



## برآورد مدل اتورگرسیو چندکی خطی با استفاده از الگوریتم EM تصادفی

محمد بهمنی \*

گروه آمار، دانشکده علوم، دانشگاه زنجان، زنجان، ایران

دبیر مسئول: غلامرضا محتشمی برزادران

تاریخ پذیرش: ۱۴۰۱/۵/۲۲

تاریخ دریافت: ۱۴۰۰/۹/۲۰

چکیده: در این مقاله، مدل سری زمانی اتورگرسیو چندکی معرفی شده است و سپس پارامترهای مدل با استفاده از الگوریتم SEM که یک روش تکراری برای محاسبه برآوردهای ماکسیمم درستنمایی است، برآورد می‌شوند. تابع درستنمایی در مدل اتورگرسیو چندکی بر اساس توزیع لاپلاس نامتقارن و همچنین آمیخته مقیاس این توزیع بیان می‌شود و با استفاده از الگوریتم SEM پارامترهای مدل برآورد می‌شوند. کارایی و کاربرد روش پیشنهادی با مطالعات شبیه‌سازی و تحلیل داده‌های واقعی مورد ارزیابی قرار می‌گیرد.

واژه‌های کلیدی: مدل اتورگرسیو چندکی، الگوریتم SEM، توزیع لاپلاس نامتقارن.

رده‌بندی ریاضی: 62G08; 62M10; 62F10.

### ۱ مقدمه

تحلیل سری‌های زمانی به طور نظری و عملی از سال ۱۹۷۰ میلادی به جهت پیش‌بینی و کنترل فرایندهای وابسته به زمان، به سرعت بسط و توسعه پیدا کرده است. در سال‌های اخیر تحلیل مدل‌های سری‌های زمانی کاربرد زیادی در علوم مختلف از جمله زیست‌شناسی، اقتصاد سنجی و محیط‌زیست پیدا کرده است. دلیل این امر نیز داده‌های جمع‌آوری شده در این حوزه‌ها می‌باشد که معمولاً یک وابستگی بین داده‌ها در طول زمان وجود دارد. در واقع تحلیل سری‌های زمانی معمولاً به داده‌هایی مربوط می‌شود که مستقل نبوده و به طور متوالی به یکدیگر وابسته‌اند یا به عبارت دیگر مشاهدات در طول زمان جمع‌آوری می‌شوند و یک وابستگی بین آنها موجود است. در واقع همین وابستگی بین مشاهدات متوالی است که مورد توجه قرار می‌گیرد و بیشتر کاربرد آن نیز در پیش‌بینی خواهد بود. این پیش‌بینی در مورد حوادثی است که در آینده احتمال وقوع دارند و معمولاً بر اساس رویدادهایی صورت می‌گیرند که در گذشته اتفاق افتاده‌اند. به بیان دیگر تحلیل سری‌های زمانی، به دنبال یک مدل براساس اطلاعات گذشته است تا از طریق این مدل امکان تصمیمات آینده‌نگر را فراهم سازد.

یکی از مدل‌های مهم در تحلیل سری‌های زمانی، مدل اتورگرسیو است. فرایندهای اتورگرسیو کاربرد زیادی در پردازش سیگنال، مخابرات، رادار و اقتصاد سنجی دارند. مدل‌های اتورگرسیو به اختصار با نماد  $AR(p)$  نمایش داده می‌شوند، که در آن  $p$  مرتبه مدل است. مرتبه

\*نویسنده مسئول مقاله

رایانامه: (M. Bahmani), bahmanimo74@znu.zc.ir

مدل را می‌توان بر اساس توابع خودهمبستگی (ACF) و خودهمبستگی جزئی (PACF) مشاهدات یا با استفاده از معیارهای عددی دیگر نظیر معیار اطلاع آکائیک (AIC) یا معیار اطلاع بیزی (BIC) تعیین نمود. به دلایل ذکر شده، مدل‌های اتورگرسیو از دیدگاه آمار بیزی و فراوانی‌گرا مورد مطالعه قرار گرفته‌اند. برای مثال [۷] برآوردهای فراوانی‌گرا و بیزی مدل اتورگرسیو مرتبه اول را با فرض بر این که خطاهای مدل از توزیع چوله نرمال پیروی می‌کنند را مورد مطالعه قرار دادند. [۱۳] پارامترهای مدل اتورگرسیو با مرتبه  $p$  را با فرض توزیع‌های دم پهن برای خطاهای مدل در حضور داده‌های گمشده در مشاهدات را با استفاده از الگوریتم EM تصادفی برآورد کردند. همچنین [۵] مدل‌های اتورگرسیو خطی جزئی با خطاهای خانواده توزیع‌های آمیخته مقیاس از توزیع نرمال را ارائه کردند. [۸] مدل اتورگرسیو نیمه‌پارامتری غیرخطی را با خطاهای خانواده توزیع‌های آمیخته از توزیع چوله نرمال را مورد بررسی قرار دادند. همچنین [۱۵] پارامترهای مدل اتورگرسیو را با استفاده از الگوریتم EM و با فرض توزیع تی روی خطاها محاسبه کردند. در کلیه مطالعات عنوان شده، مدل‌های اتورگرسیو با فرض این که مولفه خطا دارای توزیع مشخص با میانگین صفر بوده، مورد بررسی قرار گرفته‌اند. به این معنی که در مدل اتورگرسیو مرتبه  $p$ ,

$$y_t = \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + \varepsilon_t$$

همواره فرض بر این است که  $\varepsilon_t$  دارای توزیع مشخص و  $E(\varepsilon_t) = 0$  است. لذا

$$E(y_t | y_{t-1}, \dots, y_{t-p}) = \phi_1 y_{t-1} + \dots + \phi_p y_{t-p}.$$

بنابراین می‌توان گفت در این مطالعات همانند رگرسیون میانگین، هدف تحلیل متوسط مقدار شرطی  $y_t$  است. اما به خوبی معلوم است که میانگین به عنوان یک معیار گرایش به مرکز نسبت به مشاهدات دور افتاده حساس است. لذا در صورت وجود نقاط دور افتاده در مشاهدات، مدل اتورگرسیو عنوان شده ممکن است دیگر مدل کارایی نباشد. [۱۰] برای حل این مشکل مدل اتورگرسیو چندکی را پیشنهاد کردند. در این مدل به جای بررسی متوسط مقدار متغیر  $y_t$ ، هدف تحلیل چندک شرطی توزیع متغیر  $y_t$  است. مدل اتورگرسیو چندکی نسبت به مدل‌های قبلی دارای دو مزیت مهم است، نخست این که مدل اتورگرسیو چندکی نسبت به داده‌های دور افتاده استوار است و دوم این که اطلاعات جزئی‌تری در خصوص تاثیر متغیرهای  $p$  گام قبل در چندک‌های مختلف توزیع متغیر  $y_t$  را ارائه می‌دهد [۱۰]. اگر فرض کنیم چندک  $\tau$  ام  $\varepsilon_t$  برابر صفر باشد آنگاه چندک شرطی  $\tau$  ام  $y_t$  به صورت زیر است،

$$Q_\tau(y_t | y_{t-1}, \dots, y_{t-p}) = \phi_1^\tau y_{t-1} + \dots + \phi_p^\tau y_{t-p}.$$

اگر  $\tau$  را برابر  $0.5$  قرار دهیم آنگاه مدل اتورگرسیو میانه به دست می‌آید. [۱۰] نشان دادند که در مدل اتورگرسیو چندکی برآورد پارامترها نسبت به روش حداقل مربعات زمانی که داده‌ها دارای مشاهدات دور افتاده هستند یا مشاهدات از توزیع‌های چوله و دم پهن پیروی می‌کنند، دقت بیشتری دارند. مدل رگرسیون چندکی و اتورگرسیو چندکی با رویکرد آمار بیزی و آمار فراوانی‌گرا مورد توجه بسیاری از محققان قرار گرفته است، برای نمونه [۱۲] ابتدا معیار همبستگی چندکی و همبستگی جزئی چندکی را با استفاده از رگرسیون چندکی معرفی کردند سپس با تعمیم این معیارها به مدل اتورگرسیو چندکی، توابع خود همبستگی چندکی و خود همبستگی جزئی چندکی را برای تجزیه و تحلیل، تشخیص مرتبه مدل و ارزیابی کارایی مدل‌های سری زمانی پیشنهاد نمودند. [۱۴] مدل‌های رگرسیونی را بر اساس کلاس توزیع‌های لوگ نرمال مستقل برای داده‌های مثبت به صورت چندکی مدلسازی کردند و با استفاده از الگوریتم EM پارامترهای مدل را برآورد نمودند. همچنین برای مطالعه بیشتر می‌توان به [۴] و [۲۰] اشاره کرد. از این رو مدل‌هایی که توزیع مشاهدات را در چندک‌های مختلف مورد بررسی قرار می‌دهند و نسبت به ناپایداری واریانس و داده‌های دور افتاده نیز حساس نیستند، می‌توانند جایگزینی مناسب برای مدل‌های مبتنی بر میانگین در نظر گرفته شوند. اخیراً محققان زیادی از توزیع لاپلاس نامتقارن (AL) برای تحلیل مدل رگرسیون چندکی و اتورگرسیو چندکی در مطالعات خود استفاده کرده‌اند. رویکرد این مطالعات بر این پایه استوار است که مینیمم کردن تابع زیان در مدل چندکی با ماکسیمم کردن تابع درستنمایی بر اساس توزیع AL هم ارز می‌باشد. [۱۶] پارامترهای مدل رگرسیون چندکی با خطاهای اتورگرسیو مرتبه  $p$  را با استفاده از الگوریتم EM و بر پایه‌ی توزیع AL برآورد کردند، در واقع [۱۶] با فرض توزیع AL برای خطاهای مدل، تابع درستنمایی را تشکیل دادند. هر چند که فرض توزیع AL برای خطاها ممکن است در عمل نادرست باشد اما به لحاظ تجربی نتایج بسیار قابل قبول و استواری را به دست می‌دهد. [۶] مدل رگرسیون چندکی توام برای داده‌های پیوسته و گسسته را بر اساس مفصل گاوسی و با استفاده از توزیع AL بررسی کردند و از طریق الگوریتم EM پارامترهای مدل را برآورد نمودند. [۱۷] با فرض توزیع AL برای خطاها برآورد ماکزیمم درستنمایی پارامترهای مدل اتورگرسیو چندکی تبدیل مارکوفی را با استفاده از الگوریتم EM محاسبه کردند. در کلیه‌ی این مطالعات مدل‌سازی بر اساس توزیع AL برای خطاها صورت گرفته است و تابع درستنمایی تحت توزیع AL تشکیل شده است همچنین پارامترهای مدل با استفاده از الگوریتم EM برآورد شده است. الگوریتم EM روشی کارآمد برای برآورد ماکسیمم درستنمایی یک مدل آماری در حضور داده‌های سانسور شده، متغیرهای پنهان است اما بعضی از اوقات به دست آوردن مرحله E از الگوریتم EM دشوار می‌باشد، یک راه حل برای این مشکل استفاده از الگوریتم SEM است [۲]. الگوریتم SEM نسبت به الگوریتم EM ساده‌تر و دارای محاسبات راحت‌تر است.

هدف اصلی این مقاله، ارائه رویکردی جدید برای برآورد پارامترهای مدل اتورگرسیو چندکی بر اساس تابع درستنمایی است. برای این منظور، در بخش ۲ مدل اتورگرسیو چندکی معرفی شده و تابع درستنمایی با فرض توزیع AL برای خطاها تشکیل می‌شود. در بخش ۳ پارامترهای مدل اتورگرسیو چندکی از طریق الگوریتم SEM برآورد می‌شود. در بخش ۴ کارایی و کاربست مدل در مطالعات شبیه‌سازی تحت توزیع‌های

مختلف برای خطاها و معیار خطای مدل در مقایسه با مدل اتورگرسیو چندکی کلاسیک در [۱۰] سنجیده می‌شود. در بخش ۵، مدل پیشنهادی به داده‌های واقعی برازش داده می‌شود و با توجه به معیار میانگین قدر مطلق خطای پیش بینی (MAP) کارایی مدل در مقایسه با مدل اتورگرسیو چندکی کلاسیک سنجیده می‌شود. در بخش ۶ نتیجه‌گیری و پیشنهادها ارائه می‌شود.

## ۲ مدل

مدل اتورگرسیوخطی مرتبه  $p$  را به صورت زیر در نظر بگیرید،

$$y_t = \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + \varepsilon_t = \sum_{j=1}^p y_{t-j} \phi_j + \varepsilon_t, \quad t = p+1, \dots, n, \quad (1.2)$$

که در آن  $y_t$  متغیر پاسخ در زمان  $t$ ،  $\Phi = (\phi_1, \dots, \phi_p)'$  پارامترهای مدل اتورگرسیو و  $\varepsilon_t$  مولفه خطا با میانگین صفر و واریانس  $\sigma^2$  است. در رابطه (۱.۲) برای اطمینان از ایستایی فرض می‌کنیم تمام ریشه‌های چندجمله‌ای  $1 - \sum_{j=1}^p \phi_j z^j$  خارج دایره واحد هستند. با فرض این که چندک  $\tau$  ام  $\varepsilon_t$  برابر صفر است (یعنی  $\int_{-\infty}^0 f(\varepsilon_t) d\varepsilon_t = \tau$  به ازای  $t = p+1, \dots, n$ )، آنگاه

$$Q(y_t | y_{t-p}, \dots, y_{t-1}) = \phi_1 y_{t-1} + \dots + \phi_p y_{t-p}, \quad (2.2)$$

چندک  $\tau$  ام شرطی  $y_t$  به شرط  $y_{t-p}, \dots, y_{t-1}$  است. توجه کنیم مدل (۲.۲) یک مدل نیمه پارامتری است. زیرا تنها شرط مدل، صفر بودن چندک  $\tau$  ام مولفه خطا است. در حالی که در مدل اتورگرسیو معمولی مرتبه  $p$ ، فرض بر این است که مولفه‌های خطا دارای توزیع نرمال با میانگین صفر و واریانس  $\sigma^2$  می‌باشد. حال با توجه به مدل رگرسیون چندکی ارائه شده در [۹] و همچنین ایده بیان شده توسط [۱۰]، برآورد پارامترهای مدل اتورگرسیو چندکی در چندک  $\tau$  ام به صورت

$$\hat{\Phi} = \arg \min_{\Phi} \sum_{t=p+1}^n \rho_{\tau} \left( y_t - \sum_{j=1}^p \phi_j y_{t-j} \right), \quad (3.2)$$

به دست می‌آیند، که در آن  $\rho_{\tau}(u)$  تابع زیان و به صورت

$$\rho_{\tau}(u) = \begin{cases} u\tau, & u > 0 \\ u(1-\tau), & u \leq 0 \end{cases}$$

تعریف می‌شود. [۱۸] از طریق تعریف توزیع AL برای خطاها نشان دادند که مینیمم کردن تابع زیان در (۳.۲) معادل با ماکسیمم کردن تابع درستنمایی بر اساس این توزیع است. [۱۸] ثابت کردند که فرض توزیع AL برای خطاها نتایج تجربی استواری را ارائه می‌کند هر چند این فرض نادرست برای خطاهای واقعی باشد. تابع چگالی توزیع AL به صورت زیر است،

$$f(y | \mu, \sigma, \tau) = \frac{\tau(1-\tau)}{\sigma} \exp \left\{ -\rho_{\tau} \left( \frac{y-\mu}{\sigma} \right) \right\},$$

که در آن  $\mu$  پارامتر مکان،  $\sigma$  پارامتر مقیاس و  $\tau$  پارامتر شکل است. دلیل استفاده از توزیع AL در تحلیل مدل‌های رگرسیون چندکی و اتورگرسیو چندکی این است که  $\mu$  چندک  $\tau$  ام توزیع AL می‌باشد. با فرض این که در مدل (۱.۲) مولفه خطا دارای توزیع  $AL \left( \sum_{j=1}^p \phi_j y_{t-j}, \sigma, \tau \right)$  است، تابع درستنمایی به صورت

$$\begin{aligned} L(\Phi, \sigma | y_1, \dots, y_n) &= \prod_{t=p+1}^n [f(y_t | y_{t-p}, \dots, y_1)] \\ &= \prod_{t=p+1}^n \left[ \frac{\tau(1-\tau)}{\sigma} \exp \left\{ -\rho_{\tau} \left( \frac{y_t - \sum_{j=1}^p \phi_j y_{t-j}}{\sigma} \right) \right\} \right], \end{aligned}$$

است. بنابراین تابع درستنمایی مدل اتورگرسیو چندکی مرتبه  $p$  بر اساس توزیع  $AL$  را می‌توان به صورت زیر نوشت،

$$L(\Theta|\mathbf{y}) = \left(\frac{\tau(1-\tau)}{\sigma}\right)^{n-p} \exp\left\{-\sum_{t=p+1}^n \rho_{\tau} \left(\frac{y_t - \sum_{j=1}^p \phi_j y_{t-j}}{\sigma}\right)\right\}, \quad (4.2)$$

که در آن  $\Theta = (\Phi', \sigma)$  بردار پارامترهای مدل را نشان می‌دهد. هر استنباط آماری برای مدل اتورگرسیو چندکی رابطه (۱.۲) را می‌توان بر اساس تابع درستنمایی (۴.۲) انجام داد. اما به دلیل وجود تابع قدرمطلق نمی‌توان به صورت تحلیلی تابع درستنمایی را نسبت به پارامترها ماکسیمم کرد و فرم بسته‌ای برای برآوردها به دست آورد. برای حل این مشکل ابتدا با استفاده از شکل آمیخته مقیاس توزیع  $AL$ ، مدل را به صورت آمیخته‌ای از توزیع نرمال بازنویسی می‌کنیم [۱۱]. اگر فرض کنیم متغیر تصادفی  $y$  دارای توزیع  $AL$  باشد، آنگاه

$$y|v \sim N(\theta_1 v, \theta_2 \sigma v); \quad v \sim \text{Exp}\left(\frac{1}{\sigma}\right), \quad E(v) = \sigma.$$

در نتیجه با استفاده از شکل آمیخته توزیع  $AL$  رابطه (۱.۲) را می‌توان به صورت

$$y_t = \sum_{j=1}^p \phi_j y_{t-j} + \theta_1 v_t + \sqrt{\theta_2 \sigma v_t} e_t, \quad t = p+1, \dots, n, \quad (5.2)$$

در نظر گرفت، که در آن  $v_t \sim \text{Exp}\left(\frac{1}{\sigma}\right)$ ،  $e_t \sim N(0, 1)$  و  $v_t$  و  $e_t$  مستقل از هم هستند. در عبارت فوق

$$\theta_1 = \frac{1-2\tau}{\tau(1-\tau)} \quad \theta_2 = \frac{2}{\tau(1-\tau)}$$

است و  $v_t$  همان متغیر آمیخته کننده را نشان می‌دهد. بنابراین تابع درستنمایی شرطی توام برای داده‌های کامل  $\{\mathbf{y}, \mathbf{v}\}$  به صورت زیر به دست می‌آید،

$$L_C(\Theta|\mathbf{y}, \mathbf{v}) = \prod_{t=p+1}^n \left[ \frac{1}{\sqrt{2\pi} \sqrt{\theta_2 \sigma v_t}} \exp\left\{-\frac{\left(y_t - \sum_{j=1}^p \phi_j y_{t-j} - \theta_1 v_t\right)^2}{2\theta_2 \sigma v_t}\right\} \cdot \frac{1}{\sigma} \exp\left\{-\frac{1}{\sigma} v_t\right\} \right]. \quad (6.2)$$

### ۳ روش برآورد

الگوریتم EM که اولین بار توسط [۳] ارائه شد، ابزاری قدرتمند برای محاسبه برآورد ماکسیمم درستنمایی پارامترهای یک مدل آماری در حضور داده‌های ناقص، متغیرهای پنهان، توزیع‌های بریده شده، مشاهدات گروه‌بندی شده و سانسور شده است. با این حال ممکن است محاسبه متوسط مقدار در مرحله E، الگوریتم EM بسیار سخت و غیرممکن باشد، به خصوص زمانی که متوسط مقدار یک انتگرال با ابعاد بالا یا یک مجموع بزرگ باشد. یکی از راه‌هایی که برای حل این مشکل پیشنهاد شده، الگوریتم SEM است [۲]. در واقع الگوریتم SEM یک جایگزین بسیار مناسب برای الگوریتم EM می‌باشد. با توجه به وجود متغیرهای پنهان  $v_t$  در تابع درستنمایی رابطه (۶.۲)، الگوریتم SEM برای برآورد پارامترها روش کارایی است. در این قسمت با استفاده از این روش، برآورد پارامترهای مدل عنوان شده را به دست می‌آوریم. الگوریتم SEM یک روش تکراری برای محاسبه برآوردهای ماکسیمم درستنمایی است. هر تکرار از این الگوریتم شامل دو مرحله است: مرحله S و مرحله M، که با یک مقدار اولیه مانند  $\Theta^{(0)}$  آغاز می‌شود. در مرحله S با توجه به توزیع شرطی متغیر پنهان  $v_t$ ، مقدار  $v_t$  را شبیه‌سازی می‌کنیم و در لگاریتم تابع درستنمایی داده‌های کامل قرار می‌دهیم. مرحله M که به آن مرحله ماکسیمم‌سازی گویند، لگاریتم تابع درستنمایی داده‌های کامل نسبت به پارامترهای مجهول یعنی  $\Theta$  ماکسیمم می‌شود تا  $\Theta$ ی جدید به دست آید. به همین ترتیب مراحل S و M را به اندازه‌ی کافی تکرار کرده و بعد حذف نمونه‌های غیرموثر از باقی نمونه‌ها برای استنباط در مورد پارامترها استفاده می‌کنیم. برای مدل اتورگرسیو چندکی مرتبه  $p$ ، در تکرار  $h$ ام لگاریتم تابع درستنمایی برای داده‌های کامل به صورت زیر به دست می‌آید،

$$\log [L_C(\Theta|\mathbf{y}, \mathbf{v})] \propto -\frac{(n-p)}{2} \log(\theta_2) - \frac{3(n-p)}{2} \log \sigma - \frac{1}{2} \sum_{t=p+1}^n \log v_t - \sum_{t=p+1}^n \frac{1}{\sigma} \left[ \frac{\left(y_t - \sum_{j=1}^p \phi_j y_{t-j}\right)^2}{2\theta_2} v_t^{-1} + \frac{\theta_1^2 + 2\theta_2}{2\theta_2} v_t - \frac{\theta_1 \left(y_t - \sum_{j=1}^p \phi_j y_{t-j}\right)}{\theta_2} \right]. \quad (1.3)$$

مرحله S: در این مرحله مقدار  $v_t$  را با توجه به توزیع شرطی متغیر پنهان  $v_t$  به شرط مشاهدات شبیه‌سازی می‌کنیم،

$$\begin{aligned} f(v_t|y_t) &\propto \frac{1}{\sqrt{v_t}} \exp \left\{ -\frac{1}{2} \frac{\left( y_t - \sum_{t=p+1}^n \phi_j y_{t-j} - \theta_1 v_t \right)^2}{2\theta_2 \sigma v_t} \right\} \exp \left\{ -\frac{1}{\sigma} v_t \right\} \\ &\propto \frac{1}{\sqrt{v_t}} \exp \left\{ -\frac{1}{2} \left[ \frac{\left( y_t - \sum_{j=1}^p \phi_j y_{t-j} \right)^2}{\theta_2 \sigma} v_t^{-1} + \frac{\theta_1^2 + 2\theta_2}{\theta_2 \sigma} v_t \right] \right\} \\ &\sim GIG \left( \frac{1}{2}, \frac{\left( y_t - \sum_{j=1}^p \phi_j y_{t-j} \right)^2}{\theta_2 \sigma}, \frac{\theta_1^2 + 2\theta_2}{\theta_2 \sigma} \right), \quad t = p+1, \dots, n. \end{aligned} \quad (۲.۳)$$

که در آن  $GIG(\lambda, \chi, \psi)$ ، توزیع گاوسی وارون تعمیم‌یافته با پارامترهای  $\lambda \in \mathbb{R}$  و  $\psi > 0, \chi > 0$  و تابع چگالی

$$f(x|\lambda, \chi, \psi) = \frac{\chi^{-\lambda} (\sqrt{\chi\psi})^\lambda}{2K_\lambda(\sqrt{\chi\psi})} x^{\lambda-1} \exp \left\{ -\frac{1}{2} (\chi x^{-1} + \psi x) \right\}, \quad x > 0,$$

می‌باشد [۱۹]. همچنین  $K_\lambda(\sqrt{\chi\psi})$  تابع بسل تعدیل‌یافته است که به صورت زیر تعریف می‌شود [۱]:

$$K_\lambda(\sqrt{\chi\psi}) = \frac{1}{2} \int_0^\infty x^{\lambda-1} \exp \left\{ -\frac{1}{2} (\chi\psi) \{ \chi - x^{-1} \} \right\} dx.$$

مرحله M: در این مرحله مقدار  $v_t$  را که در مرحله S با استفاده از توزیع شرطی شبیه‌سازی کردیم در لگاریتم تابع درست‌نمایی داده‌های کامل قرار داده سپس از لگاریتم تابع درست‌نمایی داده‌های کامل نسبت به  $\Theta = (\Phi', \sigma)$  مشتق گرفته و مساوی صفر قرار می‌دهیم. برای پارامتر مقیاس  $\sigma$  در تکرار  $h$  ام داریم:

$$\frac{\partial Q}{\partial \sigma} = -\frac{3(n-p)}{2\sigma} + \frac{1}{\sigma^2} \sum_{t=p+1}^n \left[ \frac{\eta_t^2}{2\theta_2} v_t^{-1} + \frac{\theta_1^2 + 2\theta_2}{2\theta_2} v_t^{-1} - \frac{\theta_1 \eta_t}{\theta_2} \right] = 0 \quad (۳.۳)$$

لذا،

$$\hat{\sigma}^{(h)} = \frac{2}{3(n-p)} \sum_{t=p+1}^n \left[ \frac{\eta_t^2}{2\theta_2} \cdot \frac{1}{v_t^{(h)}} + \frac{\theta_1^2 + 2\theta_2}{2\theta_2} v_t^{(h)} - \frac{\theta_1 \eta_t}{\theta_2} \right], \quad (۴.۳)$$

که در آن  $\eta_t = y_t - \sum_{j=1}^p \hat{\phi}_j^{(h-1)} y_{t-j}$  است. همچنین برای ضرایب  $\phi_1, \dots, \phi_p$  در تکرار  $h$  ام داریم:

$$\frac{\partial Q}{\partial \phi_k} = -\frac{1}{\theta_2 \sigma} \sum_{t=p+1}^n \left[ v_t^{-1} \left( y_t - \sum_{t=p+1}^n \phi_j y_{t-j} \right) \cdot y_{t-k} + \theta_1 y_{t-k} \right] = 0, \quad k = 1, \dots, p, \quad (۵.۳)$$

لذا،

$$\sum_{j=1}^p \phi_j \sum_{t=p+1}^n y_{t-j} v_t^{-1} y_{t-k} = \sum_{t=p+1}^n y_{t-k} (v_t^{-1} y_t - \theta_1) \quad (۶.۳)$$

بنابراین برآوردهای  $\Phi = (\phi_1, \dots, \phi_p)$  به صورت زیر خواهد بود،

$$A\Phi = B \quad \Rightarrow \quad \hat{\Phi} = A^{-1}B, \quad (۷.۳)$$

که در آن

$$A = \begin{pmatrix} \sum_{t=p+1}^n y_{t-1} v_t^{-1} y_{t-1} & \cdots & \sum_{t=p+1}^n y_{t-1} v_t^{-1} y_{t-p} \\ \vdots & & \vdots \\ \sum_{t=p+1}^n y_{t-p} v_t^{-1} y_{t-1} & \cdots & \sum_{t=p+1}^n y_{t-p} v_t^{-1} y_{t-p} \end{pmatrix},$$

$$B = \begin{pmatrix} \sum_{t=p+1}^n y_{t-1} (v_t^{-1} y_t - \theta_1) \\ \vdots \\ \sum_{t=p+1}^n y_{t-p} (v_t^{-1} y_t - \theta_1) \end{pmatrix}.$$

لذا می توان رابطه (۷.۳) را به صورت

$$\hat{\Phi}^{(h)} = (E^T W^{(h)} E)^{-1} \cdot [E^T W^{(h)} (e - (W)^{-1} D)] \quad (۸.۳)$$

نوشت، که در آن

$$E^T = \begin{pmatrix} y_p & y_{p+1} & \cdots & y_{n-1} \\ y_{p-1} & y_p & \cdots & y_{n-2} \\ \vdots & \vdots & & \vdots \\ y_1 & y_2 & \cdots & y_{n-p} \end{pmatrix}_{p \times (n-p)}, \quad d = \begin{pmatrix} \theta_1 \\ \theta_1 \\ \vdots \\ \theta_1 \end{pmatrix}_{(n-p) \times 1}$$

$$W^{(h)} = \begin{pmatrix} 1/v_{p+1}^{(h)} & & & \\ & \ddots & & \\ & & \ddots & \\ & & & 1/v_n^{(h)} \end{pmatrix}_{(n-p) \times (n-p)}, \quad e = \begin{pmatrix} y_{p+1} \\ y_{p+2} \\ \vdots \\ y_n \end{pmatrix}_{(n-p) \times 1}$$

است. حال گام های S و M را به تعداد کافی و بزرگ تکرار کرده و بعد از حذف نمونه های غیر موثر، از باقی نمونه ها برای استنباط در مورد پارامترها استفاده می کنیم. در بخش های بعدی با استفاده از مطالعه شبیه سازی و تحلیل داده واقعی کارایی مدل ارزیابی می شود.

#### ۴ مطالعات شبیه سازی

در این بخش برای مقایسه کارایی روش ارائه شده در این مقاله (SEM) با مدل اتورگرسیو چندکی بر اساس تابع زیان (QA) ارائه شده در [۱۰] به مطالعه شبیه سازی می پردازیم. برای محاسبه برآوردهای مدل اتورگرسیو چندکی بر اساس تابع زیان، از تابع  $qr()$  در بسته *quantreg* نرم افزار R استفاده می کنیم. برای این منظور دو شبیه سازی به ترتیب زیر انجام می دهیم.

شبیه سازی ۱: یک مجموعه داده به حجم  $n = 100$  از مدل

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \phi_3 y_{t-3} + \phi_4 y_{t-4} + \varepsilon_t, \quad t = p + 1, \dots, n \quad (۱.۴)$$

با فرض  $(\phi_1, \phi_2, \phi_3, \phi_4) = (0.3, -0.05, 0.1, 0.1)$  تولید می کنیم.

شبیه سازی ۲: یک مجموعه داده به حجم  $n = 100$  از مدل

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t, \quad t = p + 1, \dots, n \quad (۲.۴)$$

با فرض  $(\phi_1, \phi_2) = (0.4, 0.4)$  تولید می کنیم. هم چنین  $\varepsilon_t$  ها از توزیع های زیر تولید می شوند:

$$\varepsilon_t \sim N(0, 1) \bullet$$

$$\varepsilon_t \sim \text{Laplace}(0, 1) \bullet$$

$$\varepsilon_t \sim t_{(3)} \bullet$$

توجه داشته باشیم که توزیع‌های  $\text{Laplace}(0, 1)$  و  $t_{(3)}$  دم پهن هستند و اغلب برای مدلسازی داده‌های دورافتاده مورد استفاده قرار می‌گیرند. ما با فرض این توزیع‌ها برای خطاها، استواری مدل را مورد بررسی قرار می‌دهیم. مقادیر آغازین پارامترها را در الگوریتم SEM برابر برآوردگرهای حداقل مربعات قرار می‌دهیم. به منظور لحاظ کردن تغییرپذیری نتایج حاصل از شبیه‌سازی، ۲۰۰ مرتبه روند شبیه‌سازی را تکرار می‌کنیم. در هر تکرار از شبیه‌سازی، مراحل S و M را ۴۰۰۰ مرتبه تکرار می‌کنیم، سپس ۲۰۰۰ نمونه‌ی اول را به عنوان نمونه‌های غیرموثر حذف کرده و سپس از ۲۰۰۰ نمونه‌ی بعدی برای محاسبه مقدار  $\hat{\phi}_i^{(k)} = \frac{1}{2000} \sum_{k=1}^{2000} \phi_i^{(k)}$  استفاده می‌کنیم. این روند را ۲۰۰ مرتبه تکرار کرده و سپس متوسط مقدار پارامترها را در ۲۰۰ تکرار شبیه‌سازی، مجذور میانگین مربعات خطا (RMSE) و یک فاصله اطمینان 95% برای پارامترهای مدل مورد نظر به صورت زیر به دست می‌آوریم،

$$\text{RMSE} = \sqrt{\frac{1}{200} \sum_{i=1}^{200} \left( \hat{\phi}_i^{(k)} - \phi^{true} \right)^2},$$

$$\text{mean} = \frac{1}{200} \sum_{i=1}^{200} \hat{\phi}_i^{(k)}, \quad i = 1, \dots, 200, \quad k = 1, \dots, 2000.$$

نتایج شبیه‌سازی برای سه سطح چندک در جدول‌های ۱ و ۲ گزارش شده است. با توجه به این جدول‌ها نتایج زیر حاصل می‌شود:

(۱) مدل اتورگرسیو چندکی بر اساس توزیع AL نسبت به توزیع خطاهای مختلف استوار می‌باشد، به ویژه زمانی که داده‌های دورافتاده در مشاهدات وجود دارد.

(۲) الگوریتم SEM برای برآورد پارامترها روش کارایی می‌باشد، در ۲۰۰ تکرار شبیه‌سازی میانگین برآورد پارامترها بسیار نزدیک به مقدار واقعی آنها بوده و دارای RMSE کوچک هستند.

جدول ۱: نتایج شبیه سازی ۱

$\phi_4$	$\phi_3$	$\phi_2$	$\phi_1$	$\tau$	Error
. / ۱	. / ۱	- / ۰.۵	. / ۳	true	$N(0, 1)$
. / ۱۰۱۲	- / ۱۱۸۹	- / ۰.۵۰۹	. / ۲۸۹۸	mean	. / ۲۵
. / ۱۲۶۶	. / ۱۲۹۶	. / ۱۳۴۴	. / ۱۳۵۴	RMSE	
- / ۱۶۱۱	- / ۱۴۶۸	- / ۲۹۵۱	. / ۰۱۵۹	95% C.L	
. / ۳۵۵۸	. / ۳۳۶۲	. / ۲۲۶۱	. / ۵۴۶۸	95% C.U	
. / ۰.۸۴۸	- / ۰.۲۲	- / ۰.۷۸۴	. / ۲۸۴۸	mean	. / ۵
. / ۱۲۶۵	. / ۱۱۸۹	. / ۱۲۱۳	. / ۱۱۷۳	RMSE	
- / ۱۸۳۳	- / ۱۲۷۷	- / ۳۰۸۷	. / ۰.۴۰۵	95% C.L	
. / ۳۱۴۱	. / ۳۲۰۷	. / ۱۶۱۳	. / ۵۰۲۴	95% C.U	
. / ۰.۸۱۰	- / ۰.۸۶۶	- / ۰.۶۱۳	. / ۳۰۰۱	mean	. / ۷۵
. / ۱۳۶۸	. / ۱۴۵۳	. / ۱۳۶۷	. / ۱۳۸۹	RMSE	
- / ۱۶۲۹	- / ۱۸۷۰	- / ۳۱۳۱	. / ۰.۲۱۹	95% C.L	
. / ۳۵۵۴	. / ۳۶۴۶	. / ۲۰۵۵	. / ۵۴۷۰	95% C.U	
. / ۱۰۲۱	- / ۱۱۹۶	- / ۰.۵۵۶	. / ۳۰۳۸	mean	. / ۲۵ Laplace(۰.۱)
. / ۰.۹۸۱	. / ۱۰۲۹	. / ۰.۸۶۴	. / ۱۰۳۸	RMSE	
- / ۰.۸۵۵	- / ۰.۹۷۳	- / ۲۲۶۰	. / ۰.۹۲۴	95% C.L	
. / ۳۰۰۹	- / ۳۲۱۱	. / ۰.۹۰۴	. / ۵۰۲۸	95% C.U	
. / ۰.۹۵۴	- / ۰.۹۳۹	- / ۰.۵۳۶	. / ۲۹۷۹	mean	. / ۵
. / ۰.۷۸۲	. / ۰.۸۵۸	. / ۰.۸۴۴	. / ۰.۸۶۰	RMSE	
- / ۰.۷۵۰	- / ۰.۸۴۷	- / ۲۴۳۹	. / ۱۲۰۳	95% C.L	
. / ۲۲۹۴	- / ۲۴۴۰	. / ۱۱۳۷	. / ۴۶۷۳	95% C.U	
. / ۰.۷۸۴	- / ۰.۹۷۳	- / ۰.۶۱۸	. / ۲۹۵۱	mean	. / ۷۵
. / ۱۰۴۵	. / ۱۰۶۱	. / ۰.۹۷۷	. / ۱۱۶۶	RMSE	
- / ۱۳۰۱	- / ۱۳۳۳	- / ۲۹۰۰	. / ۰.۵۷۸	95% C.L	
. / ۲۹۸۵	- / ۲۸۸۴	. / ۱۱۱۲	. / ۵۳۳۲	95% C.U	
. / ۰.۹۲۴	. / ۱۰۳۲	- / ۰.۶۲۳	. / ۲۹۹۳	mean	. / ۲۵ $t(3)$
. / ۱۰۰۰	- / ۰.۹۷۸	. / ۰.۹۵۳	. / ۰.۹۷۸	RMSE	
- / ۱۲۳۲	- / ۰.۹۶۲	- / ۲۳۴۷	. / ۰.۸۳۳	95% C.L	
. / ۳۱۸۴	. / ۲۹۶۲	. / ۱۵۲۲	. / ۴۸۰۰	95% C.U	
. / ۰.۸۱۰	. / ۱۰.۸۴	- / ۰.۵۵۹	. / ۲۹۸۳	mean	. / ۵
. / ۰.۸۷۵	- / ۰.۸۷۸	. / ۰.۸۴۸	. / ۰.۸۴۲	RMSE	
- / ۰.۹۱۲	- / ۰.۷۱۵	- / ۲۱۳۹	. / ۰.۹۶۴	95% C.L	
. / ۲۳۰۱	. / ۳۱۰۶	. / ۱۳۰۱	. / ۴۵۱۵	95% C.U	
. / ۰.۸۵۲	. / ۰.۹۰۶	- / ۰.۴۲۴	. / ۲۹۳۵	mean	. / ۷۵
. / ۰.۹۷۹	- / ۰.۹۶۸	. / ۰.۹۹۲	. / ۰.۹۱۶	RMSE	
- / ۱۳۵۶	- / ۰.۹۵۸	- / ۲۴۲۰	. / ۰.۸۲۷	95% C.L	
. / ۲۷۳۳	. / ۲۸۶۲	. / ۱۷۱۱	. / ۴۸۵۰	95% C.U	



جدول ۲: نتایج شبیه‌سازی ۲

$\phi_2$	$\phi_1$	$\tau$	Error
. / ۴	. / ۴	true	$N(0, 1)$
. / ۳۷۹۳	. / ۴۰۵۷	mean	. / ۲۵
. / ۱۴۳۰	. / ۱۵۱۳	RMSE	
. / ۰۸۸۳	. / ۱۱۰۶	95% C.L	
. / ۶۳۷۱	. / ۶۸۶۷	95% C.U	
. / ۳۸۳۳	. / ۴۰۰۱	mean	. / ۵
. / ۱۱۱۵	. / ۱۱۸۸	RMSE	
. / ۱۷۳۰	. / ۱۵۶۱	95% C.L	
. / ۵۸۰۸	. / ۶۱۲۰	95% C.U	
. / ۳۷۶۳	. / ۳۸۸۳	mean	. / ۷۵
. / ۱۳۷۰	. / ۱۴۹۳	RMSE	
. / ۱۰۳۵	. / ۰۹۳۶	95% C.L	
. / ۶۱۸۰	. / ۶۵۷۳	95% C.U	
. / ۳۹۴۸	. / ۴۰۲۶	mean	. / ۲۵ Laplace(.۱)
. / ۰۹۲۸	. / ۱۰۱۹	RMSE	
. / ۲۱۹۰	. / ۲۰۰۶	95% C.L	
. / ۵۷۸۵	. / ۵۹۶۸	95% C.U	
. / ۳۸۶۱	. / ۳۹۴۹	mean	. / ۵
. / ۰۷۱۱	. / ۰۷۱۶	RMSE	
. / ۲۲۹۳	. / ۲۴۵۱	95% C.L	
۵۰۴۰	. / ۵۳۹۰	95% C.U	
. / ۳۸۶۸	. / ۴۰۸۱	mean	. / ۷۵
. / ۱۰۹۳	. / ۱۰۵۴	RMSE	
. / ۱۲۳۷	. / ۲۰۵۸	95% C.L	
. / ۶۰۵۴	. / ۶۳۰۶	95% C.U	
. / ۳۷۵۱	. / ۴۰۳۰	mean	. / ۲۵ $t(3)$
. / ۱۰۹۵	. / ۰۹۶۹	RMSE	
. / ۱۳۴۸	. / ۲۰۱۱	95% C.L	
. / ۵۵۹۸	. / ۵۹۱۲	95% C.U	
. / ۳۹۵۳	. / ۳۸۷۸	mean	. / ۵
. / ۰۷۷۴	. / ۰۷۴۲	RMSE	
. / ۲۱۳۸	. / ۲۲۷۷	95% C.L	
. / ۵۴۳۱	. / ۵۳۱۳	95% C.U	
. / ۳۹۰۸	. / ۴۰۹۹	mean	. / ۷۵
. / ۱۰۳۹	. / ۰۹۹۰	RMSE	
. / ۱۶۵۴	. / ۲۲۲۰	95% C.L	
. / ۶۰۵۴	. / ۶۰۳۰	95% C.U	

برای مقایسه‌ی کارایی روش‌های عنوان شده از معیار خطای مدل (ME) که به صورت

$$ME(\hat{\Phi}) = (\hat{\Phi} - \Phi^{true})' (\hat{\Phi} - \Phi^{true}),$$

تعریف می‌شود، استفاده می‌کنیم. در واقع مقدار  $ME(\hat{\Phi})$  را ۲۰۰ مرتبه تحت هر مدل محاسبه کرده و سپس متوسط مقدار و انحراف استاندارد آنها به عنوان معیار مقایسه گزارش می‌شوند. واضح است که هر چه مقدار ME تحت هر مدل کمتر باشد آن مدل از کارایی بیشتری برخوردار است. با توجه به جدول ۳ مشخص است که مقادیر میانگین خطاهای مدل و انحراف استاندارد خطاهای مدل در چندک‌های ۰/۲۵، ۰/۵ و ۰/۷۵ برای روش SEM کمتر از روش QA است که این حاکی از کارایی بالای مدل پیشنهادی است.

جدول ۳: مقادیر ME و sd تحت مدل‌های مختلف

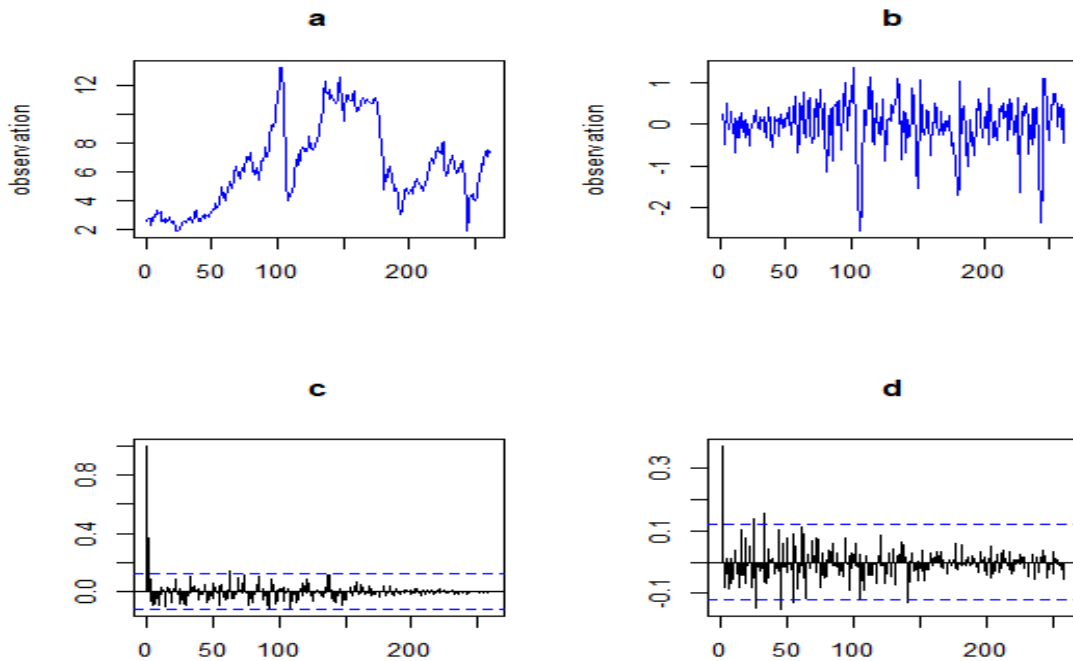
$SEM_{0.75}$	$SEM_{0.5}$	$SEM_{0.25}$	$QA_{0.75}$	$QA_{0.5}$	$QA_{0.25}$	Error	
۰/۰۷۷۸	۰/۰۵۸۶	۰/۰۶۹۲	۰/۰۸۸۸	۰/۰۷۰۲	۰/۰۷۵۸	ME	$N(0, 1)$
۰/۰۵۳۴	۰/۰۴۲۹	۰/۰۴۶۵	۰/۰۶۲۲	۰/۰۵۱۴	۰/۰۵۱۵	sd	شبیه‌سازی ۱
۰/۰۴۵۳	۰/۰۲۸۰	۰/۰۳۸۵	۰/۰۴۹۲	۰/۰۳۰۶	۰/۰۴۱۰	ME	Laplace(0, 1)
۰/۰۴۱۱	۰/۰۲۴۳	۰/۰۳۳۳	۰/۰۴۴۸	۰/۰۲۷۸	۰/۰۳۴۶	sd	
۰/۰۳۷۲	۰/۰۲۹۶	۰/۰۳۸۲	۰/۰۴۰۶	۰/۰۳۴۴	۰/۰۴۱۰	ME	$t(3)$
۰/۰۳۰۲	۰/۰۲۵۹	۰/۰۳۰۷	۰/۰۳۳۰	۰/۰۲۹۹	۰/۰۳۳۹	sd	
۰/۰۴۱۰	۰/۰۲۶۵	۰/۰۴۳۳	۰/۰۴۴۳	۰/۰۳۰۴	۰/۰۴۵۷	ME	$N(0, 1)$
۰/۰۴۳۹	۰/۰۲۸۷	۰/۰۴۳۴	۰/۰۴۷۰	۰/۰۳۲۰	۰/۰۴۵۸	sd	شبیه‌سازی ۲
۰/۰۲۳۰	۰/۰۱۰۱	۰/۰۱۹۰	۰/۰۲۳۷	۰/۰۱۰۵	۰/۰۱۹۳	ME	Laplace(0, 1)
۰/۰۲۲۵	۰/۰۱۳۲	۰/۰۲۰۵	۰/۰۲۳۴	۰/۰۱۴۲	۰/۰۲۰۷	sd	
۰/۰۲۰۶	۰/۰۱۱۵	۰/۰۲۱۴	۰/۰۲۱۸	۰/۰۱۳۵	۰/۰۲۲۴	ME	$t(3)$
۰/۰۲۰۷	۰/۰۱۳۶	۰/۰۲۴۳	۰/۰۲۲۹	۰/۰۱۷۷	۰/۰۲۵۸	sd	

## ۵ تحلیل داده‌های واقعی

در این بخش هدف برآزش مدل‌های عنوان شده به داده‌های واقعی و مقایسه عملکرد پیش‌بینی در آنها است. مجموعه داده‌ها با حجم ۲۶۱، قیمت نفت خام برنت اروپا بر حسب دلار را به صورت ماهانه از تاریخ ۱-۱-۲۰۰۰ تا ۹-۱۲-۲۰۲۱ نشان می‌دهد، که از سایت *Federal Reserve Economic Data* گرفته شده است. داده‌ها را ابتدا بر عدد ۱۰ تقسیم می‌کنیم. نمودار سری زمانی مشاهدات در شکل ۱(a) آمده است. مشخص است که داده‌ها دارای روند بوده و لذا سری ایستا نیست، برای ایستایی یک بار مشاهدات را تفاضلی می‌کنیم که نمودار آن در شکل ۱(b) نشان داده شده است و حاکی از ایستایی سری می‌باشد. همچنین شکل ۱(c) و ۱(d) به ترتیب نمودارهای توابع خودهمبستگی و خودهمبستگی جزئی مشاهدات را نشان می‌دهد. با توجه به این نمودارها مشخص است که مشاهدات دارای مدل اتورگرسیو هستند. با استفاده از تابع SelectModel در بسته FitAR در نرم افزار R مرتبه‌ی مدل اتورگرسیو را برای مشاهدات به کمک معیار اطلاع بیزی تعیین می‌کنیم، با این معیار مشخص می‌شود که مجموعه داده‌ها از مدل  $AR(1)$  پیروی می‌کند. بنابراین مدل اتورگرسیو مرتبه‌ی اول را برای داده‌های قیمت نفت برنت خام اروپا به صورت زیر در نظر می‌گیریم،

$$y_t = \phi_0 + \phi_1 y_{t-1} + \varepsilon_t, \quad (1.5)$$

که در آن فرض می‌کنیم  $\varepsilon_t$  دارای توزیع  $AL(0, \sigma, \tau)$  است.

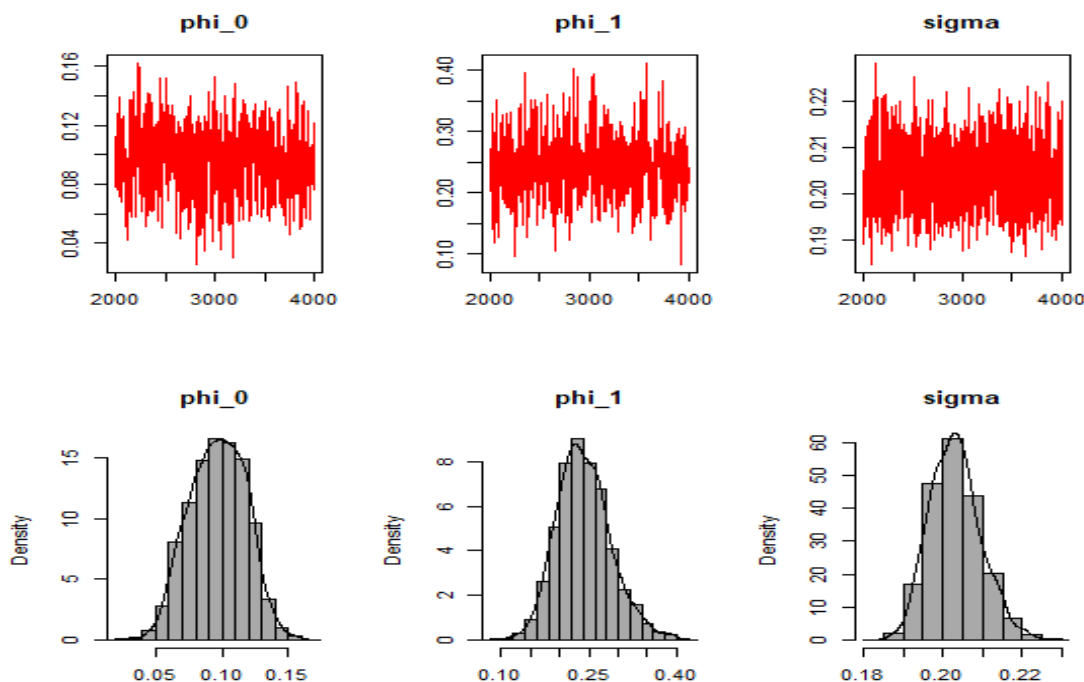


شکل ۱: ردیف اول نمودارهای سری زمانی داده‌های قیمت نفت خام برنت اروپا قبل و بعد از ایستاسازی و ردیف دوم نمودارهای ACF و PACF را نشان می‌دهد.

برای برازش داده‌ها به مدل پیشنهادی در چندک‌های  $0/25$ ،  $0/5$  و  $0/75$  مراحل S و M را  $4000$  مرتبه تکرار می‌کنیم سپس  $2000$  نمونه اول را حذف می‌کنیم و  $2000$  نمونه‌ی بعدی را برای استنباط در مورد پارامترها استفاده می‌کنیم. برای همه‌ی پارامترها تقریباً بعد از  $2000$  نمونه همگرایی مطلوب رخ می‌دهد که شکل ۲ ردیف اول نمودار مسیر پارامترها در چندک  $0/5$  را نشان می‌دهد. نمودار هیستوگرام و نمودار تابع چگالی بر اساس  $2000$  نمونه‌ی موثر در چندک  $0/5$  در ردیف دوم شکل ۲ ارائه شده است. این نمودار به ما یک توصیف بصری از توزیع نمونه‌های موثر برآورد پارامترها می‌دهد. برای اجتناب از طولانی شدن، نمودارها فقط برای چندک  $0/5$  نمایش داده شده‌اند. نتایج برآورد داده‌های نفت خام اروپا در چندک‌های  $0/25$ ،  $0/5$  و  $0/75$  برای مدل پیشنهادی در جدول ۴ گزارش شده است.

جدول ۴: مقادیر پارامترهای برآورد شده، انحراف استاندارد و فاصله‌ی اطمینان ۹۵ درصدی برای داده‌های قیمت نفت خام برنت اروپا در چندک‌های  $0/25$ ،  $0/5$  و  $0/75$

$\sigma$	$\phi_1$	$\phi_0$		$\tau$
$0/1757$	$0/3420$	$-0/3226$	mean	$0/25$
$0/0053$	$0/0273$	$0/0234$	sd	
$0/1655$	$0/3670$	$-0/3708$	95%C.L	
$0/1868$	$0/5121$	$-0/2781$	95%C.U	
$0/2347$	$0/2410$	$0/0967$	mean	$0/5$
$0/0063$	$0/0466$	$0/0211$	sd	
$0/1916$	$0/1584$	$0/0571$	95%C.L	
$0/2156$	$0/3442$	$0/1361$	95%C.U	
$0/1511$	$0/2514$	$0/3588$	mean	$0/75$
$0/0046$	$0/0247$	$0/0159$	sd	
$0/1425$	$0/1835$	$0/3287$	95%C.L	
$0/1607$	$0/3185$	$0/3913$	95%C.U	



شکل ۲: ردیف اول نمودارهای مسیر پارامترها در ۲۰۰۰ نمونه موثر در چندک ۰/۵ و ردیف دوم نمودارهای تابع چگالی و هیستوگرام ۲۰۰۰ نمونه موثر در چندک ۰/۵ برای داده‌های قیمت نفت خام برنت اروپا نشان می‌دهد.

جهت ارزیابی کارایی روش‌ها از پیش‌بینی بازگشتی خارج نمونه‌ای با مبدا زمانی  $t_0$  برای هر روش استفاده می‌کنیم (روش پیشنهادی در این مقاله با ایده بیان شده توسط [۱۰]). یعنی ابتدا مدل را به بخشی از داده‌ها برازش می‌دهیم ( $n=220$ ) و ۲ گام بعدی را پیش‌بینی می‌کنیم سپس ۲ گام به جلو رفته و قرار می‌دهیم ( $n=222$ ) و ۲ گام بعدی را پیش‌بینی می‌کنیم، این روند را تا زمان رسیدن به زمان  $t = T$  ( $n=260$ ) یعنی ادامه می‌دهیم (همین روند را برای پیش‌بینی ۴ گام نیز انجام می‌دهیم). در واقع با این روند ۴۰ مورد پیش‌بینی تحت هر روش انجام می‌شود. مقادیر متوسط قدر مطلق خطای پیش‌بینی (MAP) و انحراف استاندارد میانگین خطای پیش‌بینی (Sd) را در چندک‌های ۰/۲۵، ۰/۵ و ۰/۷۵ برای روش (QA)، همچنین روش پیشنهادی به صورت

$$\text{MAP} = \frac{1}{N} \sum_{i=1}^N |\hat{y}_{t_0+i} - y_i| = \frac{1}{N} \sum_{i=1}^N |\hat{e}_{t_0+i}|,$$

محاسبه می‌شود، که در آن  $N$  تعداد پیش‌بینی‌ها می‌باشد. واضح است که هر چه مقادیر MAP و Sd تحت یک مدل کمتر باشد، کارایی آن مدل بیشتر است. مقادیر MAP و Sd برای مدل‌های مختلف در جدول ۵ گزارش شده است. با توجه به این جدول مشاهده می‌شود که مقادیر MAP و Sd روش SEM هر چند در چندک ۰/۷۵ عملکرد تقریباً مشابهی با روش QA داشته است اما در سایر سطوح چندکی روش SEM بهتر از روش QA عمل کرده است.

جدول ۵: مقادیر ME و sd تحت مدل‌های مختلف

$SEM_{0.75}$	$SEM_{0.5}$	$SEM_{0.25}$	$QA_{0.75}$	$QA_{0.5}$	$QA_{0.25}$		
۰/۰۵۶۰	۰/۰۴۷۶	۰/۰۶۶۴	۰/۰۵۵۸	۰/۰۴۷۸	۰/۰۶۶۶	MAP	h=۴
۰/۰۵۱۵	۰/۰۴۹۵	۰/۰۴۲۰	۰/۰۵۹۵	۰/۰۴۹۶	۰/۰۴۲۰	SDP	
۰/۰۵۱۹	۰/۰۴۵۶	۰/۰۶۰۵	۰/۵۱۷	۰/۰۴۵۹	۰/۰۶۰۷	MAP	h=۲
۰/۰۵۵۶	۰/۰۴۷۳	۰/۰۴۴۲	۰/۰۵۵۶	۰/۰۴۷۴	۰/۰۴۳۶	SDP	

## ۶ نتیجه گیری

مدل‌های اتورگرسیو یکی از مدل‌های پرکاربرد و با اهمیت در تجزیه و تحلیل سری‌های زمانی است. مطالعات فراوانی بر مبنای این مدل‌ها صورت گرفته است. اما در اغلب این مطالعات همواره فرض بر این بوده که خطاها دارای توزیع مشخص می‌باشند. در واقع با توجه به این که در مدل‌های اتورگرسیو، متغیر  $y_t$  روی گذشته خود در  $p$  گام قبلی رگرسیون شده است، می‌توان با رویکرد رگرسیون میانگین به تجزیه و تحلیل و استنباط در مورد پارامترهای مدل اتورگرسیو پرداخت. اما همانگونه که اشاره شد این مدل‌ها دارای محدودیت‌هایی از جمله حساس بودن به نقاط دورافتاده و ناپایداری واریانس هستند. لذا مدل اتورگرسیو چندکی که توزیع مشاهدات را در چندک‌های مختلفی بررسی می‌کنند، نسبت به داده‌های دورافتاده حساس نیستند و نیازی به شرط نرمال بودن ندارند، به عنوان یک جایگزین مناسب برای مدل‌های مبتنی بر میانگین، در تحلیل سری زمانی می‌باشند. در این مقاله مدل اتورگرسیو چندکی با فرض توزیع AL برای خطاها معرفی گردید. سپس با استفاده از الگوریتم SEM پارامترهای مدل برآورد شدند. این الگوریتم شامل دو مرحله است. در مرحله S توزیع شرطی متغیر پنهان را به شرط مشاهدات شبیه‌سازی کرده و در مرحله M، لگاریتم تابع درستنمایی داده‌های کامل را نسبت به پارامترها ماکسیمم می‌کنیم. برای مقایسه‌ی کارایی مدل پیشنهادی در مقایسه با مدل اتورگرسیو چندکی کلاسیک در [۱۰]، مطالعه شبیه‌سازی انجام شده است. با استفاده از مطالعات شبیه‌سازی نشان داده شد مدل پیشنهادی علاوه بر این که نسبت به توزیع خطاهای مختلف استوار است از لحاظ معیار خطای مدل نیز کارایی بیشتری نسبت به مدل اتورگرسیو چندکی کلاسیک دارد. در قسمت تحلیل داده‌های واقعی توانایی پیش‌بینی مدل در مقایسه با مدل اتورگرسیو چندکی کلاسیک، با استفاده از معیارهای ارزیابی بررسی شد. نتایج به‌دست آمده حاکی از این است که مدل پیشنهادی در مقایسه با مدل اتورگرسیو چندکی کلاسیک در سطوح چندکی متفاوت به پیش‌بینی دقیق‌تری می‌انجامد. در نظر گرفتن پارامترهای توان بر روی ضرایب اتورگرسیو و انتخاب متغیر برای مدل اتورگرسیو چندکی می‌تواند هدف مطالعه آینده باشد. همچنین تحلیل مدل اتورگرسیو چندکی ترکیبی بر اساس توزیع AL می‌تواند یک زمینه پژوهشی مناسب جهت ارائه مدل جامع‌تر در آینده باشد.

## تقدیر و تشکر

نویسنده این مقاله از زحمات سردبیر محترم و دبیر تخصصی محترم و همچنین داور محترم که با نظرات و پیشنهادات ارزشمندشان باعث بهبود کیفیت مقاله شدند، کمال تشکر و قدردانی را دارد.

## فهرست منابع

- [1] Barndorff O., Nielsen E., *Non-Gaussian Ornstein-Uhlenbeck Based Models some of their uses in financial Economics*, J. R. Stat. Soc., B: Stat. Methodol., **63** (2001), 167–241.
- [2] Celeux G., Diebolt J., *The SEM algorithm: A probabilistic teacher algorithm derived from the EM algorithm for the mixture problem*, Computational Statistics Quarterly, **2** (1985), 73–82.
- [3] Dempster A., Lird N., Rubin D., *Maximum Likelihood from Incomplete Data via the EM algorithm*, J. R. Stat. Soc., B: Stat. Methodol., **39** (1977), 1–38.

- [4] Engle R.F., Manganelli S., *CAViaR: Conditional autoregressive value at risk by regression quantiles*, J. Bus. Econ. Stat., **22** (2004), 367–381.
- [5] Ferreira G., Luis M.C., Lachos V.H., *Bayesian modeling of autoregressive partial linear models with scale mixture of normal errors*, J. Appl. Stat., **40** (2013), 1796–1816.
- [6] Ghasemzadeh S., Ganjali M., Baghfalaki T., *Quantile regression via the EM algorithm for joint modeling of mixed discrete and continuous data based on Gaussian copula*, Stat. Methods Appl., **31** (2022), 1–22.
- [7] Hajrajabi A., Fallah A., *Classical and Bayesian Estimate of the AR(1) Model with Skew Symmetric innovations*, J. Iran. Stat. Soc., **18** (2019), 157–175.
- [8] Hajrajabi A., Maleki M., *Nonlinear Semiparametric autoregressive model with finite mixtures of scale mixtures of skew normal innovations*, J. Appl. Stat., **46** (2019), 2010–2029.
- [9] Koenker R., Bassett J., *Regression Quantiles*, Econometrica, **46** (1978) 33–50.
- [10] Koenker R., Xiao Z., *Quantile autoregression*, J. Am. Stat. Assoc., **101** (2006), 980–990.
- [11] Kozumi H., Kobayashi G., *Gibbs sampling methods of Bayesian quantile regression*, J. Stat. Comput. Simul., **81** (2011), 1565–1578.
- [12] Li G., Li Y., Tsai C.L., *Quantile correlations and quantile autoregressive modeling*, J. Am. Stat. Assoc., **110** (2015), 246–261
- [13] Liu J., Kumar S., Palomar D.P., *Parameter estimation of Heavy-tailed AR model with Missing Data via Stochastic EM algorithm*, IEEE Trans. Signal Process., **67** (2019), 2159–2172.
- [14] Morán-Vásquez R.A., Mazo-Lopera M.A., Ferrari S.L., *Quantile modeling through multivariate log-normal/independent linear regression models with application to newborn data*, Biom. J., **63** (2021), 1290–1308.
- [15] Nduka U.C., *EM-based algorithms for autoregressive models with t-distributed innovations*, Commun. Stat. - Simul. Comput., **47** (2018), 206–228.
- [16] Tian Y., Tang M., Zang Y., Tian M., *Quantile regression for linear models with autoregressive errors using EM algorithm*, Comput. Stat., **33** (2018), 1605–1625.
- [17] Tao Y., Yin J., *Maximum likelihood estimation for quantile autoregression models with Markovian switching*, Commun. Stat. Theory Methods, (2022), 1–29.
- [18] Yu K., Moyeed R.A., *Bayesian quantile regression*, Stat. Probab. Lett., **54** (2001), 437–447.
- [19] Zhou Y.H., Ni Z.X., Li Y., *Quantile Regression via the EM Algorithm*, Commun. Stat. - Simul. Comput., **43** (2014), 2162–2172.
- [20] Zeng Z., Li M., *Bayesian median autoregression for robust time series forecasting*, Int. J. Forecast., **37** (2021), 1000–1010.



## Linear Quantile autoregressive model estimation using Stochastic EM algorithm

Mohammad Bahmani<sup>†</sup>

Department of Statistics, Faculty of Basic Science, Zanjan University, Zanjan, Iran

Communicated by: G. R. Mohtashami Borzadaran

Received: 2021/12/11

Accepted: 2022/8/13

**Abstract:** In this paper, the quantile autoregressive time series model is introduced and then the model parameters are estimated using the Stochastic EM algorithm, which is an iterative method to compute maximum likelihood estimates. The likelihood function for the quantile autoregressive model is constructed based on the asymmetric Laplace distribution and a scale mixture representation of this distribution is used to estimate the model parameters via Stochastic EM algorithm. The efficiency and application of the proposed method are illustrated by some simulation studies and analyzing a real dataset.

**Keywords:** Quantile autoregressive model, Stochastic EM algorithm, Asymmetric Laplace distribution.



©2022 Shahid Chamran University of Ahvaz, Ahvaz, Iran. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0 license) (<http://creativecommons.org/licenses/by-nc/4.0/>).

<sup>†</sup>Corresponding author.

E-mail addresses: [bahmanimo74@znu.ac.ir](mailto:bahmanimo74@znu.ac.ir) (M. Bahmani)