Archive of SID

Journal of Statistical Sciences, Autumn and Winter, 2022
Vol. 16, No. 2, pp 417-434
DOI: 10.29252/jss.16.2.417

Coefficients Estimation of Linear Regression Models Using Liu-Type Shrinkage Estimators

Zandi, Z.[1], Bevrani, H.[2]
[1]Department of Statistics, University of Kurdistan, Sanandaj, Iran.
[2]Department Statistics, University of Tabriz, Tabriz, Iran.

Corresponding author: Z. Zandi, zahrazandi163@yahoo.com

Received: 20/2/2022 **Revised:** 8/8/2022 **Accepted and Published Online:** 10/8/2022.

Introduction

In this study, we addressed parameter estimation in the linear regression model in the presence of multicollinearity when there exists some prior information about predictor variables that appears as a linear restriction on the model parameters. We estimated the parameters based on Liu-type linear shrinkage, preliminary test, Stein, and positive Stein strategies. The performance of the proposed estimators is compared to the Liu-type estimator in terms of their relative efficiency via a Monte Carlo simulation study and an actual data set.

Material and Methods

In the linear regression model, the ordinary least squares (OLS) estimator is the best linear unbiased estimator for model parameters when the predictor variables are independent. The multicollinearity problem arises when there exists near linear dependence among the predictor variables. This problem leads to variance inflation of the OLS estimator. Thus the interpretations based on it are not true. The ridge and Liu-type estimators are two methods to combat multicollinearity. The Liu-type estimator is more efficient than the ridge estimator when there is a strong correlation between the predictor variables. We suppose that there is some prior information about parameter vector $\boldsymbol{\beta}$ under a linear restriction as $\mathbf{R}\boldsymbol{\beta} = \mathbf{r}$ where $\mathbf{R}$ is a $p_2 \times p$ matrix and $\mathbf{r}$ is a $p_2 \times 1$ vector. The restricted estimator of $\boldsymbol{\beta}$ is obtained by maximizing the log-likelihood function of the linear regression model under the linear restriction. The Liu-type restricted estimator can be defined in

www.SID.ir

Archive of SID

the presence of multicollinearity under the linear restriction. We propose the Liu-type shrinkage estimators using the Liu-type and Liu-type restricted estimators to improve the estimation of parameters. We compare the performance of the Liu-type shrinkage estimators and the Liu-type estimator in terms of their relative efficiency using a Monte Carlo simulation study. The simulation is conducted under different sample sizes, $n = 30, 50$, the correlation level between the predictor variables $\rho = 0.80, 0.90, 0.95$, $p_1 = 5$, and $p_2 = 3, 5, 7$. To investigate the behavior of the proposed estimators, we define $\Delta = \|\boldsymbol{\beta} - \boldsymbol{\beta}^0\|^2$, where $\|.\|$ is the Euclidean norm, $\boldsymbol{\beta}$ is the parameters vector in the simulated model and $\boldsymbol{\beta}^0$ is the true parameters vector in the candidate sub-model. We also apply the proposed estimation methods to a real data set.

Results and Discussion
The simulation results show that all estimators' performances become better when p_2 and ρ increase for fixed n. For all combinations of p_2, ρ, and n, the Liu-type restricted estimator has the best performance at $\Delta = 0$. As Δ moves away from zero, all estimators' simulated relative efficiencies (SREs) decrease. As λ approaches one, the performance of the Liu-type linear shrinkage estimator increases.

Conclusion
This paper suggested the Liu-type shrinkage estimators in the linear regression model in the presence of multicollinearity under the subspace information. A Monte Carlo simulation was conducted to compare the proposed estimators' performance with the Liu-type estimator. The simulation results confirm that the proposed estimators perform better than the Liu-type estimator when $\Delta = 0$ and near it for all p_2, ρ, and n.

Keywords: Linear regression model, Multicollinearity, Shrinkage estimators, Monte Carlo simulation.

Mathematics Subject Classification (2010): 62F10, 62J07.

©The Author(s). The Publisher is Iranian Statistical Society.
This is an open access article distributed under the terms and conditions of (CC BY-NC 4.0)

www.SID.ir

Archive of SID

مجله علوم آماری، پاییز و زمستان ۱۴۰۱
جلد ۱۶، شماره ۲، ص ۴۱۷ – ۴۳۴
DOI: 10.29252/jss.16.2.417
مقاله پژوهشی

برآورد ضرایب مدل رگرسیون خطی با استفاده از برآوردگرهای انقباضی نوع-لیو

زهرا زندی[۱]، حسین بیورانی[۲]
[۱]گروه آمار دانشگاه کردستان
[۲]گروه آمار دانشگاه تبریز

نویسنده مسئول : زهرا زندی، zahrazandi163@yahoo.com

تاریخ دریافت: ۱۴۰۰/۱۲/۲۰ تاریخ بازنگری: ۱۴۰۱/۵/۱۷ تاریخ پذیرش و انتشار: ۱۴۰۱/۵/۱۹

چکیده: این مقاله برآوردگرهای انقباضی نوع-لیو را برای ضرایب مدل رگرسیونی خطی با حضور هم‌خطی چندگانه تحت اطلاعات زیرفضا پیشنهاد می‌دهد. عملکرد برآوردگرهای معرفی شده از نظر کارایی نسبی آن‌ها از طریق شبیه‌سازی مونت کارلو و یک مجموعه داده واقعی با برآوردگر نوع-لیو مقایسه می‌شود. نتایج آشکار می‌کنند که برآوردگرهای معرفی شده نسبت به برآوردگر نوع-لیو عملکرد بهتری دارند.

واژه‌های کلیدی: مدل رگرسیون خطی، هم‌خطی چندگانه، برآوردگرهای انقباضی نوع-لیو، شبیه‌سازی مونت کارلو.

کد موضوع‌بندی ریاضی (۲۰۱۰): 62J07، 62F10.

۱ مقدمه

مدل رگرسیون خطی $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ را در نظر بگیرید که در آن $\mathbf{y} = (y_1, \ldots, y_n)'$ بردار متغیر پاسخ، $\mathbf{X}$ یک ماتریس طرح از مرتبه $n \times p$، $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)'$ بردار ضرایب رگرسیونی مجهول و $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_n)'$ بردار اثرات تصادفی است که مؤلفه‌های آن به‌صورت مستقل و هم‌توزیع دارای توزیع نرمال با میانگین ۰ و واریانس متناهی σ^2 هستند. بنابراین بردار متغیر پاسخ $\mathbf{y}$ دارای توزیع نرمال n-متغیره $N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\boldsymbol{I}_n)$ است که در آن

©نویسنده(گان). ناشر انجمن آمار ایران است.
این مقاله با دسترسی آزاد تحت شرایط و ضوابط (CC BY-NC 4.0) توزیع شده است.

www.SID.ir

Archive of SID

$\boldsymbol{I}_n$ ماتریس همانی از مرتبه $n \times n$ است. لگاریتم تابع درستنمایی مدل رگرسیون خطی به‌صورت

$$\mathcal{L}(\boldsymbol{\beta}; \mathbf{y}) = (-\frac{n}{۲}) \ln(۲\pi\sigma^{۲}) - \frac{۱}{۲\sigma^{۲}}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}), \quad (۱)$$

تعریف می‌شود. یک روش معمول برای برآورد ضرایب رگرسیونی در مدل رگرسیون خطی، برآوردگر کمترین توان‌های دوم عادی[1] (OLS) به‌صورت $\hat{\boldsymbol{\beta}}^{OLS} = (\mathbf{X}'\mathbf{X})^{-۱}\mathbf{X}'\mathbf{y}$ است. این برآوردگر با استفاده از اطلاعات نمونه‌ای و با مینیمم کردن مجموع توان‌های دوم خطای مدل محاسبه می‌شود که بهترین برآوردگر نااریب خطی برای $\boldsymbol{\beta}$ است وقتی که متغیرهای پیشگو (x_{ij}) مستقل از هم باشند.

گاهی ممکن است در یک مدل رگرسیون خطی علاوه بر اطلاعات به‌دست آمده از طریق نمونه تصادفی، اطلاعاتی غیر از نمونه تصادفی موجود باشد که به اطلاعات پیشین نامطمئن معروف است و به‌صورت قید خطی $\mathbf{R}\boldsymbol{\beta} = \mathbf{r}$ روی بردار ضرایب رگرسیونی ظاهر می‌شود، که در آن $\mathbf{R}$ یک ماتریس $p_۲ \times p$ با عناصر ثابت به‌طوری‌که $p_۲ < p$ و $\mathbf{r}$ یک بردار $p_۲ \times ۱$ با مؤلفه‌های ثابت است (احمد ، ۲۰۱۴). برآورد ضرایب رگرسیونی $\boldsymbol{\beta}$ بر اساس اطلاعات نمونه‌ای همان برآوردگر کمترین توان‌های دوم تحت عنوان برآوردگر نامقید[2] است. برآورد ضرایب رگرسیونی $\boldsymbol{\beta}$ تحت قید خطی را برآوردگر مقید[3] می‌نامند و با ماکسیمم کردن لگاریتم تابع درستنمایی (۱) نسبت به $\boldsymbol{\beta}$، تحت این قید به صورت

$$\hat{\boldsymbol{\beta}}^{RE} = \hat{\boldsymbol{\beta}}^{OLS} - (\mathbf{X}'\mathbf{X})^{-۱}\mathbf{R}'\Big(\mathbf{R}\,(\mathbf{X}'\mathbf{X})^{-۱}\mathbf{R}'\Big)^{-۱}\Big(\mathbf{R}\,\hat{\boldsymbol{\beta}}^{OLS} - \mathbf{r}\Big). \quad (۲)$$

به‌دست می‌آید. در صورت درست بودن اطلاعات پیشین نامطمئن، با استفاده از برآوردگرهای نامقید و مقید می‌توان برآوردگرهای کاراتری برای ضرایب رگرسیونی معرفی کرد که به برآوردگرهای انقباضی معروف‌اند و شامل برآوردگرهای انقباضی خطی[4] (تامپسون ، ۱۹۶۸)، پیش‌آزمون[5] (بنکرافت ، ۱۹۴۴)، استاین[6] و استاین مثبت[7] (استاین ، ۱۹۵۶) هستند. هر اطلاعات پیشینی قبل از به‌کارگیری در مدل از طریق فرضیه $H_۰ : \mathbf{R}\,\boldsymbol{\beta} = \mathbf{r}$ در مقابل $H_۱ : \mathbf{R}\,\boldsymbol{\beta} \neq \mathbf{r}$، با آماره آزمون T_n آزمون می‌شوند. در صورت رد نشدن $H_۰$ اطلاعات پیشین وارد مدل می‌شود (احمد ، ۲۰۱۴؛ صالح ، ۲۰۰۶). برآوردگرهای انقباضی در مدل‌های رگرسیونی مختلف به‌کار برده شده‌اند (زندی و همکاران ، ۲۰۲۱؛ یوزباشی و همکاران ، ۲۰۲۰؛ لیساوادی و همکاران ، ۲۰۲۰).

مسأله هم‌خطی چندگانه در مدل رگرسیون خطی وقتی اتفاق می‌افتد که وابستگی خطی بین متغیرهای پیشگو

[1] Ordinary least squares
[2] Unrestricted estimator
[3] Restricted estimator
[4] Linear shrinkage
[5] Pretest
[6] Stein
[7] Positive Stein

www.SID.ir

Archive of SID

وجود داشته باشد. در این حالت ستون‌های ماتریس $\mathbf{X}$ متعامد نیستند که باعث می‌شود دترمینان ماتریس $\mathbf{X}'\mathbf{X}$ به صفر نزدیک شود و در نتیجه عناصر ماتریس $(\mathbf{X}'\mathbf{X})^{-1}$ بسیار بزرگ شوند. از طرفی واضح است که $Cov(\hat{\boldsymbol{\beta}}^{OLS}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$، بنابراین هم‌خطی چندگانه باعث تورم ماتریس واریانس کوواریانس برآوردگرهای کمترین توان‌های دوم خطا می‌شود. در نتیجه استنباط‌ها براساس این برآوردگر، نامعتبر هستند. برای مقابله با مشکل هم‌خطی چندگانه، هورل و کنارد (۱۹۷۰) برآوردگر ریج را به‌صورت

$$\hat{\boldsymbol{\beta}}^{Ridge} = (\mathbf{X}'\mathbf{X} + k\,\boldsymbol{I}_p)^{-1}\mathbf{X}'\mathbf{y}, \tag{۳}$$

معرفی کردند، که در آن $k > 0$ پارامتر ریج و $\boldsymbol{I}_p$ یک ماتریس همانی از مرتبه $p \times p$ است. یک روش دیگر برای مقابله با مشکل هم‌خطی چندگانه، برآوردگر نوع-لیو به صورت

$$\hat{\boldsymbol{\beta}}^{LT} = (\mathbf{X}'\mathbf{X} + k\,\boldsymbol{I}_p)^{-1}(\mathbf{X}'\mathbf{X} - d\,\boldsymbol{I}_p)\hat{\boldsymbol{\beta}}^{OLS}, \quad -\infty < d < \infty. \tag{۴}$$

است. وقتی که همبستگی قوی (نزدیک یک) بین متغیرهای پیشگو وجود داشته باشد، ماتریس $\mathbf{X}'\mathbf{X} + k\,\boldsymbol{I}_p$ در برآوردگر ریج هنوز بدشرطیده [۱] است. در این حالت، برآوردگر نوع-لیو عملکرد بهتری نسبت به برآوردگر ریج دارد. میانگین توان‌های دوم خطای (MSE) برآوردگر نوع-لیو در مدل خطی به‌صورت

$$MSE(\hat{\boldsymbol{\beta}}^{LT}) = \sum_{i=1}^{p}\frac{(d+k)^2\alpha_i^2}{(\lambda_i+k)^2} + \sigma^2\sum_{i=1}^{p}\frac{(d-\lambda_i)^2}{\lambda_i(\lambda_i+k)^2}, \tag{۵}$$

است، که در آن $\lambda_1 \geq \ldots \geq \lambda_p > 0$ مقادیر ویژه مرتب شده ماتریس $\mathbf{X}'\mathbf{X}$، $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_p)' = \mathbf{C}'\boldsymbol{\beta}$ و $\mathbf{C}$ یک ماتریس متعامد است که ستون‌های آن تشکیل بردارهای ویژه ماتریس $\mathbf{X}'\mathbf{X}$ را می‌دهند. برآورد مقدار بهینه پارامتر d با مینیمم کردن (۵) نسبت به این پارامتر به‌صورت (لیو ، ۲۰۰۳)

$$\hat{d}_{\text{بهینه}} = \frac{\sum_{i=1}^{p}\frac{\hat{\sigma}^2 - k\hat{\alpha}_i^2}{(\lambda_i+k)^2}}{\sum_{i=1}^{p}\frac{\hat{\sigma}^2 + \lambda_i\hat{\alpha}_i^2}{\lambda_i(\lambda_i+k)^2}},$$

به‌دست می‌آید، که در آن $\hat{\boldsymbol{\alpha}} = (\hat{\alpha}_1, \ldots, \hat{\alpha}_p)' = \mathbf{C}'\hat{\boldsymbol{\beta}}^{OLS}$ و $\hat{\sigma}^2 = \frac{\left(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}}^{OLS}\right)'\left(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}}^{OLS}\right)}{n-p}$. براساس ایده هورل و همکاران (۱۹۷۰)، برآورد پارامتر k به‌صورت $\hat{k} = \frac{p\hat{\sigma}^2}{(\hat{\boldsymbol{\beta}}^{OLS})'(\hat{\boldsymbol{\beta}}^{OLS})}$ به دست می‌آید. علاوه بر این، می‌توان از روش اعتبارسنجی متقابل تعمیم‌یافته نیز برای محاسبه مقدار بهینه پارامترهای d و k استفاده کرد (امینی و روزبه ، ۲۰۱۸؛ روزبه ، ۲۰۱۸). برآوردگر نوع-لیو در مدل دوجمله‌ای منفی (آسار ، ۲۰۱۸) و در مدل پواسن

[1] Ill-conditioned

www.SID.ir

Archive of SID

آماسیده در صفر (آسار و همکاران ، ۲۰۱۸) به‌کار برده شده است. روزبه و امینی (۱۳۹۸) و روزبه و معنوی (۱۳۹۹) به بررسی مسأله هم‌خطی چندگانه در مدل‌های رگرسیونی پرداخته‌اند.

هدف اصلی ما در این مقاله، برآورد ضرایب رگرسیونی در مدل خطی تحت اطلاعات زیرفضا [۱] و وجود وابستگی خطی نزدیک بین متغیرهای پیشگو است. بنابراین، برآوردگرهای انقباضی نوع-لیو در این مدل معرفی می‌شوند که شامل برآوردگرهای انقباضی خطی، پیش‌آزمون، استاین و استاین مثبت نوع-لیو هستند. با استفاده از شبیه‌سازی مونت کارلو و یک مثال واقعی، عملکرد برآوردگرهای معرفی شده نسبت به برآوردگر نوع-لیو از طریق کارایی نسبی آن‌ها در قسمت‌های مختلف فضای پارامتر مقایسه می‌شود. در ادامه بخش‌های مقاله عبارتند از در بخش ۲ برآوردگر مقید نوع-لیو و برآوردگرهای انقباضی نوع-لیو در مدل رگرسیون خطی معرفی می‌شود. در بخش ۳ شبیه‌سازی مونت کارلو برای مقادیر مختلف از همبستگی و تعداد متغیرهای پیشگوی مختلف انجام می‌شود. در بخش ۴ عملکرد برآوردگرهای معرفی شده در یک مثال واقعی بررسی می‌شود و بخش آخر شامل بحث و نتیجه‌گیری است.

۲ برآوردگرهای معرفی شده

در حضور هم‌خطی چندگانه، برآوردگر نوع-لیو (۴) با استفاده از اطلاعات نمونه‌ای به‌دست آمده است که می‌توان آن را برآوردگر نامقید نوع-لیو نامید. در صورت وجود اطلاعات پیشین نامطمئن در مورد زیرمجموعه‌ای از متغیرهای پیشگو، برآوردگر مقید نوع-لیو برای ضرایب رگرسیونی به‌صورت

$$\hat{\boldsymbol{\beta}}^{LTR} = (\mathbf{X}'\mathbf{X} + k\,\mathbf{I}_p)^{-1}(\mathbf{X}'\mathbf{X} - d\,\mathbf{I}_p)\,\hat{\boldsymbol{\beta}}^{RE},$$

تعریف می‌شود، که در آن $\hat{\boldsymbol{\beta}}^{RE}$ برآوردگر مقید (۲) است. حال با استفاده از برآوردگرهای (نامقید) نوع-لیو و مقید نوع-لیو، برآوردگرهای انقباضی نوع-لیو در مدل رگرسیون خطی معرفی می‌شوند.

الف- برآوردگر انقباضی خطی نوع-لیو: برآوردگر انقباضی خطی نوع-لیو به‌صورت ترکیب خطی محدب از $\hat{\boldsymbol{\beta}}^{LT}$ و $\hat{\boldsymbol{\beta}}^{LTR}$ به صورت

$$\hat{\boldsymbol{\beta}}^{LTLS} = \lambda\hat{\boldsymbol{\beta}}^{LTR} + (1-\lambda)\hat{\boldsymbol{\beta}}^{LT},\ \ 0 \le \lambda \le 1,$$

تعریف می‌شود، که در آن λ پارامتر انقباضی است که میزان اطمینان از اطلاعات پیشین را نشان می‌دهد. هر چه این پارامتر به یک نزدیکتر باشند، نشان دهنده اطمینان بیشتر در مورد درست بودن اطلاعات پیشین است. واضح است به ازای $\lambda = 1$، $\hat{\boldsymbol{\beta}}^{LTLS} = \hat{\boldsymbol{\beta}}^{LTR}$ و برای $\lambda = 0$، $\hat{\boldsymbol{\beta}}^{LTLS} = \hat{\boldsymbol{\beta}}^{LT}$ است.

---
[1] Subspace information

www.SID.ir

Archive of SID

ب- برآوردگر پیش‌آزمون نوع-لیو: برآوردگر پیش‌آزمون نوع-لیو به‌صورت

$$\hat{\boldsymbol{\beta}}^{LTPT}=\hat{\boldsymbol{\beta}}^{LT}-I(T_n\leq T_{n,\alpha})\Big(\hat{\boldsymbol{\beta}}^{LT}-\hat{\boldsymbol{\beta}}^{LTR}\Big),$$

تعریف می‌شود، که در آن $I(\cdot)$ یک تابع نشانگر، T_n آماره آزمون نسبت درستنمایی برای آزمون فرض H_0 در مقابل H_1 است که براساس ایده احمد (۲۰۱۴) به‌صورت

$$T_n=\frac{(\mathbf{R}\hat{\boldsymbol{\beta}}^{OLS}-\mathbf{r})'\big(\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'\big)^{-1}(\mathbf{R}\hat{\boldsymbol{\beta}}^{OLS}-\mathbf{r})}{\hat{\sigma}^2},$$

تعریف می‌شود. ایتچسون و سیلوی (۱۹۵۸) نشان دادند، تحت درست بودن فرض H_0، وقتی $n\to\infty$، آماره آزمون T_n دارای توزیع $\chi^2(p_2)$ است و $T_{n,\alpha}$ مقدار بحرانی سطح-α مربوط به کران بالای توزیع کای-دو است. یعنی، $T_{n,\alpha}=\chi^2_{1-\alpha}(p_2)$. در برآوردگر پیش‌آزمون نوع-لیو در صورتی‌که فرض H_0 رد نشود، $I\big(T_n\leq\chi^2_{1-\alpha}(p_2)\big)=1$ است. آن‌گاه $\hat{\boldsymbol{\beta}}^{LTPT}=\hat{\boldsymbol{\beta}}^{LTR}$. اگر فرض H_0 رد شود، یعنی اطلاعات پیشین صحیح نیست. در نتیجه $I\big(T_n\leq\chi^2_{1-\alpha}(p_2)\big)=0$ و $\hat{\boldsymbol{\beta}}^{LTPT}=\hat{\boldsymbol{\beta}}^{LT}$ است. ضعف برآوردگر پیش‌آزمون نوع-لیو این است که منجر به انتخاب یکی از برآوردگرهای (نامقید) نوع-لیو یا مقید نوع-لیو می‌شود.

ج- برآوردگر استاین نوع-لیو: برآوردگر استاین نوع-لیو به‌صورت تابعی از $\hat{\boldsymbol{\beta}}^{LT}$، $\hat{\boldsymbol{\beta}}^{LTR}$ و آماره آزمون T_n تعریف می‌شود.

$$\hat{\boldsymbol{\beta}}^{LTS}=\hat{\boldsymbol{\beta}}^{LTR}+\Big(1-\frac{p_2-2}{T_n}\Big)\Big(\hat{\boldsymbol{\beta}}^{LT}-\hat{\boldsymbol{\beta}}^{LTR}\Big)\qquad p_2\geq 3.$$

د- برآوردگر استاین مثبت نوع-لیو: برای کنترل انقباض زیاد در برآوردگر استاین نوع-لیو، با در نظر گرفتن تنها مقادیر مثبت برای تابع وزن $1-\frac{p_2-2}{T_n}$ در $\hat{\boldsymbol{\beta}}^{LTS}$، برآوردگر استاین مثبت نوع-لیو به‌صورت

$$\hat{\boldsymbol{\beta}}^{LTPS}=\hat{\boldsymbol{\beta}}^{LTR}+\Big(1-\frac{p_2-2}{T_n}\Big)^{+}\Big(\hat{\boldsymbol{\beta}}^{LT}-\hat{\boldsymbol{\beta}}^{LTR}\Big),\qquad p_2\geq 3,$$

تعریف می‌شود، که در آن $a^+=\max(0,a)$. برتری برآوردگرهای استاین نوع-لیو نسبت به $\hat{\boldsymbol{\beta}}^{LTLS}$ و $\hat{\boldsymbol{\beta}}^{LTPT}$ این است که عملکرد آن‌ها وابسته به λ و α نیست (لیساوادی و همکاران، ۲۰۲۰).

۳ شبیه‌سازی مونت کارلو

این بخش به مقایسه عملکرد برآوردگرهای انقباضی نوع-لیو نسبت به برآوردگر (نامقید) نوع-لیو در مدل‌های رگرسیون خطی با اجرای شبیه‌سازی مونت کارلو در نرم‌افزار آماری R می‌پردازد. شبیه‌سازی‌ها به‌ازای حجم نمونه $n=35,\ 50$

www.SID.ir

Archive of SID

و تعداد متغیرهای پیشگوی $p = p_1 + p_2 = 8, 10, 12$ انجام می‌شود به‌طوری‌که به‌ترتیب $p_2 = 3, 5, 7$ تا از آن‌ها تأثیر منعی‌داری روی متغیر پاسخ ندارند و متغیرهای پیشگوی غیرفعال هستند. به عبارت دیگر، در قید خطی $\mathbf{R}\beta = \mathbf{r}$، فرض می‌شود $\mathbf{R} = [0_{p_2 \times p_1}, \boldsymbol{I}_{p_2 \times p_2}]$ و $\mathbf{r} = 0_{p_2 \times 1}$، که در آن 0 یک ماتریس با عناصر صفر و $\boldsymbol{I}$ یک ماتریس همانی است. بنابراین $p_1 = 5$ تعداد متغیرهای پیشگوی معنی‌دار (فعال) است. به منظور بالا بردن دقت عملکرد برآوردگرها برای همه حالت‌های شبیه‌سازی شده، تعداد تکرارها برابر ۵۰۰۰ در نظر گرفته می‌شود. متغیرهای پیشگو براساس ایده آسار و همکاران (۲۰۱۸) به‌صورت

$$x_{ij} = u_{ij}\sqrt{1-\rho^2} + \rho\, u_{ip}, \quad i = 1\ldots, n, \quad j = 1, \ldots, p,$$

تولید می‌شوند، که در آن $\mathbf{x}_i = (x_{i1}, \ldots, x_{ip})'$، i-امین سطر از ماتریس طرح $\mathbf{X}$، ρ نشان دهنده میزان همبستگی بین هر دو متغیر پیشگو و u_{ij}ها به‌صورت مستقل از توزیع نرمال استاندارد تولید و شبیه‌سازی‌ها به‌ازای مقادیر $\rho = 0/8, 0/9, 0/95$ انجام می‌شوند. همچنین متغیر پاسخ y_i به‌صورت

$$y_i = \beta_1 x_{i1} + \ldots + \beta_p x_{ip} + \varepsilon_i, \quad i = 1, \ldots, n,$$

تولید می‌شود، که در آن ε_iها به‌صورت مستقل و همتوزیع دارای توزیع نرمال $N(0, 0/9)$ هستند.

به منظور بررسی رفتار برآوردگرهای معرفی شده در قسمت‌های مختلف از فضای پارامتر، فاصله بین مدل درست و مدل شبیه‌سازی شده به‌صورت $\Delta = \|\boldsymbol{\beta} - \boldsymbol{\beta}^\circ\|^2$ تعریف می‌شود، که در آن $\|\cdot\|$ نرم اقلیدسی، $\boldsymbol{\beta}$ بردار ضرایب رگرسیونی در مدل شبیه‌سازی شده و $\boldsymbol{\beta}^\circ$ بردار ضرایب رگرسیونی در مدل درست است. برای تولید y_iها قرار داده می‌شود $\boldsymbol{\beta}^\circ = (\boldsymbol{\beta}_1', \boldsymbol{\beta}_2')' = (\boldsymbol{\beta}_1', 0'_{p_2 \times 1})'$، که در آن $\boldsymbol{\beta}_1 = (0/25, 0/53, 0/62, 0/81, 0/3)'$ و 0 یک بردار $p_2 \times 1$ با عناصر صفر است. در شبیه‌سازی‌ها همه حالت‌ها به‌ازای $\Delta = \{0, 0/5, 1, 1/5, 2\}$ انجام می‌شود. مقدار $\Delta = 0$، به معنی درست بودن فرض صفر است. در این حالت $\boldsymbol{\beta} = \boldsymbol{\beta}^\circ$ است. وقتی‌که $\Delta > 0$ است، بردار ضرایب در مدل شبیه‌سازی شده به‌صورت $\boldsymbol{\beta} = (\boldsymbol{\beta}_1', \boldsymbol{\beta}_2')' = (\boldsymbol{\beta}_1', c, 0'_{(p_2-1)\times 1})'$ در نظر گرفته می‌شود، که در آن $c = \sqrt{\Delta}$.

برای بررسی عملکرد برآوردگرهای معرفی شده نسبت به برآوردگر (نامقید) نوع-لیو از معیار کارایی نسبی شبیه‌سازی شده[1] (SRE) به‌صورت

$$SRE(\hat{\boldsymbol{\beta}}^{LT}, \hat{\boldsymbol{\beta}}^*) = \frac{SMSE(\hat{\boldsymbol{\beta}}^{LT})}{SMSE(\hat{\boldsymbol{\beta}}^*)},$$

استفاده می‌شود، که در آن $\hat{\boldsymbol{\beta}}^*$ هر یک از برآوردگرهای مقید، انقباضی خطی، پیش‌آزمون، استاین و استاین مثبت

[1] Simulated relative efficiency

www.SID.ir

Archive of SID

نوع-لیو است و $SMSE(\hat{\boldsymbol{\beta}}^*)$ میانگین توان‌های دوم خطای شبیه‌سازی شده برآوردگر $\hat{\boldsymbol{\beta}}^*$ است.

$$SMSE(\hat{\boldsymbol{\beta}}^*) = \frac{۱}{۵۰۰۰}\sum_{t=۱}^{۵۰۰۰}(\hat{\boldsymbol{\beta}}^* - \boldsymbol{\beta})'_t(\hat{\boldsymbol{\beta}}^* - \boldsymbol{\beta})_t.$$

مقدار بزرگتر از یک برای کارایی نسبی شبیه‌سازی شده، نشان دهنده برتری $\hat{\boldsymbol{\beta}}^*$ نسبت به $\hat{\boldsymbol{\beta}}^{LT}$ است.

کارایی‌های نسبی شبیه‌سازی شده برآوردگرهای در جداول ۱ و ۲ و شکل‌های ۱ تا ۶ گزارش شده‌اند. در ادامه خلاصه نتایج بیان می‌شود.

۱- برای حجم نمونه ثابت، با افزایش تعداد متغیرهای پیشگوی غیرفعال $(p_۲)$، کارایی همه برآوردگرها معرفی شده افزایش می‌یابد.

۲- برای همه ترکیب‌های $p_۲$، ρ و n، برآوردگر مقید نوع-لیو دارای بهترین عملکرد است وقتی که اطلاعات پیشین درست باشد. یعنی، وقتی که $\Delta = ۰$ است. وقتی که مقدار Δ از صفر دور می‌شود، کارایی همه برآوردگرها کاهش می‌یابد. به‌عبارت دیگر، هر چه از مدل درست دور شویم، عملکرد برآوردگرها بدتر می‌شود. با این وجود، در این حالت عملکرد برآوردگرهای استاین نوع-لیو بهتر از دیگر برآوردگرها است.

۳- کارایی نسبی شبیه‌سازی شده همه برآوردگرها با افزایش ρ افزایش می‌یابد.

۴- عملکرد برآوردگر انقباضی خطی نوع-لیو با پارامتر λ رابطه مستقیم دارد. هر چه مقدار آن به عدد یک نزدیکتر باشد، عملکرد این برآوردگر بهتر است.

۵- به ازای مقدار کوچک α، کارایی برآوردگر پیش‌آزمون نوع-لیو بیشتر است. وقتی که فرض $H_۰$ درست است، این برآوردگر بهتر عمل می‌کند.

۶- وقتی Δ برابر صفر یا نزدیک صفر است، عملکرد برآوردگر استاین مثبت نوع-لیو بهتر از برآوردگر استاین نوع-لیو است.

۴ تحلیل داده واقعی

مجموعه داده ایالت مربوط به ۵۰ ایالت آمریکا هستند که در نرم‌افزار R به‌صورت state.x۷۷ موجود است. داده‌ها شامل ۸ متغیر است که در آن متغیر امید به زندگی در سال‌های ۱۹۶۹ تا ۱۹۷۱ به‌عنوان متغیر پاسخ در نظر گرفته می‌شود. در جدول ۳ متغیرهای پیشگو معرفی شده‌اند. جدول ۴ شامل ماتریس همبستگی بین متغیرهای پیشگو است که براساس آن بین بعضی از متغیرها همبستگی قوی وجود دارد. به‌عبارت دیگر، بین متغیرهای پیشگو هم‌خطی چندگانه وجود دارد. هدف، برازش یک مدل رگرسیون خطی به این داده‌ها و برآورد ضرایب رگرسیونی براساس برآوردگرهای انقباضی نوع-لیو و مقایسه کارایی نسبی آن‌ها نسبت به برآوردگر (نامقید) نوع-لیو است. روش‌های مختلفی برای تشخیص متغیرهای معنی‌دار وجود دارد که از جمله می‌توان به روش‌های لاسو و لاسوی انقباضی و

www.SID.ir

Archive of SID

جدول ۱. کارایی نسبی برآوردگرها نسبت به برآوردگر نوع-لیو برای $n = ۳۵$

$LTPS$	LTS	$LTPT$		$LTLS$		LTR	Δ	$p_۲$	ρ
		α		λ					
		۰/۰۵	۰/۰۱	۰/۸	۰/۵				
۱/۱۳۱	۱/۱۱۹	۱/۲۴۳	۱/۴۱۵	۱/۶۱۳	۱/۴۲۰	۱/۶۵۸	۰/۰	۳	۰/۸۰
۱/۰۳۵	۱/۰۳۴	۰/۹۳۷	۰/۹۰۰	۱/۰۲۹	۱/۱۵۴	۰/۸۹۴	۰/۵		
۱/۰۰۴	۱/۰۰۴	۰/۹۷۷	۰/۹۴۰	۰/۵۶۲	۰/۸۲۸	۰/۴۲۶	۱/۰		
۰/۹۹۹	۰/۹۹۹	۰/۹۹۷	۰/۹۹۵	۰/۳۲۹	۰/۵۷۹	۰/۲۳۳	۱/۵		
۰/۹۹۸	۰/۹۹۸	۱/۰۰۰	۱/۰۰۰	۰/۲۱۰	۰/۴۱۳	۰/۱۴۴	۲/۰		
۱/۳۸۵	۱/۳۶۰	۱/۴۴۶	۱/۷۱۶	۱/۸۹۰	۱/۵۷۸	۱/۹۷۲	۰/۰	۵	
۱/۱۴۸	۱/۱۴۱	۰/۹۸۹	۱/۰۰۴	۱/۲۳۳	۱/۳۰۵	۱/۰۹۰	۰/۵		
۱/۰۳۹	۱/۰۳۸	۰/۹۶۵	۰/۹۰۸	۰/۶۹۳	۰/۹۶۶	۰/۵۳۴	۱/۰		
۱/۰۱۱	۱/۰۱۱	۰/۹۹۶	۰/۹۸۸	۰/۴۱۳	۰/۶۹۴	۰/۲۹۶	۱/۵		
۱/۰۰۳	۱/۰۰۳	۱/۰۰۰	۰/۹۹۹	۰/۲۶۷	۰/۵۰۴	۰/۱۸۴	۲/۰		
۱/۶۹۴	۱/۶۴۶	۱/۷۶۰	۲/۰۹۷	۲/۱۷۹	۱/۷۲۹	۲/۲۹۷	۰/۰	۷	
۱/۳۰۴	۱/۲۹۰	۱/۰۸۰	۱/۱۵۸	۱/۴۳۸	۱/۴۴۷	۱/۲۸۷	۰/۵		
۱/۱۰۲	۱/۱۰۱	۰/۹۴۹	۰/۸۸۸	۰/۸۲۲	۱/۰۹۲	۰/۶۴۰	۱/۰		
۱/۰۴۰	۱/۰۴۰	۰/۹۹۵	۰/۹۸۱	۰/۴۹۵	۰/۷۹۹	۰/۳۵۸	۱/۵		
۱/۰۱۷	۱/۰۱۷	۰/۹۹۸	۰/۹۹۸	۰/۳۲۱	۰/۵۸۹	۰/۲۲۴	۲/۰		
۱/۱۹۵	۱/۱۶۹	۱/۴۸۶	۱/۶۳۶	۱/۶۵۰	۱/۴۴۵	۱/۶۹۵	۰/۰	۳	۰/۹۰
۱/۰۸۷	۱/۰۷۶	۱/۰۰۹	۱/۰۵۳	۱/۲۱۳	۱/۲۵۶	۱/۱۰۳	۰/۵		
۱/۰۲۳	۱/۰۲۱	۰/۸۸۳	۰/۷۸۶	۰/۷۷۴	۱/۰۰۳	۰/۶۲۰	۱/۰		
۱/۰۰۶	۱/۰۰۵	۰/۹۵۱	۰/۸۸۳	۰/۴۹۷	۰/۷۷۲	۰/۳۶۹	۱/۵		
۱/۰۰۱	۱/۰۰۱	۰/۹۹۲	۰/۹۶۴	۰/۳۳۶	۰/۵۹۱	۰/۲۳۸	۲/۰		
۱/۶۱۴	۱/۵۱۹	۱/۹۲۵	۲/۰۵۱	۱/۹۷۹	۱/۶۲۷	۲/۰۶۸	۰/۰	۵	
۱/۳۲۰	۱/۲۷۹	۱/۲۴۸	۱/۳۱۶	۱/۴۷۰	۱/۴۳۱	۱/۳۶۰	۰/۵		
۱/۱۱۷	۱/۱۱۱	۰/۸۸۶	۰/۸۳۳	۰/۹۵۳	۱/۱۶۴	۰/۷۷۴	۱/۰		
۱/۰۴۷	۱/۰۴۶	۰/۹۱۰	۰/۸۰۵	۰/۶۲۰	۰/۹۱۳	۰/۴۶۵	۱/۵		
۱/۰۲۰	۱/۰۲۰	۰/۹۷۰	۰/۹۲۲	۰/۴۲۲	۰/۷۱۰	۰/۳۰۲	۲/۰		
۲/۱۵۳	۱/۸۴۶	۲/۴۰۳	۲/۴۸۰	۲/۳۴۶	۱/۸۱۱	۲/۴۹۰	۰/۰	۷	
۱/۶۱۹	۱/۴۸۱	۱/۵۶۷	۱/۶۲۳	۱/۷۴۹	۱/۶۰۴	۱/۶۳۹	۰/۵		
۱/۲۲۷	۱/۲۲۴	۰/۹۶۰	۰/۹۴۷	۱/۱۳۹	۱/۳۲۰	۰/۹۳۵	۱/۰		
۱/۱۱۱	۱/۱۰۸	۰/۸۶۱	۰/۷۶۲	۰/۷۴۴	۱/۰۴۸	۰/۵۶۲	۱/۵		
۱/۰۵۷	۱/۰۵۶	۰/۹۴۰	۰/۸۳۷	۰/۵۰۷	۰/۸۲۴	۰/۳۶۵	۲/۰		
۱/۳۳۵	۱/۲۶۵	۱/۷۵۹	۱/۷۹۰	۱/۷۳۷	۱/۴۹۷	۱/۷۹۰	۰/۰	۳	۰/۹۵
۱/۲۱۱	۱/۱۷۵	۱/۳۲۰	۱/۳۴۴	۱/۴۲۲	۱/۳۷۱	۱/۳۴۷	۰/۵		
۱/۰۹۷	۱/۰۸۶	۰/۹۰۲	۰/۸۹۰	۱/۰۳۸	۱/۱۸۶	۰/۸۸۲	۱/۰		
۱/۰۴۲	۱/۰۳۹	۰/۷۹۰	۰/۶۸۱	۰/۷۳۷	۰/۹۹۲	۰/۵۷۷	۱/۵		
۱/۰۲۰	۱/۰۱۶	۰/۸۴۴	۰/۶۸۷	۰/۵۳۱	۰/۸۱۶	۰/۳۹۴	۲/۰		
۲/۰۲۰	۱/۵۴۰	۲/۲۱۹	۲/۲۳۲	۲/۱۲۴	۱/۷۰۳	۲/۲۳۴	۰/۰	۵	
۱/۶۵۰	۱/۳۶۶	۱/۶۷۹	۱/۶۸۸	۱/۷۴۹	۱/۵۷۱	۱/۶۸۸	۰/۵		
۱/۳۰۳	۱/۱۹۵	۱/۱۰۳	۱/۱۱۰	۱/۲۸۷	۱/۳۷۶	۱/۱۱۱	۱/۰		
۱/۱۳۴	۱/۱۰۰	۰/۸۰۱	۰/۷۵۴	۰/۹۲۰	۱/۱۶۶	۰/۷۲۹	۱/۵		
۱/۰۶۴	۱/۰۵۴	۰/۷۳۵	۰/۶۲۱	۰/۶۶۶	۰/۹۷۱	۰/۴۹۸	۲/۰		
۲/۶۹۰	۱/۰۹۶	۲/۷۶۱	۲/۷۶۱	۲/۵۷۵	۱/۹۱۳	۲/۷۶۱	۰/۰	۷	
۲/۰۷۹	۱/۰۴۳	۲/۰۷۹	۲/۰۸۲	۲/۱۲۲	۱/۷۷۴	۲/۰۸۲	۰/۵		
۱/۴۸۸	۱/۰۲۱	۱/۳۵۸	۱/۳۶۲	۱/۵۶۰	۱/۵۶۶	۱/۳۶۳	۱/۰		
۱/۱۸۴	۱/۰۱۴	۰/۹۱۲	۰/۸۹۷	۱/۱۱۵	۱/۳۳۸	۰/۸۹۱	۱/۵		
۱/۰۵۹	۱/۰۰۷	۰/۷۱۴	۰/۶۵۲	۰/۸۰۷	۱/۱۲۴	۰/۶۰۸	۲/۰		

www.SID.ir

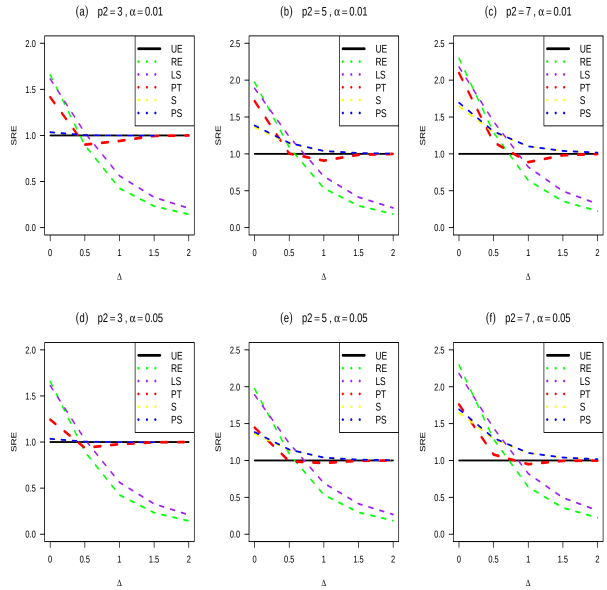
Archive of SID

جدول ۲. کارایی نسبی برآوردگرها نسبت به برآوردگر نوع-لیو برای $n = ۵۰$

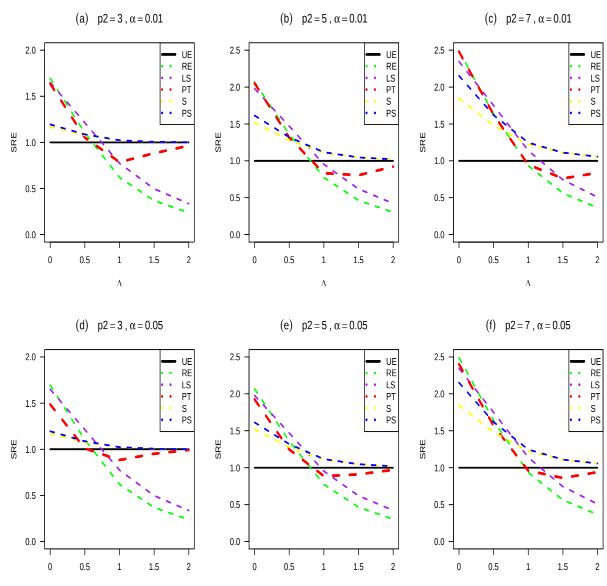
$LTPS$	LTS	$LTPT$		$LTLS$		LTR	Δ	$p_۲$	ρ
		α		λ					
		۰٫۰۵	۰٫۰۱	۰٫۸	۰٫۵				
۱٫۱۰۶	۱٫۰۹۷	۱٫۱۶۴	۱٫۲۹۸	۱٫۶۳۹	۱٫۴۳۴	۱٫۶۹۱	۰٫۰	۳	۰٫۸۰
۱٫۰۱۷	۱٫۰۱۶	۰٫۹۵۷	۰٫۹۱۷	۰٫۹۰۷	۱٫۰۸۶	۰٫۷۵۸	۰٫۵		
۱٫۰۰۱	۱٫۰۰۱	۰٫۹۹۹	۰٫۹۹۴	۰٫۴۳۸	۰٫۷۰۸	۰٫۳۲۰	۱٫۰		
۰٫۹۹۹	۰٫۹۹۹	۱٫۰۰۰	۱٫۰۰۰	۰٫۲۴۱	۰٫۴۶۰	۰٫۱۶۷	۱٫۵		
۰٫۹۹۹	۰٫۹۹۹	۱٫۰۰۰	۱٫۰۰۰	۰٫۱۴۹	۰٫۳۱۲	۰٫۱۰۰	۲٫۰		
۱٫۲۹۷	۱٫۲۸۷	۱٫۲۶۲	۱٫۴۸۲	۱٫۸۶۱	۱٫۵۶۳	۱٫۹۳۶	۰٫۰	۵	
۱٫۰۸۳	۱٫۰۸۱	۰٫۹۶۳	۰٫۹۳۸	۱٫۰۵۹	۱٫۲۱۲	۰٫۸۹۶	۰٫۵		
۱٫۰۱۷	۱٫۰۱۷	۰٫۹۹۸	۰٫۹۸۹	۰٫۵۲۷	۰٫۸۱۷	۰٫۳۸۹	۱٫۰		
۱٫۰۰۴	۱٫۰۰۴	۱٫۰۰۰	۱٫۰۰۰	۰٫۲۹۴	۰٫۵۴۴	۰٫۲۰۵	۱٫۵		
۱٫۰۰۰	۱٫۰۰۰	۱٫۰۰۰	۱٫۰۰۰	۰٫۱۸۳	۰٫۳۷۵	۰٫۱۲۴	۲٫۰		
۱٫۴۹۴	۱٫۴۷۹	۱٫۳۸۱	۱٫۶۶۹	۲٫۰۹۰	۱٫۶۸۳	۲٫۲۰۰	۰٫۰	۷	
۱٫۱۶۸	۱٫۱۶۶	۰٫۹۷۸	۰٫۹۶۹	۱٫۲۱۲	۱٫۳۲۰۰	۱٫۰۴۳	۰٫۵		
۱٫۰۴۲	۱٫۰۴۲	۰٫۹۹۵	۰٫۹۷۹	۰٫۶۱۹	۰٫۹۱۴	۰٫۴۶۴	۱٫۰		
۱٫۰۱۲	۱٫۰۱۲	۱٫۰۰۰	۰٫۹۹۹	۰٫۳۵۱	۰٫۶۲۳	۰٫۲۴۷	۱٫۵		
۱٫۰۰۲	۱٫۰۰۲	۱٫۰۰۰	۱٫۰۰۰	۰٫۲۲۱	۰٫۴۳۷	۰٫۱۵۱	۲٫۰		
۱٫۱۵۰	۱٫۱۲۸	۱٫۳۱۶	۱٫۵۰۳	۱٫۶۲۳	۱٫۴۲۷	۱٫۶۶۸	۰٫۰	۳	۰٫۹۰
۱٫۰۴۵	۱٫۰۴۳	۰٫۹۳۵	۰٫۹۲۱	۱٫۰۸۵	۱٫۱۸۴	۰٫۹۵۷	۰٫۵		
۱٫۰۰۶	۱٫۰۰۶	۰٫۹۵۲	۰٫۸۸۱	۰٫۶۲۵	۰٫۸۸۲	۰٫۴۸۳	۱٫۰		
۱٫۰۰۰	۱٫۰۰۰	۰٫۹۹۷	۰٫۹۸۹	۰٫۳۷۸	۰٫۶۳۸	۰٫۲۷۲	۱٫۵		
۰٫۹۹۸	۰٫۹۹۸	۱٫۰۰۰	۱٫۰۰۰	۰٫۲۴۶	۰٫۴۶۶	۰٫۱۷۱	۲٫۰		
۱٫۴۴۳	۱٫۴۰۵	۱٫۶۰۶	۱٫۸۳۰	۱٫۸۷۶	۱٫۵۷۴	۱٫۹۴۸	۰٫۰	۵	
۱٫۱۸۱	۱٫۱۶۹	۱٫۰۱۵	۱٫۰۵۶	۱٫۲۷۰	۱٫۳۲۴	۱٫۱۳۱	۰٫۵		
۱٫۰۵۲	۱٫۰۵۱	۰٫۹۳۳	۰٫۸۵۵	۰٫۷۴۴	۱٫۰۰۷	۰٫۵۸۰	۱٫۰		
۱٫۰۱۷	۱٫۰۱۷	۰٫۹۹۳	۰٫۹۶۸	۰٫۴۵۴	۰٫۷۴۲	۰٫۳۲۹	۱٫۵		
۱٫۰۰۵	۱٫۰۰۵	۱٫۰۰۰	۱٫۰۰۰	۰٫۲۹۸	۰٫۵۴۹	۰٫۲۰۷	۲٫۰		
۱٫۷۶۹	۱٫۶۸۷	۱٫۹۴۴	۲٫۱۶۵	۲٫۱۵۱	۱٫۷۱۵	۲٫۲۶۶	۰٫۰	۷	
۱٫۳۴۴	۱٫۳۲۲	۱٫۱۴۳	۱٫۲۴۰	۱٫۴۶۴	۱٫۴۵۱	۱٫۳۲۶	۰٫۵		
۱٫۱۱۲	۱٫۱۱۰	۰٫۹۰۹	۰٫۸۳۳	۰٫۸۷۱	۱٫۱۲۲	۰٫۶۸۹	۱٫۰		
۱٫۰۴۱	۱٫۰۴۱	۰٫۹۸۰	۰٫۹۲۹	۰٫۵۳۸	۰٫۸۴۰	۰٫۳۹۴	۱٫۵		
۱٫۰۱۵	۱٫۰۱۵	۰٫۹۹۸	۰٫۹۹۲	۰٫۳۵۵	۰٫۶۳۱	۰٫۲۵۰	۲٫۰		
۱٫۲۳۳	۱٫۱۸۵	۱٫۵۸۷	۱٫۷۰۰	۱٫۶۶۳	۱٫۴۵۳	۱٫۷۰۹	۰٫۰	۳	۰٫۹۵
۱٫۱۱۷	۱٫۱۰۰	۱٫۰۹۲	۱٫۱۴۹	۱٫۲۷۴	۱٫۲۸۸	۱٫۱۷۷	۰٫۵		
۱٫۰۳۶	۱٫۰۳۱	۰٫۸۴۰	۰٫۷۶۲	۰٫۸۵۸	۱٫۰۶۲	۰٫۷۰۳	۱٫۰		
۱٫۰۱۱	۱٫۰۱۱	۰٫۸۹۴	۰٫۷۷۱	۰٫۵۷۴	۰٫۸۴۵	۰٫۴۳۴	۱٫۵		
۱٫۰۰۴	۱٫۰۰۴	۰٫۹۶۹	۰٫۹۰۸	۰٫۳۹۷	۰٫۶۶۵	۰٫۲۸۷	۲٫۰		
۱٫۷۱۷	۱٫۵۶۳	۲٫۰۰۱	۲٫۰۴۲	۱٫۹۶۳	۱٫۶۲۲	۲٫۰۴۳	۰٫۰	۵	
۱٫۳۹۲	۱٫۳۱۹	۱٫۳۶۰	۱٫۴۰۲	۱٫۵۱۳	۱٫۴۵۱	۱٫۴۱۲	۰٫۵		
۱٫۱۵۳	۱٫۱۳۶	۰٫۸۹۲	۰٫۸۵۹	۱٫۰۲۷	۱٫۲۱۲	۰٫۸۴۸	۱٫۰		
۱٫۰۶۳	۱٫۰۶۱	۰٫۸۳۴	۰٫۷۱۳	۰٫۶۹۰	۰٫۹۷۷	۰٫۵۲۶	۱٫۵		
۱٫۰۲۹	۱٫۰۲۹	۰٫۹۲۰	۰٫۸۰۱	۰٫۴۸۰	۰٫۷۷۸	۰٫۳۴۷	۲٫۰		
۲٫۲۲۹	۱٫۷۱۶	۲٫۴۲۵	۲٫۴۳۲	۲٫۲۹۸	۱٫۷۸۸	۲٫۴۳۵	۰٫۰	۷	
۱٫۶۷۴	۱٫۴۳۱	۱٫۶۶۲	۱٫۶۸۲	۱٫۷۷۳	۱٫۶۰۵	۱٫۶۸۴	۰٫۵		
۱٫۲۵۹	۱٫۱۹۹	۱٫۰۰۵	۱٫۰۱۲	۱٫۲۱۰	۱٫۳۵۳	۱٫۰۱۴	۱٫۰		
۱٫۰۹۹	۱٫۰۹۰	۰٫۷۸۳	۰٫۶۹۹	۰٫۸۱۸	۱٫۱۰۲	۰٫۶۳۰	۱٫۵		
۱٫۰۴۳	۱٫۰۴۲	۰٫۸۲۸	۰٫۶۸۳	۰٫۵۷۱	۰٫۸۸۷	۰٫۴۱۷	۲٫۰		

www.SID.ir

Archive of SID



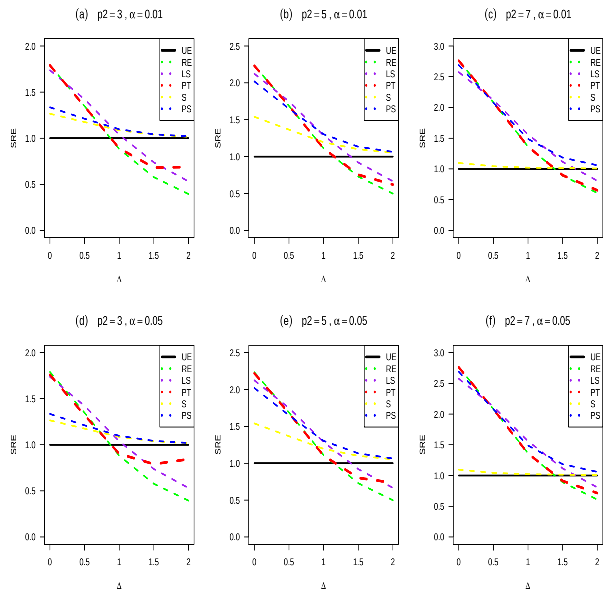
شکل ۱. کارایی نسبی برآوردگرها نسبت به برآوردگر نوع-لیو برای $n = ۳۵$، $\lambda = ۰/۸$ و $\rho = ۰/۸$



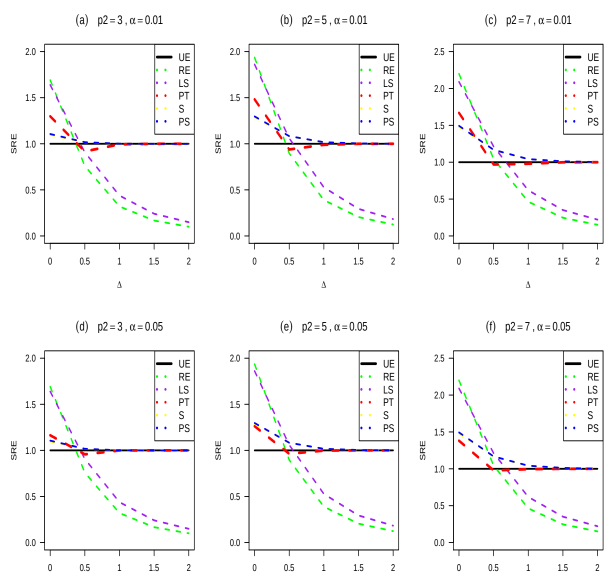
شکل ۲. کارایی نسبی برآوردگرها نسبت به برآوردگر نوع-لیو برای $n = ۳۵$، $\lambda = ۰/۸$ و $\rho = ۰/۹$

www.SID.ir

Archive of SID



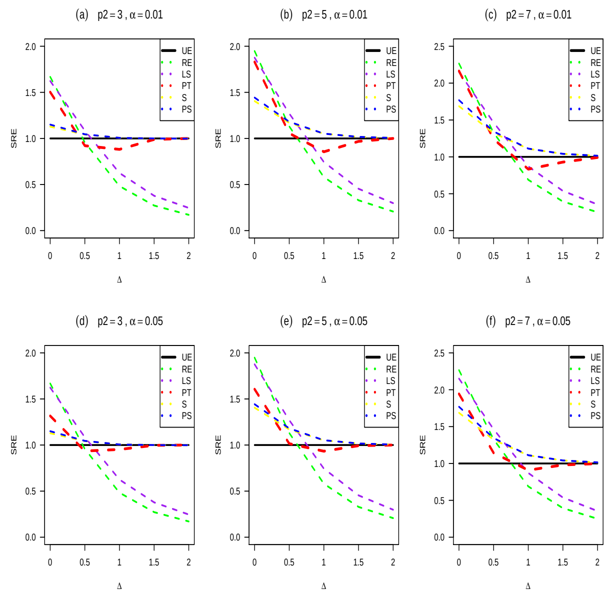
شکل ۳. کارایی نسبی برآوردگرها نسبت به برآوردگر نوع-لیو برای $n = ۳۵$، $\lambda = ۰/۸$ و $\rho = ۰/۹۵$



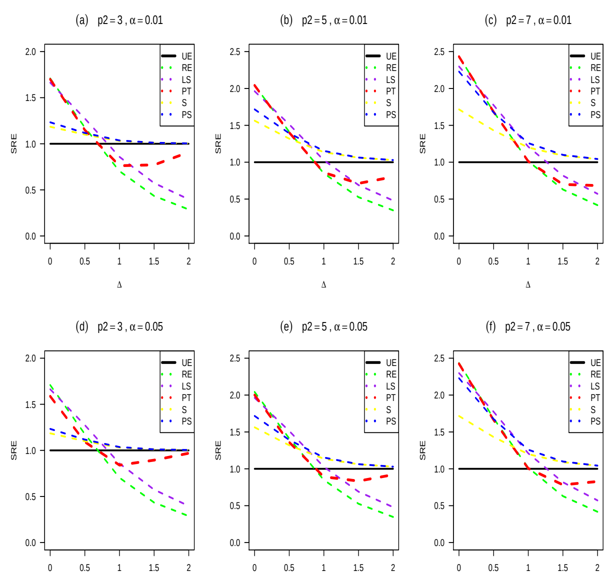
شکل ۴. کارایی نسبی برآوردگرها نسبت به برآوردگر نوع-لیو برای $n = ۵۰$، $\lambda = ۰/۸$ و $\rho = ۰/۸$

www.SID.ir

Archive of SID



شکل ۵. کارایی نسبی برآوردگرها نسبت به برآوردگر نوع-لیو برای $n =$ ۵۰، $\lambda =$ ۰/۸ و $\rho =$ ۰/۹



شکل ۶. کارایی نسبی برآوردگرها نسبت به برآوردگر نوع-لیو برای $n =$ ۵۰، $\lambda =$ ۰/۸ و $\rho =$ ۰/۹۵

www.SID.ir

Archive of SID

جدول ۳. معرفی متغیرهای پیشگو در مجموعه داده ایالت

متغیر	نام متغیر	توضیح
x_1	جمعیت	برآورد جمعیت تا جولای ۱۹۷۵
x_2	درآمد	درآمد سرانه ۱۹۷۴
x_3	بی‌سوادی	درصد جمعیت بی‌سواد
x_4	نرخ قتل	نرخ قتل عمد و قتل غیرعمد به ازای هر ۱۰۰۰۰۰ نفر جمعیت ۱۹۷۶
x_5	فارغ التحصیل	درصد فارغ التحصیلان دبیرستانی ۱۹۷۰
x_6	سرمازدگی	میانگین تعداد روزهای با حداقل دمای زیر صفر ۱۹۶۰ − ۱۹۳۱ در پایتخت یا شهرهای بزرگ
x_7	مساحت	مساحت زمین برحسب مایل مربع

همچنین معیارهای AIC و BIC اشاره کرد. در این مقاله از این دو معیار استفاده شده است که براساس آن‌ها، زیرمدل شامل متغیرهای پیشگوی x_4، x_5 و x_6 با کوچکترین مقدار $AIC = ۱۱۷/۹۷$ و $BIC = ۱۲۷/۵۳۴$ به‌دست آمده است. بنابراین بقیه متغیرها معنی‌دار نیستند و ضرایب مربوط به آن‌ها برابر صفر است و قید خطی $\mathbf{R}\boldsymbol{\beta} = \mathbf{r}$ به‌صورت

$$\mathbf{R} = \begin{bmatrix} ۱ & ۰ & ۰ & ۰ & ۰ & ۰ & ۰ \\ ۰ & ۱ & ۰ & ۰ & ۰ & ۰ & ۰ \\ ۰ & ۰ & ۱ & ۰ & ۰ & ۰ & ۰ \\ ۰ & ۰ & ۰ & ۰ & ۰ & ۰ & ۱ \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_7 \end{bmatrix}, \quad \mathbf{r} = \begin{bmatrix} ۰ \\ ۰ \\ ۰ \\ ۰ \end{bmatrix}$$

است. در نتیجه $\beta_1 = \beta_2 = \beta_3 = \beta_7 = ۰$ است. بنابراین $p_2 = ۴$ و تعداد متغیرهای پیشگوی معنی‌دار برابر با $p_1 = ۳$ هستند به‌طوری‌که تعداد کل متغیرها برابر با $p = p_1 + p_2 = ۷$ است. مقادیر λ و α در برآوردگرهای انقباضی خطی نوع-لیو و پیش‌آزمون نوع-لیو به‌ترتیب برابر با ۰/۵ و ۰/۰۱ در نظر گرفته شده‌اند. به منظور مقایسه عملکرد برآوردگرهای معرفی شده، نمونه $m = ۳۰$ تایی با ۱۰۰۰ بار باجایگذاری و به روش بوت استرپ انتخاب شده است. بردار درست ضرایب رگرسیونی با یک بار برازش مدل رگرسیون خطی چندگانه به‌صورت $\boldsymbol{\beta}^\circ = (۰, ۰, ۰, -۰/۲۸۳, ۰/۰۴۹, -۰/۰۰۶, ۰)'$ به‌دست آمده است. برآورد ضرایب رگرسیونی معنی‌دار به همراه میانگین توان‌های دوم خطا و کارایی نسبی برآوردگرها به ازای $\Delta = ۰$ در جدول ۵ داده شده است. کارایی نسبی شبیه‌سازی شده برآوردگرهای معرفی شده با نتایج شبیه‌سازی‌ها در جداول ۱ و ۲ موافق است.

بحث و نتیجه‌گیری

در این مقاله مدل رگرسیون خطی در نظر گرفته شد که بین متغیرهای پیشگو هم‌خطی چندگانه وجود داشت. همچنین فرض شد اطلاعاتی غیر از نمونه تصادفی به‌صورت قید خطی روی پارامترها در دسترس است. تحت این دو فرض، برآوردگرهای انقباضی نوع-لیو برای برآورد ضرایب رگرسیونی مدل به منظور بهبود عملکرد برآورد این ضرایب معرفی

www.SID.ir

Archive of SID

جدول ۴. ماتریس ضریب همبستگی بین متغیرهای پیشگو در مجموعه داده واقعی

متغیر	$x_۱$	$x_۲$	$x_۳$	$x_۴$	$x_۵$	$x_۶$	$x_۷$
$x_۱$	۱٫۰۰۰						
$x_۲$	۰٫۴۶۰	۱٫۰۰۰					
$x_۳$	۰٫۰۱۵	۰٫۰۰۰	۱٫۰۰۰				
$x_۴$	۰٫۰۱۸	۰٫۰۰۰	۰٫۰۰۰	۱٫۰۰۰			
$x_۵$	۰٫۸۸۰	۰٫۵۹۰	۰٫۱۱۰	۰٫۶۸۰	۱٫۰۰۰		
$x_۶$	۰٫۱۵۰	۰٫۰۰۱	۰٫۱۱۰	۰٫۱۱۰	۰٫۰۰۹	۱٫۰۰۰	
$x_۷$	۰٫۵۰۰	۰٫۰۰۰	۰٫۰۰۰	۰٫۰۰۸	۰٫۰۱۸	۰٫۰۰۰	۱٫۰۰۰

جدول ۵. برآورد ضرایب رگرسیونی متغیرهای پیشگوی معنی‌دار، میانگین توان‌های دوم خطا و کارایی نسبی شبیه‌سازی شده برآوردگرهای معرفی شده نسبت به برآوردگر (نامقید) نوع-لیو در مجموعه داده واقعی

برآوردگرها	$\hat{\beta}_۴$	$\hat{\beta}_۵$	$\hat{\beta}_۶$	MSE	SRE
نوع-لیو	۰٫۰۲۶	۰٫۰۹۷	۰٫۰۰۹	۲٫۹۹۲	۱٫۰۰۰
مقید نوع-لیو	۰٫۲۲۷	۰٫۱۵۵	۰٫۰۰۵	۰٫۲۷۹	۱۰٫۷۲
انقباضی نوع-لیو	۰٫۱۲۶	۰٫۱۲۶	۰٫۰۰۷	۰٫۹۰۰	۳٫۳۲۲
پیش‌آزمون نوع-لیو	۰٫۱۷۳	۰٫۱۴۳	۰٫۰۰۶	۱٫۲۶۸	۲٫۳۵۹
استاین نوع-لیو	۰٫۰۹۳	۰٫۱۲۸	۰٫۰۰۹	۲٫۰۳۸	۱٫۴۶۸
استاین مثبت نوع-لیو	۰٫۰۸۹	۰٫۱۱۹	۰٫۰۰۸	۱٫۸۸۷	۱٫۵۸۵

شد. کارایی‌های نسبی شبیه‌سازی شده برآوردگرها نسبت به برآوردگر (نامقید) نوع-لیو در قسمت‌های مختلف فضای پارامتر با استفاده از شبیه‌سازی مونت کارلو به‌ازای تعداد متفاوت از متغیرهای پیشگو، $p = ۸, ۱۰, ۱۲$، میزان همبستگی‌های متفاوت بین متغیرهای پیشگو، یعنی $\rho = ۰٫۸, ۰٫۹, ۰٫۹۵$ و برای حجم نمونه $n = ۳۵, ۵۰$ مقایسه شد. به‌علاوه، از یک مجموعه داده واقعی برای مقایسه عملکرد برآوردگرهای معرفی شده و همچنین برآورد ضرایب معنی‌دار استفاده شد. نتایج به‌دست آمده از شبیه‌سازی‌ها و مثال واقعی برتری برآوردگرهای معرفی شده را نسبت به برآوردگر (نامقید) نوع-لیو نشان می‌دهند.

تقدیر و تشکر

نویسندگان مقاله از سردبیر، داوران و ویراستار ارجمند که با پیشنهادات و نظرات ارزنده خود موجب بهبود این مقاله شدند، تقدیر و تشکر می‌نمایند.

مراجع

روزبه، م. و امینی، م. (۱۳۹۸)، برآوردگر استوار مرزبندی شده تعمیم‌یافته محتمل در مدل رگرسیون نیمه‌پارامتری، مجله علوم آماری، ۱۳(۲)، ۴۶۰-۴۴۱.

www.SID.ir

Archive of SID

روزبه، م. و معنوی، م. (۱۳۹۹)، مدل‌سازی سن تقویمی به روش رگرسیون ستیغی کمترین توان‌های دوم پیراسته، مجله علوم آماری، ۱۴(۲)، ۴۲۸-۴۰۹.

Ahmed, S. E. (2014), *Penalty, Shrinkage and Pretest Strategies – Variable Selection and Estimation*, Heidelberg Springer.

Aitchison, J. and Silvey, S. D. (1958), Maximum-Likelihood Estimation of Parameters Subject to Restraints, *The Annals of Mathematical Statistics*, **29**(3), 813-828.

Amini, M. and Roozbeh, M. (2015), Optimal Partial Ridge Estimation in Restricted Semiparametric Regression Models, *Journal of Multivariate Analysis*, **136**, 26-40.

Roozbeh, M. (2018), Optimal QR-Based Estimation in Partially Linear Regression Models with Correlated Errors Using GCV Criterion, *Computational Statistics & Data Analysis*, **117**, 45-61.

Bancroft, T. A. (1944), On Biases in Estimation due to the Use of Preliminary Tests of Significance, *The Annals of Mathematical Statistics*, **15** (2), 190-204.

Asar, Y. (2018), Liu-Type Negative Binomial Regression A Comparison of Recent Estimators and Applications, *In Trends and Perspectives in Linear Statistical Inference*, Springer, Cham, 23-39.

Asar, Y., Ahmed, S. E. and Yuzbasi, B. (2018). Efficient and Improved Estimation Strategy in Zero-Inflated Poisson Regression Models, *In International Conference on Management Science and Engineering Management*, Springer, Cham, 329-342.

Hoerl, A. E. and Kennard, R. W. (1970), Ridge Regression Biased Estimation for Non-Orthogonal Problems, *Technometrics*, **12**(1), 55-67.

Hoerl, A. E, Kennard, R. W. and Baldwin, K. F. (1975), Ridge Regression Some Simulation, *Communications in Statistics-Theory and Methods*, **4**, 105-123.

Lisawadi, S., Ahmed, S. E. and Reangsephet, O. (2020), Post Estimation and Prediction Strategies in Negative Binomial Regression Model, *International Journal of Modelling and Simulation*, https//doi.org/10.1080/02286203.2020.1792601.

www.SID.ir

Archive of SID

Liu, K. (2003), Using Liu-Type Estimator to Combat Collinearity, *Communications in Statistics-Theory and Methods*, **32** (5), 1009-1020.

Saleh, A. M. E. (2006), *Theory of Preliminary Test and Stein-Type Estimation with Applications*, **517**, John Wiley, New York.

Stein, C. (1956), The Admissibility of Hotelling's T^2-Test, *Mathematical Statistics*, **27**, 616-623.

Thompson, J. R. (1968), Some Shrinkage Techniques for Estimating the Mean, *Journal of the American Statistical Association*, **63** (321), 113-122.

Yuzbasi, B., Arashi, M. and Ahmed, S. E. (2020), Shrinkage Estimation Strategies in Generalized Ridge Regression Models Low/High-Dimension Regime, *International Statistical Review*, **88** (1), 229-251.

Zandi, Z., Bevrani, H. and Arabi Belaghi, R. (2021), Improved Shrinkage Estimators in Zero-Inflated Negative Binomial Regression Model, *Hacettepe Journal of Mathematics and Statistics*, **50**(6), 1855-1876. https//doi.org/10.15672/hujms.911424.

Zandi, Z., Bevrani, H. and Arabi Belaghi, R. (2021), Using Shrinkage Strategies to Estimate Fixed Effects in Zero-Inflated Negative Binomial Mixed Model, *Communications in Statistics-Simulation and Computation*, https//doi.org/10.1080/03610918.2021.1928704.

www.SID.ir