کاربردی از مدل آمیخته‌ی خطی در براورد کوچک ناحیه‌ای محصول پرتقال در استان فارس

حمیدرضا نواب‌پور[۱]

چکیده:

در سال‌های اخیر روش‌های براورد کوچک ناحیه‌ای مورد توجه قرار گرفته‌اند. این توجه به‌دلیل افزایش درخواست براوردهای معتبر برای کوچک ناحیه‌ها بوده است، زیرا این براوردها برای برنامه‌ریزی‌های توسعه‌ای دقیق‌تر، تخصیص اعتبارهای دولتی و تصمیم‌گیری‌های تجاری به‌کار می‌آیند. پرسش کلیدی در براورد کوچک ناحیه‌ای این است که وقتی اندازه‌ی نمونه‌ای کم است، چگونه می‌توان براوردهای معتبر به‌دست آورد؟ زمانی‌که اندازه‌ی نمونه‌ای به‌دست آمده از یک ناحیه، کوچک (یا حتی صفر) باشد، براوردگرهای مستقیم از یک انحراف معیار بزرگ و غیر قابل قبولی برخوردار می‌شوند. یک راه بهبود این براوردها، وام گرفتن قدرت از منبع‌های داده‌ای موجود است. برای این کار می‌توان از طریق مدل‌بندی یا به عبارتی استفاده از مدل‌های اتصالی (شامل مدل‌های صریح و ضمنی) اثر اندازه‌ی نمونه‌ای را بهبود بخشید. برای به‌دست آوردن براورد معتبر میانگین کوچک ناحیه‌ای مدل‌های آمیخته‌ی خطی تعمیم‌یافته و بهترین پیشگوگر نااریب خطی تجربی به‌طور گسترده مورد استفاده قرار گرفته‌اند. در این مقاله ابتدا براورد کوچک ناحیه‌ای را معرفی می‌کنیم. سپس برای به‌دست آوردن براوردهای معتبر برای کوچک ناحیه‌ها به معرفی مدل فی-هریوت (فی و هریوت، ۱۹۷۹)، حالت خاص مدل‌های آمیخته‌ی خطی تعمیم‌یافته می‌پردازیم. سرانجام در یک مطالعه‌ی شبیه‌سازی با استفاده از داده‌های سرشماری کشاورزی سال ۱۳۸۲، تولید پرتقال در شهرستان‌های استان فارس (کوچک ناحیه‌ها) در سال ۱۳۸۲ و بر اساس مدل فی-هریوت براورد می‌شود.

واژه‌های کلیدی: براورد کوچک ناحیه‌ای، براوردگر مستقیم، براوردگر نامستقیم، مدل فی-هریوت، مدل‌های آمیخته‌ی خطی تعمیم‌یافته، براورد ماکسیمم درستنمایی مقید، بهترین پیشگوگر نااریب خطی تجربی.

۱ مقدمه

به‌منظور صرفه‌جویی در هزینه‌ها و زمان و نیز دقت قابل قبول آماره‌ها، واحدهای تولیدکننده‌ی آمار تمایل به تولید آمار در سطح ملی دارند. اما در طول چند دهه‌ی گذشته مدیران و مسئولان، سیاست‌گذاران و برنامه‌ریزان، درخواست‌های زیادی برای براوردهای مربوط به زیرجامعه‌ها داشته‌اند. به‌طور معمول آمارگیری‌های نمونه‌ای برای تولید براوردهای پارامترهای جامعه‌ی بزرگ به‌کار گرفته می‌شوند. زیرجامعه‌ها می‌توانند:

۱- گروه‌های اجتماعی-جمعیتی، مانند: سن، جنسیت، گروه‌های قومی- قبیله‌ای درون یک ناحیه‌ی جغرافیایی بزرگ؛ و

۲- ناحیه‌های جغرافیایی مانند: استان، شهرستان، شهر و ... باشند.

این زیرجامعه‌ها، کوچک ناحیه‌ها و براوردهای مربوط به آن‌ها، براوردهای کوچک ناحیه‌ای[۲] نامیده می‌شوند. به‌دلیل کوچک بودن اندازه‌ی نمونه‌ای این ناحیه‌ها که ممکن است حتی گاهی صفر نیز باشد، اغلب براوردهای مستقیم از انحراف استاندارد بزرگ و غیر قابل قبولی برخوردار هستند. یک راه بهبود دقت براوردها، وام گرفتن قدرت[۳] از منبع‌های داده‌ای مرتبط (داده‌های به‌دست آمده از کوچک ناحیه‌های مشابه یا مجاور در طول زمان یا داده‌های حاصل از آخرین آمارگیری همان کوچک ناحیه یا ترکیبی از این دو) است. بنابراین وام گرفتن قدرت یعنی افزایش اثر اندازه‌ی نمونه‌ای برای براورد که با استفاده از مدل‌هایی (صریح یا ضمنی) برای ربط دادن داده‌های اضافی و کمکی به داده‌های نمونه‌ای انجام

[۱]عضو هیئت علمی گروه آمار، دانشگاه علامه طباطبایی

[۲]small area estimations

[۳]borrow strength

می‌شود (رائو، ۲۰۰۳). براوردهایی که از این روش به‌دست می‌آیند براوردهای نامستقیم نامیده می‌شوند. برای محاسبه‌ی براوردهای نامستقیم، دردسترس بودن داده‌های کمکی خوب و تعیین مدل‌های ربط‌دهنده‌ی مناسب بسیار مهم هستند. یک مدل مناسب، برای زمانیکه اندازه‌ی نمونه‌ای کافی نیست و یا اطلاعات لازم در دسترس نیستند، مدل آمیخته‌ی خطی تعمیم‌یافته است. از جمله روش‌های براورد نامستقیم بر اساس مدل‌های ربط‌دهنده‌ی صریح در سطح ناحیه مدل فی–هریوت است که اخیراً بسیار مورد توجه قرار گرفته است.

۲ کاربرد مدل‌های آمیخته‌ی خطی در براوردهای کوچک ناحیه‌ای

درطول سه دهه‌ی گذشته، مدل‌های آمیخته‌ی خطی، در سطح وسیعی از کاربردهای کوچک ناحیه‌ای به‌کار گرفته شده‌اند. چنین مدل‌هایی، انعطاف‌پذیری زیادی در ترکیب اطلاعات از منبع‌های مختلف دارند. یکی از کاربردهای مهم مدل‌های آمیخته‌ی خطی حل مشکل براورد کوچک ناحیه‌ای است، به‌ویژه زمانی‌که اندازه‌ی نمونه‌ای به‌دست آمده در کوچک ناحیه‌ها کافی نباشند. با استفاده از مدل‌های آمیخته‌ی خطی می‌توان اطلاعات دو یا چند کوچک ناحیه را با هم ادغام کرد تا یک نمونه‌ی مناسب حاصل شود. با استفاده از بهترین پیشگوگر نااریب خطی تجربی در مدل‌های آمیخته‌ی خطی می‌توان براوردی معتبر برای کوچک ناحیه‌ها به‌دست آورد. در این بخش مدل‌های آمیخته‌ی خطی تعمیم‌یافته معرفی می‌شوند. سپس براورد پارامترهای نامعلوم این مدل‌ها شامل بردار پارامتری β و واریانس مدل و بهترین پیشگوگر نااریب خطی تجربی مورد بررسی قرار می‌گیرند. همچنین کاربرد مدل فی–هریوت در براورد کوچک ناحیه‌ای تشریح می‌شود.

۱.۲ مدل‌های آمیخته‌ی خطی تعمیم‌یافته $(GLMM)$

مدل خطی که فقط شامل اثرهای ثابت است را در نظر بگیرید،

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$$

که در آن $\mathbf{y}$، بردار تصادفی متغیرهای پاسخ است، ماتریس $\mathbf{X_{n\times p}}$ ماتریس طرح بوده که فرض می‌شود دارای رتبه‌ی کامل ستونی است و بردار $\boldsymbol{\beta}_{p\times 1}$، بردار اثرهای ثابت است. در بعضی از موارد برخی اثرهای ثابت تصادفی فرض می‌شوند (جیانگ و لاهیری، ۲۰۰۶). این حالت زمانی اتفاق می‌افتد که مشاهده‌ها به یکدیگر وابسته باشند. بردار $\mathbf{e}$ شامل خطاهای مدل با بردار میانگین $\mathbf{0}$ و ماتریس واریانس کوواریانس $\sigma_e^2\mathbf{I_n}$ است. بردار $\mathbf{y}$ دارای توزیع $\boldsymbol{N_n}\left(\boldsymbol{X\beta}, \boldsymbol{\sigma_e^2 I_n}\right)$ است (N_n نمادی است که برای نشان دادن توزیع نرمال n بعدی به‌کار گرفته شده است). اگر به مدل رگرسیونی معمولی برداری از اثرهای تصادفی اضافه شود، مدل آمیخته‌ی خطی حاصل می‌شود که به‌صورت زیر است:

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Zu} + \boldsymbol{e}$$

که در آن ماتریس $\mathbf{Z_{n\times q}}$ مانند $\mathbf{X}$ ماتریس طرحی است. برای بردارهای $\mathbf{u}$ و $\mathbf{e}$ فرض‌های زیر وجود دارند.

$$E\left(\mathbf{u}\right) = 0; Cov\left(\mathbf{u}\right) = \mathbf{G} = \mathbf{G}\left(\sigma_{\mathbf{e}}\right); \mathbf{u} \sim \mathbf{N_q}\left(0, \mathbf{G}\right),$$

$$E\left(\mathbf{e}\right) = 0; Cov\left(\mathbf{e}\right) = \mathbf{R} = \mathbf{R}\left(\sigma_{\mathbf{e}}\right); \mathbf{e} \sim \mathbf{N_n}\left(0, \mathbf{R}\right)$$

$$Cov\left(\mathbf{u}, \mathbf{e}\right) = 0$$

که $\mathbf{G}$ و $\mathbf{R}$ ماتریس‌های کوواریانس متناهی دلخواه هستند. تحت فرض‌های بالا خواهیم داشت:

$$E\left(\mathbf{y}\right) = \mathbf{X}\boldsymbol{\beta}, \mathbf{Cov}\left(\mathbf{y}\right) = \mathbf{V} = \mathbf{V}\left(\sigma_{\mathbf{e}}\right) = \mathbf{ZGZ'} + \mathbf{R},$$

$$\mathbf{y} \sim \mathbf{N_n}\left(\boldsymbol{X\beta}, \boldsymbol{ZGZ'} + \boldsymbol{R}\right).$$

۱.۱.۲ براورد پارامترهای مدل

در مدل آمیخته‌ی خطی بالا پارامترهای $\boldsymbol{\beta}$ و $\mathbf{V}$ و $\mathbf{u}$ نامعلوم هستند که باید براورد شوند. برای براورد بردار پارامترهای $\boldsymbol{\beta}$ و ماتریس $\mathbf{V}$ روش‌های مختلفی وجود دارند که در زیر به آن‌ها اشاره می‌شود. در مدل‌های آمیخته‌ی خطی برای براورد بردار پارامتری $\boldsymbol{\beta}$ اغلب از روش کم‌ترین توان‌های دوم تعمیم‌یافته (GLS)[۴] استفاده می‌شود که از معادله‌های نرمال $X'V^{-1}X\beta = X'V^{-1}y$ به‌دست می‌آید و برابر است با:

$$\hat{\beta}_{GLS} = (\boldsymbol{X'V^{-1}X})^{-1}\boldsymbol{X'V^{-1}y}. \qquad (1)$$

[۴] generalized least square

در مدل‌های آمیخته‌ی خطی تعمیم‌یافته $\hat{\beta}_{GLS}$ بهترین براوردگر نااریب خطی برای $\boldsymbol{\beta}$ است. اگر ماتریس واریانس کوواریانس در رابطه‌ی (۱) نامعلوم باشد، براورد آن جایگزین می‌شود.

یک موضوع مهم در تحلیل مدل‌های آمیخته‌ی خطی، براورد مؤلفه‌های واریانسی[۵] است. اما برای ارزیابی تغییرات براوردگرها و پیشگوگرها لازم است که کمیت‌های دیگری نیز مانند اثرهای تصادفی و آمیخته براورد شوند. برای مدل‌ها با مؤلفه واریانسی روش کلاسیک براورد واریانس، روش تحلیل واریانس ($ANOVA$) است. همچنین دو روش براورد که به‌طور فراوان برای براورد مؤلفه‌های واریانسی به‌کار می‌روند روش‌های ماکسیمم درست‌نمایی (ML) و ماکسیمم درست‌نمایی مقیّد ($REML$) هستند. در براورد پارامترهای واریانس به روش ماکسیمم درست‌نمایی مشکل‌هایی وجود دارد. اول این‌که براورد واریانس‌های به‌دست آمده از این روش گرایش به اریبی دارند. مشکل دیگر براوردگرهای ML، مدل وابسته[۶] بودن آن‌ها است که نسبت به ماتریس مدل بسیار حساس‌اند (دیگل و همکاران، ۲۰۰۲). به‌خاطر وجود چنین مشکل‌هایی، روش براورد درست‌نمایی دیگری برای براورد پارامترهای واریانس وجود دارد که روش ماکسیمم درست‌نمایی مقید[۷] نامیده می‌شود. روش براورد ماکسیمم درست‌نمایی مقید بر اساس همان اصل درست‌نمایی است و براوردهای حاصل همان ویژگی‌ها نظیر سازگاری، کارایی و فرض نرمال بودن مجانبی براوردگرهای ML را دارا می‌باشند. از این‌رو روش $REML$ با توجه به نااریب بودن یا تقریباً نااریب بودن، در مدل‌های آمیخته‌ی خطی برای براورد پارامترهای واریانس به روش براورد ML ارجحیت دارد. با توجه به $\hat{\beta}_{GLS}$، لگاریتم تابع درست‌نمایی برای واریانس برابر است با:

$$LogL(V) = -\frac{n-p}{۲}Log(۲\pi) - \frac{۱}{۲}Log\left|V\right| + \frac{۱}{۲}Log\left|X'X\right| + \frac{۱}{۲}Log\left|X'V^{-۱}X\right| - \frac{۱}{۲}(r(\sigma))'V^{-۱}r(\sigma)$$

که در آن

$$r(\sigma) = \left(y - X\hat{\beta}_{GLS}\right) = y - X(X'V(\sigma)^{-۱}X)^{-۱}X'V(\sigma)^{-۱}y.$$

که با ماکسیمم کردن تابع لگاریتم درست‌نمایی، $\hat{\sigma}_{REML}$ حاصل می‌شود. سیرل و همکاران (۱۹۹۲) معادله‌ی براورد ML برای واریانس را به‌دست آورده‌اند. در کل، ماکسیمم‌سازی به استفاده از روش‌های عددی مانند الگوریتم نیوتن-رافسون، امتیاز فیشری و الگوریتم EM نیاز دارد.

۲.۲ مدل فی-هریوت

فی و هریوت اطلاعات کمکی از دیگر منبع‌های داده‌ای مانند داده‌های ثبتی را برای کاهش میانگین توان دوم خطای (MSE) براوردگرهای کوچک ناحیه‌ای به‌کار گرفتند. ویژگی‌های براورد کوچک ناحیه‌ای مانند اریبی و MSE به‌طور شرطی از اطلاعات کمکی نتیجه می‌شوند و فرض می‌شود که این اطلاعات برای همه‌ی ناحیه‌ها در دسترس‌اند و بدون خطا، اندازه‌گیری می‌شوند. مدل فی-هریوت یک مدل پایه‌ای در سطح ناحیه است. در این مدل واحدهای مشاهده‌ای برای هر دو متغیر پاسخ و کمکی ناحیه‌ها هستند. این مدل به صورت زیر است:

$$y_i = X_i'\beta + u_i + \varepsilon_i = \theta_i + \varepsilon_i; i = ۱, \ldots, m.$$

که $\theta_i = X_i'\beta + u_i$ میانگین کوچک ناحیه‌ی iام و غیرقابل مشاهده و m تعداد کوچک ناحیه‌ها است. بردار $\mathbf{X_i}$ یک بردار $p \times ۱$ از متغیرهای تبیینی و $\boldsymbol{\beta}$ یک بردار $p \times ۱$ از ضریب‌های رگرسیونی و ε_i خطای طرح و $\mathbf{u_i}$ اثر تصادفی خاص کوچک ناحیه‌ی iام است و همان خطای مدل هستند که به‌طور مستقل از هم توزیع شده و دارای توزیع‌های زیر هستند:

$$u_i \sim N\left(۰, A\right); \varepsilon_i \sim N\left(۰, D_i\right); Cov\left(u_i, \varepsilon_i\right) = ۰$$

$$y_i \sim N\left(X_i'\beta, V_i\right); V_i = A + D_i.$$

که A واریانس خطای مدل و نامعلوم است و D_i واریانس خطای نمونه‌گیری و معلوم فرض می‌شود. با این وجود گاهی قابل قبول است که $D_i = \frac{\sigma_e^۲}{n_i}$ در نظر گرفته شود. در اینجا n_i اندازه نمونه‌ای برای ناحیه‌ی iام و $\sigma_e^۲$ تغییرات درون ناحیه‌ای است. در ادامه به پیشگویی ترکیب خطی اثرهای آمیخته تحت مدل فی-هریوت می‌پردازیم که کاربردهای زیادی در براورد کوچک ناحیه‌ای دارد.

[۵] variance components

[۶] model-dependent

[۷] restricted maximum likelihood estimation

۱.۲.۲ پیشگویی ترکیب خطی اثرهای آمیخته

فرض کنید که اطلاعات کمکی در دسترس هستند. تحت مدل فی-هریوت ترکیب خطی $\theta_i = X_i'\beta + u_i$ را در نظر می‌گیریم. اگر پارامترهای D_i و A و $\boldsymbol{\beta}$ معلوم باشند، تحت فرض‌های نرمال، بهترین پیشگوگر نااریب خطی برای θ_i برابر است با (لوهر و یبارا، ۲۰۰۸):

$$\begin{aligned}\tilde{\theta}_{i,FH} &= X_i'\beta + \tilde{u}_i \\ &= X_i'\beta + \frac{A}{A+D_i}\left(y_i - X'_i\beta\right) \\ &= \gamma_i y_i + (۱-\gamma_i)\, X_i'\beta,\end{aligned}$$

به‌طوری‌که $\gamma_i = \frac{A}{A+D_i}$. این پیشگوگر یک ترکیب محدب از براوردگر مستقیم و مقدار $X_i'\beta$ مدل رگرسیونی است. میانگین توان دوم خطای این پیشگوگر عبارت است از:

$$\begin{aligned}MSE\left(\tilde{\theta}_{i,FH}\right) &= E(\tilde{\theta}_{i,FH} - \theta_i)^۲ \\ &= (\gamma_i - ۱)^۲ E(u_i)^۲ + \gamma_i^۲ E(\varepsilon_i)^۲ \\ &+ ۲\,(\gamma_i - ۱)\,\gamma_i E\left(u_i\varepsilon_i\right)\end{aligned}$$

با توجه به فرض استقلال u_i و ε_i و این‌که $E(u_i) = ۰$ و $E\left(\varepsilon_i\right) = ۰$ خواهیم داشت:

$$\begin{aligned}MSE\left(\tilde{\theta}_{i,FH}\right) &= \frac{D_i^۲}{\left(A+D_i\right)^۲} Var\left((u_i) + \frac{A^۲}{\left(A+D_i\right)^۲} Var(\varepsilon_i)\right) \\ &= \frac{D_i^۲ A}{\left(A+D_i\right)^۲} + \frac{A^۲ D_i}{\left(A+D_i\right)^۲} \\ &= \frac{AD_i}{A+D_i} = \gamma_i D_i.\end{aligned}$$

در اصل پارامترهای A و $\boldsymbol{\beta}$ نامعلومند و باید توسط داده‌ها براورد شوند. بنا بر این بهترین پیشگوگر نااریب خطی تجربی برای θ_i برابر است با:

$$\begin{aligned}\hat{\theta}_{i,FH} &= X_i'\hat{\beta} + \hat{u}_i, \\ &= \hat{\gamma}_i y_i + (۱-\hat{\gamma}_i)\, X'_i\hat{\beta}, \qquad (۲)\end{aligned}$$

که در آن $\hat{\gamma}_i = \frac{\hat{A}}{\hat{A}+D_i}$. رابطه‌ی (۲)، یک ترکیب محدب از براوردگر مستقیم y_i و مقدار پیشگویی $X'_i\hat{\beta}$ معادله‌ی رگرسیونی است. میانگین توان دوم خطای این پیشگویی با توجه به این‌که

$$l' = X_i'; V = A + D_i; G = A; Z = ۱; m = ۱,$$

برابر است با (داتا و لاهیری، ۲۰۰۰):

$$\begin{aligned}MSE\left(\hat{\theta}_{i,FH}\right) &= MSE\left(\tilde{\theta}_{i,FH}\right) + E(\hat{\theta}_{i,FH} - \tilde{\theta}_{i,FH})^۲ \\ &= g_{۱i}\left(A\right) + g_{۲i}\left(A\right) + g_{۳i}\left(A\right) \qquad (۳)\end{aligned}$$

به‌طوری‌که

$$\begin{aligned}g_{۱i}\left(A\right) &= \frac{AD_i}{A+D_i} = \gamma_i D_i; \\ g_{۲i}\left(A\right) &= D_i^۲ (A+D_i)^{-۲} x'_i [\sum_{u=۱}^{m}\left(A+D_u\right)^{-۱} x_u x_u']^{-۱} x_i; \\ g_{۳i}\left(A\right) &= ۲D_i^۲ ((A+D_i)^{-۳}[\sum_{u=۱}^{m}\left(A+D_u\right)^{-۲}]^{-۱}.\end{aligned}$$

روش‌های براورد کوچک ناحیه‌ای، ترکیب براوردهای مستقیم یک آمارگیری با پیشگویی‌هایی از یک مدل، در به‌دست آوردن کمیت‌های مورد نظر با کاهش میانگین توان دوم خطا هستند.

۳ کاربرد مدل فی- هریوت در براورد کوچک ناحیه‌ای

در این بخش با به‌کارگیری مدل فی-هریوت که یک مدل پایه‌ای در سطح ناحیه است، تولید کل محصول پرتقال در شهرستان‌های استان فارس (کوچک ناحیه‌ها) با استفاده از داده‌های سرشماری کشاورزی سال ۱۳۸۲ مرکز آمار ایران براورد می‌شود.

سرشماری کشاورزی توسط مرکز آمار ایران انجام می‌شود. روش گرداوری داده‌ها از طریق پرسش‌نامه‌ای که از هر بهره‌بردار پرسیده می‌شود، صورت می‌گیرد. در طراحی پرسش‌نامه مواردی از قبیل: تعداد بهره‌برداری؛ سطح کاشت محصول (برحسب هکتار)؛ تعداد درخت بارور و تعداد درخت نهال و مقدار تولید (بر حسب تن) در نظر گرفته می‌شود. نتیجه‌ی این سرشماری حکایت از وجود ۹۲۹۸ بهره‌برداری محصول پرتقال داشته و تولید کل پرتقال ۹۷۹۶۶ تن در سطح استان بوده است.

در این کاربرد برای به‌دست آوردن براورد پارامتر موردنظر (تولید کل محصول پرتقال در استان فارس در سال ۱۳۸۲) با در نظر گرفتن تولید کل محصول نارنگی به‌عنوان متغیر کمکی در شهرستان‌های استان فارس (کوچک ناحیه) از رویکرد مدل‌مبنا شامل مدل صریح (مدل فی-هریوت) استفاده می‌شود.

۱.۳ تعیین اندازه‌ی نمونه‌ای

برای به‌دست آوردن براورد کوچک ناحیه‌ای از داده‌های آمارگیری نمونه‌ای که برای اجرا در ناحیه‌ی بزرگ‌تر شامل کوچک ناحیه‌ها طراحی شده است، استفاده می‌شود. نمونه‌های قرار گرفته در هر شهرستان به عنوان اندازه‌ی نمونه‌ای هر شهرستان (کوچک ناحیه‌ها) در نظر گرفته می‌شود.

برای به‌دست آوردن اندازه‌ی نمونه‌ای کل، n، در سطح استان، با استفاده از یک نمونه‌ی تصادفی به تعداد ۲۰۰ بهره‌برداری از فهرست حاصل از سرشماری کشاورزی سال ۱۳۸۲ واریانس مجموع تولید در سطح استان را براورد و سپس از فرمول زیر مقدار n را برای $\alpha = ۰/۰۵$ محاسبه کردیم.

$$n = \frac{Z^{۲}_{\alpha/۲} s^{۲} N^{۲}}{e^{۲} + Z^{۲}_{\alpha/۲} s^{۲} N} \qquad (٤)$$

در رابطه‌ی (٤)، $Z_{\alpha/۲} \approx ۲$ ، N=۹۲۹۸ و $s^{۲}$ براورد واریانس مجموع است که برابر با ۹٤۸/۱۱٦ به‌دست آمده و e حاشیه‌ی خطا است که در این‌جا بر اساس ۰/۱۲ از تولید کل محصول پرتقال (۰/۱۲×۹۷۹٦٦=۱۱۷۵۵/۹۲) در سطح استان در نظر گرفته شده است. اندازه‌ی نمونه‌ای برای یک نمونه‌گیری در سطح استان فارس در زیر محاسبه شده است.

$$n = \frac{۴ \times ۹۴۸/۱۱۶ \times (۹۲۹۸)^{۲}}{(۱۱۷۵۵/۹۲)^{۲} + ۴ \times ۹۴۸/۱۱۶ \times (۹۲۹۸)} \approx ۱۹۰۰$$

پس از انتخاب این تعداد نمونه در سطح استان فارس، تعداد نمونه‌های قرار گرفته شده در شهرستان‌های این استان محاسبه و در جدول ۱ ارائه شده است.

برای تعیین اندازه‌ی نمونه‌ای برای هر شهرستان، با استفاده از یک نمونه‌ی تصادفی مقدماتی از هر شهرستان و با به‌کارگیری رابطه‌ی (٤) و با در نظر گرفتن $\alpha = ۰/۰۵$ و هم‌چنین با در نظر گرفتن حاشیه‌ی خطای ۰/۱۲ از تولید کل محصول پرتقال برای شهرستان داراب و ۰/۱۵ از تولید کل محصول پرتقال در بقیه‌ی شهرستان‌ها، اندازه‌ی نمونه‌ای برای شهرستان‌های استان فارس به‌دست آمد. این اندازه‌های نمونه‌ای در جدول ۲ آمده‌اند.

جدول ۱: اندازه‌ی نمونه‌ای قرار گرفته در شهرستان‌های استان فارس از نمونه‌ی کل

ردیف	شهرستان	اندازه‌ی نمونه‌ای (n_i)	واریانس ($s_i^۲$)
۱	استهبان	٤٠	۳٥/۸۲
۲	جهرم	٦۸۹	۱٦۳/۲٥
۳	داراب	٥٥٥	٤٥٤/۱۸
٤	زرین دشت	۳	۰/۱۳
٥	شیراز	٦	۰/۲٦
٦	فراشبند	۷	۰/۰۰۹
۷	فسا	۱٥۰	۳٥۲۱/۷۸
۸	فیروزآباد	٤٠	۰/٤٥
۹	قیروکارزین	۸۰	٦۱٦/۳۹
۱۰	کازرون	۱٦۰	۳۰۷/۲۷
۱۱	لارستان	۳٥	٤۸۳/٥۷
۱۲	لامرد	۳	۰/۰٥٦
۱۳	مرودشت	٥	۰/۰٥۱
۱٤	ممسنی	۱۱٥	٥۰/٤٦
۱٥	مهر	۳	۰/۰۸
۱٦	نیریز	۹	۲/٦۱

جدول ۲: اندازه‌ی نمونه‌ای بهین محاسبه شده n_i برای شهرستان‌های استان فارس

ردیف	شهرستان	تعداد بهره‌برداری N_i	حاشیه‌ی خطا e_i	اندازه‌ی نمونه‌ای مقدماتی	n_i
۱	استهبان	۱٥۳	۱٦۲/٤٥	۱٥	۷۸
۲	جهرم	۳٤٦۷	۲۳۱٤/۸	٥۰	۱۲۹۹
۳	داراب	۳۰۰۳	۷۸٥۲/٥	٥۰	۱٦٦۰
٤	زرین دشت	۸	۰/۳	٥	٦
٥	شیراز	۲۸	۲/۱	۸	٤۷
٦	فراشبند	۳۱	۲/۸٥	٥	٥۱
۷	فسا	۷٤٥	۲۳۷٦/٤٥	٤۰	٦۹٤
۸	فیروزآباد	۱٥۲	۱٤/۷	۲۰	۱٦۰
۹	قیروکارزین	۲۸۸	۳۸۸/۲	۳۰	۳۱٦
۱۰	کازرون	۷۰۰	۱۳۲۰/۱٥	۳٥	۷۱۸
۱۱	لارستان	۱٤۷	۱٥۹/۱٥	۲۰	٤۰
۱۲	لامرد	۹	۰/۳	٥	۱۹
۱۳	مرودشت	۱٤	۰/۳	٥	۲۳
۱٤	ممسنی	٤۹۹	۹۲/۲٥	۳۰	٤۹۳
۱٥	مهر	۱٦	۱/۰٥	٥	۱۷
۱٦	نیریز	٤۰	۱۸/۳	۸	۳٤

با مقایسه‌ی جدول‌های ۱ و ۲ با یکدیگر، نتیجه می‌شود که اندازه‌ی نمونه‌ای قرار گرفته در هر شهرستان وقتی اندازه‌ی نمونه‌ای در سطح استان محاسبه شده است، کوچک‌تر از اندازه‌ی نمونه‌ای بهین محاسبه شده برای هر شهرستان است. واریانس هر شهرستان در جدول ۱ به‌دلیل کوچک بودن اندازه‌ی نمونه‌ای، بزرگ است. بنا بر این در این حالت این شهرستان‌ها کوچک ناحیه به‌شمار می‌آیند و براوردگرهای مستقیم برای براورد مجموع در این کوچک ناحیه‌ها نادقیق‌اند لذا باید از فن‌های براورد کوچک ناحیه‌ای برای بهبود براوردها استفاده شود.

۲.۳ براورد تولید کل محصول پرتقال

برای به‌دست آوردن براورد تولید کل محصول پرتقال در هر شهرستان (کوچک ناحیه‌ها) با به‌کارگیری رویکرد مدل‌مبنا، باید یک مدل مناسب به داده‌ها برازش داده شود. با استفاده از منبع داده‌ای موجود (سرشماری کشاورزی ۱۳۸۲)، تولید نارنگی در هر شهرستان را به عنوان متغیر کمکی در نظر گرفتیم. از آن‌جایی‌که مجموع تولید نارنگی در سطح هر شهرستان از سرشماری کشاورزی سال ۱۳۸۲ در دسترس است، لذا می‌توانیم از مدل فی–هریوت که در آن واحدهای مشاهده‌ای برای هر دو متغیر پاسخ و کمکی در سطح ناحیه‌ها هستند، برای براورد تولید کل محصول پرتقال استفاده کنیم. همان‌طور که در بخش ۳ اشاره شد، مدل فی–هریوت به صورت زیر است:

$$y_i = X_i'\beta + u_i + \varepsilon_i = \theta_i + \varepsilon_i; i = ۱, \dots, ۱۶.$$

نمادها در بخش ۲ معرفی شده‌اند. با توجه به این‌که به ازای هر y_i فقط یک متغیر کمکی داریم، در نتیجه مدل فی–هریوت، یک مدل فی–هریوت تک متغیری است. مقدار y_iی ناحیه‌ای و X_i برای هر شهرستان در جدول ۳ آورده شده است:

با توجه به این‌که هدف، براورد تولید کل محصول پرتقال در شهرستان‌های استان فارس است، تحت مدل فی–هریوت ترکیب خطی اثر آمیخته‌ی $\theta_i = X_i'\beta + u_i$ را در نظر می‌گیریم. بهترین پیشگوگر نااریب خطی تجربی برای θ_i برابر است با:

$$\hat{\theta}_{i,FH} = \hat{\gamma}_i y_i + (۱ - \hat{\gamma}_i) X_i'\hat{\beta} (۵)$$

که در آن $\hat{\gamma}_i = \frac{\hat{A}}{\hat{A}+\hat{D}_i}$ ، $\hat{\beta} = (X'\hat{V}^{-۱}X)^{-۱}X'\hat{V}^{-۱}y$ ، $\hat{V} = \hat{A} + \hat{D}_i$ ، $X' = (X_۱, \dots, X_{۱۶})$ و $\hat{D}_i = \frac{N-n}{nN} s_i^۲$.

جدول ۳: مقدارهای محصول پرتقال، y_i، و نارنگی، X_i، بر حسب تن

ردیف	شهرستان	y_i	X_i
۱	استهبان	۲۸۷/۵۳	٦
۲	جهرم	۲۸۲۷/۸	۱۳۷٤۰
۳	داراب	۸۸۵۳/۰٤	۷۹۹
٤	زرین دشت	۰/۷٤	۰/۵
۵	شیراز	۱/۸۲	٦
٦	فراشبند	۲/۲	۰/۵
۷	فسا	٤۷۹۲/٦	۷۱۰٦
۸	فیروزآباد	۱۲/۱۱۵	۹۷
۹	قیروکارزین	۸٤۸	۱۹۹۵
۱۰	کازرون	۱٦۰۷/۲۷	۵۵۳۵
۱۱	لارستان	۳۵٦/۷۵	۳۸۷
۱۲	لامرد	۰/۵۳	۱
۱۳	مرودشت	۰/۳۷	۱
۱٤	ممسنی	۱۵۵/۵۷	۱۰٤
۱۵	مهر	۰/۵۵	۲
۱٦	نیریز	۲۵/۹٤	٤

با استفاده از فرمول (۵)، $\hat{\beta} = ۲/۱$ به‌دست آمد که در سطح ۵ درصد معنی‌دار است. همچنین بازه‌ی اطمینان تقریبی برای برابر با (۲/۸۱ و ۱/۳۱) به‌دست آمد که شامل صفر نیست. برای بررسی فرض نرمال بودن e_iها (براورد ε_iها) در سطح $\alpha = ۰/۰۵$ آزمون کولموگوروف–اسمیرونوف انجام شد. تحت این آزمون pمقدار آزمون برابر با ۰/۲۱ شد لذا فرض نرمال بودن e_iها رد نمی‌شود. برای به‌دست آوردن براورد واریانس اثرهای تصادفی، $\hat{A}$، از روش براوردگر گشتاوری ساده استفاده شد. براورد این واریانس برابر است با:

$$\hat{A} = \frac{۱}{m-p}\left\{\sum\nolimits_i (Z_i^{-۱}y_i - X'_i\hat{\beta})^۲ - \sum\nolimits_i \frac{D_i}{Z_i^۲}(۱ - \tilde{h}_{ii})\right\}$$
$$= ۱۷/۷۱$$

که در آن $m = ۱۶$ (برابر با تعداد کوچک ناحیه‌ها)، $p = ۱$ (تعداد پارامترهای براوردشده، $\hat{\beta}$) و $Z_i = ۱$ و $\tilde{h}_{ii} = X_i'(\sum_I X_i X'_i)^{-۱}X_i$. بنا بر این $\hat{\gamma}_i$ و $\hat{V}_i$ و $\hat{D}_i$ و $\hat{\theta}_{i,FH}$

برای هر شهرستان با توجه به رابطه‌ی (۵) و همچنین مقدار y_i و X_i برای هر شهرستان محاسبه و در جدول ٤ آورده شده‌اند:

جدول ٤: مقدار پارامترهای براورد شده‌ی، $\hat{D}_i$، $\hat{V}_i$، $\hat{\gamma}_i$، $\hat{\theta}_{i,FH}$

ردیف	شهرستان	$\hat{D}_i$	$\hat{V}_i$	$\hat{\gamma}_i$	$\hat{\theta}_{i,FH}$
۱	استهبان	۰/٦٦	۱۸/۳۷	۰/۹٦	٦/۹۲
۲	جهرم	۰/۱٦	۱۷/۹	۰/۹۹	٤/۱۸
۳	داراب	۰/٦۷	۱۸/۳۷	۰/۹٦	۱۵/۷
٤	زرین دشت	۰/۰۳	۱۷/۷٤	۰/۹۹	۰/۲۸
۵	شیراز	۰/۳	۱۷/۷٤	۰/۹۹	۰/۳
٦	فراشبند	۰/۰۰۱	۱۷/۷۱	۰/۹۹۹	۰/۳
۷	فسا	۱۸/۷۵	۳٦/٤٦	۰/٤۸	۲۱/۵۲
۸	فیروزآباد	۰/۰۱	۱۷/۷۲	۰/۹۹۹	۰/۳۰۱
۹	قیروکارزین	۵/۵٦	۲۳/۲۷	۰/۷٦	۹/۹٦
۱۰	کازرون	٤۸/۱	۱۹/۱۹	۰/۹۲	۱۰/۸۷
۱۱	لارستان	۱۰/۵۳	۲۸/۲٤	۰/٦۳	۹/۰۲
۱۲	لامرد	۰/۱۲	۱۷/۷۲	۰/۹۹۹	۰/۱۷۵
۱۳	مرودشت	۰/۰۰۱	۱۷/۷۱	۰/۹۹۹	۰/۰۷٤
۱٤	ممسنی	۰/۳٤	۱۸/۰۵	۰/۹۸	۱/۳٤
۱۵	مهر	۰/۰۲۲	۱۷/۷۳	۰/۹۹۸	۰/۱۸
۱٦	نیریز	۰/۲۲	۱۷/۹۳	۰/۹۹	۲/۸٦

جدول ۵: براورد فی-هریوت $\hat{Y}_{i,FH}$ براورد مستقیم، $\hat{Y}_i$، و مقدار واقعی، Y_i، حاصل از سرشماری کشاورزی سال ۱۳۸۲

ردیف	شهرستان	$\hat{Y}_{i,FH}$	$\hat{Y}_i$	Y_i
۱	استهبان	۱۰٦۰	۱۰۹۹/۱۱	۱۰۸۳
۲	جهرم	۱٤٤۸۳	۱٤۲۲۹/۳	۱۵۳٤۲
۳	داراب	٤۷۱٤۳	٤۷۹۰۲/۱۲	۵۲۳۵۰
٤	زرین دشت	۲/۲۲	۲/۲٤	۲
۵	شیراز	۸/۳۲	۸/٤۹	۱٤
٦	فراشبند	۹/۷	۹/۷	۱۹
۷	فسا	۱٦۰۳۲	۱۹۳۷۰	۱۵۸٤۳
۸	فیروزآباد	٤٦	٤٦/۰٤	۹۸
۹	قیروکارزین	۲۸٦۷/۳۳	۳۰۵۲	۲۵۸۸
۱۰	کازرون	۷٦۰۹	۷۰۳۱/۸۲	۸۸۰۱
۱۱	لارستان	۱۳۲۵	۱٤۹۸/٤	۱۰٦۱
۱۲	لامرد	۱/٦	۱/٦	۲
۱۳	مرودشت	۱/۰٤	۱/۰٤	۲
۱٤	ممسنی	٦٤٤	٦۷۲/۳۳	٦۳۵
۱۵	مهر	۲/۸۸	۲/۹۳	۷
۱٦	نیریز	۱۱٤/٤	۱۱۵/۳	۱۲۲

با به‌دست آوردن میانگین توان دوم خطای براوردهای کوچک ناحیه‌ای به‌دست آمده می‌توان دقت و اعتبار این براوردگرها را ارزیابی کرد. با توجه به رابطه‌ی (۳) میانگین توان دوم خطا برابر است با:

$$\widehat{MSE}(\hat{\theta}_{i,FH}) = g_{1i}\left(\hat{A}\right) + g_{2i}\left(\hat{A}\right) + g_{3i}\left(\hat{A}\right)$$

که در آن

$$g_{1i}\left(\hat{A}\right) = \frac{\hat{A}\hat{D}_i}{\hat{A}+\hat{D}_i} = \hat{\gamma}_i\hat{D}_i;$$

$$g_{2i}\left(\hat{A}\right) = \hat{D}_i^2(\hat{A}+\hat{D}_i)^{-2}x'_i[\sum_{u=1}^{m}(\hat{A}+\hat{D}_u)^{-1}x_u x_u']^{-1}x_i;$$

$$g_{3i}\left(\hat{A}\right) = 2\hat{D}_i^2(\hat{A}+\hat{D}_i)^{-3}[\sum_{u=1}^{m}(\hat{A}+\hat{D}_u)^{-2}]^{-1}.$$

و همچنین $\widehat{MSE}(\hat{Y}_{i,FH}) = N_i^2\widehat{MSE}(\hat{\theta}_{i,FH})$ که N_i تعداد بهره‌برداری در شهرستان iام است. مقدارهای تقریبی میانگین توان دوم خطا برای براوردهای به‌دست آمده در جدول ٦ آورده شده اند (مقدار $\widehat{MSE}(\hat{Y}_{i,FH})$ در مقیاس ۰/۰۰۰۱ تولید بیان شده است).

بر این اساس براوردگر فی-هریوت تولید کل محصول پرتقال برای هر شهرستان برابر با $\hat{Y}_{i,FH} = N_i\hat{\theta}_{i,FH}$ است که مقدار آن برای هر شهرستان (کوچک ناحیه) در جدول ۵ آورده شده است.

٤ مطالعه‌ی شبیه‌سازی

در این مطالعه با استفاده از نمونه‌ی ۱۹۰۰تایی انتخاب شده، ۱۰۰۰ نمونه‌ی خودگردان تولید شده و مقدار براوردهای تولید کل و MSEی آن‌ها برای براورد مستقیم و فی-هریوت محاسبه و مقایسه می‌شوند.

۱.٤ فرمول‌های براورد

فرمول‌های به‌کار رفته برای براورد خودگردان تولید کل محصول پرتقال برای براوردگرهای مستقیم، $\bar{\hat{Y}}_i$، و فی-هریوت، $\bar{\hat{Y}}_{i,FH}$، بر اساس ۱۰۰۰ نمونه‌ی خودگردان به صورت زیر است:

$$\bar{\hat{Y}}_{i,FH} = \frac{1}{1000}\sum_{u=1}^{1000}\left[N_i\left(\hat{\gamma}_{iu}y_{iu} + (1-\hat{\gamma}_{iu})X'_i\hat{\beta}\right)\right]; \quad i = 1,\dots,16 \qquad (٦)$$

جدول ۶: میانگین توان دوم خطای براورد محصول پرتقال در شهرستان‌های استان فارس در سال ۱۳۸۲

ردیف	شهرستان	$g_{1i}(A)$	$g_{2i}(A)$	$g_{3i}(A)$	$\widehat{MSE}(\hat{\theta}_{i,FH})$	$\widehat{MSE}(\hat{Y}_{i,FH})$
۱	استهبان	۰/۶۳	۰/۰۰۰۱	۰/۰۰۴	۰/۶۴	۱/۴۸
۲	جهرم	۰/۱۹	۰/۰۰۰۲	۰/۰۰۰۳	۰/۱۹	۲۲۸/۳۸
۳	داراب	۰/۶۴	۰/۰۰۳	۰/۰۰۴	۰/۶۵	۵۸۶/۱۷
۴	زرین دشت	۰/۰۳	۰	۰/۰۰۰۱	۰/۰۳	۰/۰۰۰۲
۵	شیراز	۰/۰۳	۰	۰/۰۰۰۱	۰/۰۳۹	۰/۰۰۳
۶	فراشبند	۰/۰۰۱	۰	۰	۰/۰۰۱	۰/۰۰۰۱
۷	فسا	۹	۲/۸۱	۰/۴۵	۱۲/۲۵	۶۷۹/۹
۸	فیروزآباد	۰/۰۱	۰	۰	۰/۰۱	۰/۰۲
۹	قیروکارزین	۴/۲۲	۰/۲۷	۰/۱۲	۴/۶۱	۲۸/۲۴
۱۰	کازرون	۱/۳۶	۰/۰۳۵	۰/۰۱	۱/۴	۶۸/۶
۱۱	لارستان	۶/۶۳	۰/۱۱۸	۰/۱۸	۶/۹۳	۱۴/۹۷
۱۲	لامرد	۰/۰۱۲	۰	۰	۰/۰۱۲	۰/۰۰۰۱
۱۳	مرودشت	۰/۰۰۱	۰	۰	۰/۰۰۱	۰/۰۰۰۰۲
۱۴	ممسنی	۰/۳۳	۰	۰/۰۰۱	۰/۳۳	۸/۱۵
۱۵	مهر	۰/۰۲۲	۰	۰	۰/۰۲۲	۰/۰۰۰۶
۱۶	نیریز	۰/۲۱۸	۰	۰/۰۰۱	۰/۲۲	۰/۰۳۵

$$\bar{\hat{Y}}_i = \frac{1}{1000}\sum_{u=1}^{1000}\left[\sum_{j=1}^{n_{iu}}\frac{N_i}{n_{iu}}y_{iju}\right]; \quad i = 1, \ldots, 16 \tag{7}$$

فرمول‌های به‌کار رفته برای براوردهای خودگردان میانگین توان دوم خطای براوردگرهای فی–هریوت $\overline{\widehat{MSE}}(\hat{Y}_{i,FH})$ و مستقیم $\overline{\widehat{Var}}\left(\hat{Y}_i\right)$ طبق رابطه‌ی (۳) بر اساس ۱۰۰۰ نمونه‌ی خودگردان به صورت زیر است:

$$\overline{\widehat{MSE}}(\bar{\hat{Y}}_{i,FH}) = \frac{1}{1000}\sum_{u=1}^{1000}\left[N_i^2\widehat{MSE}(\hat{\theta}_{i,FH})\right]; \quad i = 1, \ldots, 16$$

$$\overline{\widehat{Var}}\left(\bar{\hat{Y}}_i\right) = \frac{1}{1000}\sum_{u=1}^{1000}\left[\sum_{j=1}^{n_{iu}}\frac{N_i^2}{n_{iu}^2}\widehat{Var}(\hat{Y}_{iju})\right]; \quad i = 1, \ldots, 16$$

۲.۴ یافته‌های مطالعه

براورد تولید کل محصول پرتقال بر اساس روش‌های براورد مستقیم و فی–هریوت با ۱۰۰۰ نمونه‌ی خودگردان طبق رابطه‌های (۶) و (۷) محاسبه و در جدول ۷ ارائه شده است.

به‌منظور مقایسه‌ی عمل‌کرد براوردگرهای مستقیم و فی–هریوت، براوردهای $\overline{\widehat{MSE}}(\hat{Y}_{i,FH})$ و $\overline{\widehat{Var}}\left(\hat{Y}_i\right)$ (در مقیاس ۰/۰۰۰۱ تولید) بر اساس ۱۰۰۰ نمونه‌ی خودگردان، در جدول ۸ آورده شده‌اند.

با توجه به جدول ۸ مشاهده می‌شود که در اغلب شهرستان‌ها، براوردگر فی–هریوت برای براورد تولید کل پرتقال برای ۱۶ شهرستان استان فارس عمل‌کرد بهتری نسبت به براوردگر مستقیم داشته است. همچنین با مقایسه‌ی براورد MSEی دو براوردگر مستقیم و فی–هریوت در جدول ۸ مشاهده می‌شود که براوردگرهای مستقیم دارای براورد MSEی بزرگ‌تر و در نتیجه دارای عدم دقت بیشتری نسبت به براوردهای فی–هریوت هستند.

جدول ۷: براورد فی-هریوت $\hat{Y}_{i,FH}$ براورد مستقیم، $\hat{Y}_i$، و مقدار واقعی، Y_i، حاصل از سرشماری کشاورزی سال ۱۳۸۲

ردیف	شهرستان	$\hat{Y}_{i,FH}$	$\hat{Y}_i$	Y_i
۱	استهبان	۱۱۰۵	۱۱۹۵	۱۰۸۳
۲	جهرم	۱۵۲٤۰	۱۵۰۵۷	۱۵۳٤۲
۳	داراب	٤۷۵۷۹	٤۹٤۸۰	۵۲۳۵۰
٤	زرین دشت	۱/۹	۲/۱۱	۲
۵	شیراز	۱۰/٦	۱۰/۲	۱٤
٦	فراشبند	۱۱/۲۲	۱۱/۳	۱۹
۷	فسا	۱٦۰۹٦	۱۷۲۲۵	۱۵۸٤۳
۸	فیروزآباد	۹۰	۱۰۷	۹۸
۹	قیروکارزین	۲۹۷٤	۳۱٤٦	۲۵۸۸
۱۰	کازرون	۸۱٤۸	۷۵۹٤	۸۸۰۱
۱۱	لارستان	۱۱۲۵	۱۳۲٤	۱۰٦۱
۱۲	لامرد	۱/۸	۱/۸	۲
۱۳	مرودشت	۱/۸	۱/٦	۲
۱٤	ممسنی	٦۲٤	٦۵۹	٦۳۵
۱۵	مهر	٤/۲	۳/۹۸	۷
۱٦	نیریز	۱۱۸/٦	۱۲۹/۷	۱۲۲

جدول ۸: مقایسه‌ی عمل‌کرد براوردگرهای فی-هریوت $\hat{Y}_{i,FH}$ و مستقیم، $\hat{Y}_i$

ردیف	شهرستان	$\widehat{MSE}(\hat{Y}_{i,FH})$	$\widehat{Var}\left(\hat{Y}_i\right)$
۱	استهبان	۲/۹۸	۲/۷۳
۲	جهرم	٤۹٤/۳	۵٦٦/۸
۳	داراب	۷۵۵/۰۱	۱۵۱۳/۲
٤	زرین دشت	۰/۰۰۰۱	۰/۰۰۰۲
۵	شیراز	۰/۰۰۰۲	۰/۰۰۱
٦	فراشبند	۰/۰۰۰۱	۰/۰۳
۷	فسا	۵۷۷/۵۳	۱۱۸۷
۸	فیروزآباد	۰/۰۳۳	۰/۲
۹	قیروکارزین	۲۲/۸	٤۹/۳٦
۱۰	کازرون	٤۵/٦	۵۹/۷۲
۱۱	لارستان	۱۳/۵	۲۵/۳٦
۱۲	لامرد	۰/۰۰۰۱	۰/۰۰۰۱
۱۳	مرودشت	۰/۰۰۰۱	۰/۰۰۰۱
۱٤	ممسنی	۱۳/٦	۱۹/۱۵
۱۵	مهر	۰/۰۰۰٦	۰/۰۰۰٤
۱٦	نیریز	۰/۰۱۷	۰/۱
	میانگین	۱۰۷/۸٦	۲۱۳/۹۳

۵ نتیجه‌گیری

در این مقاله برای براورد تولید کل محصول پرتقال در شهرستان‌های استان فارس (کوچک ناحیه‌ها) از رویکرد مدل‌مبنا (مدل فی-هریوت) استفاده شد. موضوع اساسی در براورد کوچک ناحیه‌ای، به‌دست آوردن براورد معتبر برای کوچک ناحیه‌ها و ارزیابی اعتبار این براوردها از طریق میانگین توان دوم خطای براوردگر است. در مطالعه‌ی شبیه‌سازی نشان داده شد که براوردهای فی-هریوت به‌دلیل دارا بودن MSEی کوچک‌تر و در نتیجه دقت بیشتر نسبت به براوردهای مستقیم، قابل قبول‌تر از براوردهای مستقیم برای براورد تولید پرتقال در سطح شهرستان‌های استان فارس در سال ۱۳۸۲ است.

لازم به ذکر است که برای به‌کارگیری مدل مناسب توجه به رابطه متغیر پاسخ و متغیر کمکی مهم است. چرا که اگر مدل انتخابی مناسب نباشد اریبی بزرگی به وجود می‌آورد. هر چند براوردگرهای مدل‌مبنا گرایش به اریبی دارند ولی در قیاس با براوردگرهای طرح‌مبنا (براوردگر مستقیم) واریانس‌های کوچک‌تری در زمینه‌ی براوردهای کوچک ناحیه‌ای دارند. همچنین به‌کارگیری مدل صریح به‌دلیل اثر دادن تغییرات بین ناحیه‌ای در براورد پارامتر موردنظر، بهبود قابل توجه‌ای در براورد پارامترها ایجاد می‌کند.

مراجع

[1] Datta, G.S. and Lahiri, P. (2000). A Unified Measure of Uncertainty of Estimated Best Linear Unbiased Predictors in Small Area Estimation Problems, *Statistica Sinica*, **10**, 613-627.

[2] Diggle, P.J. Heagerty, P. Liang, K. Y. and Zeger, S. L. (2002). *Analysis of longitudinal data, Second edition*. Oxford: Oxford University Press.

[3] Fay, R.E. and Herriot, R.A. (1979). Estimation of Income from Small Places, An Application of James-Stein Procedures to Census Data, *Journal of the American Statistical Association*, **74**, 269-277.

[4] Jiang, J. and Lahiri, P. (2006). Mixed model Prediction and Small Area Estimation, *Sociedad de Estadsticia Investigation operativa*, **15**, No. 1.

[5] Lohr, SH. and Ybarra, L. (2008). Small Area Estimation when Auxiliary Information is Measured with Error, *Biometrika*, **95**, 4, 919–931.

[6] Rao, J.N.K. (2003). *Small Area Estimation*, John Wiley Sons Inc, New York.

[7] Searle, S.R. Casella, G. and McCulloch, C. E. (1992). *Variance components*, John Wiley Sons Inc, New York.