



Variable Selection in Small Area Estimation under AR-GARCH Models Based on the Gradient Boosting Algorithm

Bashiri Goudarzi, O.¹, Sayyareh, A.², Zamani Mehreyan, S.³

¹Department of Pure Mathematics, K.N. Toosi University of Technology, Tehran, Iran.

²Department of Computer Sciences and Statistics, K.N. Toosi University of Technology, Tehran, Iran.

³Department of Statistics, Imam Khomeini International University, Qazvin, Iran.

Corresponding author: A. Sayyareh, asayyareh@kntu.ac.ir Received: 20/4/2024

Revised: 12/5/2025 Accepted and Published Online: 13/5/2025.

Introduction

The boosting algorithm is a method to transform weak learning systems into strong systems based on the combination of different results. In this paper, mixture models with random effects are considered for small areas where the errors follow the *AR-GARCH* model and The machine learning algorithms, such as the boosting algorithm, have been proposed for variable selection. Using simulated and tax liability data, the boosting algorithm's performance is studied and compared with classical variable selection methods, such as the step-by-step method.

Material and Methods

One of the problems of small area estimation is producing reliable estimates of the features of interest. It is assumed that the observed and auxiliary values follow the linear regression model and that the sampling errors depend on and follow the autoregressive model. However, in practice, there are many situations in which the sampling errors are dependent and follow the AR-GARCH model. The classical methods for variable selection, such as the step-by-step method, sequentially delete or add explanatory variables using some criterion. A problem with the step-by-step method is that some

real explanatory variables with causal effects on the dependent variable may not be statistically significant. In contrast, nuisance variables may be coincidentally significant. Using some risk functions, the model-based boosting algorithm is a tool to fit a statistical model while performing variable selection simultaneously. For the tax liability dataset, we proposed the mixed model with random effects for small areas where sampling errors are dependent. We followed the AR-GARCH model and selected variables based on the boosting algorithm and step-by-step method.

Results and Discussion

For the tax liability dataset, the estimated intra-cluster correlation coefficient is 0.228. Although the fitted model based on the step-by-step method has a lower *MSE*, the fitted model based on the boosting algorithm has a lower *AIC*. Since the p-value of the t-test, 0.9998, is greater than 0.05, the mean of the series is not significantly different from zero. The p-value of the Ljung-Box test for the time series is 0.0027. It shows a correlation between the series values, which can be described by classical models such as the autoregressive model. Also, the p-value of the Ljung-Box test for the absolute time series, 0.0493, shows a correlation between the absolute time series, which the volatility model can describe.

Conclusion

In this work, machine learning algorithms, such as the boosting algorithm, estimate and select variables in mixed models with random effects for small areas. The performance of the boosting algorithm and the step-by-step method in variable selection were studied using simulated and tax liability data. The results show that the boosting algorithm performs more appropriately in variable selection in small areas. In some areas, the step-by-step method cannot detect the true model.

Keywords: Boosting algorithm, Variable selection, Small area estimation, *AR – GARCH* model.

Mathematics Subject Classification (2010): 62F03, 68W20.



انتخاب متغیر در کوچک نواحی تحت مدل‌های AR-GARCH با الگوریتم تقویت گرایان

ام‌البنین بشیری گودرزی^۱، عبدالرضا سیاره^۲، صدیقه زمانی مهریان^۳

^۱گروه ریاضی محض، دانشگاه صنعتی خواجه نصیرالدین طوسی

^۲گروه علوم کامپیوتر و آمار، دانشگاه صنعتی خواجه نصیرالدین طوسی

^۳گروه آمار، دانشگاه بین‌المللی امام خمینی (ره)

نویسنده مسئول: عبدالرضا سیاره، asayyareh@kntu.ac.ir

تاریخ دریافت: ۱۴۰۳/۱/۳۱ تاریخ بازنگری: ۱۴۰۴/۲/۲۲ تاریخ پذیرش و انتشار: ۱۴۰۴/۲/۲۳

چکیده: الگوریتم تقویت، الگوریتمی ترکیبی برای کاهش عدم توازن و واریانس از خانواده الگوریتم‌های یادگیری ماشین در حوزه یادگیری با نظارت است. این الگوریتم، روشی برای تبدیل سیستم‌های یادگیری ضعیف به سیستم قوی بر اساس ترکیب نتایج مختلف است. پس از انتخاب متغیرها و ساخت مدل اولیه، با تنظیم نرخ یادگیری و سایر پارامترهای الگوریتم تقویت، مدل‌های ضعیف به مدل قوی‌تری برای برازش به داده‌ها تبدیل می‌شود. در این مقاله مدل‌های آمیخته با اثرات تصادفی برای کوچک نواحی در نظر گرفته شده که در آن خطاها از مدل $AR - GARCH$ پیروی می‌کنند. به منظور انتخاب متغیر در این مدل‌ها برای کوچک نواحی، الگوریتم تقویت پیشنهاد شده است. با داده‌های شبیه‌سازی شده و داده‌های مالیاتی، عملکرد الگوریتم تقویت در انتخاب متغیر با روش‌های کلاسیک انتخاب متغیر مورد مطالعه قرار گرفته است. نتایج نشان می‌دهند الگوریتم تقویت عملکرد بهتری در انتخاب متغیر برای کوچک نواحی دارد.

واژه‌های کلیدی: الگوریتم تقویت، انتخاب متغیر، برآورد کوچک ناحیه‌ای، مدل $AR - GARCH$.

کد موضوع بندی ریاضی (۲۰۱۰): 68W20، 62F03.



©نویسندگان). ناشر انجمن آمار ایران است.

این مقاله با دسترسی آزاد تحت شرایط و ضوابط (CC BY-NC 4.0) توزیع شده است.

۱ مقدمه

در برخی از مطالعات، متغیر مورد توجه در نواحی و بخش‌های مرتبط با جمعیت بررسی می‌شود. برآوردهای مبتنی بر داده‌های نمونه‌ای ناحیه‌ویژه را برآورد مستقیم می‌نامند. اگر اندازه نمونه ناحیه، به قدر کافی بزرگ باشد که برآورد مستقیم با دقت کافی به دست آید، ناحیه بزرگ در نظر گرفته می‌شود. هم‌چنین وقتی ناحیه‌ای کوچک تلقی می‌شود که اندازه نمونه ناحیه‌ویژه به قدر کافی بزرگ نباشد که برآوردهای مستقیم با دقت کافی به دست آیند. معمولاً اندازه‌های نمونه در ناحیه‌های مورد علاقه، کوچک یا حتی صفر هستند. این امر باعث می‌شود که برآوردهای مستقیم دارای واریانس زیادی هستند. هنگامی که مشاهده‌ای در برخی از کوچک نواحی وجود ندارد، برآوردهای مستقیم حتی نمی‌توانند محاسبه شوند.

دو ایده اصلی برای توسعه مدل‌هایی در برآورد کوچک نواحی استفاده شده است که منجر به توسعه مدل‌هایی با اثر ثابت یا مدل‌های آمیخته با اثرات تصادفی می‌شود. مدل‌های اثر ثابت، تغییرات متغیر پاسخ را به طور کامل با استفاده از تغییرات عوامل شناخته شده توضیح می‌دهند. برآورد مشخصه‌های کوچک نواحی بر اساس مدل‌های اثر ثابت را برآوردهای مصنوعی می‌نامند، **لوی و فرنچ (۱۹۷۷)**. در دهه‌های اخیر مدل‌های آمیخته خطی به دلیل توانایی این مدل‌ها در پیش‌بینی ترکیب خطی اثرات ثابت و تصادفی به‌طور گسترده مورد استفاده قرار گرفته است. **فای و هریوت (۱۹۷۹)** و **پففرمن و برنارد (۱۹۹۱)** مدل‌های آمیخته را به‌منظور بهبود برآورد مشخصه‌های کوچک نواحی بر اساس نمونه‌گیری پیمایشی یا داده‌های سرشماری استفاده کردند. **رائو و یو (۱۹۹۲، ۱۹۹۴)** تعمیمی از مدل فای-هریوت را برای سری‌های زمانی و داده‌های مقطعی پیشنهاد دادند. **سینگ و همکاران (۱۹۹۴)** و **قوش و همکاران (۱۹۹۶)** از مدل‌های سری زمانی برای بهبود برآورد در کوچک نواحی استفاده کردند. **یو و رائو (۲۰۰۰)** روش‌های بیز سلسله مراتبی را برای برآورد میانگین کوچک نواحی تحت مدل‌های آمیخته با اثر تصادفی پیشنهاد دادند. **زارعی (۱۴۰۰)** توزیع α -پایدار متقارن را برای مؤلفه‌های خطاهای نمونه‌گیری پیشنهاد داد و با استفاده از رهیافت بیز تجربی پارامترهای کوچک نواحی را برآورد کرد.

انتخاب مدل به‌خصوص مدل‌سازی مدل رگرسیون یکی از مسائل مهم در بحث کوچک ناحیه‌ای است. **وایدا و بلانچارد (۲۰۰۵)** معیار اطلاع آکائیک شرطی را به‌عنوان یک برآوردگر پیشنهاد کردند که به‌طور مجانبی ناریب است و بیان کردند که معیار اطلاع آکائیک ممکن است برای پیش‌بینی مؤلفه اثرات تصادفی نامناسب باشد. **جیانگ و همکاران (۲۰۱۸)** نشان دادند که برآوردهای به‌دست آمده از روش مونت کارلو در کوچک نواحی برآوردهای نارایی هستند. **(سوکاساوا و همکاران، ۲۰۱۹)** مدل‌های فای-هریوت آشیانی را در نظر گرفتند و به‌طور همزمان به انتخاب متغیر و برآورد پارامترها بر اساس کمینه کردن میانگین مربعات خطاهای پیش‌بینی پرداختند. **سوکاساوا و کوبوکاوا (۲۰۲۰)** به بررسی روش‌های برآورد پارامتر در مدل‌های پیشنهاد شده برای کوچک نواحی پرداختند و روش‌های مختلفی را برای محاسبه میانگین مربع خطاها و فاصله اطمینان ارائه دادند.

شاپیره (۱۹۹۰) روش تقویت را به‌عنوان یک روش عمومی برای ترکیب چند طبقه‌بندی کننده برای بهبود دقت طبقه‌بندی کلی برای تقریباً هر نوع الگوریتم یادگیری معرفی کرد. الگوریتم تقویت به نوعی از الگوریتم‌های

یادگیری تحت نظارت اشاره دارد که به دلیل دقت پیش‌بینی بالا از محبوبیت زیادی در علوم داده برخوردار هستند (فروند و شاپیره، ۱۹۹۷؛ چن و گسترین، ۲۰۱۶). فریدمن (۲۰۰۱) الگوریتم تقویت گرادیان را به صورت ترکیب دو الگوریتم تقویت تطبیقی و تقویت L^2 معرفی کرد. (بوهلمن و هاتورن، ۲۰۰۷) الگوریتم تقویت را برای برآورد و انتخاب متغیر در مدل‌های رگرسیون پیشنهاد کردند. الگوریتم تقویت برای مدل‌های سری زمانی غیرخطی توسط رابینز و نونوف و همکاران (۲۰۱۲) مورد مطالعه قرار گرفته است. آدرینو و بوهلمن (۲۰۱۶) نوسانات سری‌های زمانی مالی با ابعاد بالا را از طریق الگوریتم گرادیان تابعی برآورد کردند. چن و کیو (۲۰۲۳) از الگوریتم تقویت برای انتخاب متغیر و برآورد پارامتر در داده‌های طولی و سانسور شده استفاده کردند. همچنین، چن (۲۰۲۴) از الگوریتم تقویت با روش برآورد اسپلاین مکعبی برای بازیابی مجدد رابطه غیرخطی بین متغیرهای کمکی و زمان بقا استفاده کرد. در این مقاله تعمیمی از مدل رائو و یو (۱۹۹۴) در نظر گرفته شده و به انتخاب متغیر با استفاده از روش تقویت پرداخته شده است. مدل‌های آمیخته با اثرات تصادفی برای کوچک نواحی مطالعه شده که در آن خطاها از مدل $AR - GARCH$ پیروی می‌کنند. به منظور انتخاب متغیر در این نوع از مدل‌ها برای کوچک نواحی، الگوریتم‌های یادگیری ماشین از جمله الگوریتم تقویت پیشنهاد و مورد بررسی قرار گرفته است.

در بخش ۲، تعمیمی از مدل رائو و یو (۱۹۹۴) در نظر گرفته شده و برآوردهای بیشینه درست‌نمایی پارامترهای مدل محاسبه شده است. در بخش ۳ الگوریتم تقویت گرادیان جهت انتخاب متغیر معرفی شده است. در بخش ۴ عملکرد دو روش تقویت و گام به گام در انتخاب متغیر در مدل پیشنهاد شده در کوچک نواحی با استفاده از مطالعات شبیه‌سازی مقایسه شده است.

۲ معرفی مدل و برآورد پارامترهای مدل

یکی از پرکاربردترین مدل‌های در سطح ناحیه برای برآورد کوچک ناحیه‌ای توسط فای و هریوت (۱۹۷۹) ارائه شده است. با استفاده از مدل فای-هریوت، فرض می‌شود که پارامتر کوچک ناحیه‌ای θ_i از طریق مدل خطی به داده‌های کمکی ناحیه خاص مرتبط است. به عبارت دیگر

$$\theta_i = x_i\beta + \nu_i, \quad i = 1, \dots, M$$

که در آن M نشان دهنده تعداد ناحیه‌ها است. در حقیقت در کوچک نواحی $\theta_i = \hat{\theta}_i + e_i$ برآوردگری با واریانس بزرگ است، زیرا حجم نمونه برای محاسبه $\hat{\theta}_i$ در عمل می‌تواند کوچک باشد. فای و هریوت (۱۹۷۹) با ترکیب دو معادله فوق مدل فای-هریوت

$$\theta_i = x_i\beta + \nu_i + e_i, \quad i = 1, \dots, M$$

۶ انتخاب متغیر در کوچک نواحی تحت مدل‌های AR-GARCH

را پیشنهاد دادند، که در آن x_i بردار $(1 \times k)$ بعدی از متغیرهای کمکی، β بردار $(k \times 1)$ بعدی از پارامترها، $e_i \sim N(0, \sigma_e^2)$ و اثرات تصادفی ν_i از توزیع نرمال $N(0, \sigma_\nu^2)$ پیروی می‌کند. اثرات تصادفی تغییرات بین ناحیه‌ها را توضیح می‌دهند که نمی‌توان آن‌ها را با اضافه کردن متغیرهای کمکی توضیح داد. [رائو و یو \(۱۹۹۴\)](#)، [\(۱۹۹۲\)](#) تعمیمی از مدل فای-هریوت را به صورت

$$\begin{aligned} y_{it} &= \mu_{it} + e_{it}, & i &= 1, \dots, M, & t &= 1, \dots, T \\ \mu_{it} &= x_{it}\beta + \nu_i + u_{it} \end{aligned} \quad (۱)$$

برای سری‌های زمانی و داده‌های مقطعی پیشنهاد دادند، که در آن u_{it} ها متغیرهای تصادفی مستقلی هستند که از توزیع نرمال پیروی می‌کنند. می‌توان فرض کرد که u_{it} از مدل خودهمبسته مرتبه اول $u_{it} = \rho u_{i,t-1} + \epsilon_{it}$ پیروی می‌کند، که در آن ϵ_{it} ها مستقل و هم‌توزیع با توزیع نرمال $N(0, \sigma_\epsilon^2)$ هستند. در این مقاله تعمیمی از مدل [رائو و یو \(۱۹۹۴، ۱۹۹۲\)](#) در نظر گرفته شده و پارامترهای مدل با استفاده از روش بیشینه درست‌نمایی برآورد شده است. فرض کنید داده‌های مشاهده شده برای ناحیه i ، از مدل

$$\begin{aligned} y_{it} &= \mu_{it} + e_{it}; & i &= 1, \dots, M, & t &= 1, \dots, T \\ \mu_{it} &= x_{it}\beta + \nu_i + u_{it} \end{aligned} \quad (۲)$$

پیروی می‌کنند، که در آن $\nu_1, \dots, \nu_M$ ، اثرات تصادفی مستقل و هم‌توزیع $N(0, \sigma_\nu^2)$ و از e_{it} ها مستقل هستند. همچنین u_{it} ها از مدل $AR - GARCH$ مرتبه اول $u_{it} = \rho u_{it-1} + \sigma_{it|t-1} \epsilon_{it}$ پیروی می‌کنند. مدل (۱) را می‌توان به صورت

$$\begin{aligned} y_{it} &= x_{it}\beta + \nu_i + u_{it} + e_{it}, & i &= 1, \dots, M, & t &= 1, \dots, T \\ u_{it} &= \rho u_{it-1} + \sigma_{it|t-1} \epsilon_{it} \end{aligned}$$

بازنویسی کرد. با جایگذاری $u_{it} = y_{it} - x_{it}\beta - \nu_i - e_{it}$ در $u_{it} = \rho u_{it-1} + \sigma_{it|t-1} \epsilon_{it}$ داریم:

$$y_{it} - \beta x_{it} - \nu_i - e_{it} = \rho u_{it-1} + \sigma_{it|t-1} \epsilon_{it}.$$

ام‌البین بشیری گودرزی و همکاران ۷

بنابراین برای $M, i = 1, \dots, T$ و $t = 1, \dots, T$ می‌توان مدل (۱) را به صورت

$$\begin{aligned} y_{it} &= \rho u_{it-1} + \sigma_{t|t-1} \epsilon_{it} + e_{it} + \beta x_{it} + \nu_i \\ &= \rho y_{it-1} + (x_{it} - \rho x_{it-1})\beta + \nu_i(1 - \rho) + e_{it} - \rho e_{it-1} + \sigma_{t|t-1} \epsilon_{it} \end{aligned}$$

یا به طور معادل به فرم ماتریسی

$$\mathbf{y}_t = \rho \mathbf{y}_{t-1} + (\mathbf{x}_t - \rho \mathbf{x}_{t-1})\beta + (1 - \rho)I_M \nu + \mathbf{e}_t - \rho \mathbf{e}_{t-1} + \sigma_{t|t-1} \epsilon_t \quad (3)$$

بازنویسی کرد، که در آن یک ماتریس $M \times k$ ، I_M یک ماتریس $M \times M$ ، β یک بردار $k \times 1$ ، $\sigma_{t|t-1}$ یک ماتریس $M \times M$ است و

$$\mathbf{y}_t = \begin{pmatrix} y_{1t} \\ \vdots \\ y_{Mt} \end{pmatrix}, \nu = \begin{pmatrix} \nu_1 \\ \vdots \\ \nu_M \end{pmatrix}, \mathbf{e}_t = \begin{pmatrix} e_{1t} \\ \vdots \\ e_{Mt} \end{pmatrix}, \epsilon = \begin{pmatrix} \epsilon_{1t} \\ \vdots \\ \epsilon_{Mt} \end{pmatrix}.$$

به علاوه $\nu \sim N_M(\circ, G)$ ، که در آن $G = \sigma_\nu^2 I_M$ ، $\delta \sim N_M(\circ, F)$ ، $\delta = (1 - \rho)\nu$ و $\sigma_\nu^2 = (1 - \rho)G(1 - \rho)'$ ، $F = (1 - \rho)G(1 - \rho)'$ ، $e_t \sim N_M(\circ, \Sigma)$ ، که $\Sigma = \text{diag}\{\sigma_{e_1}^2, \dots, \sigma_{e_M}^2\}$ معلوم در نظر گرفته می‌شود و همچنین $\eta_t = e_t - \rho e_{t-1} \sim N_M(\circ, S)$ ، $S = \Sigma + \rho \Sigma \rho'$.

شناساپذیری یکی از ویژگی‌های لازم برای کفایت یک مدل آماری است. اگر مدل شناساپذیر نباشد، نمی‌توان پارامترهای مدل را تعیین کرد. مدل (۲) را به ازای همه i ها، مدل توأم و به ازای یک i ثابت، مدل تکی می‌نامند. قضایای حدی زیادی در رابطه با شناساپذیری مدل‌های آمیخته با اثرهای تصادفی بیان شده است، از جمله می‌توان به قضایای زیر اشاره کرد که اثبات این قضایا در وانگ (۲۰۱۳) موجود است.

قضیه ۱. مدل توأم (۲) شناساپذیر است اگر و تنها اگر حداقل یکی از مدل‌های تکی آن شناساپذیر باشد.

قضیه ۲. پر رتبه بودن ماتریس طرح X و شناساپذیری ماتریس کواریانس Σ_u در مدل تکی شرط لازم برای شناساپذیری بردار پارامترها است.

قضیه ۳. فرض کنید $1 \leq q < T$ ، مدل (۲) شناساپذیر است اگر برای $\theta_\epsilon^* \neq \theta_\epsilon$ $\text{rank}(\Sigma_\epsilon - \Sigma_\epsilon^*) > q$ به طوری که θ_ϵ و θ_ϵ^* هر دو عضو Θ_ϵ هستند.

قضیه ۴. فرض کنید $1 < T$ ، مدل (۲) شناساپذیر است اگر و تنها اگر $\Sigma = \text{diag}\{\sigma_{e_1}^2, \dots, \sigma_{e_M}^2\}$ ماتریس کواریانس معلوم باشد.

۸ انتخاب متغیر در کوچک نواحی تحت مدل‌های AR-GARCH

بنابراین با استفاده از قضیه ۴ با معلوم در نظر گرفتن ماتریس کوواریانس Σ شناساپذیر بودن مدل (۲) نتیجه می‌شود. در ادامه روند محاسبات برآورد پارامترهای مدل با استفاده از روش بیشینه درست‌نمایی ارائه خواهد شد.

لگاریتم تابع درست‌نمایی به صورت $\ell(\theta; y_t) = C - \frac{1}{\varphi} \log |w_t| - \frac{1}{\varphi} z_t' w_t^{-1} z_t$ به دست می‌آید، که در آن $w_t = F + S + \text{diag}\{\sigma_{1,t|t-1}^2, \dots, \sigma_{M,t|t-1}^2\}$ و $z_t = y_t - \rho y_{t-1} + (x_t - \rho x_{t-1})\beta$ تابع ℓ نسبت به پارامترهای مدل مشتق گرفته و با حل معادلات، برآوردهای بیشینه درست‌نمایی حاصل می‌شوند. برآوردگر بیشینه درست‌نمایی پارامتر β شکل بسته‌ای دارد و به صورت

$$\hat{\beta} = (x_t - \rho x_{t-1})' w_t^{-1} (y_t - \rho y_{t-1}) ((x_t - \rho x_{t-1})' w_t^{-1} (x_t - \rho x_{t-1}))^{-1}$$

قابل محاسبه است. اگر $\varphi = (\omega, \alpha, \gamma)$ ، با توجه به اینکه $\sigma_{t|t-1}^2 = \omega + \alpha y_{t-1}^2 + \gamma \sigma_{t-1|t-2}^2$ می‌توان نوشت

$$\frac{\partial \ell}{\partial \varphi} = \frac{\partial \ell}{\partial w_t} \left(\frac{\partial w_t}{\partial \varphi} \right). \quad (۴)$$

$$\frac{\partial \ell}{\partial w_t} = \frac{\partial \ell}{\partial w_t^{-1}} \left(\frac{\partial w_t^{-1}}{\partial w_t} \right) \quad (۵)$$

که در آن

$$\frac{\partial w_t^{-1}}{\partial w_t} = \left((w_t^{-1})' \otimes w_t^{-1} \right) \quad (۶)$$

$$\frac{\partial \ell}{\partial w_t^{-1}} = \frac{1}{\varphi} \sum_{t=2}^T \text{vec}'(w_t') - \frac{1}{\varphi} \sum_{t=2}^T z_t' \otimes z_t'. \quad (۷)$$

با جایگذاری معادلات (۶) و (۷) در معادله (۵) داریم

$$\begin{aligned} \frac{\partial \ell}{\partial w_t} &= \frac{\partial \ell}{\partial w_t^{-1}} \left(\frac{\partial w_t^{-1}}{\partial w_t} \right) \\ &= \frac{1}{\varphi} \sum_{t=2}^T \text{vec}'(w_t') \left((w_t^{-1})' \otimes w_t^{-1} \right) - \frac{1}{\varphi} \sum_{t=2}^T (z_t' \otimes z_t') \left((w_t^{-1})' \otimes w_t^{-1} \right). \end{aligned}$$

بنابراین برآوردهای پارامترهای مدل $GARCH$ با حل معادله

$$\begin{aligned} \circ &= \frac{\partial \ell}{\partial \varphi} = \frac{\partial \ell}{\partial w_t} \frac{\partial w_t}{\partial \varphi} \\ &= \frac{1}{\varphi} \sum_{t=2}^T \text{vec}'(w_t') \left((w_t^{-1})' \otimes w_t^{-1} \right) \frac{\partial w_t}{\partial \varphi} - \frac{1}{\varphi} \sum_{t=2}^T (z_t' \otimes z_t') \left((w_t^{-1})' \otimes w_t^{-1} \right) \frac{\partial w_t}{\partial \varphi} \\ &= \frac{1}{\varphi} \sum_{t=2}^T w_t^{-1} \left(\frac{\partial w_t}{\partial \varphi} \right) - \frac{1}{\varphi} \sum_{t=2}^T (w_t^{-1} z_t)' (w_t^{-1} z_t) \frac{\partial w_t}{\partial \varphi} \end{aligned} \quad (8)$$

حاصل می‌شوند. همچنین توجه کنید که

$$\frac{\partial w_t}{\partial \varphi} = \frac{\partial (F + S + \Sigma_{t|t-1})}{\partial \varphi} = \frac{\partial \Sigma_{t|t-1}}{\partial \varphi}$$

که در آن $\sigma_{it|t-1}^2 = \omega + \alpha y_{it-1}^2 + \gamma \sigma_{it-1|t-2}^2$ و $\Sigma_{t|t-1} = \text{diag}\{\sigma_{1,t|t-1}^2, \dots, \sigma_{M,t|t-1}^2\}$ است. بنابراین مشتق‌های تابع $\sigma_{it|t-1}^2$ نسبت به $\varphi = (\omega, \alpha, \gamma)$ به صورت

$$\frac{\partial \sigma_{it|t-1}^2}{\partial \omega} = 1, \quad \frac{\partial \sigma_{it|t-1}^2}{\partial \alpha} = y_{it-1}^2, \quad \frac{\partial \sigma_{it|t-1}^2}{\partial \gamma} = \sigma_{it-1|t-2}^2$$

قابل محاسبه است. با توجه به معادلات به‌دست آمده مشاهده می‌کنیم که معادلات درست‌نمایی (۸) فرم صریح ندارند و با استفاده از روش‌های عددی مانند روش نیوتن رافسون و الگوریتم امیدگیری-بیشینه‌سازی برآورد این پارامترها را به‌دست می‌آوریم.

۳ الگوریتم تقویت

روش تقویت، یک روش یادگیری ماشین برای مسائل رگرسیون و طبقه‌بندی است. مانند سایر روش‌های تقویت، روش تقویت گرادیان ترکیب خطی از یک سری مدل‌های ضعیف برای ایجاد یک مدل قوی و کارآمد است. فرض کنید $\{(x_1, y_1), \dots, (x_n, y_n)\}$ داده‌هایی باشند که مدل برای یادگیری از آن‌ها استفاده می‌کند و هدف از یادگیری کمینه کردن یک تابع زیان مانند $L(y, F(x))$ است. به عبارت دیگر

$$\hat{F}(x) = \arg \min_{F(x)} E_{x,y} [L(y, F(x))]$$

در روش تقویت گرادیان $L(y, F(x)) = (y - F(x))^2$ است، به عبارت دیگر در این روش هدف یادگیری مدلی مانند $F(x)$ برای کمینه کردن میانگین مربعات خطا است. برای پیدا کردن $F(x)$ به صورت مرحله‌ای عمل می‌شود. در مرحله m مولفه جدید $f_m(x)$ با وزن c_m به مدل $F_{m-1}(x)$ اضافه می‌شود و تبدیل به مدل جدید $F_m(x)$ می‌شود، به عبارت دیگر، $F_m = F_{m-1} + c_m f_m(x)$. توجه کنید که مدل $c_m f_m(x)$ به گونه‌ای انتخاب می‌شود که بتواند $y - F_{m-1}(x)$ را پیش‌بینی کند، به عبارت دیگر هدف پیش‌بینی مانده‌ها، $y - F_{m-1}(x)$ ، است. هر مولفه جدید $f_m(x)$ و وزن ترکیب آن c_m را می‌توان بر اساس کمینه کردن تابع زیانی که از پیش تعریف شده، تعیین کرد. روند کلی روش تقویت گرادیان در الگوریتم زیر توضیح داده شده است. این روش تا زمانی تکرار می‌شود که برخی از شرایط همگرایی ایجاد شود.

الگوریتم تقویت گرادیان (فریدمن، ۲۰۰۱):

گام ۱: اولین مدل یک مقدار ثابت $F_0 = \arg \min_{const} \sum_{i=1}^n L(y_i, const)$ انتخاب می‌شود.

گام ۲: برای $m = 1, \dots, M$

الف. شبه-مانده‌های $r_i^m = -\frac{1}{2} \left[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)} \right]_{F(x)=F_{m-1}(x)}$ محاسبه شود.

ب. $f_m^*(x) = \arg \min_{f_m(x)} \sum_{i=1}^n (r_i^m - c_m f_m(x_i))^2$

ج. $c_m^* = \arg \min_c \sum_{i=1}^n L(y_i, F_{m-1}(x_i) + c_m f_m(x_i))$

د. $F_m(x) = F_{m-1}(x) + c_m^* f_m^*(x)$

گام ۳: مدل نهایی $F_M(x) = F_0 + \sum_{m=1}^M c_m^* f_m^*(x)$ است.

در گام ۱ هدف برآورد یک مقدار ثابت است، در این حالت $\hat{y}_i = \bar{y}$. از آنجا که هدف این الگوریتم، کمینه کردن تابع زیان $\sum_{i=1}^n L(y_i, F(x_i)) = \sum_{i=1}^n (y_i - F(x_i))^2$ است که نسبت به $F(x_i)$ پیوسته است، بنابراین $\left[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)} \right] = -2(y_i - F(x_i))$. پس در تکرار m ام داریم

$$r_i^m = (y_i - F(x_i)) \Big|_{F(x)=F_{m-1}(x)} = -\frac{1}{2} \left[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)} \right]_{F(x)=F_{m-1}(x)}$$

۴ مطالعات شبیه‌سازی

در این بخش با استفاده از مطالعات شبیه‌سازی عملکرد نسبی مطالب نظری ارائه شده در مقاله و به طور خاص، عملکرد الگوریتم تقویت و روش گام به گام را در انتخاب متغیر بررسی می‌کنیم. به طور مشابه با راتو و یو (۱۹۹۴)

داده‌ها از مدل درست

$$\begin{aligned} y_{it} &= x_{i,t}B_1 + v_i + u_{it} + e_{it} \\ u_{it} &= \rho u_{i,t-1} + \sigma_{it|t-1}\epsilon_{it}, \\ \sigma_{it|t-1}^2 &= w + \alpha y_{i,t-1}^2 + \beta \sigma_{i,t-1|t-2}^2 \\ |\rho| &< 1, \quad i = 1, \dots, M \quad \text{and} \quad t = 1, \dots, T \end{aligned} \quad (9)$$

تولید شده، که در آن v_i و $e_{i,t}$ ، $\epsilon_{i,t}$ متغیرهای مستقل و هم‌توزیع هستند و از توزیع نرمال استاندارد $N(0, 1)$ پیروی می‌کنند و $\theta = (B_1, \rho, w, \alpha, \beta) = (2, 0.6, 0.02, 0.3, 0.4)$ است. فرض می‌شود $\Sigma = \text{diag}\{\sigma_{e_1}^2, \dots, \sigma_{e_M}^2\}$ معلوم و $\sigma_{e_i}^2 = 1$ است، به رانو و یو (۱۹۹۴) مراجعه کنید. همچنین فرض می‌شود که u_{it} از مدل خودبازگشتی مرتبه ۱، $AR(1)$ و $\sigma_{it|t-1}^2$ از مدل $GARCH(1, 1)$ پیروی می‌کنند.

مدل (۹) را به‌عنوان مدل درست در نظر بگیرید. به‌منظور انتخاب متغیر با استفاده از الگوریتم تقویت و روش گام به گام فرض می‌شود که $x_{1,it}$ ، $x_{2,it}$ ، $x_{3,it}$ ، $x_{4,it}$ به‌ترتیب از توزیع یکنواخت $U(-\sqrt{3}, \sqrt{3})$ ، توزیع یکنواخت $U(-3, 3)$ ، توزیع نرمال $N(0, 3)$ و توزیع گاما $G(2, 5)$ تولید شده‌اند. هدف انتخاب متغیر مناسب از بین مجموعه متغیرهای $\{x_{1,it}, x_{2,it}, x_{3,it}, x_{4,it}\}$ با استفاده از الگوریتم تقویت و روش گام به گام است. برای مدل‌ها رگرسیون با خطای نرمال، هر دو روش حداقل مربعات خطا و بیشینه درست‌نمایی نتایج یکسانی را ایجاد می‌کنند، به این مفهوم که کمینه کردن تابع میانگین مربعات خطا معادل بیشینه کردن تابع لگاریتم درست‌نمایی $L(y, F(x)) = \sum_{i=1}^n \log(f(y; F(x)))$ است، بنابراین در الگوریتم تقویت می‌توان از برآوردهای بیشینه درست‌نمایی استفاده کرد. گام‌های الگوریتم تقویت تا زمانی تکرار می‌شود که برخی از شرایط همگرایی ایجاد شود. در این شبیه‌سازی از معیار اطلاع آکائیک جهت ایجاد شرط همگرایی استفاده شده است، به این مفهوم که $f_m(x_{it})$ با وزن c_m به مدل $F_{m-1}(x_{it})$ اضافه می‌شود به‌طوری‌که مقدار اطلاع آکائیک مدل جدید $F_m(x_{i,t})$ از مقدار اطلاع آکائیک مدل $F_{m-1}(x_{i,t})$ کمتر باشد.

نمونه‌های مستقل $\{y_{it}^{(r)}; t = 1, \dots, T; i = 1, \dots, m; r = 1, \dots, R\}$ با $M = 27$ ناحیه و اندازه نمونه‌های $T = 40, 50, 100, 250, 500, 1000$ تولید شده و نتایج مربوط به برآورد پارامترها، میانگین خطای استاندارد هر یک از برآوردها و انتخاب متغیر بر اساس الگوریتم تقویت در جداول ۱-۶ و بر اساس روش گام به گام در جداول ۷-۱۲ گزارش شده است. در این جداول، B_1 ، B_2 ، B_3 و B_4 به‌ترتیب ضرایب رگرسیونی مربوط به متغیرهای $x_{1,it}$ ، $x_{2,it}$ ، $x_{3,it}$ و $x_{4,it}$ هستند. با استفاده از جداول ۱-۶ مشاهده می‌شود که در تمام نواحی برای اندازه نمونه‌های مختلف الگوریتم تقویت به‌درستی مدل (متغیر) درست را انتخاب می‌کند و با افزایش اندازه نمونه مقادیر برآورد شده به مقادیر درست پارامترها نزدیک می‌شود. با استفاده از جداول ۷-۱۲ مشاهده می‌شود که برای برخی از نواحی، روش گام به گام مدل را به‌درستی تشخیص نمی‌دهد. اگرچه این الگوریتم عملکرد مناسبی در برآورد پارامترهای مدل $GARCH$ برای اندازه نمونه‌های کوچک ندارد اما با افزایش اندازه نمونه مقادیر برآورد شده به

مقادیر درست پارامترها نزدیک می‌شود. با مقایسه نتایج به‌دست آمده از روش گام به گام و الگوریتم تقویت مشاهده می‌شود که الگوریتم تقویت عملکرد مناسب‌تری در انتخاب متغیر دارد. همچنین مقدار میانگین خطای استاندارد و معیار اطلاع آکائیک مدل‌های برازش داده شده با استفاده از الگوریتم تقویت کمتر از مقدار میانگین خطای استاندارد و معیار اطلاع آکائیک مدل‌های برازش داده شده با استفاده از روش گام به گام است. متغیرهای انتخاب شده با استفاده از دو روش در ستون آخر جداول گزارش شده است.

جدول ۱. برازش مدل با استفاده از الگوریتم تقویت وقتی که $T = 40$.

AIC	$\hat{\sigma}_\nu^2$	$\hat{\mu}_\nu$	$\hat{\beta}$	$\hat{\alpha}$	$\hat{w}$	$\hat{\rho}$	$\hat{B}_T$	$\hat{B}_T$	$\hat{B}_T$	$\hat{B}_T$	Area
۱۲۰/۵۵۸۵	۱/۰۲۹۰	۰/۰۲۱۶	۰/۰۰۰۰	۰/۰۷۲۴	۰/۰۰۵۳	۰/۶۱۱۲	-	-	-	-	۲/۱۷۴۸ A_1
۴۲۴۷/۱۲۱	۱/۰۲۵۹	۰/۰۲۷۷	۰/۰۸۳۰	۰/۱۸۴۶	۰/۰۸۶۸	۰/۵۲۶۹	-	-	-	-	۱/۷۷۹۶ A_2
۴۵۸۰/۱۲۲	۱/۰۴۲۰	۰/۰۱۳۱	۰/۲۸۳۵	۰/۱۸۵۰	۰/۰۳۹۶	۰/۶۰۱۴	-	-	-	-	۱/۹۶۳۲ A_3
۴۹۷۴/۱۰۹	۱/۰۲۵۱	۰/۰۲۹۱	۰/۰۲۳۳	۰/۰۹۸۰	۰/۰۸۰۹	۰/۵۰۱۴	-	-	-	-	۲/۱۸۲۹ A_4
۳۷۷۸/۱۲۳	۱/۰۲۷۲	۰/۰۱۵۰	۰/۰۴۶۳	۰/۱۴۷۶	۰/۰۸۱۲	۰/۴۴۰۴	-	-	-	-	۱/۷۶۱۱ A_5
۴۱۵۵/۱۱۹	۱/۰۰۰۱	۰/۰۳۹۱	۰/۲۶۶۷	۰/۶۹۴۸	۰/۰۱۳۳	۰/۴۷۴۵	-	-	-	-	۱/۹۸۳۵ A_6
۰۱۷۴/۱۱۶	۱/۰۲۶۱	۰/۰۲۸۹	۰/۰۴۹۶	۰/۱۱۷۹	۰/۰۷۲۵	۰/۶۳۷۰	-	-	-	-	۲/۰۸۲۶ A_7
۲۹۷۵/۱۲۶	۱/۰۴۳۵	۰/۰۴۱۹	۰/۰۵۱۴	۰/۱۶۰۲	۰/۰۲۴۶	۰/۴۹۷۹	-	-	-	-	۲/۳۱۸۰ A_8
۸۰۲۰/۱۲۲	۱/۰۲۸۱	۰/۰۲۷۸	۰/۰۵۷۸	۰/۱۸۳۴	۰/۰۶۷۵	۰/۴۱۱۸	-	-	-	-	۱/۸۱۷۱ A_9
۷۸۰۸/۱۲۳	۱/۰۲۸۹	۰/۰۴۷۲	۰/۰۰۰۰	۰/۲۰۳۶	۰/۰۳۸۸	۰/۷۱۱۶	-	-	-	-	۲/۰۷۰۴ A_{10}
۱۶۸۱/۱۳۰	۱/۰۲۷۳	۰/۰۱۲۴	۰/۰۴۹۵	۰/۰۹۸۰	۰/۰۵۸۷	۰/۶۹۰۳	-	-	-	-	۲/۰۲۷۳ A_{11}
۷۳۹۸/۱۱۹	۱/۰۱۱۱	۰/۰۱۵۴	۰/۴۹۹۸	۰/۰۹۴۸	۰/۰۲۳۲	۰/۴۷۰۴	-	-	-	-	۱/۹۳۶۶ A_{12}
۳۳۰۸/۱۰۴	۰/۹۹۹۳	۰/۰۴۱۳	۰/۵۷۳۶	۰/۲۶۲۴	۰/۰۳۵۹	۰/۴۸۴۲	-	-	-	-	۱/۸۱۸۲ A_{13}
۸۰۲۹/۱۰۹	۱/۰۳۲۰	۰/۰۰۹۹	۰/۰۸۵۷	۰/۱۸۵۷	۰/۲۵۰۵	۰/۶۲۰۲	-	-	-	-	۲/۶۳۶۰ A_{14}
۵۷۱۵/۱۱۶	۱/۰۲۸۱	۰/۰۳۱۷	۰/۰۴۸۲	۰/۱۶۲۹	۰/۱۰۹۲	۰/۶۵۷۶	-	-	-	-	۲/۲۵۳۰ A_{15}
۲۵۴۹/۱۳۱	۱/۰۲۸۳	۰/۰۰۸۶	۰/۰۴۹۸	۰/۱۹۴۷	۰/۱۹۱۳	۰/۴۴۳۷	-	-	-	-	۲/۳۶۷۲ A_{16}
۴۴۰۴/۱۱۱	۱/۰۲۵۱	۰/۰۰۷۰	۰/۰۴۹۴	۰/۱۳۷۷	۰/۰۷۷۱	۰/۵۰۲۵	-	-	-	-	۱/۷۸۱۰ A_{17}
۱۶۶۹/۱۲۶	۱/۰۳۰۴	۰/۰۱۶۹	۰/۰۵۰۹	۰/۱۶۶۰	۰/۱۳۶۹	۰/۵۴۶۷	-	-	-	-	۱/۶۹۴۶ A_{18}
۶۰۶۴/۱۲۳	۱/۰۱۷۸	۰/۰۲۱۱	۰/۰۵۱۱	۰/۱۰۹۴	۰/۰۸۲۲	۰/۴۷۲۰	-	-	-	-	۲/۲۵۰۲ A_{19}
۴۲۵۴/۱۱۳	۱/۰۵۹۵	۰/۰۱۰۴	۰/۰۰۰۰	۰/۲۶۳۳	۰/۰۶۱۵	۰/۴۰۱۴	-	-	-	-	۱/۸۹۰۸ A_{20}
۶۶۴۹/۱۲۴	۱/۰۴۴۷	۰/۰۰۱۲	۰/۰۴۸۱	۰/۲۱۵۴	۰/۱۶۴۳	۰/۵۰۶۰	-	-	-	-	۲/۳۲۷۰ A_{21}
۴۲۹۲/۱۲۰	۱/۰۴۹۴	۰/۰۴۱۰	۰/۴۱۷۶	۰/۱۶۸۳	۰/۰۲۸۳	۰/۳۲۵۶	-	-	-	-	۲/۰۳۶۶ A_{22}
۱۹۶۹/۱۱۴	۱/۰۳۱۸	۰/۰۱۵۰	۰/۲۸۵۹	۰/۴۱۶۰	۰/۰۲۱۰	۰/۶۵۱۷	-	-	-	-	۱/۹۷۹۳ A_{23}
۰۶۲۴/۱۰۱	۱/۰۲۴۵	۰/۰۱۸۹	۰/۰۴۸۳	۰/۱۷۲۱	۰/۰۷۵۴	۰/۵۳۶۱	-	-	-	-	۲/۱۳۷۲ A_{24}
۶۳۵۰/۱۰۴	۱/۰۲۷۱	۰/۰۰۱۲	۰/۰۴۸۱	۰/۱۳۸۹	۰/۰۵۹۲	۰/۴۶۹۲	-	-	-	-	۲/۱۵۶۵ A_{25}
۶۴۶۷/۱۱۵	۱/۰۲۶۶	۰/۰۱۷۹	۰/۰۴۷۳	۰/۱۸۰۰	۰/۰۸۲۴	۰/۴۴۲۸	-	-	-	-	۱/۷۹۰۴ A_{26}
۱۸۸۶/۱۲۲	۱/۰۳۳۷	۰/۰۲۵۸	۰/۰۴۹۸	۰/۲۱۴۹	۰/۰۲۳۲	۰/۵۱۴۵	-	-	-	-	۱/۶۹۸۲ A_{27}
-	۰/۰۰۰۹	۰/۰۰۱۸	۰/۰۹۸۸	۰/۰۵۸۰	۰/۰۰۶۹	۰/۱۳۰۵	۰/۰۰۰۰	۰/۰۰۰۰	۰/۰۰۰۰	۰/۰۴۶۹	MSE

جدول ۲. برازش مدل با استفاده از الگوریتم تقویت وقتی که $T = 50$.

AIC	$\hat{\sigma}_\nu^2$	$\hat{\mu}_\nu^T$	$\hat{\beta}$	$\hat{\alpha}$	$\hat{w}$	$\hat{\rho}$	$\hat{B}_+$	$\hat{B}_-$	$\hat{B}_+$	$\hat{B}_-$	$\hat{B}_+$	$\hat{B}_-$	Area
۱۲۲/۸۲۷۶	۱/۰۴۲۸	-۰/۰۱۲۸	۰/۱۸۲۰	۰/۲۸۵۲	۰/۰۴۴۲	۰/۵۲۶۹	-	-	-	-	۱/۰۹۸۰۱	-	A_1
۱۳۱/۶۱۸۵	۱/۰۳۵۹	-۰/۰۱۵۶	۰/۵۹۹۸	۰/۲۰۴۲	۰/۰۲۵۹	۰/۶۱۷۹	-	-	-	-	۲/۰۹۶۴	-	A_2
۱۵۰/۸۰۶۰	۱/۰۱۹۳	-۰/۰۲۵۵	۰/۲۵۳۰	۰/۱۹۲۱	۰/۰۵۳۱	۰/۵۸۰۷	-	-	-	-	۲/۱۱۰۸	-	A_3
۱۵۰/۳۵۲۵	۰/۹۹۴۹	-۰/۰۲۰۷	۰/۲۵۰۶	۰/۲۱۷۴	۰/۱۲۸۹	۰/۴۲۵۱	-	-	-	-	۲/۳۰۸۹	-	A_4
۱۴۲/۵۲۸۴	۱/۰۰۲۸	-۰/۰۱۸۲	۰/۵۸۵۸	۰/۲۴۷۸	۰/۱۷۰۱	۰/۵۴۹۶	-	-	-	-	۱/۳۵۹۹	-	A_5
۱۴۸/۸۹۹۱	۱/۰۲۲۶	۰/۰۰۳۶	۰/۲۴۹۹	۰/۱۵۷۴	۰/۰۹۷۶	۰/۴۹۲۸	-	-	-	-	۱/۸۰۶۰	-	A_6
۱۴۴/۲۰۶۵	۱/۰۱۲۲	۰/۰۷۴۷	۰/۳۴۱۹	۰/۴۹۸۴	۰/۰۱۵۱	۰/۴۶۱۰	-	-	-	-	۲/۰۶۴۳	-	A_7
۱۳۹/۰۵۳۴	۱/۰۴۶۱	۰/۰۰۵۹	۰/۲۵۲۲	۰/۱۹۰۱	۰/۰۳۰۱	۰/۴۶۵۷	-	-	-	-	۲/۲۲۵۶	-	A_8
۱۶۱/۵۶۶۸	۱/۰۲۲۱	۰/۰۰۹۱	۰/۲۵۳۷	۰/۱۹۸۹	۰/۰۸۸۱	۰/۶۱۰۷	-	-	-	-	۲/۰۳۱۰	-	A_9
۱۵۲/۳۲۵۷	۱/۰۲۱۳	-۰/۰۰۶۲	۰/۱۶۵۵	۰/۳۴۷۰	۰/۰۴۹۵	۰/۶۶۶۲	-	-	-	-	۱/۰۹۴۹۱	-	A_{10}
۱۵۹/۷۰۹۲	۱/۰۳۸۷	-۰/۰۲۹۱	۰/۱۸۱۰	۰/۲۴۲۷	۰/۱۱۸۷	۰/۴۲۴۲	-	-	-	-	۱/۳۲۴۲	-	A_{11}
۱۶۰/۴۸۷۴	۱/۰۱۰۹	۰/۰۰۹۷	۰/۶۳۰۷	۰/۲۸۲۷	۰/۰۰۷۷	۰/۶۹۱۶	-	-	-	-	۲/۰۹۹۵	-	A_{12}
۱۳۹/۲۷۳۹	۱/۰۲۰۸	-۰/۰۱۴۸	۰/۲۳۷۶	۰/۲۸۶۴	۰/۰۶۶۶	۰/۴۲۹۶	-	-	-	-	۱/۸۱۴۵	-	A_{13}
۱۵۴/۴۵۲۳	۱/۰۱۶۲	-۰/۰۰۳۱	۰/۲۵۱۲	۰/۲۰۳۵	۰/۰۳۲۷	۰/۴۸۲۳	-	-	-	-	۱/۶۷۱۴	-	A_{14}
۱۴۲/۷۷۷۰	۱/۰۳۵۳	۰/۰۱۹۶	۰/۲۵۲۸	۰/۲۱۴۵	۰/۰۴۰۴	۰/۴۷۹۸	-	-	-	-	۱/۹۴۷۳	-	A_{15}
۱۴۸/۷۶۰۲	۱/۰۱۹۲	۰/۰۰۵۲۳	۰/۱۶۹۴	۰/۲۵۶۹	۰/۰۸۱۳	۰/۴۶۴۸	-	-	-	-	۱/۹۱۸۳	-	A_{16}
۱۶۲/۵۲۳۵	۱/۰۱۸۹	۰/۰۰۲۲	۰/۲۷۴۷	۰/۲۲۷۵	۰/۰۶۳۸	۰/۴۸۴۳	-	-	-	-	۱/۷۹۹۳	-	A_{17}
۱۴۴/۷۷۸۸	۱/۰۲۰۴	-۰/۰۳۰۶	۰/۲۵۹۳	۰/۱۹۴۴	۰/۰۴۳۹	۰/۵۲۶۱	-	-	-	-	۱/۹۸۴۱	-	A_{18}
۱۴۹/۱۳۰۸	۱/۰۱۹۷	۰/۰۴۸۴	۰/۱۷۲۲	۰/۱۸۰۴	۰/۱۰۱۵	۰/۰۱۷۹	-	-	-	-	۲/۶۷۸۳	-	A_{19}
۱۳۸/۷۳۰۰	۱/۰۳۶۶	-۰/۰۷۷۳	۰/۲۰۱۴	۰/۱۹۰۰	۰/۰۳۹۶	۰/۵۷۱۱	-	-	-	-	۱/۹۶۷۲	-	A_{20}
۱۴۱/۵۹۰۲	۱/۰۲۳۰	۰/۰۰۸۶	۰/۱۹۶۶	۰/۳۸۷۷	۰/۰۳۹۷	۰/۵۰۲۴	-	-	-	-	۲/۰۵۸۷	-	A_{21}
۱۲۳/۸۶۳۷	۱/۰۰۲۱	-۰/۰۰۰۵	۰/۲۴۸۹	۰/۱۹۲۱	۰/۰۸۳۳	۰/۴۱۸۲	-	-	-	-	۱/۷۶۶۴	-	A_{22}
۱۵۲/۹۱۹۸	۱/۰۲۲۲	-۰/۰۱۴۶	۰/۲۵۱۳	۰/۲۱۴۴	۰/۰۵۸۴	۰/۴۹۰۶	-	-	-	-	۲/۳۱۱۹	-	A_{23}
۱۵۸/۹۱۳۱	۱/۰۲۷۲	-۰/۰۰۶۶	۰/۵۸۱۵	۰/۲۸۶۰	۰/۰۱۴۶	۰/۵۹۵۷	-	-	-	-	۲/۱۳۵۰	-	A_{24}
۱۴۳/۰۸۰۷	۱/۰۳۸۱	-۰/۰۲۶۹	۰/۱۸۸۴	۰/۲۰۰۳	۰/۰۵۵۰	۰/۶۸۷۴	-	-	-	-	۱/۹۱۷۶	-	A_{25}
۱۴۱/۷۶۹۰	۱/۰۳۹۲	-۰/۰۲۰۱	۰/۲۶۹۸	۰/۱۸۰۰	۰/۰۴۱۲	۰/۳۵۹۵	-	-	-	-	۲/۰۷۶۵	-	A_{26}
۱۵۲/۹۶۸۱	۱/۰۲۱۲	-۰/۰۱۰۱	۰/۲۵۳۱	۰/۲۱۸۵	۰/۰۷۲۵	۰/۲۷۳۳	-	-	-	-	۲/۱۹۰۲	-	A_{27}
-	۰/۰۰۰۴	۰/۰۰۰۶	۰/۰۹۸۱	۰/۰۵۳۹	۰/۰۰۶۸	۰/۱۱۹۱	۰/۰۰۰۰	۰/۰۰۰۰	۰/۰۰۰۰	۰/۰۰۰۰	۰/۰۳۶۲	-	MSE

به‌منظور مطالعات بیشتر 10^3 مرتبه شبیه‌سازی تکرار شده و فراوانی نسبی تعداد دفعاتی که الگوریتم تقویت و روش گام به گام مدل را به‌درستی تشخیص داده‌اند به ترتیب در جدول ۱۳ و جدول ۱۴ ارائه شده است. نتایج این جداول نشان می‌دهد که فراوانی نسبی تشخیص مدل درست با استفاده از الگوریتم تقویت (برخلاف روش گام به گام) به یک همگرا است، بنابراین الگوریتم تقویت عملکرد مناسب‌تری در انتخاب متغیر نسبت به روش گام به گام دارد. شکل ۴ نیز تأیید کننده این نتایج است.

۵ تحلیل داده‌های بدهی‌های مالیاتی

در این بخش با ارائه یک مثال تجربی عملکرد الگوریتم تقویت و روش گام به گام را در انتخاب متغیر بررسی می‌کنیم. داده‌های بدهی مالیاتی را در نظر بگیرید که شامل اطلاعات ۲۵۸ پرداخت‌کننده مالیات در ایالت متحده است و طی سال‌های ۱۹۸۲ و ۱۹۸۷-۱۹۸۶ جمع‌آوری شده‌اند. این داده‌ها به‌منظور بررسی اثر ویژگی‌های فردی و اقتصادی افراد بر روی میزان بدهی مالیاتی آن‌ها مورد مطالعه قرار گرفته‌اند. (فریز، ۲۰۰۴؛ راب-هسکت و اسکوندال، ۲۰۰۸). مجموعه داده‌ها شامل یک متغیر پاسخ، چهار متغیر توضیحی مربوط با ویژگی‌های فردی و چهار متغیر توضیحی مربوط به ویژگی‌های اقتصادی است. متغیر لگاریتم بدهی‌های مالی فرد $LNTAX$ ، متغیر پاسخ است که بر اساس

جدول ۳. برازش مدل با استفاده از الگوریتم تقویت وقتی که $T = ۱۰۰$.

AIC	$\hat{\sigma}_\nu^2$	$\hat{\mu}_\nu^2$	$\hat{\beta}$	$\hat{\alpha}$	$\hat{w}$	$\hat{\rho}$	$\hat{B}_+$	$\hat{B}_+$	$\hat{B}_+$	$\hat{B}_+$	Area
۲۹۸/۲۲۸۲	۱/۰۰۹۶	۰/۰۱۷۵	۰/۱۳۵۹	۰/۰۶۹۸	۰/۰۷۷۸	۰/۴۳۸۶	—	—	—	۲/۱۷۰۰	A_1
۳۱۶/۵۳۳۰	۱/۰۱۰۰	۰/۰۱۴۵	۰/۲۴۴۹	۰/۱۵۴۸	۰/۰۶۵۴	۰/۴۹۳۴	—	—	—	۲/۱۴۷۹	A_2
۳۹۲/۶۰۳۰	۱/۰۱۶۰	۰/۰۱۱۰	۰/۲۵۱۵	۰/۲۳۷۵	۰/۰۴۹۰	۰/۶۶۱۹	—	—	—	۱/۹۹۸۲	A_3
۲۹۳/۷۱۸۳	۱/۰۰۸۲	۰/۰۰۰۳	۰/۲۳۵۴	۰/۱۹۵۲	۰/۰۵۴۱	۰/۶۰۱۰	—	—	—	۱/۹۹۴۳	A_4
۲۶۷/۷۱۱۴	۱/۰۰۴۵	۰/۰۲۱۴	۰/۴۱۴۷	۰/۳۱۰۹	۰/۰۲۱۳	۰/۶۲۴۲	—	—	—	۲/۰۳۵۶	A_5
۲۵۹/۵۱۸۸	۱/۰۰۷۰	۰/۰۳۲۴	۰/۵۸۰۲	۰/۲۱۹۹	۰/۰۱۲۹	۰/۴۵۵۵	—	—	—	۲/۰۶۱۴	A_6
۲۸۱/۸۳۰۰	۱/۰۱۲۴	۰/۰۰۷۶	۰/۲۷۸۴	۰/۲۰۲۶	۰/۰۶۹۲	۰/۵۰۳۵	—	—	—	۱/۸۳۴۶	A_7
۲۹۰/۱۹۹۶	۱/۰۱۰۵	۰/۰۰۲۰	۰/۲۶۷۰	۰/۲۰۱۷	۰/۰۷۵۹	۰/۴۶۷۱	—	—	—	۲/۰۶۶۰	A_8
۲۷۲/۶۱۱۵	۱/۰۰۹۵	۰/۰۰۷۲	۰/۲۹۲۶	۰/۲۲۲۵	۰/۰۵۹۴	۰/۵۷۲۶	—	—	—	۲/۰۵۷۵	A_9
۲۹۴/۸۰۹۴	۰/۹۹۲۰	۰/۰۰۵۵	۰/۴۴۲۳	۰/۲۰۷۶	۰/۰۰۲۳	۰/۶۹۵۷	—	—	—	۱/۹۹۲۷	A_{10}
۳۱۳/۴۵۳۸	۱/۰۱۹۷	۰/۰۰۵۲	۰/۲۵۲۸	۰/۲۲۰۱	۰/۰۵۷۰	۰/۳۹۰۸	—	—	—	۲/۲۴۴۱	A_{11}
۲۸۹/۸۸۵۳	۱/۰۱۱۳	۰/۰۱۴۸	۰/۲۵۰۳	۰/۲۱۵۰	۰/۱۰۴۲	۰/۴۲۷۵	—	—	—	۱/۶۸۰۰	A_{12}
۲۹۰/۶۴۲۲	۱/۰۰۸۴	۰/۰۰۱۴	۰/۴۶۱۹	۰/۲۱۱۹	۰/۰۲۲۹	۰/۶۱۲۶	—	—	—	۱/۹۹۲۷	A_{13}
۲۸۱/۹۸۶۳	۱/۰۱۱۶	۰/۰۳۷۹	۰/۴۸۹۴	۰/۲۵۸۰	۰/۰۱۹۷	۰/۵۲۹۲	—	—	—	۱/۹۵۰۸	A_{14}
۲۸۱/۵۵۰۶	۱/۰۰۴۵	۰/۰۲۴۲	۰/۳۸۸۵	۰/۲۳۴۲	۰/۰۳۹۷	۰/۴۰۶۶	—	—	—	۲/۱۴۴۱	A_{15}
۳۰۹/۷۷۵۲	۱/۰۱۰۷	۰/۰۰۴۶	۰/۳۲۵۳	۰/۱۹۴۷	۰/۰۳۸۰	۰/۵۱۷۱	—	—	—	۲/۰۹۱۲	A_{16}
۳۰۱/۸۷۰۷	۱/۰۰۵۲	۰/۰۰۲۴	۰/۵۰۳۳	۰/۲۵۸۰	۰/۰۰۵۵	۰/۵۴۲۳	—	—	—	۱/۹۲۶۵	A_{17}
۳۰۹/۳۲۷۸	۱/۰۰۵۱	۰/۰۰۶۱۸	۰/۲۱۸۷	۰/۲۸۷۱	۰/۰۶۱۲	۰/۴۹۴۹	—	—	—	۱/۹۶۸۰	A_{18}
۳۰۷/۲۴۰۹	۱/۰۰۱۹	۰/۰۰۱۷	۰/۴۵۵۲	۰/۲۳۳۲	۰/۰۲۹۷	۰/۵۸۳۰	—	—	—	۱/۹۸۸۳	A_{19}
۲۹۰/۹۸۱۲	۱/۰۰۸۸	۰/۰۱۹۷	۰/۴۳۹۹	۰/۲۷۱۱	۰/۰۱۵۷	۰/۶۳۳۵	—	—	—	۲/۰۱۶۵	A_{20}
۳۱۰/۱۹۸۴	۱/۰۴۳۸	۰/۰۰۳۲	۰/۲۹۴۵	۰/۲۵۹۸	۰/۰۴۴۶	۰/۵۳۷۹	—	—	—	۱/۹۸۵۷	A_{21}
۳۰۷/۲۳۵۱	۱/۰۲۴۳	۰/۰۱۱۰	۰/۲۳۲۵	۰/۲۱۰۶	۰/۰۰۸۰	۰/۴۷۶۶	—	—	—	۱/۸۸۰۹	A_{22}
۳۲۳/۹۰۰۵	۱/۰۱۰۳	۰/۰۰۹۱	۰/۲۹۶۳	۰/۲۴۱۲	۰/۰۸۸۲	۰/۶۳۳۴	—	—	—	۲/۰۸۸۱	A_{23}
۲۶۵/۱۸۴۵	۱/۰۰۳۹	۰/۰۰۰۷	۰/۲۶۹۹	۰/۱۹۶۶	۰/۰۷۲۰	۰/۵۷۴۷	—	—	—	۲/۰۲۶۳	A_{24}
۲۷۲/۶۲۷۳	۱/۰۰۹۳	۰/۰۰۴۵	۰/۴۸۱۴	۰/۲۶۳۱	۰/۰۴۷۵	۰/۶۷۷۴	—	—	—	۲/۱۷۲۴	A_{25}
۲۸۸/۱۴۳۹	۱/۰۱۰۱	۰/۰۱۴۹	۰/۲۷۲۱	۰/۲۸۳۱	۰/۰۵۱۹	۰/۵۶۲۸	—	—	—	۱/۹۳۹۳	A_{26}
۳۰۰/۰۳۲۲	۱/۰۰۱۹	۰/۰۶۸۶	۰/۳۰۳۷	۰/۳۷۲۶	۰/۰۳۶۹	۰/۵۵۶۵	—	—	—	۱/۹۹۰۷	A_{27}
—	۰/۰۰۰۲	۰/۰۰۰۵	۰/۰۸۱۲	۰/۰۳۳۱	۰/۰۰۵۷	۰/۰۱۴۴	۰/۰۰۰۰	۰/۰۰۰۰	۰/۰۰۰۰	۰/۰۲۰۵	MSE

دلار سال ۱۹۸۳ ثبت شده است. متغیرهای توضیحی مربوط با ویژگی‌های فردی عبارتند از: MS متغیر نشانگر برای متأهلین، HH متغیر نشانگر برای سرپرست خانوارها، AGE متغیر نشانگر برای افراد ۶۵ سال یا بیشتر از آن و $DEPEND$ نشان دهنده تعداد افراد تحت تکفل بیمه‌ای مالیات دهنده است. متغیر توضیحی مربوط به ویژگی‌های اقتصادی عبارتند از: $LNTPI$ لگاریتم مجموع درآمدهای خالص فرد بر اساس دلار سال ۱۹۸۳، MR نرخ نهایی مالیات، EMP متغیر نشانگر برای استفاده از تسهیلات تخفیفی توسط مالیات دهندگان و $PREP$ نشان دهنده میزان پرداخت‌های اولیه مالیات دهنده است.

در مطالعات مختلف، مدل

$$y_{it} = B_0 + B_1 LNTPI_{it} + B_2 MS_{it} + B_3 HH_{it} + B_4 DEPEND_{it} + B_5 AGE_{it} + B_6 MR_{it} + B_7 EMP_{it} + B_8 PREP_{it} + v_i + \epsilon_{it}$$

در نظر گرفته شده که در آن $i = 1, \dots, 258$ و $t = 1, \dots, 5$ است و ϵ_{it} از توزیع نرمال پیروی می‌کند (فریز، ۲۰۰۴؛ راب-هسکت و اسکوندال، ۲۰۰۸). برآورد ضریب همبستگی درون خوشه‌ای برابر با 0.228 به‌دست آمده است و نشان می‌دهد که حدود ۲۳ درصد از کل تغییرپذیری مشاهدات از تغییرپذیری بین افراد ناشی

جدول ۴. برازش مدل با استفاده از الگوریتم تقویت وقتی که $T = ۲۵۰$.

AIC	$\hat{\sigma}_\nu^2$	$\hat{\mu}_\nu^2$	$\hat{\beta}$	$\hat{\alpha}$	$\hat{w}$	$\hat{\rho}$	$\hat{B}_\tau$	$\hat{B}_\tau$	$\hat{B}_\tau$	$\hat{B}_\tau$	Area
۷۵۰/۶۶۳۳	۱/۰۰۳۳	۰/۰۱۰۷	۰/۲۷۹۹	۰/۲۳۲۸	۰/۰۲۰۴	۰/۵۱۳۹	—	—	—	۲/۰۳۰۸	A_1
۷۶۳/۷۳۸۲	۱/۰۰۱۲	۰/۰۴۸۰	۰/۲۶۳۲	۰/۴۲۶۷	۰/۰۳۵۵	۰/۶۳۵۵	—	—	—	۱/۹۸۴۱	A_2
۷۵۶/۳۴۶۷	۱/۰۰۵۲	۰/۰۱۵۲	۰/۳۹۳۶	۰/۲۳۸۲	۰/۰۳۹۳	۰/۷۱۲۷	—	—	—	۱/۸۵۱۹	A_3
۷۲۵/۳۳۱۸	۱/۰۰۸۱	۰/۰۰۲۳	۰/۵۴۴۲	۰/۱۹۵۱	۰/۰۰۷۷	۰/۴۸۸۰	—	—	—	۲/۱۰۶۸	A_4
۷۱۲/۹۷۰۹	۱/۰۰۳۷	۰/۰۰۹۱	۰/۲۷۵۵	۰/۲۱۸۴	۰/۰۳۳۰	۰/۵۰۰۷	—	—	—	۱/۸۹۸۴	A_5
۷۵۶/۰۹۸۳	۱/۰۰۴۵	۰/۰۰۴۹	۰/۴۷۸۶	۰/۲۸۵۸	۰/۰۱۶۸	۰/۷۱۶۰	—	—	—	۱/۹۸۲۷	A_6
۷۱۴/۸۰۳۸	۱/۰۰۵۸	۰/۰۰۶۶	۰/۲۹۳۰	۰/۲۴۲۹	۰/۰۷۳۲	۰/۴۲۳۸	—	—	—	۲/۱۱۹۰	A_7
۷۵۴/۱۱۹۴	۱/۰۰۲۷	۰/۰۱۱۷	۰/۴۵۶۴	۰/۲۵۳۳	۰/۰۲۰۵	۰/۶۱۶۸	—	—	—	۲/۰۱۵۵	A_8
۷۲۰/۹۲۷۷	۱/۰۰۳۳	۰/۰۱۴۸	۰/۵۳۸۲	۰/۲۶۰۳	۰/۰۱۵۳	۰/۶۴۴۳	—	—	—	۱/۹۷۶۹	A_9
۶۹۳/۲۶۳۴	۱/۰۰۳۹	۰/۰۰۱۲	۰/۳۷۳۰	۰/۲۵۱۴	۰/۰۲۴۴	۰/۵۶۳۷	—	—	—	۱/۹۶۶۳	A_{10}
۷۵۶/۸۴۲۷	۱/۰۰۷۴	۰/۰۰۱۲	۰/۴۸۴۵	۰/۲۰۴۱	۰/۰۲۷۷	۰/۴۲۳۹	—	—	—	۱/۸۶۲۲	A_{11}
۷۲۷/۷۶۵۱	۱/۰۰۵۲	۰/۰۰۳۷	۰/۴۰۹۵	۰/۱۹۸۲	۰/۰۴۴۸	۰/۵۰۳۴	—	—	—	۱/۸۳۸۷	A_{12}
۷۱۹/۲۸۲۲	۱/۰۰۴۱	۰/۰۰۶۵	۰/۲۹۲۸	۰/۱۹۵۵	۰/۱۴۵۸	۰/۳۳۵۸	—	—	—	۲/۲۸۲۲	A_{13}
۷۲۱/۸۴۳۷	۱/۰۰۴۰	۰/۰۰۳۹	۰/۲۹۹۳	۰/۲۱۳۶	۰/۰۵۱۴	۰/۶۷۵۹	—	—	—	۲/۰۹۴۵	A_{14}
۷۳۶/۳۰۹۱	۱/۰۰۳۴	۰/۰۰۰۱	۰/۳۰۵۳	۰/۲۴۱۰	۰/۰۳۳۸	۰/۶۱۵۷	—	—	—	۲/۰۱۴۴	A_{15}
۷۱۲/۲۵۲۷	۱/۰۰۲۶	۰/۰۰۲۵	۰/۴۱۲۶	۰/۳۵۹۰	۰/۰۲۷۴	۰/۵۴۸۴	—	—	—	۱/۹۵۹۵	A_{16}
۷۰۸/۶۱۷۱	۱/۰۰۰۳	۰/۰۰۴۶۷	۰/۳۱۶۱	۰/۴۰۷۵	۰/۰۲۰۵	۰/۵۳۶۱	—	—	—	۱/۹۳۲۵	A_{17}
۷۴۲/۱۶۶۶	۱/۰۰۷۵	۰/۰۱۲۸	۰/۲۹۴۸	۰/۲۰۱۵	۰/۰۶۲۷	۰/۵۱۷۷	—	—	—	۱/۸۴۲۷	A_{18}
۷۰۷/۶۴۸۱	۱/۰۰۲۲	۰/۰۰۴۶	۰/۳۵۰۵	۰/۳۸۰۵	۰/۰۲۲۸	۰/۴۷۰۱	—	—	—	۱/۹۸۲۴	A_{19}
۷۷۳/۶۳۹۳	۱/۰۰۰۲	۰/۰۱۲۳	۰/۵۲۸۲	۰/۳۳۲۸	۰/۰۱۱۷	۰/۴۹۱۵	—	—	—	۱/۹۶۳۵	A_{20}
۷۹۵/۵۷۸۶	۱/۰۰۴۷	۰/۰۱۲۵	۰/۳۶۴۶	۰/۳۲۸۶	۰/۰۲۵۳	۰/۴۹۲۸	—	—	—	۱/۹۵۲۶	A_{21}
۷۵۸/۲۹۲۰	۱/۰۰۵۰	۰/۰۰۰۷	۰/۵۵۸۷	۰/۲۶۶۸	۰/۰۱۹۴	۰/۴۴۵۲	—	—	—	۱/۸۰۸۶	A_{22}
۷۴۰/۸۷۹۵	۱/۰۰۴۵	۰/۰۰۲۴	۰/۲۱۹۱	۰/۲۳۴۷	۰/۰۷۱۱	۰/۴۸۵۱	—	—	—	۱/۸۵۱۶	A_{23}
۷۵۶/۶۵۵۳	۱/۰۰۴۰	۰/۰۱۹۸	۰/۳۰۳۵	۰/۳۹۰۹	۰/۰۲۹۰	۰/۶۳۴۱	—	—	—	۲/۰۲۹۴	A_{24}
۷۴۹/۴۳۲۱	۰/۹۹۷۵	۰/۰۱۶۸	۰/۵۰۴۷	۰/۲۲۵۱	۰/۰۰۴۹	۰/۶۴۵۳	—	—	—	۲/۰۲۷۲	A_{25}
۷۳۳/۰۲۵۳	۱/۰۰۳۶	۰/۰۰۲۰	۰/۳۹۰۷	۰/۲۸۰۲	۰/۰۳۵۱	۰/۵۷۳۵	—	—	—	۲/۰۲۳۵	A_{26}
۷۴۲/۴۰۳۵	۱/۰۰۱۶	۰/۰۲۹۷	۰/۴۷۱۷	۰/۲۴۰۹	۰/۰۱۷۵	۰/۶۳۱۳	—	—	—	۲/۰۴۵۵	A_{27}
—	۰/۰۰۰۰	۰/۰۰۰۲	۰/۰۷۶۱	۰/۰۱۹۹	۰/۰۰۰۶	۰/۰۰۷۳	۰/۰۰۰۰	۰/۰۰۰۰	۰/۰۰۰۰	۰/۰۰۵۸	MSE

می‌شود. ما مدل

$$\begin{aligned}
 y_{it} &= \mathbf{x}_{it}\mathbf{B} + v_i + u_{it} + e_{it} \\
 u_{it} &= \rho u_{i,t-1} + \sigma_{it|t-1}\epsilon_{it}, \\
 \sigma_{it|t-1}^2 &= w + \alpha y_{i,t-1}^2 + \beta \sigma_{i,t-1|t-2}^2 \\
 |\rho| &< 1, \quad i = 1, \dots, 258 \quad t = 1, \dots, 5
 \end{aligned} \tag{۱۰}$$

را در نظر گرفته و با استفاده از الگوریتم تقویت و روش گام به گام متغیرهای توضیحی مناسب را انتخاب می‌کنیم، که در آن

$$\mathbf{x}_j = c(LNTPI_{it}, MS_{it}, HH_{it}, DEPEND_{it}, AGE_{it}, MR_{it}, EMP_{it}, PREP_{it})$$

و $\mathbf{B} = (B_1, B_2, B_3, B_4, B_5, B_6, B_7, B_8)$ است و $e_{i,t}$ ، $\epsilon_{i,t}$ و v_i متغیرهای مستقل و هم‌توزیع هستند و از

جدول ۵. برازش مدل با استفاده از الگوریتم تقویت وقتی که $T = 500$.

AIC	$\hat{\sigma}_\nu^2$	$\hat{\mu}_\nu$	$\hat{\beta}$	$\hat{\alpha}$	$\hat{w}$	$\hat{\rho}$	$\hat{B}_+$	$\hat{B}_-$	$\hat{B}_\tau$	$\hat{B}_1$	Area
۱۴۶۱/۱۴۶	۱/۰۰۰۲۸	-۰/۰۰۰۰۷	۰/۲۸۴۵	۰/۲۷۳۵	۰/۰۵۴۶	۰/۵۵۵۳	-	-	-	۲/۰۸۵۹	A_1
۱۴۵۴/۱۵۶	۱/۰۰۰۳۲	-۰/۰۰۰۰۴	۰/۳۴۴۸	۰/۲۹۰۴	۰/۰۳۱۹	۰/۵۵۸۳	-	-	-	۲/۰۴۵۲	A_2
۱۴۶۲/۳۹۰	۱/۰۰۰۱۳	-۰/۰۰۰۰۴	۰/۳۵۵۰	۰/۲۸۵۷	۰/۰۲۲۵	۰/۶۰۰۷	-	-	-	۱/۰۹۳۷۷	A_3
۱۴۸۰/۱۷۶	۱/۰۰۰۱۸	-۰/۰۰۰۰۵	۰/۴۵۲۳	۰/۲۷۱۳	۰/۰۱۸۳	۰/۶۰۸۵	-	-	-	۲/۰۰۰۳۷	A_4
۱۴۵۷/۶۸۶	۱/۰۰۰۱۲	۰/۰۰۰۳۶	۰/۳۱۱۳	۰/۲۱۰۴	۰/۰۱۵۵	۰/۶۲۶۶	-	-	-	۱/۰۹۳۷۷	A_5
۱۴۳۴/۳۷۱	۱/۰۰۰۱۲	۰/۰۰۰۶۵	۰/۳۷۶۱	۰/۲۷۱۷	۰/۰۱۳۳	۰/۶۳۴۹	-	-	-	۱/۰۹۸۹۲	A_6
۱۴۵۶/۵۹۲	۰/۰۹۸۳	-۰/۰۰۰۶۸	۰/۳۵۲۶	۰/۲۵۸۶	۰/۰۰۵۶	۰/۵۸۷۸	-	-	-	۲/۰۵۵۶۶	A_7
۱۵۰۰/۱۹۹	۱/۰۰۰۰۲	۰/۰۰۰۹۰	۰/۴۱۲۰	۰/۳۸۵۶	۰/۰۱۹۶	۰/۶۳۶۳	-	-	-	۲/۰۳۸۵	A_8
۱۴۴۶/۱۲۳	۱/۰۰۰۱۳	-۰/۰۰۰۵۶	۰/۴۶۱۳	۰/۳۳۲۴	۰/۰۱۶۸	۰/۵۳۵۰	-	-	-	۱/۰۹۸۹۳	A_9
۱۴۸۰/۱۳۶	۱/۰۰۰۱۳	-۰/۰۰۰۴۰	۰/۳۴۴۸	۰/۲۹۱۹	۰/۰۲۵۸	۰/۵۷۸۷	-	-	-	۱/۰۹۶۳۲	A_{10}
۱۴۴۹/۵۲۰	۱/۰۰۰۰۲	۰/۰۰۰۶۳	۰/۳۴۵۵	۰/۲۶۲۷	۰/۰۱۳۸	۰/۵۸۹۲	-	-	-	۲/۰۴۴۸	A_{11}
۱۴۷۸/۴۷۶	۱/۰۰۰۱۶	-۰/۰۰۰۱۰	۰/۲۷۰۲	۰/۲۹۹۲	۰/۰۳۴۵	۰/۵۸۰۹	-	-	-	۲/۰۹۷۷	A_{12}
۱۴۸۷/۷۳۵	۱/۰۰۰۱۴	۰/۰۰۰۵۳	۰/۳۲۴۹	۰/۲۹۰۵	۰/۰۲۵۴	۰/۵۲۱۴	-	-	-	۲/۰۸۳۰	A_{13}
۱۴۲۰/۹۵۵	۱/۰۰۰۱۶	۰/۰۰۰۳۰	۰/۴۲۱۸	۰/۳۳۱۱	۰/۰۲۰۶	۰/۵۸۸۴	-	-	-	۱/۰۹۹۳۷	A_{14}
۱۴۲۹/۸۹۸	۱/۰۰۰۱۰	۰/۰۰۰۵۷	۰/۳۵۳۲	۰/۳۴۷۴	۰/۰۲۳۷	۰/۵۶۵۵	-	-	-	۲/۰۰۱۶۸	A_{15}
۱۴۶۴/۸۹۲	۱/۰۰۰۵۴	-۰/۰۰۰۲۹	۰/۳۹۲۵	۰/۲۵۰۲	۰/۰۰۸۰	۰/۵۸۱۵	-	-	-	۱/۰۸۸۲۶	A_{16}
۱۴۵۲/۵۹۸	۱/۰۰۰۰۹	۰/۰۰۰۳۵	۰/۴۴۶۷	۰/۲۶۵۱	۰/۰۱۷۸	۰/۵۲۴۵	-	-	-	۲/۰۵۱۶	A_{17}
۱۴۷۰/۵۹۷	۱/۰۰۰۱۴	-۰/۰۰۰۸۲	۰/۳۱۳۷	۰/۲۵۰۹	۰/۰۱۵۷	۰/۵۸۵۱	-	-	-	۲/۰۰۰۸۵	A_{18}
۱۴۹۰/۶۶۱	۱/۰۰۰۲۳	۰/۰۰۰۰۲	۰/۴۱۴۱	۰/۲۶۴۶	۰/۰۳۱۲	۰/۵۴۴۹	-	-	-	۲/۰۰۸۴۶	A_{19}
۱۵۱۰/۲۷۱	۱/۰۰۰۱۶	-۰/۰۰۰۹۳	۰/۴۳۹۸	۰/۳۴۶۰	۰/۰۱۸۰	۰/۵۳۳۲	-	-	-	۱/۰۹۹۸۷	A_{20}
۱۵۰۵/۵۲۵	۱/۰۰۰۱۶	-۰/۰۰۰۸۱	۰/۳۱۸۴	۰/۳۲۹۵	۰/۰۲۳۲	۰/۵۵۳۵	-	-	-	۱/۰۹۷۸۵	A_{21}
۱۴۴۵/۱۹۵	۱/۰۰۰۱۸	-۰/۰۰۰۱۳	۰/۳۴۵۷	۰/۳۵۴۸	۰/۰۳۶۹	۰/۵۲۶۳	-	-	-	۱/۰۹۵۷۷	A_{22}
۱۴۹۵/۲۷۶	۱/۰۰۰۱۹	-۰/۰۰۰۹۵	۰/۳۶۳۶	۰/۳۱۶۱	۰/۰۳۳۶	۰/۶۵۱۴	-	-	-	۲/۰۳۹۰	A_{23}
۱۴۶۴/۰۷۶	۱/۰۰۰۱۷	-۰/۰۰۰۱۰	۰/۳۴۹۷	۰/۳۸۳۳	۰/۰۱۹۲	۰/۶۱۸۹	-	-	-	۲/۰۰۱۴۸	A_{24}
۱۴۶۰/۱۹۵	۱/۰۰۰۲۳	۰/۰۰۰۵۴	۰/۳۸۸۰	۰/۳۵۷۹	۰/۰۲۱۱	۰/۶۴۵۵	-	-	-	۱/۰۹۸۲۵	A_{25}
۱۵۰۷/۸۲۳	۱/۰۰۰۱۷	۰/۰۰۰۱۶	۰/۲۸۶۱	۰/۲۶۰۱	۰/۰۳۲۱	۰/۶۴۹۲	-	-	-	۲/۰۵۳۹	A_{26}
۱۴۸۳/۲۶۶	۱/۰۰۰۱۸	۰/۰۰۰۲۳	۰/۳۰۹۶	۰/۲۸۷۲	۰/۰۲۷۵	۰/۶۲۴۵	-	-	-	۱/۰۹۸۰۸	A_{27}
-	۰/۰۰۰۰۰	۰/۰۰۰۰۲	۰/۰۰۲۵۹	۰/۰۰۰۷۴	۰/۰۰۰۰۱	۰/۰۰۰۳۹	۰/۰۰۰۰۰	۰/۰۰۰۰۰	۰/۰۰۰۰۰	۰/۰۰۰۳۴	MSE

توزیع نرمال پیروی می‌کنند. نتایج مربوط به برآورد پارامترها،

$$AIC = -2\ell(\theta; y_t) + 2k \quad (11)$$

و $MSE = \frac{1}{T-1} \sum_{t=2}^T (y_t - \hat{y}_t)^2$ مربوط به مدل‌های برازش داده شده بر اساس الگوریتم تقویت و روش گام به گام به ترتیب در شکل‌های ۲ و ۳ نمایش داده شده است که در آن $\ell(\theta; y_t)$ نشان دهنده لگاریتم تابع درستنمایی، k نشان دهنده تعداد پارامترهای مدل، y_t مقدار واقعی و $\hat{y}_t$ مقدار پیش‌بینی است. رابطه (۱۱) با این فرض به دست می‌آید که مدل درست در مدل پارامتری مشخص شده وجود داشته باشد. این رابطه نشان می‌دهد که ارببی لگاریتم تابع درستنمایی به عنوان برآوردگر ریسک کولیک-لیبیل تقریباً برابر با تعداد پارامترهای آزاد موجود در مدل است، بنابراین معیار اطلاع آکائیک به عنوان برآوردگر نا اربب برای ریسک کولیک-لیبیل پیشنهاد می‌شود. همچنین میانگین نتایج به دست آمده از ۲۵۸ ناحیه (افراد مورد مطالعه) در جدول ۱۵ ارائه شده است. از این جدول مشاهده می‌شود که متغیرهای توضیحی $LNTPI$ و MR با استفاده از الگوریتم تقویت و متغیرهای توضیحی $LNTPI$ ، MS ، $DEPEND$ و MR با استفاده از روش گام به گام انتخاب شده‌اند. اگر چه مدل برازش داده شده بر اساس روش گام به گام دارای MSE کمتری است اما مدل برازش داده شده بر اساس الگوریتم تقویت دارای AIC کمتری است.

جدول ۶. برازش مدل با استفاده از الگوریتم تقویت وقتی که $T = ۱۰۰۰$.

AIC	$\hat{\sigma}_\nu^2$	$\hat{\mu}_\nu^2$	$\hat{\beta}$	$\hat{\alpha}$	$\hat{w}$	$\hat{\rho}$	$\hat{B}_\tau$	$\hat{B}_\tau$	$\hat{B}_\tau$	$\hat{B}_\tau$	Area
۲۹۴۶/۰۰۵	۱/۰۰۱۰	۰/۰۰۳۶	۰/۳۸۱۸	۰/۲۸۹۱	۰/۰۲۳۲	۰/۶۲۳۳	—	—	—	—	۱/۹۷۱۲ A_1
۲۹۶۴/۶۷۷	۱/۰۰۰۷	۰/۰۰۲۱	۰/۳۷۰۷	۰/۲۷۴۳	۰/۰۲۳۱	۰/۵۸۴۵	—	—	—	—	۱/۹۶۲۰ A_2
۲۹۱۴/۶۱۸	۱/۰۰۰۶	۰/۰۰۳۰	۰/۴۱۸۲	۰/۳۰۴۰	۰/۰۲۰۶	۰/۶۱۷۷	—	—	—	—	۱/۹۹۳۸ A_3
۲۹۷۴/۸۳۵	۱/۰۰۰۱	۰/۰۰۱۶۷	۰/۴۰۳۸	۰/۳۱۱۲	۰/۰۱۷۷	۰/۶۱۰۹	—	—	—	—	۲/۰۱۰۵ A_4
۲۹۲۴/۵۷۵	۱/۰۰۰۳	۰/۰۰۱۱۰	۰/۴۸۴۱	۰/۲۳۳۴	۰/۰۱۷۴	۰/۶۳۳۴	—	—	—	—	۱/۹۴۱۹ A_5
۲۹۸۳/۱۹۵	۱/۰۰۰۷	۰/۰۰۰۸۸	۰/۴۰۲۸	۰/۳۱۳۸	۰/۰۱۷۸	۰/۵۱۲۸	—	—	—	—	۱/۹۵۶۲ A_6
۲۹۱۸/۷۴۹	۱/۰۰۱۰	۰/۰۰۰۹	۰/۳۱۰۴	۰/۲۸۴۶	۰/۰۲۶۴	۰/۶۰۲۴	—	—	—	—	۱/۹۹۴۵ A_7
۲۹۲۸/۶۸۶	۱/۰۰۰۵	۰/۰۰۱۶۸	۰/۲۹۲۳	۰/۳۰۶۹	۰/۰۲۵۳	۰/۵۵۵۸	—	—	—	—	۲/۰۲۹۳ A_8
۲۹۴۵/۹۳۷	۱/۰۰۱۱	۰/۰۰۰۲۹	۰/۳۹۲۷	۰/۲۴۲۵	۰/۰۲۳۹	۰/۵۵۸۱	—	—	—	—	۱/۹۵۳۱ A_9
۲۹۸۴/۸۸۶	۱/۰۰۰۹	۰/۰۰۰۶۵	۰/۳۸۰۶	۰/۳۳۶۲	۰/۰۱۸۷	۰/۵۸۸۸	—	—	—	—	۱/۹۸۰۱ A_{10}
۲۹۴۷/۳۳۸	۱/۰۰۱۰	۰/۰۰۰۴۱	۰/۳۵۸۲	۰/۳۵۰۸	۰/۰۲۶۸	۰/۶۱۹۳	—	—	—	—	۲/۰۲۳۰ A_{11}
۲۹۷۷/۴۰۲	۱/۰۰۰۲	۰/۰۰۰۲۰	۰/۴۵۸۳	۰/۲۸۹۳	۰/۰۲۱۶	۰/۵۹۶۷	—	—	—	—	۱/۹۸۲۰ A_{12}
۲۹۲۶/۵۹۳	۱/۰۰۰۹	۰/۰۰۰۷۶	۰/۴۹۷۷	۰/۲۸۵۴	۰/۰۱۸۴	۰/۵۳۹۰	—	—	—	—	۲/۰۱۶۲ A_{13}
۲۹۲۶/۹۴۰	۱/۰۰۰۲۵	۰/۰۰۰۸۸	۰/۴۹۴۱	۰/۲۵۵۴	۰/۰۱۷۵	۰/۵۹۷۰	—	—	—	—	۱/۹۹۱۹ A_{14}
۲۹۹۴/۵۹۷	۱/۰۰۱۹	۰/۰۰۱۴۲	۰/۴۹۵۱	۰/۲۵۶۶	۰/۰۱۳۷	۰/۵۴۲۷	—	—	—	—	۲/۰۲۷۸ A_{15}
۲۹۷۵/۷۲۱	۱/۰۰۱۳	۰/۰۰۰۵۷	۰/۴۱۴۵	۰/۳۲۲۶	۰/۰۱۷۳	۰/۶۴۷۷	—	—	—	—	۲/۰۴۱۳ A_{16}
۲۹۶۸/۹۳۳	۱/۰۰۱۶	۰/۰۰۰۹۴	۰/۳۶۲۳	۰/۳۰۰۸	۰/۰۲۴۵	۰/۵۵۳۲	—	—	—	—	۲/۰۰۲۰ A_{17}
۳۰۱۳/۹۱۲	۱/۰۰۰۹	۰/۰۰۱۰۰	۰/۳۵۲۳	۰/۳۳۸۷	۰/۰۲۳۵	۰/۶۲۷۹	—	—	—	—	۲/۰۰۳۹ A_{18}
۲۹۲۹/۳۰۲	۱/۰۰۰۹	۰/۰۰۰۳۷	۰/۴۲۶۰	۰/۲۷۴۷	۰/۰۲۶۱	۰/۵۹۲۲	—	—	—	—	۱/۹۵۹۸ A_{19}
۲۹۶۵/۹۹۹	۱/۰۰۰۸	۰/۰۰۰۷۰	۰/۳۹۲۹	۰/۲۷۳۰	۰/۰۲۵۸	۰/۵۹۶۱	—	—	—	—	۲/۰۰۲۸ A_{20}
۲۹۲۳/۴۵۱	۱/۰۰۰۹	۰/۰۰۰۳۵	۰/۱۲۲۱	۰/۳۴۴۱	۰/۰۳۴۶	۰/۵۹۷۲	—	—	—	—	۲/۰۴۳۳ A_{21}
۲۹۰۵/۹۳۰	۱/۰۰۰۴	۰/۰۰۲۳۴	۰/۳۱۲۸	۰/۳۱۲۷	۰/۰۲۳۲	۰/۶۱۰۲	—	—	—	—	۱/۹۷۰۳ A_{22}
۲۹۱۹/۱۰۷	۱/۰۰۲۳	۰/۰۰۲۰۲	۰/۵۱۵۵	۰/۳۱۷۳	۰/۰۲۰۲	۰/۶۳۳۸	—	—	—	—	۱/۹۱۵۷ A_{23}
۲۸۸۶/۹۸۸	۰/۹۹۹۴	۰/۰۰۲۰۹	۰/۵۱۴۸	۰/۳۰۱۷	۰/۰۱۴۶	۰/۶۰۷۷	—	—	—	—	۱/۹۶۰۹ A_{24}
۲۹۲۸/۰۲۸	۱/۰۰۰۴	۰/۰۰۰۱۴	۰/۵۰۵۹	۰/۲۵۱۱	۰/۰۱۵۳	۰/۶۱۷۰	—	—	—	—	۲/۰۲۸۴ A_{25}
۲۸۵۹/۲۶۴	۱/۰۰۱۳	۰/۰۰۱۱۶	۰/۳۸۹۰	۰/۳۰۰۱	۰/۰۱۸۸	۰/۵۷۹۴	—	—	—	—	۱/۹۹۷۲ A_{26}
۲۸۸۴/۹۱۹	۱/۰۰۱۰	۰/۰۰۰۶۰	۰/۵۷۹۹	۰/۳۲۰۷	۰/۰۱۹۵	۰/۶۲۶۸	—	—	—	—	۲/۰۶۱۹ A_{27}
—	۰/۰۰۰۰	۰/۰۰۰۱	۰/۰۰۶۷	۰/۰۰۴۱	۰/۰۰۰۰	۰/۰۰۱۲	۰/۰۰۰۰	۰/۰۰۰۰	۰/۰۰۰۰	۰/۰۰۲۱	MSE

در ادامه این مجموعه از داده‌های را به‌عنوان یک نمونه به اندازه ۱۲۹۰ در نظر گرفته و با استفاده از الگوریتم تقویت و روش گام به گام به انتخاب متغیر پرداخته و در نهایت مدل

$$\begin{aligned}
 y_j &= \mathbf{x}_j \mathbf{B} + v + u_j + e_j \\
 u_j &= \rho u_{j-1} + \sigma_{j|j-1} \epsilon_j, \\
 \sigma_{j|j-1}^2 &= w + \alpha y_{j-1}^2 + \beta \sigma_{j-1|j-2}^2 \\
 |\rho| &< 1, \quad j = 1, \dots, 1290
 \end{aligned} \tag{۱۲}$$

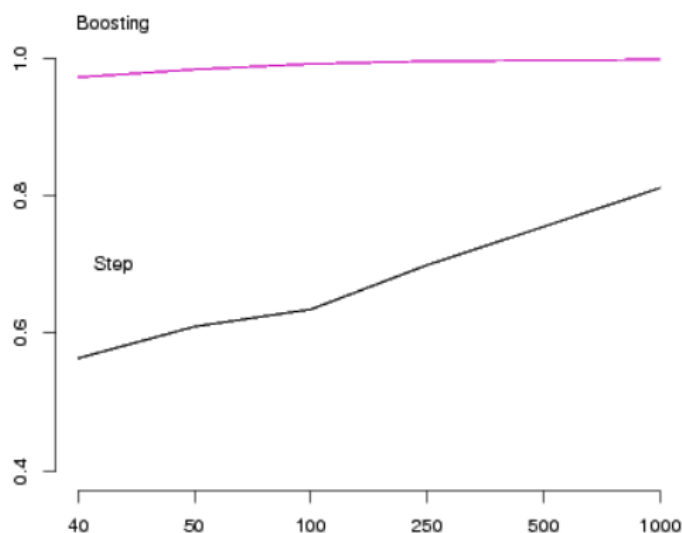
را بر روی داده‌ها برازش می‌دهیم که در آن

$$\mathbf{x}_j = c(LNTPI_j, MS_j, HH_j, DEPEND_j, AGE_j, MR_j, EMP_j, PREP_j)$$

و $\mathbf{B} = (B_1, B_2, B_3, B_4, B_5, B_6, B_7, B_8)$ است. نتایج حاصل از برآورد پارامتر و انتخاب متغیر با استفاده از الگوریتم تقویت و روش گام به گام در جدول ۱۶ ارائه شده است. از این جدول مشاهده می‌شود که متغیرهای

جدول ۷. برازش مدل با استفاده از روش گام به گام وقتی که $T = ۴۰$.

AIC	$\hat{\sigma}_L^2$	$\hat{\mu}_L^2$	$\hat{\beta}$	$\hat{\alpha}$	$\hat{w}$	$\hat{\rho}$	$\hat{B}_+$	$\hat{B}_+$	$\hat{B}_+$	$\hat{B}_+$	Area
۱۲۱/۹۱۴۲	۱/۰۳۳۸	-۰/۰۰۰۵	۰/۱۵۶۴	۰/۱۰۲۴	۰/۲۷۵۸	۰/۴۸۸۲	-	-	-۰/۱۸۲۲	۲/۲۶۰۹	A_1
۱۲۲/۶۹۶۹	۱/۰۲۹۲	۰/۰۱۰۰	۰/۲۵۱۱	۰/۱۳۱۵	۰/۰۴۷۳	۰/۷۵۸۱	-	-	-	۲/۰۵۴۳	A_2
۱۳۳/۰۰۶۸	۱/۰۴۰۶	-۰/۰۰۷۲	۰/۲۸۳۲	۰/۱۰۸۶	۰/۰۰۷۳	۰/۴۶۳۱	-	-	-	۲/۰۸۲۷	A_3
۱۱۲/۸۱۹۷	۱/۰۲۷۶	۰/۰۱۲۳	۰/۱۵۰۷	۰/۱۸۴۵	۰/۰۹۹۲	۰/۱۹۹۸	-	-۰/۰۸۷۳	-	۱/۹۹۲۷	A_4
۱۲۳/۷۱۱۳	۱/۰۲۷۲	۰/۰۱۳۹	۰/۱۴۵۳	۰/۲۰۱۰	۰/۰۸۵۸	۰/۴۳۴۵	-	-	-	۱/۷۴۸۸	A_5
۱۱۹/۱۴۴۰	۱/۰۲۳۵	۰/۰۰۵۲	۰/۱۷۷۶	۰/۴۵۱۸	۰/۰۸۸۷	۰/۰۶۳۸	-	-	۰/۱۴۸۸	۲/۰۶۶۵	A_6
۱۱۶/۰۷۸۷	۱/۰۲۶۱	-۰/۰۰۲۹۱	۰/۱۱۲۴	۰/۱۳۹۵	۰/۰۷۰۹	۰/۶۴۴۶	-	-	-	۲/۰۶۶۱	A_7
۱۲۸/۷۵۸۲	۱/۰۵۰۱	۰/۰۱۲۷	۰/۱۱۰۲	۰/۳۵۷۰	۰/۰۵۷۸	۰/۶۵۷۸	-	-	-	۱/۹۹۲۵	A_8
۱۲۲/۸۵۵۸	۱/۰۳۳۸	-۰/۰۰۲۷۲	۰/۱۵۶۳	۰/۱۳۰۲	۰/۰۶۰۷	۰/۴۴۸۱	-	-	-	۱/۸۳۷۱	A_9
۱۲۳/۷۹۳۷	۱/۰۴۲۱	۰/۰۰۴۴	۰/۱۹۹۹	۰/۱۲۵۵	۰/۰۹۷۲	۰/۳۰۶۷	-	-۰/۰۸۴۱	-	۲/۰۷۵۳	A_{10}
۱۳۰/۹۰۷۴	۱/۰۳۱۵	۰/۰۱۵۹	۰/۱۵۰۸	۰/۱۸۸۹	۰/۲۲۸۵	۰/۲۵۴۷	-	-۰/۱۱۱۹	-	۲/۰۳۹۶	A_{11}
۱۲۰/۴۲۰۹	۱/۰۰۴۸	۰/۰۱۶۳	۰/۴۵۶۹	۰/۱۴۸۱	۰/۰۲۱۶	۰/۳۹۲۱	-	-	-	۱/۹۷۴۸	A_{12}
۱۰۹/۴۷۷۹	۱/۰۰۵۶	-۰/۰۰۲۶۱	۰/۵۹۵۲	۰/۱۴۶۹	۰/۰۵۴۶	۰/۳۶۵۲	-	-	-	۱/۷۷۹۵	A_{13}
۱۰۹/۸۳۴۱	۱/۰۲۷۷	-۰/۰۰۶۰۹	۰/۱۴۲۸	۰/۱۷۴۱	۰/۱۱۷۳	-۰/۰۱۹۲	۱/۱۳۸۷	-	-۰/۱۹۵۱	۲/۱۰۸۰	A_{14}
۱۱۶/۹۰۹۵	۱/۰۳۰۹	-۰/۰۰۳۵۴	۰/۱۵۰۳	۰/۱۵۴۲	۰/۱۸۹۱	۰/۴۳۲۳	-۱/۰۳۹۴	-	-	۲/۲۹۸۳	A_{15}
۱۲۳/۸۰۹۶	۱/۰۵۷۵	۰/۰۰۵۵	۰/۱۵۶۳	۰/۱۲۵۰	۰/۱۸۵۱	۰/۳۰۳۳	۱/۲۳۵۴	-۰/۰۹۲۹	۰/۲۳۷۷	۲/۱۲۷۶	A_{16}
۱۱۲/۷۰۸۳	۱/۰۲۹۳	۰/۰۱۲۹	۰/۱۵۲۴	۰/۱۷۴۹	۰/۱۶۶۳	۰/۲۸۸۴	-	۰/۰۹۱۳	-	۱/۷۷۲۳	A_{17}
۱۳۴/۲۳۸۱	۱/۰۳۷۶	-۰/۰۰۴۶۲	۰/۱۷۸۴	۰/۱۰۵۶	۰/۱۲۶۵	۰/۳۸۴۴	-	-۰/۰۸۳۱	-	۱/۹۸۳۳	A_{18}
۱۲۴/۹۱۱۲	۱/۰۲۷۱	۰/۰۱۸۱	۰/۱۵۰۶	۰/۱۸۲۳	۰/۰۵۷۵	۰/۲۴۹۴	-	-	-	۲/۱۸۷۷	A_{19}
۱۱۷/۲۱۰۸	۱/۰۴۴۵	-۰/۰۰۹۸۷	۰/۱۰۸۶	۰/۳۱۴۹	۰/۰۶۹۰	۰/۳۶۰۷	-	-	-	۱/۸۵۸۲	A_{20}
۱۲۷/۳۶۹۴	۱/۰۲۹۷	۰/۰۰۲۸	۰/۱۵۰۳	۰/۱۰۶۳	۰/۱۹۵۶	۰/۳۴۳۰	-	-	-	۲/۳۹۳۰	A_{21}
۱۲۰/۵۹۲۶	۱/۰۴۱۷	۰/۰۰۹۹	۰/۲۲۱۹	۰/۱۳۶۴	۰/۰۴۶۳	۰/۲۹۰۵	-	-	-	۲/۰۵۹۴	A_{22}
۱۱۵/۶۱۴۴	۱/۰۲۹۳	۰/۰۱۶۴	۰/۲۸۹۱	۰/۴۰۴۱	۰/۰۲۱۸	۰/۶۵۰۵	-	-	-	۱/۹۷۱۷	A_{23}
۱۰۴/۰۰۸۵	۱/۰۲۸۳	-۰/۰۰۲۶۰	۰/۱۴۹۰	۰/۲۱۸۵	۰/۱۲۷۹	۰/۲۵۳۲	۱/۲۵۲۹	-	-	۲/۱۳۱۷	A_{24}
۱۰۸/۲۷۰۸	۱/۰۲۸۴	-۰/۰۰۱۴۴	۰/۱۵۰۴	۰/۱۲۹۰	۰/۱۶۶۷	۰/۱۵۰۲	-	-	-	۲/۳۸۴۴	A_{25}
۱۱۷/۱۲۸۸	۱/۰۲۹۳	۰/۰۱۷۱	۰/۱۴۷۵	۰/۱۳۴۱	۰/۰۹۲۴	۰/۳۹۳۸	-	-	-	۱/۷۶۶۲	A_{26}
۱۲۴/۴۵۶۱	۱/۰۳۰۱	۰/۰۰۲۵۳	۰/۱۴۹۷	۰/۱۷۱۶	۰/۱۶۸۸	۰/۲۰۴۸	-	-	-	۱/۵۸۸۲	A_{27}
-	۰/۰۰۱۱	۰/۰۰۲۵	۰/۱۱۰۰	۰/۰۶۰۵	۰/۰۱۷۶	۰/۱۴۰۸	۰/۰۰۰۰	۰/۰۰۰۰	۰/۰۰۰۰	۰/۰۵۴۰	MSE



شکل ۱. نمودار فراوانی نسبی تشخیص مدل درست با استفاده از الگوریتم تقویت و روش گام به گام برای اندازه نمونه‌های مختلف.

جدول ۸. برازش مدل با استفاده از روش گام به گام وقتی که $T = 50$.

AIC	$\hat{\sigma}_v^2$	$\hat{\mu}_v^2$	$\hat{\beta}$	$\hat{\alpha}$	$\hat{w}$	$\hat{\rho}$	$\hat{B}_+$	$\hat{B}_+$	$\hat{B}_+$	$\hat{B}_+$	Area
۱۲۹/۰۴۸۲	۱/۰۳۰۶	۰/۰۵۷۲	۰/۷۳۵۱	۰/۱۶۷۲	۰/۰۱۴۲	۰/۳۶۳۶	۰/۹۰۴۹				A_1
۱۲۱/۶۳۵۲	۱/۰۱۳۴	۰/۰۳۰۶	۰/۴۹۶۸	۰/۲۰۹۸	۰/۰۴۷۸	۰/۳۰۰۲			۰/۱۴۷۸		A_2
۱۵۰/۸۶۹۳	۱/۰۲۱۲	۰/۰۲۵۸	۰/۰۵۲۷	۰/۱۹۸۱	۰/۰۵۱۸	۰/۵۸۸۳					A_3
۱۵۱/۵۰۱۹	۱/۰۲۵۰	۰/۰۰۷۷	۰/۰۰۰۰	۰/۱۷۱۱	۰/۰۵۱۲	۰/۳۵۷۴					A_4
۱۴۶/۱۷۲۲	۱/۰۲۴۰	۰/۰۰۶۲	۰/۰۰۰۰	۰/۱۱۰۵	۰/۱۱۳۹	۰/۳۲۵۷			۰/۱۶۶۷		A_5
۱۴۹/۰۶۷۸	۱/۰۲۲۹	۰/۰۰۶۹	۰/۰۴۹۷	۰/۱۷۹۶	۰/۰۸۷۲	۰/۳۳۰۵					A_6
۱۵۰/۰۶۷۶	۰/۰۹۷۳	۰/۰۵۴۰	۰/۴۶۷۸	۰/۳۶۴۱	۰/۰۱۴۴	۰/۳۸۰۱					A_7
۱۴۷/۶۸۷۲	۱/۰۳۰۵	۰/۰۰۵۶	۰/۰۵۲۴	۰/۲۰۱۸	۰/۱۳۰۱	۰/۲۶۷۵					A_8
۱۶۲/۴۱۹۹	۱/۰۲۶۱	۰/۰۰۴۳	۰/۰۵۵۳	۰/۲۲۷۰	۰/۱۷۲۳	۰/۱۴۶۶	۰/۲۸۵۲				A_9
۱۵۸/۳۴۴۸	۱/۰۲۱۱	۰/۰۰۵۶	۰/۱۵۵۱	۰/۳۴۲۹	۰/۰۵۱۷	۰/۶۶۴۳					A_{10}
۱۶۳/۲۶۶۴	۱/۰۵۵۶	۰/۰۴۱۰	۰/۷۸۲۶	۰/۰۱۷۲۹	۰/۰۲۲۴	۰/۱۴۷۹					A_{11}
۱۶۱/۲۷۰۳	۱/۰۲۲۶	۰/۰۱۰۲	۰/۰۴۸۵	۰/۱۸۰۷	۰/۱۸۷۰	۰/۲۷۹۵	۱/۰۱۳۹				A_{12}
۱۴۱/۲۰۲۸	۱/۰۱۸۶	۰/۰۴۲۵	۰/۱۷۲۶	۰/۴۶۵۷	۰/۰۲۰۵	۰/۶۲۴۱					A_{13}
۱۵۷/۰۳۷۸	۱/۰۲۸۱	۰/۰۲۷۰	۰/۶۱۸۶	۰/۱۸۰۹	۰/۰۲۸۸	۰/۲۷۴۴	۰/۱۵۵۵				A_{14}
۱۶۲/۹۹۰۳	۱/۰۲۱۲	۰/۰۱۹۵	۰/۰۵۲۸	۰/۲۱۲۴	۰/۰۴۰۹	۰/۳۸۰۳					A_{15}
۱۴۹/۲۸۳۸	۱/۰۳۷۸	۰/۰۳۹۷	۰/۰۰۰۰	۰/۲۰۱۴	۰/۰۶۳۹	۰/۳۲۵۴					A_{16}
۱۶۴/۴۳۸۵	۱/۰۱۹۳	۰/۰۰۲۴	۰/۲۲۱۶	۰/۲۴۱۴	۰/۰۰۶۰	۰/۲۹۱۴					A_{17}
۱۴۵/۳۰۹۶	۱/۰۴۵۹	۰/۰۱۰۳	۰/۰۴۳۷	۰/۱۹۷۵	۰/۱۳۹۶	۰/۳۴۷۸					A_{18}
۱۵۱/۱۳۰۵	۱/۰۲۲۸	۰/۰۱۳۵	۰/۰۰۰۰	۰/۱۹۲۹	۰/۱۲۳۷	۰/۲۵۵۵	۰/۸۸۷۰		۰/۱۶۵۵		A_{19}
۱۴۰/۳۲۸۸	۱/۰۲۰۲	۰/۰۵۸۱	۰/۰۴۵۹	۰/۱۸۰۰	۰/۱۲۳۳	۰/۲۱۷۹	۰/۱۰۰۴۱				A_{20}
۱۴۴/۱۲۳۸	۱/۰۲۶۶	۰/۰۰۸۲	۰/۰۰۰۰	۰/۲۴۹۶	۰/۱۶۲۹	۰/۲۱۹۵	۱/۱۳۱۹				A_{21}
۱۲۵/۲۶۹۵	۱/۰۲۱۹	۰/۰۰۳۲	۰/۰۴۹۱	۰/۱۹۳۶	۰/۰۵۴۴	۰/۵۳۲۱					A_{22}
۱۵۳/۷۹۳۱	۱/۰۰۹۰	۰/۰۲۷۸	۰/۴۴۸۳	۰/۳۱۶۰	۰/۰۱۵۶	۰/۵۷۲۰					A_{23}
۱۶۱/۳۹۳۶	۱/۰۲۱۳	۰/۰۰۱۵	۰/۰۵۰۸	۰/۰۴۰۳	۰/۰۹۶۸	۰/۴۵۱۰			۰/۱۲۹۱		A_{24}
۱۴۲/۰۴۰۶	۱/۰۲۱۶	۰/۰۰۸۶	۰/۲۰۷۷	۰/۱۷۲۶	۰/۰۷۶۶	۰/۳۰۲۱	۰/۹۸۴۵				A_{25}
۱۸۳/۲۶۷۶	۱/۰۲۳۸	۰/۰۲۵۳	۰/۱۰۴۱	۰/۲۹۲۱	۰/۰۴۰۵	۰/۴۲۱۵					A_{26}
۱۶۱/۰۵۹۸	۱/۰۲۱۰	۰/۰۰۹۹	۰/۰۰۰۰	۰/۲۱۸۸	۰/۰۷۵۴	۰/۲۵۵۸					A_{27}
—	۰/۰۰۰۵	۰/۰۰۰۶	۰/۱۰۶۰	۰/۰۰۰۲	۰/۰۰۸۹	۰/۰۰۹۵	۰/۰۰۰۰	۰/۰۰۰۰	۰/۰۰۰۰	۰/۰۲۳۶	MSE

توضیحی $LNTPI$ ، MR و EMP با استفاده از الگوریتم تقویت و متغیرهای توضیحی $LNTPI$ ، HH ، MR ، HH ، MS ، HH ، MR و EMP با استفاده از روش گام به گام انتخاب شده‌اند. اگر چه مدل‌های برازش داده شده بر اساس الگوریتم تقویت و روش گام به گام دارای MSE یکسانی هستند اما مدل برازش داده شده بر اساس الگوریتم تقویت دارای AIC کمتری است.

به‌منظور بررسی مناسب بودن مدل $GARCH(1, 1) - AR(1)$ جهت برازش بر u_j ، از آنجایی که p -مقدار آزمون تی، $0/9998$ ، بزرگتر از $0/05$ است، بنابراین فرضیه صفر بودن میانگین سری رد نمی‌شود. همچنین با توجه به p -مقدار آزمون لیونگ-باکس برای سری زمانی u_j ، $0/0027$ ، می‌توان دریافت که همبستگی بین مقادیر سری وجود دارد که این همبستگی می‌تواند توسط مدل‌های کلاسیک مانند مدل خودبازگشتی توصیف شود. همچنین با توجه به p -مقدار آزمون لیونگ-باکس برای قدرمطلق سری زمانی $|u_j|$ ، $0/0493$ ، می‌توان دریافت که همبستگی بین مقادیر قدرمطلق سری وجود دارد که نشان دهنده نیاز به مدل‌بندی تلاطم است.

نمودارهای سری زمانی و خودهمبستگی سری زمانی u_j و نمودارهای خودهمبستگی و خودهمبستگی جزئی قدرمطلق سری زمانی مانده‌های استاندارد شده مدل $GARCH(1, 1) - AR(1)$ برازش داده شده در شکل ۴ رسم شده است. با توجه به نمودار سری زمانی مشاهده می‌شود که سری رفتار تصادفی حول میانگین صفر را دارد و با توجه به نمودار خودهمبستگی آن می‌توان دریافت که بین مقادیر سری همبستگی وجود ندارد و مدل $AR(1)$ مدل مناسبی جهت برازش به میانگین سری بوده است. با توجه به اینکه مقادیر خودهمبستگی و خودهمبستگی جزئی

جدول ۹. برازش مدل با استفاده از روش گام به گام وقتی که $T = ۱۰۰$.

AIC	$\hat{\sigma}_\nu^2$	$\hat{\mu}_\nu^2$	$\hat{\beta}$	$\hat{\alpha}$	$\hat{w}$	$\hat{\rho}$	$\hat{B}_\tau$	$\hat{B}_\tau$	$\hat{B}_\tau$	$\hat{B}_\tau$	Area
۲۹۹/۲۶۴۲	۱/۰۰۹۳	۰/۰۱۸۵	۰/۲۵۴۵	۰/۱۴۷۲	۰/۱۰۲۰	۰/۳۶۷۰	—	—	—	۲/۲۳۴۹	A_1
۳۱۷/۶۶۵۲	۱/۰۱۰۱	۰/۰۱۴۳	۰/۲۱۶۶	۰/۱۵۶۳	۰/۰۶۴۷	۰/۴۸۷۰	—	—	—	۲/۱۴۵۷	A_2
۲۹۳/۰۰۱۹	۱/۰۱۱۷	۰/۰۱۲۱	۰/۲۴۱۵	۰/۲۱۸۲	۰/۰۲۳۶	۰/۶۹۷۲	—	—	—	۱/۹۲۵۹	A_3
۲۹۴/۶۳۸۲	۱/۰۰۷۷	۰/۰۰۰۱	۰/۲۳۵۵	۰/۱۹۰۸	۰/۰۵۴۱	۰/۶۰۰۹	—	—	—	۱/۹۹۶۱	A_4
۲۶۸/۳۱۳۵	۱/۰۰۶۸	۰/۰۲۷۸	۰/۳۸۵۲	۰/۳۷۴۸	۰/۰۳۵۴	۰/۴۵۹۸	—	۰/۰۶۴۰	—	۲/۰۰۴۱	A_5
۲۶۶/۲۷۷۴	۱/۰۱۰۹	۰/۰۰۴۸	۰/۲۵۱۳	۰/۱۹۳۲	۰/۱۰۳۸	۰/۲۵۸۵	—	—	۰/۱۴۵۱	۲/۰۲۷۲	A_6
۲۸۳/۵۲۹۵	۱/۰۱۳۲	۰/۰۰۷۶	۰/۲۷۴۵	۰/۲۰۲۰	۰/۰۶۸۷	۰/۵۰۵۹	—	—	—	۱/۸۳۶۳	A_7
۲۹۴/۶۰۲۸	۱/۰۱۷۰	۰/۰۰۰۳	۰/۲۵۴۱	۰/۱۶۹۸	۰/۰۵۵۰	۰/۵۸۷۳	—	—	—	۲/۰۸۱۱	A_8
۲۷۷/۲۸۸۱	۱/۰۱۱۳	۰/۰۰۰۱	۰/۲۴۹۱	۰/۱۸۵۳	۰/۱۳۹۸	۰/۴۱۷۵	—	۰/۰۵۶۹	۰/۰۸۸۷	۲/۱۸۵۹	A_9
۲۹۵/۷۳۶۰	۱/۰۰۹۱	۰/۰۰۱۷	۰/۵۷۳۶	۰/۲۴۵۹	۰/۰۰۳۳	۰/۶۹۸۲	—	—	—	۲/۰۲۵۴	A_{10}
۳۱۸/۱۸۸۴	۱/۰۰۹۸	۰/۰۰۴۱	۰/۱۹۶۹	۰/۲۳۰۳	۰/۱۲۵۵	۰/۴۰۹۴	—	۰/۰۶۴۰	—	۲/۱۵۵۴	A_{11}
۲۹۱/۰۷۴۳	۱/۰۱۲۹	۰/۰۰۰۹	۰/۲۵۲۰	۰/۱۶۷۴	۰/۱۱۹۶	۰/۳۷۰۹	۰/۵۶۱۰	—	۰/۰۸۰۴	۱/۸۱۲۶	A_{12}
۲۹۱/۵۸۵۷	۱/۰۱۰۳	۰/۰۱۱۹	۰/۲۶۰۷	۰/۱۹۰۲	۰/۰۸۲۵	۰/۵۵۱۲	۰/۶۸۶۵	—	—	۲/۰۱۵۰	A_{13}
۲۸۳/۷۰۱۶	۱/۰۱۲۴	۰/۰۰۳۵	۰/۵۰۶۹	۰/۲۳۰۷	۰/۰۱۹۹	۰/۵۲۵۸	—	—	—	۱/۹۴۰۸	A_{14}
۲۸۲/۴۸۶۳	۱/۰۰۰۷	۰/۰۱۳۹	۰/۴۳۳۸	۰/۲۳۵۸	۰/۰۴۷۹	۰/۴۸۲۸	—	—	—	۲/۱۶۵۰	A_{15}
۳۱۲/۱۷۴۵	۱/۰۱۰۱	۰/۰۱۱۸	۰/۲۵۳۸	۰/۱۷۸۹	۰/۰۷۷۴	۰/۳۶۳۰	—	۰/۰۵۶۱	—	۱/۹۳۷۸	A_{16}
۳۰۳/۱۹۸۷	۱/۰۰۴۵	۰/۰۲۲۷	۰/۵۴۹۸	۰/۲۵۷۳	۰/۰۰۳۶	۰/۴۸۵۳	—	۰/۰۶۱۹	—	۱/۹۴۱۲	A_{17}
۳۱۰/۸۹۴۳	۱/۰۰۳۷	۰/۰۷۱۹	۰/۱۳۱۵	۰/۴۰۳۱	۰/۰۵۱۳	۰/۴۲۵۴	—	—	—	۱/۸۰۱۶	A_{18}
۳۱۰/۶۱۹۴	۱/۰۰۵۱	۰/۰۱۹۹	۰/۴۹۰۶	۰/۲۲۹۷	۰/۰۲۴۶	۰/۷۵۰۰	—	—	—	۲/۰۰۲۱	A_{19}
۲۹۱/۹۸۶۹	۱/۰۰۲۰	۰/۰۱۰۵	۰/۳۴۳۶	۰/۲۵۳۷	۰/۰۲۲۵	۰/۵۹۶۹	—	—	—	۲/۰۴۷۳	A_{20}
۳۱۳/۹۵۳۰	۱/۰۱۸۴	۰/۰۰۹۴	۰/۲۸۱۲	۰/۳۵۳۸	۰/۰۳۵۹	۰/۵۶۶۶	—	—	—	۱/۹۷۸۲	A_{21}
۳۰۸/۷۱۸۵	۱/۰۲۷۲	۰/۰۱۱۱	۰/۵۳۳۹	۰/۲۰۲۵	۰/۰۰۸۸	۰/۴۶۱۱	—	—	—	۱/۸۶۸۱	A_{22}
۳۲۴/۶۸۵۸	۱/۰۱۰۱	۰/۰۰۹۳	۰/۲۷۸۱	۰/۲۴۰۲	۰/۰۸۹۱	۰/۶۳۰۴	—	—	—	۲/۰۹۲۸	A_{23}
۲۶۷/۹۸۷۷	۱/۰۰۵۶	۰/۰۰۲۸	۰/۲۵۰۲	۰/۱۹۴۵	۰/۰۶۱۴	۰/۵۷۷۵	—	—	—	۲/۰۸۹۷	A_{24}
۲۷۳/۱۶۲۴	۱/۰۰۹۴	۰/۰۰۴۶	۰/۴۷۹۴	۰/۱۶۸۵	۰/۰۴۶۸	۰/۶۷۰۹	—	—	—	۲/۱۷۰۸	A_{25}
۲۹۰/۸۹۴۰	۱/۰۱۱۱	۰/۰۰۰۶	۰/۲۵۰۸	۰/۱۷۵۲	۰/۱۰۱۱	۰/۳۵۸۷	—	—	۰/۱۲۱۵	۲/۰۸۰۵	A_{26}
۳۰۵/۹۳۰۷	۱/۰۰۲۰	۰/۰۳۳۳	۰/۴۴۲۶	۰/۲۳۱۶	۰/۰۵۱۸	۰/۳۳۳۰	۰/۶۳۲۳	۰/۰۶۲۴	۰/۱۲۰۲	۱/۹۸۹۱	A_{27}
—	۰/۰۰۰۲	۰/۰۰۰۸	۰/۰۹۱۶	۰/۰۳۷۶	۰/۰۰۳۵	۰/۰۳۲۳	۰/۰۰۰۰	۰/۰۰۰۰	۰/۰۰۰۰	۰/۰۰۶۲	MSE

در تمام تأخیرها داخل باند اطمینان قرار دارند می‌توان دریافت که مدل $GARCH(1, 1)$ برازش داده شده مدل مناسبی جهت توصیف تلاطم سری بوده است.

بحث و نتیجه‌گیری

مسئله اصلی در مدل‌بندی کردن داده‌ها، انتخاب متغیر و برآورد پارامترها است به‌گونه‌ای که مدل انتخاب شده دقت مناسبی در پیش‌بینی داده‌های جدید و همچنین کمترین خطای پیش‌بینی را داشته باشد. با افزایش ابعاد متغیرها، تعداد پارامترها در مدل نیز رشد می‌کند. این می‌تواند برآورد پارامترهای مدل را برای مدل‌ها با ابعاد بالا پیچیده‌تر کند. در این مقاله به‌منظور برآورد و انتخاب متغیر در مدل‌های آمیخته با اثرات تصادفی برای کوچک نواحی از الگوریتم‌های یادگیری ماشین از جمله الگوریتم تقویت استفاده شده است. عملکرد الگوریتم تقویت و روش گام به گام در انتخاب متغیر با استفاده از داده‌های شبیه‌سازی شده و داده‌های بدهی‌های مالیاتی مورد مطالعه قرار گرفت. نتایج حاصل نشان می‌دهد که الگوریتم تقویت عملکرد مناسب‌تری در انتخاب متغیر در کوچک نواحی دارد به‌طوری که در برخی از نواحی روش گام به گام توانایی تشخیص مدل درست را ندارد. همچنین فراوانی نسبی تشخیص مدل درست با استفاده از الگوریتم تقویت به یک همگرا است.

جدول ۱۰. برازش مدل با استفاده از روش گام به گام وقتی که $T = ۲۵۰$.

AIC	$\hat{\sigma}_\nu^2$	$\hat{\mu}_\nu^2$	$\hat{\beta}$	$\hat{\alpha}$	$\hat{w}$	$\hat{\rho}$	$\hat{B}_\tau$	$\hat{B}_\tau$	$\hat{B}_\tau$	$\hat{B}_\tau$	Area
۷۵۳/۲۷۲۳	۱/۰۰۳۴	۰/۰۱۰۴	۰/۲۶۴۲	۰/۴۵۴۲	۰/۰۱۹۹	۰/۵۲۰۴	—	—	—	—	۲/۰۱۹۷ A_1
۷۶۴/۰۷۸۸	۱/۰۰۱۲	—/۰۰۴۷۸	۰/۲۹۲۲	۰/۴۳۲۴	۰/۰۳۶۹	۰/۶۳۶۴	—	—	—	—	۲/۰۰۱۱ A_2
۷۵۷/۳۰۹۲	۱/۰۰۳۹	۰/۰۲۰۱	۰/۳۲۷۸	۰/۲۹۰۷	۰/۰۵۱۹	۰/۶۳۷۹	—	—/۰۰۴۲۹	—	—	۱/۸۳۵۳ A_3
۷۲۷/۵۷۱۵	۱/۰۰۸۸	۰/۰۰۴۱	۰/۵۶۶۵	۰/۱۵۵۱	۰/۰۰۵۸	۰/۴۴۱۸	—	—	—	—	۲/۱۴۴۲ A_4
۷۱۵/۴۰۶۳	۱/۰۰۳۴	۰/۰۱۳۶	۰/۴۷۲۲	۰/۱۴۵۷	۰/۰۲۷۶	۰/۴۷۳۸	۰/۴۴۴۵	—	—	—	۱/۹۴۰۹ A_5
۷۵۶/۲۳۲۰	۱/۰۰۴۶	۰/۰۰۵۰	۰/۴۷۷۳	۰/۲۵۳۱	۰/۰۱۶۴	۰/۷۱۷۲	—	—	—	—	۱/۹۸۷۹ A_6
۷۱۵/۲۲۲۳	۱/۰۰۹۵	۰/۰۰۴۶	۰/۵۲۲۰	۰/۱۹۵۸	۰/۰۰۶۶	۰/۴۴۹۸	—	—	—	—	۲/۰۹۶۰ A_7
۷۵۵/۹۷۵۲	۱/۰۰۳۸	۰/۰۰۲۸	۰/۱۲۴۸	۰/۲۵۴۵	۰/۰۴۶۸	۰/۵۸۰۱	—	—	—	—	۲/۰۷۸۲ A_8
۷۲۱/۰۰۸۰	۱/۰۰۱۹	۰/۰۱۹۴	۰/۵۹۰۵	۰/۱۸۹۷	۰/۰۱۶۴	۰/۵۹۵۸	—	—	—	—	۱/۹۱۱۷ A_9
۶۹۴/۰۹۲۴	۱/۰۰۳۷	۰/۰۰۰۸	۰/۳۲۸۴	۰/۲۵۰۶	۰/۰۲۳۱	۰/۵۴۹۷	—	—	—	—	۱/۹۹۹۳ A_{10}
۷۵۸/۳۹۷۶	۱/۰۰۹۵	۰/۰۰۱۶	۰/۴۹۱۷	۰/۱۸۳۱	۰/۰۲۸۲	۰/۴۲۷۲	—	—	—	—	۱/۸۶۴۵ A_{11}
۷۲۸/۵۸۲۷	۱/۰۰۴۶	—/۰۰۰۳۷	۰/۳۱۵۹	۰/۱۶۴۳	۰/۰۴۸۰	۰/۴۹۹۶	—	—	—	—	۱/۸۵۳۳ A_{12}
۷۱۹/۹۶۰۱	۱/۰۰۳۳	۰/۰۰۶۶	۰/۳۶۳۱	۰/۲۳۰۸	۰/۰۲۴۳	۰/۵۷۷۳	—	—	—	—	۱/۸۹۶۶ A_{13}
۷۲۱/۹۵۶۹	۱/۰۰۴۰	—/۰۰۰۵۳	۰/۲۵۴۳	۰/۱۹۸۲	۰/۰۳۷۹	۰/۷۱۶۵	—	—	—	—	۲/۰۴۵۳ A_{14}
۷۳۸/۲۷۷۰	۱/۰۰۳۴	—/۰۰۰۰۲	۰/۳۰۵۴	۰/۱۴۱۰	۰/۰۳۳۸	۰/۶۱۵۶	—	—	—	—	۲/۰۱۴۵ A_{15}
۷۱۴/۲۲۱۸	۱/۰۰۲۱	—/۰۰۱۵۶	۰/۴۹۲۶	۰/۳۱۸۹	۰/۰۲۶۰	۰/۴۸۰۱	—/۰۳۵۸۶	—	—	—	۱/۹۷۸۳ A_{16}
۷۰۸/۷۴۳۳	۱/۰۰۰۵	—/۰۰۴۵۲	۰/۳۱۴۴	۰/۵۰۵۲	۰/۰۲۱۳	۰/۵۳۱۰	—	—	—	—	۱/۹۲۵۴ A_{17}
۷۴۶/۰۷۷۶	۱/۰۰۶۷	—/۰۰۱۸۴	۰/۲۳۴۸	۰/۱۸۴۵	۰/۰۶۰۹	۰/۵۵۰۰	—	—	—/۰۰۷۸۰	—	۱/۹۴۹۷ A_{18}
۷۰۹/۳۸۱۴	۱/۰۰۰۱	—/۰۰۱۴۱	۰/۵۵۲۶	۰/۱۴۰۸	۰/۰۱۱۲	۰/۴۶۴۲	—/۰۶۳۹۰	—	—	—	۱/۹۹۲۹ A_{19}
۷۷۷/۱۵۱۸	۱/۰۰۲۵	۰/۰۱۰۴	۰/۳۵۰۶	۰/۲۹۷۰	۰/۰۳۶۱	۰/۴۰۵۱	—/۰۵۲۶۳	—	—	—	۲/۰۳۹۹ A_{20}
۷۹۶/۸۲۴۲	۱/۰۰۴۸	—/۰۰۱۰۸	۰/۳۶۳۱	۰/۳۴۸۵	۰/۰۲۵۶	۰/۴۹۱۱	—	—	—	—	۱/۹۴۴۷ A_{21}
۷۵۸/۴۲۲۸	۱/۰۰۳۳	۰/۰۰۲۲	۰/۴۷۲۲	۰/۱۷۰۶	۰/۰۲۵۷	۰/۵۱۵۴	—	—	—	—	۱/۹۲۱۹ A_{22}
۷۴۴/۳۰۵۹	۱/۰۰۱۸	—/۰۰۰۲۳	۰/۵۵۸۰	۰/۱۶۴۸	۰/۰۱۲۲	۰/۳۸۷۷	—	—	—	—	۱/۹۰۲۹ A_{23}
۷۵۷/۰۸۷۶	۱/۰۰۸۲	—/۰۰۱۵۱	۰/۴۸۲۴	۰/۲۲۹۰	۰/۰۲۵۱	۰/۵۵۴۲	—	—/۰۰۴۱۸	—	—	۲/۰۲۰۹ A_{24}
۷۵۲/۹۳۰۶	۱/۰۰۲۴	—/۰۰۰۶۹	۰/۵۶۴۱	۰/۱۷۹۳	۰/۰۳۷۸	۰/۴۴۴۳	—/۰۵۴۵۹	—	—/۰۰۵۷۵	—	۲/۰۶۳۷ A_{25}
۷۳۳/۰۲۷۸	۱/۰۰۵۷	—/۰۰۰۶۰	۰/۴۴۴۶	۰/۲۷۰۹	۰/۰۱۸۵	۰/۶۵۶۰	—	—	—	—	۲/۰۳۹۳ A_{26}
۷۴۷/۶۰۲۷	۱/۰۰۱۴	۰/۰۲۹۸	۰/۴۷۱۲	۰/۲۴۰۵	۰/۰۱۷۷	۰/۷۳۰۷	—	—	—	—	۲/۰۴۷۱ A_{27}
—	۰/۰۰۰۰	۰/۰۰۰۲	۰/۰۷۸۸	۰/۰۲۷۶	۰/۰۰۰۷	۰/۰۱۰۷	۰/۰۰۰۰	۰/۰۰۰۰	۰/۰۰۰۰	۰/۰۰۴۵	MSE

تقدیر و تشکر

نویسندگان از سردبیر محترم و داوران گرامی که پیشنهادات آن‌ها موجب بهبود کیفیت مقاله حاضر گردید؛ کمال تشکر را دارند.

مراجع

زارعی، ش. (۱۴۰۰)، برآورد کوچک ناحیه‌ای بیز تجربی استوار با توزیع α -پایدار متقارن برای مولفه‌های خطا، مجله علوم آماری، ۱۵(۲)، ۴۶۳-۴۸۰.

Audrino, F. and Buhlmann, P. (2016), Volatility Estimation with Functional Gradient Descent for Very High-dimensional Financial Time Series, *The Journal of Computational Finance*, **6**(3), 65-89.

Buhlmann, P. and Hothorn, T. (2007), Boosting Algorithms: Regularization, Prediction and Model Fitting, *Statistical Science*, **22**(4), 477-505.

جدول ۱۱. برازش مدل با استفاده از روش گام به گام وقتی که $T = 500$.

AIC	$\hat{\sigma}_\nu^2$	$\hat{\mu}_\nu^2$	$\hat{\beta}$	$\hat{\alpha}$	$\hat{w}$	$\hat{\rho}$	$\hat{B}_\tau$	$\hat{B}_\tau$	$\hat{B}_\tau$	$\hat{B}_\tau$	Area
۱۴۶۱/۵۵۶	۱/۰۰۲۰	-۰/۰۰۱۷	۰/۲۵۴۳	۰/۱۰۸۸	۰/۰۴۹۱	۰/۵۲۹۲	-	-	-	۲/۰۵۹۱	A_1
۱۴۵۵/۶۸۱	۱/۰۰۳۴	-۰/۰۰۱۱	۰/۲۴۳۲	۰/۳۰۵۰	۰/۰۲۹۰	۰/۵۳۱۰	-	-	-	۲/۰۰۲۶	A_2
۱۴۶۲/۷۷۱	۱/۰۰۱۷	-۰/۰۰۲۲	۰/۳۹۶۹	۰/۲۶۹۸	۰/۰۲۱۷	۰/۵۲۹۱	-	-	-	۱/۹۲۱۴	A_3
۱۴۸۰/۲۴۶	۱/۰۰۱۹	-۰/۰۰۵۴	۰/۴۴۹۸	۰/۲۷۲۸	۰/۰۱۸۴	۰/۶۰۸۵	-	-	-	۱/۹۹۷۶	A_4
۱۴۶۰/۵۲۲	۱/۰۰۱۱	۰/۰۰۲۶	۰/۵۰۹۷	۰/۲۲۵۷	۰/۰۱۲۵	۰/۶۴۰۰	-	-	-	۱/۹۶۴۷	A_5
۱۴۳۵/۸۵۱	۱/۰۰۱۲	۰/۰۰۵۵	۰/۵۳۵۳	۰/۲۲۴۳	۰/۰۱۵۹	۰/۶۲۱۳	-	-	-	۱/۹۴۹۳	A_6
۱۴۵۶/۷۰۱	۰/۹۹۸۰	-۰/۰۰۶۹	۰/۸۵۰۰	۰/۰۶۱۹	۰/۰۰۵۳	۰/۵۹۶۲	-	-	-	۲/۰۴۵۲	A_7
۱۵۰۰/۴۷۶	۱/۰۰۰۲	۰/۰۰۵۹	۰/۴۵۱۶	۰/۳۴۰۷	۰/۰۲۰۰	۰/۶۲۸۹	-	-	-	۲/۰۵۴۷	A_8
۱۴۷۷/۲۵۳	۱/۰۰۱۰	۰/۰۰۱۹	۰/۴۳۹۰	۰/۳۳۶۶	۰/۰۱۹۳	۰/۵۲۰۰	-	-	-	۱/۹۲۷۸	A_9
۱۴۸۱/۶۱۶	۱/۰۰۱۶	-۰/۰۱۰۳	۰/۳۲۸۴	۰/۲۴۲۲	۰/۰۴۵۵	۰/۵۱۲۳	-	-	۰/۰۵۶۵	۱/۹۶۴۳	A_{10}
۱۴۵۲/۵۴۱	۱/۰۰۰۰	۰/۰۰۵۸	۰/۵۶۹۳	۰/۲۲۳۲	۰/۰۱۵۴	۰/۴۶۹۳	-	۰/۰۲۲۵	-	۲/۰۰۸۵	A_{11}
۱۴۷۹/۱۹۶	۱/۰۰۱۶	-۰/۰۱۱۹	۰/۴۳۲۳	۰/۳۲۳۷	۰/۰۳۳۴	۰/۵۹۹۴	-	-	-	۲/۰۷۸۴	A_{12}
۱۴۸۸/۳۷۳	۱/۰۰۱۳	۰/۰۰۷۶	۰/۵۱۰۶	۰/۲۳۲۰	۰/۰۲۲۰	۰/۵۳۵۴	-	-	-	۲/۰۵۸۷	A_{13}
۱۴۲۴/۲۵۶	۱/۰۰۰۹	۰/۰۰۳۰	۰/۴۵۷۸	۰/۲۳۳۸	۰/۰۲۷۶	۰/۵۳۴۲	-۰/۲۹۴۵	-	-	۱/۹۹۶۷	A_{14}
۱۴۳۰/۰۳۹	۱/۰۰۱۰	۰/۰۰۶۱	۰/۳۵۰۱	۰/۳۵۳۱	۰/۰۲۳۵	۰/۵۶۵۹	-	-	-	۲/۰۱۱۱	A_{15}
۱۴۶۵/۰۶۰	۱/۰۰۴۲	-۰/۰۰۲۱	۰/۷۳۹۵	۰/۱۴۲۴	۰/۰۰۸۰	۰/۵۳۱۱	-	-	-	۱/۹۳۳۲	A_{16}
۱۴۵۸/۵۱۱	۱/۰۰۱۱	۰/۰۰۱۴	۰/۵۸۴۵	۰/۱۲۰۴	۰/۰۱۷۱	۰/۳۸۱۵	۰/۳۴۵۰	-	-۰/۰۸۴۶	۲/۰۴۲۴	A_{17}
۱۴۷۲/۲۳۴	۰/۹۹۹۶	-۰/۰۰۹۶	۰/۵۴۰۸	۰/۱۷۰۸	۰/۰۱۳۰	۰/۵۱۴۸	-۰/۲۴۲۴	۰/۰۲۲۷	-	۲/۰۱۱۸	A_{18}
۱۴۹۱/۳۳۵	۱/۰۰۲۱	-۰/۰۰۰۹	۰/۲۳۸۴	۰/۲۱۰۱	۰/۰۳۹۰	۰/۵۸۶۰	-	-	-	۲/۰۴۰۱	A_{19}
۱۵۱۳/۰۹۲	۱/۰۰۱۸	-۰/۰۰۷۸	۰/۴۳۳۶	۰/۲۹۲۲	۰/۰۲۴۲	۰/۴۲۱۴	-	۰/۰۲۶۹	-	۱/۹۶۶۳	A_{20}
۱۵۰۶/۹۰۸	۱/۰۰۱۷	-۰/۰۰۸۲	۰/۳۱۲۹	۰/۴۰۳۵	۰/۰۲۲۷	۰/۵۵۱۳	-	-	-	۱/۹۹۵۸	A_{21}
۱۴۹۹/۳۴۷	۱/۰۰۱۴	-۰/۰۱۱۱	۰/۴۶۰۳	۰/۲۷۱۹	۰/۰۳۷۰	۰/۴۵۰۳	-	-۰/۰۲۵۱	-۰/۰۵۳۷	۲/۰۵۴۵	A_{22}
۱۴۹۵/۳۲۷	۱/۰۰۱۹	-۰/۰۱۰۹	۰/۲۸۷۱	۰/۳۲۷۳	۰/۰۳۰۴	۰/۶۵۸۴	-	-	-	۲/۰۲۳۳	A_{23}
۱۴۶۴/۷۰۰	۱/۰۰۱۷	-۰/۰۰۹۹	۰/۳۴۵۵	۰/۳۸۱۷	۰/۰۱۹۷	۰/۶۱۶۲	-	-	-	۲/۰۲۰۸	A_{24}
۱۴۶۴/۱۶۱	۱/۰۰۲۴	۰/۰۱۴۹	۰/۳۸۴۷	۰/۳۶۳۸	۰/۰۴۰۹	۰/۶۴۶۸	-	-	-	۱/۹۹۲۴	A_{25}
۱۵۰۷/۸۷۵	۱/۰۰۱۷	۰/۰۱۶۴	۰/۱۸۶۲	۰/۲۵۸۹	۰/۰۳۲۳	۰/۶۴۸۱	-	-	-	۲/۰۵۵۲	A_{26}
۱۴۸۴/۲۵۵	۱/۰۰۱۸	۰/۰۰۲۴	۰/۳۲۶۹	۰/۲۸۰۸	۰/۰۲۶۹	۰/۶۲۲۸	-	-	-	۱/۹۷۴۳	A_{27}
-	۰/۰۰۰۰	۰/۰۰۰۲	۰/۰۳۷۴	۰/۰۰۹۹	۰/۰۰۰۳	۰/۰۰۷۷	۰/۰۰۰۰	۰/۰۰۰۰	۰/۰۰۰۰	۰/۰۰۱۶	MSE

Chen, L.P. (2024), Accelerated Failure Time Models with Error-prone Response and Nonlinear Covariates, *Statistics and Computing*, 34,183.

Chen, T. and Guestrin, C. (2016), XGBoost: A Scalable Tree Boosting System, *KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining August*, 785–794
<https://doi.org/10.1145/2939672.2939785>

Chen, L. P. and Qiu, B. (2023), Analysis of Length-biased and Partly Interval-censored Survival Data with Mismeasured Covariates, *Biometrics*, 79, 3929-3940.

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.

Frees, E.W. (2004), Longitudinal and Panel Data, Analysis and Applications in the Social Sciences, Cambridge University Press, New York.

جدول ۱۲. برازش مدل با استفاده از روش گام به گام وقتی که $T = ۱۰۰۰$.

AIC	$\hat{\sigma}_\nu^2$	$\hat{\mu}_\nu^2$	$\hat{\beta}$	$\hat{\alpha}$	$\hat{w}$	$\hat{\rho}$	$\hat{B}_+$	$\hat{B}_-$	$\hat{B}_\tau$	$\hat{B}_\lambda$	Area
۲۹۴۷/۴۰۸	۱/۰۰۱۰	۰/۰۰۴۹	۰/۳۴۷۲	۰/۳۱۵۱	۰/۰۲۳۳	۰/۶۲۶۹	—	—	—	۱/۹۹۳۸	A_1
۲۹۶۵/۰۴۶	۱/۰۰۰۶	۰/۰۰۱۸	۰/۳۶۳۷	۰/۲۷۳۶	۰/۰۲۲۵	۰/۵۸۶۷	—	—	—	۱/۹۶۷۸	A_2
۲۹۱۵/۱۱۲	۱/۰۰۰۵	—/۰۰۰۲۸	۰/۴۱۲۵	۰/۳۰۷۴	۰/۰۲۰۵	۰/۶۳۰۸	—	—	—	۱/۹۸۵۳	A_3
۲۹۷۵/۵۸۶	۱/۰۰۰۱	—/۰۰۱۶۰	۰/۴۰۵۵	۰/۳۲۸۴	۰/۰۱۸۰	۰/۶۱۱۳	—	—	—	۲/۰۲۲۹	A_4
۲۹۲۴/۶۷۵	۱/۰۰۰۳	—/۰۰۱۱۳	۰/۴۹۸۴	۰/۲۱۷۹	۰/۰۱۷۸	۰/۶۲۵۶	—	—	—	۱/۹۳۱۸	A_5
۲۹۸۴/۳۷۳	۱/۰۰۰۵	—/۰۰۱۲۹	۰/۴۶۱۲	۰/۲۶۸۴	۰/۰۱۷۲	۰/۵۰۸۶	۰/۱۹۲۶	—	—	۱/۹۷۱۵	A_6
۲۹۱۹/۹۴۴	۱/۰۰۰۹	۰/۰۰۱۲	۰/۳۴۱۸	۰/۲۴۰۷	۰/۰۲۸۶	۰/۵۸۶۵	—	—	۰/۰۲۹۵	۲/۰۰۱۴	A_7
۲۹۲۹/۰۸۱	۱/۰۰۰۵	—/۰۰۱۷۱	۰/۲۵۶۶	۰/۳۴۱۵	۰/۰۲۵۱	۰/۵۴۰۰	—	—	—	۱/۹۹۲۱	A_8
۲۹۴۶/۴۲۳	۱/۰۰۱۱	۰/۰۰۳۵	۰/۳۸۷۱	۰/۲۵۳۰	۰/۰۲۳۲	۰/۵۴۳۶	—	—	—	۱/۹۶۱۴	A_9
۲۹۸۶/۹۹۹	۱/۰۰۰۹	۰/۰۰۳۷	۰/۴۲۶۰	۰/۲۶۲۸	۰/۰۲۲۱	۰/۵۵۴۶	۰/۲۲۲۵	—	—	۱/۹۵۵۳	A_{10}
۲۹۴۸/۳۶۰	۱/۰۰۱۰	۰/۰۰۳۶	۰/۲۶۳۲	۰/۳۴۲۰	۰/۰۲۷۳	۰/۶۱۶۶	—	—	—	۲/۰۳۳۳	A_{11}
۲۹۷۷/۶۷۶	۱/۰۰۲۲	۰/۰۰۲۴	۰/۴۷۵۶	۰/۲۱۳۷	۰/۰۲۱۹	۰/۵۹۳۰	—	—	—	۱/۹۶۵۸	A_{12}
۲۹۲۸/۶۹۹	۱/۰۰۰۸	—/۰۰۰۷۷	۰/۶۳۵۷	۰/۱۳۳۹	۰/۰۱۴۳	۰/۵۱۸۵	۰/۲۱۷۰	—	—	۱/۹۸۵۸	A_{13}
۲۹۲۹/۴۴۶	۱/۰۰۳۸	۰/۰۰۸۱	۰/۵۳۳۴	۰/۱۹۸۹	۰/۰۱۷۲	۰/۵۸۴۸	—	—/۰۱۶۳	—	۱/۹۷۴۸	A_{14}
۲۹۹۴/۹۷۵	۱/۰۰۲۰	۰/۰۱۴۶	۰/۵۳۳۳	۰/۲۳۹۵	۰/۰۱۳۶	۰/۵۴۴۴	—	—	—	۲/۰۲۱۹	A_{15}
۲۹۷۶/۴۸۸	۱/۰۰۱۲	۰/۰۰۵۶	۰/۵۱۱۷	۰/۲۲۵۸	۰/۰۱۷۲	۰/۴۹۹۶	—	—	—	۲/۰۳۸۰	A_{16}
۲۹۶۹/۲۹۳	۱/۰۰۱۹	—/۰۰۰۹۹	۰/۳۶۳۵	۰/۳۰۲۱	۰/۰۲۴۸	۰/۵۴۱۵	—	—	—	۱/۹۸۲۹	A_{17}
۳۰۱۵/۴۰۷	۱/۰۰۰۹	—/۰۰۰۷۹	۰/۴۷۷۱	۰/۲۵۰۲	۰/۰۲۲۳	۰/۵۹۷۰	—	—/۰۲۷۲	—	۲/۰۱۳۸	A_{18}
۲۹۲۹/۳۲۵	۱/۰۰۰۹	۰/۰۰۴۰	۰/۴۲۷۴	۰/۱۷۶۶	۰/۰۲۵۸	۰/۴۹۳۸	—	—	—	۱/۹۶۳۱	A_{19}
۲۹۶۷/۴۲۶	۱/۰۰۰۷	—/۰۰۰۸۰	۰/۴۵۵۲	۰/۲۵۰۹	۰/۰۲۵۴	۰/۵۶۸۵	—	—	—/۰۲۹۶	۱/۹۹۷۳	A_{20}
۲۹۲۴/۱۰۹	۱/۰۰۱۰	—/۰۰۰۳	۰/۰۴۵۷	۰/۳۲۷۶	۰/۰۴۴۴	۰/۵۶۵۲	۰/۱۸۹۸	—	—	۲/۰۴۰۸	A_{21}
۲۹۰۸/۱۵۲	۱/۰۰۰۲	۰/۰۲۴۷	۰/۳۶۰۰	۰/۲۹۵۰	۰/۰۲۲۹	۰/۵۹۱۱	—	—/۰۱۷۵	—	۱/۹۷۱۴	A_{22}
۲۹۲۲/۹۲۷	۱/۰۰۲۶	۰/۰۱۹۸	۰/۴۹۶۰	۰/۲۶۶۸	۰/۰۱۶۸	۰/۶۴۹۲	—	—	۰/۰۲۸۸	۱/۹۸۶۴	A_{23}
۲۸۸۶/۹۹۳	۰/۹۹۹۴	—/۰۲۰۷	۰/۵۲۹۵	۰/۲۸۷۸	۰/۰۱۴۶	۰/۶۰۳۲	—	—	—	۱/۹۳۳۶	A_{24}
۲۹۳۰/۸۶۴	۱/۰۰۰۶	—/۰۰۰۳۷	۰/۴۶۰۱	۰/۲۵۹۵	۰/۰۱۸۸	۰/۵۹۴۶	—	—	۰/۰۳۸۰	۱/۹۸۵۲	A_{25}
۲۸۵۹/۸۶۵	۱/۰۰۱۱	—/۰۰۱۴۳	۰/۳۸۹۷	۰/۳۷۲۰	۰/۰۱۹۹	۰/۵۷۸۱	—	—	—	۲/۰۳۱۸	A_{26}
۲۸۸۵/۶۸۹	۱/۰۰۰۶	۰/۰۰۶۹	۰/۵۸۹۲	۰/۲۳۳۳	۰/۰۱۹۶	۰/۶۰۴۱	—/۲۲۵۲	—	—	۲/۰۶۲۸	A_{27}
—	۰/۰۰۰۰	۰/۰۰۰۱	۰/۰۰۷۹	۰/۰۰۵۴	۰/۰۰۰۰	۰/۰۰۱۹	۰/۰۰۰۰	۰/۰۰۰۰	۰/۰۰۰۰	۰/۰۰۱۶	MSE

Freund, Y. and Schapire, R. (1997), A Decision Theoretic Generalization of Online Learning and an Application to Boosting, *Journal of Computer and System Sciences*, **55**(1), 119-139.

Friedman, J. H. (2001), Greedy Function Approximation: a Gradient Boosting Machine, *The Annals of Statistics*, **29**, 1189-1232.

Ghosh, M., Nangia, N. and Kim, D.H. (1996), Estimation of Median Income of Four-person Families: a Bayesian Time Series Approach, *Journal of the American Statistical Association*, **91**, 1423-1431.

Jiang, J., Nguyen, T. and Lahiri, P. (2018), A Unified Monte-Carlo Jackknife for Small Area Estimation after Model Selection, *Annals of Mathematical Sciences and Applications*, **3**, 405-438.

Levy, P.S. and French, D.K. (1977), Synthetic Estimation of State Health Charac-

جدول ۱۳. فراوانی نسبی نتایج انتخاب مدل درست با استفاده از الگوریتم تقویت.

Area	۴۰	۵۰	۱۰۰	۲۵۰	۵۰۰	۱۰۰۰
A _۱	۰/۹۹۰	۱/۰۰۰	۱/۰۰۰	۱/۰۰۰	۱/۰۰۰	۱/۰۰۰
A _۲	۰/۹۹۲	۰/۹۹۹	۱/۰۰۰	۱/۰۰۰	۰/۹۹۹	۱/۰۰۰
A _۳	۰/۹۸۹	۰/۹۸۹	۰/۹۹۳	۰/۹۹۷	۰/۹۹۸	۰/۹۹۹
A _۴	۰/۹۷۸	۰/۹۸۵	۰/۹۹۱	۰/۹۹۵	۰/۹۹۸	۰/۹۹۹
A _۵	۰/۹۲۱	۰/۹۳۷	۰/۹۸۴	۰/۹۸۹	۰/۹۹۰	۰/۹۹۵
A _۶	۰/۹۸۳	۱/۰۰۰	۱/۰۰۰	۱/۰۰۰	۰/۹۹۸	۱/۰۰۰
A _۷	۰/۹۷۵	۰/۹۸۵	۱/۰۰۰	۱/۰۰۰	۰/۹۹۹	۰/۹۹۹
A _۸	۰/۹۸۴	۰/۹۹۷	۱/۰۰۰	۱/۰۰۰	۱/۰۰۰	۱/۰۰۰
A _۹	۰/۹۸۶	۰/۹۹۳	۰/۹۹۵	۰/۹۹۷	۰/۹۹۹	۰/۹۹۹
A _{۱۰}	۰/۹۷۸	۱/۰۰۰	۱/۰۰۰	۱/۰۰۰	۱/۰۰۰	۱/۰۰۰
A _{۱۱}	۰/۹۸۶	۰/۹۸۷	۰/۹۹۱	۰/۹۹۳	۰/۹۹۶	۰/۹۹۹
A _{۱۲}	۰/۹۹۱	۰/۹۹۱	۰/۹۹۵	۰/۹۹۶	۰/۹۹۸	۰/۹۹۹
A _{۱۳}	۰/۹۷۸	۰/۹۸۷	۰/۹۶۳	۱/۰۰۰	۱/۰۰۰	۱/۰۰۰
A _{۱۴}	۰/۹۷۳	۰/۹۸۳	۰/۹۹۴	۰/۹۹۷	۰/۹۹۹	۰/۹۹۹
A _{۱۵}	۰/۹۸۰	۰/۹۹۰	۰/۹۹۲	۰/۹۹۶	۱/۰۰۰	۱/۰۰۰
A _{۱۶}	۰/۹۶۶	۰/۹۷۲	۰/۹۸۱	۱/۰۰۰	۱/۰۰۰	۱/۰۰۰
A _{۱۷}	۰/۹۳۷	۰/۹۷۳	۱/۰۰۰	۱/۰۰۰	۱/۰۰۰	۱/۰۰۰
A _{۱۸}	۰/۹۵۹	۰/۹۶۷	۰/۹۷۸	۰/۹۸۲	۰/۹۸۹	۱/۰۰۰
A _{۱۹}	۰/۹۶۴	۰/۹۸۴	۰/۹۸۵	۰/۹۹۱	۰/۹۹۴	۰/۹۹۸
A _{۲۰}	۰/۹۸۱	۰/۹۸۳	۱/۰۰۰	۱/۰۰۰	۰/۹۹۹	۰/۹۹۹
A _{۲۱}	۰/۹۸۷	۱/۰۰۰	۱/۰۰۰	۱/۰۰۰	۱/۰۰۰	۰/۹۹۹
A _{۲۲}	۰/۹۷۲	۰/۹۷۹	۰/۹۸۷	۰/۹۹۳	۱/۰۰۰	۱/۰۰۰
A _{۲۳}	۰/۹۸۱	۰/۹۹۲	۰/۹۹۸	۰/۹۹۸	۰/۹۹۹	۰/۹۹۹
A _{۲۴}	۰/۹۹۰	۰/۹۹۵	۱/۰۰۰	۱/۰۰۰	۱/۰۰۰	۱/۰۰۰
A _{۲۵}	۰/۹۷۵	۰/۹۸۴	۱/۰۰۰	۱/۰۰۰	۱/۰۰۰	۱/۰۰۰
A _{۲۶}	۰/۹۴۱	۰/۹۵۷	۰/۹۹۷	۱/۰۰۰	۱/۰۰۰	۰/۹۹۹
A _{۲۷}	۰/۹۷۸	۰/۹۹۳	۰/۹۹۸	۰/۹۹۹	۰/۹۹۹	۰/۹۹۹

teristics Based on the Health Interview Vurvey, *Vital and Health Statistics*, **2** (75), 78-1349.

Pfeffermann, D. and Bernard, C. (1991), Some New Estimators for Small Area Means with Application to the Assessment of Farmland Values, *Journal of Business & Economics Statistics*, **9**, 73-84.

Rabe-Hesketh, S. and Skrondal, A. (2008), Multilevel and Longitudinal Modeling Using Stata, 2nd , StataCorp LP.

Rao, J.N.K. and Yu, M. (1992), Small Area Estimation by Combining Time Series and Cross-Sectional Data, In Proceedings of the Section on Survey Research Method, *American Statistical Association*, 1-9.

جدول ۱۴. فراوانی نسبی نتایج انتخاب مدل درست با استفاده از روش گام به گام.

Area	۴۰	۵۰	۱۰۰	۲۵۰	۵۰۰	۱۰۰۰
A _۱	۰/۵۱۷	۰/۵۴۳	۰/۵۵۲	۰/۶۵۶	۰/۷۱۶	۰/۷۵۵
A _۲	۰/۶۴۶	۰/۶۵۳	۰/۶۸۲	۰/۷۵۳	۰/۷۸۸	۰/۸۲۴
A _۳	۰/۶۳۳	۰/۶۴۱	۰/۶۵۵	۰/۷۵۲	۰/۷۹۰	۰/۸۵۷
A _۴	۰/۵۸۲	۰/۶۱۸	۰/۶۷۱	۰/۶۹۲	۰/۷۳۷	۰/۷۷۹
A _۵	۰/۵۶۵	۰/۶۲۲	۰/۶۵۶	۰/۶۹۰	۰/۷۶۶	۰/۸۱۶
A _۶	۰/۶۲۰	۰/۶۷۸	۰/۶۸۶	۰/۷۵۵	۰/۷۸۱	۰/۸۴۱
A _۷	۰/۶۲۷	۰/۶۳۷	۰/۶۴۷	۰/۷۵۷	۰/۷۷۰	۰/۸۰۴
A _۸	۰/۶۱۹	۰/۶۲۳	۰/۶۴۴	۰/۶۹۶	۰/۷۴۸	۰/۸۲۰
A _۹	۰/۴۹۹	۰/۵۸۵	۰/۶۰۷	۰/۷۵۰	۰/۷۶۲	۰/۸۰۰
A _{۱۰}	۰/۶۰۲	۰/۶۲۸	۰/۶۰۶	۰/۶۶۹	۰/۶۹۱	۰/۷۶۱
A _{۱۱}	۰/۵۳۱	۰/۵۹۳	۰/۶۲۶	۰/۶۸۵	۰/۷۰۷	۰/۸۵۸
A _{۱۲}	۰/۵۵۵	۰/۵۶۵	۰/۶۴۷	۰/۷۵۸	۰/۷۶۹	۰/۸۳۱
A _{۱۳}	۰/۵۲۷	۰/۵۹۴	۰/۶۳۳	۰/۶۹۲	۰/۷۶۸	۰/۷۹۰
A _{۱۴}	۰/۵۶۱	۰/۵۹۹	۰/۶۳۸	۰/۷۶۰	۰/۷۸۸	۰/۸۱۶
A _{۱۵}	۰/۵۳۴	۰/۶۰۸	۰/۶۲۲	۰/۶۷۶	۰/۷۰۶	۰/۸۱۱
A _{۱۶}	۰/۵۳۹	۰/۵۶۹	۰/۵۹۷	۰/۶۵۹	۰/۷۱۶	۰/۷۶۲
A _{۱۷}	۰/۵۶۴	۰/۵۸۷	۰/۶۱۰	۰/۶۴۰	۰/۷۶۲	۰/۸۰۲
A _{۱۸}	۰/۶۱۶	۰/۶۴۰	۰/۶۵۷	۰/۶۹۵	۰/۷۶۴	۰/۸۴۱
A _{۱۹}	۰/۵۳۹	۰/۵۸۸	۰/۶۱۷	۰/۶۷۸	۰/۷۵۸	۰/۷۶۴
A _{۲۰}	۰/۵۸۳	۰/۶۱۲	۰/۶۵۷	۰/۶۹۴	۰/۷۴۸	۰/۷۵۷
A _{۲۱}	۰/۶۴۴	۰/۶۸۷	۰/۶۹۶	۰/۷۵۷	۰/۷۶۹	۰/۸۳۵
A _{۲۲}	۰/۶۰۶	۰/۶۵۵	۰/۶۶۳	۰/۷۴۸	۰/۷۶۰	۰/۸۵۱
A _{۲۳}	۰/۵۶۱	۰/۵۹۸	۰/۶۴۶	۰/۶۵۳	۰/۷۸۲	۰/۸۵۵
A _{۲۴}	۰/۶۱۰	۰/۶۶۲	۰/۶۷۷	۰/۷۱۷	۰/۷۵۶	۰/۷۶۲
A _{۲۵}	۰/۴۹۸	۰/۵۹۲	۰/۶۰۳	۰/۶۵۴	۰/۷۶۶	۰/۸۴۹
A _{۲۶}	۰/۵۵۴	۰/۵۹۵	۰/۶۲۵	۰/۷۱۸	۰/۷۷۳	۰/۸۵۲
A _{۲۷}	۰/۴۹۲	۰/۵۷۰	۰/۵۸۹	۰/۶۵۹	۰/۷۶۳	۰/۸۵۴

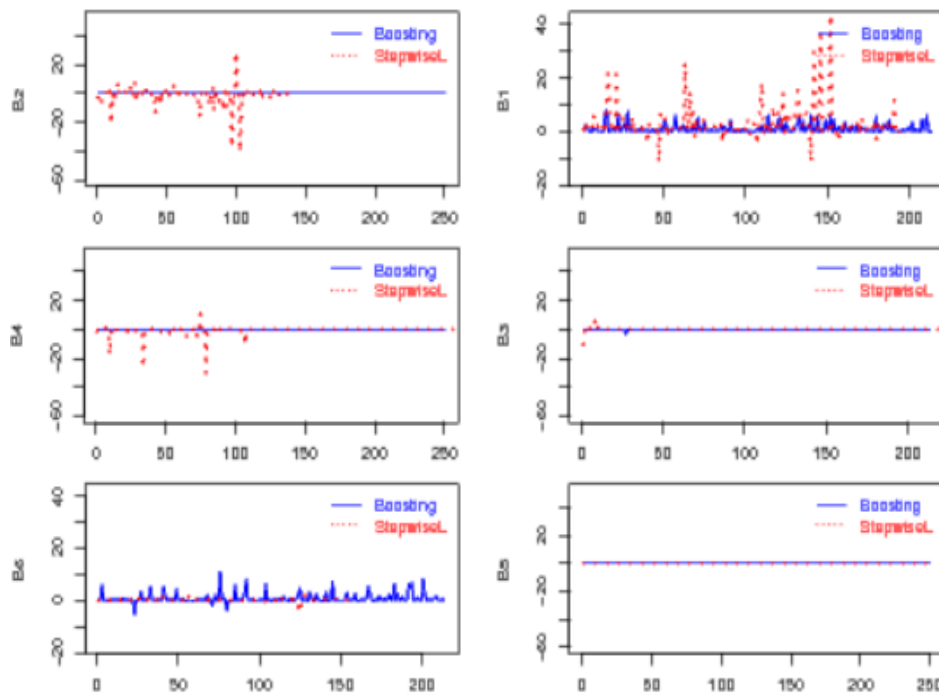
Rao, J. N. K. and Yu, M. (1994), Small Area Estimation by Combining Time Series and Cross-Sectional Data, *Canadian Journal of Statistics*, **22**, 511–528.

Robinzonov, N., Tutz, G. and Hothorn, T. (2012), Boosting Techniques for Nonlinear Time Series Models, *ASta Advances in Statistical Analysis*, **96**(1), 99-122.

Schapire, R. (1990), The Strength of Weak Learnability, *Mach. Learn.*, **5**, 197-227.

Singh, A.C., Mantel, H.J. and Thomas, B.W. (1994), Time Series EBLUPs for Small Areas using Survey Data, *Survey Methodology*, **20**, 33-43.

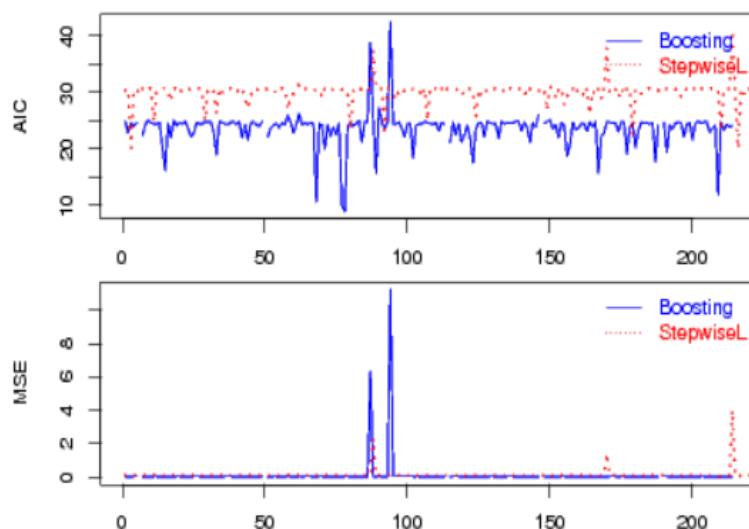
Sigrist, F. (2021), Gradient and Newton boosting for classification and regression, *Expert Systems with Applications*, **167**, 114080, ISSN 0957-4174, <https://doi.org/10.1016/j.eswa.2020.114080>.



شکل ۰۲. نمودار پارامترهای برآورد شده با الگوریتم تقویت و روش گام به گام

جدول ۰۱۵. برازش مدل با استفاده از الگوریتم تقویت و روش گام به گام

روش گام به گام	الگوریتم تقویت	
۱/۰۳۴۱	۱/۰۶۵۵	$\hat{B}_1$
-۲/۴۵۵۴	۰/۰۰۰۰	$\hat{B}_2$
۰/۰۰۰۰	۰/۰۰۰۰	$\hat{B}_3$
-۰/۲۷۸۵	۰/۰۰۰۰	$\hat{B}_4$
۰/۰۰۰۰	۰/۰۰۰۰	$\hat{B}_5$
۰/۰۲۷۳	۰/۷۷۰۰	$\hat{B}_6$
۰/۰۰۰۰	۰/۰۰۰۰	$\hat{B}_7$
۰/۰۰۰۰	۰/۰۰۰۰	$\hat{B}_8$
-۰/۴۴۲۲	۰/۰۷۴۹	$\hat{\rho}$
۰/۰۰۲۵	۱۲/۳۲۱۵	$\hat{w}$
۰/۰۲۵۰	۰/۰۳۶۸	$\hat{\alpha}$
۰/۰۴۷۱	۰/۶۳۹۲	$\hat{\beta}$
۰/۲۸۴۳	۰/۰۳۰۴	$\hat{\mu}_v$
۱/۱۲۵۷	۴/۴۷۶۷	$\hat{\sigma}_v^2$
۲۶/۶۱۱۷	۲۳/۶۳۳۹	AIC
۰/۰۹۸۸	۰/۱۰۷۲	MSE



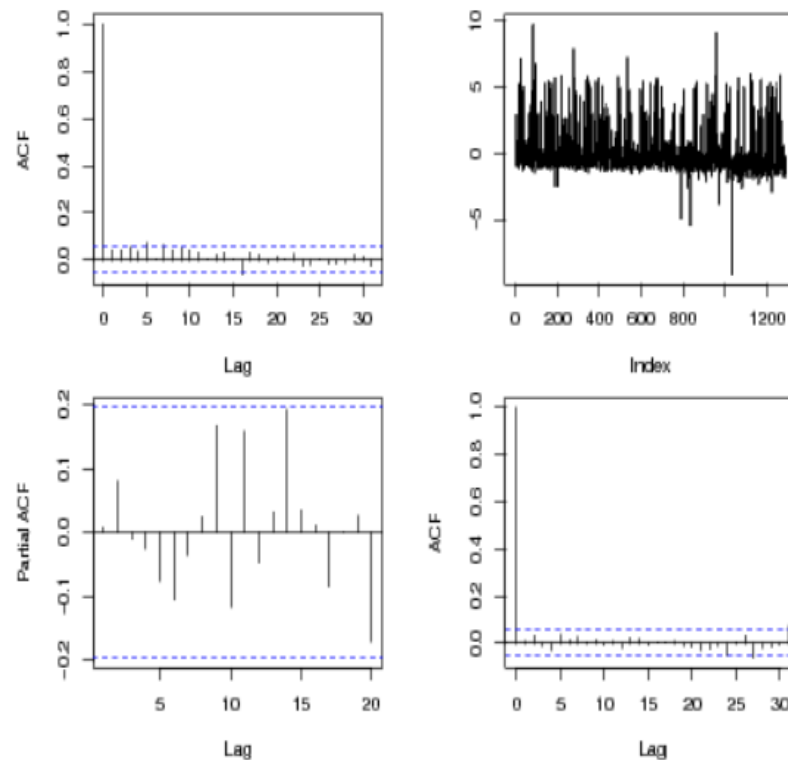
شکل ۰۳. نمودار AIC و MSE مدل‌های برازش داده شده با الگوریتم تقویت و روش گام به گام

جدول ۰۱۶. برازش مدل با استفاده از الگوریتم تقویت و روش گام به گام

روش گام به گام	الگوریتم تقویت	
۰/۴۱۲۱	۲۸/۸۲۶۶	$\hat{B}_1$
۰/۲۵۲۶	۰/۰۰۰۰	$\hat{B}_2$
-۰/۶۲۵۵	-۰/۹۸۱۷	$\hat{B}_3$
-۰/۱۲۴۵	۰/۰۰۰۰	$\hat{B}_4$
۰/۰۰۰۰	۰/۰۰۰۰	$\hat{B}_5$
۰/۱۳۵۳	۴۲/۶۵۷۵	$\hat{B}_6$
-۰/۷۲۲۶	-۱/۶۵۸۶	$\hat{B}_7$
۰/۰۰۰۰	۰/۰۰۰۰	$\hat{B}_8$
۰/۰۳۲۸	۰/۰۳۹۴	$\hat{\rho}$
۰/۰۱۰۷	۰/۰۱۰۳	$\hat{W}$
۰/۰۰۰۰	۰/۰۰۰۰	$\hat{\alpha}$
۰/۹۹۶۱	۰/۹۹۶۳	$\hat{\alpha}$
۰/۰۰۰۱	-۰/۰۰۰۹	$\hat{\mu}_v$
۱/۰۰۰۶	۱/۰۰۰۷	$\hat{\sigma}_v^2$
۳۶۷۱/۰۶۰۰	۳۶۶۳/۰۳۳۳	AIC
۰/۹۹۹۸	۰/۹۹۹۹	MSE

Sugasawa, S., Kawakubo, Y. and Datta, G. S. (2019), Observed Best Selective Prediction in Small Area Estimation, *Journal of Multivariate Analysis*, **173**, 383-392.

Sugasawa, S. and Kubokawa, T. (2020), Small Area Estimation with Mixed mod-



شکل ۴. نمودار سری زمانی (راست بالا) نمودار خودهمبستگی سری زمانی (چپ بالا)، نمودار خود همبستگی قدر مطلق سری زمانی مانده‌های استاندارد شده مدل $AR(1) - GARCH(1, 1)$ برازش داده شده (راست پایین) و نمودار خود همبستگی جزئی قدرمطلق سری زمانی مانده‌های استاندارد شده مدل $AR(1) - GARCH(1, 1)$ برازش داده شده (چپ پایین).

els: A Review, *Japanese Journal of Statistics and Data Science*, **3**, 693-720.
<https://doi.org/10.1007/s42081-020-00076-x>.

Vaida, F. and Blanchard, S. (2005), Conditional Akaike Information for Mixed-Effects Models, *Biometrika*, **92**, 351-370.

Wang, W. (2013), Identifiability of Linear Mixed Effects Models, *Electronic Journal of Statistics*, **7**, 244-263.

You, Y. and Rao, J.N.K. (2000), Hierarchical Bayes Estimation of Small Area Means using Multi-level Models, *Survey Methodology*, **26**, 173-181.