بسم الله الرحمن الرحیم

پردیس علوم
دانشکده‌ی ریاضی، آمار و علوم کامپیوتر
پروژه کارشناسی

رشته
آمار و کاربرد ها

عنوان

ارائه مدلی جهت پیش بینی نرخ رشد اقتصادی کشور چین

نگارنده

رضا صالحی

استاد پروژه

دکتر سودابه شمه سوار

تیر ۹۶

تقدیم به روح پدرم...

رضا صالحی

تیر ۱۳۹۶

چکیده

در این پروژه قصد داریم با استفاده از نرم افزارهای تحلیل آماری R و با بهره بردن از روش های سری زمانی، داده های فصلی مربوط به شاخص رشد اقتصادی کشور چین(GDP) ، (از سال ۱۹۵۴-۲۰۱۶میلادی) را مدل بندی کنیم. پیش از برازش مدل به اختصار به بیان مفاهیم اولیه مرتبط با سری زمانی می پردازیم. در فصل دوم این پروژه به تفصیل مفاهیم مهم الگوهایی برای سری‌های زمانی ایستا را بیان می کنیم. فصل سوم شامل تجزیه و تحلیل داده های نام برده با استفاده از نرم افزارهای مدل های سری زمانی خواهد بود که در آن از الگوهای فصلی و غیر فصلی AR و MA و ARMA و ARIMA استفاده می کنیم. قسمت آخر این فصل به تعبیری مهم‌ترین قسمت آن خواهد بود. زیرا قصد داریم با استفاده از مدل سری‌زمانی برازش داده شده مقادیر آینده را پیش‌بینی کنیم.

کلیدواژه: رشد اقتصادی، سری‌های زمانی، مدل‌بندی،کشور چین

پیشگفتار

مبحث ((سری‌های زمانی))، از جمله مباحث آماری است که کاربردهای بسیاری در علوم مختلف دارد. نتایج حاصل از به‌کارگیری مدل‌های سری‌های زمانی، معمولا بسیار قابل توجه‌اند و پیش‌بینی‌هایی که از این طریق به‌دست می‌آیند، نسبتا دقیق‌اند. به همین دلیل، از این نتایج می‌توان استفاده‌های بسیاری در برنامه‌ریزی‌های کلان به عمل آورد. با توجه به وجود داده‌های بسیاری به‌صورت سری‌زمانی در نظام آماری کشورها، بدیهی است که روش‌های مدل بندی سری‌زمانی می‌تواند تاثیر شگرف و عمده‌ای در راستای تولید دانش آماری به‌طور اعم و کاربردهای گسترده‌ای در نظام آماری به‌طور اخص داشته باشد.

در این پروژه تاثیرات رشد اقتصادی کشور چین مورد بحث است که تاثیرات رشد اقتصادی بر سیاست‌های کشورها یکی از مهمترین موضوعات است و بی‌شک یک چالش بزرگ برای هر کشور نرخ رشد اقتصادی آن میباشد که باعث ورود یا خروج سرمایه‌گزاران کلان و یا میلیاردها دلار منابع مالی می‌شود. در این راستا دولت‌ها برنامه‌های اقتصادی خود را به گونه‌ای تدوین میکنند که بتوانند با توجه به ظرفیت‌ها به بیشترین حد رشد اقتصادی برسند که برای اینکار قطعا چشم انداز و پیش بینی‌هایی برای آینده نیز خواهند داشت.

در ابتدای این پروژه لازم است که مفهوم رشد اقتصادی و عوامل تاثیرگزار بر آن را بیان کنیم. رشد اقتصادی به تعبیر ساده عبارت است از افزایش تولید یک کشور در یک سال خاص در مقایسه با مقدار آن در سال پایه. در سطح کلان، افزایش تولید ناخالص داخلی(GDP) در سال مورد بحث به نسبت مقدار آن در یک سال پایه، رشد اقتصادی محسوب می‌شود. علت اینکه برای محاسبه‌ی رشد اقتصادی، از قیمت‌های سال پایه استفاده می‌شود آن است که افزایش محاسبه شده در تولید ناخالص ملی، ناشی از افزایش میزان تولیدات باشد و تاثیر افزایش قیمت‌ها(تورم) حذف گردد.

منابع رشد اقتصادی عبارت‌اند از:

۱-افزایش نهاده‌های تولید(افزایش سرمایه یا نیروی کار)

۲-افزایش بهره‌وری عوامل تولید

۳-به‌کارگیری ظرفیت‌های احتمالی خالی در اقتصاد

هدف ما در این پروژه پیش بینی نرخ رشد اقتصادی کشور چین که در سال‌های اخیر نوسان‌های بسیار داشته با استفاده از مدل‌های سری‌های زمانی می‌باشد.

کتاب تجزیه و تحلیل سری های زمانی به قلم جاناتان دی کرایر از منابع مناسب این حوزه است که آقای حسینعلی نیرومند [۱] آن را ترجمه کرده اند. کتاب مقدمه ای بر تحلیل سری های زمانی از سی چتفیلد هم منبع خوبی در این حوزه است که توسط آقایان حسینعلی نیرومند و ابوالقاسم بزرگ نیا [۲] ترجمه شده است.

فهرست مطالب

فصل ۱

آشنایی با سری زمانی

۱.۱ مفهوم سری زمانی، مثال‌ها و تاریخچه

سری‌های زمانی، مشاهداتی‌اند که در طول زمان،گردآوری می‌شوند.کثرت چنین مشاهداتی تحلیل سری‌های زمانی را به یکی از کاربردی‌ترین شاخه‌های علم آمار تبدیل کرده است. به عنوان مثال، متوسط درجه‌ی حرارت، رطوبت یا سرعت باد که بسته به نیاز،به طور روزانه، هفتگی یا ماهانه ثبت می‌شوند، قیمت یک سهام خاص یا مقدار یک شاخص کل در بازار بورس، مقدار تقاضا، تولید یا فروش محصولات یک شرکت، درامد آن،و مبلغی که بابت تبلیغ محصولات خود صرف می‌کند، تعداد گردشگران و درامد حاصل از صنعت گردشگری، مشخصه‌هایی از کیفیت کالاهای تولیدی، مثل تعداد نقص‌ها، طول عمر، میزان چسبندگی یا غلظت، که در طی فرایند ساخت، اندازه‌گیری می‌شوند نمونه‌هایی از سری‌های زمانی هستند. زمان ثبت مشاهدات، غالبا گسسته است(مثال‌های فوق) و در بیشتر موارد، فاصله زمانی بین مشاهدات متوالی، ثابت است. اما این زمان می‌تواند پیوسته هم باشد.به‌عنوان مثال، وضعیتی را در نظر بگیرید که فعال (۱) یا غیرفعال (۰) بودن یک شبکه‌ی خدمات‌رسانی رایانه‌ای به‌طور پیوسته ثبت می‌شود، یا جریان حجمی آب خروجی از یک منبع به‌طور پیوسته اندازه‌گیری می‌شود. در این پروژه،آن دسته از سری‌های زمانی را مطالعه می‌کنیم که زمان ثبت گسسته با فاصله‌های زمانی ثابت دارند.

هر چند توصیف رفتار یک سری‌زمانی از لحاظ تغییرات مقطعی و درازمدت در آن، یا مطالعه‌ی وابستگی‌های موجود بین عناصر سری، از بررسی‌های متداولی است که در مورد سری‌های زمانی انجام می‌شود، می‌توان گفت که مهمترین هدف از تحلیل سری‌زمانی، پیش‌بینی مقادیر آینده آن

است. در صورتی که با بیش از یک سری سر و کار داشته باشیم، مطالعه‌ی روابط بین این سری‌ها را نیز می‌توان به این هدف اضافه کرد.این روابط، در صورت وجود، علاوه بر اینکه دست‌یابی به پیش‌بینی‌های دقیق‌تر را ممکن می‌سازد، در کنترل رفتار یک سری‌زمانی با استفاده از سری‌های مرتبط با آن نیز کاربرد دارند. در این پروژه با روش‌هایی برای تحلیل سری‌های زمانی یک متغیره آشنا می‌شویم.تجزیه و تحلیل سری زمانی به طور نظری و عملی از زمان شروع کار اصلی جورج ای.پی.باکس و ام.جیکنز در (۱۹۷۰)، وی(۱۹۹۰)، براک‌ول و دیویس(۱۹۹۶) به سرعت توسعه پیدا نموده است.

قبل از آغاز مطالعات باکس و جنکینز، و به‌خصوص قبل از انتشار کتاب برجسته‌ی آنان در سال ۱۹۷۰، غالبا شکل ظاهری سری‌زمانی، ملاک تعیین مدل قرار می‌گرفت و در ساختار مدل‌هایی که برای این منظور به کار می‌رفتند، توابعی از زمان برای بیان تغییرات درازمدت در سطح سری اضافه می‌شد و توابعی متناوب از زمان برای بیان تغییرات منظم در سطح سری. به چنین مدل‌هایی مدل‌های کلاسیک می‌گویند.

۲.۱ مفاهیم بنیادی

پیدا کردن الگوهای مناسب برای سری‌های زمانی کاری مهم می‌باشد. ما یک استراتژی چند مرحله‌ای را برای ساختن یک الگو توسعه می‌دهیم که توسط باکس و جنکینز وضع شده است.
در این روش سه مرحله‌ی عمده وجود دارد که از هر یک از آنها ممکن است چندین بار استفاده کنیم.

۱. تشخیص (یا شناسایی الگو)

۲. برازش الگو

۳. تشخیص درستی الگو

در تشخیص یا شناسایی الگو، دسته‌ای از الگوهای سری‌های زمانی را که برای سری زمانی مشاهده شده مناسب است انتخاب می‌کنیم. در این مرحله نمودار زمانی سری را مورد توجه قرار داده و آماره‌های متفاوت زیادی را از داده‌ها محاسبه می‌کنیم و همچنین از اطلاعات خود در زمینه‌ی موضوعی مانند اقتصاد، فیزیک، شیمی یا بیولوژی که داده‌ها از آن جا ناشی شده‌اند استفاده می‌کنیم. تاکید می‌کنیم که الگویی که در این مرحله انتخاب می‌شود آزمایشی است و به تجدید نظر نیاز خواهد داشت.

۳.۱ معرفی توابع و روابط مهم در مدل‌های سری زمانی

در یک مدل سری زمانی همواره دارای یک مجموعه مشاهدات یا به طور علمی‌تر متغیر تصادفی که در طول زمان ثبت شده می‌باشیم.که بصورت زیر نشان می‌دهیم.

$$\{z_t : t = 0, 1, ...\}$$

تابع میانگین: تابع میانگین همان امید ریاضی متغیر تصادفی مورد نظر می‌باشد که در طول زمان مورد بررسی قرار می‌گیرد که به صورت زیر نمایش می‌دهیم:

$$m = E(z_t)$$

تابع اتوکوواریانس: تابع اتوکوواریانس به صورت زیر تعریف می‌شود:

$$g_{t,s} = cov(z_t, z_s)$$

تابع خود همبستگی: تابع خود همبستگی به صورت زیر تعریف می شود:

$$corr(z_t, z_s) = \frac{cov(z_t,z_s)}{\sqrt{var(z_t)}\sqrt{var(z_s)}}$$

با توجه به تعاریف بالا می‌توان نتایج زیر را به‌دست آورد :

۱. $r_{t,t} = 1$ $\quad$ $g_{t,t} = var(z_t)$

۲. $r_{t,s} = r_{s,t}$ $\quad$ $g_{t,s} = g_{s,t}$

۳. $|r_{t,s}| \leq 1$ $\quad$ $|g_{t,s}| \leq \sqrt{g_{t,t} * g_{s,s}}$

۴.۱ ایستایی

مفهوم اساسی ایستایی این است که قوانین احتمالی حاکم بر فرآیند با زمان تغییر نمی‌کند، یعنی فرآیند در تعادل آماری است، به خصوص فرایند تصادفی y_t را ایستایی اکید می‌گوییم. هرگاه توزیع توام $y_{t_1}, y_{t_2}...y_{t_n}$ به ازای همه‌ی انتخاب‌های نقطه‌های زمانی $t_1, ...t_n$ و همه‌ی انتخاب‌های تاخیر k همانند توزیع توام $y_{t_1-k}, y_{t_2-k}...y_{t_n-k}$ باشد.
اگر فرآیندی اکیداً ایستا و واریانس متناهی داشته باشد آنگاه تابع کوواریانس فقط بایستی به تأخیر زمانی بستگی داشته باشد.

حالت دیگری از ایستایی به نام ایستایی ضعیف می‌باشد که دارای شرایط زیر می‌باشد.

۱. تابع میانگین در طول زمان ثابت باشد.

۲. برای تمام زمان های t و تاخیر k داریم:

$$g_{t,t-k} = g_{0,k}$$

به عنوان مثالی از فرآیندهایی که دارای ایستایی ضعیف هستند می‌توان اغتشاش خالص و موج کسینوسی را نام برد که دارای شرایط فوق هستند.

فصل ۲

الگو هایی برای سری زمانی ایستا

در این فصل می‌خواهیم به معرفی مدل‌هایی بسیار مهم در سری زمانی یعنی میانگین متحرک (MA) ، اتو رگرسیو (AR) و مدل اتو رگرسیو و میانگین متحرک (ARMA) و ارائه‌ی ویژگی‌های آن می‌پردازیم.

۱.۲ فرآیندهای میانگین متحرک (MA)

فرمول کلی میانگین متحرک به صورت زیر می‌باشد:

$$y_t = e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - ... - \theta_q e_{t-q}$$

این قبیل سری‌ها را میانگین متحرک از مرتبه‌ی q نامیده و به صورت مختصر MA(q) نشان می دهند.

اصطلاح میانگین متحرک از این حقیقت ناشی می‌شود که y_t با استفاده از وزن‌های $(1, -\theta_1, -\theta_2, ..., -\theta_p)$ برای متغیرهای $(e_t, e_{t-1}, ..., e_{t-q})$ حاصل می‌شود. و سپس با انتقال وزن‌ها و به کار بستن آنها در مورد $(e_{t-q+1}, e_{t-q+2}, ..., e_t, e_{t+1})$ متغیر y_{t+1} را به‌دست می‌آوریم و همین‌طور تا آخر،به پیش می‌رویم. این الگو اولین بار توسط سلاتسکی و والد مورد بررسی قرار گرفت. می‌توان به عنوان مثال به بررسی میانگین متحرک مرتبه‌ی اول بپردازیم:

$$y_t = e_t - \theta e_{t-1}$$

واضح است که:

- $E(y_t) = 0$
- $\gamma_0 = var(y_t) = \sigma_e^2(1+\theta^2)$
- $cov(y_t, y_{t-1}) = -q\sigma_e^2$

که σ_e^2 واریانس e_t می باشد.

و به طور خلاصه داریم:

- $E(y_t) = 0$
- $\gamma_0 = var(y_t) = \sigma_e^2(1+\theta^2)$
- $\gamma_1 = -q\sigma_e^2$
- $\rho_1 = \frac{-\theta}{1+\theta^2}$
- $\gamma_k = \rho_k = 0 \qquad k \geq 2$

به طور کلی می‌توان نشان داد که فرمول تابع خود همبستگی فرآیند MA(q) کلی را به سهولت می‌توان به‌دست آورد:

$$\rho_k = \begin{cases} \frac{-\theta_k+\theta_1\theta_{k+1}+\theta_2\theta_{k+2}+...+\theta_{q-k}\theta_q}{1+\theta_1^2\theta_2^2+...+\theta_q^2} & k = 1,2,3,...,q \\ 0 & k > q \end{cases}$$

۲.۲ فرآیندهای اتو رگرسیو (AR)

فرآیندهای اتورگرسیو به طوری که از اسمشان پیداست رگرسیون روی خودشان می‌باشد. به خصوص فرآیند اتو رگرسیو مرتبه‌ی p ام $\{y_t\}$ در معادله‌ی زیر صدق می‌کند:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} + e_t$$

که $\{e_t\}$ اغتشاش خالص می‌باشد.

به مانند قسمت قبل در این قسمت نیز به بررسی فرآیند اتو رگرسیو مرتبه‌ی اول می‌پردازیم:

- $y_t = \phi y_{t-1} + e_t$
- $\gamma_0 = \sigma_e^2(1+\phi^2)$
- $\gamma_k = \phi\gamma_{k-1}$
- $\rho_k = \frac{\gamma_k}{\gamma_0} = \phi^k$

چون $1 \leq |\phi|$ ،لذا با افزایش تعداد تأخیرهای k تابع خود همبستگی یک منحنی نمایی نزولی خواهد بود.

اگر $1 > \phi > 0$ باشد، تمام همبستگی ما مثبت و اگر $0 > \phi > -1$ همبستگی در تأخیر یک منفی است $(\rho_1 = \phi)$ و علامت‌های خود همبستگی‌های متوالی وقتی مقدارشان به طور نمایی تنزل کند ، از مثبت به منفی تغییر می‌کند. می‌توان تابع اتو رگرسیو مرتبه‌ی اول را به صورت بازگشتی به تابعی بر حسب اغتشاش خالص تبدیل کرد: $y_t = e_t + \phi_1 e_{t-1} + \phi_2 e_{t-2} + \phi_3 e_{t-3} + \ldots$

در حالت کلی می‌توان بیان کرد که:

$$\gamma_0 = \frac{\sigma_e^2}{1-\phi_1\rho_1-\phi_2\rho_2-...\phi_p\rho_p}$$

که σ_e^2 واریانس اغتشاش خالص می باشد :

$$\rho_k = \phi_1\rho_{k-1}+\phi_2\rho_{k-2}+...+\phi_p\rho_{k-p} \qquad k \geq 1$$

۳.۲ فرآیندهای اتو رگرسیو - میانگین متحرک(ARMA)

اگر قسمتی از سری را اتو رگرسیو و بخشی از آن را میانگین متحرک فرض کنیم، در این صورت یک الگوی سری زمانی کلی به‌دست می‌آید. به طور کلی اگر داشته باشیم:

$$Y_t = \phi_1Y_{t-1}+\phi_2Y_{t-2}+...+\phi_pY_{t-p}+e_t-\theta_1e_{t-1}-...-\theta_qe_{t-q}$$

گوییم $\{y_t\}$ یک فرآیند مرکب اتو رگرسیو – میانگین متحرک به ترتیب از مرتبه‌ی p و q است و آن را به صورت مختصر ARMA(p،q) نشان می دهیم.
می‌توان نشان داد که:

- $\gamma_0 = \phi\gamma_1 + [1-\theta(\phi-\theta)]\sigma_e^2$
- $\gamma_1 = \phi\gamma_0 - \theta\sigma_e^2$
- $\gamma_k = \phi\gamma_{k-1} \qquad k \geq 2$

از حل دو معادله اول γ_0 بدست می آید:

$$\gamma_0 = \frac{1-2\phi\theta^2}{1-\phi^2}\sigma_e^2$$

و از حل معادله بازگشتی ساده نتیجه می شود که :

$$\rho_k = \frac{(1-\theta\phi)}{1-2\theta\phi+\theta^2}\theta^{k-1} \qquad k \geq 1$$

در پایان این فصل ذکر این نکته ضروری می‌باشد که سه الگوی معرفی شده یعنی میانگین متحرک، اتو رگرسیو و اتو رگرسیو $-$ میانگین متحرک دارای شرایط لازم ایستایی ضعیف هستند.

۴.۲ مدل‌های ARIMA

در صورت عدم فرض ایستایی میانگین در مدل ARMA مدل جدیدی به نام ARIMA وجود دارد به این صورت که ابتدا روی داده هایی که ایستایی میانگین ندارند، یک تفاضل اعمال می کنیم و سپس مدل ARMA به داده های تفاضلی شده برازش می دهیم که به فرم زیر نمایش می دهیم:
$ARIMA(p,d,q)$

فصل ۳

تحلیل با استفاده از سری‌های زمانی

در این فصل قصد داریم با استفاده از نرم افزار R به تجزیه و تحلیل داده های مربوط به شاخص رشد اقتصادی (GDP) کشور چین طی سال های ۱۹۵۴ تا ۲۰۱۶ بپردازیم.
قبل از چگونگی استفاده از نرم‌افزار بهتر است در مورد مفهوم تفاضلی کردن توضیحاتی داده شود. زمانی که یک مجموعه داده ناایستا باشد (میانگین ثابت نباشد) از تفاضلی کردن استفاده می‌شود. همچنین در صورتی که واریانس مجموعه‌ی مشاهدات برابر نباشد با استفاده از تبدیلات باکس-کاکس می‌توان اقدام به ایستا کردن واریانس کرد. اگر با استفاده از تبدیلات باکس-کاکس واریانس ثابت نشد آنگاه دارای سری‌زمانی ناایستا در واریانس خواهیم بود که در این‌صورت با کمک مدل‌هایی همچون ARCH یا GARCH مشکل حل خواهد شد. اکنون می‌توانیم به بیان مراحل برازش مدل‌های سری زمانی به یک مجموعه مشاهدات در نرم‌افزار R بپردازیم.

۱.۳ تجزیه و تحلیل به کمک نرم افزار R

در این روش ابتدا داده های وارد شده در نرم افزار را به شکل سری زمانی در می آوریم. یعنی دوره‌ی تکرار و نقطه ی آغاز فرایند را تعیین می کنیم. سپس نمودار پراکنش داده ها acf و pacf را رسم می‌کنیم.

۱.۱.۳ وارد کردن داده های سری زمانی

در ابتدای کار ، داده های شاخص رشد اقتصادی را به صورت Vector وارد نرم افزار می کنیم:

>GDP=c$(15, 6, 4.3, 6.9, 15, 5.1, 21.3, 9, 0, -27.27, -5.58, 10.3, 18.18, 16.95, 10.65, -5.57,$
$-4.1, 16.94, 19.3, 7.06, 3.81, 7.76, 2.31, 8.72, -1.57, 7.57, 11.667, 7.6, 7.807, 5.172, 8.934,$
$10.835, 15.139, 13.443, 8.94, 11.869, 11.235, 4.186, 3.907, 9.294, 14.216, 13.868, 13.052, 10.949,$
$9.928, 9.231, 7.838, 7.667, 8.492, 8.34, 9.131, 10.111, 10.111, 11.396, 12.719, 14.231, 9.654, 9.4,$
$10.636, 9.536, 7.856, 7.758, 7.298, 6.914)$

>GDP

```
15.600   4.300   6.900  15.000   5.100  21.300   9.000   0.000 -27.270
-5.580  10.300  18.180  16.950  10.650  -5.570  -4.100  16.940  19.300
 7.060   3.810   7.760   2.310   8.720  -1.570   7.570  11.667   7.600
 7.807   5.172   8.934  10.835  15.139  13.443   8.940  11.869  11.235
 4.186   3.907   9.294  14.216  13.868  13.052  10.949   9.928   9.231
 7.838   7.667   8.492   8.340   9.131  10.111  10.111  11.396  12.719
14.231   9.654   9.400  10.636   9.536   7.856   7.758   7.298   6.914
```

سپس داده ها را به شکل سری زمانی در می آوریم. نقطه آغاز این سری هم سال ۱۹۵۴ است.

>GDP=ts(GDP،freq=$1, start = c(1954, 1)$)

>GDP

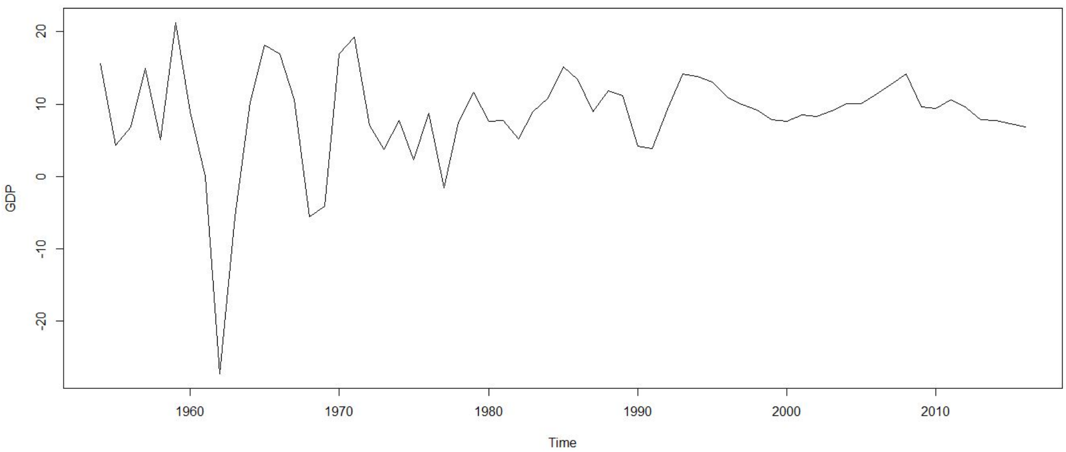
```
Time Series:
Start = 1954
End = 2016
Frequency = 1
 [1]  15.600   4.300   6.900  15.000   5.100  21.300   9.000   0.000 -27.270
[10]  -5.580  10.300  18.180  16.950  10.650  -5.570  -4.100  16.940  19.300
[19]   7.060   3.810   7.760   2.310   8.720  -1.570   7.570  11.667   7.600
[28]   7.807   5.172   8.934  10.835  15.139  13.443   8.940  11.869  11.235
[37]   4.186   3.907   9.294  14.216  13.868  13.052  10.949   9.928   9.231
[46]   7.838   7.667   8.492   8.340   9.131  10.111  10.111  11.396  12.719
[55]  14.231   9.654   9.400  10.636   9.536   7.856   7.758   7.298   6.914
```

۲.۱.۳ رسم نمودار پراکنش داده ها

در این مرحله نمودار پراکنش داده ها را رسم می کنیم.

```
>plot.ts(GDP)
```

شکل ۱.۳: نمودار پراکنش داده‌ها



۳.۱.۳ بررسی رگرسیون داده ها

قبل از برازش مدل‌های سری زمانی ابتدا بررسی می‌کنیم، بین زمان و داده رابطه ی رگرسیونی برقرار می‌باشد یا خیر. از آنجایی که زمان به عنوان متغیر کمکی یا پیشگو برای مشاهدات می‌باشد، ابتدا یک بردار به عنوان زمان از ۱ تا تعداد مشاهدات ایجاد می‌کنیم. سپس یک بردار به عنوان توان دوم ستون زمان ایجاد می‌کنیم زیرا ممکن است رابطه‌ی رگرسیونی از نوع درجه دوم باشد. بعد از اجرای رگرسیون خطی، به بررسی معنادار بودن ضرایب و استقلال باقیمانده‌ها می‌پردازیم.

>t1=1:length(GDP)

>t1

```
 1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25
26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50
51 52 53 54 55 56 57 58 59 60 61 62 63
```

>t2=t1^2

>t2

```
 1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25
26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50
51 52 53 54 55 56 57 58 59 60 61 62 63
```

حال رگرسیون داده ها را برازش می دهیم:

>regression=lm(GDP$\sim t1 + t2$)

>summary(regression)

```
Call:
lm(formula = GDP ~ t1 + t2)

Residuals:
    Min      1Q  Median      3Q     Max
-34.088  -1.796   0.066   3.132  14.802

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  5.8107149  2.7399704   2.121   0.0381 *
t1           0.1197558  0.1975504   0.606   0.5467
t2          -0.0008727  0.0029917  -0.292   0.7715
```

با توجه به مقدار پی-مقدار پارامترهای t۱ و t۲ که کمتر از مقدار استاندارد یعنی ۰.۰۵ است پس می توان نتیجه گرفت که مدل رگرسیونی مناسب نمی‌باشد.

از آنجایی که پارامتر Intercept معنادار نمی باشد و فرض صفر بودن β_0 رد می شود پس مدل ما بصورت زیر می شود:

$y = \beta_0 + \varepsilon$

۴.۱.۳ بررسی استقلال باقی مانده ها

برای چک کردن استقلال باقی مانده ها از آزمون runs از پکیج randtests استفاده می کنیم. فرض H_0 این آزمون، استقلال باقی مانده هاست. اگر باقیمانده‌ها مستقل باشند، آن‌گاه مدل رگرسیونی کفایت می‌کند.

اگر باقیمانده‌ها مستقل نباشند، به سراغ مدل‌های سری زمانی می‌رویم.

```
>R=resied(regression)
>library(randtests)
>runs.test(R)
```

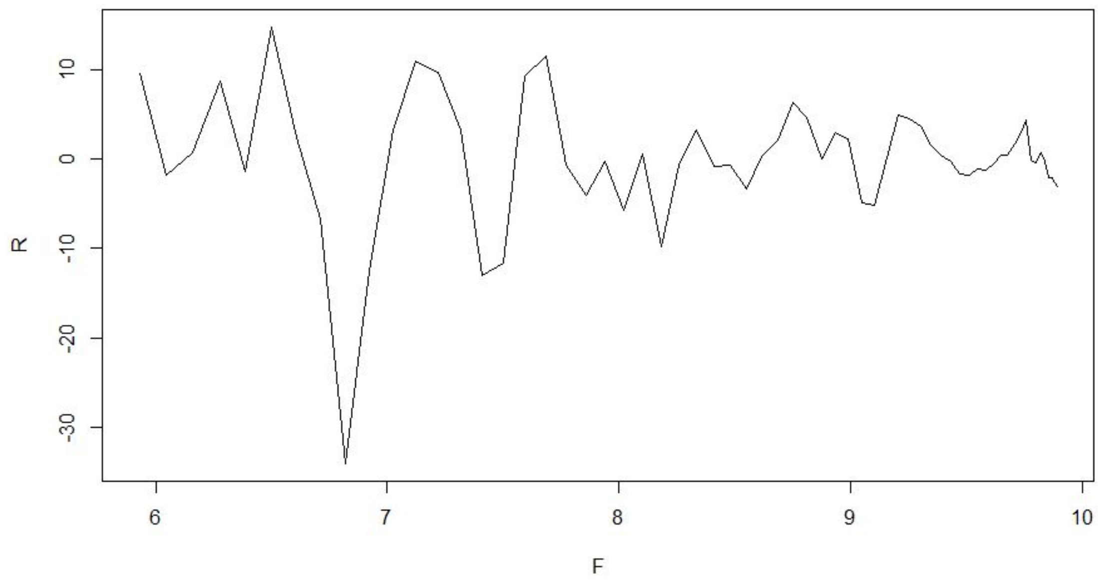
```
        Runs Test

data:  R
statistic = -2.5611, runs = 22, n1 = 31, n2 = 31, n = 62, p-value =
0.01043
alternative hypothesis: nonrandomness
```

پی - مقدار این آزمون کمتر از ۰.۰۵ است پس فرض استقلال باقیمانده‌ها رد می شود و باقیمانده‌ها به هم وابسته هستند.

در نتیجه مدل رگرسیونی مناسب نمی‌باشد و باید سراغ مدل‌های سری‌زمانی برویم. اگر بخواهیم نتیجه بالا را از روی شکل نتیجه بگیریم ابتدا مقادیر برازش داده شده را بدست می‌آوریم و این مقادیر را درمقابل باقیمانده رسم میکنیم.

```
>F=fitted(regression)
>plot(F،R،type=”l”)
```

شکل ۲.۳: شکل وابستگی مانده‌ها



۵.۱.۳ بررسی نرمال بودن باقیمانده‌ها

برای چک کردن نرمال بودن باقی مانده ها از آزمون shapiro.test استفاده می کنیم. فرض H_0 این آزمون نرمال بودن توزیع باقیمانده‌هاست.

```
>shapiro.test(R)

        Shapiro-Wilk normality test

data:  R
W = 0.84431, p-value = 1.294e-06
```

همان طور که مشخص است، پی - مقدار این آزمون کمتر از ۰.۰۵ است پس فرض نرمال بودن باقیمانده‌ها رد می شود.

۶.۱.۳ بررسی ایستایی میانگین

همانطور که در قسمت های قبل دیدیم، مدل رگرسیونی مناسب نمی باشد. پس حالا می خواهیم مدل سری زمانی مورد نظر را بدست آوریم. قبل از برازش مدل سری زمانی باید ایستایی میانگین و واریانس داده ها چک شود.

برای چک کردن ایستایی میانگین، ابتدا اعداد ۱ تا ۷ را بصورت نه تا نه تا بصورت vector با استفاده از دستور rep می‌سازیم:

>factor=rep($1:7, each=9$)

>factor

```
1 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2 2 3 3 3 3 3 3 3 3 3 4 4 4 4 4 4 4 4 4 5 5
5 5 5 5 5 5 5 6 6 6 6 6 6 6 6 6 7 7 7 7 7 7 7 7 7
```

حال Factor در نظر می گیریم و ایستایی میانگین داده های اصلی (GDP) را چک می‌کنیم.

kruskal.test(GDP $\sim factor$)

```
        Kruskal-Wallis rank sum test

data:  GDP by factor
Kruskal-Wallis chi-squared = 7.2932, df = 6, p-value = 0.2946
```

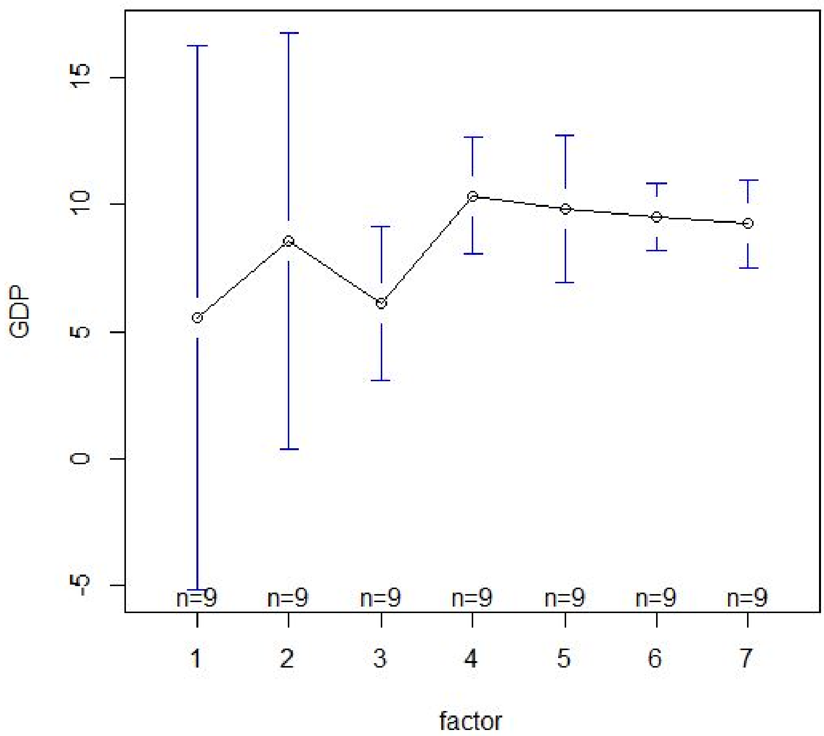
فرض H_0 آزمون فوق، ایستایی میانگین داده هاست.

با توجه به پی-مقدار آزمون فوق که بیشتر از ۰.۰۵ است، پس فرض ایستایی میانگین قبول می‌شود.
حال میخواهیم نتیجه بالا را با استفاده از شکل نیز نتیجه بگیریم.برای این کار از دستور plotmeans در پکیج gplots استفاده میکنم.

>library(gplots)

>plotmeans(GDP $\sim factor$)

شکل ۳.۳: ایستایی میانگین



۷.۱.۳ بررسی ایستایی واریانس داده‌ها

در این مرحله می‌خواهیم ایستایی واریانس داده‌ها را که در مرحله ی قبل از میانگین ایستا شد، چک کنیم. برای این کار باید از آزمون لی‌ون استفاده کنیم. برای استفاده از آزمون لی‌ون باید از پکیج Rcmdr استفاده کنیم.

>library(Rcmdr)

>levene.test(GDP،factor)

```
Levene's Test for Homogeneity of Variance (center = median)
      Df F value   Pr(>F)
group  6  3.7689 0.003201 **
      56
```

فرض H_0 آزمون فوق ، ایستایی واریانس داده هاست.

در اینجا پی-مقدار آزمون کمتر از ۰.۰۵ است فرض ایستایی واریانس پذیرفته نمی‌شود و باید سراغ تبدیلات boxcox برویم.

توجه تبدیلات boxcox شامل اعداد منفی نمی‌شود بنابراین ما باید مینیمم داده‌ها (GDP) را بگیریم و به همه ی آنها به علاوه یه مقدار مثبت(به دلیل اینکه داده ها صفر نشوند) در ستون بردار GDP.new اضافه کنیم.

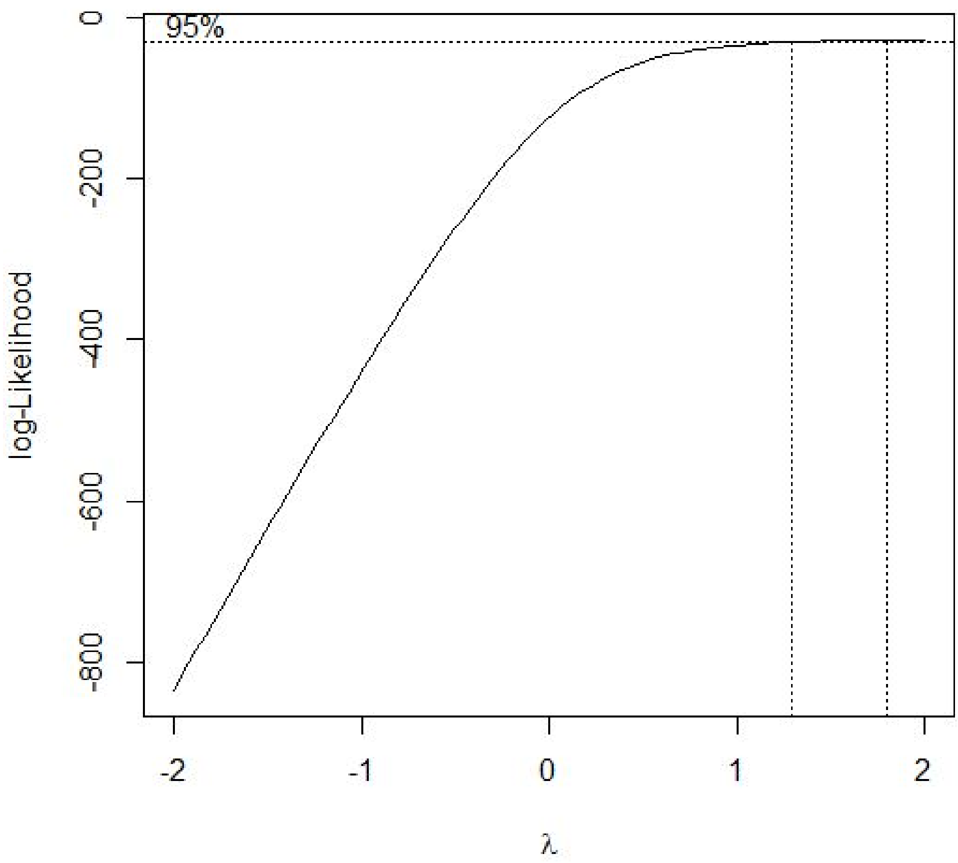
```
>m=min(GDP)
>m
-27.27
```

```
>GDPnew=GDP+27.30
>GDPnew
Time Series:
Start = 1954
End = 2016
Frequency = 1
 [1] 42.900 31.600 34.200 42.300 32.400 48.600 36.300 27.300  0.030 21.720
[11] 37.600 45.480 44.250 37.950 21.730 23.200 44.240 46.600 34.360 31.110
[21] 35.060 29.610 36.020 25.730 34.870 38.967 34.900 35.107 32.472 36.234
[31] 38.135 42.439 40.743 36.240 39.169 38.535 31.486 31.207 36.594 41.516
[41] 41.168 40.352 38.249 37.228 36.531 35.138 34.967 35.792 35.640 36.431
[51] 37.411 37.411 38.696 40.019 41.531 36.954 36.700 37.936 36.836 35.156
[61] 35.058 34.598 34.214
```

حال تبدیل boxcox را بر روی بردار GDP.new اعمال می‌کنیم. برای استفاده از تبدیل boxcox باید پکیج MASS را فراخوانی بنماییم.

```
>library(MASS)
>regression1=lm(GDPnew~ t1+t2)
>boxcox=regresssion1،lambda=seq(−2،2،0.1)
```

شکل ۴.۳: نمودار تبدیل باکس-کاکس



طبق تبدیلی که انجام شد یک مقدار از روی شکل بدست می‌آید که ما باید داده‌ها را به توان این مقدار برسانیم. این مقدار تقریبا برابر ۱.۵ است.

>GDP1=GDP$^{1.5}$

حال دوباره به بررسی ایستایی میانگین می پردازیم.

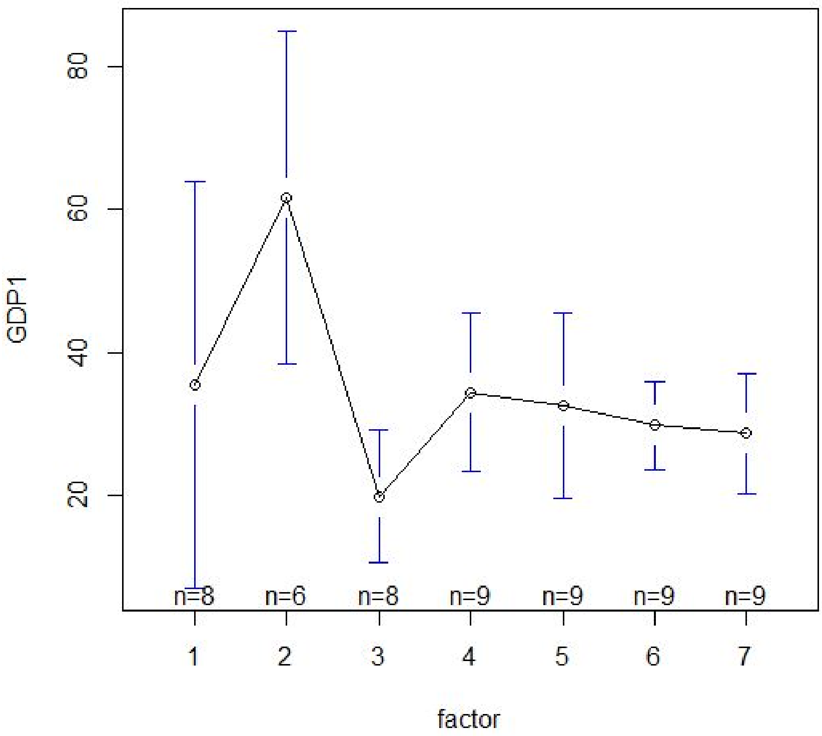
ایستایی میانگین

ایستایی میانگین داده های جدید (GDP.new) را چک می کنیم.

>kruskal.test(GDP1$\sim$ $factor$)

```
        Kruskal-Wallis rank sum test

data:  GDP1 by factor
Kruskal-Wallis chi-squared = 14.679, df = 6, p-value = 0.02291
```

شکل ۵.۳: عدم ایستایی میانگین



فرض H_0 آزمون فوق، ایستایی میانگین داده هاست.

با توجه به پی-مقدار آزمون فوق که کمتر از ۰.۰۵ است پس فرض ایستایی میانگین رد می شود.

برای رفع این مشکل و ایستا کردن میانگین، یک مرتبه داده هایمان را تفاضلی می کنیم.

>kruskal.test(GDP2=diff(GDP1))

ای دستور داده‌های راGDP۱ یک مرتبه تفاضلی کرده و در GDP۲ می‌ریزد.

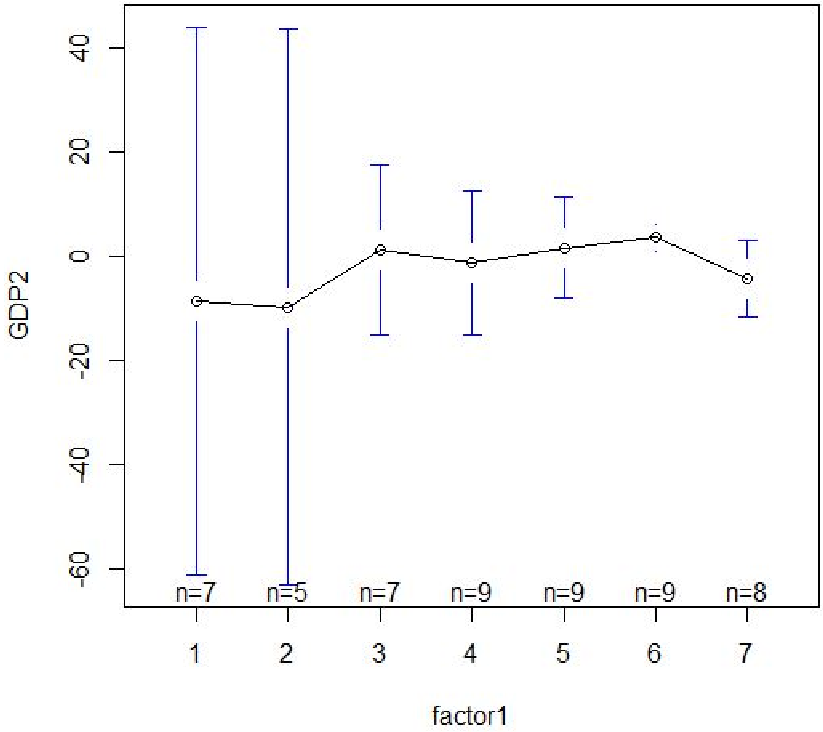
حال دوباره ایستایی میانگین را چک می‌کنیم.

>factor1=factor[1:62]

>kruskal.test(GDP2∼ $factor1$)

```
        Kruskal-Wallis rank sum test

data:  GDP2 by factor1
Kruskal-Wallis chi-squared = 3.186, df = 6, p-value = 0.7852
```

شکل ۶.۳: ایستایی میانگین



همانطور که مشاهده می‌شود با توجه به پی-مقدار آزمون که بزرگتر از ۰.۰۵ می باشد پس فرض ایستایی میانگین پذیرفته می‌شود.

ایستایی واریانس

در این مرحله دوباره می‌خواهیم ایستایی واریانس داده‌ها را که در مرحله‌ی قبل از میانگین ایستا شد، چک کنیم.

```
>levene.test(GDP2،factor1)
```

```
Levene's Test for Homogeneity of Variance (center = median)
      Df F value    Pr(>F)
group  6  6.0158 0.0001002 ***
      47
```

فرض H_0 آزمون فوق، ایستایی واریانس داده هاست.

همانطور که مشاهده می‌کنید بعد از تبدیل boxcox باز هم پی-مقدار آزمون کمتر از ۰.۰۵ است و فرض ایستایی واریانس پذیرفته نمی‌شود .
اصطلاحا در چنین مواردی می گوییم که boxcox قدرت کافی برای ایستا کردن واریانس را نداشته است.
حال که در این داده‌ها ایستایی در میانگین ثابت شد ولی واریانس دارای ناایستایی می‌باشد. باید بررسی شود که آیا مدل دارای اثر آرچ و گارچ می‌باشد یا خیر. که در این مدل ابتدا یک مدل مناسب برای میانگین ثابت شده به‌دست خواهیم‌آورد. سپس یک مدل دیگر برای واریانس که ثابت نشده‌است.

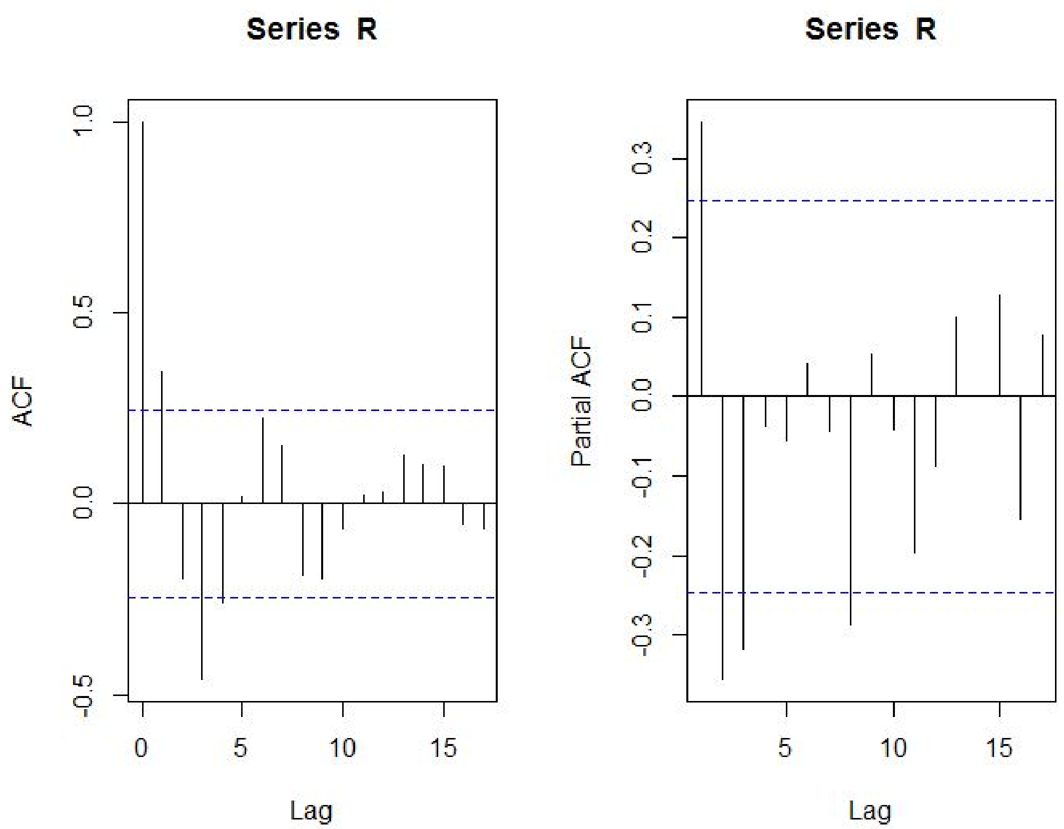
۸.۱.۳ برازش مدل سری زمانی

نمودار های acf و pacf را برای تشخیص ترم های AR و MA و زدن حدس‌هایی در مورد مدل مطلوبمان رسم می‌کنیم و سپس مدل اولیه را برازش می دهیم. در صورت پذیرش یا رد این فرض مدل را تغییر داده و دوباره آن را تست می‌کنیم. آن قدر این کار را ادامه می‌دهیم تا به بهترین مدل ممکن برای داده‌ها برسیم.

رسم نمودار acf و pacf

```
>par(mfrow=c(1،2))
>acf(R)
>pacf(R)
```

شکل ۳.۷: نمودار acf و pacf



برازش مدل

با توجه به این که در این آزمون به ما پی-مقدار داده نشده است باید به دنبال راهی برای آزمون ضرایب مدل سری‌زمانی باشیم. ما می‌خواهیم معنا داری ضرایب معادله سری‌زمانی را بررسی کنیم. توزیع برآوردگر بطور حدی نرمال می‌باشد، بنابراین برای فرض $H_0 : \psi_i = 0$ مقدار آماره آزمون برابر است با مقدار ضریب ψ_i در جدول ضرایب تقسیم بر انحراف معیار نظیر آن.

چون در آماره ی آزمون انحراف معیار را برآورد می‌کنیم توزیع آماره ی آزمون t خواهد بود. درجه آزادی این توزیع برابر است با تعداد مشاهدات منهای تعداد پارامتر های برآورد شده می‌باشد. اگر قدرمطلق آماره ی آزمون از $t_{\alpha,n-p}$ بیشتر باشد حضور ضریب مورد نظر در مدل معنادار می‌باشد. البته در تعداد بالا فرق چندانی ندارد زیرا توزیع t به نرمال استاندارد میل می‌کند.

برای مدل زدن ما نیازمند به استفاده از پکیج forecast هستیم. بنابراین پکیج forecast را فراخوانی می‌کنیم و بعد با دستور arima به مدل‌زدن می‌پردازیم.

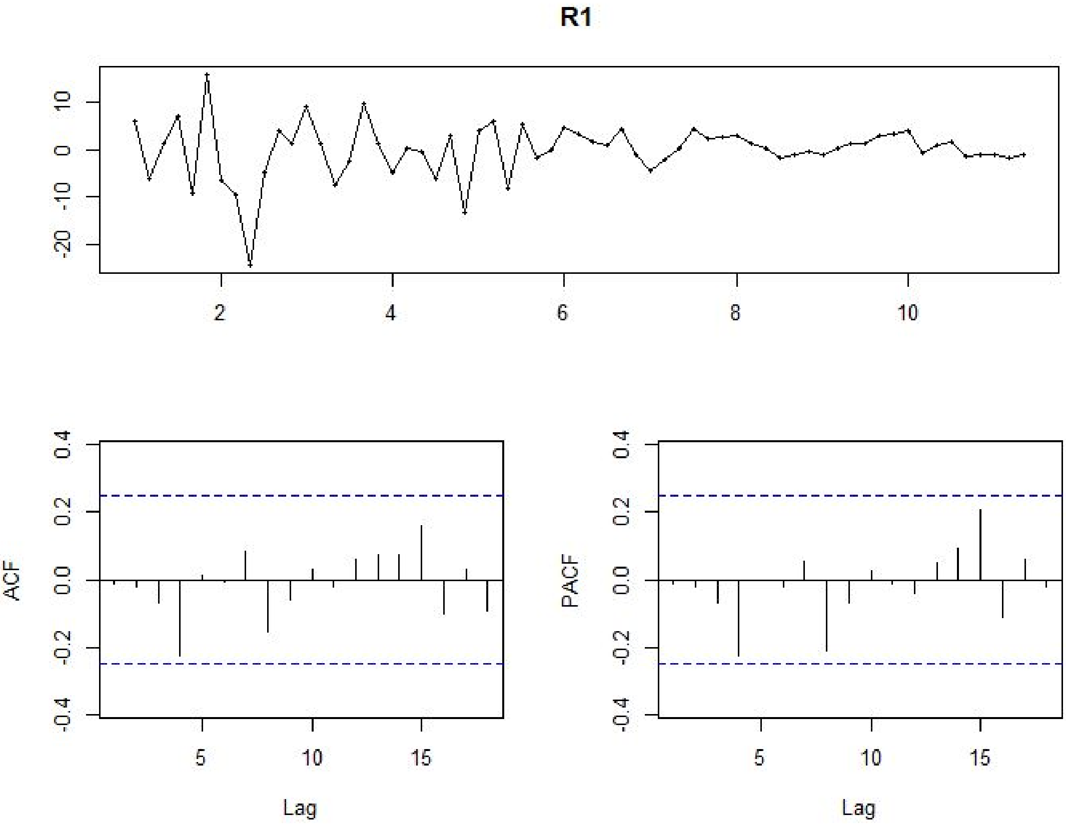
```
> library(forecast)
> GDP=ts(GDP,frequency=6)
> model= Arima(GDP, order = c(2,0,2), seasonal = c(0,0,1))
> (1-pnorm(abs(model$coef)/sqrt(diag(model$var.coef))))*2

         ar1          ar2          ma1          ma2         sma1
1.963162e-03 1.745103e-02 0.000000e+00 6.039613e-14 5.690978e-02
```

با توجه به مقادیر پی-مقدار که همه کمتر از ۰.۰۵ است پس فرض صفر بودن مدل‌های آریما رد خواهد شد. یعنی مدل می‌تواند مدل قابل‌قبولی باشد.

برای اینکه بفهمیم آیا چیزی از مدل جا مانده است یا خیر یک‌بار دیگر نمودارهای acf و pacf را روی باقیمانده‌ها می‌کشیم.

```
>R1=resied(model)
>F1=fitted(model)
>tsdisplay(R1)
```

شکل ۸.۳: نمودار acf و pacf مدل



با توجه به اینکه تمام ترم‌ها در باند قرار گرفته است پس نتیجه خواهیم گرفت که مدل بدست‌آمده می‌تواند مدل مناسبی برای ایستایی میانگین باشد.

پس مدل $ARIMA(2,0,2)*(0,0,1)$ پذیرفته می‌شود.

۹.۱.۳ بررسی وجود اثر آرچ و گارچ

برای این‌کار ابتدا باید نمودارهای ACF و PACF را برای باقیمانده‌های مدل آریما بدست‌آوریم سپس آزمون Archtest را بر روی باقیمانده‌های مدل تست می‌کنیم که آیا مدل دارای اثر آرچ میباشد یا خیر

برای این‌کار از دستور ArchTest در پکیجFinTS استفاده خواهیم‌کرد

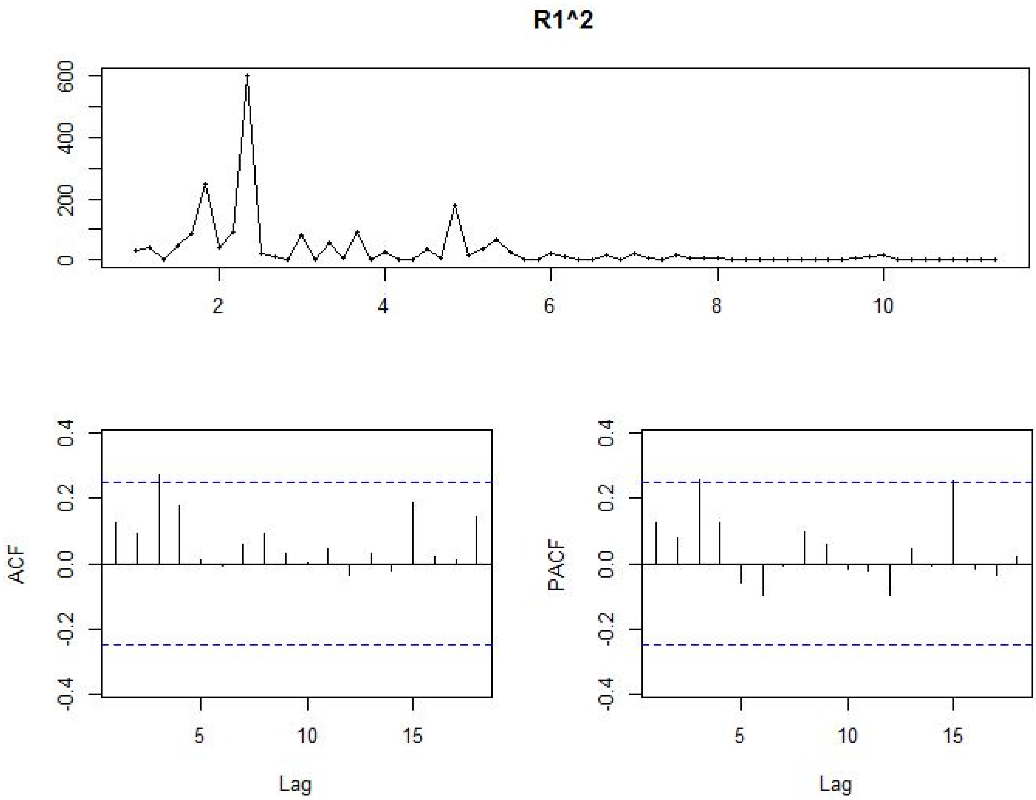
```
>library(FinTs)
>ArchTest(R1)
```

```
        ARCH LM-test; Null hypothesis: no ARCH effects

data:  R1
Chi-squared = 17.522, df = 12, p-value = 0.131
```

رسم نمودار acf و pacf

$>$tsdisplay($R1^2$)

شکل ۹.۳: نمودار acf و pacf



فرض H_0 آزمون فوق، نداشتن اثر آرچ روی باقیمانده‌های مدل است.

با توجه به پی-مقدار که بیشتر از ۰.۰۵ است. پس فرض نداشتن اثر مدل آرچ رد نخواهد شد پس داده‌های این پروژه اثر مدل آرچ را نخواهند داشت. با توجه به شکل ACF و PACF که در ترم سوم هردو شکل از باند بیرون زده است. پس روند فصلی مدل را از شش به سه تغییر می‌دهیم.

۱۰.۱.۳ برازش مدل اصلاح یافته

برای مدل زدن ما نیازمند به استفاده از پکیج forecast هستیم. بنابراین پکیج forecast را فراخوانی می‌کنیم و بعد با دستور arima به مدل زدن می‌پردازیم.

```
> GDP=ts(GDP,frequency=3)
> model1= Arima(GDP, order = c(2,0,1), seasonal = c(0,0,3))
> (1-pnorm(abs(model1$coef)/sqrt(diag(model1$var.coef))))*2
```

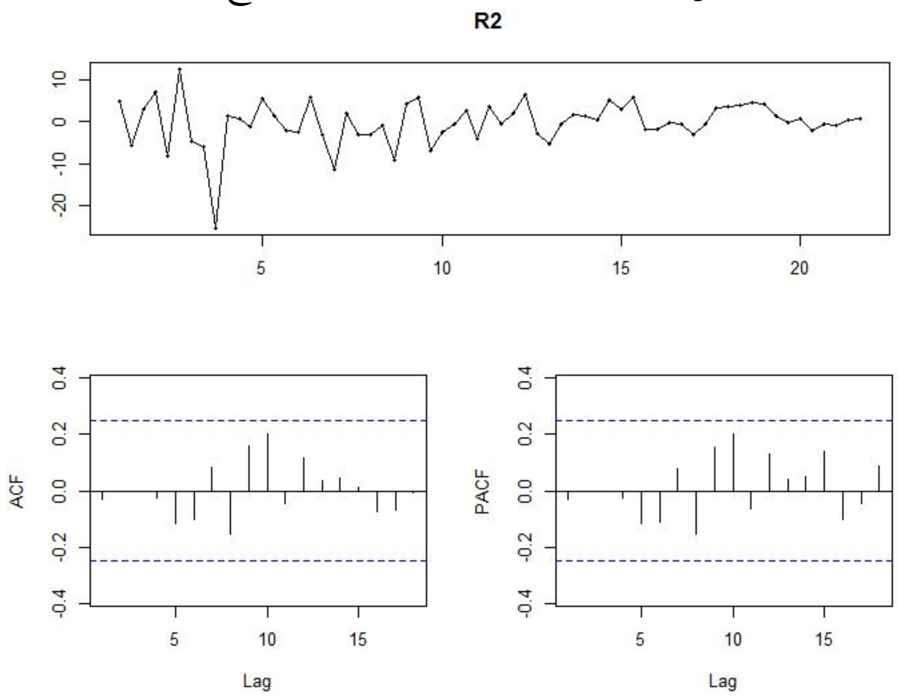
```
         ar1          ar2          ma1         sma1         sma2         sma3 
1.660876e-04 1.139068e-02 2.464695e-14 1.467996e-03 3.806047e-03 7.877402e-04 
   intercept 
0.000000e+00 
```

با توجه به مقادیر پی-مقدار که همه کمتر از ۰.۰۵ است پس فرض صفر بودن مدل‌های آریما رد خواهد شد یعنی مدل می‌تواند مدل قابل‌قبولی باشد.

برای اینکه بفهمیم آیا چیزی از مدل جا مانده است یا خیر یک‌بار دیگر نمودارهای acf و pacf را روی باقیمانده‌ها می‌کشیم.

>R$2 = resied(model)$

شکل ۱۰.۳: نمودار acf و pacf مدل اصلاح یافته



با توجه به اینکه تمام ترم‌ها در باند قرار گرفته‌اند پس نتیجه میگیریم که مدل، مدل خوبی می‌باشد
پس مدل مناسب برای این داده‌ها $ARIMA(2,0,1) * (0,0,3)$ می‌باشد

۱۱.۱.۳ پیش بینی

یکی از اهداف اصلی سری‌های زمانی پیش‌بینی است. در این مرحله قصد داریم رشد اقتصادی کشور چین را در سال ۲۰۱۷ تا ۲۰۲۰ پیش‌بینی کنیم. برای پیش‌بینی از دستور forecast استفاده می‌کنیم.

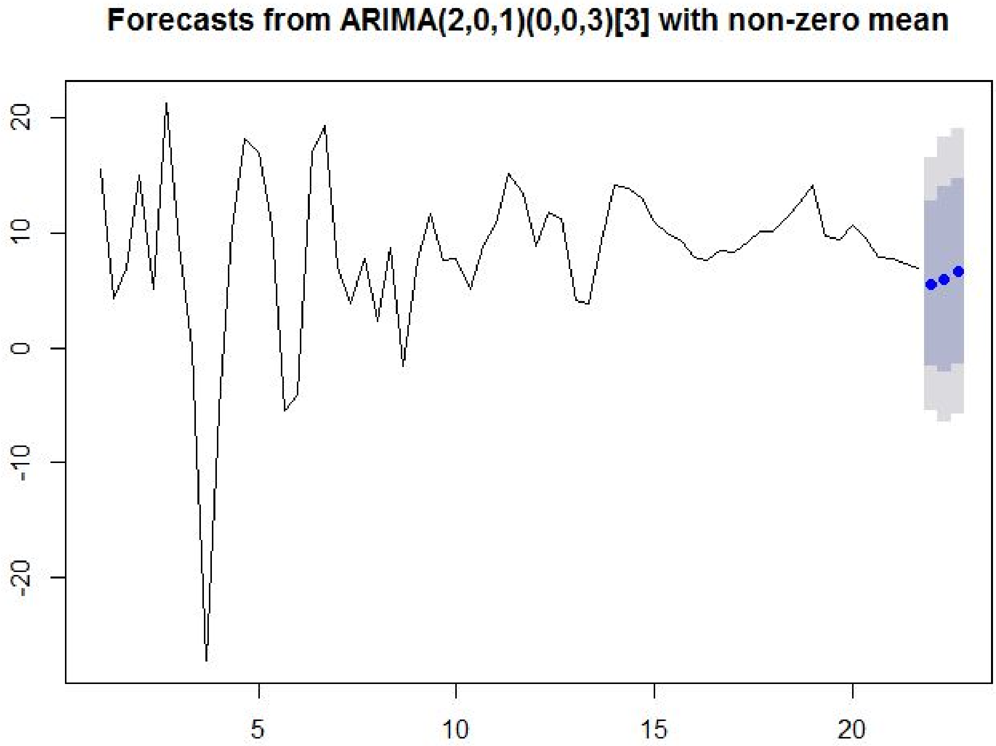
>forecast(model1،h=3)

```
      Point Forecast      Lo 80    Hi 80     Lo 95    Hi 95
22.00       5.591358 -1.627100 12.80982 -5.448321 16.63104
22.33       5.951768 -2.175190 14.07873 -6.477341 18.38088
22.67       6.647348 -1.499585 14.79428 -5.812311 19.10701
```

همانطور که می بینیم مقادیر پیش‌بینی شده‌ی ۳ گام بعد برآورد شده‌است. نمودار این سری را به همراه مقادیر پیش‌بینی شده می‌توانیم بکشیم:

```
>plot(forecast(model1،h=3))
```

شکل ۳.۱۱: نمودار پیش‌بینی داده‌ها



۱۲.۱.۳ نتیجه‌گیری

پژوهش حاضر به‌منظور پیش‌بینی نرخ رشد اقتصادی(GDP) کشور چین انجام گرفت. که با استفاده از روش‌های سری‌های زمانی و با استفاده از الگوهای AR،MA،ARMA،ARIMA به این نتیجه دست پیدا کردیم. داده‌های به‌کار رفته در این پروژه از سایت بانک اطلاعات جهانی استخراج شده است. پیش‌بینی نرخ رشد اقتصادی کشور چین بر اساس آمار و اطلاعات و پیش‌بینی‌هایی که فعالان اقتصادی این کشور انجام داده‌اند چیزی حدود ۶درصد در سال ۲۰۱۷ بوده‌است.که نتیجه‌ی پیش‌بینی این پروژه ۵.۶ است برای سال ۲۰۱۷ که به مقدار پیش‌بینی شده توسط دولت و فعالان اقتصادی کشور چین بسیار نزدیک است. که نشان می‌دهد پیش‌بینی بدست‌آمده مورد قبول می‌باشد.

واژه‌نامه فارسی-انگلیسی

اتورگرسیو	$autoregressive(AR)$
اکیدا ایستا	$strictly\ stationary$
الگو	$model$
ایستا	$stationary$
پیش‌بینی	$forecast$
تابع خودهمبستگی	$autocorrelation\ function(ACF)$
تابع خودهمبستگی جزیی	$partial\ autocorrelation\ function(PACF)$
ترم	lag
تفاضلی کردن	$differencing\ series$
سری‌های زمانی	$time\ series$
سری‌های زمانی اقتصادی	$economic\ time\ series$
سری‌های زمانی پیوسته	$continuous\ time\ series$
سری‌های زمانی فصلی	$seasonal\ time\ series$
سری‌های زمانی گسسته	$discrete\ time\ series$
سری‌های زمانی مانا	$stationary\ time\ series$
فرآیند خطی	$general\ linear\ process$
کوواریانس	$covariance$
مانده‌ها	$residual$

میانگین	*mean*
میانگین متحرک	*moving average(MA)*
نمودار سری‌زمانی	*time series plot*
نوفه‌ی سفید	*white noise*
واریانس	*variance*

منبع

تجزیه و تحلیل سری‌های زمانی/ تالیف جاناتان دی. کرایر،ترجمه حسین‌علی نیرومند/ مشهد:دانشگاه فردوسی/۱۳۷۱

مقدمه‌ای بر تحلیل سری‌های زمانی/ تالیف سی‌چتفیلد/ ترجمه حسین‌علی نیرومند، ابوالقاسم بزرگ‌نیا/ مشهد:دانشگاه فردوسی، ۱۳۷۲

آشنایی با تحلیل سری‌های زمانی به کمک نرم‌افزار s-plus /مجتبی خزایی/ تهران:پژوهشکده آمار/ ۱۳۸۲

سری‌های زمانی/ ابوالقاسم بزرگ‌نیا، حسین‌علی نیرومند/ انتشارات پیام‌نور/ ۱۳۹۰

مباحث ویژه در R / تالیف سیدسعید موسوی ندوشنی/ تهران:دانشگاه شهید عباسپور/ ۱۳۹۱

R در عمل و تحلیل داده و ترسیم نمودارها/ تالیف:رابرت کاباکوف، ترجمه:رمضان خسروی، امیرحسین امیری/ ۱۳۹۵

Abstract

The aim of this Project is to model the seasonal data of economic growth index(GDP) in china using Time Series method and R-Statistic Analysis software from 1954 to 2016 . before processing the model the basic concepts related to Time Series are demonstrated briefly. in second chapter of this Project main concepts of models for static Time Series are presented in detail. The data using software of Time Series models in which seasonal and non-seasonal models of AR، MA، ARMA،ARIMA، are used. The last section of this chapter in the most important part because we are willing to forecast the future values of GDP by the fitted Time Series model.

keyword: Economic Growth، Time Series، Modeling، China Country

College of Science

School of Mathematics , Statistics , and Computer Science

Bachelor Project

Title

A model for predicting economic growth rate in china

Author

Reza Salehi

Advisor

Dr.Shemme Savar

Jul 2017