



Comparison of Three Moment-Based Estimation Methods For FGM Copula In the Presence of Outlier

Taheri, B. M. , Jabbari, H. , Amini, M. 

Department of Statistics, Ferdowsi University of Mashhad, Mashhad, Iran.

Corresponding author: H. Jabbari, jabbarinh@um.ac.ir

Received: 5 March 2021 **Revised:** 22 October 2021 **Accepted and Published Online:** 21 November 2021.

Introduction

Using the copulas function is a particular form of introducing variables and their dependency. Also, paying attention to copulas for estimating the dependence parameters has become popular in recent decades. As a semi-parametric technique, Berahimi and Necir (2012), introduced copula moment (CM) and compared it with PMLE and ρ and τ inversion methods, also Kojadinovic and Yan (2010) used three semi-parametric methods based on copula models. Taheri et al. (2018) studied the dependence of bivariate copulas in the presence of outliers. This article used three moment-based estimation methods in the presence of outliers. The moment method and two other estimation methods are related to moment, copula moment (CM) and their mixture.

Material and Methods

Let (X, Y) be a random vector with copula function C and dependence parameter θ . Also, let $(X_1, Y_1), \dots, (X_n, Y_n)$ are a random sample from (X, Y) . We assume that the random vector (X, Y) is in the presence of outliers. In other words, assume that $n - k$ (k is an unknown and a random integer) elements of the random sample have true copula function C_1 and dependence parameter θ and the remind elements have another copula function C_2 and dependence parameter $\beta\theta$, where β is an unknown real value called as a noised parameter. The copula functions C_1 and C_2 can have completely different structures. For estimating θ in the presence of outliers, we may obtain the joint density function of a random sample. A simulation study is used to

select the best estimation method, and the estimators are compared based on MSE. Also, to illustrate the results achieved in a simulation study, we applied a real example for the tourists who visit the "Tomb of Ayub Prophet" (TAP) and the "Imamzadeh Asgari Tomb" (IAT) in North Khorasan, IRAN. Here, we test the dependence between the number of visitors to TAP and IAT at weekends and holidays.

Results and Discussion

For estimating parameters and choosing which value of β is good for FGM copula in the presence of outliers, we test various β and use them in the likelihood function for copulas in the presence of outliers. Using ۱۰۰۰ independent repetitions of the likelihood function for FGM copula in a simulation study suggests that $\beta = 0.1$ is the best value. The dependence parameter is estimated by substituting $\beta = 0.1$ in the likelihood function of copulas in the presence of outliers. The simulation results show that when we use CM and mixed methods in the presence of outliers, the empirical MSEs are reasonable. Also, CM is the best estimator based on MSE.

Conclusion

Taheri et al. (۲۰۱۸) showed that the best methods for estimating the dependence parameter in the presence of outliers are the MLE, PMLE, the inverse of ρ , and the inverse of τ and CM, respectively. In this paper, based on three moment methods, MM, Mixture and CM, the CM method is the best one according to their MSEs. From a practical point of view, it was also concluded that the estimations of the dependence parameter in the presence of outliers do not show a big difference in MLE and CM methods.

Keywords: Copula, Outlier, Moment-based estimate, Dependence parameter, Bivariate analysis.

Mathematics Subject Classification (2010): 47A55, 39B52.



©The Author(s). The Publisher is Iranian Statistical Society.

This is an open access article distributed under the terms and conditions of [\(CC BY-NC 4.0\)](https://creativecommons.org/licenses/by-nc/4.0/)

مقایسه سه روش گشتاور-مبنای برآورد پارامتر مفصل برای خانواده FGM در حضور داده دورافتاده

بی‌بی‌مریم طاهری، هادی جباری‌نوقابی، محمد امینی
گروه آمار، دانشکده علوم ریاضی، دانشگاه فردوسی مشهد

نویسنده مسئول: هادی جباری‌نوقابی، Jabbarinh@um.ac.ir

تاریخ دریافت: ۱۳۹۹/۱۲/۱۵ تاریخ بازنگری: ۱۴۰۰/۰۸/۲۰ تاریخ پذیرش و انتشار: ۱۴۰۰/۰۸/۳۰

چکیده: توجه به تابع مفصل به منظور مدل‌سازی ساختار وابستگی داده‌ها در دهه‌های اخیر بسیار رایج شده است. سه روش گشتاوری، ترکیبی و گشتاور مفصل برای برآورد پارامتر وابستگی تابع مفصل در حضور داده دورافتاده در این مقاله مورد نظر است. هرچند روش گشتاوری یک روش قدیمی است، اما گاهی اوقات این روش منجر به برآورد نامناسبی می‌گردد. در نتیجه، دو روش دیگر برآورد پارامتر بر پایه گشتاوری برای بهبود برآورد پارامتر در نظر گرفته شده‌اند. نتایج مطالعه شبیه‌سازی نشان داد که وقتی از روش گشتاور مفصل و روش ترکیبی برای مفصل در حضور داده دورافتاده استفاده می‌کنیم، میانگین مربع خطای به دست آمده کوچک‌تر است. همچنین روش گشتاور مفصل بهترین برآورد براساس میانگین مربع خطا است. در نهایت، نتایج عددی به دست آمده در یک مثال کاربردی به کار گرفته می‌شود. واژه‌های کلیدی: مفصل، برآورد گشتاور-مبنا، داده دورافتاده، پارامتر وابستگی، تحلیل دومتغیره کد موضوع بندی ریاضی (۲۰۱۰): 47A55, 39B52.

۱ مقدمه

یکی از مهم‌ترین موضوعاتی که در استنباط آماری با آن مواجه هستیم، برآورد پارامتر است. به دلیل خواص و ویژگی‌های بسیار خوب مفصل در مدل‌سازی روابط بین متغیرها از جمله جداسازی ساختار وابستگی از رفتار کناری متغیرها و امکان ساخت مدل‌های متنوع و منعطف در تولید بهتر ساختار وابستگی داده‌ها، اخیراً این تابع به عنوان وسیله‌ای

برای معرفی وابستگی بین متغیرها به شکل خاص و مدل‌سازی روابط بین آن‌ها بسیار مورد استفاده قرار گرفته است (اسکلار، ۱۹۵۹). **طاهری و همکاران (۲۰۱۸)** مفصل دومتغیره در حضور داده‌های دورافتاده را معرفی کردند و پارامتر وابستگی را براساس روش‌های متعدد برآورد کردند. آن‌ها برای برآورد پارامتر وابستگی روش‌های ماکسیمم درستنمایی (ML) و گشتاوری (MM) و مدل‌های خطی تعمیم‌یافته (GLM) و شبه ماکسیمم درستنمایی^۱ (PML) و روش معکوس را به کار بردند. آن‌ها متوجه شدند که بر اساس میانگین مربع خطا (MSE) روش MM بدترین روش است. در عمل، روش‌های نیمه‌پارامتری به منظور غلبه بر محدودیت‌هایی که روش‌های پارامتری دارند، مورد استفاده قرار می‌گیرند. به عنوان روش‌های نیمه‌پارامتری، دو روش گشتاوری مفصل^۲ (CM) که توسط **براهیمی و نصیر (۲۰۱۲)** معرفی شد و روش ترکیبی (Mix) مورد استفاده قرار گرفته‌اند. در روش Mix از برآوردهای به دست آمده به روش ML استفاده کرده و آن‌ها را در روش MM جایگزین می‌کنیم. **براهیمی و نصیر (۲۰۱۲)** روش گشتاور مفصل را معرفی و از آن برای برآورد پارامترهای مفصل دو پارامتری ارشمیدسی استفاده کردند. آن‌ها از CM استفاده کرده و کاربرد آن را در یک مثال شبیه‌سازی از مفصل‌های ارشمیدسی نشان دادند. **ناداراجا (۲۰۱۵)** بسط‌هایی از مفصل را تعریف کرد، چیزی که ما آن را MM می‌نامیم و تمامی این بسط‌ها را برای مفصل FGM مورد استفاده قرار داد. **کجادی‌نوچ و یان (۲۰۱۰)** سه روش نیمه‌پارامتری برآورد را با یکدیگر مقایسه کردند. روش تصویرسازی توسط **تاویل (۲۰۱۱)** برای برآورد پارامتر مفصل مورد استفاده قرار گرفت. **کیم و همکاران (۲۰۰۷)** روش‌های پارامتری و نیمه‌پارامتری را برای برآورد پارامترهای مفصل مقایسه کردند. **جباری و همکاران (۲۰۱۱)** سه روش MM و ML و Mix را برای داده‌های دورافتاده یک متغیره مورد مقایسه قرار دادند.

توابع مفصل کاربردهای فراوانی در آب و هوا، مدیریت ریسک، اقتصاد، هیدرولوژی، صنعت گردشگری و غیره دارند (برای کسب اطلاعات بیشتر، **شمس و حقیقی (۲۰۱۳)**؛ **تانگ و همکاران (۲۰۱۶)**)؛ **ژانگ و همکاران (۲۰۱۳)** را ملاحظه نمایید). **اسکلار (۱۹۵۹)** نخستین بار و به صورت رسمی ارتباط بین توابع توزیع کناری متغیرها را از طریق تابع مفصل معرفی کرد. بعد از او، بسیاری از روش‌های پارامتری و ناپارامتری به منظور مدل‌سازی مفصل‌ها و برآورد پارامترهای وابستگی آن‌ها، توسط نویسندگان مختلفی مورد استفاده قرار گرفته است. **جنست (۱۹۸۷)** و **جنست و همکاران (۱۹۹۵)** یک روش نیمه‌پارامتری را برای برآورد پارامترهای وابستگی در خانواده توزیع‌های چند متغیره تعریف کردند. برای کسب اطلاعات بیشتر در زمینه مدل‌های چند متغیره و مفاهیم وابستگی چند متغیره می‌توان به **جو (۱۹۹۷)** مراجعه نمود. **سوکاهازا (۲۰۰۵)** از تابع توزیع تجربی و مفصل تجربی به منظور برآورد پارامتر وابستگی با روش مینیمم فاصله استفاده کرد. **جو (۲۰۰۵)** از روش برآورد دو مرحله‌ای برای برآورد پارامترها استفاده کرد. در این روش، پارامترهای توزیع‌های کناری برآورد می‌شوند و سپس مقادیر آن‌ها به جای پارامترها قرار گرفته و پارامتر وابستگی برآورد می‌شود. **کیم و همکاران (۲۰۰۷)** از روش‌های پارامتری و نیمه‌پارامتری برای برآورد پارامتر وابستگی استفاده کردند. آن‌ها روش‌های استنباط بر اساس توابع حاشیه‌ای (IFM) و ML را همانند آن چه که **جنست و همکاران (۱۹۹۵)** انجام

¹Pseudo Maximum Likelihood

²Copula Moment

داده بودند، به عنوان روش‌های پارامتری و نیمه‌پارامتری به کار بردند و سپس این روش‌ها را با یکدیگر مقایسه کردند. جنسیت و فاوور (۲۰۰۷) توضیحات بیش‌تری در مورد مدل‌های مفصل بر پایه روش‌های رتبه‌ای ارائه دادند. گریگور (۲۰۱۱) پارامتر مفصل را با استفاده از دو روش ML و کم‌ترین فاصله (MD) در یک مطالعه شبیه‌سازی برآورد کرد. کیم و همکاران (۲۰۱۱) یک کلاس جدید از مفصل‌های دو متغیره و خصوصیات آن‌ها را در نظر گرفتند و سپس پارامتر وابستگی را از طریق مفصل دو متغیره تعمیم یافته برآورد کردند.

هدف اصلی این مقاله برآورد پارامتر وابستگی در حالت دو متغیره در حضور داده دورافتاده با استفاده از مفصل است که با سه روش بر پایه گشتاوری مقایسه می‌شوند. واژه دورافتاده به داده‌های دور از مجموعه داده‌ها یا به بخشی از داده‌ها که توزیع متفاوتی با مابقی آن‌ها دارد، گفته می‌شود. در این مقاله از مفهوم دوم استفاده شده است. هرچند تاکنون روش‌های برآورد زیادی در مورد پارامتر مفصل انجام شده است، لکن تاکنون مطالعات زیادی در مورد توزیع‌های دو متغیره در حضور داده‌های دورافتاده صورت نگرفته است. در این مقاله سه روش پارامتری و نیمه پارامتری برای برآورد پارامتر وابستگی با استفاده از تابع مفصل به کار برده می‌شود. در بخش ۲، توزیع توأم $(X_1, Y_1), \dots, (X_n, Y_n)$ در حضور k داده دورافتاده معرفی می‌شود. در بخش ۳، سه روش MM و Mix و CM در حضور داده‌های دورافتاده دو متغیره معرفی شده و برآوردگر پارامتر وابستگی θ برای تابع مفصل FGM دو متغیره با استفاده از سه روش مذکور به دست آورده می‌شود. سپس کارایی آن‌ها در یک مطالعه شبیه‌سازی مورد مقایسه قرار می‌گیرد. در بخش ۴، مثال کاربردی معرفی شده در طاهری و همکاران (۲۰۱۸) به منظور بررسی عملی روش‌ها، مورد استفاده قرار می‌گیرد و در نهایت نتایج به دست آمده از تمامی روش‌های برآوردیابی با یکدیگر مقایسه می‌گردند.

۲ توزیع توأم نمونه تصادفی

اسکلار (۱۹۵۹) نشان داد که اگر H تابع توزیع بردار تصادفی پیوسته (X, Y) با توابع توزیع کناری G و F باشد، به طور یکتا می‌توان رابطه $H(x, y; \theta) = C(F(x), G(y); \theta)$ را برای هر $x, y \in R$ نوشت که در آن $C : [0, 1]^2 \rightarrow [0, 1]$ یک تابع توزیع دو متغیره با توابع کناری دارای توزیع یکنواخت استاندارد است و θ پارامتر وابستگی است. اگر C عضوی از خانواده مفصل پارامتری مطلقاً پیوسته باشد و توابع توزیع مجهول آن نیز مطلقاً پیوسته باشند، آنگاه رابطه $h(x, y; \theta) = c(F(x), G(y); \theta)f(x)g(y)$ بین تابع چگالی توأم (h) و توابع چگالی کناری $(g$ و $f)$ برقرار است، که در آن c تابع چگالی مفصل C است.

به منظور برآورد پارامتر وابستگی θ در حضور داده‌های دورافتاده باید تابع توزیع توأم نمونه تصادفی مشخص شود. فرض کنید $(X_1, Y_1), \dots, (X_n, Y_n)$ یک نمونه تصادفی از بردار تصادفی (X, Y) در معرض داده‌های دورافتاده باشد و هر دو متغیر تحت تأثیر علل ایجاد داده‌های دورافتاده در نمونه قرار گیرند. برای دسترسی سریع‌تر، مفروضات مسأله به صورت زیر معرفی می‌شوند:

(A1) نمونه تصادفی $(X_{i_1}, Y_{i_1}), \dots, (X_{i_{n-k}}, Y_{i_{n-k}})$ مستقل با تابع توزیع توأم

- $H_1(x, y; \theta) = C_1(F_1(x), G_1(y); \theta)$ ، تابع چگالی توأم $h_1(x, y; \theta)$ و پارامتر وابستگی θ است.
- (A2) بردارهای تصادفی مانده مستقل با توزیع توأم $H_2(x, y; \theta, \beta) = C_2(F_2(x), G_2(y); \theta, \beta)$ ، چگالی $h_2(x, y; \theta, \beta)$ ، پارامتر وابستگی θ و پارامتر مزاحم β هستند.
- (A3) هر زوج از بردارهای $(X_{i_1}, Y_{i_1}), \dots, (X_{i_k}, Y_{i_k})$ و $(X_{i_{k+1}}, Y_{i_{k+1}}), \dots, (X_{i_{n-k}}, Y_{i_{n-k}})$ مستقل از یکدیگر هستند.
- (A4) هر ترکیب (i_1, \dots, i_{n-k}) با احتمال $\frac{1}{\binom{n}{k}}$ از اعداد طبیعی $(1, \dots, n)$ انتخاب شده است که در آن $\binom{n}{k} = \frac{n!}{k!(n-k)!}$.
- (A5) به عنوان یک حالت خاص، فرض می‌کنیم همه توزیع‌های کناری یکنواخت استاندارد باشند.

قضیه ۱. (طاهری و همکاران، ۲۰۱۸). فرض کنید $Z_1 = (X_1, Y_1), Z_2 = (X_2, Y_2), \dots$ برای هر ترکیب (i_1, \dots, i_{n-k}) یک زوج نمونه تصادفی از بردار تصادفی $Z = (X, Y)$ باشد. اگر مفروضات (A1)-(A4) برقرار باشند، آنگاه تابع چگالی توأم نمونه (Z_1, \dots, Z_n) عبارت است از:

$$h(z_1, \dots, z_n; \theta, \beta) = \frac{1}{\binom{n}{k}} \prod_{i=1}^n h_1(z_i; \theta) \sum_{(i_1, \dots, i_k)} \prod_{j=1}^k \frac{h_2(z_{i_j}; \theta, \beta)}{h_1(z_{i_j}; \theta)},$$

که در آن $\sum_{(i_1, \dots, i_k)} = \sum_{i_1=1}^{n-k+1} \dots \sum_{i_k=i_{k-1}+1}^n$

بنا به قضیه ۱، تابع چگالی توأم در حضور داده دورافتاده بر اساس تابع چگالی مفصل به صورت

$$h(z_1, \dots, z_n; \theta, \beta) = \frac{1}{\binom{n}{k}} \prod_{i=1}^n c_1(F_1(x), G_1(y); \theta) f_1(x) g_1(y) \times \sum_{(i_1, \dots, i_k)} \prod_{j=1}^k \frac{c_2(F_2(x), G_2(y); \theta, \beta) f_2(x) g_2(y)}{c_1(F_1(x), G_1(y); \theta) f_1(x) g_1(y)}. \quad (1)$$

است. همچنین، توابع توزیع و چگالی مفصل دومتغیره در حضور k داده دورافتاده با فرض این که توزیع‌های کناری یکنواخت استاندارد باشند، عبارتند از:

$$C(u, v; \theta, \beta) = \frac{k}{n} C_2(u, v; \theta, \beta) + \frac{n-k}{n} C_1(u, v; \theta),$$

$$c(u, v; \theta, \beta) = \frac{k}{n} c_2(u, v; \theta, \beta) + \frac{n-k}{n} c_1(u, v; \theta).$$

۳ روش‌های برآورد بر اساس گشتاور در معرض داده دورافتاده

بر اساس نمونه تصادفی $(X_1, Y_1), \dots, (X_n, Y_n)$ ، برآورد پارامتر وابستگی θ و پارامتر مزاحم β در حضور داده دورافتاده بر پایه تابع مفصل و زوج نمونه تصادفی $(U_1, V_1), \dots, (U_n, V_n)$ مدنظر است. فرض کنید برای $\hat{U}_i = \hat{F}(x_i)$ و $\hat{V}_i = \hat{G}(y_i)$ توابع توزیع تجربی کناری باشند.

۳.۱ روش گشتاوری

ناداراجا (۲۰۱۵) تعمیم‌هایی (حدود بیست تعمیم) را برای مفصل‌های دومتغیره معرفی و تمامی این تعمیم‌ها را برای مفصل FGM بررسی کرد. بر اساس تابع مفصل، گشتاور حاصل ضرب جامعه که توسط طاهری و همکاران (۲۰۱۸) در معرض داده دورافتاده به دست آمده است، برای هر $p, q \geq 1$ به صورت

$$\begin{aligned} E(U^p V^q) &= \int_0^1 \int_0^1 u^p v^q \left(\frac{k}{n} c_{\gamma}(u, v; \theta, \beta) + \frac{n-k}{n} c_1(u, v; \theta) \right) dudv \\ &= \frac{k}{n} \int_0^1 \int_0^1 u^p v^q c_{\gamma}(u, v; \theta, \beta) dudv \\ &\quad + \frac{n-k}{n} \int_0^1 \int_0^1 u^p v^q c_1(u, v; \theta) dudv. \end{aligned}$$

است. به عنوان نمونه، اگر مفصل‌های C_1 و C_{γ} به ترتیب برای بخش‌های غیر دورافتاده و دورافتاده هر دو از نوع خانواده FGM دومتغیره به صورت

$$C_1(u, v; \theta) = uv[1 + \theta(1-u)(1-v)]; \quad -1 \leq \theta \leq 1 \quad (2)$$

$$C_{\gamma}(u, v; \theta, \beta) = uv[1 + \theta\beta(1-u)(1-v)]; \quad -1 \leq \theta\beta \leq 1 \quad (3)$$

باشند، آنگاه گشتاور حاصلضرب در حضور داده دورافتاده عبارت است از

$$E(UV) = \frac{1}{4} + \frac{k}{n} \frac{\theta\beta}{36} + \frac{n-k}{n} \frac{\theta}{36}$$

با فرض معلوم بودن β ، برآورد پارامتر θ به روش گشتاور (MM) به صورت $\hat{\theta}_{MM} = \frac{36uv-9}{b\beta+b}$ است. اگر θ معلوم در نظر گرفته شود، آنگاه برآورد MM پارامتر مزاحم β به صورت $\hat{\beta}_{MM} = \frac{36uv-9-\bar{b}\theta}{b\theta}$ است. در حالتی که هر دو پارامتر θ و β نامعلوم باشند، برآورد MM آنها به صورت

$$\hat{\theta}_{MM} = -9 \frac{nO_{\gamma} - 2O_1^2 \pm \sqrt{\Delta}}{nO_{\gamma}(b\beta + \bar{b})}$$

$$\hat{\beta}_{MM} = \frac{4O_1 - 9O_2 + \hat{\theta}_{MM}\bar{b}(O_1 - O_2)}{\hat{\theta}_{MM}b(O_2 - O_1)}$$

هستند، که در آن‌ها

$$\Delta = \left[\frac{b\beta + \bar{b}}{36} \left(\frac{nO_2}{2} - O_1 \right) \right]^2 - 4 \frac{nO_2}{16 \times 81} (b\beta + \bar{b})^2 \frac{nO_2}{16} - \frac{O_1^2}{9},$$

$$O_1 = \sum_{i=1}^n u_i v_i, \quad O_2 = \sum_{i=1}^n u_i^2 v_i^2, \quad b = \frac{k}{n}, \quad \bar{b} = \frac{n-k}{n}.$$

۳.۲ روش ترکیبی

با ماکسیم کردن تابع درستنمایی ساخته شده از روی توزیع توأم نمونه در (۱)، برآوردهای ML پارامترهای θ و β به دست می‌آید. اگر مفصل‌های C_1 و C_2 به ترتیب برای بخش‌های غیر دورافتاده و دورافتاده هر دو از نوع خانواده FGM دو متغیره به صورت روابط (۲) و (۳) باشند، برآوردگر ML پارامتر β ($\hat{\beta}_{ML}$) از برابر صفر قرار دادن رابطه حاصل می‌شود، که در آن $\frac{\partial \ln S(u_{i_j}, v_{i_j}; \theta, \beta)}{\partial \beta}$

$$S(u_{i_j}, v_{i_j}; \theta, \beta) = \sum_{(i_1, i_2, \dots, i_k)} \prod_{j=1}^k \frac{1 + \beta\theta(1 - 2u_{i_j})(1 - 2v_{i_j})}{1 + \theta(1 - 2u_{i_j})(1 - 2v_{i_j})}.$$

با جایگذاری آن در $\frac{1}{4} \left(1 + \frac{\theta(k\beta + n - k)}{9n} \right)$ ، $\bar{u}\bar{v}$ برآوردگر θ به روش ترکیبی (Mix) برای مفصل FGM دو متغیره در معرض داده دورافتاده به صورت $\hat{\theta}_{Mix} = \frac{36\bar{u}\bar{v} - 9}{b\hat{\beta}_{ML} + \bar{b}}$ خواهد بود.

۳.۳ روش گشتاور مفصل

برای هر عدد طبیعی r و تابع مفصل $C(u, v)$ ، گشتاور مفصل (CM) به صورت

$$M_r(C(\theta)) = E[(C(U, V))^r] = \int_{[0,1]^2} (C(u, v))^r dC(u, v). \quad (4)$$

تعریف می‌شود، که در آن θ پارامتر همبستگی U و V است. چنانچه به جای تابع مفصل $C(u, v)$ در (۴)، مفصل تجربی $C_n(\hat{u}_i, \hat{v}_i)$ جایگذاری شود، که در آن $\hat{u}_i = \frac{1}{n} \sum_{i=1}^n I(X_i \leq x)$ و $\hat{v}_i = \frac{1}{n} \sum_{i=1}^n I(Y_i \leq y)$ ، گشتاور مفصل تجربی مرتبه r به صورت

$$\hat{M}_r = \frac{1}{n} \sum_{i=1}^n (C_n(\hat{u}_i, \hat{v}_i))^r. \quad (5)$$

خواهد شد. با حل معادله $M_r(C(\theta)) = \hat{M}_r$ ، برآورد پارامتر θ به روش CM به دست می‌آید. در حضور داده دورافتاده، روابط (۴) و (۵) به صورت

$$M_r(C(u, v; \theta, \beta)) = \int_{[c, 1]^r} [\bar{b}C_1(u, v; \theta) + bC_2(u, v; \theta, \beta)]^r dC(u, v; \theta, \beta)$$

$$\hat{M}_r = \frac{1}{n} \sum_{i=1}^n (\bar{b}C_{1n}(\hat{u}_i, \hat{v}_i) + bC_{2n}(\hat{u}_i, \hat{v}_i))^r$$

بازنویسی خواهند شد، که در آن C_{1n} و C_{2n} به ترتیب مفصل‌های تجربی متناظر با C_1 و C_2 هستند. اگر مفصل‌های C_1 و C_2 به ترتیب برای بخش‌های غیر دورافتاده و دورافتاده هر دو از نوع خانواده FGM دومتغیره معرفی شده در (۲) و (۳) باشند، آنگاه بنا به روش CM برای $r = 1, 2$ داریم:

$$M_1(C(u, v; \theta, \beta)) = \frac{2\theta(b\beta + \bar{b}) + 9}{36}$$

$$M_2(C(u, v; \theta, \beta)) = \frac{-\theta^2(b\beta + \bar{b})^2}{600} + \frac{\theta(b\beta + \bar{b})}{24} + \frac{1}{9}.$$

اگر β معلوم باشد برآورد θ به روش CM با حل معادله $M_1(C(u, v; \theta, \beta)) = \hat{M}_1$ به صورت $\hat{\theta}_{CM} = \frac{36\hat{M}_1 - 9}{2(b\beta + \bar{b})}$ خواهد بود.

۳.۴ مقایسه کارایی سه روش برآورد

با فرض معلوم بودن پارامتر مزاحم β ، کارایی سه برآورد گشتاوری MM و Mix و CM، در مفصل FGM دومتغیره در معرض داده دورافتاده به ازای مقادیر مختلف n ، β و k در یک مطالعه شبیه‌سازی با ۱۰۰۰ تکرار مستقل و بر اساس ملاک MSE تجربی مورد ارزیابی و مقایسه قرار می‌گیرد. نتایج حاصل در جدول ۱ بیانگر است که: با افزایش n ، مقدار MSE تجربی برآوردهای دو روش برآورد Mix و CM کاهش می‌یابد. با افزایش مقدار پارامتر مزاحم β ، مقدار MSE تجربی در همه روش‌ها به‌ازای هر k و هر n کاهش می‌یابد. اما MSE تجربی برآوردهای CM کاهش بسیار ناچیزی دارد. روش CM کوچک‌ترین MSE را در بین روش‌ها دارد، لذا بهترین روش برآورد است. مقدار MSE تجربی پارامتر وابستگی θ به‌ازای $k = 5$ در روش CM دارای بیش‌ترین نواسانات در n ‌های مختلف است. به ویژه، وقتی $\beta = 0.8$ ، مقدار MSE تجربی برآوردهای θ به روش CM به‌ازای $n = 10$ برابر ۶.۸۶ است که با افزایش حجم نمونه به $n = 100$ به میزان قابل توجهی (یعنی عدد ۰.۸۶) کاهش می‌یابد. از آن‌جا که روش MM عملکرد مناسبی در برآورد پارامتر وابستگی ندارد از جمله حافظ دامنه نیست، MSE تجربی برآوردهای پارامتر θ به روش MM بسیار بزرگ به دست آمده است. از طرفی دیگر، MSE در روش MM با افزایش n ، افزایش زیادی دارد. در روش‌های MM و Mix با افزایش k شاهد افزایش مقدار MSE تجربی هستیم. هرچند روش Mix

جدول ۰۱. مقادیر MSE تجربی برآوردگر θ

n	روش برآورد	$k=5$		$k=2$		$k=1$	
		$\beta=0.3$	$\beta=0.8$	$\beta=0.3$	$\beta=0.8$	$\beta=0.3$	$\beta=0.8$
۱۰	MM	۶۸۰۶۳۰۱	۶۹۷۲۱۹۵	۶۴۶۴۲۲۳	۶۵۲۸۷۰۱	۶۳۵۲۱۵۷	۶۳۸۴۰۷۶
	Mix	۱۹۱۷۱۵۲	۲۶۷۷۶۹۴	۱۰۹۵۱۹۴	۱۲۰۴۶۴۱	۹۳۶۵۲۳۷	۹۷۸۱۴۳
	CM	۶۶۷۲۳۳۶	۶۸۶۰۹۷۶	۲۳۲۷۳۳۶	۲۳۷۷۶۰۱	۱۶۳۱۱۶۱	۱۶۴۹۶۲
۲۰	MM	۲۶۴۰۶۴۸	۲۶۷۳۲۴۷	۲۵۷۲۸۴۲	۲۵۸۵۶۸۹	۲۵۵۰۴۳۶	۲۵۵۶۸۲۷
	Mix	۱۱۹۰۰۷۹	۱۳۴۸۵۸۳	۹۳۶۵۲۳۱	۹۷۸۱۴۳۲	۸۶۹۸۲۲۵	۸۸۸۱۳۴۴
	CM	۱۴۷۷۸۷۸	۱۴۹۷۹۶	۱۷۱۲۰۷۴	۱۷۲۵۱۱۹	۱۳۱۰۷۹۳	۱۳۲۳۴۷
۳۰	MM	۵۸۸۰۶۸	۵۹۲۹۲۷۸	۵۷۷۹۲۷۴	۵۷۹۸۵۲۲	۵۷۴۵۶۶۹	۵۷۵۵۲۶
	Mix	۱۰۳۸۰۷۹۶	۱۱۲۰۱۱۳۴	۸۹۱۲۳۹۲	۹۱۶۷۰۳۹	۸۴۹۱۶۵	۸۶۰۸۸۰۶
	CM	۱۱۲۶۳۶۷	۱۱۲۶۵۸۹	۱۲۶۳۵	۱۲۶۶۶۵۵	۱۰۵۰۷۵۳۲	۱۰۵۱۴۰۴۵
۵۰	MM	۱۶۷۶۹۱۶	۱۷۰۱۵۷۵	۱۶۴۸۳۷۵	۱۶۶۴۶۵۴	۱۵۹۷۶۱۷	۱۵۹۹۲۱۷
	Mix	۹۳۶۵۲۱۲	۹۷۸۱۴۳۷	۸۵۷۳۳۶۴	۸۷۱۶۳۱۴	۸۳۳۱۶۴۹	۸۳۹۹۶۲۵
	CM	۰۹۵۳۶۰۰۸	۰۹۵۵۸۰۹۳	۱۰۲۱۹۳۷	۱۰۲۴۳۵۱	۱۰۰۹۰۴۶	۱۰۰۸۴۱۶
۸۰	MM	۵۱۲۸۳	۵۲۴۶۳	۴۱۰۱۱۶۳	۴۱۰۶۲۸۸	۴۰۹۲۲۰۳	۴۰۹۴۷۶۲
	Mix	۸۸۵۸۱۷۴	۹۰۹۴۳۶۱	۸۳۹۱۰۶۹	۸۴۷۷۱۶۳	۸۲۴۳۶۱۱	۸۲۸۵۳۶۹
	CM	۰۹۰۱۲۵۸	۰۹۰۲۱۹۱	۰۹۵۲۳۱	۰۹۶۱۱۹	۰۹۲۶۰۴	۰۹۲۶۸۹۹
۱۰۰	MM	۶۶۰۹۷۴۸	۶۶۷۴۹۴۶	۶۴۰۶۴۶۷	۶۴۱۲۸۷۲	۶۳۹۵۲۶۷	۶۳۹۸۴۶۶
	Mix	۸۶۹۸۲۴۴	۸۸۸۱۳۷۷	۸۳۳۱۶۴۵	۸۳۹۹۵۸۵	۸۲۱۴۵۸۲	۸۲۴۷۸۶۱
	CM	۰۸۳۰۰۳۴	۰۸۵۶۵۱۳	۰۹۵۱۵۶۰۱	۰۹۵۲۰۸۷۲	۰۸۴۱۳۲۸	۰۸۵۱۶۹۴

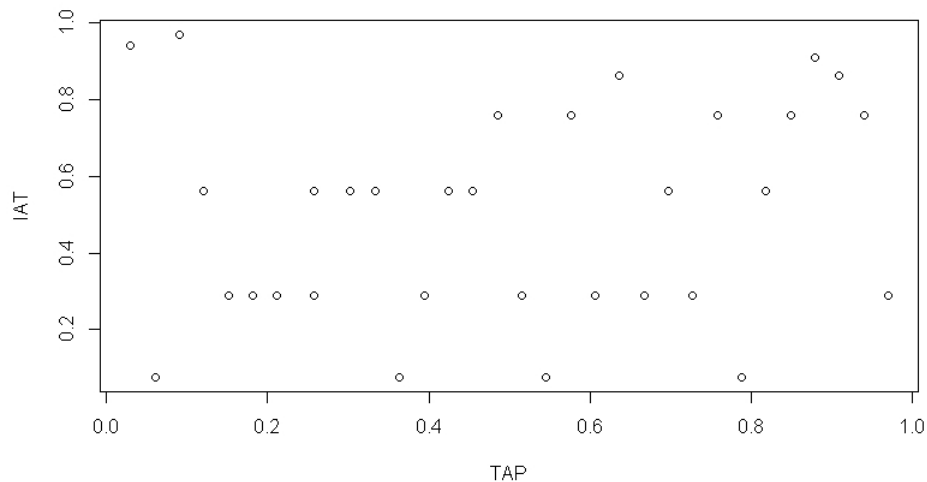
نسبت به MM بهتر است، لیکن بر اساس MSE تجربی، نسبت به روش CM کارایی مناسبی ندارد. مقایسه نتایج شبیه‌سازی مطالعه طاهری و همکاران (۲۰۱۸) با نتایج مطالعه حاضر نشان می‌دهد که گرچه روش برآوردیابی CM با روش‌های MLE و PMLE و معکوس ρ و τ قابل رقابت نیست، ولی روش CM بهتر از روش GLM است.

۴ مثال واقعی

به منظور مقایسه نتایج عملی به دست آمده از مطالعه طاهری و همکاران (۲۰۱۸) با نتایج مطالعه حاضر از داده‌های مثال واقعی مطالعه آنان استفاده می‌کنیم. در این مثال دو جاذبه گردشگری مقبره ایوب نبی (ع) و امامزاده عسگری (ع) از استان خراسان شمالی مد نظر است. دو متغیر مورد نظر، تعداد گردشگران مقبره ایوب نبی (ع) (TAP) و تعداد گردشگران امامزاده عسگری (ع) (IAT) است که در طول روزهای مختلف سال از این دو مکان گردشگری بازدید می‌کنند. مقبره ایوب نبی (ع) یک مکان مذهبی است که در نزدیکی آن یک چشمه معدنی قرار دارد و تمامی اوقات سال شلوغ بوده و بازدیدکنندگان مخصوص خود را دارد. امامزاده عسگری (ع) مکان مذهبی دیگری است که بازدیدکنندگان کمتری با دید مذهبی از آنجا بازدید می‌کنند. بیشتر بازدیدکنندگان امامزاده عسگری (ع) را دانشجویان هنر و معماری تشکیل می‌دهند زیرا از معدود بناهای بجا مانده‌ای است که دارای دو گنبد است.

حالت دورافتاده در هر یک از متغیرهای مورد نظر به هدف بازدید برمی‌گردد. در متغیر اول با دو نوع بازدید کننده مواجه هستیم: مردمی مذهبی که قصدشان بازدید از مقبره ایوب نبی (ع) است و مردمی که تمایل به استفاده از چشمه معدنی دارند. بازدیدکنندگان مقبره ایوب نبی (ع) داده‌های واقعی و افرادی که از چشمه‌ها استفاده می‌کنند، مشاهدات دورافتاده را تشکیل می‌دهند. در متغیر دوم دو نوع دیگر از بازدیدکنندگان را مشاهده می‌کنیم: مردم محلی که به زیارت مقبره بستگان‌شان می‌روند داده‌های واقعی محسوب شده و دانشجویانی که قصد بازدید از جاذبه‌های علمی و هنری این مکان را دارند، مشاهدات دورافتاده را تشکیل می‌دهند. در نتیجه برای هر دو متغیر امکان وجود مشاهدات دورافتاده وجود دارد و اگر تعداد بازدیدکنندگان در یک روز خاص بیشتر از مقدار خاصی باشد، با داده دورافتاده مواجه هستیم. هر چند این عدد ثابت و معلوم نیست و برای دو متغیر یکسان نیست، لیکن تعداد داده‌های دورافتاده و واقعی به روز مورد مطالعه (روز تعطیل یا ایام اول، وسط یا انتهای هفته) بستگی دارند. بنابراین، با توجه به روزهای مورد مطالعه می‌توانیم تعداد مشاهدات دورافتاده را مشخص نماییم.

به منظور مطالعه و مدل‌سازی وابستگی بین تعداد گردشگران بازدیدکننده از این دو مکان، یک نمونه تصادفی به حجم $n = 32$ از روزهای مختلف سال انتخاب شده و تعداد گردشگران بازدیدکننده از دو مکان در این روزها ثبت می‌شوند. شکل ۱ میزان همبستگی بین دو متغیر را توصیف می‌کند. با توجه به روزهای مورد مطالعه، مقدار مناسب برای تعداد داده‌های دورافتاده باید در بازه $[1, 5]$ باشد. در مطالعه طاهری و همکاران (۲۰۱۸) با ذکر دلایل علمی و عملی بهترین مقدار تعداد مشاهدات دورافتاده برابر $k = 1$ و مناسب‌ترین مقدار پارامتر مزاحم برابر $\beta = 0.1$ مشخص شده است. از دید علمی، با توجه به تأثیر مقادیر مختلف پارامتر بتا در تابع درست‌نمایی در حضور داده‌های



شکل ۰۱. نمودار پراکنش برای TAP و IAT

دورافتاده، مشخص شده است که مقدار بهینه پارامتر بتا عدد ۰/۱ است. از دید عملی باید به فلسفه و روزهای جمع‌آوری مشاهدات توجه کرد. با توجه به آن که جفت داده‌های مورد بررسی در بازه زمانی اسفندماه جمع‌آوری شده بودند (داده‌ها مربوط به قبل از شرایط کرونا بوده است)، جفت داده‌های دورافتاده در روزهایی در نظر گرفته شده‌اند که بازدیدکنندگان افزایش یا کاهش چشمگیر داشته‌اند. بنابراین با توجه به رسم ایرانیان برای زیارت بیشتر اهل قبور و مرقد امامزادگان در روزهای پایانی سال یا روزهای اولیه سال جدید مشاهده گردید که در آخرین روز سال ($k = 1$) در تعداد بازدیدکنندگان تفاوت زیادی وجود دارد. از طرف دیگر، روزهای پنج‌شنبه هر هفته نیز تغییراتی در تعداد بازدیدکنندگان وجود داشت ($k = 5$) که پس از بررسی و انجام آزمون مناسب، مشاهده گردید که آخرین روز سال بهترین واحد نمونه‌ای است که در آن مشاهدات دورافتاده در تعداد بازدیدکنندگان رخ داده است. از آنجا که داده‌های مورد استفاده در این مقاله گسسته هستند، با توجه به مقاله **مادسن (۲۰۰۹)** ابتدا از توزیع یکنواخت (۱, ۰) اعداد تصادفی تولید شده، سپس از داده‌های اصلی این مقادیر کسر شده و داده‌های پیوسته جدید تولید شده برای برازش مفصل FGM دومتغیره مورد بررسی قرار گرفته است. با توجه به نتیجه آزمون‌های همبستگی رو و تاو داده‌های پیوسته شده در جدول ۵ وابستگی بین داده‌ها معنی‌دار نیست و به نظر می‌رسد مفصل FGM برای مدل‌سازی آن‌ها مناسب باشد. نتیجه آزمون نیکویی برازش کلموگروف-اسمیرنوف برای داده‌ها پیوسته شده نشان داد که مفصل FGM

جدول ۰۲. آزمون همبستگی متغیرها

k	τ_n	ρ_n	$p - value(\tau)$
۱	۰/۱۵۷	۰/۹۲	۰/۰۱۲۴
۲	۰/۱۴۵	۰/۱۸۹	۰/۰۱۲۵
۵	۰/۱۰۰	۰/۱۱۰	۰/۰۳۱۲

برای مدل‌سازی داده‌ها مناسب است (کمترین p -مقدار به ازای سه مقدار مختلف k برابر ۰/۳۵۷۱ است). برآوردهای پارامتر وابستگی θ در جدول ۴ نشانگر آن است که نتایج حاصل برای داده‌های پیوسته شده و داده‌های اصلی تفاوت چشم‌گیری ندارند (رقم ۵ بعد از اعشار تفاوت دارد) لذا از تکرار آن‌ها و رسم جدول جداگانه خودداری شده است. لازم به ذکر است که انحراف معیارهای برآوردگر θ براساس روش بازنمونه‌گیری با ۱۰۰۰ بار تکرار مستقل فرایند برآوردیابی به دست آمده است.

جدول ۰۳. برآورد θ (انحراف معیار برآوردگر) به روش‌های مختلف

روش	$k = 1$	$k = 2$	$k = 5$
MM	(۴۶،۵۱۲) - ۹،۸۹۰	(۳۰،۵۴۱) - ۹،۵۷۴	(۳۰،۳۰۶) - ۱۰،۴۷۳
Mix	(۲۴،۴۶۴) - ۴،۲۶۴	(۲۳،۹۵۹) - ۴،۵۴۳	(۲۱،۶۵۵) - ۵،۴۹۳۷۱
CM	(۱،۲۶۷) ۰/۲۸۳	(۱،۲۲۸) ۰/۲۴۰	(۱،۰۸۰) ۰/۰۸۴

بنا به نتایج شبیه‌سازی ارایه شده در جدول ۱ و نتایج مطالعه طاهری و همکاران (۲۰۱۸) به‌ازای $n = 30$ و

بر این روش‌های برآوردیابی تمرکز می‌کنیم. با توجه به جدول ۴ به‌ازای همه مقادیر k ، برآورد به روش‌های MLE و CM و PMLE و CM منطقی است. با افزایش k مقدار $\hat{\theta}$ کاهش می‌یابد. انحراف معیار برآوردگر $\hat{\theta}$ در روش CM به طور فزاینده‌ای از سایر روش‌ها کم‌تر است. این خود دلیل خوبی بر ارجحیت این روش نسبت به دو روش دیگر است. به‌ازای $k = 1$ و $\beta = 0.1$ ، برآورد پارامتر وابستگی به روش MLE برابر $\hat{\theta} = 0.195$ و به روش CM برابر $\hat{\theta} = 0.283$ به دست آمده است که تفاوت چندانی ندارند. لازم به ذکر است که از هر دو دیدگاه مطالعات شبیه‌سازی و دلایل عملی، برآوردیابی تنها در حالت $k = 1$ و $\beta = 0.1$ برای مثال کاربردی منطقی است.

بحث و نتیجه‌گیری

طاهری و همکاران (۲۰۱۸) به مطالعه وابستگی مفصل‌های دومتغیره در حضور داده‌های دورافتاده، و برآورد پارامتر وابستگی خانواده مفصل FGM در حضور داده‌های دورافتاده به روش‌های مختلف پارامتری، نیمه‌پارامتری و ناپارامتری پرداختند. روش گشتاوری یکی از روش‌های دیرینه به منظور برآورد پارامتر است. در این مقاله، برای بهبود و مقایسه عملکرد، سه روش برآورد گشتاور-مینا در حضور داده‌های دورافتاده مورد بررسی قرار گرفتند تا بهترین روش برآورد پارامتر وابستگی معرفی شود. نتایج تجربی این مطالعه نشان داد که بهترین روش‌های برآوردیابی پارامتر وابستگی در حضور داده‌های دورافتاده به ترتیب MLE و PMLE و معکوس ρ و τ و CM هستند. از دیدگاه عملی نیز این نتیجه به دست آمد که برآوردهای پارامتر وابستگی در حضور داده‌های دورافتاده به روش‌های MLE و CM تفاوت زیادی ندارند.

تقدیر و تشکر

برخود لازم می‌دانیم از نکات ارزنده داوران، سردبیر و ویراستار محترم مجله که باعث غنای بیشتر و بهتر علمی این مقاله شده است، تشکر و قدردانی نماییم.

مراجع

Taheri, B.M., Jabbari, H. and Amini, M. (2018), Parameter Estimation of Bivariate Distributions in Presence of Outliers: An Application to FGM Copula, *Journal of Computational and Applied Mathematics*, **343**, 155-173.

- Sklar, M. (1959), Fonctions de Repartition An Dimensions et Leurs Marges, *Publ. Inst. Statist. Univ. Paris*, **8**, 229-231.
- Brahimi, B. and Necir, A. (2012), A Semiparametric Estimation of Copula Models Based on the Method of Moments, *Statistical Methodology*, **9**, 467-477.
- Dixit, U. and Nooghabi, M. J. (2011), Efficient Estimation of the Parameters of the Pareto Distribution in the Presence of Outliers, *Communications of the Korean Statistical Society*, **18**, 817-835.
- Nadarajah, S. (2015), Expansions for Bivariate Copulas, *Statistics & Probability Letters*, **100**, 77-84.
- Kim, G., Silvapulle, M. J., and Silvapulle, P. (2007), Comparison of Semiparametric and Parametric Methods for Estimating Copulas, *Computational Statistics & Data Analysis*, **51**, 2836-2850.
- Shams, S., Haghghi, F. K., et al. (2013), A Copula-GARCH Model of Conditional Dependencies: Estimating Tehran Market Stock Exchange Value-at-Risk, *Journal of Statistical and Econometric Methods*, **2**, 39-50.
- Tang, J., Sriboonchitta, S., Ramos, V., and Wong, W.-K. (2016), Modelling Dependence Between Tourism Demand and Exchange Rate Using the Copula-Based GARCH Model, *Current Issues in Tourism*, **19**, 876-894.
- Genest, C. (1987), Frank's Family of Bivariate Distributions, *Biometrika*, **74**, 549-555.
- Genest, C., Ghoudi, K., and Rivest, L.-P. (1995), A Semiparametric Estimation Procedure of Dependence Parameters in Multivariate Families of Distributions, *Biometrika*, **82**, 543-552.
- Joe, H. (1997), *Multivariate Models and Multivariate Dependence Concepts*, CRC Press.

- Tsukahara, H. (2005), Semiparametric Estimation in Copula Models, *Canadian Journal of Statistics*, **33**, 357-375.
- Joe, H. (2005), Asymptotic Efficiency of the Two-Stage Estimation Method for Copula-Based Models, *Journal of multivariate Analysis*, **94**, 401-419.
- Genest, C. and Favre, A.-C. (2007), Everything You Always Wanted to Know About Copula Modeling But Were Afraid to Ask, *Journal of Hydrologic Engineering*, **12**, 347-368.
- Kim, J. M., Sungur, E. A., Choi, T., and Heo, T.-Y. (2011), Generalized Bivariate Copulas and their Properties, *Model Assisted Statistics and Applications*, **6**, 127-136.
- Kojadinovic, I. and Yan, J. (2010), Comparison of three Semiparametric Methods for Estimating Dependence Parameters in Copula Models, *Insurance: Mathematics and Economics*, **47**, 52-63.
- Madsen, L. (2009), Maximum Likelihood Estimation of Regression Parameters with Spatially Dependent Discrete Data, *Journal of Agricultural, Biological, and Environmental Statistics*, **14**, 375-391.
- Touboul, J. (2011), Goodness-of-Fit Tests For Elliptical and Independent Copulas through Projection Pursuit, *Algorithms*, **4**(2), 87-114.
- Weiß, G. (2011), Copula Parameter Estimation by Maximum-Likelihood and Minimum-Distance Estimators: A Simulation Study, *Computational Statistics*, **26**, 31-54.
- Zhang, J., Nakagawa, T. and Kuwano, M. (2013), Copula Modelling Approaches to Jointly Represent Tourists Time Use and Expenditure Behaviour, *13 WCTR, July 15-18, Rio, Brazil*.