مجله مدل‌سازی پیشرفته ریاضی، دوره ۶، شماره ۱، بهار و تابستان ۱۳۹۵

برآورد کوچک ناحیه‌ای و پیشگویی فضایی

زهرا ناصری و محسن محمدزاده[1]

گروه آمار، دانشگاه تربیت مدرس

تاریخ دریافت: ۱۳۹۴/۶/۲۴ تاریخ پذیرش: ۱۳۹۵/۲/۱۷

چکیده : اندک بودن اندازه داده‌ها در آمارگیری از کوچک نواحی، موجب دقت کم برآوردهای مستقیم ویژگی‌های مختلف در این نواحی می‌شود. با توجه به افزایش وسیع تقاضا برای تولید آمارهای معتبر و دقیق برای کوچک نواحی، مطالعات زیادی انجام شده است که با ارائه رهیافت‌های مناسب این مشکل حل شود. معمولاً مدل‌های آمیخته خطی اساس بسیاری از روش‌های برآورد کوچک ناحیه‌ای هستند که با استفاده از منابع مختلف، اطلاعاتی کمکی به برآوردگرهای مستقیم وام می‌دهند تا دقت آن‌ها را افزایش دهند. چنانچه داده‌های کوچک نواحی وابستگی فضایی داشته باشند، می‌توان از مدل رگرسیونی که شامل متغیرهای کمکی و خطاهای همبسته فضایی به نواحی مجاور است، استفاده کرد. در این مقاله برآورد کوچک ناحیه‌ای بر اساس مدل خطی با اثرات کوچک ناحیهٔ همبسته فضایی بررسی می‌شود که در آن ساختار همسایگی فضایی داده‌ها از طریق ماتریس مجاورت در مدل لحاظ می‌گردد و اطلاعات کمکی فضایی نیز در برآورد کوچک ناحیه‌ای به‌کار گرفته می‌شود. سپس برآورد کوچک ناحیه‌ای میزان محصولات کشاورزی در شهرستان‌های استان فارس با دو روش متداول EBLUP و MBDE و با دو رویکرد معمولی (غیرفضایی) و فضایی در سطح واحد آماری به‌دست آورده خواهد شد آنگاه دقت آن‌ها مورد ارزیابی و مقایسه قرار می‌گیرند.

واژه‌های کلیدی: کوچک ناحیه‌ای، مدل آمیخته خطی، پیشگوی فضایی، برآورد مدل‌مبنا.

رده‌بندی ریاضی (۲۰۱۰): ۶۲F۱۰، ۶۲H۱۱.

۱- مقدمه

معمولاً سیاست‌گذاران برای تصمیم‌گیری و برنامه‌ریزی‌های خود نیازمند برآوردهایی در سطوح کشوری یا دیگر حوزه‌های بزرگ هستند. اما توسعه جوامع و پیچیده‌تر شدن مؤلفه‌های برنامه‌ریزی

۱-آدرس الکترونیکی نویسنده مسئول مقاله:mohsen_m@modares.ac.ir

و همچنین ضرورت توجه به ویژگی‌ها و مطالبات محلی، آن‌ها را وادار به استفاده از برآوردهایی در حوزه‌های کوچک‌تر یا کوچک نواحی نموده است. از نقطه نظر آماری به سطوحی از جامعه که برای به‌دست آوردن برآوردها دارای اندازه نمونه کافی نیستند، کوچک ناحیه‌ای گفته می‌شود. اما معمولاً اندازه نمونه در کوچک ناحیه‌ای گاهی صفر یا ناچیز است. بنابراین یکی از مسائل مهمی که اخیراً آمارگیری نمونه‌ای[۱] با آن مواجه شده است، به‌دست آوردن برآوردهایی دقیق برای کوچک ناحیه‌ای با استفاده از اطلاعات نواحی مجاور است. در زمینه آمار رسمی نیز، با توجه به اینکه مشخصه‌های جامعه یا نمونه از لحاظ جغرافیایی تغییر می‌کنند، برنامه‌ریزی‌های اقتصادی معمولاً برای حوزه‌های کوچک صورت می‌گیرند. به‌عنوان مثال ناحیه جغرافیایی کوچک موردنظر می‌تواند استان، شهرستان، بخش، دهستان یا یک بخش سرشماری باشد. بنابراین بسیاری از آمارگیری‌های نمونه‌ای برای به‌دست آوردن برآوردهای مستقیم[۲] با دقت لازم برای حوزه‌های کوچک، نیاز به حجم بالایی از نمونه خواهند داشت که ممکن است فراتر از زمان، منابع تأمین بودجه و ظرفیت‌های آمارگیری باشد. روش‌های برآوردیابی کوچک ناحیه‌ای[۳](SAE) برآوردگرهایی نامستقیم[۴] را با استفاده از منابع اطلاعات کمکی برای حوزه‌های کوچکی که حجم نمونهٔ کافی از آن‌ها در اختیار نداریم، فراهم می‌سازند به‌طوری‌که از دقت بیشتری نسبت به برآوردگرهای مستقیم برخوردارند.

روش کلی استفاده از مدل‌های آمیخته خطی با اثرات تصادفی نواحی است که برای برآورد از بهترین پیشگوی خطی نااریب تجربی[۵] (EBLUP) استفاده می‌شود [۱]. رهیافت دیگر، برآوردگر مستقیم مدل‌مبنا[۶] (MBDE) است [۲]، که از میانگین موزون نمونه‌های متعلق به ناحیه با وزن‌هایی مبتنی بر سطح جامعه برای برآورد کوچک ناحیه‌ای استفاده می‌شود. این وزن‌ها قدرت قرضی[۷] را که شامل اثرات تصادفی نواحی هستند وارد مدل می‌کنند.

اکثر روش‌های مدل‌مبنای برآورد کوچک ناحیه‌ای بر اساس مدل آمیخته خطی با اثرات تصادفی ناحیه‌ای، تغییرپذیری بین نواحی دورتر از طریق متغیرهای کمکی در قسمت ثابت مدل تخمین می‌زنند. در این روش‌ها اثرات تصادفی نواحی، به‌طور معمول مستقل در نظر گرفته می‌شوند، درحالی‌که در عمل مرزهای کوچک ناحیه‌ای قراردادی و دلخواه هستند و دلیلی وجود ندارد که

۱-Sample survey
2-Direct estimation
۳-Small area estimation
4-Indirect estimator
۵-Empirical best linear unbiased predictor
6-Model based direct estimator
7-Borrow strength

واحدهای جامعه دو طرف مرز همبسته نباشند. بدیهی است وقتی فاصله بین نواحی زیاد می‌شود همبستگی بین ویژگی‌های مشترک آن‌ها کاهش می‌یابد. به‌همین علت باید در مدل‌های کوچک ناحیه‌ای وابستگی فضایی اثرات تصادفی نواحی موردتوجه قرار گیرد. تا بتوان همبستگی اثرات نواحی مجاور را از طریق ملاک مجاورت در تحلیل داده‌ها لحاظ نمود. برای این منظور، در این مقاله، ابتدا برآورد کوچک ناحیه‌ای در سطح واحد آماری با به‌کارگیری مدل آمیخته خطی با ساختار کوواریانس قطری بلوکی استفاده می‌شود و سپس این روش‌ها برای پیشگویی فضایی در کوچک نواحی به‌کار گرفته خواهند شد. در بخش ۲ مدل با اثرات مستقل بررسی می‌شود، در بخش ۳ مدل با اثرات همبسته فضایی معرفی می‌شود. در بخش ۴ در مطالعه‌ای شبیه‌سازی روش‌های مختلف برآورد کوچک ناحیه‌ای مورد ارزیابی و مقایسه قرار می‌گیرند. سپس در بخش ۵ کاربردی از برآورد کوچک ناحیه‌ای با استفاده از مدل در سطح واحد آماری برای برآورد محصول نارنگی در کوچک ناحیه‌های استان فارس (شهرستان‌ها) با استفاده از اطلاعات سرشماری کشاورزی سال ۱۳۸۲ ارائه خواهد شد. در انتها نیز به بحث و نتیجه‌گیری پرداخته می‌شود.

۲- مدل با اثرات تصادفی مستقل

برای برآورد کوچک ناحیه‌ای مدل‌های مختلفی قابل استفاده هستند. در این بخش مدل با اثرات تصادفی مستقل معرفی می‌شوند که رائو [۱] روش EBLUP و چمبرز و چاندرا [۲] روش MBD را برای برآورد پارامترهای آن پیشنهاد دادند. مدل رگرسیونی با خطای آشیانی به‌صورت

$$y_{ij} = x_{ij}\beta + \varepsilon_{ij} \qquad i = 1,\ldots,m \quad j = 1,\ldots,N_i$$
$$\varepsilon_{ij} = u_i + e_{ij}$$

را در نظر بگیرید [۳]، که در آن y_{ij} مقدار متغیر پاسخ موردنظر برای jامین واحد در کوچک ناحیه‌ای iام، x_{ij} بردار p بعدی متغیرهای کمکی مرتبط با این واحدها، β بردار p بعدی اثرات ثابت، u_i اثر تصادفی کوچک ناحیه‌ای iام با میانگین صفر و واریانس σ_u^2 و e_{ij} خطای تصادفی سطح واحد با میانگین صفر و واریانس σ_e^2 است. فرض می‌شود u_{ij} و e_{ij} دو به دو مستقل و دارای توزیع نرمال هستند در این صورت فرم ماتریسی مدل به‌صورت

$$Y_i = X_i\beta + u_i I_{N_i} + e_i \quad , i = 1,\ldots,m \qquad (1)$$

حاصل می‌شود، که در آن $Y_i = (y_{i1},\ldots,y_{iN_i})^t$ بردار مشاهدات متغیر پاسخ، N_i تعداد واحدهای جامعه در کوچک ناحیه‌ای iام، $X_i = (x_{i1},\ldots,x_{iN_i})^t$ ماتریس $N_i \times p$ متشکل از متغیرهای کمکی و $e_i = (e_{i1},\ldots,e_{iN_i})^t$ بردار خطا است. بر اساس مدل (۱) میانگین جامعه در کوچک ناحیه‌ای iام به‌صورت

$$\bar{Y}_i = \bar{X}_i \beta + u_i + \bar{e}_i$$

است، که در آن $\bar{X}_i = \frac{1}{N_i}\sum_{j=1}^{N_i} x_j$ معلوم است. ماتریس کواریانس Y_i نیز به‌صورت

$$V_i = Var\left(Y_i\right) = \sigma_e^2 I_{N_i} + \sigma_u^2 J_{N_i} J_{N_i}^t ,\ i = 1,\ldots,m$$

حاصل می‌شود، که در آن J_{N_i} بردار واحد با طول N_i و I_{N_i} ماتریس همانی از مرتبه N_i است. این ماتریس کواریانس به مؤلفه‌های واریانس یعنی بردار پارامتر نامعلوم $\theta = (\sigma_u^2, \sigma_e^2)^t$ بستگی دارد. چنانچه برآورد θ با $\hat{\theta} = (\hat{\sigma}_u^2, \hat{\sigma}_e^2)^t$ نشان داده شود آنگاه برآورد V_i به‌صورت $\hat{V}_i = \hat{\sigma}_e^2 I_{N_i} + \hat{\sigma}_u^2 J_{N_i} J_{N_i}^t$ می‌شود. مدل‌های ناحیه‌ای (۱) را می‌توان به مدل در سطح جامعه به‌صورت

$$Y = X\beta + Zu + e \tag{۲}$$

بیان کرد، که در آن $Y = (Y_1^t,\ldots,Y_m^t)^t$، $X = (X_1^t,\ldots,X_m^t)^t$، $e = (e_1^t,\ldots,e_m^t)^t$، $u = (u_1,\ldots,u_m)^t$ و $Z = (Z_1,\ldots,Z_m)^t = diag\,(J_{N_1},\ldots,J_{N_m})^t$. درصورتی‌که نمونه‌های مستخرج از نواحی مختلف مستقل باشند، ماتریس کواریانس Y دارای ساختار قطری بلوکی به‌صورت $V = diag\,(V_1,\ldots,V_m)$ خواهد بود.با فرض آن‌که X ماتریس پر رتبه ستونی از مرتبه p است، تجزیه V، Z، X و Y به مؤلفه‌های نمونه و غیرنمونه را در نظر بگیرید، که در آن Y_s بردار $n \times 1$ مشاهدات نمونه، X_s ماتریس $n \times p$ مقادیر نمونه متغیرهای کمکی، Z_s ماتریس $n \times m$ مؤلفه‌های نمونه Z، V_{ss} ماتریس کواریانس $n \times n$ مربوط به بردار نمونه Y_s است. در ادامه همانند مشاهدات نمونه‌ای Y_s، مابقی $r = N - n$ واحد غیر نمونه‌ای جامعه با Y_r نمایش داده می‌شود و از J_N، J_n و J_r برای نمایش بردارهای واحد به ترتیب N، n و r بعدی استفاده می‌شود. همچنین نمادهای I_N، I_n و I_r نیز نشان‌دهنده ماتریس‌های واحد با مراتب مندرج در اندیس آن‌ها هستند. از نمادهای مشابه با اضافه کردن اندیس i برای سطح کوچک نواحی استفاده می‌شود. برای مثال، s_i متناظر با مجموعه n_i واحد نمونه در کوچک ناحیه‌ای i و r_i متناظر با مجموعه $N_i - n_i$ واحدهای غیرنمونه‌ای در کوچک ناحیه‌ای i با واریانس‌های $V_{iss} = \sigma_e^2 I_{is} + \sigma_u^2 Z_{is} Z_{is}^t$ و کواریانس‌های $V_{isr} = \sigma_u^2 Z_{is} Z_{ir}^t$ است.

با در نظر گرفتن (۲)، بهترین پیشگوی خطی ناریب تجربی (EBLUP) برای میانگین کوچک ناحیه‌ای iام به‌صورت

$$\hat{\bar{Y}}_{\text{iEBLUP}} = f_i \bar{Y}_{is} + \left(1 - f_i\right)\left[\bar{X}_{ir}^t \hat{\beta} + \hat{\gamma}_i \left(\bar{Y}_{is} - \bar{X}_{is}^t \hat{\beta}\right)\right]$$

حاصل می‌شود [۱]، که در آن $\hat{\gamma}_i = \hat{\sigma}_u^2\left(\hat{\sigma}_u^2 + n_i^{-1}\hat{\sigma}_e^2\right)^{-1}$، $f_i = \frac{n_i}{N_i}$، $\bar{Y}_{is} = \frac{1}{n_i}\sum_{s_i} Y_j$ و $\bar{X}_{is} = \frac{1}{n_i}\sum_{s_i} X_j$ میانگین‌های نمونه واحد i، $\bar{X}_{ir} = \left(N_i - n_i\right)^{-1}\left(N_i\bar{X}_i - n_i\bar{X}_{is}\right)$ میانگین متناظر X برای $N_i - n_i$ واحد غیرنمونه در ناحیه i و بهترین برآوردگر خطی نااریب تجربی β به‌صورت $\hat{\beta} = \left[\sum_i (X_{is}^t \hat{V}_{iss}^{-1} X_{is})\right]^{-1}\left[\sum_i (X_{is}^t \hat{V}_{iss}^{-1} Y_{is})\right]$ است. با در نظر گرفتن مدل خطی آمیخته سطح جامعه (۲)، وزن‌های نمونه که EBLUP را برای کل جامعه تعیین می‌کنند، عبارتند از [۴]

$$w_{EBLUP} = (w_{j,EBLUP}) = J_n + \hat{H}^t\left(X^t J_N - X_s^t I_n\right) + \left(I_n - \hat{H}^t X_s^t\right)\hat{V}_{ss}^{-1}\hat{V}_{sr} I_r \qquad (۳)$$

که در آن $\hat{H} = \left(\sum_i X_{is}^t \hat{V}_{iss}^{-1} X_{is}\right)^{-1}\left(\sum_i X_{is}^t \hat{V}_{iss}^{-1}\right)$. چمبرز و چاندرا [۲] برآوردگر مستقیم مدل‌مبنای میانگین کوچک ناحیه‌ای iام را به‌صورت

$$\hat{\bar{Y}}_{i,MBD} = \sum_{j\in s_i} w_{j,EBLUP}\, y_j \,/\, \sum_{j\in s_i} w_{j,EBLUP} \qquad (۴)$$

پیشنهاد دادند که برآورد استوار میانگین توان دوم خطای آن به‌صورت

$$\begin{aligned} M\left(\hat{\bar{Y}}_{i,MBD}\right) &= Var\left(\hat{\bar{Y}}_{i,MBD}\right) + bias^2\left(\hat{\bar{Y}}_{i,MBD}\right) \\ &= \sum_{j\in s_i}\lambda_j\left(y_j - x_j^t\hat{\beta}\right)^2 + \left[\left(\hat{\bar{X}}_{i,MBD} - \bar{X}_i\right)^t\hat{\beta}\right]^2 \end{aligned} \qquad (۵)$$

است، که در آن $\hat{\bar{X}}_{i,MBD}$ میانگین موزون متغیرهای کمکی در ناحیه i مبتنی بر وزن‌های (۳) و $\lambda_j = N_i^{-2}[((\sum_{k\in s_i} w_k)^{-1}(N_i w_j - \sum_{k\in s_i} w_k))^2 + (N_i - n_i)(n_i - 1)^{-1}]$ است.

۳- مدل با اثرات تصادفی همبسته فضایی

معمولاً برای منظور کردن همبستگی بین نواحی مجاور از مدل‌های فضایی با اثرات تصادفی ناحیه استفاده می‌شود [۵]. اما چاندرا و همکاران [۶] استفاده از مدل رگرسیونی خطی با وابستگی

فضایی در ساختار خطا را پیشنهاد نمودند. برای این منظور یک فرایند خود رگرسیو هم‌زمان[۱] (SAR) برای خطا در نظر گرفتند [۷]، به‌طوری‌که بردار اثرات تصادفی ناحیه به‌صورت

$$v = (v_i) = \rho W v + u \tag{۶}$$

است، که در آن ρ ضریب خودهمبستگی فضایی، W ماتریس شباهت از مرتبه m، u دارای توزیع نرمال با میانگین صفر و واریانس $\sigma_u^2 I_m$ است. از آنجا که رابطه (۶) را می‌توان به‌صورت $v = \left(I_m - \rho W\right)^{-1}$ بازنویسی نمود، بنابراین u دارای توزیع نرمال با میانگین صفر و واریانس به‌صورت

$$G = \sigma_u^2 \left[\left(I_m - \rho W\right)\left(I_m - \rho W^t\right)\right]^{-1}$$

خواهد بود. ماتریس W بیانگر نحوه ارتباط اثرات تصادفی در نواحی مجاور است، در حالی‌که ρ شدت این ارتباط فضایی را نشان می‌دهد. انتخاب W به‌عنوان ماتریس مجاورت، ساده‌ترین تعریف آن است، یعنی عناصر W برای جفت نواحی مجاور، مقادیر غیر صفر می‌گیرند. برای راحتی تفسیر، این ماتریس به فرم استانداردشده سطری تعریف می‌شود [۸]. اگر ناحیه j با ناحیه k مرز مشترک داشته باشد، درایه (i, j)ام ماتریس مجاورت یک و در غیر این صورت صفر است. این درایه‌ها در فرم استانداردشده سطری به‌صورت

$$w_{jk} = \begin{cases} d_j^{-1} & j, k \text{ رواجم} \\ 0 & o.w \end{cases} \tag{۷}$$

هستند، که در آن d_j تعداد کل نواحی دارای مرز مشترک با ناحیه j و شامل خود ناحیه j است. همچنین ماتریس مجاورت می‌تواند تابعی از فاصله بین موقعیت‌های خاص در هر ناحیه به‌صورت

$$w_{jk} = \begin{cases} d_{jk}^{-1} & j \neq k \\ 0 & o.w \end{cases} \tag{۸}$$

تعریف شود، که در آن d_{jk}^{-1} فاصله بین مراکز دو ناحیه j و k است. برای تعریف EBLUP به‌جای مدل (۲) از مدل خطی آمیخته به فرم

$$Y = X\beta + Zv + e \tag{۹}$$

1-Simultaneous Autoregressive

استفاده می‌شود، که در آن v بردار m بعدی اثرات ناحیه همبسته فضایی است به‌طوری‌که از مدل SAR ارائه شده در (۶) با $Var(e)=\sigma_e^2 I_N$ و $Var(v)=G$ پیروی می‌کند. این مدل را می‌توان به‌صورت

$$Y = X\beta + Z(I-\rho W)^{-1}u + e \qquad (10)$$

نیز بازنویسی نمود، که در این صورت داریم $Var(Y)=V=\sigma_e^2 I_N + ZGZ^t$. بردار پارامترهای $\theta=(\sigma_u^2,\sigma_e^2,\rho)^t$ در عمل نامعلوم است. با جایگذاری برآورد سازگار مجانبی $\hat\theta=(\hat\sigma_u^2,\hat\sigma_e^2,\hat\rho)^t$ و با فرض برقراری (۱۰)، EBLUP فضایی (SEBLUP) برای میانگین کوچک ناحیه‌ای iام، یعنی $\overline{Y}_i$ به‌صورت

$$\hat{\overline{Y}}_{i,\text{SEBLUP}} = f_i\overline{Y}_{is} + (1-f_i)\left(\overline{X}_{ir}^t\hat\beta^s + m_i^t\hat v\right)$$

حاصل می‌شود، به‌طوری‌که $\hat v = \hat\sigma_u^2\left[(I_m-\hat\rho W)(I_m-\hat\rho W^t)\right]^{-1} Z_s^t\hat V_{ss}^{-1}\left(Y_s - X_s\hat\beta^s\right)$، که در آن

$$\hat V_{ss} = \hat\sigma_e^2 I_n + Z_s\hat\sigma_u^2\left[(I_m-\hat\rho W)(I_m-\hat\rho W^t)\right]^{-1} Z_s^t$$

$$\hat\beta^s = \left(X_s^t\hat V_{ss}^{-1}X_s\right)^{-1}\left(X_s^t\hat V_{ss}^{-1}Y_s\right)$$

m_i، بردار m تایی $(0,\ldots,1,\ldots,0)^t$ با درایهٔ یک در iامین موقعیت است.

وقتی تمام اثرات تصادفی دارای توزیع نرمال باشند، بردار پارامتر θ را می‌توان به روش ماکسیمم درستنمایی[1] (ML) یا ماکسیمم درستنمایی مقید[2] (REML) برآورد کرد [۹]. تقریب‌های عددی برآورد REML پارامترهای σ_e^2، σ_u^2 و ρ از طریق فرایندی دو مرحله‌ای به‌دست می‌آید. در اولین مرحله، الگوریتم نلدر-مید [۱۰] برای تخمین این برآوردگرها استفاده می‌شود. در مرحله دوم از تخمین برآوردگرهای مرحله قبل، به‌عنوان مقادیر اولیه در الگوریتم امتیازبندی فیشر[3] استفاده می‌شود زیرا تابع لگاریتم درستنمایی نقاط ماکسیمم موضعی زیادی دارد. در برآورد مستقیم مدل‌مبنا طبق رابطه (۱۰)، وزن‌های نمونه EBLUP برای (۳) به ساختار اثرات تصادفی ناحیه در مدل آمیخته (۲) فقط از طریق نمونه و ساختار کوواریانس جامعه بستگی دارد. در نتیجه برای تعمیم به ساختارهای کوواریانس پیچیده‌تر فقط نیاز است $\hat V_{ss}^{-1}$ و $\hat V_{sr}$ با این مدل‌های

1-Maximum likelihood
2-Restricted maximum likelihood
3-Fisher scoring algorithm

پیچیده‌تر قابل محاسبه باشند. وقتی رابطه (۱۰) برقرار باشد،وزن‌های متناظر با SEBLUP یعنی W_{SEBLUP} با جایگذاری

$$\hat{V}_{ss}^{-1}=\left\{\hat{\sigma}_e^{2} I_s+Z_s\hat{\sigma}_u^{2}\left[\left(I_m-\hat{\rho}W\right)\left(I_m-\hat{\rho}W^{t}\right)\right]^{-1}Z_s^{t}\right\}^{-1}$$

$$\hat{V}_{sr}^{-1}=\hat{\sigma}_u^{2}Z_s\left[\left(I_m-\hat{\rho}W\right)\left(I_m-\hat{\rho}W^{t}\right)\right]^{-1}Z_r^{t}$$

در رابطه (۳) به‌دست می‌آیند. برآوردگر فضایی مستقیم مدل‌مبنای $\overline{Y}_i$ و برآوردگر میانگین توان دوم خطاهای آن نیز به‌ترتیب از روابط (۴) و (۵) حاصل می‌شوند [۶].

۴- مطالعه شبیه‌سازی

در این بخش در مطالعه‌ای شبیه‌سازی کارایی چهار روش مختلف برآورد کوچک ناحیه‌ای که در بخش‌های ۲ و ۳ معرفی شده‌اند، براساس دو ماتریس مجاورت معرفی شده، مورد ارزیابی و مقایسه قرار می‌گیرند. این روش‌ها شامل EBLUP و MBDE تحت مدل‌های خطی آمیخته با اثرات نواحی مستقل و SEBLUP و SMBE تحت مدل‌های خطی آمیخته با اثرات نواحی همبسته فضایی هستند. بر اساس نقشه استان فارس شبیه‌سازی شکل گرفته است، به‌طوری‌که موقعیت کوچک نواحی منطبق بر موقعیت شهرستان‌های موردنظر در استان فارس، در نظر گرفته‌شده است.

برای تولید واحدهای جامعه از مدل رگرسیونی با خطای آشیانی (۹) به‌ازای $\beta=(۱,۵)$ و $m=۱۷$ استفاده شده، که در آن اثرات تصادفی نواحی مجاور با مدل SAR توزیع شده‌اند. مقادیر متغیر کمکی x_{ij} از توزیع پیرسون VI با تابع چگالی

$$f\left(x,\alpha_1,\alpha_2,\beta,\gamma\right)=\frac{\left(\left(x-\gamma\right)/\beta\right)^{\alpha_1-1}}{\beta B\left(\alpha_1,\alpha_2\right)\left(1+\left(x-\gamma\right)/\beta\right)^{\alpha_1+\alpha_2}}$$

و به‌ازای $\left(\alpha_1,\alpha_2,\beta,\gamma\right)=\left(۳۳۴/۷۴,۰/۴۳,۰/۰۰۴۳,۰\right)$ تولیدشده است و $v=\left(v_i\right)$ بردار m تایی اثرات فضایی ناحیه، $u=u_i$ بردار m تایی تحقق‌های مستقل از توزیع $N\left(۰,۱\right)$، e_{ij} خطای فردی از توزیع $N\left(۰,۲\right)$ تولیدشده و درایه‌های ماتریس مجاورت سطر استانداردشده بر اساس روابط (۷) و (۸) و نقشه شهرستان‌های استان فارس به‌صورت

$$\begin{pmatrix}
۱/۶ & ۰ & ۱/۶ & ۱/۶ & ۱/۶ & ۰ & ۰ & ۰ & ۰ & ۰ & ۱/۶ & ۰ & ۱/۶ & ۰ & ۰ & ۰ & ۰ \\
۰ & ۱/۷ & ۰ & ۱/۷ & ۱/۷ & ۱/۷ & ۰ & ۱/۷ & ۰ & ۰ & ۰ & ۰ & ۰ & ۱/۷ & ۱/۷ & ۰ & ۰ \\
۱/۶ & ۰ & ۱/۶ & ۰ & ۱/۶ & ۰ & ۰ & ۱/۶ & ۰ & ۰ & ۱/۶ & ۰ & ۰ & ۱/۶ & ۰ & ۰ & ۰ \\
۱/۹ & ۱/۹ & ۰ & ۱/۹ & ۱/۹ & ۱/۹ & ۱/۹ & ۰ & ۱/۹ & ۱/۹ & ۰ & ۰ & ۱/۹ & ۰ & ۰ & ۰ & ۰ \\
۱/۶ & ۱/۶ & ۱/۶ & ۱/۶ & ۱/۶ & ۰ & ۰ & ۰ & ۰ & ۰ & ۰ & ۰ & ۰ & ۱/۶ & ۰ & ۰ & ۰ \\
۰ & ۱/۶ & ۰ & ۱/۶ & ۰ & ۱/۶ & ۱/۶ & ۰ & ۰ & ۰ & ۰ & ۰ & ۰ & ۰ & ۱/۶ & ۰ & ۱/۶ \\
۰ & ۰ & ۰ & ۱/۵ & ۰ & ۱/۵ & ۱/۵ & ۰ & ۰ & ۱/۵ & ۰ & ۰ & ۰ & ۰ & ۰ & ۰ & ۱/۵ \\
۰ & ۱/۸ & ۱/۸ & ۰ & ۰ & ۰ & ۰ & ۱/۸ & ۰ & ۰ & ۰ & ۱/۸ & ۰ & ۱/۸ & ۱/۸ & ۱/۸ & ۱/۸ \\
۰ & ۰ & ۰ & ۱/۳ & ۰ & ۰ & ۰ & ۰ & ۱/۳ & ۰ & ۰ & ۰ & ۱/۳ & ۰ & ۰ & ۰ & ۰ \\
۰ & ۰ & ۰ & ۱/۳ & ۰ & ۰ & ۱/۳ & ۰ & ۰ & ۱/۳ & ۰ & ۰ & ۰ & ۰ & ۰ & ۰ & ۰ \\
۱/۴ & ۰ & ۱/۴ & ۰ & ۰ & ۰ & ۰ & ۰ & ۰ & ۰ & ۱/۴ & ۰ & ۱/۴ & ۰ & ۰ & ۰ & ۰ \\
۰ & ۰ & ۰ & ۰ & ۰ & ۰ & ۰ & ۱/۳ & ۰ & ۰ & ۰ & ۱/۳ & ۰ & ۰ & ۰ & ۱/۳ & ۰ \\
۱/۵ & ۰ & ۰ & ۱/۵ & ۰ & ۰ & ۰ & ۰ & ۱/۵ & ۰ & ۱/۵ & ۰ & ۱/۵ & ۰ & ۰ & ۰ & ۰ \\
۰ & ۱/۵ & ۱/۵ & ۰ & ۱/۵ & ۰ & ۰ & ۱/۵ & ۰ & ۰ & ۰ & ۰ & ۰ & ۱/۵ & ۰ & ۰ & ۰ \\
۰ & ۱/۵ & ۰ & ۰ & ۰ & ۱/۵ & ۰ & ۱/۵ & ۰ & ۰ & ۰ & ۰ & ۰ & ۰ & ۱/۵ & ۰ & ۱/۵ \\
۰ & ۰ & ۰ & ۰ & ۰ & ۰ & ۰ & ۱/۳ & ۰ & ۰ & ۰ & ۱/۳ & ۰ & ۰ & ۰ & ۱/۳ & ۰ \\
۰ & ۰ & ۰ & ۰ & ۰ & ۱/۵ & ۱/۵ & ۱/۵ & ۰ & ۰ & ۰ & ۰ & ۰ & ۰ & ۱/۵ & ۰ & ۱/۵
\end{pmatrix}$$

در نظر گرفته‌شده است. داده‌های تولیدشده برای مقادیر پارامتر خودهمبستگی فضایی مختلف با تعداد تکرار ۱۰۰۰ مورد ارزیابی قرار گرفتند. برای ارزیابی برآوردهای نواحی کوچک از سه ملاک قدر مطلق اریبی نسبی[1] (ARB)

$$ARB = \frac{\bar{Y}_{Method} - \bar{Y}}{\bar{Y}} \times ۱۰۰$$

ریشه میانگین توان دوم خطاهای نسبی[2] (RRMSE)

$$RRMSE = \left\{ \frac{۱}{m} \bar{Y}^{-۱} \sum_{i=۱}^{m} \left(\bar{Y}_{Method} - \bar{Y} \right)^{۲} \right\}^{\frac{۱}{۲}}$$

و نرخ پوشش[3] (CR)

$$Mean_{\text{Method}} - ۲\sqrt{MSE_{\text{Method}}} < \bar{Y} < Mean_{\text{Method}} + ۲\sqrt{MSE_{\text{Method}}}$$

استفاده می‌شود.

1-Absolute relative biased
2-Relative root mean square error
3-Coverage rate

شکل (۱): نقشه استان فارس

جدول (۱): نتایج شبیه‌سازی برای میانگین ۱۷ کوچک ناحیه بر اساس ماتریس مجاورت (۷)

ملاک ارزیابی	روش	ρ				
		۰	۰/۲	۰/۵	۰/۷	۰/۹
ARB	EBLUP	۰/۰۲۳	۰/۰۱۷	۰/۰۲۷	۰/۰۹۴	۰/۱۴۱
	SEBLUP	۰/۰۰۵	۰/۰۲۰	۰/۰۲۴	۰/۰۵۴	۰/۱۳۷
	MBDE	۰/۰۲۷	۰/۰۶۳	۰/۰۴۴	۰/۰۷۲	۰/۱۲۳
	SMBDE	۰/۰۲۷	۰/۰۶۳	۰/۰۴۴	۰/۰۷۲	۰/۱۲۳
RRMSE	EBLUP	۱۲/۹۹۴	۱۱/۶۹۸	۸/۱۹۰	۸/۵۵۱	۸/۴۰۳
	SEBLUP	۱۳/۲۷۰	۱۱/۹۱۲	۸/۱۶۶	۸/۳۲۵	۸/۱۹۹
	MBDE	۱۶/۰۱۷	۱۳/۸۹۳	۹/۷۷۲	۹/۸۵۹	۹/۵۷۵
	SMBDE	۱۶/۰۱۸	۱۳/۸۹۴	۹/۷۷۲	۹/۸۵۹	۹/۵۷۴
CR	EBLUP	۰/۳۴۴	۰/۵۹۱	۰/۹۵	۰/۹۴۹	۰/۹۵۱
	SEBLUP	۰/۳۴۵	۰/۵۹۲	۰/۹۵۳	۰/۹۵۶	۰/۹۵۶
	MBDE	۰/۳۴۹	۰/۵۹۸	۰/۹۶	۰/۹۶۴	۰/۹۶۴
	SMBDE	۰/۳۴۹	۰/۵۹۸	۰/۹۶۲	۰/۹۶۶	۰/۹۶۶

در جداول ۱ و ۲ میانگین این معیارها گزارش شده است. همان‌طور که در جدول ۱ ملاحظه می‌شود وقتی ضریب همبستگی برابر با صفر است مطابق با حالتی که وابستگی فضایی در مدل لحاظ نشود، روش‌های غیر فضایی با ملاک RRMSE نتایج بهتری را نشان می‌دهند در غیر این صورت اختلاف چشمگیری بین دو روش دیده نمی‌شود. برای مقدار کوچک ρ برابر با ۰/۲، ملاک RRMSE در دو حالت فضایی و غیر فضایی با دو رهیافت EBLUP و MBD نزدیک به‌هم عمل می‌کنند. به‌ازای $\rho = ۰/۵$ برآورد فضایی نسبت به غیرفضایی اریبی کمتری را نشان می‌دهد. همچنین ملاک RRMSE در هر دو روش SEBLUP و SMBD مقادیر کمتری نسبت به دو روش EBLUP و MBD را دارد. به‌ازای $\rho = ۰/۷$ هر دو ملاک RRMSE و ARB نتایج بهتری را برای روش‌های فضایی نشان می‌دهند. در نهایت با بیشترین مقدار ρ که در این شبیه‌سازی ۰/۹ در نظر گرفته‌شده است، RRMSE در روش‌های فضایی EBLUP و MBD مقادیر کمتری نسبت به روش‌های غیر فضایی متناظرشان نشان می‌دهد. با توجه به معیار CR با افزایش ضریب خودهمبستگی فضایی درصد بیشتری از بازه‌های اطمینان ساخته شده مقدار صحیح میانگین را در دو روش فضایی EBLUP و MBD در بر می‌گیرد.

جدول (۲): نتایج شبیه‌سازی برای میانگین ۱۷ کوچک ناحیه بر اساس ماتریس مجاورت (۸)

ملاک ارزیابی	روش	ρ				
		۰	۰/۲	۰/۵	۰/۷	۰/۹
ARB	EBLUP	۰/۱۷۶	۰/۰۹۴	۰/۴۳۵	۰/۲۸۳	۰/۳۱۲
	SEBLUP	۰/۱۴۷	۰/۰۸۱	۰/۴۰۲	۰/۲۳۰	۰/۳۰۷
	MBDE	۰/۲۵۵	۰/۲۶۲	۰/۳۳۸	۰/۴۰۱	۰/۴۲۸
	SMBDE	۰/۲۵۵	۰/۲۶۲	۰/۳۳۸	۰/۴۰۱	۰/۴۲۸
RRMSE	EBLUP	۷/۶۲۸	۷/۷۸	۷/۶۱۲	۷/۶۸۱	۷/۶۹۱
	SEBLUP	۷/۹۰۵	۸/۰۲۵	۷/۸۰۳	۷/۹۲۱	۷/۹۸۹
	MBDE	۹/۸۹۴	۱۰/۲۰۱	۹/۳۲۱	۹/۵۲۳	۹/۵۹۴
	SMBDE	۹/۸۹۴	۱۰/۲۰۱	۹/۳۲۱	۹/۵۲۳	۹/۵۹۴
CR	EBLUP	۰/۳۲۵	۰/۵۴۱	۰/۹۵۰	۰/۹۳۹	۰/۹۴۹
	SEBLUP	۰/۳۲۵	۰/۵۴۲	۰/۹۵۵	۰/۹۴۰	۰/۹۵۱
	MBDE	۰/۳۳۲	۰/۵۴۸	۰/۹۵۹	۰/۹۵۰	۰/۹۵۹
	SMBDE	۰/۳۳۲	۰/۵۴۷	۰/۹۶۰	۰/۹۵۱	۰/۹۶۱

طبق جدول ۲، با ملاک ARB روش فضایی EBLUP نسبت به غیرفضایی آن، نتایج دقیق‌تری ارائه می‌دهد. در روش MBD بین دو روش فضایی و غیر فضایی بر اساس ملاک ARB تفاوت فاحشی وجود ندارد. مطابق ماتریس وزن در نظر گرفته‌شده که تابعی از فاصله بین موقعیت‌هاست، با ملاک RRMSE روش EBLUP نسبت به روش فضایی آن، نتایج بهتری را نشان می‌دهد. از آنجایی در روش‌های به‌کار گرفته، ماتریس مجاورت مبتنی بر وجود مرز مشترک بین نواحی طبق ملاک‌های ارزیابی نتایج بهتری را نشان می‌دهد، بنابراین در ادامه برای برآورد میانگین کوچک نواحی داده‌های کشاورزی از ماتریس مجاورت (۷) استفاده می‌شود.

۵- برآورد کوچک ناحیه‌ای فضایی محصولات کشاورزی

در این بخش بر اساس مجموعه داده‌های سرشماری کشاورزی سال ۱۳۸۲ استان فارس با استفاده از روش‌های برآورد کوچک ناحیه‌ای که در بخش‌های ۲ و ۳ معرفی شدند، مجموع تولید محصول نارنگی در شهرستان‌های استان فارس برآورد می‌شوند سپس روش‌های بهترین پیشگوی نااریب خطی، برآوردگر مستقیم مدل‌مبنا و روش‌های فضایی مورد ارزیابی و مقایسه عددی قرار می‌گیرند. برای ارزیابی نتایج به‌دست آمده از ملاک‌های قدرمطلق اریبی نسبی و ریشه میانگین توان دوم خطاهای نسبی استفاده می‌شود. داده‌ها شامل متغیرهای پاسخ و کمکی از فایل اصلی اطلاعات سرشماری کشاورزی متشکل از ۱۰۵۵۳ بهره‌برداری در استان فارس استخراج شده‌اند. از بین داده‌ها ۷۵۹۳ بهره‌بردار محصول نارنگی هستند و بقیه بهره‌برداری‌ها درخت نهال بودند که هنوز محصولی برداشت نکرده‌اند. اندازه نمونه‌ای بهین برای برآورد میزان کل تولید این محصول با انتخاب نمونه‌ای به‌اندازهٔ $n_0 = ۵۰$ و محاسبه واریانس نمونه‌ای، $s^2 = ۱۴۹/۸۸$، در سطح ۰/۰۵ و خطای یک تن بر اساس رابطه $n = \frac{s^2 Z^2_{\alpha/2}}{d^2}\left(1 - \frac{n_0}{N}\right)$ تقریباً ۵۷۲ حاصل شد. متغیر پاسخ موردنظر مجموع کل محصول نارنگی برداشت شده در سطح استان فارس و شهرستان‌های این استان است. اطلاعات کمکی قابل‌دسترس برای این مطالعه، تعداد درخت بارور نارنگی برای هر بهره‌برداری است. اندازه نمونه‌ای با روش تصادفی ساده بدون جایگذاری از چارچوب در دست از بهره‌برداری‌ها گزینش شدند.

همان‌طور که در جدول ۳ ملاحظه می‌شود، اندازهٔ نمونه‌های مستخرج از شهرستان‌های استان فارس همه به‌جز شهرستان‌های جهرم و داراب کمتر از نمونه‌های بهین برای برآورد میزان برداشت در هر شهرستان است. بنابراین برآوردهای مستقیم مجموع تولید نارنگی در آن‌ها نمی‌توانند از دقت لازم برخوردار باشند، لذا برای بهبود دقت برآوردها از روش‌های برآورد کوچک ناحیه‌ای استفاده می‌کنیم. در مطالعه حاضر سعی می‌شود که با برازش یک مدل مناسب در سطح واحد

آماری حداکثر قدرت را برای برآوردگر مستقیم وام گرفت و با افزایش دقت آن، نتیجه معتبری به‌دست آورد.

جدول (۳): اندازه‌های نمونه‌ای تعداد بهره‌بردار شهرستان‌های استان فارس

شهرستان	تعداد بهره‌بردار	اندازه سهم از نمونه کل	اندازه بهین
استان فارس	۷۵۹۳	۵۷۲	۵۷۲
استهبان	۷۳	۵	۱۱
جهرم	۲۲۱۷	۱۶۱	۱۰۶
داراب	۷۵۹	۶۵	۶۳
شیراز	۴۶۶	۲۹	۳۶
فسا	۸۱۸	۶۲	۹۸
فیروزآباد	۴۷۱	۳۷	۴۹
کازرون	۶۶۹	۵۸	۶۶
لارستان	۱۰۸	۹	۴۷
مرودشت	۱۸۵	۱۱	۳۶
ممسنی	۱۰۳۲	۷۱	۷۶
نی‌ریز	۳۳۲	۲۷	۳۲
لامرد	۸	۱	۵
ارسنجان	۷۸	۷	۱۷
زرین‌دشت	۲۲	۵	۹
قیروکارزین	۲۲۵	۱۶	۵۶
مهر	۴۷	۴	۱۳
فراشبند	۸۳	۴	۱۹

منبع: [۱۲]

آزمون همبستگی فضایی موران [۱۱] با p مقدار ۰/۰۵ معنی‌دار شده است، که حاکی از وجود ارتباط فضایی بین میزان محصول نارنگی در سطح شهرستان‌های استان فارس است، بنابراین به‌دلیل آن‌که فرض استقلال برقرار نیست، روش غیرفضایی برای آن‌ها ناکارآمد خواهد بود و در ادامه بررسی این امر به‌خوبی نشان داده خواهد شد.با برازش مدل و برآورد ماکسیمم درستنمایی مقید پارامترهای واریانس اثرهای تصادفی ناحیه‌ها σ_u^2 و واریانس خطاهای نمونه‌گیری σ_e^2 که اطلاعاتی در موردشان در دسترس نیست به‌دست آورده شده‌اند.

جدول (۴): پیشگویی و برآورد فضایی میزان محصول نارنگی در شهرستان‌های استان فارس

شهرستان	مقدار واقعی	برآوردگر SMBD	پیشگوی SEBLUP
استهبان	۱۱/۱	۱۱/۱۵	۱۱/۲۰
جهرم	۴۰۲/۵۰	۴۰۴/۱۱	۴۰۷/۳۳
داراب	۱۶۱/۲۰	۱۶۲/۵۰	۱۵۷/۹۵
شیراز	۵۵/۱۰	۵۵/۱۰	۵۵/۱۰
فسا	۱۵۸/۷۲	۱۵۹/۳۴	۱۶۱/۲۰
فیروزآباد	۷۳/۶۳	۷۳/۶۳	۷۳/۶۳
کازرون	۱۲۹/۹۲	۱۳۱/۶۶	۱۲۸/۱۸
لارستان	۲۲/۱۴	۲۲/۲۳	۲۱/۲۴
مرودشت	۲۱/۵۶	۲۱/۵۶	۲۱/۶۷
ممسنی	۱۰۴/۳۷	۱۰۴/۳۷	۱۰۵/۰۸
نی‌ریز	۶۱/۰۲	۶۱/۰۲	۶۱/۰۲
لامرد	۲/۱۴	۲/۱۳	۲/۱۰
ارسنجان	۱۴/۲۸	۱۴/۳۵	۱۴/۴۲
زرین دشت	۱۱/۲۰	۱۱/۲۰	۱۳
قیروکارزین	۳۷/۷۶	۳۷/۷۶	۳۹/۲۰
مهر	۸/۱۶	۸/۱۶	۸/۲۴
فراشبند	۷/۱۲	۷/۱۲	۷/۶۴

جدول (۵): ملاک‌های ارزیابی روش‌های مختلف

ملاک ارزیابی	MBDE	SMBD	EBLUP	SEBLUP
ARB	۰/۳۷۴	۰/۳۷۷	۰/۳۹۹	۰/۱۴۰
RRMSE	۱/۴۹۱	۱/۴۸۹	۱/۶۵۷	۱/۶۳۲
CR	۰/۹۹۹	۰/۹۹۹	۰/۹۴۱	۰/۹۴۱

برای مدل با اثرات همبسته فضایی با فرض نرمال بودن اثرات تصادفی برآورد پارامترها به‌روش ماکسیمم درستنمایی مقید به‌صورت $\theta=(\sigma_e^{۲},\sigma_u^{۲},\rho)=(۰/۰۷,۰/۰۲,۰/۷۲)$ حاصل شده است. براساس مقادیر میانگین واقعی و میانگین‌های حاصل از دو روش SEBLUP و SMBDE در جدول ۴، ملاحظه می‌شود که دو روش فضایی برآورد میانگین‌های دقیق‌تری را نتیجه

می‌دهند. همان‌طور که در جدول ۵ ملاحظه می‌شود براساس ملاک RRMSE روش‌های فضایی نتایج دقیق‌تری را ارائه می‌کند و براساس ملاک ARB نیز روش SEBLUP نسبت به EBLUP اریبی کمتری را موجب می‌شود. بنابراین هر دو روش برآورد کوچک ناحیه‌ای توسط مدل با اثرات تصادفی وابسته فضایی نتایج بهتری ارائه می‌کنند هرچند ملاک اریبی نسبی برای روش MBDE اختلاف ناچیزی را نشان می‌دهد. از آنجا که برآورد مقدار وابستگی در داده‌ها ۰/۷۲ است، به‌کار بردن نتایج حاصل از مدل‌های با اثرات وابسته فضایی از توجیه بیشتری برخوردارند.

بحث و نتیجه‌گیری

در این مقاله از مدل در سطح واحد آماری با اثرات همبسته فضایی نواحی در برآورد کوچک ناحیه‌ای استفاده شده است. همچنین نحوهٔ انجام محاسبات روش‌های MBDE و EBLUP موردبررسی قرار گرفتند و در مطالعهٔ شبیه‌سازی و همچنین مثال کاربردی تأثیرگذاری مدل‌های فضایی برای درنظرگرفتن ساختار فضایی در برآورد کوچک ناحیه‌ای موردبررسی قرار گرفت.
موضوعات مختلفی در زمینه استفاده از مدل در سطح واحد آماری با اثرات همبسته فضایی نواحی در برآورد کوچک ناحیه‌ای وجود دارد که می‌توان به آن‌ها پرداخت. شناسایی موقعیت‌هایی که دارای اطلاعات فضایی مؤثر هستند و روش مناسب وارد کردن این اطلاعات در فرایند مدل‌بندی کوچک ناحیه‌ای از مهم‌ترین این موضوعات هستند. گزارش بار محاسباتی در مدل‌های فضایی مناسب با مجموعه داده‌های آمارگیری از موضوعات کاربردی است که می‌توان به آن پرداخت زیرا با حجم داده‌های آمارگیری به امکانات محاسباتی پیشرفته‌ای نیاز است.
موضوع مهم ارتباط بین داده‌های آمارگیری و اطلاعات فضایی است به‌طوری‌که، در این مقاله همه نواحی دارای نمونه فرض شده‌اند، در بسیاری از مواقع داده‌های آمارگیری در دسترس فقط از یک ناحیه هستند درحالی‌که اطلاعات فضایی برای تمامی نواحی موجود هستند. در این مقاله برای لحاظ کردن اثر فضایی داده‌ها در مدل از ساختار همسایگی نواحی مجاور استفاده شد، اگر بتوان ساختار همبستگی فضایی داده‌ها را در قالب تابع کوواریانس فضایی تعیین نمود، آنگاه به‌صورت یک مسئله زمین آمار قابل تحلیل است.

تقدیر و تشکر

نویسندگان از زحمات سردبیر و همچنین داوران محترم مجله که نظرات ارزنده آن‌ها موجب بهبود مقاله شد و از حمایت قطب علمی داده‌های ترتیبی و فضایی دانشگاه فردوسی مشهد قدردانی می‌نماید.

مراجع

[1] Rao, J. N. K. (2003), *Small Area Estimation*, John Wile, New York.

[2] Chambers, R. L. and Chandra, H. (2006), Improved Direct Estimators for Small Area, Centre for Survey Statistics and Methodology, University of Wollongong, Working Paper 03-08, 2008, 26p.

[3] Fuller, W.A. and Harter, R.M. (1987), The Multivariate Components of Variance Model for Small Area Estimation, In Platek, R., Rao, J. N. K., Sarndal, C. E. and Singh, M.P. (Eds.), Small Area Statistics, John Wiley and Sons, New York.

[4] Royall, R.M. and Cumberland, W.G. (1978), Variance Estimation in Finite PopulationSampling, *Journal of the American Statistical Association*, **73**, 351-358.

[5] Cressie, N. (1991), Small Area Prediction of Undercount Using the General Linear Model, *Proceeding of Statistics Canada Symposium*, **90**, 93-105.

[6] Chandra, H., Salvati, N. and Chambers, R. (2007), Small Area Estimation for Spatially Correlated Populations-A Comparison of Direct and Indirect Model Based Methods, *Statistics in Transition*, **8**, 331-350.

[7] Anselin, L. (1992), *Method and Models*, Kluwer Academic Publishers, Boston.

[8] Banerjee, S., Carlin, B. and Gelfand, A. (2004), *Hierarchical Modelling and Analysisfor Spatial Data*, Chapman and Hall, New York.

[9] Petrucci, A. and Salvati, N. (2006), Small Area Estimation for Spatial Correlation in Watershed Erosion Assessment, *Journal of Agricultural, Biological and Environmental Statistics*, **11**, 169-182.

[10] Nelder, J. and Mead, R. (1965), A Simplex Method for Function Minimization, *Computer Journal*, **7**, 308-313.

[11] Moran, P. A. P. (1950), Notes on Continuous Stochastic Phenomena, *Biometrika*, **37**, 17-23.

[12] باذل، ف. (۱۳۹۱)، روش وزن‌دهی در برآورد کوچک ناحیه‌ای، پایان‌نامه کارشناسی ارشد، دانشگاه تربیت مدرس.

Small area Estimation and Spatial Prediction

Zahra Naseri, Mohsen Mohammadzadeh

Department of Statistics, Trabiat Modares University, Tehran, Iran

Abstract

Direct estimators of parameters are not precise because of little surveys unit in small areas. Regarding the wide spread increase of demand for providing valid and accurate statistics for small areas, attempts have been made to present proper solutions for the problems. Small area estimation approaches provide the direct estimators with borrowing strength to increase their precision based on a model, especially about those estimators that are based on the linear mixed model including random area effects and using various auxiliary sources. Data associated with spatially contiguous small areas may be modeled via covariates, with error terms that are spatially dependent according to neighbor areas. In this paper we investigate small area estimation based on linear models with spatially correlated small area effects where the neighborhood structure is described by a contiguity matrix. Such models allow efficient use of spatial auxiliary information in small area estimation. Then estimation for small areas will be achieved for the amount of agronomy production in Fars province, according to the two common EBLUP and MBDE methods and two usual (non spatial) and spatial approaches based on the unit level model. Then the accuracy of them have been compared.

***Keywords*:** Small Area; Linear Mixed Model; Spatial Model; Spatial Prediction; Model Based Estimation.

Mathematics Subject Classification (2010): 62F10, 62H11.