وابستگی سویی در مفصل های نامتقارن

شقایق قمری مود[۱] ، محمد امینی[۲] ، ابوالقاسم بزرگ‌نیا[۱]

[۱] گروه آمار، دانشگاه آزاد اسلامی واحد مشهد
[۲] گروه آمار، دانشکده علوم ریاضی، دانشگاه فردوسی مشهد

چکیده: وابستگی سویی بین متغیرها با استفاده از رگرسیون مفصل نامتقارن در سال های اخیر بسیار مورد توجه قرار گرفته است و در زمینه‌های زیادی از جمله تحقیقات پزشکی، امور مالی، بیمه، آب شناسی و ... کاربرد دارد. در این مقاله ابتدا مفهوم وابستگی سویی را تعریف کرده و در روشی چند مرحله ای وابستگی سویی در رفتار توام متغیرها بر اساس رگرسیون مفصل نامتقارن را به دست می‌آوریم. بدین منظور نوع خاصی از مفصل های نامتقارن پارامترهای توزیع را با استفاده از روش‌های نیمه پارامتری که به توزیع حاشیه ای معلوم متغیر ها نیاز ندارد، برآورد می‌کنیم. روش مونت کارلو را زمانی که فرم بسته تابع رگرسیون مفصل متغیرها در دست نیست برای محاسبه اندازه وابستگی سویی به کار خواهیم برد و سپس یک مدل مناسب به داده‌های واقعی برازش داده می‌شود.

واژه‌های کلیدی: مفصل نامتقارن، تابع رگرسیون مفصل، وابستگی سویی، اندازه وابستگی سویی.

۱ مقدمه

مفصل ها به عنوان ابزاری انعطاف پذیر در مدل‌سازی وابستگی داده‌های چند متغیره استفاده می‌شوند. با کمک مفصل‌ها می‌توان وابستگی توزیع توام و توزیع حاشیه‌ای را جداگانه بررسی کرد، همچنین در تابع مفصل متغیرها به وسیله تبدیلی صعودی مثل رتبه به متغیرهای دیگری تبدیل می‌شوند که مفصل متناظر با این متغیرهای جدید مشابه مفصل داده‌های اولیه است. با توجه به این مزیت‌ها اخیرا استفاده از مفصل‌ها برای تعیین وابستگی سویی[۱] در بین متغیرها مورد توجه قرار گرفته است.
منظور از وابستگی سویی جهت احتمالی اثر بین دو متغیر است. مثلا تاثیر بازگشتی (تاثیر یک متغیر بر دیگر متغیرها) یا تاثیر غیربازگشتی (تاثیر دو متغیر بر یکدیگر). سانگور[۲] در سال ۲۰۰۵ بهترین راه تعیین وابستگی سویی بین متغیرها را مطالعه ساختار وابستگی براساس مدل رگرسیون مفصل معرفی کرد.او وابستگی سویی را نه فقط بر اساس رفتار حاشیه‌ای متغیرها بلکه براساس رفتار توام آن ها با استفاده از مفصل نامتقارن تعریف و ویژگی‌های آن را بیان کرد.

۲ مفصل ها و قضیه اسکلار

فرض کنید X و Y دو متغیر تصادفی پیوسته تعریف شده روی فضای احتمال مشابه باشند. F_X و F_Y تابع توزیع یک متغیره X و Y و $F_{X,Y}$ تابع توزیع توام (X,Y) است. مفصل تابع توزیعی است که وابستگی بردار

[۱] Directional dependence
[۲] Sungur

تصادفی (X,Y) را بیان می‌کند به عبارت دیگر یک مفصل دو متغیره تابعی است مثل $C:[0,1]^2 \to [0,1]$ با شرایط زیر:

$$C(u,0)=C(0,v)=0 \;\; , \;\; C(u,1)=u \;\; , \;\; C(1,v)=v \;\; \forall u,v \in [0,1]$$

$$C(u_2,v_2)-C(u_2,v_1)-C(u_1,v_2)+C(u_1,v_1) \geq 0 \;\; \forall u_1<u_2, v_1<v_2$$

براساس قضیه اسکلار (۱۹۵۹) اگر $F(x,y)$ یک توزیع توام با توابع توزیع حاشیه‌ای F_X و F_Y باشد، آن‌گاه مفصل یکتای C وجود دارد به طوری که:

$$F_{XY}(x,y)=C(F_X(x),F_Y(y))$$

و در نتیجه با قرار دادن $u=F_X(x)$ و $v=F_Y(y)$ مفصل C به صورت زیر تعیین می‌شود:

$$C(u,v)=H(F_X^{-1}(u),F_Y^{-1}(v)).$$

۳ مفهوم و اندازه وابستگی سویی

سانگور (۲۰۰۵) دو نوع وابستگی سویی را بر اساس تابع رگرسیون مفصل پیشنهاد داد: جهت وابستگی (بر پایه رفتار حاشیه‌ای متغیرها) و وابستگی سویی (بر پایه رفتار توام متغیرها) که در این مقاله وابستگی سویی را بررسی خواهیم کرد.

تعریف: اگر (U,V) متغیرهای تصادفی یکنواخت با توزیع حاشیه‌ای یکنواخت $(0,1)$ باشند آن‌گاه:

$$C_u(v)=P(V\leq v|U\leq u)=\frac{\partial C(u,v)}{\partial u}$$

$$r_C(u)=r^C_{V|U}=E_C(V|U=u)=1-\int_0^1 C_u(v)dv$$

که در آن $r_C(u)$ را تابع رگرسیون مفصل V نسبت به U نامند.

تعریف: الف) زوج تصادفی (U,V) در رفتار توام وابستگی سویی دارند اگر تابع رگرسیون مفصل $r^C_{U|V}(w) \neq r^C_{V|U}(w)$.

ب) زوج تصادفی (X,Y) در حاشیه‌ای ها وابستگی سویی دارند اگر تابع رگرسیون مفصل $r^C_{V|U}(w)=r^C_{U|V}(w)$ و $r^C_{Y|X}(z)\neq r^C_{X|Y}(z)$.

سانگور (۲۰۰۵) اندازه‌های کلی وابستگی سویی را به صورت زیر پیشنهاد داد:

$$\rho^2_{U\to V}=\frac{Var(r^C_{V|U}(U))}{Var(V)}=12E[(r^C_{V|U}(u))^2]-3 \qquad (1)$$

$$\rho^2_{V\to U}=\frac{Var(r^C_{U|V}(U))}{Var(U)}=12E[(r^C_{U|V}(u))^2]-3 \qquad (2)$$

۴ برآورد وابستگی سویی

لایبسچر[۳] (۲۰۰۸) و دورانت[۴] (۲۰۰۹) فرمول کلی خانواده نسبتا بزرگی از مفصل‌های نامتقارن را به منظور بررسی ساختار وابستگی در داده‌ها معرفی کردند. اعضای نامتقارن این خانواده از ضرب دو مفصل متقارن C_1 و C_2 به دست می آید.

$$C(u,v;\phi) = C_1(u^{\alpha}, v^{\beta})C_2(u^{\bar{\alpha}}, v^{\bar{\beta}}) \qquad (۳)$$

ϕ نشان دهنده مجموعه‌ای از پارامترهای نامتقارن α و β و پارامترهای وابستگی درمفصل‌های متقارن است که $\alpha,\beta \in (0,1)$، $\alpha,\beta \neq \frac{1}{2}$، $\beta+\bar{\beta}=1$ و $\alpha+\bar{\alpha}=1$. جدول ۱ مفصل‌های نامتقارن کها در این مقاله تحلیل می‌شود را نشان می‌دهد.

جدول ۱: مفصل‌های نامتقارن C_1 و C_2 و مفصل نامتقارن متناظر

$C(u,v;\phi)$	$C_1(u^{\alpha},v^{\beta})$	$C_2(u^{\bar{\alpha}},v^{\bar{\beta}})$
M_1	II	$Clyton(\theta)$
M_2	II	$Gumble(\theta)$
M_3	II	$Frank(\theta)$
M_4	II	$AMH(\theta)$
M_5	II	$Plackett(\theta)$
M_6	$Gumble(\theta)$	$Gumble(\theta)$
M_7	$Clyton(\theta_1)$	$Clyton(\theta_2)$
M_8	$Gumble(\theta_1)$	$Gumble(\theta_2)$
M_9	$Frank(\theta_1)$	$Frank(\theta_2)$

۱.۴ برآورد پارامترها

در برآورد پارامترهای مفصل از سه روش زیر عموما استفاده می‌شود:
۱– روش درستنمایی ماکزیمم[۵] (ML)
۲– روش تابع استنباط برای حاشیه‌ای‌ها[۶] (IFM)
۳– روش شبه درستنمایی ماکزیمم[۷] (MPL)
روش‌های ML و IFM روش‌های پارامتریک معلوم هستند و در مقابل مدل انتخاب شده برای حاشیه‌ای‌ها حساس‌اند اما روش MPL به توزیع معلوم حاشیه‌ای‌ها نیاز ندارد، بنابراین، این روش برای استنباط وابستگی سویی در رفتار توام متغیرها پیشنهاد می‌شود.
روش MPL: ابتدا R_i و S_i را که به ترتیب X_iها و Y_iها در n مشاهده وابسته $(u_1,v_1),\cdots(u_n,v_n)$ که $(u_i,v_i)\in[0,1]$ متناظر با مشاهدات وابسته و یکنواخت (u_i,v_i) هستند را نشان می‌دهند، به صورت

$$u_i = \frac{R_i}{n+1}, \quad v_i = \frac{S_i}{n+1}, \quad \forall i = 1,2,\cdots,n$$

[۳] Liebscher
[۴] Duracte
[۵] Maximum likelihood
[۶] Inference functions for margins
[۷] Maximum pseudo-liklihood

محاسبه می‌شوند، سپس با n زوج $\{(u_i, v_i); i = ۱, ۲, \cdots, n\}$ و مفصل متقارن $C(u, v; \phi)$ تابع شبه درستنمایی ماکزیمم به صورت زیر است:

$$l(\phi) = \log \Pi_{i=۱}^{n} c(u_i, v_i; \phi) = \sum_{i=۱}^{n} \log c(u_i, v_i; \phi) \qquad (۴)$$

برآورد شبه درستنمایی ماکزیمم برای پارامترها با ماکزیمم کردن شبه درستنمایی لگاریتمی در رابطه (۴) نسبت به ϕ به دست می‌آید.

۲.۴ محاسبه تابع رگرسیون مفصل و اندازه‌های وابستگی سویی

پیدا کردن فرم بسته تابع رگرسیون مفصل در مفصل‌های نامتقارن جدول ۱ کار ساده‌ای نیست، چرا که شکل تابع توزیع شرطی در مفصل‌های نامتقارن کاملا پیچیده است، بنابراین تابع رگرسیون مفصل واندازه‌های وابستگی سویی را با روش مونت—کارلو گاوسی[۸] که بر پایه نقاط گاوسی تصادفی[۹] است محاسبه می‌کنیم.در واقع محاسبه تابع رگرسیون مفصل واندازه‌های وابستگی سویی محاسبه انتگرال‌های یگانه روی بازه $[۰, ۱]$ و محاسبه انتگرال‌های دوگانه روی مربع واحد $[۰, ۱]^۲$ است. در مرحله اول به کمک نرم‌افزار R دو مجموعه مستقل از نقاط از بازه $[۰, ۱]$ تولید می‌کنیم انتگرال‌های یگانه و دوگانه بالا را در نقاط تولید شده با میانگین‌های تجربی تقریب می‌زنیم یعنی:

$$r^C_{U|V}(v) \approx \tilde{r}^C_{U|V}(v) = ۱ - \frac{۱}{S} \sum_{s=۱}^{S} C_v(u_s) \qquad (۵)$$

$$\rho^۲_{U \to V} \approx \tilde{\rho}^۲_{U \to V} = \frac{۱۲}{S} (\tilde{r}^C_{U|V}(u_s))^۲ - ۳ \qquad (۶)$$

که در آن $(u_۱, ..., u_s)$ و $(v_۱, ..., v_s)$ نقاط تصادفی از بازه $[۰.۱]$ است.

۳.۴ انتخاب مدل مفصل

یکی از مشکلات مهم انتخاب مفصلی است که به طور مناسب‌تری با داده ها منطبق باشد. مفصل نامتقارنی که بیشترین مقدار اندازه وابستگی سویی را دارد در نظر می‌گیریم و سپس آزمون نیکویی برازش را انجام می‌دهیم. چندین روش بر اساس رتبه برای آزمون GOF وجود دارد: دو آزمون بر پایه تبدیل کندال و آزمونی بر پایه مفصل تجربی که در آن از آماره میسز کرامر—وان[۱۰] استفاده می‌شود.

۵ تشخیص و تعیین وابستگی سویی

با داشتن تابع رگرسیون مفصل تقریبی می‌توان اندازه وابستگی سویی را تعیین کرد. اگر نمودار $\tilde{r}_{U|V}(w)$ را در مقابل $\tilde{r}_{V|U}(w)$ را رسم و یک خط ۴۵ درجه را روی صفحه نمودار در نظر بگیریم، آن‌گاه با بررسی انحراف

---

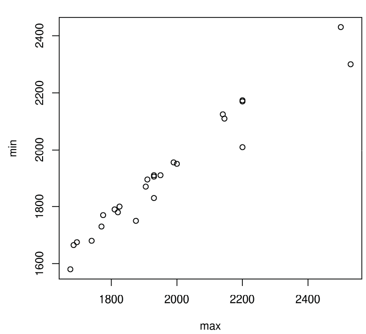
[۸] Quasi Monte Carlo
[۹] Quasi random
[۱۰] Cramer-Von Mises

نمودار از این خط می‌توان درباره وابستگی سویی نظر داد.این نمودار را CR می‌نامیم. وقتی به طور تجربی وابستگی در داده‌ها را تشخیص دادیم، آن‌گاه اندازه وابستگی سویی را به دست می آوریم و می‌توان روش بوت استر پارامتری را برای به دست آوردن فواصل اطمینان به کار برد.

۶ مثال

داده‌های این مثال را از صفحه اقتصادی روزنامه خراسان جمع آوری کرده‌ایم. در هر روز از شش ماهه دوم سال ۱۳۹۱ نرخ دلار را بر حسب ریال در نظر گرفته و بیشترین و کمترین مقدار نرخ دلار در هر هفته را بررسی کرده‌ایم که شامل ۲۵ مشاهده برای بیشترین مقدار و ۲۵ مشاهده برای کمترین مقدار نرخ دلار است. نمودار پراکنش مشاهدات و رتبه های نرمال شده آن‌ها نشان دهنده وابستگی در بین این دو مجموعه داده است. ضریب همبستگی داده شده در جدول ۲ نیز تاییدی بر این مطلب است.

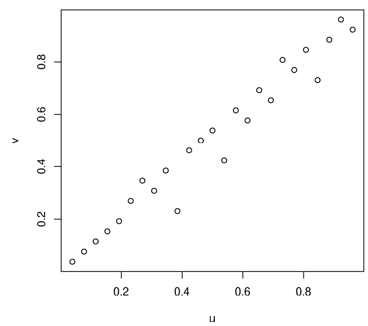
برای مدل‌سازی وابستگی سویی بین این دو گروه داده برخی مفصل‌های نامتقارن جدول ۱ را بررسی می‌کنیم. ابتدا با استفاده از ضریب همبستگی به دست آمده و ارتباط بین ضریب همبستگی و پارامتر در مفصل‌ها، در نرم‌افزار R پارامتر θ را برای مفصل‌های انتخاب شده برآورد می‌کنیم (جدول ۳).



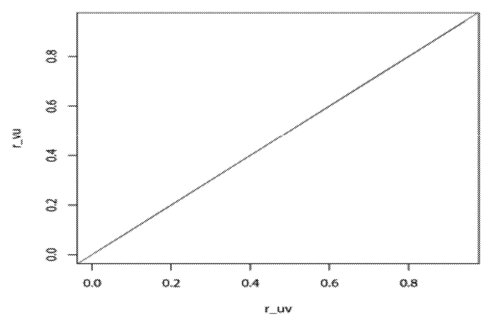
شکل ۱: نمودار پراکنش مشاهدات

در نرم‌افزار R پارامترها را باروش MPL برآورد کرده (جدول ۴) و سپس تابع رگرسیون مفصل را تقریب می‌زنیم و از این تقریب اندازه وابستگی سویی را به طور تقریبی به دست می آوریم (جدول ۵).

در بین این مدل ها مدل $M_۷$ (حاصل ضرب دو مفصل کلایتون) بیشرین مقدار $\tilde{\rho}^{۲}_{U\to V}$ و $\tilde{\rho}^{۲}_{V\to U}$ و مدل $M_۱$ (حاصل ضرب مفصل استقلال در مفصل کلایتون) کمترین مقدار اندازه وابستگی سویی را دارد. پس مدل $M_۷$ را به عنوان مدل نهایی برای تجزیه و تحلیل های بعدی در نظر می‌گیریم.
نمودار CR را رسم کرده، انحراف نسبت به خط ۴۵ درجه به علت اختلاف کم بین $\tilde{\rho}^{۲}_{U\to V}$ و $\tilde{\rho}^{۲}_{V\to U}$ در این نمودار واضح نیست اما وابستگی سویی بین بیشترین و کمترین مقدار نرخ دلار فقط زمانی وجود دارد که هر دو متغیر خیلی بزرگ یا خیلی کوچک باشند.



شکل ۲: نمودار پراکنش رتبه‌های نرمال شده مشاهدات



شکل ۳: نمودار CR برای مدل M_7

جدول ۲: مقدار ضریب همبستگی

ضریب همبستگی	مقدار
ضریب پیرسون	۰.۹۷۲۷
ضریب کندال	۰.۹۰۷۲
ضریب اسپیرمن	۰.۹۷۲۳

جدول ۳: برآورد پارامتر θ

$C(u,v;\phi)$	ϕ
M_1	۱۹.۵۰
M_3	۴۱.۳۹۰
M_4	۰.۱۵۱
M_7	۱۹.۵۰

جدول ۴: برآورد پارامترهای مفصل‌های نامتقارن

$C(u,v;\phi)$	ϕ
M_1	$(4.033981\times 10^{-2}, 2.344288\times 10^{-2}, 1.971005\times 10)$
M_3	$(0.1000, 0.1000, 28.000)$
M_4	$(1.348770\times 10^{-11}, 4.679866\times 10^{-10}, 9.997133\times 10^{-1})$
M_7	$(0.16789976, 0.15694482, 30.1955322, 27.343048)$

جدول ۵: اندازه وابستگی سویی در مفصل‌های نامتقارن

$C(u,v;\phi)$	$\tilde{\rho}^2_{V\to U}$	$\tilde{\rho}^2_{U\to V}$
M_1	2.0272229×10^{-4}	-1.128175×10^{-5}
M_3	0.7258467	0.7258467
M_4	2.653490×10^{-1}	2.653490×10^{-1}
M_7	0.9456437	0.9454558

بحث و نتیجه‌گیری

در این مقاله وابستگی سویی دو متغیر را که حاصل رفتار توام آن‌هاست، بررسی کردیم و روشی را برای استنباط کامل در باره وابستگی توام بر پایه رگرسیون مفصل نامتقارن بیان کردیم و با تحلیل داده‌های واقعی، نتایج را تشرح نمودیم. این روش را برای داده‌های دو متغیره به کار بردیم اما از این روش در ساختن مفصل‌های نامتقارن چند متغیره و استنباط روی داده‌های چند متغیره نیز استفاده می‌شود. همچنین وابستگی سویی بین دو متغیر را بدون دخالت متغیر خارجی در نظر گرفتیم (متغیر خارجی متغیر مشاهده نشده‌ای است که متغیرهای مورد نظر ما را تحت تاثیر قرار می‌دهد). در حالت کاربردی اغلب چنین متغیرهایی وجود دارند و بررسی آن‌ها ممکن است اثر غیرواقعی بین دو متغیر را نشان دهد. امید است در آینده با کار بیشتر بر روی مدل‌های رگرسیونی بر پایه مفصل با در نظر داشتن اثر متغیر خارجی روی دیگر متغیرها این چالش در زمینه وابستگی سویی برطرف شود.

مراجع

Sungur EA., A ote directional dependence in regression setting, Commun Stat: Theory Method. 2005; 34 (9 and 10): 1957-1965.

Uhm D, Kin J.M., Jung Y., Large asymmetry and directional dependence by using copula modeling to currency exchange rates, Model Assist Stat Appl. 2012; 7(4): 327-340.

Sungur EA., Some observations on copula regression setting. Commun Stat: Theory Method. 2005; 34 (9 and 10): 1967-1978.

Liebscher E., Construction of asymmetric multivariate copulas, J. Multivariate Anal. 2008; 99(10): 2234-2250.

Daeyoung K., Jong M.K., Analysis of directional dependence using asymmetric copula-based regression modelings, Jornal of statistical computation and simulation, 2013 Apr.

Durante F., Construction of non-exchangeeable bivariate distributions functions, Stat Pap. 2009; 50(2): 383-391.