B_{*p} = \Sigma_{*p} = \Sigma_{0p}$, and $(\hat{\alpha}_n^p, \hat{\theta}_n^p)$ is efficient for (α_0, θ^*):

$$\sqrt{n}\left((\hat{\alpha}_n^p, \hat{\theta}_n^p) - (\alpha_0, \theta^*)\right) \to_d N\left(0, \Sigma_{0p}^{-1}\right).$$

Moreover $\sqrt{n}\left(\hat{\alpha}_n^p - \alpha_0\right) \to_d N\left(0, \mathcal{I}_{*p}(\alpha_0)^{-1}\right)$ with

$$\mathcal{I}_{*p}(\alpha_0) \equiv \min_{\mathbf{b}} E_0\left(\begin{array}{c}\left(\frac{\partial\log c(U_{t-1},U_t;\alpha_0)}{\partial\alpha} - \frac{\partial\ell(\alpha_0,\theta^*,Z_t)}{\partial\theta}\mathbf{b}\right)\times \\ \left(\frac{\partial\log c(U_{t-1},U_t;\alpha_0)}{\partial\alpha} - \frac{\partial\ell(\alpha_0,\theta^*,Z_t)}{\partial\theta}\mathbf{b}\right)'\end{array}\right).$$

6.1.4 Ideal (or infeasible) MLE

We denote $\hat{\alpha}_n^{Ideal}$ as the ideal (or infeasible) MLE of the copula parameter α_0 when the marginal $G_0(\cdot)$ is assumed to be completely known. Proposition 6.2 implies the following result:

Proposition 6.3 (ideal MLE): Let $\hat{\alpha}_n^{Ideal} = \arg\max_{\alpha\in\mathcal{A}}\frac{1}{n}\sum_{t=2}^n\log c\left(U_{t-1},U_t;\alpha\right)$. Suppose that Assumption M holds with a completely known $G(\cdot,\theta) = G_0(\cdot)$. Let assumptions 4.1, 4.2 and 4.4' hold. Then: $B_0 \equiv -E_0\left(\frac{\partial^2\log c(U_{t-1},U_t;\alpha_0)}{\partial\alpha\partial\alpha'}\right) = \Sigma_{ideal}$ is finite and nonsingular, and $\hat{\alpha}_n^{Ideal}$ is efficient:

$$\sqrt{n}\left(\hat{\alpha}_n^{Ideal} - \alpha_0\right) \to_d N\left(0, \Sigma_{ideal}^{-1}\right).$$

Remark 6.1: Since $\mathcal{I}_*(\alpha_0) \leq \mathcal{I}_{*p}(\alpha_0) \leq \Sigma_{ideal}$, we have: $\mathcal{I}_*(\alpha_0)^{-1} \geq \mathcal{I}_{*p}(\alpha_0)^{-1} \geq \Sigma_{ideal}^{-1}$. Also Proposition 4.1 immediately implies that $\sigma_{2sp}^2 \geq \mathcal{I}_*(\alpha_0)^{-1}$.

Example 6.1' (the ideal MLE of Gaussian copula parameter): For the Gaussian copula Example 6.1, the Gaussian copula density function is

$$c(u_1, u_2; \alpha) = \frac{\phi_\alpha(\Phi^{-1}(u_1), \Phi^{-1}(u_2))}{\phi(\Phi^{-1}(u_1))\phi(\Phi^{-1}(u_2))}, \quad |\alpha| < 1.$$

where ϕ_α is the bivariate standard normal density with correlation α, and ϕ is the scalar standard normal density. Thus one can easily verify that

$$\Sigma_{ideal} = B_0 = -E_0\left(\frac{\partial^2 \log c(U_{t-1}, U_t; \alpha_0)}{\partial\alpha\partial\alpha}\right) = \frac{1+\alpha_0^2}{(1-\alpha_0^2)^2} < \infty \text{ if } \alpha_0^2 \neq 1$$

Consequently, $\sqrt{n}\left(\hat{\alpha}_n^{Ideal} - \alpha_0\right) \to_d N\left(0, \Sigma_{ideal}^{-1}\right)$ with $\Sigma_{ideal}^{-1} = (1-\alpha_0^2) \times \frac{1-\alpha_0^2}{1+\alpha_0^2}$.

We note that $Avar(\hat{\alpha}_n^{Ideal}) = \Sigma_{ideal}^{-1} < (1-\alpha_0^2) = Avar(\hat{\alpha}_n^{2sp})$, and $Avar(\hat{\alpha}_n^{Ideal}) = Avar(\hat{\alpha}_n^{2sp})$ if and only if $\alpha_0 = 0$ (i.e., independence). Also $Avar(\hat{\alpha}_n^{Ideal})$ is decreasing in $|\alpha_0|$.

Example 2.1' (the ideal MLE of Clayton copula parameter): For the Clayton copula Example 2.1, the Clayton copula density function is given by

$$c(u_1, u_2, \alpha) = (1+\alpha)u_1^{-(1+\alpha)}u_2^{-(1+\alpha)}(u_1^{-\alpha} + u_2^{-\alpha} - 1)^{-(1/\alpha+2)}, \ \alpha > 0.$$

By some tedious calculation,

$$\begin{aligned} \Sigma_{ideal} &= B_0 = -E_0\left(\frac{\partial^2 \log c(U_{t-1}, U_t; \alpha_0)}{\partial\alpha\partial\alpha}\right) \\ &= \frac{1}{\alpha(1+\alpha)} + \frac{1}{\alpha(1+\alpha)^2(1+2\alpha)} + \frac{(1+\alpha)(1+2\alpha)}{\alpha^5} \times Int(\alpha) \end{aligned}$$

where $Int(\alpha) = \int_1^\infty \int_1^\infty \frac{xy(\log x - \log y)^2 - x(\log x)^2 - y(\log y)^2}{(x+y-1)^{4+1/\alpha}} dxdy$, which is a small number bounded in $[-1, 1]$. Therefore, $\Sigma_{ideal} \in (0, \infty)$ provided that $\alpha_0 > 0$. Hence $\sqrt{n}\left(\hat{\alpha}_n^{Ideal} - \alpha_0\right) \to_d N\left(0, \Sigma_{ideal}^{-1}\right)$, where the asymptotic variance Σ_{ideal}^{-1} is increasing in α_0 and is $O(\alpha_0^2)$.

6.2 Simulations

We consider two Markov models generated by different copulas (Clayton copula and Gumbel copula) but with same marginal distribution (the Student's t distribution with 5 degree of freedom, $t[5]$). We simulate a strictly stationary first-order Markov process $\{Y_t\}$ from a specified bivariate copula $C(u_1, u_2; \alpha_0)$ with given invariant cdf G_0 as follows:

Step 1: Generate an IID sequence of uniform random variables $\{V_t\}_{t=1}^n$

Step 2: Set $U_1 = V_1$ and $U_t = C_{2|1}^{-1}[V_t|U_{t-1}, \alpha_0]$.

Step 3: Set $Y_t = G_0^{-1}(U_t)$

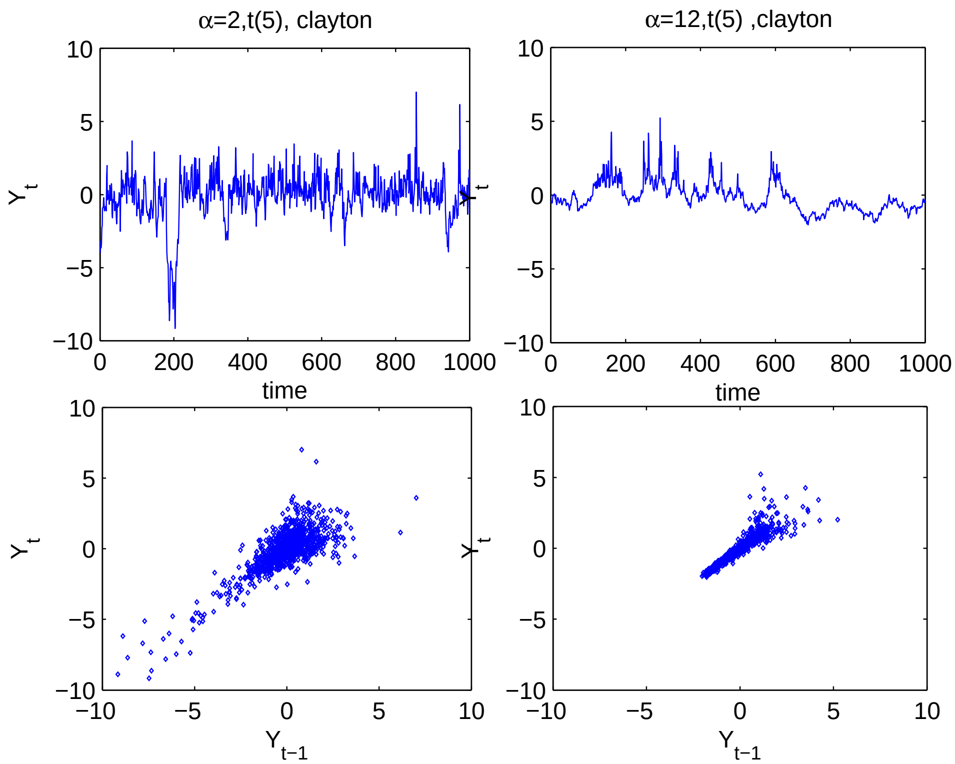


Figure 2: Clayton copula ($\alpha = 2$ *and* 12) and Student's t(5) distribution

In our simulation, the true marginal distribution is $t[5]$, with density $g_0(y) = \frac{\Gamma(3)}{\sqrt{5\pi}\Gamma(5/2)}(1+\frac{y^2}{5})^{-3}$. For each specified copula $C(u_1, u_2; \alpha_0)$, we generate a long time series but delete the first 2000, and keep the last 1000 observation as our simulated data sample data $\{Y_t\}$ (i.e., simulated sample size $n = 1000$). Figure 2 reports typical simulated Clayton-copula Markov time series with parameter values $\alpha = 2$, 12 (the corresponding Kendall's tau values are $\tau = 0.5$, 0.857) respectively. Figure 3 reports typical simulated Gumbel-copula Markov time series with parameter values $\alpha = 2$, 7 (the corresponding Kendall's tau values are $\tau = 0.5$, 0.857) respectively.

For both copula-based Markov models and for each simulated sample, we compute five estimators of α_0: sieve MLE, ideal (or infeasible) MLE, two-step estimator, correctly specified parametric MLE (functional form of g is correctly specified) and misspecified parametric MLE (functional form of g is misspecified). Sieve MLEs are computed by maximizing the joint log-likelihood $L_n(\alpha, g)$ in (3.1) using either polynomial sieve or polynomial spline sieve to approximate the log-marginal density ($\log g$). The selection of number of sieve terms $\widehat{K}$ is based on the so-called small sample AIC of Burnham and Anderson (2002): $\widehat{K} = \arg\max_K \{L_n(\widehat{\gamma}_n(K)) - K/(n-K-1)\}$, where $\widehat{\gamma}_n(K)$ is the sieve MLE of $\gamma_0 = (\alpha_0, g_0)$ using K as the sieve number of terms.

We compare the estimates of copula dependence parameter, and the estimates of 1/3 and 2/3 marginal quantiles in terms of monte Carlo mean, bias, variance, mean squared errors and confidence

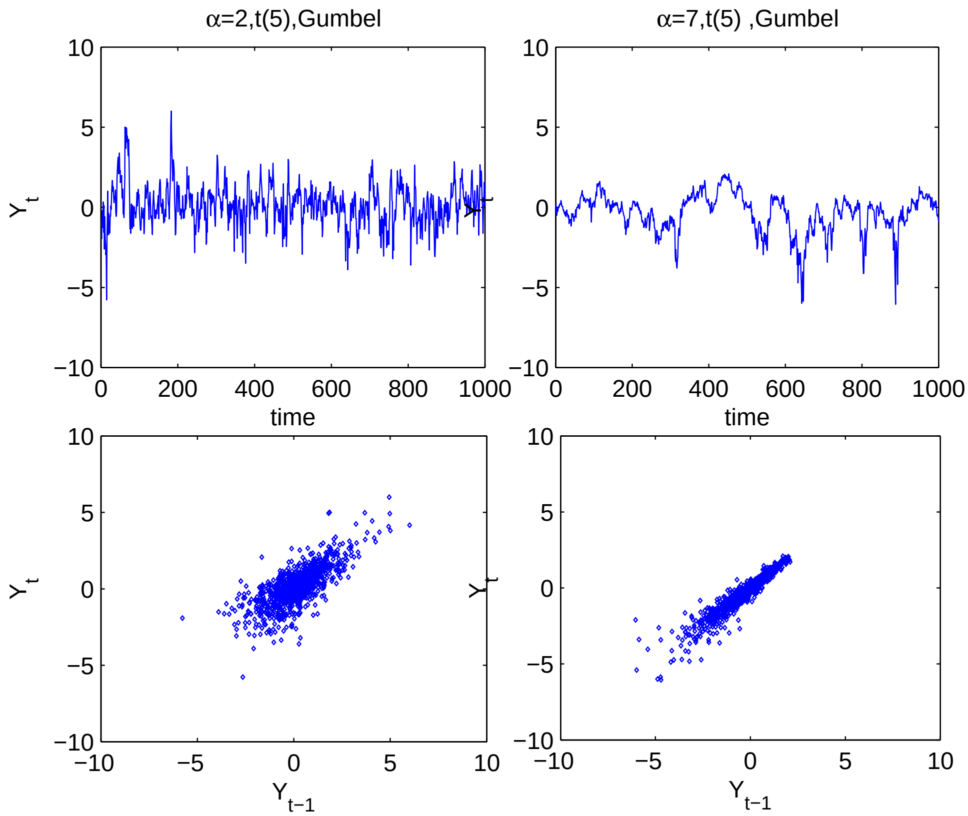


Figure 3: Gumbel copula ($\alpha = 2$ *and* 7) and Student's t(5) distribution

region. We also illustrate the performance of sieve MLE of the marginal density function. We run Monte Carlo simulation MC times ($MC = 1000$ in most of the reported results) and summarize the results in tables and figures listed in Appendix B.

For Clayton copula generated Markov model, we also construct χ^2 inverted confidence interval (based on 500 Monte Carlo simulations) and report the estimates of 0.01 conditional quantile function.

Since the two step estimator of Chen and Fan (2006) performs terribly for the Clayton copula generated Markov model when α is big, we also compute and compare several other 2step estimators that differ from each other by different ways of estimating marginal cdf in the first step. 2step-sieve estimator estimate marginal density via sieve marginal maximum likelihood in the first step; 2step-para estimator computes the marginal density via parametric marginal maximum likelihood with a correctly specified marginal; 2step-mis estimator computes the marginal density via parametric marginal maximum likelihood with a misspecified marginal. Our simulation results show that all these 2step procedures perform worse than the correctly one step procedures (such as parametric MLE and sieve MLE).

Brief summary of MC results: In Appendix B we present many tables and figures to report the Monte Carlo findings in details. Here we give a brief summary of the overall patterns: (1) sieve MLE always perform better than two step estimator in terms of bias and MSE; (2) for strong tail dependence case, the two step based estimators of copula dependence parameter perform poorly, having big bias and big MSE; (3) for strong tail dependence case, empirical cdf estimator of marginals perform poorly in terms of bias and variance; (4) extreme conditional quantiles estimated via sieve MLE is much more precise than those estimated via two-step estimators; (5) parametric MLE with correctly specified marginals is the most efficient one and also has the smallest bias, but misspecified parametric MLE could lead to inconsistent estimation of copula dependence parameter (in addition to inconsistent estimation of marginal parameter). In summary we recommend sieve MLE to estimate copula-based Markov models and its implied conditional quantiles (VaRs).

7 Conclusions

In this paper, we first show that several widely used tail dependent copula generated Markov models are in fact geometric ergodic (and geometric beta mixing), albeit their time series plots may look highly persistent and 'long memory alike'. We then propose a sieve MLE for the class of first order strictly stationary copula-based semiparametric Markov models that are characterized by the parametric copula dependence parameter α_0 and the unknown invariant density $g_0()$. We show that the sieve MLE of any smooth functionals of (α_0, g_0) are root-n consistent, asymptotically normal and efficient; and that the sieve likelihood ratio statistics is chi-square distributed. Monte Carlo studies indicate that, even for tail dependent copula based semiparametric Markov models, the sieve MLEs of the copula dependence parameter, the marginal cdf and the conditional quantiles all perform very well in finite samples.

In this paper we propose either consistent plug-in estimation of asymptotic variance or by inverting profiled likelihood criterion function to construct confidence region for the sieve MLE $\widehat{\alpha}$ of α_0. In another paper, we extend the result of Andrews (2001) on parametric bootstraps for parametric Markov models to a semiparametric bootstrap for our copula-based semiparametric Markov models.

In this paper we assume that the parametric copula function is correctly specified. We could test this assumption by performing a sieve likelihood ratio test; see e.g., Fan and Jiang (2007) for a recent review about generalized likelihood ratio tests. Alternatively, we could also consider a joint sieve ML estimation of nonparametric copula and nonparametric marginal. Recently Chen et al (2009) provide an empirical likelihood estimation of nonparametric copula using a bivariate random sample; their method could be extended to out time series setting.

A Mathematical Proofs

We first recall some definitions and results for Markov processes.from Chen and Tong (2001).

Definition A.1. *Let $\{Y_t\}$ be an irreducible Markov Chain on with transition measure $P^n(y;A) = P(Y_{t+n} \in A|Y_t = y)$, $n \geq 1$. A non-null set C is called small if there exists a positive integer n, a constant $b > 0$, and a probability measure $\nu(\cdot)$ such that $P^n(y;A) \geq b\nu(A)$ for all $x \in C$ and all measurable set A.*

Theorem A.1. *(Theorem B.1.4 in Chan and Tong, 2001) Let $\{Y_t\}$ be an irreducible and aperiodic Markov Chain. Suppose there exists a small set C, a nonnegative measurable function g which is bounded away from 0 and ∞ on C, and constants $r > 1$, $\gamma, K > 0$ such that*

$$rE[L(Y_{t+1})|Y_t = y] \leq L(y) - \gamma, \text{ for all } y \notin C,$$

and, let C' be the complement of C,

$$\int_{C'} L(w)P(y, dw) < K, \text{ for all } y \in C.$$

Then $\{Y_t\}$ is geometric ergodic. Here L is called the Lyapunov function.

Proof. of Theorem 2.1: We establish the results by applying Theorem B.1.4 of Chan and Tong (2001, the famous Tweedie's drift criterion) or applying Proposition 2.1(i) of Chen and Fan (2006).

(1) For Clayton copula, let $\{Y_t\}$ be a stationary Markov process of order 1 generated from a bivariate Clayton copula and a marginal cdf $G_0(\cdot)$. Then the transformed process $\{U_t \equiv G_0(Y_t)\}$ has uniform marginals and Clayton copula joint distribution of (U_{t-1}, U_t). When $\alpha = 0$ Clayton copula becomes the independence copula; hence the process $\{U_t \equiv G_0(Y_t)\}$ is i.i.d. and trivially geometric ergodic.

Let $\alpha > 0$, and let $\{V_t\}$ be a sequence of i.i.d. uniform(0,1) random variables. Then the following nonlinear AR(1) model is generated by the Clayton copula:

$$X_t = (V_t^{-\alpha/(1+\alpha)} - 1)X_{t-1} + 1 \quad \text{with } X_t^{-1/\alpha} \equiv U_t.$$

Thus X_t has stationary distribution and $X_t^{-1/\alpha} = U_t \sim$ uniform(0,1). Note that the state space of $\{X_t\}$ is $(1, \infty)$. Since

$$E_0[(V_t^{-\alpha/(1+\alpha)} - 1)^{1/\alpha}] = 1,$$

we can let $p \in (0, 1/\alpha)$, and $L(x) = x^p > 1$ be the Lyapunov function. Then $\rho \equiv E_0[L(V_t^{-\alpha/(1+\alpha)} - 1)] < 1$. Let $r = \rho^{-1/2} > 1$ and

$$x_0 = \max\{x \geq 1 : \quad rE_0[|x(V_t^{-\alpha/(1+\alpha)} - 1) + 1|^p] \geq x^p - 1\}.$$

Such x_0 always exists since

$$\lim_{x\to\infty} \frac{rE_0[|x(V_t^{-\alpha/(1+\alpha)} - 1) + 1|^p]}{x^p - 1} = r\rho = \rho^{1/2} < 1.$$

Choose the small set $S = [1, x_0]$. Clearly g is bounded away from 0 and ∞ on S. We now show that S is a small set. Let $f(\cdot|x)$ be the conditional density function of X_1 given $X_0 = x$. Then

$$f(y|x) = \frac{1+\alpha}{\alpha(y-1+x)^{2+1/\alpha}} \geq \frac{1+\alpha}{\alpha(y-1+x_0)^{2+1/\alpha}}$$

if $x \leq x_0$. Choose the probability measure ν on $(1, \infty)$ as $\nu(dy) = f(y|x_0)dy$. Then

$$\Pr(X_1 \in A | X_0 = x) \geq \nu(A), \text{ for all } x \in S \text{ and } A \in \mathcal{B}.$$

Hence S is indeed a small set (cf page 255 in Chan and Tong (2001)). Notice that

$$rE_0[L(X_1)|X_0 = x] \leq L(x) - 1, \text{ for all } x > x_0,$$

$$E_0[L(X_1)|X_0 = x] < \infty, \text{ for all } x \in S = [1, x_0],$$

thus all conditions in Theorem B.1.4 of Chan and Tong (2001) are satisfied; hence $\{X_t\}$ is geometric ergodic, and geometric beta mixing (or absolutely regular with geometrically decaying coefficients).

(2) For Gumbel copula, let $\{Y_t\}$ be a stationary Markov process of order 1 generated from a bivariate Gumbel copula and a marginal cdf $G_0(\cdot)$. Then the transformed process $\{U_t \equiv G_0(Y_t)\}$ has uniform marginals and (U_{t-1}, U_t) has the following Gumbel copula joint distribution:

$$C(u_1, u_2; \alpha) = \exp\{-[(-\log u_1)^\alpha + (-\log u_2)^\alpha]^{1/\alpha}\}, \quad 0 < u_1, u_2 < 1, \ \alpha \geq 1.$$

When $\alpha = 1$ Gumbel copula becomes the independence copula; hence the process $\{U_t \equiv G_0(Y_t)\}$ is i.i.d. and trivially geometric ergodic.

Let $\alpha > 1$. Let $X_t = (-\log U_t)^\alpha$. Then $U_t = F(X_t)$, with $F(x) = \exp\{-x^{1/\alpha}\}$. Let $f(x) = -F'^{-1} x^{1/\alpha - 1} \exp\{-x^{1/\alpha}\}$. Then for X_t we have

$$\Pr(X_{t+1} \geq x_2 | X_t = x_1) = \frac{f(x_1 + x_2)}{f(x_1)}, \quad x_1, x_2 > 0.$$

Hence

$$\begin{aligned} E_0(X_{t+1}|X_t = x_1) &= \int_0^\infty \Pr(X_{t+1} \geq x_2 | X_t = x_1)\, dx_2 = \int_0^\infty \frac{f(x_1 + x_2)}{f(x_1)} dx_2 \\ &= \frac{F(x_1)}{f(x_1)} = \alpha x_1^{1-(1/\alpha)}. \end{aligned}$$

Note that as $x_1 \to 0$,

$$\begin{aligned} E_0\left(X_{t+1}^{-1/(2\alpha)}|X_t = x_1\right) &= \int_0^\infty x_2^{-1/(2\alpha)} \frac{-f'(x_1+x_2)}{f(x_1)} dx_2 \\ &= x_1^{1-1/(2\alpha)} \int_0^\infty u^{-1/(2\alpha)} \frac{-f'(x_1+x_1u)}{f(x_1)} du \\ &\sim x_1^{-1/(2\alpha)}(1-1/\alpha) \int_0^1 t^{-1/(2\alpha)}(1-t)^{-1/(2\alpha)} dt \end{aligned}$$

where the last relation is due to

$$\lim_{x_1 \to 0} \frac{-f'(x_1+x_1u)}{f(x_1)} \times x_1 = (1-1/\alpha)(1+u)^{1/\alpha-2}.$$

Observe that, as $\alpha > 1$,

$$\kappa_\alpha \equiv (1-1/\alpha) \int_0^1 t^{-1/(2\alpha)}(1-t)^{-1/(2\alpha)} dt = (1-1/\alpha) \times B\left(1-1/(2\alpha), 1-1/(2\alpha)\right) < 1$$

where $B()$ is the beta function.

Let $L(x) = x^{-1/(2\alpha)} + x$ be the Lyapunov function. Let $r = \inf_{x>0} L(x)/2$. Then:

$$\lim_{x \to \infty} \frac{E_0(L(X_{t+1})|X_t = x)}{L(x) - r} = 0,$$

and

$$\lim_{x \to 0} \frac{E_0(L(X_{t+1})|X_t = x)}{L(x) - r} = \kappa_\alpha < 1.$$

Let $S = [1/\lambda, \lambda]$ with sufficient large $\lambda > 0$. Then S is a small set. So all conditions in Theorem B.1.4 of Chan and Tong (2001) are satisfied; hence $\{X_t\}$ is geometric ergodic and geometric beta mixing.

(3) For Student's t copula, let $\{Y_t\}$ be a stationary Markov process of order 1 generated from a bivariate t-copula and a marginal cdf $G_0(\cdot)$. Then the transformed process $\{U_t \equiv G_0(Y_t)\}$ satisfies the following:

$$t_\nu^{-1}(U_t) = \rho t_\nu^{-1}(U_{t-1}) + \sigma(U_{t-1})e_t, \quad \sigma(U_{t-1}) = \sqrt{\frac{\nu + (t_\nu^{-1}(U_{t-1}))^2}{\nu+1}(1-\rho^2)},$$

where $e_t \sim t_{\nu+1}$, and is independent of $U^{t-1} \equiv (U_{t-1}, ..., U_1)$ (see, e.g., Chen et al. 2008). Let $X_t \equiv t_\nu^{-1}(U_t)$. Then

$$X_t = \rho X_{t-1} + \sigma(X_{t-1})e_t, \quad \sigma(X_{t-1}) = \sqrt{\frac{\nu + (X_{t-1})^2}{\nu+1}(1-\rho^2)},$$

where $e_t \sim t_{\nu+1}$, and is independent of $X^{t-1} \equiv (X_{t-1}, ..., X_1)$. Let $L(x) = |x| + 1 \geq 1$ be the Lyapunov function. Then $E_0\{L(X_t)\} = \sqrt{\nu}\frac{\Gamma(\frac{\nu-1}{2})}{\sqrt{\pi}\Gamma(\nu/2)} + 1 < \infty$ provided that $\nu > 1$. Then:

$$\begin{aligned} E_0\left(L(X_t)|X_{t-1} = x\right) &= E_0\left(|\rho X_{t-1} + \sigma(X_{t-1})e_t| \; |X_{t-1} = x\right) + 1 = E_0\left(|\rho x + \sigma(x)e_t|\right) + 1 \\ &< \sqrt{E_0\left(|\rho x + \sigma(x)e_t|^2\right)} + 1 = \sqrt{\left(\rho^2 x^2 + \sigma^2(x)E_0[e_t^2]\right)} + 1, \end{aligned}$$

where the strict inequality is due to $e_t \sim t_{\nu+1}$ and for fixed x,

$$0 < Var\left(|\rho x + \sigma(x) e_t|^2\right) = E\left(|\rho x + \sigma(x) e_t|^2\right) - \left[E_0\left(|\rho x + \sigma(x) e_t|\right)\right]^2 .$$

Since

$$\sigma^2(x) = \frac{1-\rho^2}{\nu+1}(\nu + x^2),$$

we have

$$\begin{aligned}
\lim_{|x|\to\infty} \frac{E_0\left(L(X_t)|X_{t-1}=x\right)}{L(x)} &= \lim_{|x|\to\infty} \frac{E_0\left(|\rho x + \sigma(x) e_t|\right) + 1}{|x|+1} \\
&< \lim_{|x|\to\infty} \frac{\sqrt{\left(\rho^2 x^2 + \sigma^2(x) E_0[e_t^2]\right)} + 1}{|x|+1} \\
&= \sqrt{\rho^2 + \frac{1-\rho^2}{\nu+1} E_0[e_t^2]} \\
&\leq \sqrt{\rho^2 + \frac{1-\rho^2}{2+1} E_0[t_3^2]} = 1,
\end{aligned}$$

where the last inequality is due to $E_0[e_t^2]/(\nu+1)$ decreasing in $\nu \in [2, \infty]$, and the last equality is due to $E_0[t_3^2] = 3$. Then we can choose a small set $S = [-x_0, x_0]$ with sufficiently large $x_0 > 0$. Clearly the density of e_t is bounded from above and below on a compact set. Hence, all conditions in Theorem B.1.4 of Chan and Tong (2001) or in Proposition 2.1(i) of Chen and Fan (2006) are satisfied, and $\{X_t\}$ is geometric ergodic (hence geometric beta-mixing). □

Proof. of Proposition 3.1: Since most of the conditions of consistency theorem 3.1 of Chen (2007) are already assumed in our assumptions M, 3.1 and 3.2, it suffices to verify condition 3.5 (uniform convergence over sieves) of Chen (2007). Assumptions M implies that $\{Y_t\}_{t=1}^n$ is stationary ergodic. This and assumption 3.2 imply that Glivenko-Cantelli theorem for stationary ergodic processes is applicable, and hence:

$$\sup_{\gamma \in \Gamma_n} |L_n(\gamma) - E\{L_n(\gamma)\}| = o_p(1).$$

The result now follows from theorem 3.1 of Chen (2007). □

Proof. of Lemma 4.1: For **(1)**, recall that $Z_t = (Y_{t-1}, Y_t)$, under assumption M, for all $s < t$,

$$\begin{aligned}
& E_0\left(\left(\frac{\partial \ell(\gamma_0, Z_t)}{\partial \gamma'}[v]\right)\left(\frac{\partial \ell(\gamma_0, Z_s)}{\partial \gamma'}[\tilde{v}]\right)\right) \\
= \;& E_0\left(E_0\left(\left(\frac{\partial \ell(\gamma_0, Z_t)}{\partial \gamma'}[v]\right)\left(\frac{\partial \ell(\gamma_0, Z_s)}{\partial \gamma'}[\tilde{v}]\right) \mid Y_1, \cdots, Y_{t-1}\right)\right) \\
= \;& E_0\left(\left(\frac{\partial \ell(\gamma_0, Z_s)}{\partial \gamma'}[\tilde{v}]\right) E_0\left(\frac{\partial \ell(\gamma_0, Z_t)}{\partial \gamma'}[v] \mid Y_{t-1}\right)\right).
\end{aligned}$$

Recall that the true conditional density function is: $p^0(Y_t|Y^{t-1}) = g_0(Y_t) \times c\left(G_0(Y_{t-1}), G_0(Y_t); \alpha_0\right) = h(Y_t|Y_{t-1}; \gamma_0)$. We have:

$$\begin{aligned} E_0\left(\frac{\partial \ell(\gamma_0, Z_t)}{\partial \gamma'}[v] \mid Y_{t-1}\right) &= \int \frac{\frac{\partial h(y_t|Y_{t-1};\gamma_0)}{\partial \gamma'}}{h(y_t|Y_{t-1};\gamma_0)}[v] h(y_t|Y_{t-1};\gamma_0) dy_t \\ &= \int \frac{\partial h(y_t|Y_{t-1};\gamma_0)}{\partial \gamma'}[v] dy_t \\ &= \frac{d\left(\int h(y_t|Y_{t-1};\gamma_0 + sv) dy_t\right)}{ds}|_{s=0} = \frac{d(1)}{ds}|_{s=0} = 0, \end{aligned}$$

where the order of differentiation and integration can be reversed due to Assumption 4.3.

For **(2)**, the above equality also implies the sequence is a martingale difference process.

For **(3)**, Since $\int h(y|Y_{t-1}; \gamma_0 + sv) dy \equiv 1$, by differentiating this equation with respect to s twice and evaluating it at $s = 0$, we get $E_0\left(\left(\frac{\partial \ell(\gamma_0, Z_t)}{\partial \gamma'}[v]\right)^2 |Y_{t-1}\right) = -E_0\left(\frac{\partial^2 \ell(\gamma_0, Z_t)}{\partial \gamma \partial \gamma'}[v, v]|Y_{t-1}\right)$, where the interchange of differentiation and integration is guaranteed by Assumption 4.3. This we obtain (3). $\square$

Proof. of Theorem 4.1: Let ϵ_n be any positive sequence satisfying $\epsilon_n = o(n^{-1/2})$. Denote $r[\gamma, \gamma_0, Z_t] \equiv \ell(\gamma, Z_t) - \ell(\gamma_0, Z_t) - \frac{\partial \ell(\gamma_0, Z_t)}{\partial \gamma'}[\gamma - \gamma_0]$. Then by the definition of sieve MLE $\widehat{\gamma}_n$ (with abuse of notation, we denote it as $\hat{\gamma}$ in the following),

$$\begin{aligned} & 0 \le \frac{1}{n}\sum_{t=2}^{n}[\ell(\hat{\gamma}, Z_t) - \ell(\hat{\gamma} \pm \epsilon_n \Pi_n v^*, Z_t)] \\ = \;& \mu_n\left(\ell(\hat{\gamma}, Z_t) - \ell(\hat{\gamma} \pm \epsilon_n \Pi_n v^*, Z_t)\right) + E_0\left(\ell(\hat{\gamma}, Z_t) - \ell(\hat{\gamma} \pm \epsilon_n \Pi_n v^*, Z_t)\right) + o_p(n^{-1}) \\ = \;& \mp \epsilon_n \frac{1}{n}\sum_{t=2}^{n} \frac{\partial \ell(\gamma_0, Z_t)}{\partial \gamma'}[\Pi_n v^*] + \mu_n\left(r[\hat{\gamma}, \gamma_0, Z_t] - r[\hat{\gamma} \pm \epsilon_n \Pi_n v^*, \gamma_0, Z_t]\right) \\ & + E_0\left(r[\hat{\gamma}, \gamma_0, Z_t] - r[\hat{\gamma} \pm \epsilon_n \Pi_n v^*, \gamma_0, Z_t]\right) + o(n^{-1}). \end{aligned}$$

Claim 1: $\frac{1}{n}\sum_{t=2}^{n} \frac{\partial \ell(\gamma_0, Z_t)}{\partial \gamma'}[\Pi_n v^* - v^*] = o_p(n^{-1/2})$. This claim is true due to Chebyshev inequality, serially uncorrelated (Lemma 4.1) and identically distributed data, and $\|\Pi_n v^* - v^*\| = o(1)$.
Claim 2: $\mu_n\left(r[\hat{\gamma}, \gamma_0, Z_t] - r[\hat{\gamma} \pm \epsilon_n \Pi_n v^*, \gamma_0, Z_t]\right) = \epsilon_n \times o_p(n^{-1/2})$. This claim holds since

$$\begin{aligned} & \mu_n\left(r[\hat{\gamma}, \gamma_0, Z_t] - r[\hat{\gamma} \pm \epsilon_n \Pi_n v^*, \gamma_0, Z_t]\right) \\ = \;& \mu_n\left(\ell(\hat{\gamma}, Z_t) - \ell(\hat{\gamma} \pm \epsilon_n \Pi_n v^*, Z_t) \pm \epsilon_n \frac{\partial \ell(\gamma_0, Z_t)}{\partial \gamma'}[\Pi_n v^*]\right) \\ = \;& \mp \epsilon_n \mu_n\left(\frac{\partial \ell(\tilde{\gamma}, Z_t)}{\partial \gamma'}[\Pi_n v^*] - \frac{\partial \ell(\gamma_0, Z_t)}{\partial \gamma'}[\Pi_n v^*]\right) = \epsilon_n \times o_p(n^{-1/2}) \end{aligned}$$

where $\tilde{\gamma} \in \Gamma_n$ lies between $\hat{\gamma}$ and $\hat{\gamma} \pm \epsilon_n \Pi_n v^*$, and the last equality is implied by assumption 4.7.

Claim 3: $E_0\left(r[\hat{\gamma},\gamma_0,Z_t]-r[\hat{\gamma}\pm\epsilon_n\Pi_n v^*,\gamma_0,Z_t]\right)=\pm\epsilon_n\left\langle\hat{\gamma}-\gamma_0,v^*\right\rangle+\epsilon_n o_p(n^{-1/2})+o_p(n^{-1})$.

Note that

$$
\begin{aligned}
& E_0(r[\gamma,\gamma_0,Z_t]) = E_0\left(\ell(\gamma,Z_t)-\ell(\gamma_0,Z_t)-\frac{\partial\ell(\gamma_0,Z_t)}{\partial\gamma'}[\gamma-\gamma_0]\right) \\
= \; & \frac{1}{2}E_0\left(\frac{\partial^2\ell(\tilde{\gamma},Z_t)}{\partial\gamma\partial\gamma'}[\gamma-\gamma_0,\gamma-\gamma_0]-\frac{\partial^2\ell(\gamma_0,Z_t)}{\partial\gamma\partial\gamma'}[\gamma-\gamma_0,\gamma-\gamma_0]\right) \\
& +\frac{1}{2}E_0\left(\frac{\partial^2\ell(\gamma_0,Z_t)}{\partial\gamma\partial\gamma'}[\gamma-\gamma_0,\gamma-\gamma_0]\right)+\epsilon_n\times o_p(n^{-1/2}) \\
= \; & \frac{1}{2}E_0\left(\frac{\partial^2\ell(\gamma_0,Z_t)}{\partial\gamma\partial\gamma'}[\gamma-\gamma_0,\gamma-\gamma_0]\right)+\epsilon_n\times o_p(n^{-1/2})+o_p(n^{-1})
\end{aligned}
$$

where $\tilde{\gamma}\in\Gamma_n$ is located between γ and γ_0, and the last equality is due to assumption 4.6. By Lemma 4.1 (3), we have:

$$
\|\gamma-\gamma_0\|^2\equiv E_0\left[\left(\frac{\partial\ell(\gamma_0,Z_t)}{\partial\gamma'}[\gamma-\gamma_0]\right)^2\right]=-E_0\left(\frac{\partial^2\ell(\gamma_0,Z_t)}{\partial\gamma\partial\gamma'}[\gamma-\gamma_0,\gamma-\gamma_0]\right).
$$

Therefore,

$$
\begin{aligned}
& E_0\left(r[\hat{\gamma},\gamma_0,Z_t]-r[\hat{\gamma}\pm\epsilon_n\Pi_n v^*,\gamma_0,Z_t]\right) \\
= \; & -\frac{\|\hat{\gamma}-\gamma_0\|^2-\|\hat{\gamma}\pm\epsilon_n\Pi_n v^*-\gamma_0\|^2}{2}+o_p(\epsilon_n n^{-1/2})+o_p(n^{-1}) \\
= \; & \pm\epsilon_n\langle\hat{\gamma}-\gamma_0,\Pi_n v^*\rangle+\frac{1}{2}\|\epsilon_n\Pi_n v^*\|^2+o_p(\epsilon_n n^{-1/2})+o_p(n^{-1}) \\
= \; & \pm\epsilon_n\times\langle\hat{\gamma}-\gamma_0,v^*\rangle+\epsilon_n\times o_p(n^{-1/2})+o_p(n^{-1}).
\end{aligned}
$$

In summary, Claims 1, 2 and 3 imply that

$$
\begin{aligned}
& 0\le\frac{1}{n}\sum_{t=2}^{n}[\ell(\hat{\gamma},Z_t)-\ell(\hat{\gamma}\pm\epsilon_n\Pi_n v^*,Z_t)] \\
= \; & \mp\epsilon_n\frac{1}{n}\sum_{t=2}^{n}\frac{\partial\ell(\gamma_0,Z_t)}{\partial\gamma'}[v^*]\pm\epsilon_n\times\langle\hat{\gamma}-\gamma_0,v^*\rangle+\epsilon_n\times o_p(n^{-1/2})+o_p(n^{-1}) \\
= \; & \mp\epsilon_n\mu_n\left(\frac{\partial\ell(\gamma_0,Z_t)}{\partial\gamma'}[v^*]\right)\pm\epsilon_n\times\langle\hat{\gamma}-\gamma_0,v^*\rangle+\epsilon_n\times o_p(n^{-1/2})+o_p(n^{-1})
\end{aligned}
$$

Thus we obtain:

$$
\sqrt{n}\langle\hat{\gamma}-\gamma_0,v^*\rangle=\sqrt{n}\mu_n\left(\frac{\partial\ell(\gamma_0,Z_t)}{\partial\gamma'}[v^*]\right)+o_p(1)\Rightarrow N(0,\|v^*\|^2),
$$

where the asymptotic normality is guaranteed by Billingsley's (1961) ergodic stationary martingale difference CLT, and the asymptotic variance being equal to $\|v^*\|^2\equiv\|\frac{\partial\rho(\gamma_0)}{\partial\gamma'}\|^2$ is implied by Lemma 4.1 (1) and the definition of the Fisher norm $\|\cdot\|$. $\square$

Proof. of Theorem 4.2: Given our normality results in Theorem 4.1, for our model we can take $\Sigma_n(v) = \frac{1}{\sqrt{n}}\sum_{t=2}^{n}\frac{\partial l(\gamma_0,Z_t)}{\partial\gamma'}[v]$, which is linear in v and converges in distribution to $N(0,||v||^2)$, and $\frac{1}{2n}\sum_{t=2}^{n}\left(\frac{\partial l(\gamma_0,Z_t)}{\partial\gamma'}[v]\right)^2 = \frac{1}{2}||v||^2 + o_p(1)$ hence LAN holds. Notice that the proof in Wong (1992) allows for time series data, following his proof, under LAN, we obtain that $\rho(\widehat{\gamma}_n)$ achieves the lower efficiency bound.

Actually, from the last equation in our proof of Theorem 4.1, we have:

$$\rho(\widehat{\gamma}_n) - \rho(\gamma_0) = \langle\widehat{\gamma}_n - \gamma_0, v^*\rangle + o_p(n^{-1/2}) = \mu_n\left(\frac{\partial\ell(\gamma_0,Z_t)}{\partial\gamma'}[v^*]\right) + o_p(n^{-1/2}),$$

which means $\rho(\widehat{\gamma}_n)$ is a regular asymptotically linear estimate and its influence function equals to $\frac{\partial\ell(\gamma_0,\cdot)}{\partial\gamma'}[v^*]$ that belongs to the tangent space of the model. So we can also conclude that $\rho(\widehat{\gamma}_n)$ is semiparametrically efficient by applying the result of Bickel and Kwon (2001), which allows for strictly stationary semiparametric Markov models. $\square$

Proof. of Proposition 4.1: Thanks to Lemma 4.1, the score space in this time-series setup acts in much the same way as the score space when data are i.i.d. So the semiparametric efficiency bound for α_0 is $\mathcal{I}_*(\alpha_0) = E_0\left\{\mathcal{S}_{\alpha_0}\mathcal{S}'_{\alpha_0}\right\}$, where $\mathcal{S}_{\alpha_0}$ is the *efficient score* function for α_0, which is defined as the ordinary score function for α_0 minus its population least squares orthogonal projection onto the closed linear span (clsp) of the score functions for the nuisance parameters g_0. And α_0 is $\sqrt{n}$*-efficiently estimable* if and only if $E_0\left\{\mathcal{S}_{\alpha_0}\mathcal{S}'_{\alpha_0}\right\}$ is *non-singular*; see e.g. Bickel et al. (1993) (we can directly extend the result for i.i.d. case to this Markov time-series setting). Hence (4.3) is clearly a necessary condition for $\sqrt{n}$-normality and efficiency of $\widehat{\alpha}_n$ for α_0. Under Assumptions 4.2, 4.3 and 4.4', Propositions 4.7.4 and 4.7.6 of Bickel, et al. (1993, pages 165 - 168) for bivariate copula models apply. Therefore with $\mathcal{S}_{\alpha_0}$ defined in (4.4), we have that $\mathcal{I}_*(\alpha_0) = E_0\left\{\mathcal{S}_{\alpha_0}\mathcal{S}'_{\alpha_0}\right\}$ is finite, positive-definite. This implies that Assumption 4.4 is satisfied with $\rho(\gamma) = \lambda'\alpha$ and $\omega = \infty$ and $||v^*||^2 = ||\frac{\partial\rho(\gamma_0)}{\partial\gamma'}||^2 = \lambda'\mathcal{I}_*(\alpha_0)^{-1}\lambda < \infty$. Hence Theorem 4.1 implies, for any $\lambda \in \mathcal{R}^d, \lambda \neq 0$, we have $\sqrt{n}(\lambda'\widehat{\alpha}_n - \lambda'\alpha_0) \Rightarrow \mathcal{N}\left(0, \lambda'\mathcal{I}_*(\alpha_0)^{-1}\lambda\right)$. This implies Proposition 4.1. $\square$

Proof. of Theorem 5.1: The proof basically follows from that of Shen and Shi (2005), except using our definition of joint log-likelihood, our definition of Fisher norm $||\cdot||$, and applying Billingsley's CLT for ergodic stationary martingale difference processes. These modifications are the same as the ones in our proof of Theorem 4.1. Detailed proof is omitted due to the length of the paper, but is available upon request. $\square$

References

Ai, C. and X. Chen, 2003, Efficient Estimation of Models with Conditional Moment Restrictions Containing Unknown Functions, Econometrica, 71, 1795-1843.

Andrews, D., 2005, Higher-order Improvements of the Parametric Bootstrap for Markov Processes, In D.W.K. Andrews and J.H. Stock, eds., Identification and Inference for Econometric Models: A Festschrift in Honor of Thomas J. Rothenberg, Cambridge University Press, pp. 171-215.

Atlason, O., 2008, Generalized Parametric Models. Technical report, Department of Statistics, University of Chicago.

Beare, B., 2008, Copulas and Temporal Dependence. UCSD Economics Working Paper No. 2008-10.

Bickel, P.; Klaassen, C., Ritov, Y. and Wellner, J. A., 1993, Efficient and adaptive estimation for semiparametric models. Johns Hopkins University Press, Baltimore, MD.

Bickel, P. and J. Kwon, 2001, Inference for semiparametric models: some questions and an answer, Statistica Sinica 11, 863-960.

Billingsley, P., 1961, The Lindeberg-Levy theorem for martingales, in Proceedings of the American Mathematical Society, 12, 788-792.

Bouyé, E., Salmon, M., 2008, Dynamic copula quantile regressions and tail area dynamic dependence in forex markets, Manuscript, Financial Econometrics Research Centre, Warwick Business School, UK.

Bradley, R., 2005, Basic Properties of Strong Mixing Conditions. A Survey and Some Open Questions, Probability Surveys 2, 107-144.

Burnham, K. and D. Anderson, 2002, Model selection and multimodel inference. Springer-Verlag, New York

Chan, K. and H. Tong, 2001, Chaos: A statistical perspective, Springer-Verlag, New York.

Chen, J., L. Peng and Y. Zhao, 2009, Empirical likelihood based confidence intervals for copulas. Journal of Multivariate Analysis, 100, 137-151.

Chen, X. (2007) Large Sample Sieve Estimation of Semi-Nonparametric Models, Handbook of Econometrics, Chapter 76, 2007, J.J. Heckman & E.E. Leamer (ed.)

Chen, X. and Y. Fan, 2006, Estimation of copula-based semiparametric time series models, Journal of Econometrics, 130, 307-335.

Chen, X., Y. Fan and V. Tsyrennikov, 2006, Efficient Estimation of Semiparametric Multivariate Copula Models, Journal of the American Statistical Association, 101, 1228-1240.

Chen, X., R. Koenker and Z. Xiao, 2008, Copula-based nonlinear quantile autoregression, forthcoming in Econometrics Journal.

Chen, X. and Shen, X. (1998) Sieve extremum estimates for weakly dependent data. Econometrica

66, 289–314.

Darsow, W., B. Nguyen and E. Olsen, 1992, Copulas and Markov processes, Illinois Journal of Mathematics, 36, 600-642.

Davydov, Y., 1973, Mixing conditions for Markov chains, Theory of Probability and Its Applications, 312-328.

de la Pena, V., R. Ibragimov and S. Sharakhmetov, 2006, Characterizations of joint distributions, copulas, information, dependence and decoupling, with applications to time series, IMS Lecture Notes.

Doukhan, P., Massart, P. and Rio, E. 1995, Invariance principles for absolutely regular empirical processes, Annales de l'Institut Henri Poincaré. Probabilités et Statistiques, 31, 393–427.

Embrechts, P., 2008, Copulas: A personal view. forthcoming in Journal of Risk and Insurance.

Embrechts, P., A. McNeil, and D. Straumann, 2002, Correlation and Dependence Properties in Risk Management: Properties and Pitfalls, in: M. Dempster, eds., Risk Management: Value at Risk and Beyond (Cambridge University Press) 176-223.

Fan, J. and J. Jiang, 2007, Nonparametric inference with generalized likelihood ratio test. Test 16, 409-478.

Fan, J. and Q. Yao, 2003, Nonlinear Time Series: Nonparametric and Parametric Methods. Springer, New York.

Gao, J., 2007, Nonlinear Time Series: Semiparametric and Nonparametric Methods. Chapman & Hall/CRC, London.

Genest, C., M. Gendron and M. Bourdeau-Brien, 2008, The Advent of Copulas in Finance, forthcoming in European Journal of Finance.

Genest, C., K. Ghoudi and L. Rivest, 1995, A Semiparametric Estimation Procedure of Dependence Parameters in Multivariate Families of Distributions, Biometrika, 82, 543-552.

Genest, C. and B. Werker, 2001, Conditions for the asymptotic semiparametric efficiency of an omnibus estimator of dependence parameters in copula models. In C. M. Cuadras, J. Fortiana and J. Rodríguez-Lallena, eds., Distributions With Given Marginals and Statistical Modeling (Springer, NY) 103-112

Granger, C.W.J., 2003, Time series concepts for conditional distributions, Oxford Bulletin of Economics and Statistics, 65, supplement, 689-701.

Hayashi, F., 2000, Econometrics, Princeton University Press, New Jersey.

Ibragimov, R., 2009, Copulas-based characterizations and higher-order Markov processes. Forthcoming in Econometric Theory.

Ibragimov, R., and G. Lentzas, 2008, Copulas and long memory, working paper, Harvard University.

Joe, H., 1997, Multivariate models and dependence concepts. Chapman & Hall/CRC.

Klaassen, C. and J. Wellner, 1997, Efficient Estimation in the Bivariate Normal Copula Model: Normal Margins Are Least Favourable, Bernoulli, Vol. 3, pp. 55-77.

Kosorok, M., 2008, Introduction to Empirical Processes and Semiparametric Inference. Springer, New York.

Li, Q. and J. Racine, 2007, Nonparametric Econometrics. Princeton University Press, New Jersey.

Murphy, S. and A. van der Vaart, 2000, On Profile Likelihood, Journal of the American Statistical Association, 95, 449-465.

Nelsen, Roger B, 2006, An Introduction to Copulas :second edition, Springer Series in Statistics

Patton, A., 2002, Applications of Copula Theory in Financial Econometrics , Journal of Applied Econometrics, 21(2), 147-173.

Patton, A., 2006, Modelling Asymmetric Exchange Rate Dependence, International Economic Review, 47(2), 527-556.

Patton, A., 2008, Copula models for financial time series, forthcoming in Handbook of Financial Time Series, Springer-Verlag.

Robinson, P., 1983, Non-parametric estimation for time series models. Journal of Time Series Analsis 4, 185-208.

Shen, X., 1997, On methods of sieves and penalization, The Annals of Statistics, 25, 2555-2591.

Shen, X., H. Huang and J. Ye, 2004, Inference after model selection, Journal of the American Statistical Association, 99, 751-762.

Shen, X. and J. Shi, 2005. Sieve Likelihood ratio inference on general parameter space. Science in China, 48, 67-78.

Wong, W. H, 1992, On asymptotic effciency in estimation theory, Statist. Sinica 2 47-68.

B Figures and Tables

Table 1: Clayton copula–unknown true $\alpha = 12$, unknown true marginal=$t_{[5]}$: 2-step estimates of copula parameter

	2step-sieve	2step-empirical	2step-para	2step-misN	2step-misEV
Mean	11.370	7.896	12.098	10.709	13.185
Bias	-0.631	-4.104	0.098	-1.291	1.185
Var	3.584	5.656	6.801	14.469	23.827
MSE	3.982	22.5	6.811	16.135	25.231
$\alpha^{MC}_{(2.5,97.5)}$	(8.91,16.52)	(4.35,13.6)	(10.18, 18.42)	(5.65, 20.33)	(7.19, 26.81)

Results are based on 1000 MC replications of estimates using $n = 1000$ time series simulation. 2step-sieve=2step procedure while estimating marginal by sieves in 1st step; 2step-empirical= Chen-Fan; 2step-para=2step procedure while estimating marginal by student t distribution in 1st step; 2step-misN=2step procedure while estimating marginal assuming Normal distribution in 1st step; 2step-misEV=2step procedure while estimating marginal assuming extreme value distribution in 1st step.

Table 2: Clayton copula–unknown true $\alpha = 12$, unknown true marginal=$t_{[5]}$:2-step estimates of marginal quantities

	2step-sieve		2step-empirical		2step-para		2step-misN		2step-misEV	
	$Q_{1/3}$	$Q_{2/3}$	$Q_{1/3}$	$Q_{2/3}$	$Q_{1/3}$	$Q_{2/3}$	$Q_{1/3}$	$Q_{2/3}$	$Q_{1/3}$	$Q_{2/3}$
Mean	0.326	0.664	0.331	0.665	0.332	0.668	0.329	0.642	0.340	0.607
$Bias^2_{10^3}$	0.014	0.039	0.001	0.023	0.003	0.003	0.000	0.786	0.104	4.012
Var_{10^3}	2.151	1.196	28.83	12.08	0.039	0.039	25.763	15.154	16.729	15.208
MSE_{10^3}	2.165	1.235	28.83	12.10	0.041	0.041	25.764	15.941	16.833	19.220

Results are based on 1000 MC replications of estimates using $n = 1000$ time series simulation.

Table 3: Clayton copula–unknown true marginal=$t_{[5]}$: Estimation of copula parameter

		Sieve	Ideal	2step	Para	Mis-N	Mis-EV
$\alpha = 2$	Mean	2.001	2.005	1.920	2.001	2.111	2.907
τ	Bias	0.001	0.005	-0.080	0.001	0.111	0.907
(0.500)	Var	0.020	0.008	0.102	0.012	0.015	0.019
λ	MSE	0.020	0.008	0.109	0.012	0.027	0.841
(0.707)	$\alpha^{MC}_{(2.5,97.5)}$	(1.74,2.28)	(1.84,2.18)	(1.40, 2.63)	(1.78, 2.23)	(1.92,2.37)	(2.67,3.16)
$\alpha = 5$	Mean	4.970	5.006	4.400	5.002	5.379	6.026
	Bias	-0.030	0.006	-0.600	0.002	0.379	1.026
τ	Var	0.139	0.027	1.257	0.044	0.054	0.186
(0.714)	MSE	0.140	0.027	1.617	0.044	0.198	1.239
λ	$\alpha^{MC}_{(2.5,97.5)}$	(4.40, 5.77)	(4.69, 5.33)	(2.71,6.93)	(4.60 , 5.43)	(4.96,5.83)	(5.47,6.50)
(0.871)	$\alpha^{\mathcal{X}^2}_{(0,95)}$	(4.41, 5.45)					
$\alpha = 10$	Mean	9.889	10.01	7.169	10.01	10.77	11.75
	Bias	-0.111	0.01	-2.831	0.01	0.77	1.75
τ	Var	0.483	0.086	4.620	0.143	0.247	0.568
(0.833)	MSE	0.495	0.086	12.63	0.143	0.841	3.637
λ	$\alpha^{MC}_{(2.5,97.5)}$	(8.83 ,11.25)	(9.44,10.6)	(4.02,12.5)	(9.29,10.8)	(9.78,11.7)	(10.4,12.8)
(0.933)	$\alpha^{\mathcal{X}^2}_{(0,95)}$	(8.96, 10.8)					
$\alpha = 12$	Mean	11.85	12.01	7.896	12.00	12.94	14.04
	Bias	-0.149	0.01	-4.104	0.00	0.94	2.04
τ	Var	1.623	0.119	5.656	0.206	0.405	0.960
(0.857)	MSE	1.646	0.120	22.5	0.207	1.285	5.112
λ	$\alpha^{MC}_{(2.5,97.5)}$	(10.6,13.6)	(11.3, 12.7)	(4.35,13.6)	(11.2 , 13.0)	(11.7 , 14.2)	(12.4, 15.3)
(0.944)	$\alpha^{\mathcal{X}^2}_{(0,95)}$	(10.8, 12.9)					

Results are based on 1000 MC replications of estimates using $n = 1000$ time series simulation, except that $\mathcal{X}^2$ inverted confidence intervals are based on 500 MC replications. τ = Kendall's τ, λ = lower tail dependence index. Sieve=Sieve MLE; Ideal=Ideal MLE; 2step=Chen-Fan; Para=correctly specified parametric MLE; Mis-N=parametric MLE using mis-specified normal; Mis-EV=parametric MLE using misspecified extreme value distribution.

Table 4: Gumbel copula–unknown marginal=$t_{[5]}$: Estimation of copula parameter

		Sieve	Ideal	2step	Para	Mis-N	Mis-EV
$\alpha = 2$	Mean	2.003	1.999	1.982	1.996	2.110	1.991
	Bias	0.003	-0.001	-0.018	-0.004	0.110	-0.009
τ	Var	0.007	0.002	0.013	0.004	0.020	0.030
(0.5)	MSE	0.007	0.002	0.014	0.004	0.032	0.030
	$\alpha^{MC}_{(2.5,97.5)}$	(1.85, 2.17)	(1.91,2.10)	(1.78, 2.23)	(1.87, 2.13)	(1.94, 2.55)	(1.69,2.35)
$\alpha = 3.5$	Mean	3.477	3.498	3.352	3.491	3.672	4.028
	Bias	-0.023	-0.002	-0.148	-0.009	0.172	0.528
τ	Var	0.066	0.008	0.130	0.018	0.050	0.245
(0.714)	MSE	0.066	0.008	0.152	0.018	0.0794	0.524
	$\alpha^{MC}_{(2.5,97.5)}$	(3.03, 4.06)	(3.34, 3.68)	(2.76, 4.20)	(3.25, 3.77)	(3.35, 4.26)	(3.06, 4.91)
$\alpha = 6$	Mean	5.778	5.998	5.253	5.994	6.220	7.439
	Bias	-0.222	-0.002	-0.747	-0.006	0.220	1.439
τ	Var	0.315	0.023	0.676	0.062	0.107	1.230
(0.833)	MSE	0.365	0.023	1.235	0.062	0.155	3.302
	$\alpha^{MC}_{(2.5,97.5)}$	(4.72, 6.96)	(5.72, 6.31)	(3.92,7.17)	(5.54, 6.51)	(5.55,6.79)	(5.03, 9.46)
$\alpha = 7$	Mean	6.622	6.997	5.873	6.993	7.250	8.775
	Bias	-0.378	-0.003	-1.127	-0.007	0.250	1.775
τ	Var	0.457	0.032	0.968	0.086	0.142	1.833
(0.857)	MSE	0.600	0.032	2.238	0.086	0.204	4.983
	$\alpha^{MC}_{(2.5,97.5)}$	(5.31, 8.04)	(6.67,7.37)	(4.23, 8.20)	(6.47, 7.59)	(6.50, 7.94)	(5.75, 11.3)

Results are based on 1000 MC replications of estimates using $n = 1000$ time series simulation. $\tau =$ Kendall's τ.

Table 5: Clayton copula–unknown marginal=$t_{[5]}$: Estimation of marginal quantities

		Sieve		2step		Para		Mis-N		Mis-EV	
		$Q_{1/3}$	$Q_{2/3}$	$Q_{1/3}$	$Q_{2/3}$	$Q_{1/3}$	$Q_{2/3}$	$Q_{1/3}$	$Q_{2/3}$	$Q_{1/3}$	$Q_{2/3}$
$\alpha = 2$	Mean	0.327	0.671	0.334	0.667	0.333	0.667	0.349	0.619	0.346	0.595
	$Bias^2_{10^3}$	0.007	0.002	0.015	0.008	0.011	0.011	0.357	2.558	0.258	5.703
$\tau(0.500)$	Var_{10^3}	0.061	0.059	1.282	0.719	0.002	0.002	0.678	1.865	0.503	0.824
$\lambda(0.707)$	MSE_{10^3}	0.067	0.061	1.297	0.727	0.012	0.012	1.035	4.423	0.761	6.527
$\alpha = 5$	Mean	0.326	0.670	0.333	0.667	0.333	0.667	0.337	0.600	0.334	0.590
	$Bias^2_{10^3}$	0.017	0.000	0.012	0.009	0.011	0.011	0.046	4.874	0.019	6.421
$\tau(0.714)$	Var_{10^3}	0.101	0.105	6.018	2.686	0.002	0.002	1.093	3.734	1.293	3.134
$\lambda(0.871)$	MSE_{10^3}	0.117	0.105	6.030	2.695	0.013	0.013	1.139	8.608	1.312	9.554
$\alpha = 10$	Mean	0.323	0.663	0.331	0.666	0.333	0.667	0.356	0.627	0.362	0.633
	$Bias^2_{10^3}$	0.046	0.054	0.002	0.014	0.011	0.011	0.657	1.857	1.014	1.404
$\tau(0.833)$	Var_{10^3}	0.142	0.123	20.93	8.944	0.002	0.002	0.690	2.364	1.345	1.810
$\lambda(0.933)$	MSE_{10^3}	0.188	0.177	20.93	8.958	0.013	0.013	1.347	4.221	2.359	3.214
$\alpha = 12$	Mean	0.322	0.660	0.331	0.665	0.333	0.667	0.363	0.638	0.367	0.642
	$Bias^2_{10^3}$	0.069	0.102	0.001	0.023	0.011	0.011	1.116	1.038	1.389	0.810
$\tau(0.857)$	Var_{10^3}	0.243	0.140	28.83	12.08	0.002	0.002	1.158	2.149	1.632	2.473
$\lambda(0.944)$	MSE_{10^3}	0.312	0.243	28.83	12.10	0.013	0.013	2.274	3.188	3.022	3.283

Results are based on 1000 MC replications of estimates using $n = 1000$ time series simulation. The values of $Bias^2$, variance and MSE have been multiplied by 1000.

Table 6: Gumbel copula–unknown marginal=$t_{[5]}$: Estimation of marginal quantities

		Sieve		2step		Para		Mis-N		Mis-EV	
		$Q_{1/3}$	$Q_{2/3}$	$Q_{1/3}$	$Q_{2/3}$	$Q_{1/3}$	$Q_{2/3}$	$Q_{1/3}$	$Q_{2/3}$	$Q_{1/3}$	$Q_{2/3}$
$\alpha = 2$	Mean	0.329	0.672	0.333	0.666	0.333	0.667	0.363	0.633	0.402	0.650
	$Bias^2_{10^3}$	0.002	0.005	0.007	0.018	0.010	0.010	1.055	1.376	5.236	0.384
$\tau(0.500)$	Var_{10^3}	0.053	0.057	0.755	1.025	0.002	0.002	1.059	1.414	3.459	4.357
	MSE_{10^3}	0.055	0.062	0.762	1.043	0.012	0.012	2.114	2.790	8.694	4.742
$\alpha = 3.5$	Mean	0.328	0.674	0.332	0.665	0.333	0.667	0.407	0.670	0.429	0.648
	$Bias^2_{10^3}$	0.005	0.017	0.005	0.030	0.010	0.010	5.964	0.000	9.694	0.487
$\tau(0.714)$	Var_{10^3}	0.134	0.140	2.353	3.482	0.003	0.003	8.112	4.451	14.32	11.69
	MSE_{10^3}	0.139	0.158	2.358	3.511	0.013	0.013	14.08	4.451	24.01	12.18
$\alpha = 6$	Mean	0.324	0.680	0.330	0.664	0.333	0.667	0.391	0.651	0.394	0.606
	$Bias^2_{10^3}$	0.034	0.100	0.000	0.036	0.011	0.011	3.761	0.375	4.042	4.139
$\tau(0.833)$	Var_{10^3}	0.241	0.239	6.840	10.37	0.003	0.003	24.82	13.31	22.64	18.28
	MSE_{10^3}	0.275	0.339	6.840	10.41	0.014	0.014	28.58	13.69	26.68	22.42
$\alpha = 7$	Mean	0.322	0.683	0.329	0.665	0.333	0.667	0.370	0.630	0.378	0.591
	$Bias^2_{10^3}$	0.066	0.177	0.000	0.029	0.011	0.011	1.593	1.576	2.341	6.219
$\tau(0.857)$	Var_{10^3}	0.285	0.272	9.362	13.79	0.004	0.004	28.87	16.86	24.44	20.39
	MSE_{10^3}	0.352	0.449	9.362	13.82	0.014	0.014	30.46	18.43	26.78	26.61

Results are based on 1000 MC replications of estimates using $n = 1000$ time series simulation. The values of $Bias^2$, variance and MSE have been multiplied by 1000 .

Table 7: Clayton copula–unknown marginal=$t_{[5]}$: Estimation of unknown marginal density

	$\alpha = 2$ ($\tau = 0.50$)	$\alpha = 5$ ($\tau = 0.714$)	$\alpha = 10(\tau = 0.833)$	$\alpha = 12(\tau = 0.857)$
Int$Bias^2_{10^3}$	0.4691	0.4733	0.3650	0.3136
IntVar_{10^3}	0.4000	0.6562	0.6212	0.6341
IntMSE_{10^3}	0.8692	1.1295	0.9862	0.9477

Results are based on 1000 MC replications of estimates using $n = 1000$ time series simulation. Evaluation is based on the common support of 1000 MC simulated data. Int$Bias^2_{10^3}$ = integrated $Bias^2$ multiplied by 1000; IntVar_{10^3} = integrated variance multiplied by 1000; IntMSE_{10^3} = integrated MSE multiplied by 1000.

Table 8: Gumbel copula–unknown marginal=$t_{[5]}$: Estimation of unknown marginal density

	$\alpha = 2$ ($\tau = 0.50$)	$\alpha = 3.5$ ($\tau = 0.714$)	$\alpha = 6$($\tau = 0.833$)	$\alpha = 7$($\tau = 0.857$)
Int$Bias^2_{10^3}$	0.4214	0.4529	0.7851	1.1303
IntVar_{10^3}	0.4047	0.9513	1.5620	1.7284
IntMSE_{10^3}	0.8261	1.4042	2.3471	2.8587

Results are based on 1000 MC replications of estimates using $n = 1000$ time series simulation. Evaluation is based on the common support of 1000 MC simulated data.

Table 9: Clayton copula–unknown marginal=$t_{[5]}$: Estimation of 0.01 conditional quantile function

		Sieve	Ideal	2step	Para	Mis-N	Mis-EV
$\alpha = 5$	Int$Bias^2_{10^3}$	5.409	0.001	80.71	0.004	102.5	482.2
$\tau(0.714)$	IntVar_{10^3}	14.03	3.362	336.1	5.751	85.38	127.3
$\lambda(0.871)$	IntMSE_{10^3}	19.44	3.363	416.8	5.755	187.8	609.5
$\alpha = 10$	Int$Bias^2_{10^3}$	0.951	0.000	288.9	0.000	81.28	201.0
$\tau(0.833)$	IntVar_{10^3}	10.35	1.463	353.2	2.113	41.15	40.26
$\lambda(0.933)$	IntMSE_{10^3}	11.31	1.464	642.1	2.114	122.4	241.3
$\alpha = 12$	Int$Bias^2_{10^3}$	0.689	0.000	227.0	0.000	17.35	80.14
$\tau(0.857)$	IntVar_{10^3}	4.459	0.650	329.6	0.890	13.55	14.89
$\lambda(0.944)$	IntMSE_{10^3}	5.148	0.650	556.6	0.890	30.90	95.03

Results are based on 1000 MC replications of estimates using $n = 1000$ time series simulation. Evaluation is based on the common support of 1000 MC simulated data.

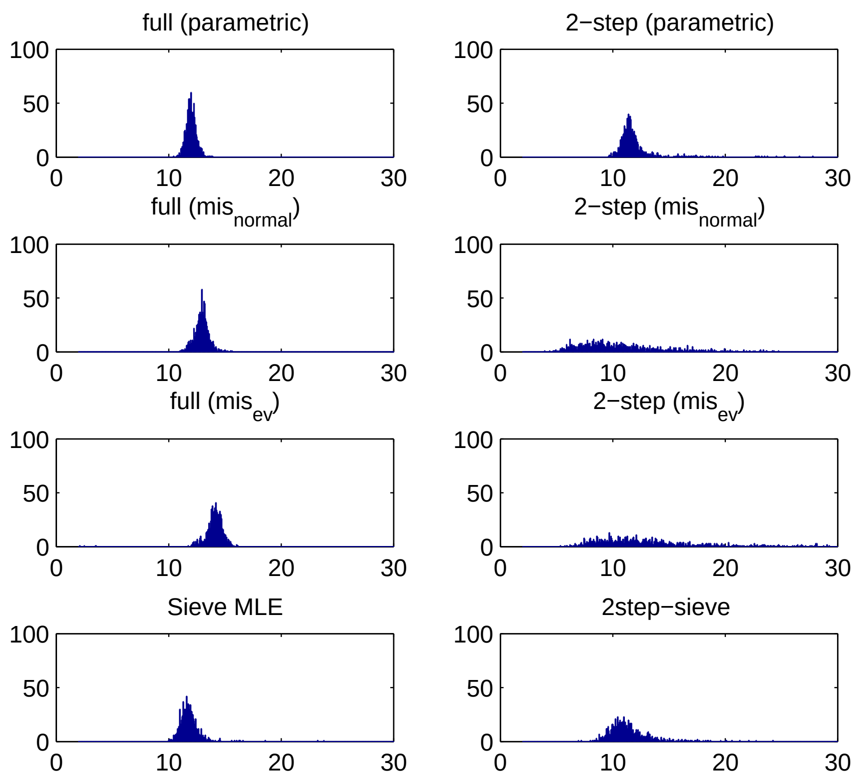
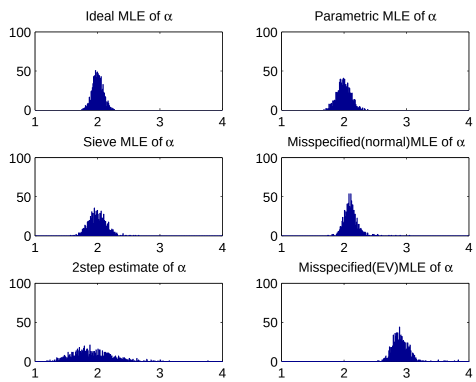
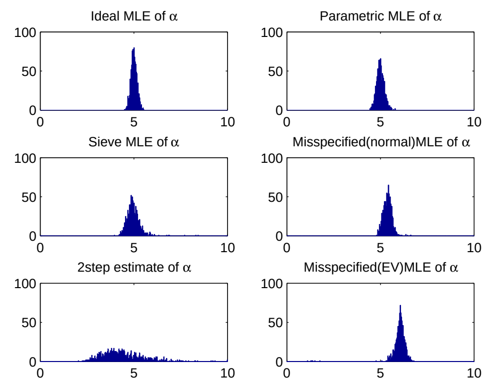


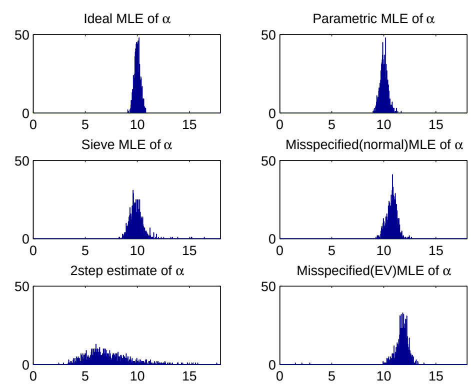
Figure 4: Clayton copula (*true* $\alpha = 12$, marginal=t(5)): Histograms of α estimates, 2-step v.s. full-Likelihood estimators



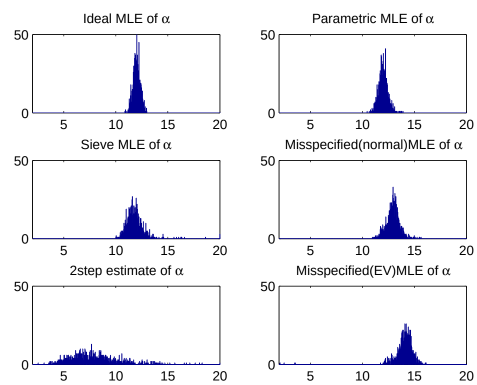
(a)



(b)



(c)



(d)

Figure 5: Clayton copula: Histograms of α estimates (a)true α=2,(b)true α=5, (c)true α=10, (d)true α=12

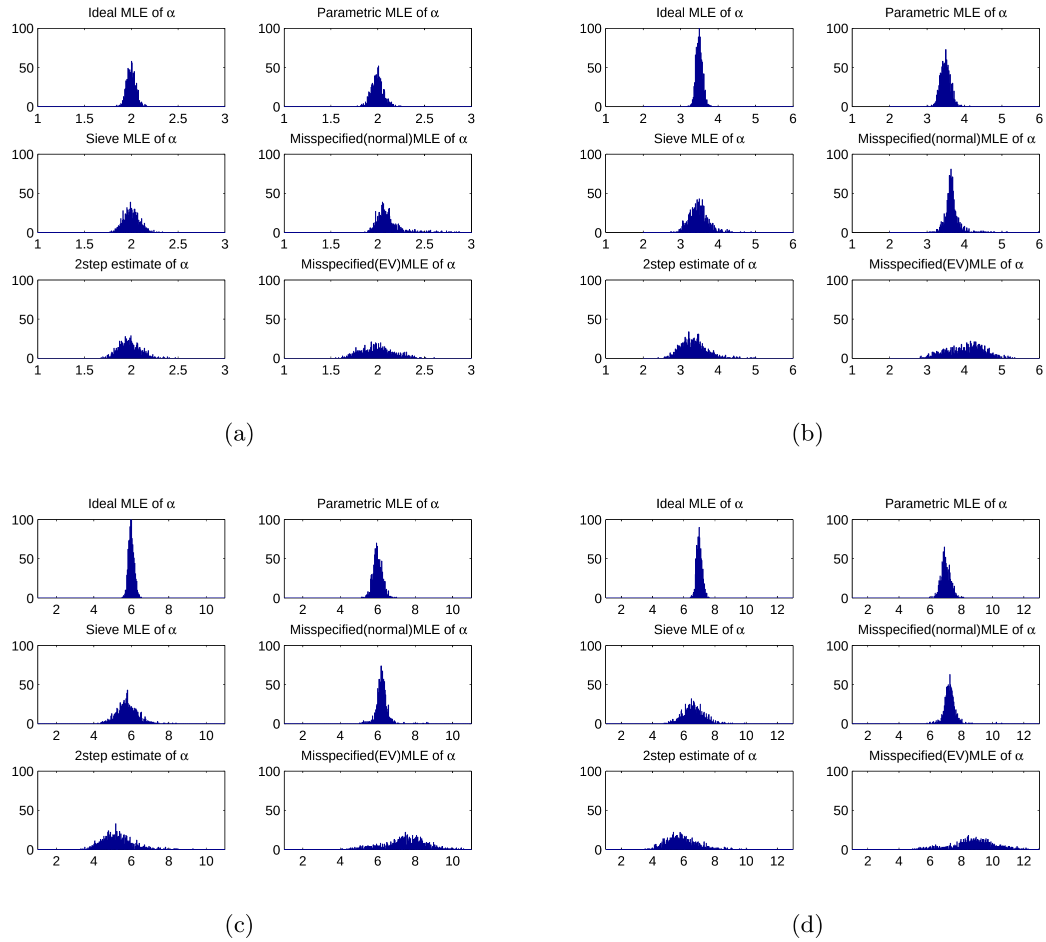


Figure 6: Gumbel copula : Histograms of α estimates (a) true α=2,(b) true α=3.5, (c) true α=6,(d) true α=7

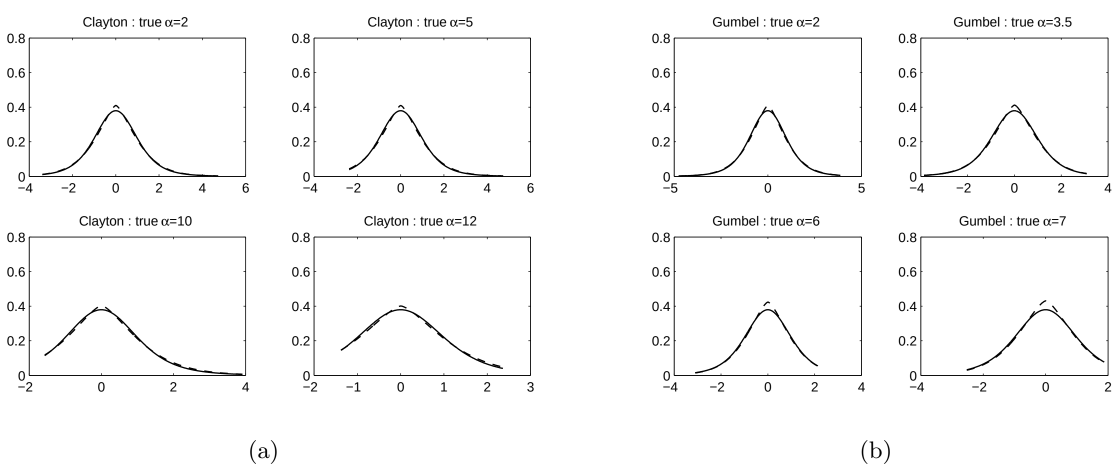
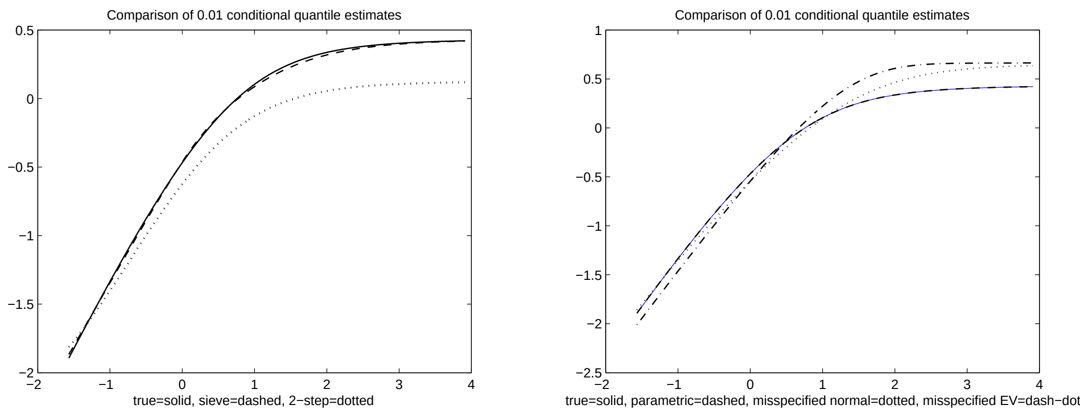
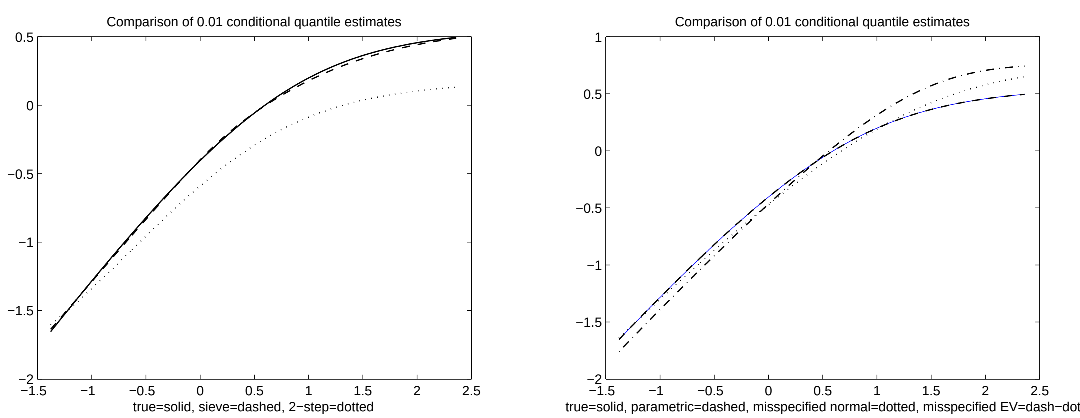


Figure 7: Sieve MLE of marginal density function for (a) Clayton copula, (b) Gumbel copula. True=solid, Sieve MLE=dashed. Evaluation is based on the common support of 1000 MC simulated data.



(a) (b)

Figure 8: Clayton copula ($true\ \alpha = 10$,marginal=t(5)): estimation of 0.01 conditional quantile function. Evaluation is based on the common support of 1000 MC simulated data.



(a) (b)

Figure 9: Clayton copula ($true\ \alpha = 12$,marginal=t(5)) : estimation of 0.01 conditional quantile function. Evaluation is based on the common support of 1000 MC simulated data.