مجله علوم آماری، پاییز و زمستان ۱۳۹۰
جلد ۵، شماره ۲، ص ۱۸۹-۲۰۲

مدل‌های اتورگرسیو فضایی و تحلیل داده‌های معاملات مسکونی شهر تهران

حمیدرضا رسولی
گروه آمار، دانشگاه تربیت مدرس

تاریخ دریافت: ۱۳۹۰/۷/۲ تاریخ آخرین بازنگری: ۱۳۹۰/۱۲/۲۵

چکیده : در این مقاله انواع مدل‌های اتورگرسیو برای تحلیل داده‌های فضایی بیان شده و پارامترهای مدل‌ها با ماکسیمم کردن تابع درستنمایی نیمرخ با فرض آن که بین متغیرهای وابسته یا خطاهای مدل رابطه اتورگرسیو فضایی برقرار باشد، برآورد شده است. سپس مدل‌های مختلف مورد ارزیابی قرار گرفته و در انتها نحوه کاربست آن‌ها در مثالی کاربردی نشان داده شده است.

واژه‌های کلیدی : مدل تأخیر فضایی، مدل خطافضایی، آزمون موران، ماتریس وزن فضایی

۱ مقدمه

داده‌هایی که از لحاظ موقعیت قرار گرفتن به یکدیگر وابستگی دارند داده‌های فضایی نامیده می‌شوند. یکی از روش‌های مدل‌بندی داده‌های فضایی، مدل‌های اتورگرسیو فضایی هستند که با در نظر گرفتن ضریب اتورگرسیو فضایی وابستگی فضایی در مدل لحاظ می‌شود. اولین بار ویتل (۱۹۵۴) با فرض اینکه متغیرهای

آدرس الکترونیک مسئول مقاله: حمیدرضا رسولی، h.rasouli@modares.ac.ir
کد موضوع‌بندی ریاضی (۲۰۱۰): ۹۱H۷۲ و ۶۲H۱۱

پاسخ و تبیینی بر روی یک مستطیل توری مشاهده شده و خطاهای هر موقعیت فقط وابسته به موقعیت‌های قبلی، بعدی، بالا و پایین باشند، وابستگی فضایی را در یک مدل اتورگرسیو لحاظ کرد. بسج (۱۹۷۴)، اورد (۱۹۷۵)، انسلین (۱۹۸۸)، هاینینگ (۱۹۹۰)، کرسی (۱۹۹۳) و والر و گاتوی (۲۰۰۴) نیز به تحلیل مدل‌بندی داده‌های فضایی، با استفاده از مدل‌های اتورگرسیو پرداختند. صفایی (۱۳۹۰) با استفاده از مدل‌های اتورگرسیو تبدیلی مارکف به تحلیل داده‌های نرخ ارز ایران پرداخت. انسلین (۱۹۸۸) نشان داد برآورد پارامترهای مدل اتورگرسیو همزمان (SAR) به روش کمترین توان‌های دوم خطا اریب و ناسازگارند و برآورد پارامترهای مدل را با ماکسیمم کردن تابع درستنمایی‌نما[۱] ارائه کرد. انسلین و لسج (۱۹۹۹) از درستنمایی نیمرخ[۲] برای برآورد پارامترها استفاده نمودند، که با جای‌گذاری برآورد تعدادی از پارامترها برحسب یکی از پارامترها در تابع درستنمایی، تعداد پارامترهای مسئله را کاهش داده، سپس با ماکسیمم کردن تابع درستنمایی نیمرخ بر اساس آن پارامتر و جای‌گذاری مقدار آن در تابع درستنمایی برآورد سایر پارامترها به‌دست آورده می‌شود. لی (۲۰۰۴) نشان داد برآوردهای حاصل از ماکسیمم کردن تابع درستنمایی نیمرخ نیز سازگار هستند. پیس و باری (۱۹۹۷) آزمون نسبت درستنمایی را برای معنی‌دار بودن ضرایب اتورگرسیو پیشنهاد دادند. اسمیرنوو (۲۰۰۵) از ماتریس اطلاع برای آزمون معنی‌دار بودن ضرایب استفاده کردند. لسج و پیس (۲۰۰۹) برآورد بیزی پارامترهای مدل‌های اتورگرسیو ارائه کردند و نشان دادند توزیع پسین پارامترها نیز ناسره هستند. در این مقاله انواع مدل‌های اتورگرسیو فضایی و آزمون موران برای تشخیص خودهمبستگی فضایی معرفی شده، سپس برآورد پارامترها با استفاده از تابع درستنمایی نیمرخ ارائه می‌شود. در انتها نحوه کاربرد این مدل‌ها در تحلیل داده‌های معاملات مسکونی شهر تهران نشان داده شده و با روش اعتبار سنجی متقابل و ارائه ضریب تعیین مدل‌ها مورد ارزیابی قرار می‌گیرند.

---

۱ Pseudo Likelihood

۲ Profile Likelihood

۲ مدل اتورگرسیو فضایی

مدل اتورگرسیو فضایی به‌صورت

$$Y = \rho W_1 Y + X\beta + U, \ \ U = \lambda W_2 U + \epsilon, \ \ \epsilon \sim N(0, \sigma^2 I), \qquad (1)$$

تعریف می‌شود، که در آن $Y = (y_1, \cdots, y_n)$ بردار متغیرهای پاسخ، $X_{n\times k}$ ماتریس طرح متشکل از k متغیر تبیینی، ρ و λ ضرایب مدل اتورگرسیو، β بردار $k \times 1$ ضرایب رگرسیونی، W_1 و W_2 ماتریس‌های وزن فضایی با بعد $n \times n$ است. ماتریس وزن فضایی را می‌توان براساس طول و عرض جغرافیایی یا مجاورت موقعیت داده‌ها تعیین کرد. در ماتریس مجاورت عناصر w_{ij} برای هر دو مشاهدهٔ همسایهٔ i و j مقدار مثبت و برای مشاهدات غیر همسایه مقدار صفر اختیار می‌کند. به علاوه $w_{ii} = 0$، یعنی وزن همسایگی هر مشاهده با خودش صفر در نظر گرفته می‌شود. در ماتریس مجاورت می‌توان وزن‌ها را به‌صورت استاندارد شده که از تقسیم وزن‌های در سطر بر مجموع آن سطر به‌دست می‌آید، استفاده کرد. در این‌صورت مجموع عناصر هر سطر برابر یک است و WY میانگین موزون مشاهدات خواهد شد. درایه‌های ماتریس وزن فاصله نیز معمولاً براساس معکوس فاصله نقاط از یکدیگر به‌صورت $w_{ij} = d_{ij}^{-\alpha}$ تعیین می‌شوند، که در آن $\alpha > 0$ و d_{ij} اندازه فاصله دو مشاهده در موقعیت‌های i و j با مختصات جغرافیایی (x_i, y_i) و (x_j, y_j) هستند. دو موقعیت i و j همسایه نامیده می‌شوند اگر فاصله بین آن‌ها کمتر از یک حد مشخص باشد (اربیا، ۲۰۰۶).

۱.۲ مدل اتورگرسیو همزمان

با قرار دادن $X = 0$ و $W_2 = 0$ در مدل کلی (۱)، مدل اتورگرسیو فضایی همزمان[۳] (SAR) به‌صورت

$$Y = \rho W_1 Y + \epsilon, \ \ \epsilon \sim N(0, \sigma^2 I),$$

[۳] Simultaneous Autoregresive

حاصل می‌شود، که در آن تغییرات Y برحسب ترکیب خطی از همسایگی‌ها است. برای برآورد ماکسیمم درستنمایی پارامترهای مدل SAR، چون درایه‌های ϵ دارای توزیع $N(\circ, \sigma^{۲})$ و مستقل هستند، لگاریتم تابع درستنمایی به‌صورت

$$\ell(\sigma^{۲}; U) = -n/۲ \ln(۲\pi\sigma^{۲}) - \frac{۱}{\sigma^{۲}}\epsilon'\epsilon$$

است. با تغییر متغیر $\epsilon = (I - \rho W_{۱})y$، لگاریتم تابع درستنمایی به‌صورت

$$\ell(\rho, \sigma, y) = -n/۲ \ln(۲\pi\sigma^{۲}) + \ln|I - \rho W_{۱}| - \frac{۱}{\sigma^{۲}}(y - \rho W_{۱}y)'(y - \rho W_{۱}y) \quad (۲)$$

حاصل می‌شود، که به خاطر پیچیده بودن، تحلیل مستقیم آن میسر نیست. انسلین و لسج (۱۹۹۹) استفاده از درستنمایی نیمرخ را برای برآورد پارامترها پیشنهاد کردند، که با جای‌گذاری مقدار

$$\hat{\sigma^{۲}} = \frac{۱}{n}\{(y - \rho W_{۱}y)'(y - \rho W_{۱}y)\} \quad (۳)$$

در (۲) به‌صورت

$$\ell(\rho, y) = -\frac{n}{۲}\ln\{(y - \rho W_{۱}y)'(y - \rho W_{۱}y)\} + \ln|I - \rho W_{۱}| \quad (۴)$$

حاصل می‌شود. حال با روش‌های بهینه‌سازی معادله (۴) نسبت به ρ ماکسیمم می‌شود و با جای‌گذاری مقدار حاصل برای ρ در عبارت (۳)، برآورد $\sigma^{۲}$ به‌دست آورده می‌شود. با توجه به اینکه $(I - \rho W_{۱})Y = \epsilon$، انسلین (۱۹۸۸) پیشگویی با مدل اتورگرسیو همزمان را به‌صورت $\hat{Y} = (I - \hat{\rho} W_{۱})^{-۱}\hat{\epsilon}$ نشان داد، که در آن $\hat{\epsilon} = Y - \hat{\rho} W_{۱} Y$.

۲.۲ مدل تأخیر فضایی

اگر در مدل (۱)، $W_{۲} = \circ$ قرار داده شود، مدل آمیخته رگرسیون اتورگرسیو فضایی به‌صورت

$$Y = \rho W_{۱}Y + X\beta + \epsilon, \quad \epsilon \sim N(\circ, \sigma^{۲}I),$$

به دست می‌آید، که مدل تأخیر فضایی[۴] نامیده می‌شود. لگاریتم تابع درستنمایی برای مدل رگرسیون تأخیر فضایی به‌صورت

$$\begin{aligned}\ell(\beta,\rho,\sigma^۲,y) &= \ln c(y) - \frac{n}{۲}\ln\sigma^۲ + \ln|I-\rho W_۱| \\ &- \frac{۱}{۲\sigma^۲}[(I-\rho W_۱)y - X\beta]'[(I-\rho W_۱)y - X\beta].\end{aligned}$$

است. با استفاده از تابع درستنمایی نیمرخ برآورد پارامترها در مراحل زیر به‌دست آورده می‌شوند:

مرحله اول: یک مدل رگرسیونی به‌فرم

$$y = X\beta_۰ + \epsilon_۰ \qquad (۵)$$

برازش داده و $\beta_۰$ به روش کمترین توان‌های دوم به‌صورت $\hat{\beta}_۰ = (X'X)^{-۱}X'y$ برآورد می‌شود.

مرحله دوم: یک مدل رگرسیونی به‌فرم

$$W_۱y = X\beta_L + \epsilon_L \qquad (۶)$$

برازش داده و β_L به روش کمترین توان‌های دوم به‌صورت $\hat{\beta}_L = (X'X)^{-۱}X'W_۱y$ برآورد می‌شود.

مرحله سوم: باقی مانده‌های دو مدل (۵) و (۶) به‌صورت

$$\hat{\epsilon}_۰ = y - X\hat{\beta}_۰ \quad \hat{\epsilon}_L = W_۱y - X\hat{\beta}_L \qquad (۷)$$

محاسبه می‌شوند.

مرحله چهارم: با استفاده از مانده‌های (۷) لگاریتم تابع درستنمایی به‌صورت

$$\ell(\rho;\hat{\epsilon}_۰,\hat{\epsilon}_L) = c(\hat{\epsilon}_۰,\hat{\epsilon}_L) - \frac{n}{۲}\ln[\frac{۱}{n}(\hat{\epsilon}_۰-\rho\hat{\epsilon}_L)'(\hat{\epsilon}_۰-\rho\hat{\epsilon}_L)] + \ln|I-\rho W_۱| \qquad (۸)$$

بازنویسی می‌شود و از ماکسیمم کردن آن برآورد ρ حاصل می‌شود.

مرحله پنجم: با استفاده از $\hat{\rho}$، پارامترهای $\sigma^۲$ و β به‌صورت

$$\hat{\sigma}^۲ = \frac{۱}{n}(\hat{\epsilon}_۰-\hat{\rho}\hat{\epsilon}_L)'(\hat{\epsilon}_۰-\hat{\rho}\hat{\epsilon}_L) \quad \hat{\beta} = (\hat{\beta}_۰ - \hat{\rho}\hat{\beta}_L)$$

[۴] Spatial Lag Model

برآورد می‌شوند. هاینینگ (۱۹۹۰) نشان داد چون $(I-\rho W_1)y = X\beta+\epsilon$، در نتیجه $\hat{y} = (I-\hat{\rho}W_1)^{-1}X\hat{\beta}$، بنابراین $\hat{\rho}W_1\hat{y} = \hat{\rho}W_1(I-\hat{\rho}W_1)^{-1}X\hat{\beta}$ و پیشگویی عبارتست از

$$\hat{y} = X\hat{\beta} + \hat{\rho}W_1\hat{y} = X\hat{\beta} + \hat{\rho}W_1(I-\hat{\rho}W_1)^{-1}X\hat{\beta}$$

۳.۲ مدل خطافضایی

اگر در مدل (۱)، $W_1 = 0$ قرار داده شود، مدل رگرسیون با خطاهای خودهمبسته فضایی به‌صورت

$$Y = X\beta + U,\ \ U = \lambda W_2 U + \epsilon,\ \ \epsilon \sim N(0, \sigma^2 I), \qquad (9)$$

حاصل می‌شود، که مدل خطافضایی[5] نامیده می‌شود. برای مدل خطا فضایی (۹) با در نظر گرفتن تابع درستنمایی

$$L(\sigma^2, \epsilon) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left(\frac{-1}{\sigma^2}\epsilon'\epsilon\right)$$

و قرار دادن $\epsilon = (I-\rho W_2)(y-X\beta)$، لگاریتم تابع درستنمایی به‌صورت

$$\ell(\beta, \rho, \sigma^2, y) = -\frac{n}{2}\ln(\pi\sigma^2) + \ln|I-\rho W_2| - \frac{1}{\sigma^2}\epsilon'\epsilon \qquad (10)$$

حاصل می‌شود. ماکسیمم کردن تابع درستنمایی (۱۰) معادل با مینیمم کردن مجموع توان‌های دوم مانده‌های مدل رگرسیونی با متغیرهای وابسته $Z(\rho) = (y-\rho W_2 y)$ و متغیرهای تبیینی $X(\rho) = (X-\rho W_2 X)$ است. در نتیجه برآورد ماکسیمم درستنمایی پارامترهای β و σ^2 به ترتیب به‌صورت

$$\begin{aligned} \hat{\beta} &= [(X-\rho W_2 X)'(X-\rho W_2 X)]^{-1}(X-\rho W_2 X)'(y-\rho W_2 y) \\ \hat{\sigma^2} &= \frac{1}{n}(y(\rho)-X(\rho)\beta(\rho))'(y(\rho)-X(\rho)\beta(\rho)) \qquad (11) \end{aligned}$$

[5] Spatial Error Model

به‌دست می‌آیند. با جای‌گذاری $\hat{\beta}$ و $\hat{\sigma^{۲}}$ در (۱۰) و ماکسیمم کردن آن برحسب ρ برآورد آن به‌دست آورده می‌شود. سپس با جای‌گذاری برآورد ρ در روابط (۱۱) برآوردهای $\sigma^{۲}$ و β برحسب مشاهدات تعیین می‌شود. جنسون و دینااردو (۱۹۹۷) نیز برای پیشگویی مدل خطافضایی نشان دادند با فرض اینکه $Y = X\beta + U$ باشد، که در آن $U = \lambda W_{۲} U + \epsilon$، آنگاه مدل را می‌توان به‌صورت

$$U = (I - \rho W_{۲})^{-۱}\epsilon$$

نوشت. در نتیجه مدل پیشگویی عبارت خواهد شد از

$$Y = X\beta + (I - \rho W_{۲})^{-۱}\epsilon$$

همچنین برای سادگی می‌توان با ضرب طرفین مدل خطا فضایی ارائه شده در (۹) در $\rho W_{۲}$ و سپس کم کردن طرفین از مدل (۹) این مدل را به‌صورت

$$Y - \rho W_{۲} Y = X\beta - \rho W_{۲} X\beta + U - \rho W_{۲} U$$

نوشت. بنابراین داریم

$$\hat{Y} = X\hat{\beta} + \hat{\rho} W_{۲}(X\hat{\beta} - Y)$$

۳ آزمون خودهمبستگی فضایی

عدم توجه به وابستگی فضایی داده‌ها، ممکن است به برآوردهای اریب، ناسازگار و ناکارا منتهی شود. از طرفی برای به کارگیری مدل‌های پیچیده فضایی، باید ضرورت آن‌ها مورد بررسی قرار گیرد. کلیف و اُورد (۱۹۷۲) براساس روش ارائه شده در موران (۱۹۵۰) آزمونی برای تشخیص خودهمبستگی فضایی مانده‌ها معرفی کرده و آن را آزمون موران نامیدند. آماره این آزمون به‌صورت $I = \frac{n}{S}\frac{e^T W e}{e^T e}$ است، که در آن W ماتریس وزن، $S = \sum_i \sum_j w_{ij}$ و $e = y - X\beta$ است. چنانچه مانده‌ها دارای

توزیع نرمال باشند، آماره موران به طور مجانبی دارای توزیع نرمال با میانگین و واریانس

$$\begin{aligned} E(I) &= \frac{n}{S}\frac{tr(MW)}{(n-k)} \\ Var(I) &= (\frac{n}{S})^{۲}\frac{[tr(MWMW^T)+tr(MW)^{۲}+(tr(MW))^{۲}]}{(n-k)(n-k+۲)} - (E(I))^{۲} \end{aligned}$$

است، که در آن‌ها $tr(\cdot)$ اثر ماتریس و $M = (I - X(X^TX)^{-۱}X^T$ است. اگر ماتریس وزن W استاندارد شده باشد، آماره آزمون به‌صورت $I = \frac{e^TWe}{e^Te}$ با میانگین و واریانس

$$\begin{aligned} E(I) &= \frac{tr(MW)}{(n-k)} \\ Var(I) &= \frac{[tr(MWMW^T)+tr(MW)^{۲}+(tr(MW))^{۲}]}{(n-k)(n-k+۲)} - (E(I))^{۲} \end{aligned}$$

خواهد شد، که بنابر قضیه حد مرکزی، وقتی $n \to \infty$، می‌توان این آماره را به‌صورت

$$\frac{I-E(I)}{Var(I)^{\frac{۱}{۲}}} \sim N(۰,۱)$$

استاندارد شده نوشت و با مقایسه مقادیر جدول نرمال استاندارد درباره آزمون قضاوت کرد.

۴ مثال کاربردی

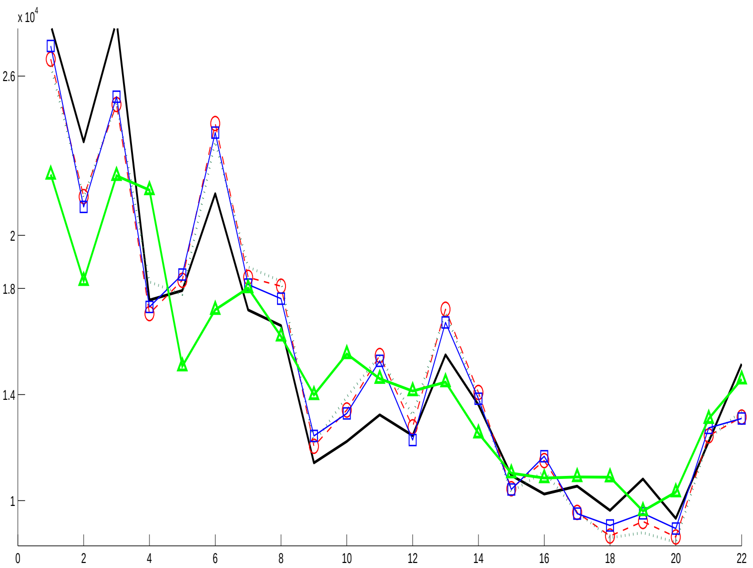
بنگاه‌های معاملات ملکی کلیه معاملات خرید و فروشی که سند آن‌ها از طریق دفتر ثبت اسناد انتقال پیدا می‌کند را در سامانه اطلاعات مدیریت معاملات املاک و مستغلات کشور ثبت می‌کنند. داده‌های این مثال شامل قیمت خرید، فروش و اجارۀ به علاوه سه درصد ودیعه پرداختی بابت یک متر مربع زیر بنای واحدهای مسکونی شهر تهران است که از طریق بنگاه‌های معاملات ملکی در سامانه مذکور ثبت شده‌اند. متوسط قیمت خرید و فروش هر مترمربع زیربنای مسکونی معامله شده در سال ۱۳۸۹ در مناطق ۲۲گانه شهر تهران (شکل ۱) به‌عنوان متغیر پاسخ و متوسط اجاره ماهانه به‌علاوه سه درصد ودیعه پرداختی بابت اجاره یک متر مربع زیربنای

شکل ۱: نقشه مناطق ۲۲گانه شهر تهران

مسکونی بنگاه‌ها به‌عنوان متغیر تبیینی در نظر گرفته شده است. انوع مدل‌های تأخیر فضایی، خطافضایی، اتورگرسیو همزمان و رگرسیون خطی ساده به داده‌ها برازش داده شده، برای مقایسه مدل‌ها از دو ملاک ضریب تعیین تعدیل‌یافته و جذر میانگین توان دوم خطا[۶](RMSE) حاصل از روش اعتبارسنجی متقابل مدل‌ها استفاده شده است. هر یک از مناطق یک ناحیهٔ فضایی و مناطق دارای مرز مشترک به‌عنوان همسایه در نظر گرفته شده‌اند. درایه‌های w_{ij} در ماتریس وزن فضایی برابر یک است اگر مناطق i و j مرز مشترک داشته باشند، و در غیر این‌صورت $w_{ij} = ۰$ در نظر گرفته شده است. با تقسیم هر سطر بر مجموع درایه‌های آن سطر ماتریس وزن استاندارد تعیین شده است.

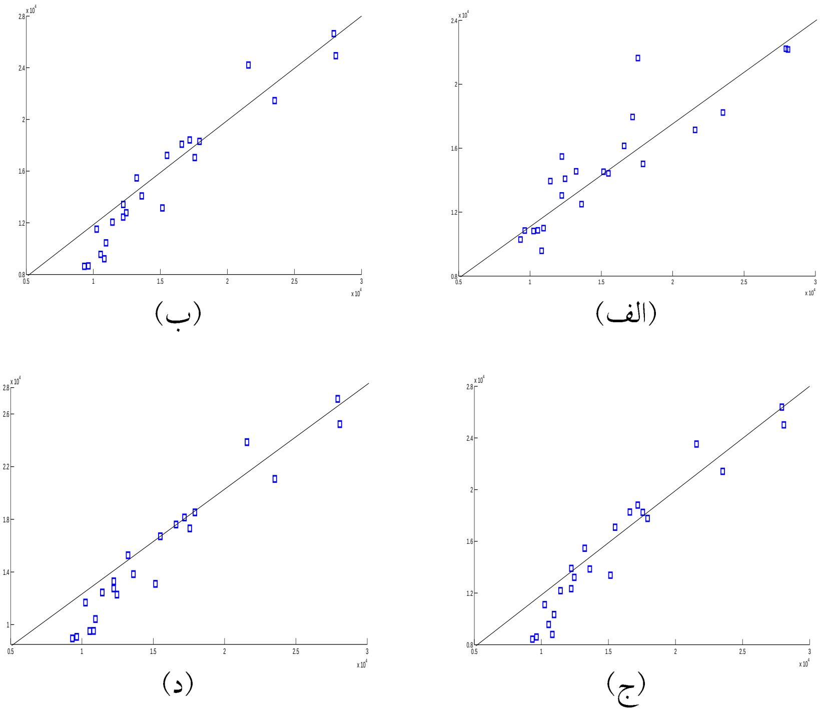
مقدار آزمون موران ۰/۱۲۱ و p–مقدار آن ۰/۰۱ محاسبه شده است که نشان‌دهنده وجود وابستگی فضایی بین داده‌ها است. برآورد پارامترها و p–مقدار آزمون معنی‌دار بودن آن‌ها به همراه ضریب تعیین تعدیل‌یافته برای مدل‌های تأخیر فضایی، خطافضایی، اتورگرسیو همزمان و رگرسیون خطی ساده در جدول ۱ ارائه شده‌اند. همان‌طور که ملاحظه می‌شود. برآورد پارامتر متغیر تبیینی اجاره در تمامی مدل‌ها مثبت و معنی‌دار است که نشان‌دهنده این است که رابطه میزان اجاره با قیمت

---
[۶] Root Mean Squared Error



شکل ۲: نمودار مقادیر واقعی متوسط قیمت مسکن در سال ۱۳۸۹ (خط ممتد) و مقادیر پیشگویی‌شده با مدل‌های خطافضایی (خط ممتد با مربع)، تاخیر فضایی (نقطه‌چین)، رگرسیون خطی ساده (خط‌چین با دایره توخالی) و اتورگرسیو همزمان (خط ممتد با مثلث)

مسکن به‌صورت مثبت و مستقیم است. در میان مدل‌های برازانده شده مدل خطافضایی بیشترین ضریب تعیین تعدیل‌یافته با مقدار ۰/۹۳۶ را دارا است و ضریب اتورگرسیو فضایی در سطح ۰/۰۵ معنی‌دار است در مقایسه با مدل رگرسیون خطی ساده با ضریب تعیین تعدیل‌یافتهٔ ۰/۹۲ نشان‌هندهٔ این است که لحاظ کردن وابستگی فضایی در مدل باعث افزایش کارایی مدل شده است. ضریب تعیین تعدیل‌یافته مدل تأخیر فضایی ۰/۹۲ است و ضریب اتورگرسیو فضایی در این مدل در سطح ۰/۰۵ معنی‌دار نیست. در مدل اتورگرسیو همزمان ضریب تعیین تعدیل‌یافته ۰/۷۴ به دست آمده و ضریب اتورگرسیو فضایی نیز معنی‌دار می‌باشد. بنابراین مدل خطافضایی از سایر مدل‌ها از دقت بیشتری برخوردار است.



شکل ۳: نمودارهای مقادیر واقعی در مقابل مقادیر پیشگوییشده الف: مدل اتورگرسیو همزمان، ب: مدل رگرسیون خطی ساده، ج: مدل تاخیر فضایی و د: مدل خطافضایی

جدول ۱: برآورد پارامترهای مدل‌های تأخیر فضایی، خطافضایی، اتورگرسیو همزمان و رگرسیون خطی

	مدل							
	تأخیر فضایی		خطافضایی		اتورگرسیو همزمان		رگرسیون خطی	
پارامتر	برآورد	p−مقدار	برآورد	p−مقدار	برآورد	p−مقدار	برآورد	p−مقدار
β_0	−۲۳۷۳/۸۳	۰/۰۴۹	−۶۳۳/۰۳	۰/۶۴	–	–	−۱۴۱۶/۱۳	۰/۲۱
β_1	۰/۱۸۳	۰/۰۰۱	۰/۲	۰/۰۰۱	–	–	۰/۲۱	۰/۰۰۱
ρ	۰/۲۱۴	۰/۱۲۸	–	–	۰/۹۵	۰/۰۰۱	–	–
λ	–	–	۰/۴۹	۰/۰۲	–	–	–	–
R^2	۰/۹۲۱۴		۰/۹۳۶		۰/۷۴		۰/۹۲۲	

۱.۴ اعتبارسنجی متقابل مدل‌ها

معیار RMSE با روش اعتبارسنجی متقابل مدل‌ها محاسبه‌و بر اساس آن مدل‌ها با هم مقایسه می‌شوند. مدل دارای RMSE کوچکتر نسبت به سایر مدل‌ها از عملکرد بهتری برخوردار است.

مدل خطا فضایی کمترین RMSE با مقدار ۱۳۵۲ را دارا است و پس از آن به ترتیب مدل‌های رگرسیون خطی ساده، تأخیر فضایی و اتورگرسیو همزمان با مقادیر ۱۴۷۵/۵، ۱۴۸۷/۸ و ۲۷۵۳/۹ قرار دارند. این نتایج مشابه نتایج حاصل از براساس معیار ضریب تعیین تعدیل‌یافته در جدول ۱ است. بنابراین مدل خطافضایی از سایر مدل‌ها دقیق‌تر عمل کرده است. مقادیر پیشگویی‌شده با مدل‌های خطافضایی (خط ممتد با مربع)، تأخیر فضایی (نقطه‌چین)، مدل رگرسیون خطی ساده (خط‌چین با دایره توخالی)، مدل اتورگرسیو همزمان (خط ممتد با مثلث) و داده‌های واقعی (خط ممتد) در شکل ۲ نشان داده شده است. همان‌طور که ملاحظه می‌شود مقادیر پیشگویی‌شده توسط مدل خطافضایی به مقادیر واقعی از سایر مقادیر پیشگویی‌شده نزدیکتر است. همچنین نمودار مقادیر واقعی در مقابل مقادیر پیشگویی‌شده همراه با یک خط ممتد با زاویه ۴۵ درجه برای همه مدل‌ها در شکل ۳ نشان داده شده است. هر چه نقاط به این خط نزدیکتر باشند نشان‌دهنده این است که پیشگویی‌ها دقیق‌تر هستند. نمودار مقادیر واقعی در مقابل مقادیر پیشگویی‌شده مدل خطافضایی که در شکل ۳-د نشان داده شده نسبت به سایر مدل‌ها عملکرد بهتری را نشان می‌دهد.

بحث و نتیجه‌گیری

بین داده‌های معاملات مسکونی شهر تهران وابستگی فضایی وجود دارد بنابراین در تحلیل‌ها باید این وابستگی را لحاظ نمود و برای مدل‌بندی این داده‌ها انواع مدل‌های اتورگرسیو فضایی معرفی گردید. همچنین مشخص شد لحاظ کردن وابستگی فضایی داده‌ها باعث افزایش کارایی و معتبر بودن تحلیل‌ها می‌شود.

تقدیر و تشکر

نویسنده مقاله کمال تشکر و قدردانی را از داوران محترم مجله که با پیشنهادهای ارزنده خود باعث بهبود مقاله شدند، دارد. همچنین از حمایت قطب علمی داده‌های ترتیبی و فضایی دانشگاه فردوسی مشهد تقدیر و تشکر می‌شود.

مراجع

صفایی، م. (۱۳۹۰)، براورد احتمال تغییر وضعیت رفتار سری‌های زمانی مالی با مدل اتورگرسیو تبدیلی مارکف، مجله علوم آماری، جلد **۵**، شماره ۱، ۱۱۸-۱۰۷.

Anselin, L. (1988), *Spatial Econometrics: Methods and Models*, Klumer Academic Publishers, The Netherlands.

Arbia, G. (2006), *Spatial Econometrics*, Springer.

Besag, J. (1974), Spatial Interaction and the Statistical Analysis on Lattice Systems, *Journal of Royal Statistical Society*, Series B, **36**, 192-236.

Cressie, N. (1993), *Statistics for Spatial Data,* New York, Wiley.

Haining, R. P. (1990), *Spatial Data Analysis in the Social and Environmental Sciences*, Cambridge, Cambridge University Press.

Cliff, A. D. and Ord, J. K., (1972), Testing for Spatial Autocorrelation Among Regression Residuals. *Geographical Analysis*, **4**, 267-84.

LeSage, J. P. and Pace R. K. (2009), *Introduction to Spatial Econometrics*, CRC Press Taylor and Francis Group New York.

LeSage, J. P. (1999), *Spatial Econometrics*, The Web Book of Regional Science, Regional Research Institute, West Virginia University, Morgantown, WV.

Lee L. F. (2002), Consistency and Efficiency of Least-Squares Estimation for Mixed Regressive Spatial Autoregressive Models, *Econometric Theory*, **18**, 252-277.

Moran, P. A. P, (1950), The Interpretation of Statistical Maps. *Journal of the Royal Statistical Society*, Series B, **10**, 243-51.

Ord, J. K. (1975), Estimation Methods for Models of Spatial Interaction, *Journal of the American Statistical Association*, **70**, 120-126.

Pace R. K. and Barry R. (1997), Sparse Spatial Autoregressions, *Statistics and Probability Letters*, *33*, 291-297.

Johnston J. and Dinardo N. (1997), *Econometric Methods, (Fourth Edition)*, McGraw-Hill, New York.

Waller, L. A. and Gotway, C. A. (2004), *Applied Spatial Statistics for Public Health Data*, John Wiley, Hoboken, N. J.

Whittle, P. (1954), On Stationary Processes in the Plane, *Biometrika*, **41**, 434-449.