دانشگاه پیام نور

دانشکده علوم پایه

گروه آمار

برآورد های استوار برای مدلهای خود برگشت در قلمرو های زمان و فرکانس

پایان نامه

برای دریافت درجه کارشناسی ارشد

در رشته آمار ریاضی

نگارش:

مجتبی پاکدل توچایی

استاد راهنما:

دکتر مسعود یارمحمدی

دی 1389

چکیده

تحلیل طیفی یک روش توسعه یافته برای تحلیل سری‌های زمانی در قلمرو فرکانس است. این روش از متداولترین شیوه‌های مورد استفاده برای کاهش مناسب داده‌ها و متعاقباً مقایسه این گونه از ثبت داده‌هاست. اشتباهات ذاتی یا سهوی در داده‌های یک سری زمانی تحت بررسی می‌تواند نتایج گمراه کننده و غیر سودمندی در نتایج مربوط به طیف آن داده‌ها برجای گذارد. به این منظور ضمن بیان انواع داده‌های دورافتاده در سری‌های زمانی و تبیین نقش حساس آن‌ها در نتایج، به ارائه برآورد استوار تابع چگالی طیفی می‌پردازیم. یکی از این روش‌ها یافتن تابع چگالی طیفی استوار با استفاده از تبدیل فوریه تابع اتوکوواریانس استوار شده می باشد. روش دیگر پیش‌وزن‌دارکردن برای‌کاهش ریسک داده‌های دورافتاده در تابع چگالی طیفی می‌باشد. سپس با انجام روش‌های شبیه‌سازی کارایی این روش‌ها در رابطه با وجود و عدم وجود نقاط دورافتاده از نوع جمع‌پذیر مورد مقایسه قرار می گیرد.

واژگان کلیدی: نقاط دورافتاده، برآوردگر استوار، تابع چگالی طیفی، طیف خودبرگشت، پیش‌وزن‌دارکردن.

پیشگفتار

تحلیل سری های زمانی به طور نظری و عملی از سالهای 1970 به بعد برای پیش بینی و کنترل به سرعت توسعه پیدا کرده است. این تحلیل معمولا به داده هایی مربوط می شود که مستقل نبوده و بطور متوالی به هم وابسته اند معمولاً وابستگی بین مشاهدات متوالی مورد توجه قرار گرفته و کاربرد اصلی آن مدل بندی داده و در نهایت بیش بینی آینده سری می باشد . در مدل بندی سری های زمانی برای سادگی عموما فرض می شود که خطای مدل دارای توزیع نرمال یا مجانبا نرمال است. اما به دلیل آنکه داده های جمع آوری شده در برخی موارد، ممکن است توسط نوعی از داده‌ها بنام "داده های دورافتاده [1]" که به دلایل متعدد روی می دهند و با داده های اصلی سنخیتی ندارند، آلوده شود. این موضوع باعث می شود که فرض نرمال بودن خطاها برقرار نشده و در نتیجه هرگونه استنباط آماری و تحلیل سری های زمانی را غیر قابل اعتماد و پیش بینی های به دست آمده را متاثر می سازد. لذا برای یافتن برآورد نسبتا درستی از چنین داده هایی عموما به نمونه‌هایی بزرگ نیازمندیم. از اینرو قادر نخواهیم بود که بطور کامل به برآوردگرهای کارا دست یابیم، چرا که ساختار دقیق یک چگالی دم کلفت [2] تقریباً نرمال، کاملاً شناخته شده نیست. از این رو، محققان هنگام تجزیه و تحلیل داده‌های جمع آوری‌شده، هنگامی که به نقاط دورافتاده برمی‌خورند، بایستی روش‌های مناسبی همچون حذف دورافتاده ها(پیراستن) و یا کم وزن کردن اثر این مشاهدات را به کار برند. تلاش‌های مفید و اثربخشی در نیل به این اهداف صورت پذیرفته‌است که می توان از آنها به دانیل و وود[3][1971] اشاره کرد.

---

[1] Outlier Data

[2] Heavy Tailed

[3] Daniel and Wood

البته در این خصوص ضعف‌هایی هم دامن‌گیر است: نخست اینکه ذات روش‌های مقابله با داده‌های دورافتاده، امکان تهیه دقیق مشخصات مربوط به توزیع برآوردگرهای پارامترها را سخت می‌کند. دوم آنکه عدم قابلیت درک روشن از نقاط دورافتاده و پردازش داده ها، بدون در نظر گرفتن و بررسی نقاط دورافتاده، ذاتاً محقق را از انجام یک سری از رهیافت‌های موثر دور نگه می‌دارد.

موضوع پایان نامه حاضر، برآوردهای استوار برای مدل های خودبرگشت در قلمروهای زمان و فرکانس است. از این رو، در فصل اول پس از ارائه مقدمه ای از روشهای تحلیل سری های زمانی در قلمرو زمان و معرفی مدل های معروف باکس و جنکینز که به نام مدل های خودبرگشت و میانگین متحرک مشهور است، انواع نقاط دورافتاده مهم در سری‌های زمانی و روشهای تشخیص این نقاط را مورد بحث و بررسی قرار می دهیم.

در فصل دوم، که به برآوردهای استوار سری های زمانی در قلمرو زمان اختصاص دارد، تاریخچه و مفهوم استواری و تعاریف مربوط به آن مورد بحث قرار می گیرد. در ادامه، برآوردگرهای استوار در مدل های سری های زمانی به طور کامل بررسی خواهند شد و در پایان این فصل، استوار سازی مدل های خودبرگشت بر اساس ایده مدل رگرسیون خطی، بحث و الگوریتم جدیدی برای این منظور، معرفی می شود.

در فصل سوم، به برآوردهای استوار سری های زمانی در قلمرو فرکانس پرداخته و مفاهیمی چون تحلیل دوره‌نگار، طیف نمونه و طیف همواره شده بحث خواهند شد و برآورد ناپارامتری تابع چگالی طیف معرفی خواهد شد. در فصل چهارم با استفاده از روش های شبیه سازی به مقایسه روش های استوار می پردازیم.

فصل اول

فرآیند های خودبرگشت و انواع نقاط دور افتاده و تاثیر آنها بر سری های زمانی

1-1: مقدمه

تحلیل سری های زمانی به طور نظری و عملی از سالهای 1970 به بعد برای پیش بینی و کنترل به سرعت توسعه پیدا کرده است. این تحلیل معمولا به داده هایی مربوط می شود که مستقل نبوده و بطور متوالی به هم وابسته اند. همین وابستگی بین مشاهدات متوالی است که مورد توجه قرار می گیرد و بیشتر کاربرد آن در پیش بینی خواهد بود. این پیش بینی در مورد حوادثی است که در آینده احتمال وقوع دارند و معمولا براساس رویدادهایی صورت می گیرند که در گذشته اتفاق افتاده اند. یکی از مهمترین روشهای استنباط نتایج برای آینده براساس آنچه در گذشته اتفاق افتاده است، تحلیل سری های زمانی در پیش بینی است. در این بخش الگوریتم معروف باکس – جنکینز را برای مدلهای مخلوط خودبرگشت و میانگین متحرک مورد بحث و بررسی قرارداده ایم که شامل شناسایی مدل، برآورد پارامترها، بازرسی تشخیصی و به دنبال آن پیش بینی است. در مرحله بازرسی تشخیصی که توسط معیارهای مناسبی صورت می گیرد در صورت عدم پذیرش مدل مورد نظر، مراحل شناسایی و برآورد پارامترها مجددا تکرار می گردند تا مدل مناسب به‌دست آید.

2-1: اهداف تجزیه و تحلیل سری های زمانی

تجزیه وتحلیل سری های زمانی معمولا در دو قلمرو مطالعه می گردند. تحلیل در قلمروزمان [1] و تحلیل در قلمرو فرکانس [2]. تحلیل در قلمرو زمان در بین آماردانان به لحاظ استفاده از توابع شناخته شده میانگین، واریانس، اتوکوواریانس و خود همبستگی از محبوبیت ویژه ای برخوردار است. در حالی که تحلیل در قلمرو فرکانس در بین مهندسین و فیزیکدانان و کارشناسان علوم زمین به خاطر کاربرد وسیع و گسترده تابع چگالی طیفی در مسائل کاربردی و عملی از طرفداران بیشتری برخوردار است. این تابع را در مهندسی بیشتر به عنوان تابع چگالی توانی می شناسند. به عنوان مثال در مخابرات تجزیه وتحلیل سیگنالها با استفاده از تحلیل فرکانسها امکان پذیر است و این کار با استفاده از تابع چگالی طیفی انجام می پذیرد. (باکس-جنکینز [3]، 1976)

در تحقیقاتی که از تحلیل سری زمانی استفاده می شود چندین هدف ممکن وجود دارد. این اهداف را می توانیم به صورت توصیف، تشریح، پیش بینی و کنترل رده بندی کنیم وقتی یک سری زمانی داده می شود، معمولا اولین مرحله در تحلیل، آن است که اندازه های توصیفی ساده را به دست آوریم. با رسم نمودار داده ها نیز به وجود روند و تغییرات فصلی و همچنین اثرات بیرونی می توان پی برد. جنبه دیگر اینکه در یک سری زمانی ارائه شده می خواهیم مقادیر آینده سری را پیش بینی کنیم. این کار در پیش بینی فروش و در تجزیه وتحلیل سری های زمانی اقتصادی و صنعتی از اهمیت زیادی برخوردار است.

---

[1]- Time Domain

[2]-Frequency Domain

[3]-Box-Jenkins

3-1: روشهای تحلیل سری های زمانی

یک روش معقول برای تحلیل سری ها، تجزیه آن به روند، تغییرات فصلی و سایر نوسانات نامنظم است. مشخصه روند در سری زمانی را می توان به صورت غیر دقیق به عنوان تغییر دراز مدت در میانگین، تعریف کنیم. اشکال این تعریف در معنی دراز مدت است. برای مثال بعضی اوقات متغیرهای مربوط به آب و هوا تغییرات دوره ای را در زمانی بسیار طولانی مانند 5 سال نشان می دهند. با وجود این در نظر گرفتن این قبیل نوسانات در کوتاه مدت هنوز دارای معنی است. وقتی از یک روند صحبت می کنیم باید تعداد مشاهدات موجود را به حساب آورد و از دراز مدت ارزیابی ذهنی به عمل آوریم.

برخی سری های زمانی، تغییراتی در یک دوره ثابت را نشان می دهند. تغییر درجه حرارت روزانه مثالی در این مورد است. به علاوه برخی سری های زمانی نوساناتی را نشان می دهند که دوره ثابتی ندارند، لیکن تا حدودی قابل پیشگویی هستند.

بعد از آنکه روند و تغییرات دوره ای از مجموعه داده ها حذف شوند، سری باقیمانده حاصل می شود که ممکن است تصادفی باشد. چندین روش را برای تجزیه و تحلیل این نوع سری ها امتحان می کنیم تا ببینیم آیا می توانیم بعضی تغییرات نامنظم ظاهری را برحسب الگوهای احتمالی مانند الگوهای « میانگین متحرک »[۱] یا « خودبرگشت »[۲] و یا ترکیب این دو الگو بیان کنیم.

---

[1]-Moving Average

[2]-Auto Regressive

4-1: فرآیندهای تصادفی

یک فرآیند تصادفی ، خانواده ای از متغیرهای تصادفی $Z(W, t)$ است که W به فضای نمونه و t به یک مجموعه اعداد صحیح متعلق می باشد. برای یک ثابت t، $Z(W, t)$ یک متغیر تصادفی است. مجموعه ای متناهی از متغیرهای تصادفی $\{Z_{t_1}$ و Z_{t_2} و... و $Z_{t_n}\}$ را، از یک فرآیند تصادفی $\{Z(w,t): t = o, \pm 1, ...\}$ در نظر می گیریم. برای یک فرآیند با مقدار حقیقی $\{Z_t : t = 0, \pm 1, \pm 2, ...\}$ تابع میانگین را با:

$$\mu_t = E(Z_t) \tag{1-1}$$

تابع واریانس فرآیند را با $\sigma_t^2 = E(Z_t - \mu_t)^2$ و تابع کوواریانس بین Z_{t_1} و Z_{t_2} را با:

$$\gamma(t_1, t_2) = E(Z_{t_1} - \mu_{t_1})(Z_{t_2} - \mu_{t_2}) \tag{2-1}$$

و تابع خود همبستگی بین Z_{t_1} و Z_{t_2} را با:

$$\rho(t_1, t_2) = \frac{\gamma(t_1, t_2)}{\sqrt{\delta_{t_1}^2}\sqrt{\delta^2_{t_2}}} \tag{3-1}$$

تعریف می کنیم.

برای یک فرآیند مانای اکید، چون تابع توزیع برای تمام t ها یکسان است، تابع میانگین $\mu_t = \mu$ یک ثابت است، مشروط به اینکه $E(|Z_t|) < \infty$. بنابراین کوواریانس و همبستگی بین Z_t و Z_{t+k} فقط به تفاضل زمانی K بستگی دارد. (باکس-جنکینز، 1976)

5-1: توابع اتوکوواریانس و خود همبستگی

برای یک فر آیند مانا ی $\{Z_t\}$ میانگین $E(Z_t)=\mu$ و واریانس $Var(Z_t)=E(Z_t-\mu)^2$ ثابت اند.

وکوواریانس های $Cov(Z_t و Z_s)$ توابعی از اختلاف زمانی $|t-s|$ می باشند. بنابراین کواریانس بین Z_t و Z_{t+k} را به صورت:

$$\gamma_K = cov(Z_t و Z_{t+k}) = E(Z_t-\mu)(Z_{t+k}-\mu) \qquad (4\text{–}1)$$

و همبستگی بین Z_t و Z_{t+k} را به صورت:

$$\rho_K = \frac{Cov(Z_t, Z_{t+k})}{\sqrt{var(Z_t)}\sqrt{Var(Z_{t+K})}} = \frac{\gamma_K}{\gamma_0} \qquad (5\text{–}1)$$

می نویسیم، و توجه داریم که $Var(Z_t)=Var(Z_{t+K})=\gamma_0$ در تحلیل سری های زمانی γ_K که تابعی از K است تابع اتوکواریانس و ρ_K را تابع خودهمبستگی[1] (ACF) می نامند.

در تحلیل سری های زمانی، همبستگی بین Z_t و Z_{t+k} (بعد از اینکه وابستگی خطی مشترک متغیرهای Z_{t+1} و Z_{t+2} و...و Z_{t+k-1} را حذف نموده ایم) به‌عنوان خود همبستگی جز یی نامیده می شود که در واقع همبستگی شرطی زیر است.

$$Corr(Z_t, Z_{t+k} | Z_{t+1}, ..., Z_{t+k-1}) \qquad (6\text{–}1)$$

بنابراین خودهمبستگی جزیی بین Z_t و Z_{t+k} برابر خودهمبستگی معمول بین $(Z_t-\hat{Z}_t)$ و $(Z_{t+K}-\hat{Z}_{t+K})$ خواهد بود. لذا اگر ϕ_{KK} خودهمبستگی جزیی[1] بین Z_t و Z_{t+k} باشد داریم:

---
[1]-Auto correlation Function

$$\phi_{kk} = \frac{\text{cov}(Z_t - \hat{Z}_t)(Z_{t+k} - \hat{Z}_{t+k})}{\sqrt{\text{var}(Z_t - \hat{Z}_t)}\sqrt{\text{var}(Z_{t+k} - \hat{Z}_{t+k})}} \tag{7-1}$$

6-1: فرآیند تصادفی محض

هر فرآیند $\{U_t\}$ را که به صورت دنباله ای از متغیرهای تصادفی ناهمبسته از توزیعی با میانگین ثابت $E(U_t) = \mu$ (که معمولا صفر فرض می شود)، با واریانس ثابت $Var(U_t) = \sigma^2$ و تابع اتوکو واریانس $\gamma_K = \text{cov}(u_t, u_{t+k}) = 0$ (بازای $k \neq 0$) است یک فرآیند تصادفی محض می نامند. بنابراین نتیجه می شود که فرآیند تصادفی محض $\{U_t\}$ ماناست با تابع اتوکوواریانس

$$\gamma_k = \begin{cases} \sigma^2 & k = 0 \\ 0 & k \neq 0 \end{cases} \tag{8-1}$$

و تابع خودهمبستگی

$$\rho_k = \begin{cases} 1 & k = 0 \\ 0 & k \neq 0 \end{cases} \tag{9-1}$$

و تابع خودهمبستگی جزئی:

$$\phi_{kk} = \begin{cases} ١ & k = 0 \\ 0 & k \neq 0 \end{cases}$$

7-1: نمایش فرآیندهای خودبرگشت

فرض می کنیم $\{U_t\}$ یک فرآیند تصادفی محض با میانگین صفر و واریانس σ^2 باشد، فرآیند $\{Z_t\}$ را فرآیند خودبرگشت مرتبه p گویند هرگاه:

$$Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + ... + \phi_p Z_{t-p} + U_t$$

$$\left(1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p\right) Z_t = U_t \qquad (10-1)$$

$$\phi_p(B) Z_t = U_t$$

الگوی بالا در واقع یک الگوی رگرسیون چندگانه است با این تفاوت که در اینجا Z_t روی متغیرهای مستقل، رگرسیون نشده بلکه روی مقادیر گذشته Z_t رگرسیون شده است. به این دلیل است که فرآیند $\{Z_t\}$ را خودبرگشت نامیده اند. با توجه به اینکه $\sum_{j=1}^{p} |\phi_j| < \infty$ لذا فرآیند، همواره وارون پذیر است .در فرآیند خودبرگشت برای ماناییی، باید ریشه های $\phi_p(B) = 0$ از لحاظ قدر مطلق خارج دایره واحد، واقع شوند.

تابع خودبرگشت جزئی ϕ_{kk}بعد از تاخیر p صفر می شود. این یک خاصیت مفید در تشخیص الگوی خودبرگشت است.(باکس-جنکینز، 1976)

8-1: فرآیند میانگین متحرک

در نمایش میانگین متحرک یک فرآیند، $\Pi(B)Z_t = U_t$ اگر فقط تعدادی متناهی از وزنهای Π مخالف صفر باشند، یعنی $\Pi_1 = -\beta_1$ و $\Pi_2 = -\beta_2$ و... و $\Pi_q = -\beta_q$ و برای $\Pi_k = 0, k > q$ آنگاه فرآیند حاصل یک فرآیند یا الگوی میانگین متحرک از مرتبه q نامیده می شود، و به صورت MA(q) نشان می‌دهیم.

$$Z_t = U_t - \beta_1 U_{t-1} - .... - \beta_q U_{t-q} \quad (11-1)$$

یا

$$Z_t = \theta_q(B) U_t$$

$$\theta_q(B) = (1 - \beta_1 B - ... - \beta_q B^q) \quad (12-1)$$

چون $1 + \beta_1^2 + .. + \beta_q^2 < \infty$ لذا یک فرآیند میانگین متحرک، همواره مانااست. اگر ریشه های $\theta_q(B) = 0$ خارج دایره واحد باشد، فرآیند میانگین متحرک وارون پذیر است. به طور کلی تابع خود همبستگی یک فرآیند MA(q) بعد از تاخیر q قطع می شود. از این خاصیت برای تشخیص اینکه آیا سری زمانی معلوم به‌وسیله یک فرآیند میانگین متحرک تولید شده است، می توان استفاده کرد.

به طور خلاصه یک فرآیند AR(p) با مرتبه متناهی، با یک فرآیند MA با مرتبه نامتناهی، متناظر است و یک فرآیندMA(q) وارون پذیر متناهی، با یک فرآیند AR با مرتبه نامتناهی، متناظر است. این رابطه دوگانگی بین فرآیندهای AR(p) و MA(q) در توابع خودهمبستگی و خودهمبستگی جزئی نیز وجود دارد.

1-9: الگوهای مرکب

کلاس مهمی از الگوها برای سری های زمانی از ترکیب فرآیندهای $MA(q)$ و $AR(p)$ تشکیل می شوند.

یک فرآیند خودبرگشت – میانگین متحرک مرکب ARMA(p,q) شامل p جمله AR و q جمله MA است که به صورت زیر نمایش داده می شود:

$$Z_t = \phi_1 Z_{t-1} + ... + \phi_p Z_{t-p} + U_t - B_1 U_{t-1} - ... - B_q U_{t-q}$$

$$\phi_p(B) X_t = \theta_q(B) Z_t \qquad (13\text{-}1)$$

اهمیت فرآیندهای ARMA در این حقیقت، نهفته است که اغلب می توان یک سری زمانی مالی را با یک الگوی ARMA بیان نمود که نسبت به فرآیند MA یا AR به تنهایی، پارامترهای کمتری دارد. بنابراین تابع خود همبستگی یک الگوی ARMA(p,q) بعد از تاخیر q به سمت صفر میل می کند. درست مانند فرآیند AR(P) که فقط به پارامترهای خودبرگشت مدل، بستگی دارد. شرط مانایی و وارون پذیری مدل ARMA این است که ریشه های $\phi(B)$ و $\theta(B)$ خارج از دایره واحد باشند و $\phi(B)$ و $\theta(B)$ هیچ ریشه مشترکی نداشته باشند.

10-1: الگوهای سری زمانی زامانا

بسیاری از سری های زمانی کاربردی، مانا هستند. در این بخش، ساختار مفیدی از سری های زمانی ناماناى همگن، نظیر الگوهای خودبرگشت میانگین متحرک تلفیق شده (ARIMA) را بیان می کنیم. برای مرتبط ساختن الگوهای سری زمانی مانا و نامانا، چند تبدیل مفید برای تفاضلی کردن و پایدار کردن واریانس، وجود دارد.

1-10-1: نامانایی در میانگین

الگوهای روند قطعی و الگوهای روند تصادفی و تفاضلی کردن دو رده از الگوهایی هستند که در به الگو در آوردن سری های زمانی نامانا در میانگین مفیدند. تابع میانگین یک فرآیند نامانا را با یک روند قطعی از زمان، می توان نشان داد. در چنین حالتی، از یک الگوی رگرسیون استاندارد، برای بیان پدیده استفاده می شود. اگر تابع میانگین μ_t از یک روند خطی $\mu_t = \alpha_0 + \alpha_1 t$ پیروی کند از الگوی خطی قطعی زیر که Z_t تصادفی محض با میانگین صفر است می توان استفاده کرد:

$$Z_t = \alpha_0 + \alpha_1 t + U_t \tag{14-1}$$

برای یک تابع میانگین درجه دوم قطعی $\mu_t = a_0 + a_1 t + a_2 t^2$ از رابطه زیر استفاده می‌شود.

$$Z_t = a_0 + a_1 t + a_2 t^2 + U_t \tag{15-1}$$

و همچنین با یک روند قطعی از یک چند جمله ای از مرتبه k از زمان، فرآیند را می توان مانند بالا به الگو در آورد. که اغلب با توجه به تحلیل رگرسیون می توان الگوها را بررسی نمود.

گرچه بسیاری از سری های زمانی، نامانا هستند اما رفتار بخشهای مختلف، بسیار شبیه به هم بوده و اختلاف آنها در سطوح میانگین محلی است. باکس و جنکینز این نوع رفتار نامانایی را نامانایی همگن می نامند. با توجه به الگوهای ARMA، فرآیند ماناست هرگاه ریشه های چند جمله ای AR آن، خارج دایره واحد واقع نشوند. با وجود این، رفتار این نوع سری های نامانای همگن مستقل از سطح آن است. بنابراین، اگر $\varphi(B)$ عملگر خودبرگشت باشد، برای هر ثابت C داریم:

$$\varphi(B)(Z_t + C) = \phi(B)Z_t \qquad (16\text{-}1)$$

نتیجه این می شود که برای d>0 که $\phi(B)$ یک عملگر خودبرگشت مانا باشد، داریم که:

$$\varphi(B) = \phi(B)(1 - B) \qquad d > 0$$

بنابراین یک سری نامانای همگن را با تفاضلی کردن مناسب به یک سری مانا تبدیل می‌کنیم. اگر سری $\{Z_t\}$ نامانا باشد سری تفاضلی شده مرتبه d یعنی $\{(1-B)^d Z_t\}$، $d \geq 1$ ماناست.

1–10–2:الگوی کلی ARIMA

فرآیند مانای حاصل از یک سری نا مانای همگن تفاضلی شده، الزاما یک فرآیند تصادفی محض نیست. به طور کلی سری تفاضلی شده $(1-B)^d Z_t$ از یک فرآیند مانای ARMA(p,q) تبعیت می کند. بنابراین:

$$\phi_P(B)(1-B)^d Z_t = \theta_0 + \theta_q(B)U_t \qquad (17-1)$$

که در آن، عملگر AR مانای $\phi_P(B)=(1-\phi_1 B-...-\phi_P B^P)$ و عملگر MA وارون پذیر $\theta_q(B)$ که $\theta_q(B)=(1-\beta_1 B-...-\beta_q B^q)$ است دارای ریشه مشترک نیستند. پارامتر θ_0 برای d=0 و d>0 نقش های زیادی بازی می کند. اگر d=0، فرآیند اولیه، مانااست و یادآور می شویم که θ_0 به میانگین فرایند وابسته است. اگر $d \geq 1$ ، θ_0 را جمله روند قطعی گویند و اغلب، آن را از الگو حذف می کنند، مگر واقعا مورد نیاز باشد. الگوی ناماناى همگن در (17-1) را یک الگوی خودبرگشت تلفیق شده با میانگین متحرک از مرتبه (p,d,q) می نامند و به صورت ARIMA(p,d,q) نشان می دهند وقتی p=0 الگوی ARIMA(p,d,q) را یک الگوی میانگین متحرک تلفیق شده از مرتبه (d,q) می نامند و به‌صورت IMA(d,q) نشان داده می شود.

1-10-3: نامانایی در واریانس و اتوکوواریانس

فرآیندی که مانای در میانگین است، الزاما مانای در واریانس و اتوکوواریانس نیست. با وجود این، یک فرایند نامانای در میانگین، نامانای در واریانس و کوواریانس نیز خواهد بود. با توجه به اینکه تمام مسائل نامانایی را، با تفاضلی کردن نمی توان انجام داد، در مورد سری های زمانی که مانا در میانگین ولی نا مانا در واریانسند، از یک تبدیل پایداری واریانس استفاده می نمائیم. به طور کلی، برای تبدیل واریانس، از تبدیل توانی که به‌وسیله باکس و کاکس (1964) معرفی شده است، می توان استفاده نمود.

11–1: مراحل شناسایی الگو

جهت شناخت الگو، الگوی ARIMA(p,d,q) را در نظر می گیریم.

$$(1-\alpha_1 B-...-\alpha_p B^p)(1-B)^d Z_t = \theta_0 + (1-\beta_1 B-...-\beta_q B^q)U_t \qquad (18\text{–}1)$$

شناخت الگو، به تبدیلات پایداری واریانس و تبدیلات تفاضلی کردن و تصمیم در مورد منظور کردن پارامتر غیر تصادفی θ. وقتی $d \geq 1$ و مرتبه های درست p ، q برای الگو مربوط می شود. اگر یک سری زمانی معلوم باشد، برای شناخت الگوی آزمایشی مراحل زیر را در نظر می گیریم که باکس و جنکینز از این روش به عنوان مرحله شناسایی الگو یاد کرده و پیشنهاد کرده اند که این کار با استفاده از آزمون ACF و PACF انجام شود. اولین مرحله در تحلیل یک سری زمانی، رسم نمودار داده هاست. معمولاً از روی نمودار به وجود روند، فصلی بودن، نقاط بیرونی، واریانس غیر ثابت و پدیده های غیر طبیعی و ناماناى دیگر می توان پی برد. که در این صورت یک تبدیل داده ها، فراهم می شود. معمولاً تبدیلات پایدار کردن واریانس و تفاضلی کردن مورد استفاده قرار می گیرند. غالباً قبل از تفاضلی کردن، از تبدیلات پایداری واریانس استفاده می شود. زیرا، تفاضلی کردن، ممکن است چند مقدار منفی را به وجود آورد. داده های تبدیل شده (تبدیل باکس و کاکس) به عنوان سری اولیه در نظر گرفته می شود. در مرحله بعد، با استفاده از ACF و PACF سری اولیه درجه لازم تفاضلی کردن را معین می نمائیم. همچنین برای تعیین مرتبه های p و q (بالاترین درجه خودبرگشت و میانگین متحرک) ACF و PACF نمونه سری که تبدیل و تفاضلی شده است را محاسبه می نماییم. همانطور که قبلاً بیان شده است یک دوگانگی قوی، بین الگوهای AR و MA بر حسب ACF و PACF های آنها وجود دارد.

برای ساختن یک الگوی ARIMA کمال مطلوب، این است که حداقل 50 مشاهده داشته باشیم و تعداد ACF و PACF های نمونه که باید محاسبه شوند، در حدود $\frac{n}{4}$ باشد.

وقتی $d > 0$ جمله روند قطعی را آزمون می کنیم. برای یک الگوی ن امانا $\phi(B)(1-B)^d Z_t = \theta_0 + \theta(B)U_t$ معمولا پارامتر θ_0 را حذف می کنند. به طوری که توانایی نمایش دادن سری با تغییرات تصادفی در سطح، ضریب زاویه یا روند را داشته باشد.

روش باکس– جنکینز برای آزمون مطلوبیت مدل براساس خاصیت های ACF و PACF در مدل های ARIMA است و عبارتند از:

1) نامانایی: برای یک سری زمانی نامانا مقادیر ACF کاهش می یابند و PACF دارای یک مقدار بزرگ مثبت یا منفی در تاخیر اول است . با تفاضلی کردن مناسبی (معمولا $d \le 2$) می توان به یک سری مانا دست یافت. تفاضلی کردن بیش از حد، باعث ایجاد نامانایی کاذب می باشد.

2) رفتار PACF برای یک فرآیند خودبرگشت: برای یک فرآیند خودبرگشت با مرتبه p مقدار PACF در تاخیر های 1, 2,…,p غیر صفر و در جاهای دیگر صفر است.

3) رفتارACF یک فرآیند میانگین متحرک: برای یک فرآیند میانگین متحرک از مرتبه q مقدار ACF در تاخیر های 1,2,…,q غیر صفر ودر جاهای دیگر صفر است.

4)نامانایی فصلی: برای یک سری زمانی نامانای فصلی ، ACF در تاخیرهایی به جز S,2S,… صفر است و به تدریج کاهش می یابد. این نامانایی را می توان با تفاضلی کردن یعنی $\nabla_s Z_t, \nabla_s^2 Z_t, \ldots, \nabla_s^D Z_t$ (معمولا D=1 است) برطرف کرد.

5) رفتار PACF برای یک فرآیند خودبرگشت فصلی: برای یک سری خودبرگشت فصلی با مرتبه P مقدار PACF در تاخیرهای S,2S,...,PS غیر صفر و برای تاخیر بیشتر از PS صفر است.

6) رفتارACF یک فرآیند میانگین متحرک فصلی: برای یک فرآیند میانگین متحرک فصلی از مرتبه Q مقدار ACF در تاخیرهای S,2S,...,QS غیر صفر و برای تاخیرهای بزرگتر از QS صفر است.

12-1: برآورد پارامترها

روشهای معمول که برای برآورد پارامترها در مدلهای سری زمانی کاربرد دارند شامل روشهای گشتاوری کمترین مربعات ، حداکثر درستنمایی و کمترین مربعات شرطی و غیر شرطی هستند.

روش گشتاوری همانطور که باکس-جنکینز (1976) نشان داده اند . برآوردهای رضایت بخشی را در مورد مدل های میانگین متحرک به‌دست نمی ده ند بنابراین در اینجا از آنها استفاده نخواهد شد. روشهای دیگر نیز سرانجام به یک نتیجه کلی که همان حداقل کردن مجموع مربعات خطاها است خواهند رسید و بنابراین تنها روش حداکثر درستنمایی را بررسی کرده و جزئیات آن را صرفا برای الگوهای خودبرگشت ، بیان می‌کنیم.