برآورد بیزی پارامترهای مدل رگرسیون با خطاهای خودهمبسته فضایی

امید کریمی، محسن محمدزاده*
گروه آمار، دانشکده علوم پایه، دانشگاه تربیت مدرس، تهران، ایران
*مسئول مکاتبات- آدرس الکترونیکی: *mohsen_m@modares.ac.ir*

(دریافت: ۸۵/۹/۲۰؛ پذیرش:۸۶/۶/۲۴)

چکیده

معمولاً در آنالیز رگرسیون فرض بر این است که خطاهای الگو مستقل هستند، اما در عمل گاهی با مواردی مانند داده‌های فضایی مواجه می‌شویم که خطاهای مدل همبسته هستند و ساختار همبستگی آنها تابعی از موقعیت قرار گرفتن مشاهدات در فضای مورد مطالعه است. از اینگونه مدلها که رگرسیون فضایی نام دارند، برای تعیین رویه‌ها در زمین‌شناسی، باستان شناسی، همه‌گیر شناسی و پردازش تصاویر استفاده می‌شود. در این مقاله مدل رگرسیون فضایی با خطاهای خودهمبسته فضایی مرتبه اول با استفاده از رهیافت بیزی مورد بررسی قرار می‌گیرد. از آنجا که تعیین توزیع پسین پارامترها دشوار می‌باشد، برای برآورد بیزی پارامترها و پیش‌بینی بیزی مشاهدات از روش MCMC استفاده شده است. سپس نحوه اجرا و کارائی روشهای ارائه شده در یک مطالعه شبیه‌سازی برای حجم نمونه و اندازه شبکه‌های مختلف مورد بررسی قرار گرفته است.

واژه‌های کلیدی: خطاهای خود همبسته فضایی، روشهای MCMC، نمونه‌گیری گیبس، الگوریتم متروپلیس- هاستینگس.

مقدمه

داده‌های فضایی مشاهداتی هستند که به یکدیگر وابسته بوده و همبستگی آنها ناشی از موقعیت قرار گرفتن داده‌ها در فضای مورد مطالعه است. در سالهای اخیر، مقالات متعددی برای بررسی جنبه‌های مختلف داده‌های فضایی ارائه شده است. از جمله می‌توان به مدلهای رگرسیون با خطاهای خودهمبسته فضایی اشاره کرد که برای تعیین رویه‌ها در زمین‌شناسی، باستان‌شناسی، همه‌گیر شناسی، جغرافیا و پردازش تصاویر استفاده می‌شود. وین سِک و راینسل (Wincek & Reincel 1986) برآورد ماکسیمم درستنمایی (ML) را برای مدل رگرسیون ARMAی سری زمانی مورد استفاده قرار داده‌اند. گریفت (Griffth 1988) مدل ساده $y_{ij} = \mu + z_{ij}$ را مورد مطالعه قرار داد، که در آن همبستگی در هر دو جهت (i, j) یکسان فرض شده است. کرسی (Cressie 1993) برآورد مدلهای رگرسیونی با خطاهای همبسته فضایی را بطور مفصل مورد بررسی قرار داده است. باسو و راینسل (Basu & Reincel 1994) برآورد ML را برای مدل رگرسیونی با خطاهای خودهمبسته فضایی دو جهتی، که از مدل ARMAی مرتبه یک پیروی می‌کنند، مورد توجه قرار داده اند. شین و سرکار (Shin & Sarkar 1994) بر ورد ML را برای مدل رگرسیونی با خطاهای سری زمانی با مقادیر گمشده مورد استفاده قرار داده‌اند. شین و سونگ (Shin & Song 2000) کارایی مجانبی برآوردگر کمترین توان‌های دوم معمولی را برای مدل رگرسیونی با خطاهای همبسته فضایی بررسی کرده‌اند. راینسل و چانگ (Reinsel & Cheang 2003) برآورد تقریبی ML و REML را برای مدلهای رگرسیونی با خطاهای سری زمانی خودهمبسته (AR: Autoregressive) مرتبه اول بدست آورده‌اند. نظریه‌های مطرح شده در این مقالات با فرض ثابت بودن پارامترهای مدل بنا شده‌اند. اما در عمل موارد زیادی وجود دارد که پارامترها می‌توانند در شرایط مختلف مقادیر متفاوتی داشته باشند. در این وضعیت بکارگیری روشهای کلاسیک برای استنباط آماری در رگرسیون با فرض ثابت بودن پارامترها تناسب ندارد. بعلاوه وقتی پارامترها ثابت هستند بدلیل چند بعدی بودن تابع درستنمایی بکارگیری روشهای ML و REML با مسائل دشواری مواجه می‌گردد که برآورد پارامترها را ناممکن می‌سازد. برای رفع این مشکل بی‌سگ و گرین (Besage & Green 1993) و مه‌کولاک و تسای (McCulloch & Tsay 1994) رهیافت بیزی را برای آنالیز داده‌های فضایی و سری زمانی بکار برده‌اند. بی‌سگ و گرین (Besage & Green 1993) نحوه استفاده از روشهای مونت کارلوی زنجیر مارکوفی (MCMC: Markov Chain Monte Carlo) را برای برآورد پارامترهای مدل در تحلیل مجموعه داده‌های کشاورزی بیان کردند. مه‌کولاک و تسای (McCulloch & Tsay 1994) کاربردهای روش نمونه‌گیری گیبس (Gibbs Sampling)را در برآورد مدل سری زمانی AR مورد بررسی قرار دادند. اوه و همکاران (Oh *et al.* 2002) نیز رهیافت بیزی را برای مدل رگرسیونی با خطاهای خودهمبسته فضایی (SAR: Spatial

Autoregressive) با داده‌های گمشده در متغیر پاسخ مورد مطالعه قرار دادند. بدنبال آنها در این مقاله نحوۀ استفاده از روشهای MCMC خصوصاً الگوریتمهای نمونه‌گیری گیبس و متروپلیس-هاستینگس در تحلیل بیزی مدلهای رگرسیونی با خطاهای خودهمبسته فضایی مرتبه اول (SAR(1)) بر شبکه مستطیلی $\{(i,j)\,|\,i=1,2,...,m,\ j=1,2,...,n\}$ بصورت

$$y_{ij}=x'_{ij}\beta+z_{ij}\quad i=1,...,m,\ \ j=1,...,n \tag{۱}$$

مورد بررسی قرار می‌گیرد، که در آن

$$z_{ij}=\alpha_1 z_{i-1,j}+\alpha_2 z_{i,j-1}+\alpha_3 z_{i-1,j-1}+\varepsilon_{ij} \tag{۲}$$

و ε_{ij} ها خطاهای نرمال ناهمبسته با میانگین صفر و واریانس σ^2 و مستقل از x_{ij}ها هستند و $|\alpha_i|<1$ ، $i=1,2,3$ می‌باشد. برای سادگی محاسبات و کاهش بعد فضای پارامتری می‌توان با قرار دادن $\alpha_3=-\alpha_1\alpha_2$ خطاهای (۲) را بصورت

$$z_{ij}=\alpha_1 z_{i-1,j}+\alpha_2 z_{i,j-1}-\alpha_1\alpha_2 z_{i-1,j-1}+\varepsilon_{ij} \tag{۳}$$

در نظر گرفت، که در اینصورت (۳) مدل رگرسیون با خطاهای SAR مرتبه اول ضربی نامیده می‌شود. معمولاً در آمار فضایی ساختار همبستگی داده‌ها با استفاده از تابع کواریانس یا همتغییرنگار(Covariogram) در تجزیه و تحلیل داده‌ها لحاظ می‌گردد. در این مقاله که مدل رگرسیونی با خطاهای SAR (1) ضربی در نظر گرفته می‌شود، ساختار همبستگی خطاها توسط تابع همبستگی $\rho_{st}=Corr(z_{ij},z_{i-s,j-t})$ ، با ویژگی تقارن $\rho_{st}=\rho_{-s,-t}=\rho_{s,-t}=\rho_{-s,t}$ تعیین می‌شود. باسو و راینسل (Basu & Reincel 1993) خصوصیات این تابع را برای مدلهای SAR مورد بررسی قرار دادند و نحوه برآورد پارامترهای مدل به روشهای ML و REML در باسو و راینسل (Basu & Reincel 1994) ارائه گردیده است. اما در عمل بکارگیری این روشها پیچیده و طولانی است. بعلاوه پارامترهای مدل اغلب ثابت نیستند و در وضعیت‌های مختلف تغییر می‌کنند. با بکارگیری رهیافت بیزی برای برآورد پارامترهای مدل می‌توان ضمن استفاده از اطلاعات پیشین محدودیتهایی را روی پارامترها نیز منظور نمود. اما مشکل اصلی در تحلیل بیزی، تعیین توزیع پسین پارامترهای نامعلوم است که بدلیل نیاز به حل انتگرالهای چندگانه، کاری بس دشوار است. برای رفع این مشکل، می‌توان از الگوریتم نمونه‌گیری گیبس گلفاند و اسمیت (Gelfand & Smith 1990) استفاده نمود. این الگوریتم خصوصاً برای پیش‌بینی مفید است، زیرا فقط به توزیع شرطی کامل هر پارامتر نیاز دارد، که فرم تحلیلی آن اغلب به آسانی بدست می‌آید. چون بعضی از توزیع‌های پسین شرطی کامل پارامترها ممکن است فرم ساده‌ای نداشته باشد، در چنین مواردی نیز می‌توان از «الگوریتم متروپلیس- هاستینگس» (Metropolis-Hastings) (Hastings 1970) برای تولید نمونه از توزیع‌های شرطی کامل استفاده نمود. در ادامه برآورد پارامترهای مدل با استفاده از رهیافت بیزی ارائه و همچنین نحوه استفاده از الگوریتم‌های نمونه‌گیری گیبس و متروپلیس- هاستینگس شرح داده می‌شود.

برآورد بیزی پارامترهای مدل

برای تحلیل مدل رگرسیونی با خطاهای SAR(1) ضربی (۳) با قرار دادن

$$Y=(y_{11},y_{21},...,y_{m1},...,y_{1n},...,y_{mn})'$$
$$Z=(z_{11},z_{21},...,z_{m1},...,z_{1n},...,z_{mn})'$$
$$X=(x_{11},x_{21},...,x_{m1},...,x_{1n},...,x_{mn})'$$

تابع درستنمایی پارامترهای مدل بصورت

$$L(\beta,\sigma^2,\alpha_1,\alpha_2\,|\,y)=(2\pi\sigma^2)^{-mn/2}|\Gamma|^{-1/2}\exp[-0.5\sigma^{-2}h(\beta)'\Gamma^{-1}h(\beta)] \tag{۴}$$

خواهد شد، که در آن $h(\beta)=Y-X\beta$ و $\Gamma=\sigma^2 Var(Z)$ است. باسو و راینسل (Basu & Reincel, 1994) تحت فرض $\alpha_3=-\alpha_1\alpha_2$ نشان دادند:

$$|\Gamma|=(1-\alpha_1^2)^n(1-\alpha_2^2)^m \quad و \quad h(\beta)'\Gamma^{-1}h(\beta)=h^*(\beta)'h^*(\beta)$$

که درآن $h^*(\beta)=H'h(\beta)$ و $H=H_1\otimes H_2$ و

$$H_k=\begin{pmatrix}\sqrt{(1-\alpha_k^2)} & 0 & 0 & \cdots & 0\\ -\alpha_k & 1 & 0 & \cdots & 0\\ 0 & -\alpha_k & 1 & \cdots & 0\\ \cdots & \cdots & \cdots & \cdots & \cdots\\ 0 & 0 & 0 & \cdots & 1\end{pmatrix},\qquad k=1,2$$

و $\otimes$ ضرب کرونه‌کر است. همبستگی فضایی موجود در مشاهدات ارزیابی تحلیلی تابع درستنمایی (۴) را برای تعیین برآورد پارامترها بسیار دشوار می‌سازد. هر چند در این حالت روشهای عددی می‌توانند برای تقریب پارامترها مفید باشند، اما استفاده از رهیافت بیزی و بکارگیری روشهای MCMC شرایط بهتری را مهیا می‌سازد. برای این منظور، لازم است توزیع پیشین پارامترها تعیین شود. با فرض استقلال β ، σ^2 ، α_1 و α_2 ، توزیع پیشین توأم پارامترها بصورت

$$\pi(\beta,\sigma^2,\alpha_1,\alpha_2)=\pi(\beta)\pi(\sigma^2)\pi(\alpha_1)\pi(\alpha_2)$$

خواهد شد. در حالت کلی تعیین توزیع پسین پارامترها بصورت تحلیلی مقدور نمی‌باشد، ولی با انتخاب پیشین‌های مزدوج نرمال p متغیره $N_p(\beta^0,A)$ برای β ، گامای معکوس $IG(\nu,\tau)$ برای σ^2 ، که چگالی آن متناسب با $(\sigma^2)^{-(\nu+1)}\exp[-\frac{\tau}{\sigma^2}]$ است و توزیع نرمال بریده $N(\alpha_i^0,\psi_i^2)I(|\alpha_i|<1)$ برای α_i ، $i=1,2$ ، توزیع پسین شرطی کامل پارامترها بصورت

$$\beta \mid others \sim N[(\frac{1}{\sigma^2}X^{*'}X^* + A^{-1})^{-1}(\frac{1}{\sigma^2}X^{*'}Y + A^{-1}\beta^0), (\frac{1}{\sigma^2}X^{*'}X^* + A^{-1})^{-1}],$$

$$\sigma^2 \mid others \sim IG((\frac{mn}{2}+\nu), h^*(\beta)'h^*(\beta)+\tau),$$

$$\pi(\alpha_1 \mid others) \propto (1-\alpha_1^2)^{\frac{n}{2}}\phi(\alpha_1;\mu_1,\sigma_1^2)I(|\alpha_1|<1),$$

$$\pi(\alpha_2 \mid others) \propto (1-\alpha_2^2)^{\frac{m}{2}}\phi(\alpha_2;\mu_2,\sigma_2^2)I(|\alpha_2|<1),$$

بدست می‌آیند (Oh *et al.*, 2002)، که در هر یک از آنها others نشان دهنده بردار مشاهدات و سه پارامتر دیگر از مجموعه پارامترهای $\{\beta,\sigma_1^2,\alpha_1,\alpha_2\}$ است،

$$X^* = H'X, \qquad Y^* = H'Y,$$
$$\sigma_1^2 = (S_{11}/\sigma^2 + 1/\psi_1^2)^{-1}, \qquad \mu_1 = \sigma_1^2[(S_{11}/\sigma^2 + (1/\psi_1^2)\alpha_1^0],$$
$$\sigma_2^2 = (S_{21}/\sigma^2 + 1/\psi_2^2)^{-1}, \qquad \mu_2 = \sigma_2^2[(S_{22}/\sigma^2 + (1/\psi_2^2)\alpha_2^0],$$
$$S_{11} = \sum_{i=2}^{m-1} z_{i.}(\beta)'H_2H_2'z_{i.}(\beta), \qquad S_{21} = \sum_{i=2}^{m} z_{i.}(\beta)'H_2H_2'z_{i-1.}(\beta),$$
$$S_{21} = \sum_{j=2}^{n-1} z_{.j}(\beta)'H_1H_1'z_{.j}(\beta), \qquad S_{22} = \sum_{j=2}^{n} z_{.j}(\beta)'H_1H_1'z_{.j-1}(\beta),$$
$$z_{i.}(\beta) = [z_{i1}(\beta),\ldots,z_{in}(\beta)]', \qquad z_{.j}(\beta) = [z_{1j}(\beta),\ldots,z_{mj}(\beta)]',$$
$$z_{ij}(\beta) = y_{ij} - x_{ij}'\beta,$$

و $\phi(x;a,b)$ تابع چگالی توزیع $N(a,b)$ است. اگر پیشین‌های ناآگاهی بخش $\pi(\sigma^2) = \frac{1}{\sigma^2}$ و $\pi(\beta) = 1$ و $i=1,2,\quad \pi(\alpha_i) = I(|\alpha_i|<1)$ در نظر گرفته شوند، توزیع‌های پسین شرطی کامل باجایگذاری $\nu = \tau = \frac{1}{\psi_1} = \frac{1}{\psi_2} = 0$ و $A^{-1} = (0)_{p\times p}$ در روابط بالا بدست می‌آیند، که می‌توان نشان داد توزیع پسین متناظر با آنها سره می‌باشد (جعفری، ۲۰۰۴). همانطور که ملاحظه می‌شود، توزیع پسین شرطی کامل β و σ^2 دارای فرم مشخصی هستند. بنابراین تولید مقادیر تصادفی از آنها با استفاده از الگوریتم نمونه‌گیری گیبس بعنوان یکی از روش‌های MCMC بصورت تکراری و به روش زیر قابل انجام است.

برای مقدار شروع داده شده $(\beta^{(t)},\sigma^{2(t)},y_{ij}^{(t)})$، مقادیر پارامترها در تکرار $(t+1)$ام الگوریتم در سه مرحله بصورت زیر تولید می‌شوند:

$$i)\ \beta^{(t+1)} \sim f_1(\beta \mid \sigma^{2(t)}, y_{ij}^{(t)})$$
$$ii)\ \sigma^{2(t+1)} \sim f_2(\sigma^2 \mid \beta^{(t+1)}, y_{ij}^{(t)})$$
$$iii)\ y_{ij}^{(t+1)} \sim f_1(y_{ij} \mid \beta^{(t+1)}, \sigma^{2(t+1)})$$

به خاطر همگرایی نمونه‌گیری گیبس و عدم وابستگی به مقادیر اولیه معمولاً تکرارهای اول نمونه‌ها کنار گذاشته می‌شوند. یعنی، برای تولید یک نمونه M تایی از هر پارامتر، $M+M_0$ مقدار تولید و M_0 تای اول آن که تکرارهای سوخته (Burn-in) نام دارند، کنار گذاشته می‌شوند. چون چگالی‌های پسین شرطی کامل α_1 و α_2 فرم مشخصی ندارند، تولید نمونه از توزیع آنها با الگوریتم متروپلیس-هاستینگس امکان پذیر است. برای این منظور ابتدا یک تابع چگالی هدف $g(\alpha_i)$ بعنوان تقریبی از پسین $\pi(\alpha_i \mid others)$، $i=1,2$ انتخاب می‌شود. سپس برای مقدار شروع داده شده $\alpha_i^{(t)}$ یک نمونه نامزد α_i^* از چگالی $g(\alpha_i)$ را تولید نموده و با قرار دادن $w(\alpha_i) = \pi(\alpha_i \mid others)/g(\alpha_i)$، مقدار $\alpha_i^{(t+1)}$ بصورت زیر تعیین می‌شود، که در آن t تکرار tام الگوریتم است.

$$\alpha_i^{(t+1)} = \begin{cases} \alpha_i^* & with\ probability\ \min\{1, \frac{w(\alpha_i^*)}{w(\alpha_i^{(t)})}\} \\ \alpha_i^t & with\ probability\ 1-\min\{1, \frac{w(\alpha_i^*)}{w(\alpha_i^{(t)})}\} \end{cases}$$

کارایی این الگوریتم وابسته به تولید نمونه از چگالی $g(\alpha_i)$ است. اگر $|\alpha_1|$ بزرگ نباشد و n به اندازه کافی بزرگ باشد، آنگاه داریم

$$\begin{aligned}\pi(\alpha_1 \mid others) &\propto (1-\alpha_1^2)^{\frac{n}{2}}\phi(\alpha_1;\mu_1,\sigma_1^2)I(|\alpha_1|<1)\\ &\approx \exp\left\{-\frac{n}{2}\alpha_1^2\right\}\phi(\alpha_1;\mu_1,\sigma_1^2)I(|\alpha_1|<1)\\ &\propto \phi(\alpha_1;0,\frac{1}{n})\phi(\alpha_1;\mu_1,\sigma_1^2)I(|\alpha_1|<1)\\ &\propto \phi(\alpha_1;\mu_1/\{n\sigma_1^2+1\},\sigma_1^2/\{n\sigma_1^2+1\})I(|\alpha_1|<1)\end{aligned}$$

اکنون با استفاده از توزیع نرمال بریده

$$\phi(\alpha_1;\mu_1/\{n\sigma_1^2+1\},\sigma_1^2/\{n\sigma_1^2+1\})I(|\alpha_1|<1)$$

می‌توان نمونه‌ای از چگالی $g(\alpha_1)$ برای α_1 تولید نمود. بطور مشابه به کمک توزیع نرمال بریده

$$\phi(\alpha_2;\mu_2/\{m\sigma_2^2+1\},\sigma_2^2/\{m\sigma_2^2+1\})I(|\alpha_2|<1)$$

می‌توان نمونه‌ای از چگالی $g(\alpha_2)$ برای α_2 تولید نمود. برای تولید نمونه از توزیع نرمال یک متغیره بریده، الگوریتم‌های کارای زیادی مانند روش cdf معکوس (Devroye 1986) و الگوریتم ردی آمیخته (Mixed Rejected Algorithm) (Geweke 1990) وجود دارند. با شروع از مقدارهای اولیهٔ $\{\beta,\sigma_1^2,\alpha_1,\alpha_2\}$، نمونه‌گیر گیبس، نمونه‌های تکراری از توزیع‌های پسین شرطی کامل تولید می‌کند و نمونه‌های حاصل از کنار گذاشتن تکرارهای سوخته می‌توانند برای برآورد پارامترها و پیش‌بینی استفاده شوند. با توجه به اینکه نمونه‌های قبلی برای تولید نمونه‌های جدید استفاده می‌شود، نمونه‌های تولید شده از روش MCMC ممکن است خودهمبسته باشند. یک روش برای بدست آوردن نمونه‌های مستقل، گرفتن نمونه‌هایی از هر l تکرار (l اندازه تأخیر) است. اکنون بر اساس الگوریتم‌های MCMC نمونه‌ای به اندازه M را برای هر یک از پارامترهای مدل تولید نموده و با قرار دادن $(\theta_{1k},\theta_{2k},\theta_{3k},\theta_{4k}) = \{\beta_k,\sigma_k^2,\alpha_{1k},\alpha_{2k}\}$ بعنوان k-امین عضو نمونه تولید شده، میانگین پسین θ_i را می‌توان بصورت

$$\hat{\theta}_i = \hat{E}(\theta_i \mid y) = \frac{1}{M}\sum_{k=1}^{M}\theta_{ik}, \quad i = 1,2,3,4 \qquad (۵)$$

و خطای استاندارد آنرا نیز بصورت

$$M\hat{S}E(\theta_i) = (\sum_{k=1}^{M}\theta_{ik}{}^2 - M\hat{\theta}_i{}^2)^{\frac{1}{2}} / M, \quad i = 1,2,3,4$$

برآورد نمود. مقادیر شروع الگوریتم‌ها را برای هر پارامتر می‌توان به صورت زیر تعیین نمود:

۱- برآورد کمترین توان‌های دوم معمولی $\hat{\beta}_0 = (x'x)^{-1}(x'y)$ بعنوان مقدار آغازین β انتخاب شود.

۲- برآورد کمترین توان‌های دوم معمولی $\hat{\sigma}_0^2$ بعنوان مقدار آغازین σ^2 در نظر گرفته شود.

۳- بوسیله رگرسیون $z_{ij}(\hat{\beta}_0)$ روی $Z_{i-1,j}(\hat{\beta}_0)$، $Z_{i,j-1}(\hat{\beta}_0)$، $Z_{i-1,j-1}(\hat{\beta}_0)$ برای (i, j)، برآوردهای $\hat{\alpha}_1$ و $\hat{\alpha}_2$ بعنوان مقادیر آغازین α_1 و α_2 به کار گرفته شود.

مثال

برای نمایش کارایی، مطلوبیت و راحتی استفاده از روشهای MCMC در برآورد بیزی پارامترهای مدل رگرسیون فضایی با خطاهای خودهمبسته ضربی، مدل $Y = X\beta + Z$ را در نظر بگیرید، که در آن خطای Z از مدل SAR(1) ضربی پیروی می‌کند. ارزیابی روشهای مطروحه با استفاده از معیار میانگین توان دوم خطای برآورد پارامترها که بر اساس داده‌های شبیه‌سازی شده حاصل می‌شوند، انجام می‌پذیرد. برای این منظور، n مقدار برای $X = (1, X_1, X_2)$ که در آن ۱ یک بردار ستونی واحد است و X_1 و X_2 بترتیب از توزیعهای $N(5,1)$ و $N(10,1)$ پیروی می‌کنند با استفاده از توابع تولید اعداد تصادفی نرم‌افزار S-plus شبیه‌سازی می‌شوند. سپس مقادیر اولیه $z_{00} = z_{10} = z_{01} = 0.07$ و $\alpha_1 = \alpha_2 = 0.1$ را اختیار نموده و با تولید n مقدار ε_{ij} از توزیع $N(0,\sigma^2)$ بر یک شبکه $n \times m$ و جایگذاری در مدل (۳) مقادیر خطاهای Z ایجاد می‌شوند. اکنون برای مقادیراختیاری $\beta = (10, 2, 1)$ بردار مشاهدات Y از رابطه $Y = X\beta + Z$ محاسبه می‌شوند. با انتخاب پیشین‌های ناآگاهی بخش

$$\pi(\beta,\sigma^2,\alpha_1,\alpha_2) = \pi(\beta)\pi(\sigma^2)\pi(\alpha_1)\pi(\alpha_2) = \frac{1}{\sigma^2} I(|\alpha_1| < 1)$$

و استفاده از الگوریتم‌های نمونه‌گیری گیبس و متروپلیس – هاستینگس بطور همزمان نمونه‌هایی از توزیع‌های شرطی کامل پارامترها تولید می‌شوند. مقادیر اولیه برای α_1 و α_2، صفر و برای β و σ^2 برآورد OLS آنها در نظر گرفته شده و برآورد پارامترها براساس رابطه (۵) صورت می‌پذیرد. این عمل را ۱۰۰۰ بار تکرار نموده میانگین توان دوم خطاهای برآورد (MSE) هریک از پارامترها بصورت $MSE(\hat{\theta}_i) = \sum_{k=1}^{M}(\hat{\theta}_{ik} - \theta_{ik})^2 / M$، که در آن M حجم نمونه حاصل از کنار گذاشتن مقادیر سوخته است، محاسبه می‌شود. جدول۱، برآورد پارامترها را برای یک شبکه ۵×۵ با اندازه‌های نمونه ۵، ۱۰، ۱۰۰، ۹۰۰ و ۹۷۰ برای M و متناظر با مقادیر سوخته ۳۰، ۱۰۰، ۹۰۰ و ۹۹۵ برای M_0 نشان می‌دهد. جدول۲ مقادیر MSE برآورد پارامترها را برای نمونه‌های مختلف نشان می‌دهد. همانطور که ملاحظه می‌شود با افزایش حجم نمونه مقدار MSE برای پارامترهای مدل کاهش می‌یابد. یعنی با افزایش حجم نمونه برآوردهای بهتری بدست می‌آوریم. حال اندازه شبکه را ۸×۸ در نظر گرفته و نتایج حاصل برآورد پارامترها و MSE آنها در جداول ۳ و ۴ خلاصه شده‌اند. از جدول ۴ می‌توان دریافت که مقادیر MSE پارامترها نسبت به حالت قبل کاهش یافته و برآوردهای بهتری برای پارامترهای مدل بدست آمده است که بیانگر افزایش دقت برآوردها با افزایش اندازهٔ شبکه است. با توجه به تفاوت کم مقادیر MSE برای نمونه‌های انتهایی الگوریتم و نمونه‌های بزرگ ($M =$ ۹۷۰)، و از طرفی کاهش چشمگیر زمان محاسبه برآورد پارامترها در حالتهای ۱۰ یا ۵ $= M$ ، توصیه می‌شود به منظور سرعت بخشیدن محاسبات، تعداد تکرار الگوریتم‌ها ($M + M_0$) و مقادیر سوخته (M_0) بزرگ اختیار شوند، بگونه‌ای که M حدود ۳۰ باشد، زیرا دنباله سری حاصل از الگوریتم‌های نمونه‌گیری گیبس و متروپلیس – هاستینگس برای پارامترهای α_1و α_2 و σ^2 و β بعد از ۳۰ تکرار اول همگرا می‌شوند.

جدول ۱- برآورد بیزی پارامترها برای نمونه های مختلف در شبکه ۵×۵

M	$\hat{\sigma}^2$	$\hat{\beta}_2$	$\hat{\beta}_1$	$\hat{\beta}_0$	$\hat{\alpha}_2$	$\hat{\alpha}_1$
۵	۱/۵۴۱۹۹	۰/۶۷۴۴۷	۱/۷۹۲۸۴	۱۴/۴۹۲۰۱	۰/۰۳۹۱۶	۰/۰۴۳۴۱
۱۰	۱/۶۲۹۶۷	۰/۶۹۳۶۵	۱/۸۰۷۳۵	۱۴/۲۷۵۳۹	۰/۰۴۴۴۸	۰/۰۴۴۵۴۷
۱۰۰	۱/۶۷۸۶۶	۰/۷۱۸۵۲	۱/۸۳۵۸۸	۱۴/۲۲۵۴۴	۰/۰۴۵۲۴	۰/۰۵۱۰۹
۹۰۰	۱/۶۸۷۵۲	۰/۶۸۸۷۵	۱/۸۳۹۶۰	۱۴/۱۸۹۳۱	۰/۰۶۰۷۱	۰/۰۵۸۴۴
۹۷۰	۱/۶۹۲۱۹	۰/۶۹۸۹۱	۱/۸۳۷۱۰	۱۴/۱۷۱۵۷	۰/۰۶۹۰۵	۰/۰۵۸۶۷

جدول ۲- MSE برآوردهای بیزی پارامترها برای نمونه های مختلف در شبکه ۵×۵

$MSE\hat{\alpha}_1$	$MSE\hat{\alpha}_2$	$MSE\hat{\beta}_0$	$MSE\hat{\beta}_1$	$MSE\hat{\beta}_2$	$MSE\hat{\sigma}^2$	M
۰/۰۰۸۴۴	۰/۰۰۶۶۷	۲۰/۱۷۸۲۳	۰/۰۴۲۹۱	۰/۱۰۵۹۷	۰/۵۹۳۷۶	**۵**
۰/۰۰۳۰۷	۰/۰۰۳۳۱	۱۸/۲۷۹۰۲	۰/۰۳۷۱۱	۰/۰۹۳۸۴	۰/۴۹۶۴۸	**۱۰**
۰/۰۰۲۳۹	۰/۰۰۲۹۹	۱۷/۹۶۶۹۹	۰/۰۲۶۹۳	۰/۰۹۹۲۳	۰/۴۸۰۵۸	**۱۰۰**
۰/۰۰۱۷۲	۰/۰۰۱۵۴	۱۷/۵۵۰۲۸	۰/۰۲۵۷۲	۰/۰۹۶۸۷	۰/۴۷۹۶۸	**۹۰۰**
۰/۰۰۱۷۱	۰/۰۰۱۴۷	۱۷/۵۵۰۳۹	۰/۰۲۴۵۳	۰/۰۹۶۷۷	۰/۴۷۲۱۳	**۹۷۰**

جدول۳- برآورد بیزی پارامترها برای نمونه های مختلف در شبکه ۸×۸

$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}$	$\hat{\sigma}^2$	M
۰/۰۵۴۱۸	۰/۰۶۷۷۳	۱۱/۱۵۳۸۲	۱/۸۳۷۵۹	۰/۹۸۷۸۴	۲/۰۷۰۸۱	**۵**
۰/۰۵۵۸۶	۰/۰۶۰۹۹	۱۱/۲۵۳۱۱	۱/۸۳۶۵۱	۰/۹۷۷۴۲	۱/۹۳۷۵۱	**۱۰**
۰/۰۶۲۱۱	۰/۰۶۶۸۴	۱۰/۵۲۹۹۹	۱/۸۷۹۹۰	۱/۰۲۰۰۵	۱/۹۵۵۲۴	**۱۰۰**
۰/۰۶۸۹۶	۰/۰۷۱۱۵	۱۰/۳۹۶۱۳	۱/۸۷۵۰۹	۱/۰۳۷۱۵	۱/۹۸۲۸۳	**۹۰۰**
۰/۰۶۹۲۸	۰/۰۷۱۹۶	۱۰/۳۸۹۰۸	۱/۸۷۴۹۴	۱/۰۳۷۹۸	۱/۹۸۲۴۶	**۹۷۰**

جدول ۴- MSE برآوردهای بیزی پارامترها برای نمونه های مختلف در شبکه ۸×۸

$MSE\hat{\alpha}_1$	$MSE\hat{\alpha}_2$	$MSE\hat{\beta}_0$	$MSE\hat{\beta}_1$	$MSE\hat{\beta}_2$	$MSE\hat{\sigma}^2$	M
۰/۰۰۲۱۰	۰/۰۰۱۰۴	۱/۳۳۱۳۰	۰/۰۲۶۳۸	۰/۰۰۵۱۰	۱/۱۴۶۶۲	**۵**
۰/۰۰۱۹۵	۰/۰۰۱۵۲	۱/۵۷۰۲۸	۰/۰۲۶۷۳	۰/۰۰۱۵۰	۰/۸۷۸۹۴	**۱۰**
۰/۰۰۱۴۴	۰/۰۰۱۱۰	۰/۲۸۰۸۹	۰/۰۱۴۴۲	۰/۰۰۱۴۰	۰/۹۱۲۴۵	**۱۰۰**
۰/۰۰۰۹۶	۰/۰۰۰۸۳	۰/۱۵۶۹۱	۰/۰۱۵۶۰	۰/۰۰۱۳۸	۰/۹۶۵۹۶	**۹۰۰**
۰/۰۰۰۹۴	۰/۰۰۰۷۹	۰/۱۵۱۳۸	۰/۰۱۵۶۰	۰/۰۰۱۳۴	۰/۹۶۵۲۳	**۹۷۰**

بحث و نتیجه‌گیری

آنالیز بیزی مدلهای رگرسیونی با خطاهای خودهمبسته فضایی ضربی مرتبه اول مطرح و برای محاسبه انتگرال‌ها در استنباط پسین، از الگوریتم‌های نمونه‌گیری گیبس و متروپلیس-هاستینگی که از روشهای MCMC هستند، استفاده شد و در یک مطالعه شبیه‌سازی برآورد بیزی پارامترها محاسبه شدند. با استفاده از ملاک MSE نشان داده شد با افزایش حجم نمونه و اندازه شبکه برآوردهای دقیقتری بدست می‌آیند. با بکارگیری نمونه‌های انتهایی تکرار الگوریتم‌های MCMC برای حصول برآوردهای با دقت لازم کفایت می‌کند. چون داده‌های فضایی نزدیک بهم دارای وابستگی قویتری هستند، استفاده از مدل خودهمبسته فضایی مرتبه اول برای لحاظ نمودن همبستگی داده‌ها از توجیه کافی برخوردار است. اما مدلهای خودهمبستهٔ مرتبه دوم و بالاتر را نیز می‌توان بکار گرفت که در این صورت هرچند وابستگی دو یا چند تأخیر قبل هم وارد مدل می‌شود، اما انجام آن نیاز به مطالعه بیشتر دارد.

تشکر و قدردانی

نویسندگان از نظرات و پیشنهادات اصلاحی داوران محترم مجله که موجب بهبود این مقاله گردید و همچنین از حمایت قطب علمی داده‌های ترتیبی و فضایی دانشگاه فردوسی مشهد کمال تشکر و قدردانی را دارند.

منابع:

Basu S., Reinsel G.C. 1993: Properties of the Spatial Unilateral First-Order ARMA Model, *Advances in Applied Probability*, **25**:

631-634.

Basu S., Reinsel G.C. 1994: Regression Models with Spatially Correlated Errors, *J. Amer. Statist. Assoc*, **89**: 88-99.

Besag J.E., Green P.J. 1993: Spatial Statistics and Bayesian Computation with Discussions. *J. R. Statist. Soc*, **55**: 25-38.

Cressie N. 1991: Statistical for Spatial Data. Wiley, New York.

Devroye L. 1986: Non-uniform Random Variate Generation. Springer, New York.

Gelfand A.E., Smith A.F. M. 1990: Sampling-Based Approaches to Calculating Marginal Densities, *J. Amer. Statist. Assoc*, **85**: 398-409.

Geweke M. 1991: Efficient Simulation from the Multivariate Normal and Student t-Distributions Subject to Linear Constraints. In: E. M. Keramidas(Ed.), *Computing Science and Statistics*, Proceedings of the 23rd Symposium on the Interface, 571-578. Fairfax,VA: Interface Foundation of North America.

Griffth D.A. 1988: Advanced Spatial Statistics. Spatial Topics in the Exploration of Quantitative Spatial Data Series. Kluwer Academic Publishers, Dordrecht.

Hastings W.K. 1970: Monte Carlo Sampling Methods Using Markov Chains and their Applications, *Biometrika*, **57:** 97-109.

Jafari K. M. 2004: Spatial Prediction with Bayesian Approach, PhD Thesis, Department of Statistics, Tarbiat Modares University.

McCulloch R.E., Tsay R.S. 1994: Bayesian Analysis of Autoregressive Time Series Via the Gibbs Sampler, *J. Time Series Anal*, **15**: 235-250.

Oh M.S., Shin D.W., Kim H.J. 2002: Bayesian Analysis of Regression Models with Spatially Correlated Errors and Missing Observations, *Computational Statistics and Data Analysis*, **39**: 387-400.

Reinsel G.C., Cheang W.K. 2003: Approximate ML and REML Estimation for Regression Models with Spatial or Time Series AR(1) Noise. *Statistics and Probability Letters*, **62:** 123-135.

Shin D.W., Song S.H. 2000: Asymptotic Efficiency of the OLSE for Polynomial Regression Models with Spatially Correlated Errors, *Statistics & Probability Letters*, **47**: 1-10.

Shin D.W., Sarkar S. 1994: Parameter Estimation in Regression Models with Autocorrelated Errors, Using Irregular Data, *Comm. Statist. Theory Methods*, **23**: 3567-3580.

Wincek M.A. Reincel G.C. 1986: An Exact Maximum Likelihood Estimation Procedure for Regression ARMA Time Series Models with Possibly Nonconsecutive Data. *J. R. Statist. Soc*, **48**: 303-31.