همبستگی در سریهای زمانی

مهمت گنجه‌لی
ترجمهٔ هدایت منتخب *

چکیده

در این مقاله ضمن اشاره به خطاهای ناشی از استفاده از فرمولهای همبستگی بدون زدودن اثر روند از متغیرها، مقایسه‌ای در خصوص نتایج حاصل از داده‌هایی که تحت تأثیر روند بوده و داده‌هایی که روندزدایی شده‌اند صورت می‌گیرد.

۱. اگر همبستگی بین متغیرهای X و Y برای دوره‌ای معین محاسبه شود، با داده‌های زمانی X_t و Y_t $(t = ۱, ۲, \ldots, t)$ روبه‌رو خواهیم بود. در داده‌های زمانی به علت اثر محتمل روند و ناکافی بودن فرمولهای همبستگی خطی برای نشان دادن درجه خطی رابطه، نتایج نادرستی به دست می‌آید. دلیل این مسئله آن است که ضریب همبستگی محاسبه شده از این فرمولها، علاوه بر نشان دادن رابطه بین متغیرهای X_t و Y_t رابطهٔ بین روندها را نیز نشان می‌دهد [۱].

هدف این نوشته اشاره به خطاهای حاصل از استفاده از فرمولهای همبستگی در داده‌های زمانی و روشهای مربوط به تعیین همبستگی بین داده‌های زمانی می‌باشد. در یک سری زمانی مرکب $t = ۱, ۲, \ldots, T$ و (X_t, Y_t)، رابطه بین متغیرها و زمان با معادلات خطی روند

$$X = a + bt \quad , \quad Y = c + dt$$

به چهار گروه تقسیم می‌شود:

الف- فرض کنیم $\tilde{X}_t$ و $\tilde{Y}_t$ نشان دهندهٔ متغیرهایی باشند که روندزدایی شده‌اند، و $(\tilde{X}_t, \tilde{Y}_t)$ در یک جهت تغییر می‌کنند $\cdot < \operatorname{cov}(\tilde{X}_t, \tilde{Y}_t)$، امّا متغیرهای X_t و Y_t اثر روند را دارند. روندهای این متغیرها می‌توانند در جهت عکس هم باشند. $\cdot < b$ و $\cdot < d$ یا $\cdot > b$ و $\cdot > d$.

در این حالت در یک سری زمانی مرکب روند نازدوده $t = ۱, ۲, \ldots, T$ و (X_t, Y_t) همبستگی خطی r_{XY} که بین X_t و Y_t محاسبه شود، به علت اثر روند، مقدار و حتی علامت همبستگی اصلی $r_{\tilde{X}\tilde{Y}}$ را تغییر خواهد داد.

ب- فرض کنیم $\cdot < \operatorname{cov}(\tilde{X}, \tilde{Y})$ بوده ولی روندها در یک جهت باشند؛ $b, d > \cdot$ در این حالت روندها اثری فزاینده بر همبستگی خواهند داشت. ضریب همبستگی r_{XY} که از داده‌های روند نازدوده محاسبه شده، از همبستگی اصلی $r_{\tilde{X}\tilde{Y}}$ یعنی از مقدار واقعی بیشتر خواهد شد.

ج- گیریم متغیرهای $\tilde{Y}_t$ و $\tilde{X}_t$ یعنی داده‌هایی که تحت تأثیر روند نیستند، در جهت عکس تغییر کند، یعنی $\cdot < \operatorname{cov}(\tilde{X}_t, \tilde{Y}_t)$. برای X_t و Y_t روندها در یک جهت‌اند. به علت اثر مثبت روندها خواهیم داشت $\cdot < r_{XY} < r_{\tilde{X}\tilde{Y}}$، حتی در مواقعی که اثر روند خیلی قوی است، $\cdot < r_{\tilde{X}\tilde{Y}}$ و یا $\cdot > r_{\tilde{X}\tilde{Y}}$ خواهد شد.این امر نشان می‌دهد که همبستگی‌های داده‌هایی که تحت تأثیر روند بوده و داده‌هایی که تحت تأثیر روند نبوده‌اند، r_{XY} و $r_{\tilde{X}\tilde{Y}}$ دارای دو علامت متفاوت خواهند بود.

---

*) دکتر هدایت منتخب، عضو هیأت علمی دانشگاه ارومیه

[Downloaded from irstat.ir on 2024-09-28]

د- فرض کنید بازهم $\text{cov}(\tilde{X}_t, \tilde{Y}_t) < 0$. ولی این دفعه روندها در جهت عکس همدیگرند. بدین جهت $r_{\tilde{X}\tilde{Y}} < 0$، به دلیل اثر روند به جهت ۱- سوق داده خواهد شد. $r_{XY} < r_{\tilde{X}\tilde{Y}} < 0$. چنانکه دیده می‌شود ضریب همبستگی به علت اثر روند، گاهی مقادیر و علامتهای متفاوت را در بر می‌گیرد.

از این رو باید از داده‌های روند زدوده استفاده کرد.

۲. اگر در یک سری زمانی متغیرهای X و Y تحت تأثیر روند واقع شوند، همبستگی محاسبه شده علاوه بر نشان دادن جهت و شدت رابطه، اثرات روندها را نیز دربر خواهد داشت. برای زدودن روند، چهار روش مورد استفاده قرار می‌گیرد (۲):

۱) روش نسبتها به روند
۲) محاسبهٔ ضریب همبستگی پیرسون از اطلاعات روند زدوده
۳) روش تفاوتها از روند
۴) ثابت گرفتن زمان در ضریب همبستگی جزیی

از بین چهار روش بالا برخلاف بقیه، در روش ضریب همبستگی جزیی، نیاز به پیدا کردن روند نیست. قبل از استفاده از یکی از این روشها، در یک سری زمانی فرمولهای همبستگی را به کار می‌گیریم.

جدول ۱

خارجیان وارد شده به ترکیه و درآمد جهانگردی ($۱۰۰۰)

سالها	خارجیان وارد شده	درآمد
۱۹۷۱	۹۲۶۰۱۹	۶۲۸۵۷
۱۹۷۲	۱۰۳۴۹۵۵	۱۰۳۷۳۱
۱۹۷۳	۱۳۴۱۵۲۷	۱۷۱۴۷۷
۱۹۷۴	۱۱۱۰۲۹۸	۱۹۳۶۸۴
۱۹۷۵	۱۵۴۰۹۰۴	۲۰۰۸۶۱
۱۹۷۶	۱۶۷۵۸۴۶	۱۸۰۴۵۶
۱۹۷۷	۱۶۶۱۴۱۶	۲۰۴۸۷۷
۱۹۷۸	۱۶۴۴۱۷۷	۲۳۰۳۹۸
۱۹۷۹	۱۵۲۳۶۵۸	۲۸۰۷۲۷
۱۹۸۰	۱۲۸۸۰۶۰	۳۲۶۶۵۴
۱۹۸۱	۱۴۰۵۳۱۱	۳۸۱۲۶۹
۱۹۸۲	۱۳۹۱۷۱۷	۳۷۰۳۲۰
۱۹۸۳	۱۶۲۵۰۹۹	۴۱۱۰۸۸
۱۹۸۴	۲۱۱۷۰۹۴	۸۴۰۰۰۰
۱۹۸۵	۲۶۱۴۹۲۴	۱۴۸۲۰۰۰
۱۹۸۶	۲۳۹۱۰۸۵	۱۲۱۵۰۰۰

منبع: وزارت فرهنگ و جهانگردی ۱۹۸۷

ضریب همبستگی محاسبه شده برای درامد خارجیان وارد شده بین سالهای (۱۹۸۶-۱۹۷۱) برابر است با $r_{XY} = ۰/۹۰۱۸$. حال آزمون فرضیه را انجام می‌دهیم.

$$H_0 : \rho = 0 \qquad H_1 : \rho > 0$$

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{(۰/۹۰۱۸)(۳/۷۴۱۷)}{۰/۴۳۲۲} = \frac{۳/۳۷۴۳}{۰/۴۳۲۲} = ۷/۸۰۷$$

$$۷/۸۰۷ > t_{۰/۰۱;۱۴} = ۲/۶۲۴$$

می‌توان گفت که رابطهٔ بسیاری قوی مثبت بین دو متغیر وجود دارد و $\hat{\rho}_{xy} = ۰/۹۰۱۸$. بازهٔ اطمینان محاسبه شده برای $(\alpha = ۰/۰۵)$ به قرار زیر است

$$j = \frac{1}{2}\ln\frac{1+r}{1-r} = \frac{1}{2}\ln\frac{1+۰/۹۰۱۸}{1-۰/۹۰۱۸} = \frac{1}{2}\ln ۱۹/۳۶۶۶$$

$$= \frac{1}{2}۲/۹۶۳۵۵ = ۱/۴۸۱۵$$

$$S_j = \frac{1}{\sqrt{r-3}} = \frac{1}{\sqrt{13}} = \frac{1}{3/۶۰۵۵} = ۰/۲۷۷$$

$$j - ۱/۹۶S_j \le \varphi \le ۱/۴۸۱۸ + ۱/۹۶(۰/۲۷۷)$$

$$۰/۹۳۸۲ \le \varphi \le ۲/۰۲۵۴$$

$$۰/۷۳۲۹ \le \rho \le ۰/۹۶۵۷$$

به منظور تحقیق اثر روند در سریهای زمانی، طبق پیشنهاد مالینوود (۳) از ضریب همبستگی اسپیرمن استفاده می‌کنیم.

جدول ۲

سالها	۱۹۷۱	۱۹۷۲	۱۹۷۳	۱۹۷۴	۱۹۷۵	۱۹۷۶	۱۹۷۷	۱۹۷۸	۱۹۷۹
خارجیان وارد شده	۹۲۶	۱۰۳۴	۱۳۴۱	۱۱۱۰	۱۵۴۰	۱۶۷۵	۱۶۶۱	۱۶۴۴	۱۵۲۳
t	۱	۲	۳	۴	۵	۶	۷	۸	۹
θ	۱	۲	۵	۳	۹	۱۳	۱۲	۱۱	۸
$(t-\theta)=d$	۰	۰	-۲	۱	-۴	-۷	-۵	-۳	۱
d^2	۰	۰	۴	۱	۱۶	۴۹	۲۵	۹	۱

سالها	۱۹۸۰	۱۹۸۱	۱۹۸۲	۱۹۸۳	۱۹۸۴	۱۹۸۵	۱۹۸۶
خارجیان وارد شده	۱۲۸۸	۱۴۰۵	۱۳۹۱	۱۶۲۵	۲۱۰۷	۲۶۱۵	۲۳۱۹
t	۱۰	۱۱	۱۲	۱۳	۱۴	۱۵	۱۶
θ	۴	۷	۶	۱۰	۱۴	۱۶	۱۵
$(t-\theta)=d$	۶	۴	۶	۳	۰	-۱	+۱
d^2	۳۶	۱۶	۳۶	۹	۰	۱	۱

$$\Sigma d^2 = ۲۰۴$$

[Downloaded from irstat.ir on 2024-09-28]

$$r_s = ۱ - \frac{۶\Sigma d^۲}{n(n^۲ - ۱)} = ۱ - \frac{۶(۲۰۴)}{۱۶(۲۵۵)} = ۱ - \frac{۱۲۲۴}{۴۰۸۰} = ۰/۰۷$$

- اکنون فرض زیر را می‌آوریم:

$H_۰$: در سری خارجیان وارد شده اثر روند وجود ندارد.

$H_۱$: اثر روند وجود دارد.

$$t = \frac{۰/۷۰\sqrt{۱۴}}{\sqrt{۱ - (۰/۷)^۲}} = \frac{۲/۶۱۹}{۰/۷۱۴} = ۳/۶۷$$

پس وجود اثر روند مورد قبول واقع می‌شود.

حال آزمون فرضیه را این دفعه برای درآمد به کار می‌گیریم.

جدول ۳

سالها	۱۹۷۱	۱۹۷۲	۱۹۷۳	۱۹۷۴	۱۹۷۵	۱۹۷۶	۱۹۷۷	۱۹۷۸	۱۹۷۹
درآمد	۶۲	۱۰۳	۱۷۱	۱۹۳	۲۰۱	۱۸۰	۲۰۵	۲۳۰	۲۸۰
t	۱	۲	۳	۴	۵	۶	۷	۸	۹
θ	۱	۲	۳	۵	۶	۴	۷	۸	۹
d	۰	۰	۰	-۱	-۱	۲	۰	۰	۰
$d^۲$	۰	۰	۰	۱	۱	۲	۰	۰	۰

سالها	۱۹۸۰	۱۹۸۱	۱۹۸۲	۱۹۸۳	۱۹۸۴	۱۹۸۵	۱۹۸۶
درآمد	۳۲۶	۳۸۱	۳۷۰	۴۱۱	۸۴۰	۱۴۸۲	۱۲۱۵
t	۱۰	۱۱	۱۲	۱۳	۱۴	۱۵	۱۶
θ	۱۰	۱۲	۱۱	۱۳	۱۴	۱۶	۱۵
d	۰	-۱	+۱	۰	۰	-۱	+۱
$d^۲$	۰	۱	۱	۰	۰	۱	۱

$\Sigma d^۲ = ۱۰$

$$r_s = ۱ - \frac{۶(۱۰)}{۴۰۸۰} = ۰/۹۸۵ \qquad t = \frac{۰/۹۸۵\sqrt{۱۴}}{۰/۱۷۲۶} = ۲۱/۳۵$$

به علت خیلی بزرگ بودن مقدار t در سری درآمد، وجود اثر روند مورد قبول واقع می‌شود.

در این حالت در سریهای X و Y به زدودن اثر روند از سریهای درآمد و خارجیان وارد شده نیاز هست. در اینجا دو روش پیش رو خواهیم داشت:

- به علت وجود اثر روند در هر دو متغیر، اثر زمان ثابت فرض شده ضریب همبستگی جزیی بین X و Y محاسبه می‌شود؛
- استفاده از روشهای مورد بحث.

بدون شک مساوی بودن جوابهای هر دو انتخاب می‌تواند مورد نظر باشد.

انتخاب اول را در داده‌های جدول شماره ۱ به کار گرفته، از فرمول زیر استفاده می‌کنیم.

$$r_{XYt} = \frac{r_{XY} - r_{XY}.r_{Xt}}{[(۱ - r^۲_{Xt})(۱ - r^۲_{Yt})]^{\frac{۱}{۲}}}$$

$$r_{Xt} = \frac{\Sigma_{Xt} - n\bar{X}\bar{t}}{\sqrt{\Sigma X^۲}.\sqrt{\Sigma t^۲}} = \frac{\Sigma Xt}{[\Sigma X^۲.\Sigma t^۲]^{\frac{۱}{۲}}}$$

$$\bar{t} = ۰$$

$$r_{Xt} = \frac{۵۲۰۳۸}{\sqrt{(۳۱۹۷۳۹۷)(۱۳۶۰)}} = \frac{۵۲۰۳۸}{(۶۵۹۴۲/۶۶۷۸۳)} = ۰/۷۸۹۱$$

$$r_{Yt} = \frac{\Sigma Yt}{[\Sigma Y^۲.\Sigma t^۲]^{\frac{۱}{۲}}} = \frac{۴۷۱۴۴}{۵۸۲۹۲/۸۴} = ۰/۸۰۸۷$$

$$r_{XY.t} = \frac{۰/۹۰۱۸ - (۰/۷۸۹۱)(۰/۸۰۸۷)}{\{[۱ - (۰/۷۸۹۱)^۲][۱ - (۰/۸۰۸۷)^۲]\}^{\frac{۱}{۲}}}$$

$$= \frac{۰/۹۰۱۸ - ۰/۶۳۶۶}{۰/۳۶۱۳}$$

$$r_{XY.t} = \frac{۰/۲۶۵۲}{۰/۳۶۱۳} = ۰/۷۳۴$$

$H_۰ : \rho_{XY.t} = ۰ \qquad H_۱ : \rho_{XY.t} > ۰$

$$t = \frac{r_{XY.t}\sqrt{n-۳}}{\sqrt{۱ - r^۲_{XY.t}}} = \frac{۰/۷۳۴\sqrt{۱۳}}{(۱ - ۰/۵۳۸۸)^{\frac{۱}{۲}}} = ۳/۸۹۷$$

$$t_{۰/۰۵;۱۳} = ۲/۱۶$$

به علت بزرگ بودن مقدار t از مقدار جدول یعنی ۲/۱۶ < ۳/۸۹۷فرض $H_۰$ رد شده و $\hat{\rho}_{XY.t} = ۰/۷۳۴$ مورد توجه قرار می‌گیرد. بازهٔ اطمینان با احتمال ۰/۹۵ به صورت زیر می‌باشد.

$$Z_{XY.t} = \frac{۱}{۲}\ln\frac{۱ + r_{XY.t}}{۱ - r_{XY.t}}$$

$$= \frac{۱}{۲}\ln\frac{۱ + ۰/۷۳۴}{۱ - ۰/۷۳۴}$$

$$= \frac{۱}{۲}\ln ۶/۵۱۸۸۰ = ۰/۹۳۷۳$$

$$S_{XY.t} = \frac{۱}{\sqrt{n-۳}} = \frac{۱}{۱۳} = ۰/۲۷۷$$

$$۰/۹۳۷۳ - (۱/۹۶)۰/۲۷۷ \le \varphi_{XY.t} \le ۰/۹۳۷۳ + (۱/۹۶)۰/۲۷۷$$

$$۰/۳۹۴۴ \le \varphi_{XY.t} \le ۱/۴۸۰۲$$

$$۰/۳۷۱۴ \le \rho_{XY.t} \le ۰/۹۰۱۵$$

از طرف دیگر اگر $r_{XY.t}$ با روشهای دیگر حل شود، رسیدن به همان نتیجه بستگی به وجود اثر روند در هر دو متغیر دارد.

[Downloaded from irstat.ir on 2024-09-28]

۳. نتیجه

مقایسهٔ ضرایب $\hat{\rho}_{XY}$ و $\hat{\rho}_{XY.t}$ و در عین حال مقایسه بازه‌های اطمینان، اختلاف فراوانی بین نتایج حاصله را نشان می‌دهد. در بسیاری از تحقیقات به علت در نظر نگرفتن اثر زمان، ضریب همبستگی‌های خیلی بزرگی دیده می‌شود. اگر چه Kendir در سال ۱۹۶۸ به این خطا اشاره کرده ولی مورد توجه قرار نگرفته بود. چنانکه در این نوشته هم دیده می‌شود، در سریهای زمانی محاسبهٔ همبستگی خطی بدون انجام آزمون وجود روند، نتایج به دست آمده اشتباه و باعث تفسیرهای نادرستی می‌شود.

مراجع

[1] Croxton F, Cowden D., Klein L., Applied General Statistics Prentice Hall of India, 1973.

[2] Kar., Kendir Saim, "Zaman Serilerinde Korelasyon ve Regresyon analizleri"Istatistik, Türk Istatistik Derneği Dergisi 1 (1968), Say 12, S. 25.

[3] Malinvaud, Statistical Methods of Econometrics, Rand Mc Nally, Chicago 1966, S. 45.

[Downloaded from irstat.ir on 2024-09-28]