Archive of SID

سیزدهمین کنفرانس آمار ایران
۴-۲ شهریور ۱۳۹۵

آزمون همبستگی فضایی برای مدل‌های رگرسیونی پانلی

سارا ساسانی [۱]، محسن محمدزاده

گروه آمار دانشگاه تربیت مدرس

چکیده: در این مقاله ضمن معرفی مدل رگرسیون داده‌های پانلی فضایی، نحوه بررسی همبستگی فضایی خطاها از طریق آزمون ضریب لاگرانژ، به شرط وجود و عدم وجود اثرات تصادفی ، مورد مطالعه قرار می‌گیرد. برای این منظور آماره آزمون شرطی و حاشیه‌ای ضریب لاگرانژ و توزیع مجانبی آن معرفی می‌شود. یک مطالعه شبیه‌سازی به منظور بررسی توان و اندازه این آزمون‌ها برای فرضیه‌های شرطی و حاشیه‌ای انجام شده است. سپس نحوه کاربست این آزمون در مثالی کاربردی با بررسی همبستگی فضایی خطاها و اثرات تصادفی موجود در داده‌های صادرات محصولات کشاورزی کشورهای عضو اکو نشان داده می‌شود. در انتها نیز بحث و نتیجه‌گیری ارائه می‌گردد.

واژه‌های کلیدی: آزمون ضریب لاگرانژ، مدل رگرسیون پانلی، داده‌های پانلی فضایی.

کد موضوع‌بندی ریاضی (۲۰۱۰): 91Bxx، 62FXX.

۱ مقدمه

داده‌های پانلی از **واحدهای مشخص** در **زمان‌های متوالی** به‌دست می‌آیند. وقتی واحدها موقعیت‌های فضایی باشند و داده‌ها از واحدی به واحد دیگر در طول زمان مستقل باشند، داده‌های **پانلی فضایی** گویند. در این مقاله به بررسی وجود همبستگی فضایی خطاهای مدل رگرسیون داده‌های پانلی که توسط بالتاجی (۲۰۰۵) ارائه شده است، با آزمون LM شرطی و حاشیه‌ای که همبستگی فضایی خطاها را به‌شرط امکان وجود و عدم وجود اثرات تصادفی آزمون می‌کند، پرداخته می‌شود. در بخش ۲ مدل رگرسیون داده‌های پانلی فضایی و آماره آزمون‌های LM شرطی و حاشیه‌ای ارائه می‌شوند. در بخش ۳ به مطالعه شبیه‌سازی در خصوص توان آزمون برای بررسی وجود همبستگی فضایی خطاها پرداخته می‌شود. در بخش ۴ مدل رگرسیونی مناسب به داده‌های صادرات محصولات کشاورزی کشورهای عضو اکو، برازش داده می‌شود. در بخش ۵ بحث و نتیجه‌گیری و پیشنهادات ارائه می‌شود.

[۱] سارا ساسانی : s.sasani@modares.ac.ir

www.SID.ir

Archive of SID

۲ مدل رگرسیون داده‌های پانلی فضایی

مدل رگرسیونی به‌صورت

$$y_{ti} = X'_{ti}\beta + u_{ti}, \quad i = ۱, \ldots, N, \quad t = ۱, \ldots, T \tag{۱.۲}$$

را در نظر بگیرید (بالتاجی، ۲۰۰۵) که در آن y_{ti} مشاهده متغیر پاسخ در موقعیت iام و دوره زمانی tام، X_{ti} بردار $K \times ۱$ متغیرهای تبیینی، β بردار پارامترهای رگرسیونی و u_{ti} عبارت خطا است. آنسلین (۱۹۸۸) فرم برداری (۱.۲) با اثرات تصادفی و خطاهای خودهمبسته فضایی را به‌صورت

$$u_t = \mu + \epsilon_t \tag{۲.۲}$$

$$\epsilon_t = \lambda W \epsilon_t + \nu_t \tag{۳.۲}$$

بیان کرد، که در آن $u'_t = (u_{t۱}, \ldots, u_{tN})$، $\epsilon'_t = (\epsilon_{t۱}, \ldots, \epsilon_{tN})$ و $\mu' = (\mu_۱, \ldots, \mu_N)$ بردار اثرات تصادفی است و فرض می‌شود $\mu' \overset{iid}{\sim} N(۰, \sigma^۲_\mu)$، λ ضریب اتورگرسیو فضایی با قید $|\lambda| < ۱$ و دارای توزیع $\nu'_t = (\nu_{t۱}, \ldots, \nu_{tN})$، $\nu'_t \overset{iid}{\sim} N(۰, \sigma^۲_\nu)$ و به‌ازای هر i و t درایه $\{\nu_{ti}\}$ از درایه $\{\mu_i\}$ مستقل است، W ماتریس وزن فضایی $N \times N$ با درایه‌های قطر اصلی صفر است، به‌طوری که ماتریس $(I_N - \lambda W)$ برای همه $|\lambda| < ۱$، ناتکین[۱] است. مدل (۳.۲) را می‌توان به‌صورت

$$\epsilon_t = (I_N - \lambda W)^{-۱}\nu_t = B^{-۱}\nu_t \tag{۴.۲}$$

بازنویسی کرد، که در آن $B = I_N - \lambda W$ و I_N ماتریس همانی $N \times N$ است. فرم ماتریسی مدل (۱.۲)به‌صورت

$$y = X\beta + u \tag{۵.۲}$$

است، که در آن y بردار مشاهدات $NT \times ۱$، X ماتریس طرح $NT \times K$، β بردار ضرایب $K \times ۱$ و u بردار خطای $NT \times ۱$ است. همچنین فرض می‌شود X ماتریس پررتبه ستونی است و قدر مطلق درایه‌های آن به‌طور مجانبی کراندار هستند. مدل (۲.۲) را نیز می‌توان به‌صورت برداری

$$u = (j_T \otimes I_N)\mu + (I_T \otimes B^{-۱})\nu \tag{۶.۲}$$

نوشت، که در آن $\nu' = (\nu'_۱, \ldots, \nu'_T)$، j_T بردار $T \times ۱$ با درایه‌های یک، I_T ماتریس همانی $T \times T$ و $\otimes$ نماد ضرب کرونکر است. ماتریس کوواریانس u را می‌توان به‌صورت (وانزبک و کاپتین، ۱۹۸۳)،

$$\Sigma_u = \sigma^۲_\nu[\bar{J}_T \otimes (T\phi I_N + (B'B)^{-۱}) + E_T \otimes (B'B)^{-۱}] = \sigma^۲_\nu \Sigma_u \tag{۷.۲}$$

نوشت، که در آن

$$\Sigma_u = [\bar{J}_T \otimes (T\phi I_N + (B'B)^{-۱}) + E_T \otimes (B'B)^{-۱}]$$

$$|\Sigma_u| = |T\phi I_N + (B'B)^{-۱}| \cdot |(B'B)^{-۱}|^{T-۱}$$

$\phi = \frac{\sigma^۲_\mu}{\sigma^۲_\nu}$، $\bar{J}_T = \frac{J_T}{T}$، $E_T = I_T - \bar{J}_T$. آنسلین (۱۹۸۸) آزمون LM را برای حالت خاص $\lambda = ۰$ در این مدل ارائه کرد. در این مقاله، آزمون ارائه شده توسط آنسلین با به‌دست آوردن آزمون شرطی و حاشیه‌ای برای همبستگی فضایی خطاها تعمیم داده می‌شود.

[۱] Nonsingular

www.SID.ir

Archive of SID

۱.۲ آزمون LM شرطی برای همبستگی خطافضایی به‌شرط $\sigma_\mu^2 \geqslant 0$

فرضیه‌های $H_0 : \lambda = 0$ در مقابل $H_1 : \lambda \neq 0$ را در نظر بگیرید. با قرار دادن $\theta = (\sigma_\nu^2, \sigma_\mu^2, \lambda)'$ آماره آزمون LM به‌صورت $LM = \tilde{q}'_\theta \tilde{J}_\theta^{-1} \tilde{q}_\theta$ ارائه می‌شود، که در آن $\tilde{q}_\theta = (\frac{\partial \ell}{\partial \theta})(\tilde{\theta})$ بردار 3×1 مشتقات جزئی نسبت به هر یك از پارامترهای θ است که با برآوردگرهای ماکسیمم درستنمایی مقید، یعنی $\tilde{\theta}$، جایگذاری شده‌اند. همچنین $\tilde{J} = E[-\frac{\partial^2 \ell}{\partial \theta \partial \theta'}](\tilde{\theta})$ ماتریس اطلاع θ در نقاط $\tilde{\theta}$ است. تحت فرضیه صفر H_0، ماتریس کوواریانس u به فرم $\Sigma_0 = \sigma_\mu^2 (J_T \otimes I_N) + \sigma_\nu^2 I_{NT}$ حاصل می‌شود. این مدل مشابه مدل آزمون مؤلفهٔ خطای یک‌طرفه (بالتاجی، ۲۰۰۵)، با معکوس ماتریس کوواریانس به‌صورت $\Sigma_0^{-1} = (\sigma_1^2)^{-1}(\bar{J}_T \otimes I_N) + (\sigma_\nu^2)^{-1}(E_T \otimes I_N)$ است، که در آن $E_T = I_T - \bar{J}_T$ و $\sigma_1^2 = T\sigma_\mu^2 + \sigma_\nu^2$.

همرل و هارتلی (۱۹۷۳) یك فرمول کلی مفید برای به‌دست آوردن $\tilde{q}_\theta$ به‌صورت

$$\frac{\partial \ell}{\partial \theta_r} = -\frac{1}{2} tr[\Sigma_u^{-1}(\frac{\partial \Sigma_u}{\partial \theta_r})] + \frac{1}{2}[u' \Sigma_u^{-1}(\frac{\partial \Sigma_u}{\partial \theta_r})\Sigma_u^{-1} u], \quad r = 1, 2, 3 \tag{۸.۲}$$

ارائه دادند. با استفاده از (۸.۲) تحت فرضیه صفر داریم:

$$\hat{q} = q(\hat{\theta}) = (0, 0, \hat{q}_\lambda)'$$

که در آن $\hat{\sigma}_\nu^2 = \frac{\hat{u}'(E_T \otimes I_N)\hat{u}}{N(T-1)}$، $\hat{\sigma}_1^2 = \frac{\hat{u}'(\bar{J}_T \otimes I_N)\hat{u}}{N}$ و $\hat{u}$ به‌ترتیب برآوردگرهای ماکسیمم درستنمایی σ_ν^2، σ_1^2 و مانده‌های ماکسیمم درستنمایی u تحت فرضیه صفر هستند. همچنین $\hat{q}_\lambda = \frac{1}{2\hat{\sigma}_\nu^2}[[\hat{u}'(E_T \otimes (W + W'))\hat{u}] + \frac{\hat{\sigma}_\nu^2}{2\hat{\sigma}_1^4}[\hat{u}'(\bar{J}_T \otimes (W + W'))\hat{u}]$.

هارویل (۱۹۷۷)، ماتریس اطلاع را به‌صورت

$$J_{rs} = E[-\frac{\partial^2 \ell}{\partial \theta_r \partial \theta_s}] = \frac{1}{2} tr[\Sigma_u^{-1}(\frac{\partial \Sigma_u}{\partial \theta_r})\Sigma_u^{-1}(\frac{\partial \Sigma_u}{\partial \theta_s})], \quad r, s = 1, 2, 3 \tag{۹.۲}$$

ارائه نمود. با استفاده از (۹.۲) و تحت فرضیه صفر ماتریس اطلاع به‌صورت

$$\hat{j} = \begin{pmatrix} \frac{N}{2}(\frac{1}{\hat{\sigma}_1^4} + \frac{T-1}{\hat{\sigma}_\nu^4}) & \frac{NT}{2\hat{\sigma}_1^4} & 0 \\ \frac{NT}{2\hat{\sigma}_1^4} & \frac{NT^2}{2\hat{\sigma}_1^4} & 0 \\ 0 & 0 & (T - 1 + \frac{\hat{\sigma}_\nu^4}{2\hat{\sigma}_1^4})b \end{pmatrix}$$

ارائه می‌شود و در پی آن

$$LM_\lambda = \hat{q}'_\lambda \hat{j}_\theta^{-1} \hat{q}_\lambda = \frac{\hat{q}_\lambda^2}{[(T-1) + \frac{\hat{\sigma}_\nu^4}{\hat{\sigma}_1^4}]b}$$

در نتیجه آماره آزمون برای فرضیه $H_0 : \lambda = 0$ در مقابل $H_1 : \lambda \neq 0$ به‌شرط $\sigma_\mu^2 \geqslant 0$ به‌صورت

$$LM_c = \frac{\hat{q}_\lambda^2}{[(T-1) + \frac{\hat{\sigma}_\nu^4}{\hat{\sigma}_1^4}]b} \tag{۱۰.۲}$$

قابل ارائه است، که در آن $b = tr(W^2 + W'W)$. به‌علاوه آماره آزمون فرضیه $H_0 : \lambda = 0$ در مقابل $H_1 : \lambda > 0$ به‌شرط $\sigma_\mu^2 \geqslant 0$، به‌صورت

$$LM_c^* = \frac{\hat{q}_\lambda}{\sqrt{[(T-1) + \frac{\hat{\sigma}_\nu^4}{\hat{\sigma}_1^4}]b}} \tag{۱۱.۲}$$

www.SID.ir

Archive of SID

ارائه می‌شود، که تحت فرضیه صفر برای T ثابت وقتی $N \longrightarrow \infty$، دارای توزیع مجانبی $N(۰,۱)$ است. تحت شرایط فوق، با نادیده گرفتن اثرات تصادفی آماره آزمون LM حاشیه‌ای برای همبستگی خطافضایی با فرض $\sigma_\mu^۲ = ۰$ به‌صورت

$$LM_m^۲ = \frac{N^۲T}{b}H^۲$$

حاصل می‌شود که در آن $H = \frac{\tilde{u}'(I_T \otimes W)\tilde{u}}{\tilde{u}'\tilde{u}}$ و تحت فرضیهٔ صفر به‌طور مجانبی دارای توزیع $\chi^۲$ با یک درجه آزادی است (آنسلین، ۱۹۸۸a). از طرفی این آماره را می‌توان به‌صورت

$$LM_m = \sqrt{\frac{N^۲T}{b}}H \qquad (۱۲.۲)$$

نوشت، که تحت فرضیهٔ صفر دارای توزیع مجانبی $N(۰,۱)$ است.

۳ مطالعه شبیه‌سازی

مدل رگرسیونی پانلی با خطاهای فضایی به‌صورت

$$y = \alpha j_{NT} + \beta_۱ x_۱ + \beta_۲ x_۲ + u \qquad (۱.۳)$$

$$u = (j_T \otimes I_N)\mu + (I_T \otimes (I_N - \lambda W)^{-۱})\nu \qquad (۲.۳)$$

درنظر گرفته شده است که در آن j_{NT} بردار NT تایی با درایه‌های یک، $\alpha = ۴$، $\beta_۱ = ۱$، $\beta_۲ = -۰/۵$، $x_۱$، $x_۲$ بردارهای NT تایی متغیرهای مستقل هستند که به‌ترتیب از توزیع‌های $x_۱ \sim u(-۰/۵, ۰/۵)$ و $x_۲ \sim N(۰,۱)$ تولید شده‌اند. j_T بردار T تایی با درایه‌های یک، I_N ماتریس همانی N بعدی، I_T ماتریس همانی T بعدی و $\otimes$ نماد ضرب کرونکر است. μ بردار اثرات تصادفی N تایی است که از توزیع $N(۰,\sigma_\mu^۲)$ و ν بردار NT تایی خطاها است که از توزیع $N(۰,۱)$ تولید شده است. ماتریس وزن W از روش مجاورت استاندارد شده برای هر دو معیار راک و کوئین حاصل شده است. روش مجاورت راک[۲] برای موقعیت‌هایی که مرز مشترک راست، چپ، بالا و پائین با موقعیت تحت بررسی را دارند تعریف می‌شود (آنسلین و ری، ۱۹۹۱). روش مجاورت کوئین[۳] مشتمل بر هشت موقعیت مجاور به موقعیت تحت بررسی است. در این روش علاوه بر چهار موقعیت مذکور از روش راک، موقعیت‌هایی که دارای رأس مشترک هستند نیز همسایه محسوب می‌شوند (آنسلین و ری، ۱۹۹۱). برای λ ضریب همبستگی خطافضایی دنباله مقادیر ۰ تا ۰/۹ با قدر نسبت ۰/۱ و برای $\sigma_\mu^۲$ واریانس اثرات تصادفی، مقادیر ۰، ۱، ۴، ۹ و ۱۶ در نظر گرفته شده است. همچنین دو مقدار ۲۵ و ۴۹ برای N، موقعیت‌های فضایی و۳ و ۷ برای T دورهٔ زمانی اختیار شده است. شکل‌های ۱ و ۲ نمودار فراوانی نسبی رد $H_۰ : \lambda = ۰$ را در ۲۰۰۰ مورد به‌ازای مقادیر $N = ۲۵, ۴۹$ و $T = ۳, ۷$ برای هر دو ماتریس وزن کوئین و راک نشان می‌دهند که در آنها محور افقی مقادیر λ و محور عمودی مقادیر توان آزمون است. توان آماره آزمون‌های LM حاشیه‌ای و شرطی ارائه شده در (۱۲.۲) و (۱۱.۲) به‌ازای مقادیر مختلف λ، در سطح معنی‌داری ۰/۰۵ به‌ترتیب با خطوط کم‌رنگ و پررنگ رسم شده است. همان‌طور که از نمودارها پیداست، توان هر دو آماره آزمون به‌ازای $\sigma_\mu^۲ \leqslant ۴$ تقریباً بر هم منطبق هستند. اما با افزایش واریانس اثرات تصادفی آماره LM_m و LM_c^* از حیث اندازه و توان اختلاف قابل توجهی دارند. این اختلاف در توان و اندازه آزمون‌ها نشان می‌دهد که هنگام

[۲] Rook Contiguity

[۳] Queen Contiguity

www.SID.ir

Archive of SID

آزمون همبستگی فضایی خطاها، نادیده گرفتن اثرات تصادفی در حالی‌که واقعاً وجود دارند، می‌تواند منجر به استنباط‌های گمراه‌کننده‌ای شود.

۴ بررسی عوامل مؤثر بر صادرات محصولات کشاورزی کشورهای عضو اکو

به‌منظور پیدا کردن شواهد آماری در تعیین عوامل مؤثر بر صادرات محصولات کشاورزی کشورهای عضو اکو داده‌هایی طی ۱۷ سال متوالی از سال ۱۹۹۲ تا ۲۰۰۸ برای هر یک از کشورهای عضو اکو شامل افغانستان، جمهوری آذربایجان، ایران، قزاقستان، قرقیزستان، پاکستان، تاجیکستان، ترکیه، ترکمنستان و ازبکستان با متغیرهای پاسخ y صادرات محصولات کشاورزی، x_1 تولید ناخالص داخلی، x_2 نرخ ارز و x_3 جمعیت از سایت بانک جهانی جمع‌آوری شده‌اند. در این زیربخش همبستگی فضایی خطاها و اثرات تصادفی با آزمون‌های LM شرطی و حاشیه‌ای انجام شده است. مقدار آماره آزمون $LM_e = ۲۲/۰۰۹$ با p-مقدار کمتر از ۰/۰۰۰۱ حاصل شده است. پس فرضیه وجود همبستگی فضایی خطاها به‌شرط امکان وجود اثرات تصادفی معنی‌دار است. بنابراین خطاها در این مدل به‌طور قابل توجهی همبستگی فضایی دارند و برای تحلیل آن‌ها باید همبسته بودن خطاها مد نظر قرار گیرند. برای این منظور مدل (۱.۳) با عبارت خطای ارائه شده در رابطه (۲.۳) به داده‌ها برازانده شده است. همان‌طور که در جدول ۱ ملاحظه می‌شود، همه متغیرها به‌جز نرخ ارز، بر صادرات محصولات کشاورزی کشورهای عضو اکو به‌شدت تأثیرگذارند. λ ضریب همبستگی فضایی خطاها با برآورد ۰/۴۷۹ به‌میزان قابل توجهی معنی‌دار است. همچنین ϕ نسبت واریانس اثرات تصادفی به واریانس خطاها، با برآورد ۱۲/۲۱۱ در سطح $\alpha = ۰/۰۵$ معنی‌دار است.

جدول ۱: برآورد پارامترهای مدل رگرسیونی برای صادرات محصولات کشاورزی کشورهای عضو اکو.

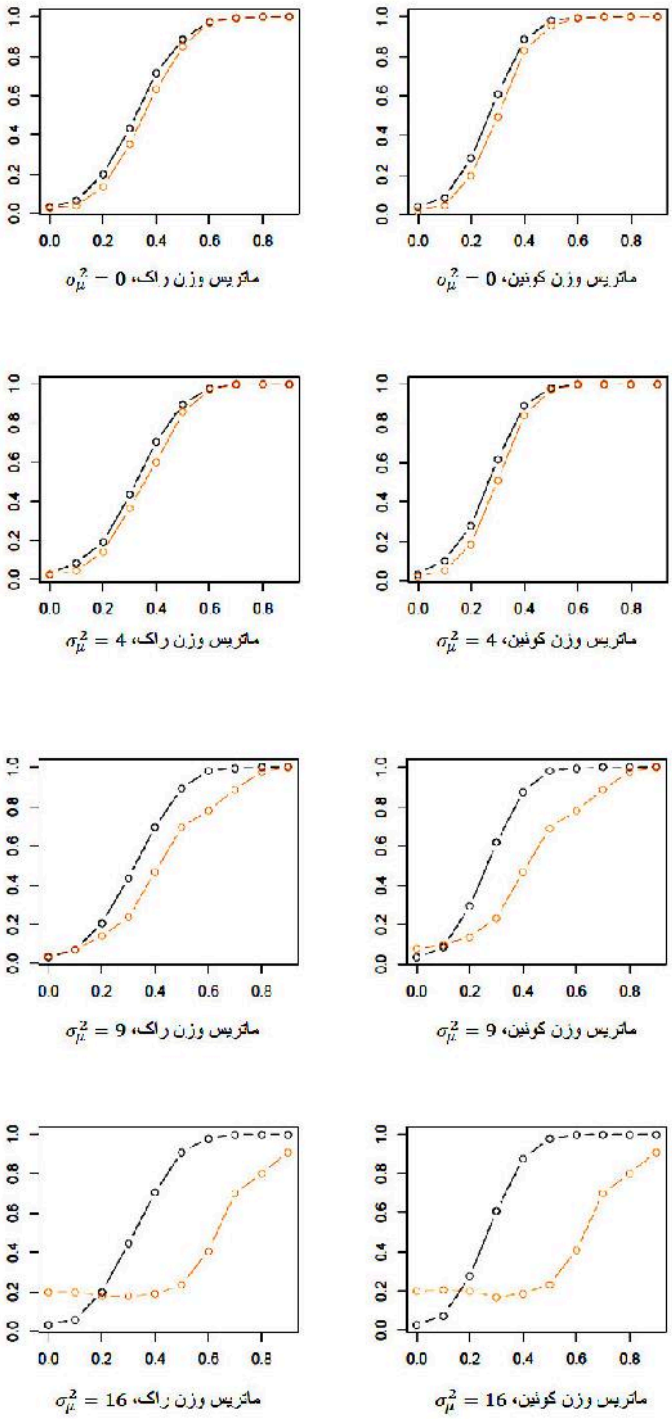
پارامتر	برآورد	انحراف معیار	p-مقدار	مقدار t
β_0	۷/۶۹۱	۲/۵۴۴	۱/۵۱۸	۰/۱۲۸
β_1	۰/۸۰۶	۰/۱۲۴	۶/۵۲۴	۰/۰۰۰۱
β_2	-۰/۰۳۷	۰/۰۱۹	-۱/۹۱۴	۰/۰۵۵
β_3	-۰/۸۱۶	۰/۳۶۷	-۲/۲۲۵	۰/۰۲۶
ϕ	۱۲/۲۱۱	۵/۷۳۲	۲/۱۳۰	۰/۰۳۳
λ	۰/۴۷۹	۰/۱۰۰	۴/۷۸۳	۰/۰۰۰۱

۵ بحث و نتیجه‌گیری

در هر مدل رگرسیونی آزمون استقلال یا وابستگی خطاها از اهمیت ویژه‌ای برخوردار است. از جمله مسائل مهم در مدل‌بندی داده‌های پانلی فضایی، وجود همبستگی فضایی مشاهدات در هر نقطه از زمان و تغییرپذیری بین واحدهای آزمایشی است که باید به نحوی در مدل لحاظ شوند. مدل‌های رگرسیونی متنوعی برای مدل‌بندی این داده‌ها ارائه شده است. از جمله این مدل‌ها مدل خطافضایی با اثرات تصادفی است که در این مقاله به آن اشاره شد. برای بررسی همبستگی فضایی خطاها و اثرات تصادفی این مدل، آزمون ضریب لاگرانژ

www.SID.ir

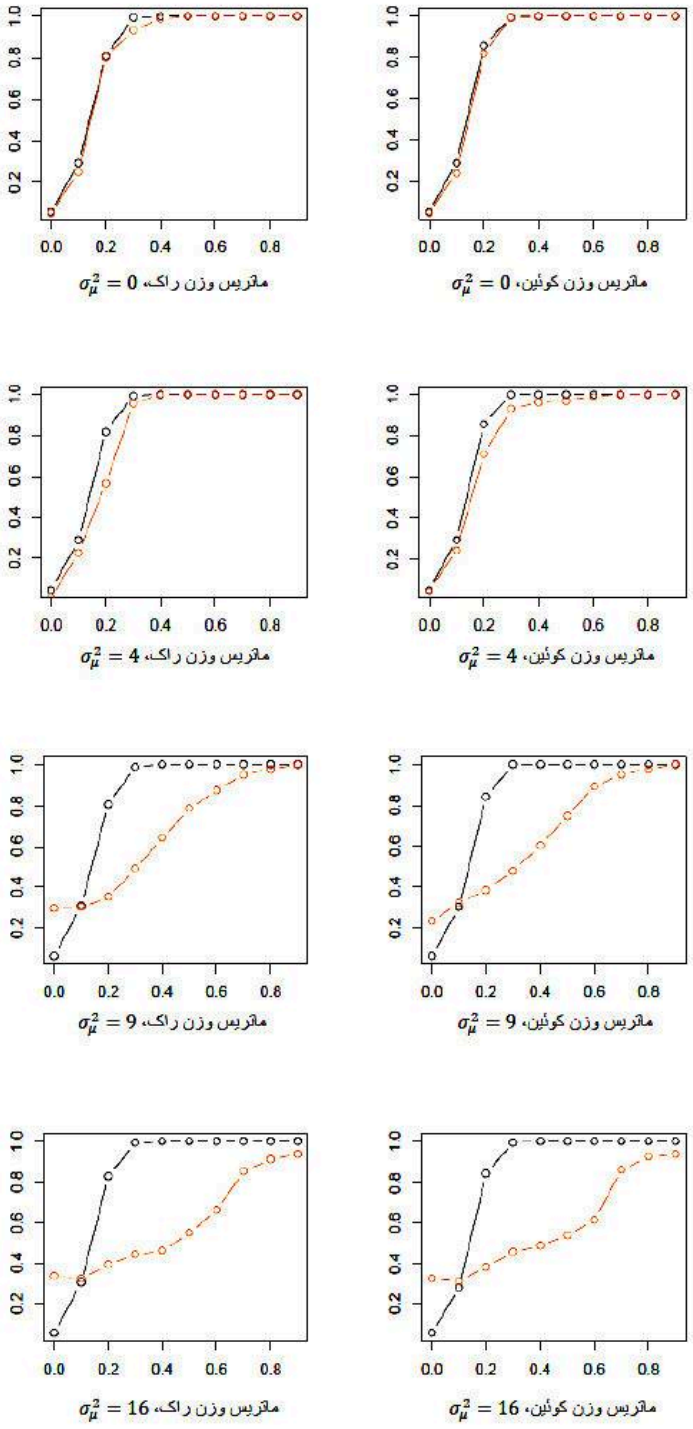
Archive of SID



شکل ۱: توان آزمون‌های LM شرطی (خط پررنگ) و حاشیه‌ای (خط کم‌رنگ) برای $H_0 : \lambda = 0$ به‌ازای $N = 25$ و $T = 3$ در ۲۰۰۰ مورد.

www.SID.ir

Archive of SID



شکل ۲: توان آزمون‌های LM شرطی (خط پررنگ) و حاشیه‌ای (خط کم‌رنگ) برای $H_0 : \lambda = 0$ به‌ازای $N = 49$ و $T = 7$ در ۲۰۰۰ مورد.

www.SID.ir

Archive of SID

متناظر با فرضیه‌های شرطی و حاشیه‌ای معرفی و ارائه شد. در مطالعه شبیه‌سازی و یك مثال کاربردی در خصوص داده‌های صادرات محصولات کشاورزی کشورهای عضو اکو نحوه کاربست این آزمون نشان داده شد.

مراجع

[1] Anselin, L. (1988), *Spatial Econometrics: Methods and Models* (Kluwer Academic Publishers, Dordrecht).

[2] Anselin, L. and Rey, S. (1991), Properties of Tests for Spatial Dependence in Linear Regression Models *Geographical Analysis*, **23**, 112-131.

[3] Baltagi, B. H. (2005), *Econometrics Analysis of Panel Data*, New York, Wiley.

[4] Harville, D. A. (1977), Maximum Liklihood Approaches to Variance Component Estimation and to Related Problems, *Journal or the American Statistical Association*, **72**, 320-338.

[5] Hemmerle, W. J. and Hartley, H. O. (1973), Computing Maximum Liklihood Estimates for the Mixed A. O. V. Model Using the W-Transformation, *Technometrics*, **15**, 819-831.

[6] Wansbeek, T. J. and Kapteyn, A. (1983), A Note on Spectral Decomposition and Maximum Likelihood Estimation of ANOVA Models with Balanced Data, *Statistics and Probability Letters*, **1**, 213-215.

www.SID.ir