جمهوری اسلامی ایران
وزارت علوم تحقیقات و فناوری

دانشگاه شیخ بهایی اصفهان

دانشکده تحصیلات تکمیلی

پایان نامه کارشناسی ارشد رشته ریاضی مالی

کاربرد تبدیل موجک هار در پیش‌بینی داده‌های سری‌زمانی مالی براساس مدل سری زمانی ARIMA

پژوهشگر

مهدی ضیایی

استاد راهنما

دکتر علی داوری

استاد مشاور

دکتر ایرج کاظمی

سال تحصیلی ۹۲-۹۱

بسم الله الرحمن الرحيم

دانشگاه شیخ بهایی اصفهان

دانشکده تحصیلات تکمیلی

پایان نامه کارشناسی ارشد رشته ریاضی مالی

کاربرد تبدیل موجک هار در پیش‌بینی داده‌های سری‌زمانی مالی براساس مدل سری زمانی ARIMA

پژوهشگر

مهدی ضیایی

استاد راهنما

دکتر علی داوری

استاد مشاور

دکتر ایرج کاظمی

سال تحصیلی ۹۲-۹۱

تشکر و قدردانی

حمد و سپاس بی‌پایان خداوند متعال را، که توفیق انجام این پژوهش را به من ارزانی داشت.
در اینجا لازم می‌دارم از زحمات فراوانی که پدرو مادرم در تمام طول سال‌های تحصیل برایم متحمل شده‌اند، کمال تشکر و قدردانی را داشته باشم. همین‌طور از همراهی همسرم در انجام این پژوهش، کمال تشکر و سپاس‌گذاری را دارم.

مهدی ضیایی

مهرماه ۹۲

چکیده

این تحقیق قصد دارد تا میزان توانایی تبدیل موجک هار (HAAR) را در پیش‌بینی داده‌های سری‌زمانی مالی مورد ارزیابی قرار دهد. به‌همین منظور داده‌های بورس نزدک را از سایت یاهو فایننس انتخاب کرده و سری‌زمانی بازدهی این داده‌ها را ابتدا توسط مدل GARCH پیش‌بینی، و نتایج را ثبت کرده ایم و سپس همان سری داده‌ها را با استفاده از تبدیل موجک هار (HAAR) تبدیل و با استفاده از مدل ARIMA این داده‌های تبدیل یافته را نیز، پیش‌بینی کرده ایم.

سپس نتایج بدست آمده از این دو روش را با هم مقایسه کرده، و روشی را که نتایج بهتری ارائه می‌دهد، مشخص کرده ایم. به این صورت که برای هر روش، میانگین مربعات خطا (MSE) یا مجذور میانگین مربعات خطا (RMSE) و همچنین میانگین خطای مطلق (MAE) را با هم مقایسه و روش مطلوب‌تر را شناسایی کرده ایم.

کلمات کلیدی:

سری‌زمانی مالی (FINANCIAL TIME SERIES) ، پیش‌بینی (FORECASTING) ، مدل ARIMA ، تبدیل موجک هار (HAAR WAVELET)

« فهرست مطالب »

« فهرست اشکال »

عنوان صفحه

مقدمه

نظر به اینکه نیاز مستمر و روز افزونی به پیش‌بینی سری‌زمانی مالی مانند بورس اوراق بهادار وجود دارد و همچنین نیاز به پیش‌ببینی‌های دقیق‌تر و کاراتر کاملا محسوس است، در نتیجه این تحقیق به دنبال یک راه حل کارا است تا بتواند بازارهای مالی را به شکل دقیق‌تری پیش‌بینی کند. در این تحقیق یک روش جدید برای پیش‌بینی سری‌زمانی مالی بر اساس تبدیلات موجک هار، پیشنهاد می‌شود.

تبدیلات موجک در شاخه‌های مختلف علوم ریاضی مورد استفاده قرار گرفته‌است. اولین کاربرد موجک‌ها در ژئوفیزیک، برای تحلیل داده‌های نقشه‌برداری شده از زلزله بود که در اکتشافات معدن و نفت برای تصویر گرفتن از لایه‌های زیر سطحی صخره‌ها استفاده می‌شد. در سال‌های بعد از آن، استفاده از موجک‌ها به عنوان یک ابزار دقیق‌تر از سری‌فوریه برای پردازش سیگنال استفاده شد و امروزه کاربردهای وسیع‌تری برای موجک در نظر گرفته شده است که از آن جمله می‌توان به کاربرد موجک در پیش‌بینی سری‌زمانی مالی که هدف این تحقیق نیز هست، اشاره کرد.

در سال ۲۰۰۲ محمد و همکاران با استفاده از داده‌های بورس عربستان توضیح دادند که برای داده‌های سری‌زمانی مالی، تبدیلات موجک بهتر از تکنیک‌های رایج، می‌تواند پیش‌بینی کند.

در سال ۲۰۰۸ آگاروال کارآیی پیش‌بینی داده‌ها براساس تبدیلات موجک و مدل‌های رایج دیگر را مقایسه کرد، نتیجه این بود که مدل پیشنهادی بر اساس تبدیل موجک، نتایج بهتری داشت.

در سال ۲۰۱۱ اسماعیل و همکاران برای داده‌های بورس عمان نشان دادند که استفاده از دو موجک هار و دوبشی، نتایج بهتری را برای پیش‌بینی سری‌زمانی مالی ارائه می‌دهد.

این تحقیق نیز نشان می‌دهد که برای داده‌های بورس نزدک استفاده از تبدسل موجک هار، نتایج مطلوب‌تری را برای پیش‌بینی سری‌زمانی مالی ارائه می‌دهد.

فصل اول

سری زمانی

۱-۱ آشنایی با مفاهیم بنیادی

فرض کنید (Ω, F, P) یک فضای احتمالی و T یک مجموعه اندیس گذار باشد. سری زمانی حقیقی تابعی است مانند $X(t,\omega)$

$$X(t,\omega): T \times \Omega \to \mathbb{R}$$

به قسمی که به ازای هر مقدار ثابت t، $X(t)$ یک متغیر تصادفی باشد. معمولاً این تابع را با X_t یا $X_t(\omega)$ و سری زمانی را به صورت $\{X_t : t \in T\}$ نشان می‌دهند.

به ازای هر مقدار ثابت ω تابع $X(t,\omega)$، یک تابع حقیقی از t است، که آن را یک مصداق یا یک تابع نمونه ای گویند. در واقع هر نمودار یک سری زمانی تصویر $X(t,\omega)$ به ازای یک مقدار ثابت ω است.

اگر مجموعه اندیس گذار فقط شامل یک عضو باشد آنگاه فرآیند تصادفی تبدیل به یک متغیر تصادفی معمولی می‌شود.

تابع توزیع توأم $\{X_t : t \in T\}$ را به صورت زیر تعریف می‌کنند.

$$F_{X_{t_1} \dots X_{t_n}}(x_{t_1}, \dots, x_{t_n}) = P(X(t_1,\omega) \le x_{t_1}, \dots, X(t_n,\omega) \le x_{t_n})$$

که $x_{t_1}, \dots, x_{t_n}$ اعداد حقیقی دلخواه اند.

۱-۱-۱ تعریف

سری زمانی $\{X_t, t \in T\}$ را مانای اکید گویند اگر

$$F_{X_{t_1} \dots X_{t_n}}(x_{t_1}, \dots, x_{t_n}) = F_{X_{t_1+k} \dots X_{t_n+k}}(x_{t_1}, \dots, x_{t_n})$$

به ازای هر مجموعه از اندیس های $(t_1, \dots, t_n)$ و $(t_1 + k, \dots, t_n + k)$ متعلق به T و تمام $(x_{t_1}, \dots, x_{t_n})$ متعلق به حوزه مقادیر متغیر تصادفی X_t برقرار باشد.

بخصوص به ازای $n = 1$ و هر k داریم.

$$F_{X_{t_1}}(x_{t_1}) = F_{X_{t_1+k}}(x_{t_1})$$

یعنی تابع توزیع در هر نقطه مجموعه اندیس گذار یکسان است. در نتیجه تمام X_t ها هم توزیع اند.

اگر سری $\{X_t, t \in T\}$ اکیدا مانا باشد و $E(X_t) < \infty$، آنگاه میانگین X_t به ازای هر مقدار t ثابت است زیرا تابع توزیع آن به ازای تمام t ها یکسان است.

$$E(X_t) = \mu, \qquad \forall t \in T$$

همین طور اگر $E(X_t^2) < \infty$ واریانس X_t ها به ازای هر t ثابت خواهد بود.

$$V(X_t)=\sigma^2, \quad \forall t \in T$$

اگر به ازای هر مجموعه متناهی از متغیرهای ($X_{t_1},...,X_{t_n}$) تابع توزیع توأم آنها معلوم باشد ، سری زمـانی از نظر احتمال کاملاً معین است. ولی در بیشتر کاربردها تابع توزیع مجهول است.

در این صورت مشخصات لازم سری زمانی، به کمک دو گشتاور اول و دوم به دست می‌آید.

۱-۱-۲ تعریف

سری زمانی $\{ X_t, t \in T \}$ را مانای ضعیف گویند اگر

۱- میانگین X_t به t بستگی نداشته باشد. $E(X_t)=\mu$.

۲- ماتریس کوواریانس $X_{t_1},...,X_{t_n}$ با ماتریس کوواریانس ($X_{t_1+k},...,X_{t_n+k}$) به ازای هـر مجموعـه ($t_1,...,t_n$) از مجموعه اندیس گذار و هر n برابر باشد.

چون میانگین X_t ثابت است، می توان آن را برابر صفر فـرض کـرد. همچنـین مـاتریس کوواریـانس تـابعی از فاصله بین مشاهدات است. یعنی کوواریانس X_t و X_{t+k} فقط به k بستگی دارد و می‌توان نوشت

$$\text{cov}(X_t, X_{t+k}) = E(X_t.X_{t+k}) = \gamma(k)$$

۱-۱-۳ تعریف

تابع $\gamma(k)$ را تابع اتوکوواریانس سری $\{ X_t \}$ می‌نامند و آن را با نماد اختصاری $acv.f$ نشان می‌دهند.

معمولاً برای مقایسه خواص اساسی سـری‌هـای زمـانی اغلـب مفیـد اسـت تـابعی داشـته باشـیم کـه از واحـد اندازه‌گیری مستقل باشد. برای این منظور تابع خود همبستگی را به صورت زیر تعریف می‌کنند.

$$\rho(k)=\frac{\gamma(k)}{\gamma(0)}$$

و آن را با نماد اختصاری $ac.f$ نشان می‌دهند. همان‌طور که ملاحظه مـی‌شـود مقـدار $\rho(k)$ در نقطـه $k=0$ برابر ۱ است.

۱-۱-۴ تابع خود همبستگی جزئی

وسیله مهم دیگری که برای مطالعه سری‌های زمانی استفاده می‌شود، تابع خود همبسـتگی جزئـی اسـت. و آن‌را با نماد اختصاری ($pac.f$) نشان می‌دهند.

۱-۱-۵ ماتریس خود همبستگی

ماتریس خود همبستگی برای یک سری مانا به طول N به صورت زیر تعریف می‌شود.

$$P_N=\begin{bmatrix}1 & \rho_1 & \rho_2 & \cdots & \rho_{N-1}\\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{N-2}\\ . & . & . & \cdots & .\\ . & . & . & \cdots & .\\ \rho_{N-1} & . & . & \cdots & 1\end{bmatrix}$$

که در آن سطر اول به صورت زیر نوشته می‌شود.

$$\rho(1,k), k=1,\ldots,N,\qquad \rho(11)=\rho_{11}=\frac{\gamma(0)}{\gamma(0)}=1,\qquad \rho(1,2)=\rho_{12}=\frac{\gamma(1)}{\gamma(0)}=\rho_1\quad,\ldots$$

سطر دوم با محاسبه $\rho(2,k)$ ، .... ، سطر آخر با محاسبه $\rho(n,k)$ به دست می‌آید.

۱-۱-۶ تعریف

تابع همبستگی جزئی را به صورت زیر تعریف کنیم.

$$\{\varphi_{kk}, k=1,2,\ldots\}$$

که آن را مجموعه همبستگی جزئی با تأخیر k نامیم. هر یک از این مقادیر به شکل زیر تعریف می‌شود.

$$\varphi_{kk}=\frac{|P_k^*|}{|P_k|}$$

که در آن P_k ماتریس خود همبستگی $k\times k$ و P_k^* همان ماتریس قبلی است که ستون آخر آن با بردار $(\rho_1,\rho_2,\ldots,\rho_k)'$ عوض می‌شود . بنابراین:

$$\phi_{11}=\rho_1\quad,\quad \phi_{22}=\frac{\begin{vmatrix}1 & \rho_1\\ \rho_1 & \rho_2\end{vmatrix}}{\begin{vmatrix}1 & \rho_1\\ \rho_1 & 1\end{vmatrix}}=\frac{\rho_2-\rho_1^2}{1-\rho_1^2}\quad,\quad\ldots$$

۱-۱-۷ برآورد توابع خود همبستگی

فرض کنید سری زمانی $X_{t_1},\ldots,X_{t_N}$ داده شده است. ابتدا مقادیر زیر را به کمک این نمونه محاسبه می‌کنیم

$$\hat{\mu}=\overline{X_t}=\frac{1}{N}\sum_{v=1}^{N}X_{t_v}$$

$$\hat{\gamma}(k)=C_k=\frac{1}{N}\sum_{t=k+1}^{N}(X_t-\overline{X_t})(X_{t-k}-\overline{X_t})$$

$$=\frac{1}{N}\sum_{t=1}^{N-k}(X_t-\overline{X_t})(X_{t-k}-\overline{X_t})$$

بدیهی است که برای دقت بیشتر در برآورد، باید N به اندازهٔ کافی بزرگ باشد (در عمل $N > ۵۰$). ضمناً مقادیر C_k را به ازای $k > N-۱$ نمی‌توان محاسبه کرد و ضمناً برای $k=۰$ داریم:

$$\hat{\gamma}(۰) = C_۰ = \frac{۱}{N}\sum_{t=۱}^{N}(X_t - \overline{X_t})^۲$$

در نتیجه

$$\hat{\rho}(k) = r_k = \frac{C_k}{C_۰} = \frac{\hat{\gamma}(k)}{\hat{\gamma}(۰)}$$

۲-۱ فرآیندهای تصادفی و الگوهای سری زمانی مانا

۱-۲-۱ تعریف

هر فرآیند گسسته $\{ Z_t, t \in T \}$ را که به صورت دنباله ای از متغیرهای تصادفی نا همبسته از توزیعی با میانگین $E(Z_t) = \mu_Z$ (که معمولاً صفر فرض می شود) ، با واریانس ثابت $\text{var}(Z_t) = \sigma_z^۲$ و تابع اتوکوواریانس $\gamma_k = \text{cov}(Z_t, Z_{t-k}) = ۰$ (به ازای $k \neq ۰$) است یک فرآیند تصادفی محض می‌نامند. از این تعریف نتیجه می‌گیریم که فرآیند تصادفی محض $\{ Z_t \}$ ماناست با تابع اتوکواریانس:

$$\gamma_k = \text{cov}(Z_t, Z_{t-k}) = \begin{cases} \sigma_z^۲ & k = ۰ \\ ۰ & k \neq ۰ \end{cases}$$

و تابع خود همبستگی:

$$\rho_k = \frac{\gamma_k}{\gamma_۰} = \begin{cases} ۱ & k = ۰ \\ ۰ & k \neq ۰ \end{cases}$$

و تابع خود همبستگی جزئی:

$$\phi_{kk} = \frac{|P_k^*|}{|P_k|} = \begin{cases} ۱ & k = ۰ \\ ۰ & k \neq ۰ \end{cases}$$

۲-۲-۱ فرآیند اتورگرسیو

فرض کنیم (Z_t) یک فرآیند تصادفی محض با میانگین صفر و واریانس $\sigma_z^۲$ باشد ، فرآیند $\{X_t\}$ را فرآیند اتورگرسیو مرتبه p گویند هرگاه

$$X_t = \alpha_۱ X_{t-۱} + \alpha_۲ X_{t-۲} + \ldots + \alpha_p X_{t-p} + Z_t$$

اگر به مسئله رگرسیون فکر کنیم می‌بینیم که الگوی بالا در واقع یک الگوی رگرسیون چندگانه است با این تفاوت که در اینجا روی متغیرهای مستقل رگرسیون نشده بلکه روی مقادیر گذشته X_t رگرسیون شده است و به

این دلیل است که فرآیند $\{X_t\}$ را اتورگرسیو نامیده‌اند. یک فرآیند اتورگرسیو مرتبه p را با نماد اختصاری $AR(p)$ نمایش می‌دهند. این قبیل فرآیند ها را یول در سال ۱۹۲۰ معرفی کرده‌است. این فرآیند همواره وارون‌پذیر است ، یعنی همواره بدون گذاشتن شرط یا شرایطی روی α_i ها می‌توان Z_t ها را بصورت یک ترکیب خطی موزون از مشاهدات حال و گذشته فرآیند نوشت. این فرآیند را با توجه به عملگر انتقال پسرو می‌توان بصورت زیر نوشت.

$$\phi(B)X_t = Z_t$$

که در آن

$$\phi(B) = ۱ - \alpha_۱ B - \alpha_۲ B^۲ - \ldots - \alpha_p B^p$$

برای مانایی فرآیند باید ریشه های $\phi(B) = ۰$ از لحاظ قدر مطلق بزرگتر از واحد باشند. (خارج دایره واحد باشند). از این فرآیند در مواقعی استفاده می‌شود که مقدار حال سری زمانی به مقادیر بلافاصله قبل از آن به علاوه یک خطای تصادفی بستگی دارد.

به‌دلیل کاربرد فراوان فرآیندهای اتوگرسیو مرتبه اول و دوم ابتدا این الگوهای ساده را بررسی می‌کنیم.

۱-۲-۳ فرآیند اتورگرسیو مرتبۀ اول (۱) AR

اگر در فرآیند $AR(p)$ قرار دهیم $p = ۱$ ، فرآیند $AR(۱)$ حاصل می‌شود که بصورت زیر است

$$X_t = \alpha_۱ X_{t-۱} + Z_t$$

با استفاده از عملگر پسرو می‌توان نوشت

$$(۱ - \alpha_۱ B)X_t = Z_t$$

به طوری که در بالا متذکر شدیم این فرآیند همواره وارون‌پذیر است، برای این که فرآیند مانا باشد باید ریشه‌های $۱ - \alpha_۱ B = ۰$ از لحاظ قدر مطلق بزرگتر از واحد باشند که از آن $|\alpha_۱| < ۱$ نتیجه می‌شود. به‌دلیل این که مقدار X_t بطور کامل به وسیلۀ $X_{t-۱}$ تعیین می‌شود، گاهی اوقات فرآیند $AR(۱)$ را فرآیند مارکوف نیز می‌نامند.

۱-۲-۴ تابع خود همبستگی فرآیند $AR(۱)$

ساده‌ترین راه برای تعیین تابع خود همبستگی فرآیندهای اتورگرسیو این است که: فرآیند را مانا فرض کرده و سپس طرفین معادلۀ

$$X_t = \alpha_۱ X_{t-۱} + \alpha_۲ X_{t-۲} + \ldots + \alpha_p X_{t-p} + Z_t$$

را در X_{t-k} ضرب کنیم و امیدهای ریاضی جملات طرفین تساوی را بدست آوریم و از این واقعیت استفاده کنیم که به ازای هر k، $\rho_k = \rho_{-k}$ (تابع خود همبستگی تابعی زوج است). پس در مورد فرآیند $AR(۱)$ داریم:

$$E(X_{t-k}X_t) = E(\alpha_۱ X_{t-۱}X_{t-k}) + E(Z_t X_{t-k})$$

برای $k \geq ۱$ چون خطای زمان حال مستقل از گذشتهٔ فرآیند است $(E(Z_t X_{t-k}) = ۰ \quad , \quad k > ۰)$، به صورت

$$\gamma_k = \alpha_۱ \gamma_{k-۱} \quad , \quad k \geq ۱$$

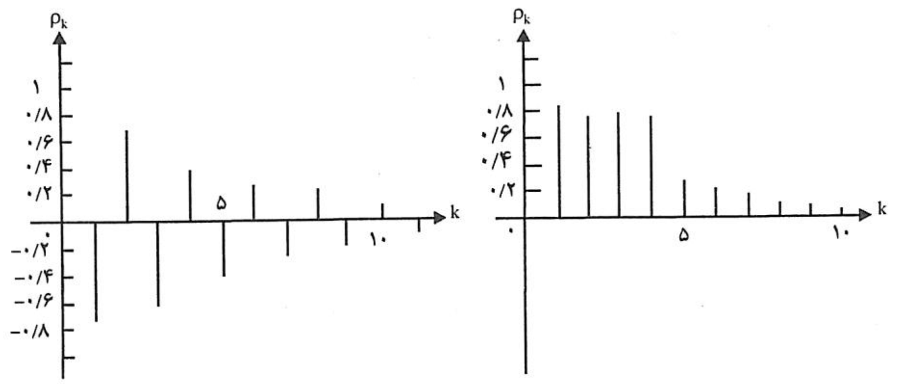
در می‌آید. اگر طرفین این تساوی را بر واریانس فرآیند (یعنی $\sigma_x^۲ = \gamma_.$) تقسیم کنیم، تابع خود همبستگی فرآیند $AR(۱)$ به دست می‌آید.

$$\frac{\gamma_k}{\gamma_.} = \alpha_۱ \frac{\gamma_{k-۱}}{\gamma_.} \quad , \quad k \geq ۱$$

یا

$$\rho_k = \alpha_۱ \rho_{k-۱} \quad , \quad k \geq ۱$$

که از آن $\rho_k = \alpha_۱^k$ و $k = ۰,۱,۲,...$ نتیجه می‌شود. بنابراین وقتی فرآیند ماناست $(|\alpha_۱| < ۱)$ تابع خود همبستگی فرآیند به طور نمایی تنزل می‌کند. اگر $۰ < \alpha_۱ < ۱$، تمام خود همبستگی ها مثبت‌اند و اگر $-۱ < \alpha_۱ < ۰$. علامت خود همبستگی یک طرح تناوبی را نشان می‌دهد که با یک مقدار منفی شروع می‌شود. اندازهٔ این خود همبستگی ها که در هر دو حالت به طور نمایی تنزل می‌کنند در شکل‌های زیر نشان داده شده است.



شکل ۱-۱ : تابع خود همبستگی فرآیند اتورگرسیو

۱-۲-۵ میانگین فرآیند اتورگرسیو مرتبهٔ اول

اگر از طرفین تساوی $X_t = \alpha_۱ X_{t-۱} + Z_t$ امید ریاضی بگیریم داریم

$$E(X_t) = \alpha_۱ E(X_{t-۱}) + ۰$$

یا

$$\mu_x - \alpha_۱ \mu_x = ۰ \rightarrow (۱ - \alpha_۱)\mu_x = ۰$$

ولی چون $\phi(1)=1-\alpha_1\neq 0$، باید داشته باشیم $E(X_t)=\mu_x=0$

۱-۲-۶ واریانس فرآیند اتورگرسیو مرتبهٔ اول

برای به دست آوردن واریانس فرآیند $AR(1)$ در عبارت

$$E(X_tX_{t-k})=E(\alpha_1X_{t-1}X_{t-k})+E(X_{t-k}Z_t)$$

مقدار $k=0$ را قرار داده و با توجه به $E(X_t^2)=\sigma_x^2$ (چون میانگین صفر است) و $E(Z_tX_t)=\sigma_z^2$ داریم:

$$\sigma_x^2=\alpha_1\gamma_1+\sigma_z^2$$

از طرفی به ازاء $k=1$ عبارت بالا نتیجه می‌دهد $\gamma_1=\alpha_1\gamma_0=\alpha_1\sigma_x^2$، پس

$$\sigma_x^2=\alpha_1(\alpha_1\sigma_x^2)+\sigma_z^2$$

و با $\alpha_1\neq 1$ داریم: $\sigma_x^2=\gamma_0=\dfrac{\sigma_z^2}{1-\alpha^2}$

به طور خلاصه میانگین، واریانس، تابع اتوکواریانس و تابع خود همبستگی فرآیند $AR(1)$ بصورت زیرند:

$$\mu_x=EX_t=0$$

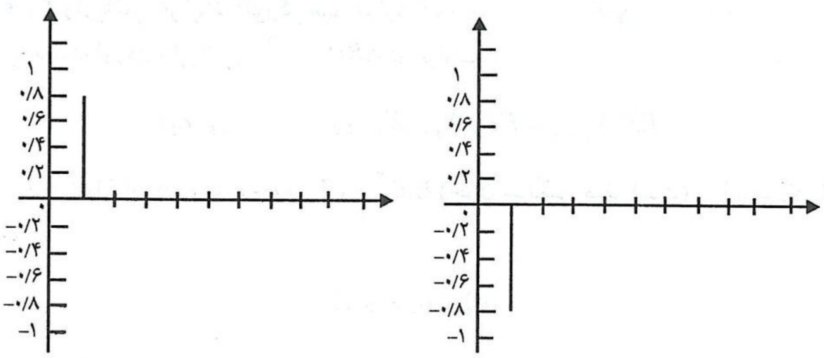
$$\begin{aligned}
\sigma_x^2=\gamma_0=\frac{\sigma_z^2}{1-\alpha_1^2} \quad &, \quad \alpha_1\neq 1\\
\gamma_k=\alpha_1\gamma_{k-1} \quad &, \quad k\geq 1\\
\rho_k=\alpha_1^k \quad &, \quad k=0,1,2,\ldots
\end{aligned}$$

۱-۲-۷ تابع خود همبستگی جزئی فرآیند اتورگرسیو مرتبهٔ اول

با توجه به تعریف تابع خود همبستگی جزئی($\phi_{kk}=|P_k^*|/|P_k|$) می‌توان نوشت:

$$\phi_{kk}=\begin{cases}\rho_1=\alpha_1 & , \quad k=1\\ 0 & , \quad k\geq 2\end{cases}$$

ملاحظه می‌کنیم که وقتی تأخیر بین مشاهدات سری زمانی مساوی ۲ یا بیش از ۲ است تابع خود همبستگی جزئی برابر صفر است (قطع می‌شود) ولی تابع خود همبستگی جزئی وقتی $k=1$ است برابر ضریب همبستگی بین مشاهدات متوالی سری زمانی است که با توجه به علامت α_1 مقدار آن مثبت یا منفی است. تابع خود همبستگی جزئی فرآیند $AR(1)$ در شکل زیر نشان داده شده است.



شکل ۱-۲ : تابع خودهمبستگی جزئی فرآیند اتورگرسیو مرتبهٔ اول

اگر در فرآیند $AR(1)$ فرض کنیم $\alpha = 1$، فرآیندی بدست می‌آید که قدم زدن تصادفی نام دارد و ناماناست. به ازاء مقادیر $\alpha \geq 2$ مثلاً $\alpha = 2$ سابقهٔ X_t نه فقط نامانایی فرآیند را نشان می‌دهد بلکه یک رشد انفجاری را نیز نشان می‌دهد. در حقیقت می‌توان گفت بعد از مدت زمانی، سری در واقع تصادفی نیست بدین معنی که مشاهدات به اندازه‌ای بزرگ‌اند که از جملهٔ تصادفی محض Z_t می‌توان صرف نظر کرد.

۱-۲-۸ تابع خود همبستگی فرآیند *AR(p)*

برای به دست آوردن تابع اتوکوواریانس یک فرآیند اتورگرسیو ساده‌ترین راه این است که دو طرف تساوی

$$X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \ldots + \alpha_p X_{t-p} + Z_t$$

را در X_{t-k} ضرب کرده و امید ریاضی جملات طرفین تساوی را به دست آوریم ، که نتیجه می‌شود.

$$E(X_t X_{t-k}) = E(\alpha_1 X_{t-1} X_{t-k}) + E(\alpha_2 X_{t-2} X_{t-k}) + \ldots + E(\alpha_p X_{t-p} X_{t-k}) + E(Z_t X_{t-k})$$

به ویژه برای $k > 0$ داریم ،

$$\rho_k = \alpha_1 \rho_{k-1} + \alpha_2 \rho_{k-2} + \ldots + \alpha_p \gamma_{k-p} \quad , \quad k > 0$$

از این معادله ملاحظه می کنیم که تابع خود همبستگی ρ_k با توجه به معادله تفاضلی زیر تعیین می‌شود.

$$\phi_p(B)\rho_k = (1 - a_1 B - a_2 B^2 - \ldots - a_p B^p)\rho_k = 0 \quad , \quad k > 0$$

۱-۲-۹ تابع خود همبستگی جزئی فرآیند کلی *AR(p)*

با توجه به معادله $\rho_k = \alpha_1 \rho_{k-1} + \alpha_2 \rho_{k-2} + \ldots + \alpha_p \rho_{k-p} \;,\; k > 0$ به سهولت می‌بینیم که برای $k > p$ ستون آخر ماتریس P_k^* در ϕ_{kk} را می توان به صورت یک ترکیب خطی از ستون‌های قبلی آن نوشت. بنابراین تابع خود همبستگی جزئی بعد از تأخیر p صفر می‌شود و این یک خاصیت مهم برای شناسایی یک الگوی AR در ساختن الگوی سری زمانی است.

۱-۲-۱۰ معادلات یول – والکر

اگر در معادلهٔ $\rho_k = \alpha_1 \rho_{k-1} + \alpha_2 \rho_{k-2} + ... + \alpha_p \rho_{k-p}, k \geq 1$ قرار دهیم، $k = 1, 2, ..., p$ و توجه کنیم کـه $\rho_0 = 1$ و $\rho_{-k} = \rho_k$، خواهیم داشت:

$$
\begin{aligned}
&\rho_1 = \alpha_1 + \alpha_2 \rho_1 + ... + \alpha_p \rho_{p-1} \\
&\rho_2 = \alpha_1 \rho_1 + \alpha_2 + ... + \alpha_p \rho_{p-2} \\
&... \\
&\rho_p = \alpha_1 \rho_{p-1} + \alpha_2 \rho_{p-2} + ... + \alpha_p
\end{aligned}
$$

این دستگاه معادلات را معادلات یول – والکر می‌نامند. با داشتن مقادیر $\alpha_1, \alpha_2, ..., \alpha_p$ از حل ایـن معـادلات خطی می‌توانیم $\rho_1, \rho_2, ..., \rho_p$ را به دست آوریم، با استفاده از معادلهٔ

$$\rho_k = \alpha_1 \rho_{k-1} + \alpha_2 \rho_{k-2} + ... + \alpha_p \rho_{k-p} \quad , \quad k \geq 1$$

می‌توانیم ρ_k را برای تأخیرهای بالاتر به دست آوریم.

۱-۲-۱۱ واریانس فرآیند ***AR(p)***

با توجه به رابطه

$$
\begin{aligned}
&E(X_t Z_t) = E\left[Z_t(\alpha_1 X_{t-1} + \alpha_2 X_{t-2} + ... + \alpha_p X_{t-p} + Z_t)\right] \\
&= E(Z_t^2) \\
&= \sigma_z^2
\end{aligned}
$$

اگر طـرفین معادلـه $X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + ... + \alpha_p X_{t-p} + Z_t$ را در X_t ضـرب کـرده و امیـد ریاضـی بگیریم خواهیم داشت ،

$$E(X_t^2) = E(\alpha_1 X_t X_{t-1}) + E(\alpha_2 X_t X_{t-2}) + ... + E(\alpha_p X_t X_{t-p}) + E(Z_t X_t)$$

و از آن نتیجه می‌شود

$$\gamma_0 = \alpha_1 \gamma_1 + \alpha_2 \gamma_2 + ... + \alpha_p \gamma_p + \sigma_z^2$$

که با استفاده از $\rho_k = \frac{\gamma_k}{\gamma_0}$ به صورت زیر نوشته می‌شود ،

$$\gamma_0 = (1 - \alpha_1 \rho_1 - \alpha_2 \rho_2 - ... - \alpha_p \rho_p) = \sigma_z^2$$

و یا

$$\gamma_0 = \frac{\sigma_z^2}{1 - \alpha_1 \rho_1 - \alpha_2 \rho_2 - ... - \alpha_p \rho_p}$$