© Statistical Research and Training Center
مجله‌ی پژوهش‌های آماری ایران
J. Statist. Res. Iran 6 (2009): 57–71
دوره‌ی ۶، شماره‌ی ۱، بهار و تابستان ۱۳۸۸، صص ۵۷–۷۱

مدل اتورگرسیو ضریب-تابعی و کاربرد آن در پیش‌بینی قیمت نفت خام سنگین ایران

پروین جلیلی[†] و مجتبی خزایی[‡,*]

[†] بانک مرکزی جمهوری اسلامی ایران

[‡] دانشگاه شهید بهشتی

چکیده. مدل اتورگرسیو ضریب-تابعی یکی از مهم‌ترین مدل‌های ناپارامتری در تحلیل سری‌های زمانی است. انعطاف‌پذیری بالای این مدل در برازش به مشاهدات واقعی سبب کاربرد گسترده این مدل در بررسی‌های اقتصادی، آب‌شناسی و غیره شده است. مدل‌های پارامتری بسیاری مانند مدل اتورگرسیو، مدل اتورگرسیو آستانه‌ای و مدل اتورگرسیو نمائی به‌عنوان حالت‌های خاصی از مدل اتورگرسیو ضریب-تابعی به دست می‌آیند. در این مقاله ضمن آشنایی با این مدل، روش‌های معمول، برای برازش مدل، بررسی کفایت مدل و پیش‌بینی معرفی می‌شوند. یک روش پیش‌بینی خودگردان برای پیش‌بینی m گام بعد معرفی می‌شود که با استفاده از آن می‌توان علاوه بر پیش‌بینی نقطه‌ای، پیش‌بینی فاصله‌ای و توزیع نمونه‌ای پیش‌بین را نیز محاسبه کرد. با استفاده از مدل اتورگرسیو ضریب-تابعی، سری زمانی متوسط قیمت ماهانه‌ی نفت خام سنگین ایران از جولای ۱۹۹۴ تا دسامبر ۲۰۰۷ تحلیل شده است.

واژگان کلیدی. مدل اتورگرسیو با ضرایب تابعی؛ روش رگرسیون خطی موضعی؛ روش خودگردان؛ تابع هسته؛ پیش‌بینی قیمت نفت.

---

*نویسنده‌ی عهده‌دار مکاتبات

Archive of SID
www.SID.ir

۱ مقدمه

سری‌های زمانی از جمله مشاهداتی هستند که تقریباً در تمامی حوزه‌ها وجود داشته و تحلیل و بررسی آن‌ها یکی از زمینه‌های مهم در علم آمار و برخی علوم دیگر است. شاید بتوان گفت مهم‌ترین هدف در مطالعه‌ی سری‌های زمانی کمک به کشف مکانیزم حاکم بر سیستمی است که منجر به تولید مشاهدات می‌شود. در راستای نیل به این هدف از سالیان دراز تلاش‌های بسیاری صورت گرفته است، به خصوص در تولد و تکامل مدل‌های سری‌های زمانی خطی گاوسی می‌توان به یول (۱۹۲۷) در مدل‌بندی اتورگرسیو (AR) مشاهدات لکه‌های خورشیدی، تا کار باکس و جنکینز (۱۹۷۰) که به‌طور برجسته‌ای مدل‌بندی ARMA را هم از بعد نظری و هم از نظر روش‌شناسی تکامل بخشیدند، اشاره کرد. در هر حال مدل‌بندی رفتارهای غیر استاندارد دیگر مانند غیر نرمال بودن، دوره‌های غیر متقارن، روابط ناخطی بین متغیرهای تأخیری، خارج از محدوده‌ی مدل‌های سری‌های زمانی خطی است. چنین رفتارهایی در بسیاری از مشاهدات واقعی دیده می‌شوند. از جمله مشاهداتی که در این زمینه بسیار معروف هستند می‌توان به مشاهدات لکه‌های خورشیدی و مشاهدات گربه‌های سیاه‌گوش کانادایی (تانگ، ۱۹۹۰) اشاره کرد. خارج از محدوده‌ی روابط خطی تنوع روابط ناخطی بسیار زیاد است. پیشرفت‌های اولیه‌ی تحلیل سری‌های زمانی ناخطی روی بعضی روابط ناخطی پارامتری تمرکز یافت. در میان مدل‌های ارایه‌شده می‌توان به مدل‌های موفقی چون مدل دوخطی (گرنجر و اندرسون، ۱۹۷۸)، مدل اتورگرسیو نمائی (EXPAR) (هاگان و اوزاکی، ۱۹۸۱ و اوزاکی، ۱۹۸۲) و مدل‌های ARCH و GARCH (انگل، ۱۹۸۲ و بلرسلف، ۱۹۸۶) و مدل آستانه‌ای (تانگ، ۱۹۹۰) اشاره کرد.

در تحلیل سری‌های زمانی ناخطی وقتی که پیشینه‌ای از ساختار مشاهدات در دست نیست شناسایی یک مدل ناخطی پارامتری مناسب برای برازش به مشاهدات با توجه به گستردگی چنین روابط ناخطی می‌تواند بسیار سخت باشد. یک جایگزین مناسب برای مدل‌های ناخطی پارامتری، استفاده از مدل‌ها و روش‌های تحلیل ناپارامتری است. این روش‌ها علاوه بر این که شیوه‌ی مستقلی برای تحلیل سری‌های زمانی فراهم می‌کنند، می‌توانند به درک و تشخیص بهتری از یک مدل پارامتری مناسب نیز منجر شوند. همچنین در مواجهه با سری‌های درازمدت با پیچیدگی‌های زیاد، روش‌های ناپارامتری توانایی تشخیص و تطبیق با ساختار مناسب و پیش‌بینی با دقت قابل قبولی را دارند. البته استفاده از مدل‌های ناپارامتری عام، زمانی که تعداد متغیرهای تأخیری و یا سری‌های کمکی بیش‌تر از دو تا باشند نیاز به مشاهدات بسیار زیادی دارند، این مشکل تحت عنوان مسئله‌ی بلای بعد زیاد توسط بل‌من (۱۹۶۱) مطرح شد. برای غلبه بر این مشکل محدودیت‌هایی را در مدل ناپارامتری عام اعمال می‌کنند که سبب به وجود آمدن مدل‌های مفیدی مانند

www.SID.ir

مدل جمعی (هاستی و تیبشیرانی، ۱۹۹۰) و مدل رگرسیون ضریب-تابعی (چن و تسی، ۱۹۹۳ و کای و همکاران، ۲۰۰۰) شده است. این مدل‌ها ضمن غلبه بر مشکل بلای بعد زیاد، انعطاف‌پذیری قابل قبولی در برازش به مشاهدات واقعی دارند. در این مقاله با تاسی از چن و تسی (۱۹۹۳) و کای و همکاران (۲۰۰۰) مدل اتورگرسیو ضریب-تابعی (Functional−coefficient autoregressive)(FAR) معرفی و از روش رگرسیون خطی موضعی برای برازش این مدل استفاده شده است. برای این منظور در بخش دوم مدل اتورگرسیو ضریب-تابعی معرفی و در بخش سوم، نحوه‌ی برازش مدل با استفاده از روش رگرسیون خطی موضعی ارایه شده است. در بخش چهارم، ملاکی برای تعیین پارامترهای مدل معرفی شده است. در بخش پنجم، روش‌هایی برای پیش‌بینی، به خصوص یک روش پیش‌بینی خودگردان برای پیش‌بینی m گام بعد سری زمانی معرفی شده است که با استفاده از آن می‌توان علاوه بر پیش‌بینی نقطه‌ای، پیش‌بینی فاصله‌ای و توزیع نمونه‌ای پیش‌بین را نیز محاسبه کرد. در انتها با استفاده از مدل اتورگرسیو ضریب-تابعی، متوسط قیمت ماهانه‌ی نفت خام سنگین ایران از جولای ۱۹۹۴ تا دسامبر ۲۰۰۷ تحلیل شده است.

۲ مدل اتورگرسیو ضریب-تابعی

یکی از مدل‌های پرکاربرد که انعطاف‌پذیری قابل قبولی در برازش به مشاهدات واقعی به خصوص در زمینه‌ی سری‌های زمانی دارد، مدل رگرسیون ضریب-تابعی است. از کاربردهای اخیر این مدل در تحلیل سری‌های زمانی می‌توان به مقاله‌ی هیوانگ و شن (۲۰۰۴) اشاره کرد که از این مدل در برازش به سری زمانی تولید ناخالص داخلی آمریکا استفاده شده است و مقاله‌ی ونگ و همکاران (۲۰۰۷) که این مدل نیز در برازش به مشاهدات طولی در پیش‌بینی‌های آب‌شناسی به کار رفته است.

این مدل به‌صورت زیر تعریف می‌شود:

$$Y_t = \sum_{j=1}^{p} a_j(\boldsymbol{U}_t) X_{jt} + \epsilon_t, \qquad (1)$$

که در آن $a_j(\cdot)$ها توابعی هموار (توابعی پیوسته، دارای مشتق حد اقل از مرتبه‌ی دوم) هستند و $\boldsymbol{U}_t = (U_{1t}, \ldots, U_{kt})'$ به متغیرهای مدل-وابسته معروف هستند. در عمل $k = 1$ یا حد اکثر $k = 2$ اختیار می‌شود به‌طوری که توابع $a_j(\cdot)$ بدون مواجهه با مسئله‌ی بلای بعد زیاد قابل براورد باشند.

در این مقاله مدل رگرسیون ضریب-تابعی در تحلیل سری زمانی تک‌متغیره بررسی می‌شود که برای این منظور با مدل اتورگرسیو ضریب-تابعی آشنا می‌شویم. این مدل اولین بار توسط چن و تسی (۱۹۹۳)

www.SID.ir

معرفی شد. آن‌ها یک الگوریتم تکراری، بر اساس رگرسیون موضعی مرتب‌شده را برای برازش مدل ارایه کردند و خواص احتمالی مدل را نیز به‌عنوان یک فرایند تصادفی مورد بررسی قرار دادند. کای و همکاران (۲۰۰۰) روش ساده‌تری جهت برازش مدل ارایه کرده و خواص مجانبی براوردگرهای خود را اثبات کردند. در ادامه، ضمن معرفی مدل FAR با روش پیشنهادی کای و همکاران برای برازش مدل و خواص مجانبی براوردگرهای حاصل آشنا می‌شویم. طی سال‌های اخیر تعمیم مدل FAR و روش برازش برای مطالعه‌ی رابطه‌ی بین چند سری زمانی انجام شده است. کاربردی از این مباحث را می‌توان در مقاله‌ی ونگ و همکاران (۲۰۰۷) دید.

مدل $\mathrm{FAR}(p,d)$ به‌صورت زیر تعریف می‌شود:

$$Y_t = a_1(Y_{t-d})Y_{t-1} + \cdots + a_p(Y_{t-d})Y_{t-p} + \epsilon_t, \qquad t = p+1, p+2, \ldots, T. \tag{۲}$$

که $\{\epsilon_t\}$ دنباله‌ای از متغیرهای تصادفی مستقل با میانگین صفر و واریانس σ^2 بوده و برای هر t، ϵ_t مستقل از $Y_{t-1}, Y_{t-2}, \ldots$ است.

مدل $\mathrm{FAR}(p,d)$ همان مدل (۱) به ازای

$$X_{1t} = Y_{t-1}, \ldots, X_{pt} = Y_{t-p}, \quad U_t = Y_{t-d}, \quad t = p+1, \ldots, T. \tag{۳}$$

است. مدل FAR مدل‌های پارامتری مهمی را به‌عنوان حالت خاص شامل می‌شود. از مهم‌ترین آن‌ها یکی مدل اتورگرسیو آستانه‌ای است که به ازای $a_i(Y_{t-d}) = \phi_i^{(1)} I(Y_{t-d} \leqslant c) + \phi_i^{(2)} I(Y_{t-d} > c)$ و دیگری مدل اتورگرسیو نمائی که به ازای $a_i(Y_{t-d}) = \alpha_i + \beta_i \exp(-\gamma_i Y_{t-d}^2)$ نتیجه می‌شوند.

۳ برازش مدل

در این بخش با روش رگرسیون وزنی موضعی برای برازش مدل اتورگرسیو ضریب–تابعی آشنا می‌شویم. با توجه به اینکه $a_j(\cdot)$ها توابعی هموارند، می‌توان آن‌ها را به‌صورت موضعی در $Y_{t-d} = y_0$ با تابع خطی $a_j(y) \approx a_j + b_j(y - y_0)$ تقریب کرد. در این صورت داریم $a_j(y_0) \approx a_j$ و براورد a_j را می‌توان براورد $a_j(y)$ در نقطه‌ی y_0 در نظر گرفت یعنی $\hat{a}_j(y_0) = \hat{a}_j$. برای براورد a_jها، مدل (۲) را می‌توان به‌صورت زیر نوشت:

$$Y_t \approx \sum_{j=1}^{p} \{a_j + b_j(Y_{t-d} - y_0)\} Y_{t-j} + \epsilon_t, \qquad t = p+1, p+2, \ldots, T$$

www.SID.ir

و پارامترهای مدل را با روش کم‌ترین توان‌های دوم موزون با وزن‌هایی که توسط هسته‌ی مناسبی اعمال می‌شوند، براورد کرد. تابع هسته‌ی $K(u)$، تابعی پیوسته، متقارن حول صفر و نزولی برای $u > ۰$ تعریف می‌شود. این ویژگی‌ها باعث می‌شود که تابع هسته انتخاب مناسبی برای وزن باشد. در عمل از $K_h(u) = \frac{۱}{h}K(\frac{u}{h})$ استفاده می‌شود که در آن h، معروف به پهنای نوار است و اندازه‌ی آن اثر وزن‌ها را مشخص می‌کند. هسته‌ی گاوسی و اپانچنیکوف، به‌ترتیب با ضابطه‌های $(\frac{۱}{\sqrt{۲\pi}})\exp(-\frac{u^۲}{۲})$ و $(\frac{۳}{۴})(۱-u^۲)I(|u| \leqslant ۱)$، از معروف‌ترین و پرکاربردترین هسته‌ها هستند.

در بحث حاضر به ازای هسته‌ی مفروض $K(u)$، مجموع توان دوم موزون عبارت است از

$$\sum_{t=p+۱}^{T}\left[Y_t - \sum_{j=۱}^{p}\{a_j + b_j(Y_{t-d} - y_۰)\}Y_{t-j}\right]^۲ K_h(Y_{t-d} - y_۰).$$

با کمی محاسبات جبری می‌توان نشان داد

$$\hat{a}_j(y_۰) = \sum_{k=p+۱}^{T} K_{T,j}(Y_{k-d} - y_۰, \boldsymbol{Y}_k)Y_k, \tag{۴}$$

که در آن $\boldsymbol{Y}_k = (Y_{k-۱}, \ldots, Y_{k-p})'$،

$$K_{T,j}(y_{t-d}, \boldsymbol{y}) = \boldsymbol{e}'_{j,۲p}(\tilde{\boldsymbol{Y}}'W\tilde{\boldsymbol{Y}})^{-۱}\begin{pmatrix}\boldsymbol{y}\\ y_{t-d}\boldsymbol{y}\end{pmatrix}K_h(y_{t-d}), \tag{۵}$$

$\boldsymbol{e}_{j,۲p}$ یک بردار با $۲p$ مؤلفه است که مقدار jامین مؤلفه‌ی آن ۱ و سایر مولفه‌ها صفرند، $\tilde{\boldsymbol{Y}}$ یک ماتریس $(T-p)\times ۲p$ با سطرهای $(\boldsymbol{Y}'_t, \boldsymbol{Y}'_t(Y_{t-d} - y_۰))'$، $t = p+۱, \ldots, T$ و

$$W = \operatorname{diag}\{K_h(Y_{p+۱-d} - y_۰), \cdots, K_h(Y_T - y_۰)\}.$$

قضیه‌های مربوط به توزیع مجانبی این براوردگرها را می‌توان در کای و همکاران (۲۰۰۰) یافت.

۴ شناسایی مدل

همان‌طور که از ساختار مدل (۲) مشخص است یکی از نکات مهم در استفاده از این مدل تعیین مرتبه‌ی p و d است. از سوی دیگر، در بخش قبل ملاحظه شد که برای براورد ضرایب مدل از تابع هسته برای وزن‌دهی استفاده می‌شود، در نتیجه پهنای نوار h، به‌عنوان پارامتر هموارساز باید براورد شود چرا که اندازه‌ی پهنای نوار در میزان اریبی و واریانس براوردگرها نقش مهمی دارد. در این بخش با روشی برای تعیین مقادیر h، d و p آشنا می‌شویم.

Archive of SID
www.SID.ir

معیارهای اعتبارسنجی متقابل (Cross Validation) و اعتبارسنجی متقابل تعمیم‌یافته (Generalized Cross Validation) از معیارهای معمول در انتخاب پهنای نوار می‌باشند. اما چون این معیارها برای مشاهدات ناهمبسته گسترش یافته‌اند از اطلاع اضافی وجود همبستگی بین مشاهدات استفاده نمی‌کنند. معیارهایی نیز وجود دارد که از این اطلاع استفاده می‌کنند که از معروف‌ترین آن‌ها معیار میانگین خطای پیش‌بینی (Average Prediction Error) است. این معیار توسط کای و همکاران (۲۰۰۰) پیشنهاد شده است. این معیار را می‌توان به‌عنوان معیار اعتبارسنجی متقابل تعدیل‌شده‌ای در نظر گرفت که به ساختار مشاهدات سری‌های زمانی مانا نیز توجه می‌کند. در بیانی ساده معیار مورد نظر را می‌توان به‌صورت زیر تشریح کرد.

فرض کنید m و Q دو عدد صحیح مثبت معلوم باشند به قسمی که $n > mQ$. ایده‌ی اصلی به این صورت است که Q تا زیرسری، هر یک به‌طول $n-qm$، $q = ۱, ۲, \ldots, Q$ از ابتدای سری جدا می‌کنیم. برای هر یک از این زیرسری‌ها مدل FAR برازش می‌شود. سپس بر اساس مدل براورد شده‌ی خطاهای پیش‌بینی گام بعد ادامه‌ی سری زمانی، به‌طول m محاسبه می‌گردد. برای مثال فرض کنید $Q = ۴$ باشد، در این حالت از ۴ زیرسری از سری مورد نظر برای محاسبه‌ی معیار APE استفاده می‌شود. بنا بر این در ابتدا، مشاهدات به‌طول $n - ۴m$ مورد استفاده قرار می‌گیرد تا مدل برازش شود. خطای پیش‌بینی‌های گام بعد برای m مشاهده‌ی بعدی محاسبه می‌شود. سپس مشاهدات به‌طول $n - ۳m$ به‌منظور براورد توابع، مورد استفاده قرار می‌گیرند و خطای پیش‌بینی‌های گام بعد برای m مشاهده بعدی محاسبه می‌گردد و الی آخر. APE، میانگین توان‌های دوم این خطاها است. در حالت کلی APE تابعی از پارامترهای هموارساز است.

در مورد مدل اتورگرسیو با ضرایب تابعی، فرض کنید $\{\hat{a}_{j,q}(\cdot)\}$ ضرایب براورد شده با استفاده از qامین $(q = ۱, ۲, \ldots, Q)$ زیرسری $\{(Y_{t-d}, \boldsymbol{Y}_t, Y_t), p+۱ \leqslant t \leqslant T - qm\}$ به ازای پهنای نوار h باشد.

میانگین خطای پیش‌بینی برای qامین زیرسری به‌صورت زیر است:

$$\mathrm{APE}_q(h) = \frac{۱}{m} \sum_{t=T-qm+۱}^{T-qm+m} \left\{ Y_t - \sum_{j=۱}^{p} \hat{a}_{j,q}(Y_{t-d}) Y_{t-j} \right\}^۲ .$$

و در نهایت APE نیز به‌صورت زیر محاسبه می‌شود:

$$\mathrm{APE}(h) = \frac{۱}{Q} \sum_{q=۱}^{Q} \mathrm{APE}_q(h). \qquad (۶)$$

www.SID.ir

hی که مربوط به کم‌ترین مقدار $\mathrm{APE}(h)$ است، به‌عنوان پهنای نوار پیشنهاد می‌شود. در کاربرد عموماً $m = [۰/۱n]$ و $Q = ۴$ در نظر گرفته می‌شود.

تعیین مناسب d نیز بسیار اهمیت دارد. به این منظور اگر اطلاعات پیشینی در مورد مشاهدات وجود داشته باشد که به ما در انتخاب d کمک کند بسیار مفید خواهد بود. در غیر این صورت می‌توان از معیارهایی مانند اعتبارسنجی متقابل، APE و غیره بهره جست. به‌عنوان مثال فرض کنید $\mathrm{APE}(h,d)$ میانگین خطای پیش‌بینی تعریف‌شده در (۶) با استفاده از متغیر تأخیری Y_{t-d} باشد. یک روش عملی و ساده، حد اقل کردن $\mathrm{APE}(h,d)$ به‌طور همزمان برای h در یک دامنه‌ی معین و d روی مجموعه‌ی $\{۱,۲,\ldots,p\}$ است. APE را می‌توان به‌منظور تعیین مرتبه‌ی اتورگرسیو (p)، نیز با همین ساختار تعمیم داد.

۵ پیش‌بینی

پیش‌بینی به‌کمک مدل‌های سری‌های زمانی ناخطی به‌طور قابل ملاحظه‌ای مشکل‌تر از مدل‌های خطی است. در این بخش چند روش پیش‌بینی به‌کمک مدل‌های FAR را معرفی می‌کنیم. در این میان چگونگی پیش‌بینی گام بعد، پیش‌بینی چند گام بعد به‌صورت گام به گام و مستقیم معرفی می‌شوند. با توجه به مدل (۲) پیش‌بینی یک گام بعد به‌صورت زیر تعریف می‌شود:

$$\hat{Y}_{t+۱} = \hat{a}_۱(Y_{t+۱-d})Y_t + \cdots + \hat{a}_p(Y_{t+۱-d})Y_{t-p+۱}\,. \tag{۷}$$

با جایگزین کردن مقدار $\hat{Y}_{t+۱}$، در رابطه‌ی زیر، پیش‌بینی دو گام بعد به‌صورت گام به گام حاصل می‌شود.

$$\hat{Y}_{t+۲} = \hat{a}_۱(Y^*_{t+۲-d})\hat{Y}_{t+۱} + \hat{a}_۲(Y^*_{t+۲-d})Y_t + \cdots + \hat{a}_p(Y^*_{t+۲-d})Y_{t-p+۲}. \tag{۸}$$

که در آن

$$Y^*_{t+۲-d} = \begin{cases} Y_{t+۲-d}, & d > ۱ \\ \hat{Y}_{t+۱}, & d = ۱ \end{cases}$$

به‌همین ترتیب می‌توان پیش‌بینی‌های گام به گام برای گام‌های بعد را محاسبه کرد.

در شیوه‌ی دیگر برای پیش‌بینی دو گام بعد که به‌روش پیش‌بینی مستقیم معروف است، مدل FAR زیربرازش می‌شود.

$$Y_{t+۲} = b_۱(Y_{t+۱-d})Y_t + \cdots + b_p(Y_{t+۱-d})Y_{t-p+۱} + \epsilon'_t,$$

www.SID.ir

Archive of SID

در این حالت پیش‌بینی دو گام بعد عبارت است از مقدار حاصل از برازش،

$$\hat{Y}_{t+۲} = \hat{b}_۱(Y_{t+۱-d})Y_t + \cdots + \hat{b}_p(Y_{t+۱-d})Y_{t-p+۱}. \qquad (۹)$$

شبیه‌سازی‌های انجام‌شده توسط کای و همکاران (۲۰۰۰) نشان می‌دهند عملکرد روش اول بهتر از روش پیش‌بینی مستقیم است.

روش دیگری که در پیش‌بینی سری‌های ناخطی کاربرد زیادی دارد روش خودگردان است. این روش توسط هیوانگ و شن (۲۰۰۴) برای پیش‌بینی مدل FAR برازش‌شده با استفاده از هموارساز اسپلاین پیشنهاد شده است. در این مقاله از این روش برای پیش‌بینی مدل FAR برازش‌شده با استفاده از رگرسیون خطی موضعی استفاده می‌کنیم.

فرض کنید مشاهدات $\{Y_۱, \ldots, Y_T\}$، از یک مدل FAR حاصل شده‌اند و هدف پیش‌بینی Y_{T+k} است. برای این منظور، ابتدا مدل FAR را برازش داده مانده‌ها را حساب می‌کنیم. با استفاده از مدل برازش‌شده k مشاهده‌ی $Y_{T+۱}, \ldots, Y_{T+k}$ را به‌صورت زیر شبیه‌سازی می‌کنیم:

$$Y_{t+۱} = \hat{a}_۱(Y_{t+۱-d})Y_t + \cdots + \hat{a}_p(Y_{t+۱-d})Y_{t+۱-p} + \hat{\epsilon}_{t+۱}, \qquad t = T, \ldots, T+k-۱$$

که در آن $\hat{\epsilon}_{T+۱}, \ldots, \hat{\epsilon}_{T+k}$ یک نمونه‌ی تصادفی با جایگذاری از مانده‌ها است.

اگر این کار را B بار تکرار کنیم، پیش‌بینی k گام بعد، به‌صورت میانگین مقادیر شبیه‌سازی‌شده‌ی $\{Y_{T+k,۱}, \ldots, Y_{T+k,B}\}$ تعریف می‌شود. در این روش همچنین از سری‌های شبیه‌سازی‌شده می‌توان برای ساخت پیش‌بینی فاصله‌ای استفاده کرد. برای این منظور اگر به ازای $۰ < \alpha < ۱$، صدک نمونه‌ای $۱۰۰ \times \alpha$ام، $\{Y_{T+k,۱}, \ldots, Y_{T+k,B}\}$ با Y^{α}_{T+k} نشان داده شود، یک پیش‌بینی فاصله‌ای برای Y_{T+k} با ضریب اطمینان $۱۰۰(۱-\alpha)$٪ برابر است با $(Y^{\frac{\alpha}{۲}}_{T+k}, Y^{۱-\frac{\alpha}{۲}}_{T+k})$.

۶ کاربرد مدل FAR در مدل‌بندی مشاهدات قیمت نفت خام سنگین ایران

نفت به‌عنوان مهمترین کالای صادراتی ایران با در اختیار داشتن ۸۰ تا ۹۰ درصد از سهم صادرات کشور، نقش مهمی در اقتصاد ایران داشته و تغییرات قیمت و نوسانات آن، تاثیر مستقیم در اقتصاد کشور دارد. با توجه به عوامل متعددی که در قیمت نفت دخالت دارند، پیش‌بینی دقیق و بلندمدت آن کاری بسیار مشکل

www.SID.ir

جدول ۱. مقادیر ملاک APE، برای تعیین مرتبه‌ی p، d و مقدار h.

p	d	h	APE
۲	۱	۱/۳۰۱	۰/۰۳۴
۲	۲	۱/۲۵۲	۰/۰۳۵
۳	۱	۱/۴۰۱	۰/۰۳۴
۳	۲	۱/۶۰۱	۰/۰۳۴
۳	۳	۱/۶۰۱	۰/۰۳۴
۴	۱	۱/۶۰۲	۰/۰۳۷
۴	۲	۱/۶۰۲	۰/۰۳۶
۴	۳	۱/۶۰۲	۰/۰۳۶
۴	۴	۱/۶۰۲	۰/۰۳۵
۵	۱	۱/۶۰۲	۰/۰۴۵
۵	۲	۱/۶۰۲	۰/۰۴۳
۵	۳	۱/۶۰۲	۰/۰۴۳
۵	۴	۱/۶۰۲	۰/۰۴۱
۵	۵	۱/۶۰۲	۰/۰۴۰

و تا حدی، غیر ممکن به نظر می‌رسد. با توجه به این‌که قیمت نفت از رفتار منظمی پیروی نمی‌کند استفاده از روش‌های ناپارامتری که کم‌تر تحت تأثیر رفتارهای غیر استاندارد سری‌ها هستند مناسب به نظر می‌رسد.

در این بخش از مدل FAR و روش رگرسیون خطی موضعی برای مدل کردن سری زمانی متوسط قیمت نفت خام سنگین ایران از جولای ۱۹۹۴ تا دسامبر ۲۰۰۷ استفاده می‌کنیم.

نمودار این سری در شکل ۱$-(a)$، رسم شده است. همان‌طور که ملاحظه می‌شود در برخی از دوره‌ها مقدار تغییر قیمت نفت بسیار زیاد بوده و این مسئله سبب می‌شود تا تغییر قیمت در دوره‌های دیگر آن‌طور که باید دیده نشود به‌همین دلیل تبدیل لگاریتمی مناسب به نظر می‌رسد.

به‌منظور تعیین مرتبه‌ی p، d و مقدار h، از ملاک APE استفاده می‌کنیم . مقادیر APE به‌ازای مقادیر مختلف p، d و h در جدول ۱ ارایه شده است. لازم به ذکر است که در هر مرتبه از p و d، $\mathrm{APE}(h)$ در شبکه‌ای از مقادیر h محاسبه و h مربوط به کم‌ترین مقدار $\mathrm{APE}(h)$ و مقدار APE متناظر با آن در جدول درج شده است.

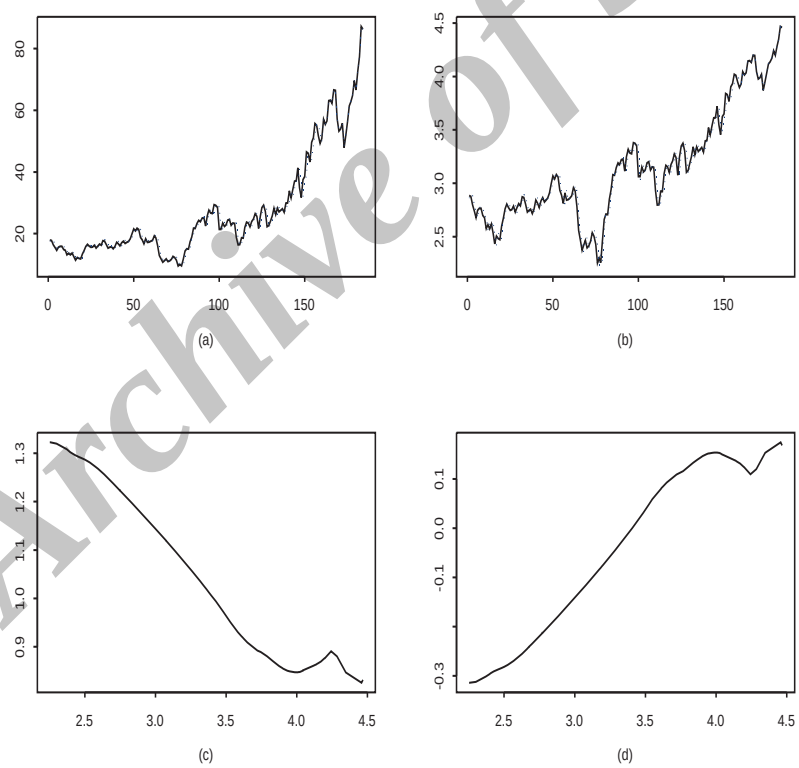
با توجه به نتایج جدول ۱ مدل $\mathrm{FAR}(۲,۱)$ با $h = ۱/۳$ را به‌عنوان مدل بهینه انتخاب می‌کنیم. نمودار توابع براورد شده توسط روابط (۴) و (۵) و استفاده از هسته‌ی اپانچنیکوف در شکل ۱$-(c)$ و

www.SID.ir

۱−(d) رسم شده‌اند. به‌کمک آن‌ها مقادیر برازش‌شده و پادلگاریتم آن‌ها محاسبه و در نمودارهای ۱−(b) و ۱−(a) به‌ترتیب با نقطه‌چین رسم شده‌اند.

مطالعه‌ی مانده‌ها جهت ارزیابی مدل برازش‌شده نیز انجام شد و مانده‌های مدل برازش‌شده از نظر وجود همبستگی و همبستگی جزئی، نرمال بودن و ثبات واریانس مورد بررسی قرار گرفتند که نتایج حاصل از ارزیابی مدل برازش‌شده و همچنین تطابق مقادیر برازش‌یافته و مقادیر واقعی لگاریتم مشاهدات در شکل ۱−(a)، حاکی از آن است که مدل برازش‌شده مدل خوبی است.

به‌منظور بررسی قابلیت پیش‌بینی مدل برازش‌شده، ۶ مشاهده‌ی آخر سری را کنار گذاشته و با استفاده



شکل ۱. (a) نمودار سری زمانی قیمت نفت خام سنگین ایران (خط) و نمودار مقادیر برازش‌یافته‌ی آن (نقطه‌چین)، (b) نمودار سری زمانی لگاریتم قیمت نفت خام سنگین ایران (خط) و نمودار مقادیر برازش‌یافته‌ی آن (نقطه‌چین)، با استفاده از مدل FAR(۲, ۱)، (c) و (d) نمودار براورد توابع a_1 و a_2

www.SID.ir

جدول ۲. قدر مطلق خطاهای پیش‌بینی و میانگین قدر مطلق خطاهای پیش‌بینی برای پیش‌بینی یک گام بعد، دو گام بعد به‌صورت گام به گام و مستقیم با استفاده از مدل FAR(۲, ۱)، برای ۶ مشاهده‌ی آخر لگاریتم قیمت نفت خام سنگین ایران

دوره	Y_t	یک گام	دو گام	مستقیم
۲۰۰۷/۷	۴/۲۴۳	۰/۰۶۶۳	۰/۱۰۳۳	۰/۱۰۴۴
۲۰۰۷/۸	۴/۱۹۹	۰/۰۵۳۴	۰/۰۱۴۸	۰/۰۱۵۲
۲۰۰۷/۹	۴/۲۸۵	۰/۰۸۴۲	۰/۰۳۰۰	۰/۰۳۲۵
۲۰۰۷/۱۰	۴/۳۴۷	۰/۰۴۸۱	۰/۱۲۷۱	۰/۱۱۷۷
۲۰۰۷/۱۱	۴/۴۶۹	۰/۱۱۳۳	۰/۱۶۲۱	۰/۱۶۹۱
۲۰۰۷/۱۲	۴/۵۷۹	۰/۰۴۰۵	۰/۰۹۲۷	۰/۰۸۳۳
میانگین		۰/۰۶۷۵	۰/۰۸۲۹	۰/۰۸۷۰

جدول ۳. قدر مطلق و میانگین قدر مطلق خطاهای پیش‌بینی خودگردان یک گام بعد، دو گام بعد و ۶ گام بعد برای ۶ مشاهده‌ی آخر لگاریتم قیمت نفت خام سنگین ایران به‌همراه کران پایین و بالای پیش‌بینی فاصله‌ای گام بعد خودگردان با ضریب ۹۵ درصد حاصل از مدل FAR(۲, ۱)

دوره	یک گام بعد	دو گام بعد	۶ گام بعد	کران پایین	کران بالا
۲۰۰۷/۷	۰/۰۶۶۵۶	۰/۰۹۵۲	۰/۰۶۸۶	۳/۹۷۳	۴/۳۴۰
۲۰۰۷/۸	۰/۰۵۲۹	۰/۰۱۳۵	۰/۰۱۵۱	۳/۹۳۰	۴/۴۰۶
۲۰۰۷/۹	۰/۰۸۰	۰/۰۱۸۰	۰/۰۹۱۷	۳/۸۷۶	۴/۴۹۲
۲۰۰۷/۱۰	۰/۰۵۴۱	۰/۱۳۹	۰/۱۴۳	۳/۸۳۰	۴/۵۶۹
۲۰۰۷/۱۱	۰/۱۰۵	۰/۱۵۹	۰/۲۵۰	۳/۸۰۲	۴/۶۶
۲۰۰۷/۱۲	۰/۰۴۱	۰/۰۷۸	۰/۲۲۸	۳/۷۶۸	۴/۷۵۳
میانگین	۰/۰۶۷	۰/۰۸۴	۰/۱۳۳		

از روابط (۷)، (۸) و (۹) پیش‌بینی این مشاهدات به‌صورت گام بعد، دو گام بعد به‌صورت گام به گام و مستقیم محاسبه می‌شوند. قدر مطلق خطای این پیش‌بینی‌ها و میانگین قدر مطلق خطای این پیش‌بینی‌ها محاسبه و در جدول ۲ ارایه شده‌اند. مقادیر کوچک میانگین قدر مطلق خطای پیش‌بینی‌ها تأکیدی بر قابلیت مدل برازش‌شده در پیش‌بینی است.

جهت ارزیابی عملکرد پیش‌بینی‌های حاصل از روش خودگردان و مقایسه‌ی آن‌ها با پیش‌بینی‌های فوق علاوه بر پیش‌بینی‌های گام بعد و دو گام بعد پیش‌بینی‌های شش گام بعد را نیز برای شش مشاهده‌ی کنار گذاشته‌شده به‌کمک نمونه‌های خودگردان $B = ۱۰۰۰$تائی محاسبه کرده‌ایم. قدر مطلق و میانگین قدر مطلق خطای این پیش‌بینی‌ها در جدول ۳ ارایه شده‌اند. مقایسه‌ی این مقادیر برای پیش‌بینی‌های یک و دو گام بعد

Archive of SID

www.SID.ir

جدول ۴. پیش‌بینی‌های خودگردان ۳ مشاهده‌ی آینده‌ی لگاریتم قیمت نفت خام سنگین ایران و قیمت نفت خام سنگین ایران به‌همراه پیش‌بینی‌های فاصله‌ای خودگردان ۹۵ درصد

دوره	پیش‌بینی لگاریتم قیمت نفت			پیش‌بینی قیمت نفت		
	کران بالا	کران پایین	۳ مشاهده‌ی بعدی	کران بالا	کران پایین	۳ مشاهده‌ی بعدی
۲۰۰۸/۱	۴/۶۴	۴/۲۸	۴/۴۸	۱۰۳/۷۴	۷۲/۱۶	۸۸/۱۵
۲۰۰۸/۲	۴/۷۲	۴/۲۲	۴/۵۰	۱۱۲/۰۰	۶۸/۲۲	۸۹/۶۶
۲۰۰۸/۳	۴/۸۴	۴/۱۷	۴/۵۲	۱۲۶/۰۵	۶۴/۷۶	۹۱/۹۰

نشان می‌دهد اختلاف اساسی بین این روش‌های پیش‌بینی وجود ندارد و پیش‌بینی‌های خودگردان به‌خوبی پیش‌بینی‌های حاصل از روابط (۷)، (۸) و (۹) هستند. با این مزیت که در روش خودگردان پیش‌بینی‌های فاصله‌ای به‌سادگی قابل محاسبه‌اند. کران‌های بالا و پایین یک پیش‌بینی فاصله‌ای با ضریب اطمینان ۹۵ درصد برای پیش‌بینی گام بعد محاسبه در جدول ۳ آمده‌اند. مقایسه‌ی مقادیر میانگین قدر مطلق خطاهای پیش‌بینی در جدول ۳ نشان می‌دهد با افزایش گام‌ها، همان‌طور که انتظار می‌رفت، از دقت پیش‌بینی‌ها کاسته می‌شود.

در انتها پیش‌بینی قیمت نفت خام سنگین ایران در سه ماه اول سال ۲۰۰۸ را به‌کمک مدل $\mathrm{FAR}(۲,۱)$ برازش‌شده و روش خودگردان محاسبه کرده و آن‌ها را با مقادیر واقعی‌شان که به‌ترتیب ۸۶/۳۶، ۸۸/۵۱ و ۹۶/۶۸ مشاهده شده‌اند مقایسه خواهیم کرد. برای این منظور، $B = ۱۰۰۰$ نمونه‌ی خودگردان در هر مورد شبیه‌سازی شده است. پیش‌بینی‌های خودگردان به‌همراه پیش‌بینی‌های فاصله‌ای ۹۵ درصدی برای لگاریتم قیمت به‌کمک مدل برازش‌شده محاسبه شده و در جدول ۴ آمده‌اند. پادلگاریتم این مقادیر به‌عنوان پیش‌بینی قیمت نفت خام سنگین ایران در سه ماهه‌ی اول سال ۲۰۰۸ محاسبه و در جدول ۴ آمده‌اند. این مقادیر در نمودارهای شکل ۲ نیز رسم شده‌اند. مقایسه‌ی بین پیش‌بینی‌های حاصل یعنی ۸۸/۱۵، ۸۹/۶۶ و ۹۱/۹۰ با مقادیر واقعی‌شان ۸۶/۳۶، ۸۸/۵۱ و ۹۶/۶۸ نشان می‌دهد که بیشینه‌ی قدر مطلق اختلاف برابر ۴/۷۵ دلار مربوط به مارچ ۲۰۰۸ و به‌طور نسبی حدود ۵ درصد مقدار واقعی‌اش است.

به‌کمک نمونه‌های خودگردان، فراوانی نسبی پیش‌بینی‌های خودگردان قیمت نفت خام سنگین ایران برای سه ماه اول سال ۲۰۰۸ محاسبه و هیستوگرام مربوط به آن در شکل ۳ رسم شده است. با استفاده از این توزیع‌های نمونه‌ای می‌توان احتمال این‌که قیمت مورد نظر مقادیر خاصی را اختیار کند براورد کرد. به‌عنوان مثال براورد احتمال این‌که قیمت نفت خام سنگین ایران در فوریه‌ی سال ۲۰۰۸ عددی بین ۹۰ تا ۹۵ دلار باشد با استفاده از توزیع شکل ۳$-(b)$ برابر است با ۰/۱۷۵.

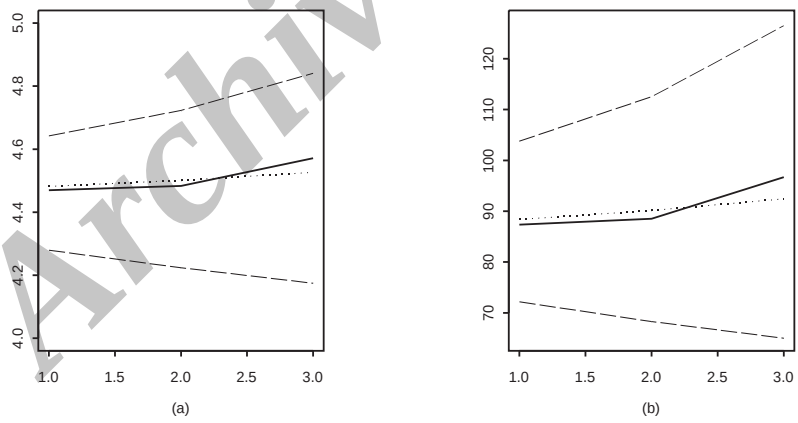
www.SID.ir

۷ نتیجه‌گیری و پیشنهادات

همان‌طور که ملاحظه شد مدل اتورگرسیو ضریب–تابعی با پیش‌فرض‌های کم روی ساختار مدل و انعطاف‌پذیری بالا از جمله مدل‌های کارآمد در تحلیل سری‌های زمانی است. روش رگرسیون خطی موضعی مورد استفاده در برازش این مدل هم از بعد کاربردی و هم از بعد نظری دارای ویژگی‌های قابل توجهی است. روش پیش‌بینی خودگردان معرفی‌شده نیز علاوه بر دقت قابل قبول در پیش‌بینی نقطه‌ای، امکان پیش‌بینی فاصله‌ای و همچنین براورد توزیع پیش‌بین را نیز فراهم می‌کند. لازم به ذکر است با توجه به تأثیر مستقیم برخی عوامل بر قیمت نفت، استفاده از مدل رگرسیون ضریب–تابعی در سری‌های زمانی با در نظر گرفتن متغیرهای کمکی مناسب، می‌تواند نتایج مفیدی در پی داشته باشد.

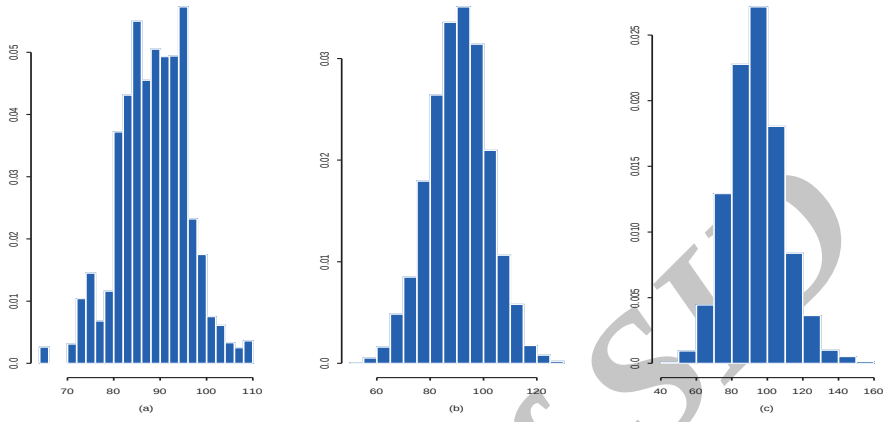
سپاس‌گزاری

نویسندگان مقاله از داوران محترم که نظرها و پیشنهادهای ارزنده‌ی آن‌ها باعث ارایه‌ی بهتر مقاله شده است، کمال تشکر را دارند.



شکل ۲. (a) نمودار پیش‌بینی ۳ مشاهده‌ی آینده‌ی لگاریتم قیمت نفت خام سنگین ایران (خط) به‌همراه پیش‌بینی فاصله‌ای خودگردان ۹۵ درصد آن‌ها (خط‌چین). (b) نمودار پیش‌بینی ۳ مشاهده‌ی آینده‌ی قیمت نفت خام سنگین ایران (خط)، نمودار مقادیر واقعی قیمت نفت خام سنگین ایران در این سه دوره (نقطه‌چین)، به‌همراه پیش‌بینی فاصله‌ای خودگردان ۹۵ درصد آن‌ها (خط‌چین).

Archive of SID
www.SID.ir



شکل ۳. (a) هیستوگرام فراوانی نسبی پیش‌بینی‌های خودگردان برای ژانویه (b)، فوریه (b) و مارچ (c) سال ۲۰۰۸

مرجع‌ها

Bellman, R.E. (1961). *Adaptive Control Processes.* Princeton university Press, Princeton.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics,* **31**, 307-327.

Box, G.E.P. and Jenkins, G.M. (1970). *Time Series Analysis, Forecasting and Control.* Holden-Day, San Francisco.

Cai, Z., Fan, J. and Yao, Q. (2000). Functional-coefficient regression models for nonlinear time series. *Journal of the American Statistical Association,* **95**, 941-956.

Chen, R. and Tsay, R.S. (1993). Functional-coefficient autoregressive models. *Journal of the American Statistical Association,* **88**, 298-308.

Engel, R.F. (1982). Autoregressive conditional heteroscedasticity with estimate of the variance of U. K. inflation. *Econometrica,* **50**, 987-1008.

Granger, C.W.J. and Anderson, A.P. (1978). *An Introduction to Bilinear Models.* Van der-hoeck & Ruprecht, Gottingen.

Haggan, V. and Ozaki, T. (1981). Modelling nonlinear vibrations using an amplitude-dependent autoregressive time series model. *Biometrica,* **68**, 189-196.

www.SID.ir

Hastie, T.J. and Tibshirani, R. (1990). *Generalized Additive Models.* Chapman and Hall, London.

Huang, J.Z. and Shen, H. (2004). Functional coefficient regression models for nonlinear time series: A polynomial spline approach. *Scandinavian Journal of Statistics,* **31**, 515-534.

Ozaki, T. (1982). The statistical analysis of perturbed limit cycle processes using nonlinear time series models. *Journal of Time Series Analysis,* **3**, 29-41.

Tong, H. (1990). *Nonlinear Time Series: A Dynamical Systems Approach*, Oxford University Press, Oxford.

Wong, H., Ip, W.C., Zhang, R.Q. and Xia, J. (2007). Non-parametric time series models for hydrological forecasting. *Journal of Hydrology,* **332**, 337-347.

Yule, G.U. (1927). On a method of investigating periodicities in distorbed series with special reference to Wolfer's sunspot numbers. *Philisophical Transactions of the Royal Society (London),* **A, 226**, 267-298.

پروین جلیلی
بانک مرکزی جمهوری اسلامی ایران،
خیابان فردوسی،
تهران، ایران.
رایانشانی: *P.jalili@cbi.ir*

مجتبی خزایی
گروه آمار، دانشکده‌ی علوم ریاضی،
دانشگاه شهید بهشتی،
تهران، ایران.
رایانشانی: *m_khazaei@sbu.ac.ir*

Archive of SID

www.SID.ir