

همبستگی در سریهای زمانی

مهمت گنجه‌لی

* ترجمه هدایت منتخب*

چکیده

اما متغیرهای X_t و Y_t اثر روند را دارند. روندهای این متغیرها می‌توانند در جهت عکس هم باشند. $b > d$ و $b < d$ و یا $b = d$. در این حالت در یک سری زمانی مرکب روند نازدوده $t = r_{XY} = 1, 2, \dots, T$ محاسبه شود، به علت اثر روند، مقدار و حتی علامت همبستگی اصلی $r_{\tilde{X}\tilde{Y}}$ را تغییر خواهد داد.

ب. فرض کیم $\text{cov}(\tilde{X}_t, \tilde{Y}_t) > 0$ بوده ولی روندها در یک جهت باشند: $b > d$ 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 X_t و Y_t یعنی داده‌هایی که تحت تأثیر روند نیستند، در جهت عکس تغییر کند، یعنی $\text{cov}(\tilde{X}_t, \tilde{Y}_t) = 0$. برای X_t و Y_t روندها در یک جهت‌اند. به علت اثر مثبت روندها خواهیم داشت $r_{XY} < r_{\tilde{X}\tilde{Y}}$ حتی در مواقعي که اثر روند خیلی قوی است. $r_{\tilde{X}\tilde{Y}} > 0$ و یا $r_{\tilde{X}\tilde{Y}} < 0$ 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.

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_t و Y_t برای دوره‌ای معین محاسبه شود، با داده‌های زمانی X_t و Y_t ($t = 1, 2, \dots, T$) 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 رابطه 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 ($t = 1, 2, \dots, T$)، 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{Y}_t و \tilde{X}_t نشان دهنده متغیرهایی باشند که روندزدایی شده‌اند، و $(\tilde{Y}_t, \tilde{X}_t)$ در یک جهت تغییر می‌کنند. $\text{cov}(\tilde{X}_t, \tilde{Y}_t) > 0$.

* دکتر هدایت منتخب، عضو هیأت علمی دانشگاه ارومیه

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

$$H_0 : \rho = ۰ \quad H_1 : \rho > ۰$$

$$t = \frac{r \cdot \sqrt{n - 2}}{\sqrt{1 - r^2}} = \frac{(۰,۹۰۱۸)(۲,۷۴۱۷)}{\sqrt{۱ - ۰,۹۰۱۸^2}} = \frac{۲,۳۷۴۳}{\sqrt{۰,۰۹۸۲}} = ۷,۸۰۷$$

$$۷,۸۰۷ > t_{0,001,12} = ۲,۶۴۴$$

می‌توان گفت که رابطه بسیاری قوی مثبت بین دو متغیر وجود دارد و $r_{xy} = ۰,۹۰۱۸$. بازه اطمینان محاسبه شده برای $\alpha = ۰,۰۵$ به قرار زیر است

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

$$= \frac{1}{2} \cdot ۲,۹۶۳۵۵ = ۱,۴۸۱۵$$

$$S_j = \frac{1}{\sqrt{r - ۳}} = \frac{1}{\sqrt{۱۳}} = \frac{1}{\sqrt{۳,۶۰۵۵}} = ۰,۲۷۷$$

$$j - ۱,۴۸۱۵ \leq \varphi \leq ۱,۴۸۱۸ + ۰,۲۷۷$$

$$۰,۹۳۸۲ \leq \varphi \leq ۲,۰۲۵۴$$

$$۰,۷۳۲۹ \leq \rho \leq ۰,۹۶۵۷$$

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

جدول ۲

سالها	خارجیان	وارد شده	
۱۹۷۹ ۱۹۷۸ ۱۹۷۷ ۱۹۷۶ ۱۹۷۵ ۱۹۷۴ ۱۹۷۳ ۱۹۷۲ ۱۹۷۱			
۱۵۲۳ ۱۶۴۴ ۱۶۶۱ ۱۶۷۵ ۱۵۴۰ ۱۱۱۰ ۱۳۴۱ ۱۰۳۴ ۹۲۶			
۹ ۸ ۷ ۶ ۵ ۴ ۳ ۲ ۱			t
۸ ۱۱ ۱۲ ۱۳ ۹ ۳ ۵ ۲ ۱			θ
۱ -۳ -۵ -۷ -۴ ۱ -۲ ۰ ۰			$(t-\theta) = d$
۱ ۹ ۲۵ ۴۹ ۱۶ ۱ ۴ ۰ ۰			d'

سالها

خارجیان

وارد شده

 t θ $(t-\theta) = d$ d'

$$\Sigma d' = ۲۰۴$$

د فرض کنید بازهم $\langle \text{cov}(\bar{X}_t, \bar{Y}_t) \rangle$. ولی این دفعه روندها در جهت عکس همیگردند. بدین جهت $\langle r_{\bar{X}\bar{Y}} \rangle$ به دلیل اثر روند به جهت ۱- سوق داده خواهد شد. $\langle r_{\bar{X}\bar{Y}} \rangle < r_{XY}$. چنانکه دیده می‌شود ضریب همبستگی به علت اثر روند، گاهی مقادیر و علاوه‌های 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) محاسبه ضریب همبستگی پرسون از اطلاعات روند زدوده

۳) روش تفاوتها از روند

۴) ثابت گرفتن زمان در ضریب همبستگی جزئی

از بین چهار روش بالا برخلاف بقیه، در روش ضریب همبستگی جزئی، 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 \$)

درآمد	خارجیان وارد شده	سالها
۶۲۸۵۷	۹۲۶۰۱۹	۱۹۷۱
۱۰۳۷۲۱	۱۰۳۴۹۵۵	۱۹۷۲
۱۷۱۴۷۷	۱۳۴۱۵۲۷	۱۹۷۳
۱۹۳۶۸۴	۱۱۱۰۲۹۸	۱۹۷۴
۲۰۰۸۶۱	۱۵۴۰۹۰۴	۱۹۷۵
۱۸۰۴۵۶	۱۶۷۵۸۴۶	۱۹۷۶
۲۰۴۸۷۷	۱۶۶۱۴۱۶	۱۹۷۷
۲۲۰۳۹۸	۱۶۴۴۱۷۷	۱۹۷۸
۲۸۰۷۲۷	۱۵۲۳۶۵۸	۱۹۷۹
۳۲۶۶۵۴	۱۲۸۸۰۶۰	۱۹۸۰
۳۸۱۲۶۹	۱۴۰۵۳۱۱	۱۹۸۱
۳۷۰۲۲۰	۱۳۹۱۷۱۷	۱۹۸۲
۴۱۱۰۸۸	۱۶۲۵۰۹۹	۱۹۸۳
۸۴۰۰۰۰	۲۱۱۷۰۹۴	۱۹۸۴
۱۴۸۲۰۰۰	۲۶۱۴۹۲۴	۱۹۸۵
۱۲۱۵۰۰۰	۲۲۹۱۰۸۵	۱۹۸۶

منبع: وزارت فرهنگ و جهانگردی ۱۹۸۷

استفاده می‌کنیم:

$$r_{XYt} = \frac{r_{XY} - r_{XY} \cdot r_{Xt}}{[(1 - r_{Xt}^2)(1 - r_{Yt}^2)]^{\frac{1}{2}}}$$

$$r_{Xt} = \frac{\Sigma X_t - n\bar{X}t}{\sqrt{\Sigma X^2} \cdot \sqrt{\Sigma t^2}} = \frac{\Sigma X_t}{[\Sigma X^2 \cdot \Sigma t^2]^{\frac{1}{2}}}$$

$$\bar{t} = 0$$

$$r_{Xt} = \frac{52038}{\sqrt{(3117397)(1360)}} = \frac{52038}{(659422/66782)} = 0,7891$$

$$r_{Yt} = \frac{\Sigma Y_t}{[\Sigma Y^2 \cdot \Sigma t^2]^{\frac{1}{2}}} = \frac{47144}{58292/84} = 0,8087$$

$$r_{XY.t} = \frac{0,9018 - (0,7891)(0,8087)}{\{(1 - (0,7891)^2)[1 - (0,8087)^2]\}^{\frac{1}{2}}} \\ = \frac{0,9018 - 0,6266}{0,3613}$$

$$r_{XY.t} = \frac{0,2652}{0,3613} = 0,734$$

$H_0 : \rho_{XY.t} = 0$ $H_1 : \rho_{XY.t} > 0$

$$t = \frac{r_{XY.t} \sqrt{n-3}}{\sqrt{1 - r_{XY.t}^2}} = \frac{0,734 \sqrt{13}}{(1 - 0,5388)^{\frac{1}{2}}} = 2,897$$

$$t_{0,05;13} = 2,16$$

به علت بزرگ بودن مقدار t از مقدار جدول یعنی $2,16 < 2,897$ فرض H_0 رد شده و $0,734 = \hat{\rho}_{XY.t}$ مورد توجه قرار می‌گیرد. بازه اطمینان با احتمال 95% به صورت زیر می‌باشد.

$$Z_{XY.t} = \frac{1}{2} \ln \frac{1 + r_{XY.t}}{1 - r_{XY.t}} \\ = \frac{1}{2} \ln \frac{1 + 0,734}{1 - 0,734} \\ = \frac{1}{2} \ln 6,51880 = 0,93723$$

$$S_{XY.t} = \frac{1}{\sqrt{n-3}} = \frac{1}{13} = 0,277$$

$$0,93723 - 0,277 \leq \varphi_{XY.t} \leq 0,93723 + (1,96)0,277$$

$$0,3944 \leq \varphi_{XY.t} \leq 1,4802$$

$$0,3714 \leq \rho_{XY.t} \leq 0,9015$$

از طرف دیگر اگر $r_{XY.t} > 0$ با روش‌های دیگر حل شود، رسیدن به همان نتیجه مستگی به وجود اثر روند در هر دو متغیر دارد.

$$r_s = 1 - \frac{6 \sum d^2}{n(n^2 - 1)} = 1 - \frac{6(204)}{16(255)} = 1 - \frac{1224}{4080} = 0,907$$

نم اکنون فرض زیر را می‌آوریم:

در سری خارجیان وارد شده اثر روند وجود ندارد.

H_0 : اثر روند وجود ندارد.

$$t = \frac{0,70 \sqrt{14}}{\sqrt{1 - (0,7)^2}} = \frac{2,619}{0,714} = 3,67$$

پس وجود اثر روند مورد قبول واقع می‌شود.

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

جدول ۳

سالها									
۱۹۷۹	۱۹۷۸	۱۹۷۷	۱۹۷۶	۱۹۷۵	۱۹۷۴	۱۹۷۳	۱۹۷۲	۱۹۷۱	
۲۸۰	۲۳۰	۲۰۵	۱۸۰	۲۰۱	۱۹۳	۱۷۱	۱۰۳	۶۲	درآمد
۹	۸	۷	۶	۵	۴	۲	۲	۱	t
۹	۸	۷	۴	۶	۵	۲	۲	۱	θ
۰	۰	۰	۲	-۱	-۱	۰	۰	۰	d
۰	۰	۰	۲	۱	۱	۰	۰	۰	d'
سالها									
۱۹۸۶	۱۹۸۵	۱۹۸۴	۱۹۸۳	۱۹۸۲	۱۹۸۱	۱۹۸۰			
۱۲۱۵	۱۴۸۲	۸۴۰	۴۱۱	۳۷۰	۳۸۱	۳۲۶			درآمد
۱۶	۱۵	۱۴	۱۳	۱۲	۱۱	۱۰			t
۱۵	۱۶	۱۴	۱۳	۱۱	۱۲	۱۰			θ
+1	-1	۰	۰	+1	-1	۰			d
1	1	۰	۰	1	1	۰			d'

$$\Sigma d^2 = 10$$

$$r_s = 1 - \frac{6(10)}{4080} = 0,985 \quad t = \frac{0,985 \sqrt{14}}{0,1726} = 21,35$$

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 جزئی 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

۳. نتیجه

مورد توجه قرار نگرفته بود. چنانکه در این نوشه هم دیده می‌شود، در سریهای زمانی محاسبه همبستگی خطی بدون انجام آزمون وجود روند، نتایج به دست آمده اشتباه و باعث تفسیرهای نادرستی می‌شود.

مقایسه ضرایب $\hat{\rho}_{XY,t}$ و $\hat{\rho}_{XYZ,t}$ 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. اگرچه Kendir در سال ۱۹۶۸ به این خطأ اشاره کرده ولی

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.