The Korean Journal of Applied Statistics (2016) DOI: http://dx.doi.org/10.5351/KJAS.2016.29.1.267
29(1), 267–277

Residual-based copula parameter estimation

Okyoung Na[a] · Sunghoon Kwon[b,1]

[a]Department of Applied Information Statistics, Kyonggi University;
[b]Department of Applied Statistics, Konkuk University

(Received January 14, 2016; Revised January 16, 2016; Accepted January 16, 2016)

Abstract

This paper considers we consider the estimation of copula parameters based on residuals in stochastic regression models. We prove that a semiparametric estimator using residual empirical distributions is consistent under some conditions and apply the results to the copula-ARMA model. We provide simulation results for illustration.

Keywords: copula function, stochastic regression model, semiparametric estimation, residual empirical distribution, copula-ARMA model, AR approximation

1. 서론

코플라 함수는 다변량 자료에서 변수들 사이의 종속적 구조를 분석하고 모형화하는데 유용한 함수다. 다변량 정규분포를 기반으로 하는 전통적인 기법과 달리 코플라 함수를 이용한 분석 기법은 개별 변수에 대한 모형과 변수들간의 종속적 구조에 대한 모형을 분리해서 모형화할 수 있으므로, 금융 및 경제 자료, 보험 자료, 기후 자료 등을 분석하는데 많이 사용되고 있다. 이와 관련된 자세한 내용은 Brechmann과 Czado (2015), Chan 등 (2009), Embrechts 등 (2002), Jondeau와 Rockinger (2006), Li 등 (2013), Patton (2009), Schölzel과 Friederichs (2008), Sun 등 (2009) 등을 참조하기 바란다.

코플라 함수를 추정하는 방법은 변수들의 결합분포에 대한 가정 정도에 따라 모수적(parametric) 추정방법, 준모수적(semiparametric) 추정방법, 비모수적(nonparametric) 추정방법으로 분류할 수 있고, 모수적 추정방법은 다시 ML(maximum likelihood) 추정방법과 IFM(inference functions for margins) 추정방법으로 나누어진다. IFM 추정방법과 준모수적 추정방법은 개별 변수의 주변분포함수를 추정하는 단계와 코플라 모수를 추정하는 단계로 구성되어 있으므로, 2단계 추정방법이라고 부르기도 한다. Kim 등 (2007)의 연구 결과에 따르면 코플라 모수에 대한 추정량의 효율성, 로버스트성, 계산 편이성, 실용성 등을 모두 고려했을 때 모수적 추정방법보다 준모수적 추정방법이 자료 분석에 사용하기 더 적합하다. 그러므로 본 연구에서는 Genest 등 (1995)가 제시한 준모수적 추정방법을 이용하여 코플라 모수를 추정하고자 한다.

본 연구에서는 d-차원 확률과정 $\{(y_{1t},\ldots,y_{dt})^T\}$에서 변수들간의 종속적 구조에 관심 있으며, 다음과 같은 모형을 고려한다. 각 i에 대하여 확률적 회귀모형(stochastic regression model)

$$y_{it} = \mathbf{x}_{int}^T \boldsymbol{\beta}_{in} + r_{int} + \epsilon_{it}, \quad t = 1,2,\ldots,n,\ i = 1,2,\ldots,d \tag{1.1}$$

[1]Corresponding author: Department of Applied Statistics, Konkuk University, 120 Neungdong-ro, gwangjin-gu, Seoul 05029, Korea. E-mail: shkwon0522@gmail.com

을 만족한다. 여기서 n은 표본의 크기, $\mathbf{x}_{int}$는 관측가능한 q_{in}-차원 확률벡터, $\boldsymbol{\beta}_{in}$는 미지의 q_{in}-차원 벡터, r_{int}는 관측불가능한 확률변수이고, 모두 표본의 크기에 따라 변할 수 있다. ϵ_{it}는 t번째 오차항으로 관측불가능하며 $E(\epsilon_{it}) = 0$, $\mathrm{Var}(\epsilon_{it}) = \sigma_i^2$이고, r_{int}와 달리 표본의 크기와 무관한 확률변수다. 그리고 $(\epsilon_{it}, \ldots, \epsilon_{dt})^T$, $t = 1, 2, \ldots, n$는 서로 독립이고, 동일한 분포를 갖는(IID) 확률벡터이며, $C(F_1(x_1), \ldots, F_d(x_d); \boldsymbol{\theta}_0)$를 결합누적분포함수로 가진다. 여기서 $F_i(x) = P(\epsilon_{it} \le x)$, $i = 1, 2, \ldots, d$이고 코플라 모수 $\boldsymbol{\theta}_0$는 $\boldsymbol{\Theta}$에 속하는 값이다.

식 (1.1)에서 정의한 확률적 회귀모형은 선형회귀모형의 확장으로 $r_{int} = 0$이면 선형회귀모형이 된다. 특히 $r_{int} = 0$이면서 $\mathbf{x}_{int} = (y_{i,t-1}, \ldots, y_{i,t-q_{in}})^T$이면 식 (1.1)은 차수가 q_{in}인 자기회귀형(AR) 모형이 된다. 그러나 식 (1.1)에는 선형회귀모형과는 달리 모형 편의(model bias)를 의미하는 r_{int}가 존재하기 때문에 무한 차수를 갖는 AR(∞) 모형이나 가역성을 만족하는 자기회귀이동평균(ARMA) 모형 등 여러 모형을 표현할 수 있다.

본 연구에서는 시계열 모형 중 식 (1.1)의 응용사례로 copula-ARMA 모형을 다룰 예정이다. copula-ARMA 모형은 SCOMDY(semiparametric copula-based multivariate dynamic) 모형의 한 예로 Chen과 Fan (2006a)이 제시한 방식으로 코플라 모수를 추정할 수도 있다. 그러나 본 연구에서는 copula-ARMA 모형을 AR(q_{in}) 모형으로 근사시켜 잔차를 구하여 코플라 모수를 추정하고, 이 경우 추정량의 성질에 대해 살펴보고자 한다.

논문의 구성은 다음과 같다. 2절에서 IID 표본을 이용하여 코플라 함수를 추정하는 대표적인 방법들에 대해 정리하였다. 3절에서는 식 (1.1)에서 관측불가능한 오차항 대신 잔차를 이용하여 계산한 코플라 모수에 대한 추정량의 성질에 대해 이론적으로 살펴보았으며, 특히 잔차들의 경험적 분포함수를 이용하여 코플라 모수를 추정하는 준모수적 추정량이 일치성을 만족하기 위해 필요한 조건들에 대해 알아보았다. 시계열 분야에서의 응용사례로 copula-ARMA 모형을 고려하였고, 모의실험을 실시하여 얻은 분석 결과를 4절에 보고하였다. 마지막으로 5절에 연구 결과에 대한 결론이 있다.

2. IID 표본에서의 코플라 모수 추정

먼저 IID 표본에서 코플라 함수를 추정하는 대표적인 방법에 대해 간단히 살펴보고자 한다. 이를 위해 잠시 $(\epsilon_{1t}, \ldots, \epsilon_{dt})^T$, $t = 1, 2, \ldots, n$가 관측가능하다고 가정하겠다.

2.1. 모수적 추정방법

모수적 추정방법은 코플라 함수가 속해 있는 함수족 $\{C(u_1, \ldots, u_d; \boldsymbol{\theta}) : \boldsymbol{\theta} \in \boldsymbol{\Theta}\}$과 ϵ_{it}의 주변분포함수 $F_i(x)$가 속해있는 분포족 $\{F_i(x; \boldsymbol{\lambda}_i) : \boldsymbol{\lambda}_i \in \boldsymbol{\Lambda}_i\}$가 알려져있을 때 사용할 수 있는 방법으로 ML 추정방법과 IFM 추정방법이 있다.

먼저 ML 방법은 코플라 모수 $\boldsymbol{\theta}$와 주변분포의 모수 $\boldsymbol{\lambda}_i$, $i = 1, \ldots, d$를 동시에 추정하는 방법으로, 추정량 $(\hat{\boldsymbol{\theta}}_n^{ML}, \hat{\boldsymbol{\lambda}}_{1n}^{ML}, \ldots, \hat{\boldsymbol{\lambda}}_{dn}^{ML})$은 로그우도함수

$$\mathcal{L}_1(\boldsymbol{\theta}, \boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_d) = \frac{1}{n} \sum_{t=1}^{n} \left(\log c\left(F_1(\epsilon_{1t}; \boldsymbol{\lambda}_1), \ldots, F_d(\epsilon_{dt}; \boldsymbol{\lambda}_d); \boldsymbol{\theta}\right) + \sum_{i=1}^{d} \log f_i(\epsilon_{it}; \boldsymbol{\lambda}_i) \right)$$

를 최대로 하는 값이다. 여기서 $c(u_1, \ldots, u_d; \boldsymbol{\theta})$는 코플라 밀도함수이고, $f_i(x; \boldsymbol{\lambda}_i)$는 $F_i(x; \boldsymbol{\lambda}_i)$에 대응되는 확률밀도함수이다.

이와 달리 IFM 추정방법은 2단계로 나누어 모수를 추정한다. 1단계에서 ϵ_{it}, $t = 1, 2, \ldots, n$을 이용하

여 $\boldsymbol{\lambda}_i$의 추정량

$$\hat{\boldsymbol{\lambda}}_{in} = \arg\max_{\boldsymbol{\lambda}_i \in \boldsymbol{\Lambda}_i} \frac{1}{n} \sum_{t=1}^{n} \log f_i(\epsilon_{it}; \boldsymbol{\lambda}_i)$$

을 구한다. 그리고 2단계에서 코플라 모수 $\boldsymbol{\theta}$의 추정량

$$\hat{\boldsymbol{\theta}}_n^{IFM} = \arg\max_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \frac{1}{n} \sum_{t=1}^{n} \log c\left(F_1\left(\epsilon_{1t}; \hat{\boldsymbol{\lambda}}_{1n}\right), \ldots, F_d\left(\epsilon_{dt}; \hat{\boldsymbol{\lambda}}_{dn}\right); \boldsymbol{\theta}\right) \tag{2.1}$$

을 계산한다.

2.2. 준모수적 추정방법

모수적 추정방법과 달리 준모수적 추정방법은 코플라 함수족만 가정할 뿐 주변분포함수에 대해선 특정한 분포족을 가정하지 않는다. 준모수적 추정방법에서는 코플라 모수를 IFM 추정방법처럼 2단계에 걸쳐 추정한다. 1단계에서 $F_i(x)$에 대한 추정량으로 경험적 분포함수

$$F_{in}(x) = \frac{1}{n+1} \sum_{t=1}^{n} I(\epsilon_{it} \le x), \quad -\infty < x < \infty \tag{2.2}$$

를 계산하고, 2단계에서 코플라 모수를 추정할 때 식 (2.1)에서 $F_i(x; \hat{\boldsymbol{\lambda}}_{in})$ 대신 경험적 분포함수 $F_{in}(x)$를 사용한다. 즉, 코플라 모수에 대한 준모수적 추정량은

$$\hat{\boldsymbol{\theta}}_n^{SP} = \arg\max_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \frac{1}{n} \sum_{t=1}^{n} \log c(F_{1n}(\epsilon_{1t}), \ldots, F_{dn}(\epsilon_{dt}); \boldsymbol{\theta}) \tag{2.3}$$

이다.

2.3. 비모수적 추정방법

비모수적 추정방법은 코플라 함수와 주변분포함수에 대해 모두 별다른 가정을 하지 않으며, 경험적 코플라 함수

$$C_n(u_1, \ldots, u_d) = \frac{1}{n} \sum_{t=1}^{n} I(F_{1n}(\epsilon_{1t}) \le u_1, \ldots, F_{dn}(\epsilon_{dt}) \le u_d), \quad 0 \le u_1, \ldots, u_d \le 1$$

가 코플라 함수에 대한 대표적인 비모수적 추정량이다. 여기서 $F_{in}(x)$는 식 (2.2)에서 정의한 경험적 분포함수이다.

3. 잔차를 이용한 코플라 모수 추정

본 연구에서는 앞에서 소개한 방법 중 준모수 추정방법을 이용하여 코플라 모수를 추정하고자 한다. 그러나 모형 (1.1)의 오차항이 관측불가능하므로, 식 (2.2)와 (2.3)에서 오차항 ϵ_{it} 대신 최소제곱법에 의해 얻은 잔차 e_{int}를 사용한다. 코플라 모수를 추정하는 절차는 다음과 같다.

(S1) 자료 $(y_{it}, \mathbf{x}_{int}^T)$, $t = 1, 2, \ldots, n$를 이용하여 $\boldsymbol{\beta}_{in}$에 대한 최소제곱추정값 $\hat{\boldsymbol{\beta}}_{in}$을 구하고, 잔차 $e_{int} = y_{it} - \mathbf{x}_{int}^T \hat{\boldsymbol{\beta}}_{in}$, $t = 1, 2, \ldots, n$를 계산한다.

(S2) 잔차 e_{int}, $t = 1, 2, \ldots, n$를 이용하여 경험적 분포함수

$$\hat{F}_{in}(x) = \frac{1}{n+1} \sum_{t=1}^{n} I(e_{int} \le x), \quad -\infty < x < \infty \tag{3.1}$$

를 계산한다.

(S3) 코플라 모수의 추정값

$$\hat{\boldsymbol{\theta}}_n = \arg\max_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \hat{L}_n(\boldsymbol{\theta}) \tag{3.2}$$

을 구한다. 여기서, $\hat{L}_n(\boldsymbol{\theta}) = (1/n) \sum_{t=1}^{n} \log c(\hat{F}_{1n}(e_{1nt}), \ldots, \hat{F}_{dn}(e_{dnt}); \boldsymbol{\theta})$이다.

이제 코플라 모수의 추정량 $\hat{\boldsymbol{\theta}}_n$이 일치추정량이 되기 위한 조건을 알아보고, 이 추정방법을 코플라를 이용한 대표적인 시계열 모형인 copula-ARMA 모형에 응용해보려고 한다.

3.1. 추정량의 일치성

식 (3.2)에서 정의한 추정량의 일치성을 증명하기 위해 식 (1.1)에서 정의한 모형이 다음에 나열된 조건들을 만족한다고 가정하자.

(A1) 모든 i에 대하여 $\max_{1 \le t \le n} |r_{int}| = o_P(1/\sqrt{n})$이다.

(A2) 모든 i에 대하여 $F_i(x)$는 두번 미분가능하고, $\sup_{x \in R} |F_i'(x)| < \infty$와 $\sup_{x \in R} |F_i''(x)| < \infty$를 만족한다.

(A3) 모든 i에 대하여 다음을 만족하는 $q_{in} \times q_{in}$ 상수행렬 A_{in}이 존재한다:

$$\left\| A_{in}^T \left(\hat{\boldsymbol{\beta}}_{in} - \boldsymbol{\beta}_{in} \right) \right\| = O_P(1), \quad \left\| \sum_{t=1}^{n} A_{in}^{-1} \mathbf{x}_{int} \right\| = O_P\left(\sqrt{n}\right), \quad \sum_{t=1}^{n} \left\| A_{in}^{-1} \mathbf{x}_{int} \right\|^2 = o_P\left(\sqrt{n}\right). \tag{3.3}$$

그리고 $\sum_{t=1}^{n} \|A_{in}^{-1} \mathbf{x}_{int}\| = O_P(b_{in})$을 만족하는 실수열 $\{b_{in}\}$, $\max_{1 \le t \le n} \|A_{in}^{-1} \mathbf{x}_{int}\| = O_P(c_{in})$을 만족하면서 0으로 수렴하는 실수열 $\{c_{in}\}$, $\lim_{n \to \infty} m_{in} = \infty$, $\lim_{n \to \infty} \sqrt{n q_{in}} c_{in} / m_{in} = 0$이고, 모든 양의 실수 a에 대하여 $\lim_{n \to \infty} m_{in}^{q_{in}} \sqrt{n} \exp(-na/(\sqrt{q_{in}} b_{in} + \sqrt{n})) = 0$을 만족하는 정수열 $\{m_{in}\}$이 존재한다.

(A4) 코플라 밀도함수 $c(u_1, \ldots, u_d; \boldsymbol{\theta})$는 $(0,1)^d \times \boldsymbol{\Theta}$에서 연속이고,

$$\begin{aligned} \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} |\log c(u_1, \ldots, u_d; \boldsymbol{\theta})| \le M \left(\min_{1 \le i \le d} u_i \right)^{-c_0} I \left(\min_{1 \le i \le d} u_i \le 1 - \max_{1 \le i \le d} u_i \right) \\ + M \left(1 - \max_{1 \le i \le d} u_i \right)^{-c_0} I \left(\min_{1 \le i \le d} u_i > 1 - \max_{1 \le i \le d} u_i \right) \end{aligned} \tag{3.4}$$

을 만족하는 상수 $M \in (0, \infty)$과 $c_0 \in (0,1)$이 존재한다. 여기서 $\boldsymbol{\Theta}$는 옹골찬 집합이다. 그리고

$$\mathcal{L}(\boldsymbol{\theta}) = \int_0^1 \cdots \int_0^1 \log c(u_1, \ldots, u_d; \boldsymbol{\theta}) c(u_1, \ldots, u_d; \boldsymbol{\theta}_0)\, du_1 \cdots du_d$$

는 $\boldsymbol{\theta} = \boldsymbol{\theta}_0$에서 유일하게 최대값을 갖는다.

조건 (A1)–(A3)는 잔차와 오차항의 차이를 줄여주기 위한 조건으로, 이 조건들이 성립하면 Lee와 Wei (1999)에 기술된 것처럼 $\sup_{x\in R}|\hat{F}_{in}(x) - F_i(x)| = o_P(1)$이 성립하게 된다. 그리고 조건 (A4)는 추정량의 일치성을 증명하기 위해 필요한 조건으로 Genest 등 (1995)과 Chan 등 (2009)가 제시한 가정과 유사하다. 만약 코플라 함수 $C(u_1,\ldots,u_d;\boldsymbol{\theta})$가 (A4)를 만족하면, $\mathcal{L}(\boldsymbol{\theta})$는 연속함수가 되고, $E(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}|\log c(F_1(\epsilon_{1t}),\ldots,F_d(\epsilon_{dt});\boldsymbol{\theta})|) < \infty$이 성립한다. 그러므로 고른 대수의 법칙과 잔차의 성질을 조합하면 $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}|\hat{L}_n(\boldsymbol{\theta}) - \mathcal{L}(\boldsymbol{\theta})| = o_P(1)$이 성립하고, 나아가 $\hat{\boldsymbol{\theta}}_n$이 $\boldsymbol{\theta}_0$로 확률수렴한다.

정리 3.1 조건 (A1)–(A4)를 만족하면, 식 (3.2)에서 정의한 추정량 $\hat{\boldsymbol{\theta}}_n$은 $\boldsymbol{\theta}_0$에 대한 일치추정량이다. 즉,

$$\hat{\boldsymbol{\theta}}_n = \boldsymbol{\theta}_0 + o_P(1) \quad as\ n \to \infty. \tag{3.5}$$

증명: 먼저 $L_n(\boldsymbol{\theta}) = (1/n)\sum_{t=1}^n \log c(F_1(\epsilon_{1t}),\ldots,F_d(\epsilon_{dt});\boldsymbol{\theta})$에 대해 살펴보자. 가정 (A4)가 성립하므로, $\log c(F_1(\epsilon_{1t}),\ldots,F_d(\epsilon_{dt});\boldsymbol{\theta})$, $t = 1,2,\ldots,n$에 고른 대수의 법칙을 적용하면,

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}|L_n(\boldsymbol{\theta}) - \mathcal{L}(\boldsymbol{\theta})| = o_P(1) \tag{3.6}$$

이 성립하고, $\mathcal{L}(\boldsymbol{\theta})$는 연속함수가 된다.

다음으로 $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}|\hat{L}_n(\boldsymbol{\theta}) - L_n(\boldsymbol{\theta})|$에 대해 살펴보자. 가정 (A1)과 (A3)가 성립하므로, 잔차와 오차항의 차이는

$$\max_{1\le t\le n}|e_{int} - \epsilon_{it}| \le \left\|A_{in}^T\left(\hat{\boldsymbol{\beta}}_{in} - \boldsymbol{\beta}_{in}\right)\right\| \max_{1\le t\le n}\left\|A_{in}^{-1}\mathbf{x}_{int}\right\| + \max_{1\le t\le n}|r_{int}| = o_P(1)$$

을 만족한다. 그리고 가정 (A1)–(A3)가 성립하므로, Lee와 Wei (1999)의 정리 2.2와 Glivenko-Cantelli Theorem을 이용하면

$$\begin{aligned}
\left|\hat{F}_{in}(e_{int}) - F_i(\epsilon_{it})\right| &\le \left|\hat{F}_{in}(e_{int}) - F_i(e_{int})\right| + |F_i(e_{int}) - F_i(\epsilon_{it})| \\
&\le \sup_{x\in R}\left|\hat{F}_{in}(x) - F_i(x)\right| + \sup_{x\in R}\left|F_i'(x)\right||e_{int} - \epsilon_{it}| \\
&\le \sup_{x\in R}\frac{1}{n+1}\left|\sum_{t=1}^n I(e_{int}\le x) - F_i(x + \epsilon_{it} - e_{int}) + F_i(x) - I(\epsilon_{it}\le x)\right| \\
&\quad + \sup_{x\in R}\left|\frac{1}{n+1}\sum_{t=1}^n I(\epsilon_{it}\le x) - F_i(x)\right| \\
&\quad + 2\sup_{x\in R}\left|F_i'(x)\right|\max_{1\le t\le n}\left|e_{int} - \epsilon_{it}\right| \\
&= o_P(1)
\end{aligned}$$

이 성립한다. 그러므로 임의의 실수 $\eta \in (0, 1/4)$에 대해

$$\lim_{n\to\infty} P\left(\max_{1\le i\le d}\max_{1\le t\le n}\left|\hat{F}_{in}(e_{int}) - F_i(\epsilon_{it})\right| \le \eta\right) = 1 \tag{3.7}$$

이 성립한다. 임의의 실수 $\delta \in (0, 1/4)$에 대하여 $\{1,2,\ldots,n\}$의 부분집합들 $B_{j\delta}^m$, $B_{j\delta}^M$, $j = 1,\ldots,d$을 다음처럼 정의하자:

$$\begin{aligned}
B_{j\delta}^m &= \left\{t : \min_{1\le i\le d}\hat{F}_{in}(e_{int}) = \hat{F}_{jn}(e_{jnt}), \hat{F}_{jn}(e_{jnt}) \le 1 - \max_{1\le i\le d}\hat{F}_{in}(e_{int}), \hat{F}_{jn}(e_{jnt}) \le \delta\right\}, \\
B_{j\delta}^M &= \left\{t : \max_{1\le i\le d}\hat{F}_{in}(e_{int}) = \hat{F}_{jn}(e_{jnt}), \hat{F}_{jn}(e_{jnt}) > 1 - \min_{1\le i\le d}\hat{F}_{in}(e_{int}), \hat{F}_{jn}(e_{jnt}) \ge 1 - \delta\right\}.
\end{aligned}$$

그러면, 임의의 실수 $\eta \in (0, \delta/2)$에 대하여

$$
\begin{aligned}
&\left|\hat{L}_n(\boldsymbol{\theta}) - L_n(\boldsymbol{\theta})\right| I\left(\max_{1\le i\le d}\max_{1\le t\le n}\left|\hat{F}_{in}(e_{int}) - F_i(\epsilon_{it})\right| \le \eta\right) \\
&\le \frac{1}{n}\sum_{j=1}^{d}\left(\sum_{t\in B_{j\delta}^m} M\left(\hat{F}_{jn}(e_{jnt})\right)^{-c_0} + \sum_{t\in B_{j\delta}^M} M\left(1-\hat{F}_{jn}(e_{jnt})\right)^{-c_0}\right) \\
&\quad + \frac{1}{n}\sum_{j=1}^{d}\sum_{t=1}^{n}\left(M\left(F_j(\epsilon_{jt})\right)^{-c_0} I\left(F_j(\epsilon_{jt}) \le \frac{3\delta}{2}\right) + M\left(1-F_j(\epsilon_{jt})\right)^{-c_0} I\left(F_j(\epsilon_{jt}) \ge 1 - \frac{3\delta}{2}\right)\right) \\
&\quad + \max\left\{\left|\log c(u_1,\ldots,u_d;\boldsymbol{\theta}) - \log c(v_1,\ldots,v_d;\boldsymbol{\theta})\right| : \delta \le u_i \le 1-\delta,\ |u_i - v_i| \le \eta,\ i = 1,\ldots,d\right\}
\end{aligned}
$$

이다. 위 부등식 우변의 세항을 각각 D_{in}, $i = 1, 2, 3$이라고 하면,

$$\lim_{\delta\to 0} D_{1n} \le \lim_{\delta\to 0} 2d\int_0^{\delta} Mx^{-c_0}dx = \lim_{\delta\to 0}\frac{2dM\delta^{1-c_0}}{1-c_0} = 0, \tag{3.8}$$

$$\lim_{\delta\to 0} E(D_{2n}) = \lim_{\delta\to 0}\frac{2dM(3\delta/2)^{1-c_0}}{1-c_0} = 0 \tag{3.9}$$

이 성립하고, $\log c(u_1, \ldots, u_d; \boldsymbol{\theta})$의 연속성에 의해

$$\lim_{\eta\to 0} D_{3n} = 0 \tag{3.10}$$

이다. 따라서 식 (3.7)–(3.10)을 이용하면

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|\hat{L}_n(\boldsymbol{\theta}) - L_n(\boldsymbol{\theta})\right| = o_P(1) \tag{3.11}$$

의 결과를 얻을 수 있다.

앞의 식 (3.6)과 (3.11)의 내용을 종합하면, $\mathcal{L}(\boldsymbol{\theta})$는 $\boldsymbol{\theta} = \boldsymbol{\theta}_0$에서 유일한 최대값을 갖는 연속함수이고,

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|\hat{L}_n(\boldsymbol{\theta}) - \mathcal{L}(\boldsymbol{\theta})\right| \le \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|\hat{L}_n(\boldsymbol{\theta}) - L_n(\boldsymbol{\theta})\right| + \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|L_n(\boldsymbol{\theta}) - \mathcal{L}(\boldsymbol{\theta})\right| = o_P(1)$$

이 성립한다. 따라서 $\hat{\boldsymbol{\theta}}_n = \arg\max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \hat{L}_n(\boldsymbol{\theta})$은 $\boldsymbol{\theta}_0 = \arg\max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \mathcal{L}(\boldsymbol{\theta})$로 확률 수렴한다. $\square$

3.2. Copula-ARMA 모형

이 절에서는 식 (1.1)의 응용으로 $\{y_{it}\}$가 각각 인과성과 가역성을 모두 만족하는 정상 자기회귀이동평균(ARMA) 과정을 따르는 경우를 고려하겠다. 구체적으로 모형을 정의하면 다음과 같다:

$$
\begin{aligned}
y_{it} &= \sum_{k=1}^{P_i}\phi_{ik}y_{i,t-k} + \sum_{k=1}^{Q_i}\pi_{ik}\epsilon_{i,t-k} + \epsilon_{it}, \quad E(\epsilon_{it}) = 0,\ \mathrm{Var}(\epsilon_{it}) = \sigma_i^2,\ i = 1,\ldots,d, \\
&(\epsilon_{1t},\ldots,\epsilon_{dt})^T \overset{IID}{\sim} C(F_1(x_1),\ldots,F_d(x_d);\boldsymbol{\theta}_0)
\end{aligned}
\tag{3.12}
$$

이다. 단, P_i와 Q_i는 0 이상의 정수이고, AR 특성다항식 $\mathcal{A}_i(z) = 1 - \sum_{k=1}^{P_i}\phi_{ik}z^k$와 MA 특성방정식 $\mathcal{B}_i(z) = 1 + \sum_{k=1}^{Q_i}\pi_{ik}z^k$은 공통근을 갖지 않으며, 모든 $|z| \le 1$에 대하여 $\mathcal{A}_i(z) \ne 0$와 $\mathcal{B}_i(z) \ne 0$을 만족한다.

식 (3.12)의 모형은 가역성을 만족하므로 AR(∞) 모형으로 표현할 수 있다. 따라서, 식 (3.12)는

$$y_{it} = \sum_{k=1}^{q_{in}} \beta_{ik} y_{i,t-k} + r_{int} + \epsilon_{it}, \quad r_{int} = \sum_{k=q_{in}+1}^{\infty} \beta_{ik} y_{i,t-k} \tag{3.13}$$

라고 다시 쓸 수 있다.

Brockwell과 Davis (2006)의 Theorem 3.1.2에 의하면, $|\beta_{ik}| \le K_i \rho_i^k$을 만족하는 상수 $K_i \in (0,\infty)$와 $\rho_i \in (0,1)$가 존재하므로,

$$E\left(\max_{1\le t\le n} |r_{int}|\right) \le nE(|y_{i1}|) \sum_{k=q_{in}+1}^{\infty} |\beta_{ik}| \le \frac{nE(|y_{i1}|)K_i \rho_i^{q_{in}+1}}{1-\rho_i}$$

이 성립한다. 그러므로, 만약 $q_{in}/\log n \to \infty$이면,

$$\max_{1\le t\le n} |r_{int}| = O_P\left(n\rho_i^{q_{in}}\right) = o_P\left(\frac{1}{\sqrt{n}}\right) \tag{3.14}$$

이고, 따라서 (A1)이 성립한다.

만약 $E(y_{it}^4) < \infty$이면, $\mathbf{x}_{int} = (y_{i,t-1},\ldots,y_{i,t-q_{in}})^T$는 다음을 만족한다:

$$\left\|\sum_{t=1}^{n} \mathbf{x}_{int}\right\| = \sqrt{\sum_{k=1}^{q_{in}} \left(\sum_{t=1}^{n} y_{i,t-k}\right)^2} = O_P\left(\sqrt{nq_{in}}\right), \tag{3.15}$$

$$\sum_{t=1}^{n} \|\mathbf{x}_{int}\|^2 = \sum_{t=1}^{n}\sum_{k=1}^{q_{in}} y_{i,t-k}^2 = O_P(nq_{in}), \tag{3.16}$$

$$\sum_{t=1}^{n} \|\mathbf{x}_{int}\| = \sum_{t=1}^{n} \sqrt{\sum_{k=1}^{q_{in}} y_{i,t-k}^2} = O_P(n\sqrt{q_{in}}), \tag{3.17}$$

$$\max_{1\le t\le n} \|\mathbf{x}_{int}\| \le \sqrt{q_{in}} \max_{-q_{in}<t<n} |y_{it}| = O_P\left(n^{\frac{1}{4}}\sqrt{q_{in}}\right). \tag{3.18}$$

그리고 $q_{in}^2/n \to 0$이면, $\|(\sum_{t=1}^n \mathbf{x}_{int}\mathbf{x}_{int}^T/n)^{-1}\| = O_P(1)$이고, $\|\sum_{t=1}^n \mathbf{x}_{int}\epsilon_{it}/n\| = O_P(\sqrt{q_{in}/n})$이므로, $\boldsymbol{\beta}_{in} = (\beta_{i1},\ldots,\beta_{iq_{in}})^T$의 최소제곱추정량 $\hat{\boldsymbol{\beta}}_{in}$은

$$\begin{aligned}\left\|\hat{\boldsymbol{\beta}}_{in} - \boldsymbol{\beta}_{in}\right\| &\le \left\|\left(\sum_{t=1}^{n} \frac{\mathbf{x}_{int}\mathbf{x}_{int}^T}{n}\right)^{-1}\right\| \left(\max_{1\le t\le n} |r_{int}| \sum_{t=1}^{n} \frac{\|\mathbf{x}_{int}\|}{n} + \left\|\sum_{t=1}^{n} \frac{\mathbf{x}_{int}\epsilon_{it}}{n}\right\|\right) \\ &= O_P(1) \times \left(o_P\left(\frac{1}{\sqrt{n}}\right) \times O_P\left(\sqrt{q_{in}}\right) + O_P\left(\sqrt{\frac{q_{in}}{n}}\right)\right) = O_P\left(\sqrt{\frac{q_{in}}{n}}\right)\end{aligned} \tag{3.19}$$

을 만족한다. 그러므로 $A_{in} = \sqrt{n/q_{in}} I_{q_{in}}$, $I_{q_{in}}$는 q_{in}-차원 단위행렬, $b_{in} = \sqrt{n} q_{in}$, $c_{in} = q_{in}^2/\sqrt{n}$, $m_{in} = n$이라고 할 때, $q_{in}^5(\log n)^2/n \to 0$이면 (A3)가 성립한다.

식 (3.13)–(3.19)의 내용을 종합하면 다음과 같은 결과를 얻을 수 있다.

보조정리 3.1 식 (3.12)에서 정의한 모형이 $E(\epsilon_{it}^4) < \infty$, (A2)과 (A4)을 만족한다고 가정하자. 만약

$$\frac{q_{in}}{\log n} \to \infty, \qquad \frac{q_{in}^5 (\log n)^2}{n} \to 0 \tag{3.20}$$

이면, 식 (3.5)가 성립한다.

4. 모의실험

여기에서는 3.2절에서 다룬 copula-ARMA 모형에 대한 모의실험을 실시하고, 그 결과를 분석해보기로 한다.

모의실험은 다음 두가지 모형을 고려하였다.

- 모형 1: ARMA(1, 1)+MA(1)

$$
\begin{aligned}
y_{1t} &= 0.5y_{1,t-1} + \epsilon_{1t} + 0.5\epsilon_{1,t-1}, \quad \epsilon_{1t} \overset{IID}{\sim} N(0,1), \\
y_{2t} &= \epsilon_{2t} + 0.5\epsilon_{2,t-1}, \quad \epsilon_{2t} \overset{IID}{\sim} N(0,1), \\
&(\epsilon_{1t}, \epsilon_{2t})^T \overset{IID}{\sim} C(F_1(x_1), F_2(x_2); \boldsymbol{\theta}_0).
\end{aligned}
$$

- 모형2: ARMA(1, 1)+AR(1)

$$
\begin{aligned}
y_{1t} &= 0.5y_{1,t-1} + \epsilon_{1t} + 0.5\epsilon_{1,t-1}, \quad \epsilon_{1t} \overset{IID}{\sim} N(0,1), \\
y_{2t} &= 0.5y_{2,t-1} + \epsilon_{2,t-1}, \quad \epsilon_{2t} \overset{IID}{\sim} N(0,1), \\
&(\epsilon_{1t}, \epsilon_{2t})^T \overset{IID}{\sim} C(F_1(x_1), F_2(x_2); \boldsymbol{\theta}_0)
\end{aligned}
$$

에 대해 실시하였다. 각각의 모형에 Gaussian, Frank, Glayton의 세 가지 코플라 함수를 적용하였고, 코플라 모수 θ_0의 값은 Kendall의 τ값 0.2, 0.5, 0.8에 대응되는 값으로 설정하였다. 즉, Gaussian 코플라의 경우엔 $\theta_0 \in \{0.309, 0.707, 0.951\}$, Frank 코플라의 경우엔 $\theta_0 \in \{1.86, 5.74, 18.20\}$, Clayton 코플라의 경우엔 $\theta_0 \in \{0.5, 2, 8\}$를 고려하였다. 표본의 크기는 $n \in \{100, 200, 400, 800, 1600\}$으로 하였으며 자료 생성시 초기값으로 $y_{i0} = 0$을 사용하고, 초기치의 영향을 줄이기 위해 처음 생성한 100개의 자료는 버리고 101번째 자료부터 사용하였다. 또한 잔차를 구하기 위한 모형은, 보조정리 3.1의 결과를 응용하여 AR(q_n), $q_n = [(\log n)^2]$을 사용하였다. 여기서, $[\,\cdot\,]$은 가우스 함수를 나타낸다.

참고로, 각 코플라 함수의 $\log c(u_1, u_2; \theta)$는

$$
\begin{aligned}
\text{Gaussian}: &\quad -\frac{1}{2}\log\left(1-\theta^2\right) - \frac{2\theta\Phi^{-1}(u_1)\Phi^{-1}(u_2) - \theta^2\left(\Phi^{-1}(u_1)^2 + \Phi^{-1}(u_2)^2\right)}{2(1-\theta^2)}, \\
\text{Frank}: &\quad \log\left(-\theta\left(e^{-\theta}-1\right)\right) - \theta(u_1+u_2) - 2\log\left(\left(e^{-\theta}-1\right) + \left(e^{-\theta u_1}-1\right)\left(e^{-\theta u_2}-1\right)\right), \\
\text{Clayton}: &\quad \log(1+\theta) - (\theta+1)(\log u_1 + \log u_2) - (\theta^{-1}+2)\log\left(u_1^{-\theta} + u_2^{-\theta} - 1\right)
\end{aligned}
$$

이며, 모두 식 (3.4)를 만족한다. 좀 더 자세한 사항은 Chen과 Fan (2006b)의 Section 5를 참조하기 바란다.

모의실험은 총 300번 반복하였으며, 각각의 모의 실험에서 코플라 모수 θ_0를 추정하고 추정값들의 평균과 표준오차를 Tables 4.1, 4.2, 4.3에 정리하였다. 첫째, 세 가지 코플라 모두 자료의 수 n이 증가함에 따라 추정값이 참 값으로 점점 가까워 지는 것을 확인할 수 있고 이는 모든 모형에서 동일하게 관측된다. 이는 본문의 정리 3.1과 보조정리 3.1의 내용을 수치적으로 입증하는 결과이다. 다만 Frank 코플라와 Clayton 코플라의 경우 더 큰 표본의 수를 고려하여 추정량의 평균이 참 값으로 수렴하는 지를 좀더 분명하게 확인할 필요가 있을 것으로 보인다. 하지만 시간상의 문제로 본 연구에서는 생략하였다. 둘째, 세 가지 코플라 모두 참 모수 θ_0를 과소추정하는 것을 확인 할 수 있다. 특히 Frank 코플라와 Clayton 코플라의 경우 표본의 크기가 1000 이하인 경우 그 차이가 매우 심하다는 것을 알 수 있다. 이는 잔차

Table 4.1. Simulation results for the Normal copula. The numbers in parentheses are standard errors

n	Model 1			Model 2		
	$\theta_0 = 0.309$	$\theta_0 = 0.707$	$\theta_0 = 0.951$	$\theta_0 = 0.309$	$\theta_0 = 0.707$	$\theta_0 = 0.951$
100	0.253(0.007)	0.622 (0.005)	0.912 (0.002)	0.237 (0.007)	0.613 (0.005)	0.911 (0.001)
200	0.275(0.004)	0.653 (0.003)	0.931 (0.001)	0.276 (0.004)	0.652 (0.003)	0.931 (0.001)
400	0.289(0.003)	0.677 (0.002)	0.940 (0.290)	0.288 (0.006)	0.675 (0.004)	0.939 (0.001)
800	0.297(0.002)	0.690 (0.001)	0.945 (0.000)	0.298 (0.002)	0.689 (0.001)	0.945 (0.000)
1600	0.299(0.001)	0.695 (0.001)	0.947 (0.000)	0.301 (0.001)	0.695 (0.001)	0.948 (0.000)
3200	0.304(0.001)	0.700 (0.001)	0.949 (0.000)	0.305 (0.001)	0.701 (0.001)	0.949 (0.000)

Table 4.2. Simulation results for the Frank copula. The numbers in parentheses are standard errors

n	Model 1			Model 2		
	$\theta_0 = 1.86$	$\theta_0 = 5.74$	$\theta_0 = 18.20$	$\theta_0 = 1.86$	$\theta_0 = 5.74$	$\theta_0 = 18.20$
100	1.382 (0.043)	4.178 (0.057)	10.935 (0.117)	1.356 (0.038)	4.161 (0.055)	10.848 (0.110)
200	1.558 (0.028)	4.631 (0.035)	12.838 (0.077)	1.586 (0.029)	4.676 (0.037)	12.842 (0.080)
400	1.671 (0.019)	5.014 (0.025)	14.461 (0.057)	1.641 (0.018)	5.042 (0.028)	14.394 (0.058)
800	1.726 (0.014)	5.306 (0.018)	15.642 (0.042)	1.756 (0.013)	5.285 (0.019)	15.674 (0.041)
1600	1.796 (0.009)	5.491 (0.012)	16.543 (0.030)	1.785 (0.009)	5.459 (0.011)	16.562 (0.029)
3200	1.815 (0.007)	5.582 (0.009)	17.137 (0.021)	1.826 (0.006)	5.561 (0.009)	17.186 (0.019)

Table 4.3. Simulation results for the Clayton Copula. The numbers in parentheses are standard errors

n	Model 1			Model 2		
	$\theta_0 = 0.5$	$\theta_0 = 2$	$\theta_0 = 8$	$\theta_0 = 0.5$	$\theta_0 = 2$	$\theta_0 = 8$
100	0.340 (0.008)	0.987 (0.015)	3.932 (0.052)	0.471 (0.012)	1.293 (0.014)	3.966 (0.045)
200	0.367 (0.008)	1.347 (0.006)	4.253 (0.023)	0.518 (0.007)	1.619 (0.036)	3.658 (0.041)
400	0.437 (0.004)	1.565 (0.012)	4.957 (0.051)	0.543 (0.011)	1.596 (0.004)	4.663 (0.015)
800	0.483 (0.003)	1.548 (0.004)	5.769 (0.021)	0.491 (0.007)	1.641 (0.007)	5.696 (0.018)
1600	0.512 (0.003)	1.796 (0.006)	6.438 (0.018)	0.506 (0.001)	1.846 (0.008)	6.284 (0.006)
3200	0.504 (0.002)	1.903 (0.003)	6.948 (0.017)	0.486 (0.001)	1.823 (0.004)	6.822 (0.001)

를 이용한 코플라 모수의 추정이 유한표본에서 무시할 수 없을 만큼의 음의 편이를 가진다는 것을 의미한다. 셋째, 모형간의 차이와 참 모수의 크기에 따른 차이도 거의 없다는 것을 확일 할 수 있다. 다만 Clayton 코플라의 경우, 표본의 크기가 작을 때 모형 1 보다는 모형 2에서 참 값과의 오차가 훨씬 작다는 것을 확인할 수 있다.

5. 결론

본 연구에서는 잔차를 이용하여 오차항의 코플라 함수를 추정하는 문제를 고려하였다. 특히 준모수적 추정량이 일치추정량이 되기 위한 조건을 구하였다. 응용사례로 코플라-AR 모형을 다루었으며, 모의실험을 통해 AR 모형으로의 근사를 통해 얻은 잔차를 이용하여 계산한 추정량의 성질도 살펴보았다.

잔차를 구하는 방법으로, 근사된 AR 모형 대신 적당한 정보량 기준(BIC 혹은 AIC)을 적용하여 각 시계열 자료를 직접 ARMA 모형으로 적합하는 방법을 고려할 수도 있다. 하지만 근사 AR 모형을 사용하는 것이 잔차를 구하는 과정이 좀 더 간단하고, 또한 근사 AR 모형의 차수를 이론적 결과를 바탕으로 쉽게 결정할 수 있다는 장점이 있다. 만약 정보량 기준으로 ARMA 모형의 차수를 정확히 추정할 수 있다면 ARMA 모형을 직접 사용하는 것이 더 좋겠지만, 차수가 정확하지 않은 경우는 근사 AR 모형을

사용하는 경우에 비하여 모형 편이가 더 심할 것으로 예상된다.

본 연구의 자연스러운 후속 연구로 추정량의 극한분포를 연구할 수 있으며, 다양한 시계열 모형으로의 확장도 고려할 수 있다. 특히 모의 실험결과에서, 준모수적 추정법이 코플라 모수를 과소추정한다는 경험적 사실을 확인하였으며 이러한 단점을 해결하기 위한 새로운 방법을 연구하는 것도 중요하다고 판단된다.

References

Brechmann, E. C. and Czado, C. (2015). COPAR-multivariate time series modeling using the copula autoregressive model, *Applied Stochastic Models in Business and Industry*, **31**, 495–514.

Brockwell, P. J. and Davis, R. A. (2006). *Time Series: Theory and Methods*, Second edition, Springer.

Chan, N. H., Chen, J., Chen, X., Fan, Y., and Peng, L. (2009). Statistical inference for multivariate residual copula of GARCH models, *Statistica Sinica*, **19**, 53–70.

Chen, X. and Fan, Y. (2006a). Estimation and model selection of semiparametric copula-based multivariate dynamic models under copula misspecification, *Journal of Econometrics*, **135**, 125–154.

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

Embrechts, P., McNeil, A., and Straumann, D. (2002). Correlation and dependence in risk management: properties and pitfalls, In *Risk Management: Value at Risk and Beyond*, ed. M.A.H. Dempster, Cambridge University Press, Cambridge, 176–223.

Genest, C., Ghoudi, K., and Rivest, L. P. (1995). A semiparametric estimation procedure of dependence parameters in multivariate families of distributions,

Jondeau, E. and Rockinger, M. (2006). The copula-GARCH model of conditional dependencies: An international stock market application, *Journal of International Money and Finance*, **25**, 827–853.*Biometrika*, **82**, 545–552.

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

Lee, S. and Wei, C. Z. (1999). On residual empirical processes of stochastic regression models with applications to time series, *The Annals of Statistics*, **27**, 237–261.

Li, Y., Xie, K., and Hu, B. (2013), Copula-ARMA model for multivariate wind speed and its applications in reliability assessment of generating systems, *Journal of Electrical Engineering and Technology*, **8**, 421–427.

Patton, A. J. (2009). Copula-Based Models for Financial Time Series, In *Handbook of Financial Time Series*, ed. Andersen, T.G., Davis, R.A., Kreiss, J.P. and Mikosch, T., Springer, Berlin Heidelberg, 767–785.

Schölzel, C. and Friederichs, P. (2008). Multivariate non-normally distributed random variables in climate research-introduction to the copula approach, *Nonlinear Processes in Geophysics*, **15**, 761–772.

Sun, W., Rachev, S., Fabozzi, F. J., and Kalev, P. S. (2009). A new approach to modeling co-movement of international equity markets: evidence of unconditional copula-based simulation of tail dependence, *Empirical Economics*, **36**, 201–229.

잔차를 이용한 코플라 모수 추정

나옥경[a] · 권성훈[b,1]

[a]경기대학교 응용정보통계학과, [b]건국대학교 응용통계학과

(2016년 1월 14일 접수, 2016년 1월 16일 수정, 2016년 1월 16일 채택)

요 약

본 연구에서는 잔차를 이용하여 오차항의 코플라 함수를 추정하는 문제를 고려하였다. 확률적 회귀모형을 개별모형으로 갖는 경우, 오차항 대신 잔차들의 경험적 분포함수를 이용하여 구한 코플라 모수에 대한 준모수적 추정량의 성질을 살펴보았으며, 이 추정량이 일치추정량이 되기 위한 조건을 구하였다. 응용사례로 코플라-자기회귀이동평균 모형을 다루었으며, 모의실험을 통해 자기회귀 근사를 통해 얻은 잔차를 이용하여 계산한 추정량의 성질도 살펴보았다.

주요용어: 코플라 함수, 확률적 회귀모형, 준모수적 추정방법, 잔차들의 경험적 분포함수, 코플라-자기회귀이동평균 모형, 자기회귀 근사

[1]교신저자: (05029) 서울 광진구 능동로 120, 건국대학교 응용통계학과. E-mail: shkwon0522@gmail.com