به نام خدا

نام و نام خانوادگی: مژگان قنبرعلی

شماره دانشجویی : 9411351033

پروژه کارشناسی : سری های زمانی با مینی تب

دانشگاه سمنان دانشکده علوم پایه رشته آمار

استاد راهنما : دکتر زال زاده

سری های زمانی به همراه نرم افزار مینی تب 18

*تعریف سری زمانی

دنباله ای از داده ها که در یک محدوده زمانی جمع آوری شده اند ، یک سری زمانی را تشکیل می دهند . به عبارت دیگر یک سری زمانی عبارت است از داده هایی که از مشاهده یک پدیده در طول زمان به دست آمده اند .

*انواع داده های سری زمانی

۱-پیوسته : مشاهدات به طور پیوسته برحسب زمان در نظر گرفته شود مانند ولتاژ و . .

۲- گسسته : مشاهدات به طور منظم در فاصله های مساوی ثبت کنیم مانند نرخ بهره و . .

*هدف از تجزیه و تحلیل سری زمانی

۱-توصیف داده ها (که بتوانیم یک الگو را کشف کنیم)

۲-مدل سازی (رفتار تصادفی یک متغیر در طی زمان را با استفاده از یک مدل احتمال مدل سازی کنیم)

۳-پیش بینی (برای یک پیش بینی درست نیاز به یک مدل درست داریم)

۴-کنترل (که اگر یک شک اتفاق به سیستم ما وارد شد ما بتوانیم اثرات آن را تخمین بزنیم و پیش بینی کنیم)

*مولفه های یک سری زمانی

الگوی رفتاری یک سری زمانی را معمولا میتوان به ۴ مولفه تفکیک کرد :

۱-روند Trend

عبارت است از تغییرات دراز مدت در میانگین سری زمانی ، سیر طبیعی سری زمانی در دراز مدت و یا به عبارت دیگر حرکات رو به بالا و پایین یک سری زمانی.

۲-تغییرات فصلی Seasonal variation

تغییراتی هستند که در دوره ای تناوبی کوتاه پیش می آیند به عبارتی این تغییرات مربوط به عواملی هستند که به طریقی منظم و چرخه ای روی یک دوره کمتر از یک سال عمر می کنند.

۳-تغییرات دوره ای Cyclical variation

تکرار حرکات رو به بالا و پایین حول سطوح روند و این نوع تغییرات دارای دوره نوسان بیشتر از یک سال می باشند.

۴-تغییرات نامنظم Irregular variation

عبارت است از حرکات پراکنده در یک سری زمانی که از الگوی منظم و مشخصی پیروی نمی کنند که این تغییرات کاملا تصادفی بوده و به وسیله عواملی مانند سیل ، زلزله و . . . . به وجود می آید.

*مدل های سری زمانی

نحوه ارتباط و اثراتی که مولفه های گفته شده در سری زمانی دارند، مدل سری زمانی را تعیین می کند که دو الگوی کلی را می توان در نظر گرفت :

۱-<u>الگوی جمعی</u> در این الگو فرض می شود که ۴ مولفه ی گفته شده از یکدیگر مستقل هستند و به این صورت نوشته می شود

$$Yi = Ti + Ci + Si + Ii$$

۲-<u>الگوی ضربی</u> در این الگو فرض می شود که ۴ مولفه ی فوق به یکدیگر وابسته هستند و به این صورت نوشته می شود

$$Yi = Ti * Ci * Si * Ii$$

*انواع سری های زمانی

۱-سری زمانی ایستا(مانا) : ساکن اند و دنباله ای از مقادیر وابسته به زمان است که میانگین و واریانس آن به زمان وابسته نباشند.

۲-سری زمانی ناایستا(نامانا) : ساکن نیستند و میانگین و واریانس و کوواریانس آن به زمان مرتبط هستند و وابسته اند.

تابع اتوکوواریانس در تاخیر K : این تابع، همبستگی بین مشاهداتی را که k واحد زمانی با یکدیگر اختلاف دارند اندازه گیری می گیرد.

**نکته: در ابتدا باید یک سری زمانی ناایستا را به یک سری زمانی ایستا تبدیل کرد تا امکان پیش بینی و مدل سازی به وجود بیاید زیرا امکان پیش بینی برای سری های زمانی ناایستا به راحتی امکان پذیر نیست.

یک سری زمانی ممکن است در میانگین یا در واریانس یا در هر دو مورد ناایستا باشد که به این موضوع می پردازیم:

ناایستایی درمیانگین

اگر یک سری زمانی در میانگین ناایستا باشد ، مهمترین روش تبدیل این سری زمانی به یک سری ایستا تفاضلی کردن می باشد.

را ∇ عملگر تفاضلی می نامند. تفاضلی کردن مرتبه d را با $\nabla^d x_t$ نشان می دهیم .

ناایستایی درواریانس

اگر به مرور زمان تغییر پذیری یک سری زمانی افزایش یابد بدین معنی است که سری زمانی نسبت به واریانس اش ناایستا است .

به طور کلی برای تبدیل واریانس از تبدیلات باکس_کاکس استفاده می کنیم

$$T(x_t) = (x_t^{y} _ 1) / y$$

که y را پارامتر تبدیل می نامند .

تبدیلات مربوط به چند مقدار y که معمولا مورد استفاده قرار می گیرد ، به صورت زیر می باشد و اگر مقدار y برابر یک شود ، نیاز به تبدیل نیست.

** نکته : تبدیلات واریانس فقط برای سری های مثبت به کار می رود.

تبدیل مناسب	مقدارy
$1/xt$	_1
$1/\sqrt{xt}$	_0.5
$\ln xt$	0
$\sqrt{xt}$	0.5
نیاز به تبدیل نیست	1

*همبستگی بین مشاهدات سری زمانی :

معمولا در سری های زمانی مشاهدات متوالی مستقل نیستند و برای بررسی این وابستگی از تابع خود همبستگی Autocorrelation function و تابع خود همبستگی جزیی Partial autocorrelation function استفاده می کنیم.

** نکته : از طریق نمودار میتوان گفت زمانی که مقادیر تابع خود همبستگی به کندی به صفر میل کند ناایستایی سری زمانی مربوطه را نشان می دهد.

برای محاسبه ی تعداد تاخیرهای که تابع خود همبستگی برای آن ها محاسبه می شود به صورت زیر خواهد بود :

تعداد مشاهدات (n)	تعداد تاخیرهای که تابع خود همبستگی برای آنها محاسبه خواهد شد (K)
کمتر مساوی ۲۴۰	$n/4$
بیشتر از ۲۴۰	$\sqrt{n} + 45$

*مدل سازی سری زمانی

۱-مدل های غیر فصلی ایستا : مدل های که در ادامه معرفی می کنیم برای مدل سازی سری های غیر فصلی ایستا استفاده می شوند .

این مدل ها عبارت اند از مدل اتو رگرسیو ، مدل میانگین متحرک ، مدل مرکب اتو رگرسیو و میانگین متحرک

الف) مدل اتو رگرسیو مرتبه P یا AR(P) : این مدل به صورت زیر است

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

** نکته : یکی از ویژگی های فرآیند AR(P) این است که pacf (خود همبستگی جزیی) آن بعد از فاصله p قطع می شود . یعنی تایع خود همبستگی جزیی آن بلافاصله پس از تاخیر p صفر می شود .

ب) مدل میانگین متحرک مرتبه q یا MA(q) : این مدل به صورت زیر است

$$X_t = z_t + \beta_1 z_{t-1} + \ldots\ldots + \beta_q z_{t-q}$$

**نکته : تابع خود همبستگی فرآیند MA(q) بعد از تاخیر q قطع می شود . یعنی تابع خود همبستگی acf آن برای مقادیر بزرگتر از q صفر خواهد بود.

پ) مدل مرکب یا ARMA(p,q) : فرآیندهای اتو رگرسیو_میانگین متحرک که آن ها را فرآیندهای مرکب می نامند از ترکیب این دو فرآیند بدست می آیند و به صورت زیر نوشته می شود (h_0 جمله ثابت)

$$X_t = h_0 + \alpha_1 x_{t-1} + \ldots\ldots\ldots + \alpha_p x_{t-p} + z_t + \beta_1 z_{t-1} + \ldots\ldots\ldots + \beta_q z_{t-q}$$

۲-مدل غیر فصلی ناایستا : بسیاری از سری های زمانی که مشاهده می شوند دارای رفتاری هستند که ناایستایی را نشان می دهند در نتیجه از مدلی که عمل تفاضل گیری نیز انجام می دهد استفاده می کنیم.

مدل ARIMA(p,d,q) یک مدل کلی است که فرآیند تلفیقی اتو رگرسیو-میانگین متحرک با درجه (p,d,q) به همراه تفاضل گیری است . که معادله آن به صورت زیر

$$W_t = \alpha_1 w_{t-1} + ... + \alpha_p w_{t-p} + z_t + \beta_t z_{t-1} + ... + \beta_q z_{t-q}$$

_ مدل IMA(d,q) یک فرآیند اریما که جملات اتو رگرسیو نداشته باشد.

_ مدل ARI(p,d) یک فرآیند اریما که جملات میانگین متحرک نداشته باشد.

** نتیجه : مدل سازی یک سری زمانی شامل سه مرحله است:

۱-تشخیص مدل آزمایشی ۲-برازش مدل یا تخمین پارامترهای مدل ۳-بررسی مناسبت مدل

که در مراحل اول تشخیص مدل آزمایشی باید ایستایی در میانگین و ایستایی در واریانس را انجام دهیم. در مراحل بعد با استفاده از رفتار acf , pacf مدل را تشخیص می دهیم.

در قسمت برازش مدل بعد از شناسایی مدل آزمایشی با استفاده از ARIMA مدلمان را به نرم افزار مینی تب معرفی کرده و نرم افزار پارامترها را برآورد می کند که در بخش Session مینی تب Final estimates of parameters لیست می شوند.

اما برای مناسبت مدل باید ۱-فرض نرمال بودن باقی مانده ها ۲-فرض استقلال باقی مانده ها ۳-فرض ثابت بودن واریانس باقی مانده ها بررسی شود که با استفاده از نمودار در بخش graphs پنجره ARIMA مینی تب و آزمون پرت-مانتو در خود خروجی مینی تب استفاده کرد.

درنهایت به این نتیجه میرسیم که آیا مدل مناسبی برای پیش بینی هست یا خیر.

*پیش بینی سری های زمانی

دو نوع روش برای پیش بینی وجوددارد ۱-روش کمی ۲-روش کیفی

در روش کیفی پیش بینی مبتنی بر دانش تجربی و درک مستقیم و سایر اطلاعات مرتبط با موضوع می باشد.

اما در روش های کمی بدین گونه است که به دو رده روش های یک متغیره (داده های که مربوط به گذشته ، مورد تجزیه و تحلیل قرار می گیرد) و روش های چند متغیره (با فرض ادامه داشتن الگوی تاریخی در آینده و با استفاده از رابطه میان متغیر مورد نظر و سایر متغیرها ، ارزش آتی متغیر مورد نظر پیش بینی می شود مانند الگوهای اقتصادسنجی و رگرسیون) تقسیم می شوند.

که مهم ترین روش های پیش بینی یک متغیره بدین صورت است:

۱-برون یابی منحنی روند۲- هموار کردن نمایی۳- روش هلث-وینترز۴-روش باکس-جنکینز

۱-برون یابی منحنی روند : یک سری دارای روند در واقع یک سری ناایستا است زیرا میانگین این سری ثابت نیست به دلیل تغییرات دراز مدت که برای تجزیه وتحلیل یک سری زمانی دارای روند (فاقد الگوی فصلی) در مینی تب به این گونه عمل می کنیم

Stat>>Time series>>Trend analysis

اگر برای تجزیه و تحلیل یک سری زمانی دارای روند رفتار فصلی نیز مشاهده شد میتوان در مینی تب به این طریق عمل کرد

Stat>>Time series>>Decomposition

که در این بخش باید طول عامل فصلی و با توجه به ماهیت داده هایمان یکی از دو عامل جمعی و ضربی را انتخاب کرد. که در مدل ضربی همانطور که قبلا پرداختیم عامل روند و عامل فصلی مستقل از هم نیستند به عبارتی با افزایش سطح داده ها اندازه مولفه فصلی نیز افزایش می یابد و در مدل جمعی عامل روند و عامل فصلی مستقل از یک دیگر هستند

همچنین برای پیش بینی دوره های آینده میتوان از این منو استفاده کرد و تعداد پیش بینی ها و نقطه پیش بینی ها را باید مشخص کرد.

۲-هموارکردن نمایی: به منظور کاهش نوسان موجود در داده ها به کار می رود و به چند شیوه میانگین متحرک ساده و هموار سازی نمایی می پردازیم.

الف) میانگین متحرک ساده که در این روش داده ها باید فاقد روند و یا الگوی فصلی مشخصی باشند و روشی است که با متوسط گیری از مشاهدات متوالی در هر لحظه ، داده ها را هموار می کند. و در مینی تب از طریق

Stat>>Time series>>Moving average

که در این بخش باید دامنه یا طول میانگین متحرک و مقادیر میانگین متحرک در مرکز را مشخص کرد و همچنین برای پیش بینی های کوتاه مدت نیز میتوان استفاده کرد .

ب) هموارسازی نمایی که این روش نیز برای سری های زمانی غیر فصلی که روندی را نشان نمی دهند به کار می رود و روشی است برای تخمین ضرایب در یک مدل سری زمانی چند جمله ای. که در مینی تب بدین گونه عمل می کنیم

Stat>>Time series>>Exp smoothing یا Double Exp smoothing

۳- روش پیش بینی هلت-وینترز : در این روش می توان از سری های که شامل روند و تغییرات فصلی می باشند استفاده کرد و برای پیش بینی های کوتاه مدت و میان مدت استفاده کرد. که در مینی تب داریم

Stat>>Time series>>Winter's method

که در این بخش باید عامل فصلی ، و وزن های برای مولفه روند ، مولفه سطح و مولفه فصلی مشخص کرد که پیش فرض خود مشخص کرده.

۴- روش پیش بینی باکس-جنکینز : در این روش یک مدل اریما به داده ها برازش می دهیم پس از مشخص کردن مدل نهایی یعنی مرتبه تفاضلی کردن و مرتبه AR و MA که در نرم افزار مینی تب داریم

Stat>>Time series>>ARIMA

در این بخش اگر مدل مشخص شده فصلی باشد دوره و مرتبه اتو رگرسیو فصلی و میانگین متحرک فصلی و درجه تفاضلی کردن را مشخص می کنیم در غیر این صورت مدل غیر فصلی را انتخاب می کنیم.

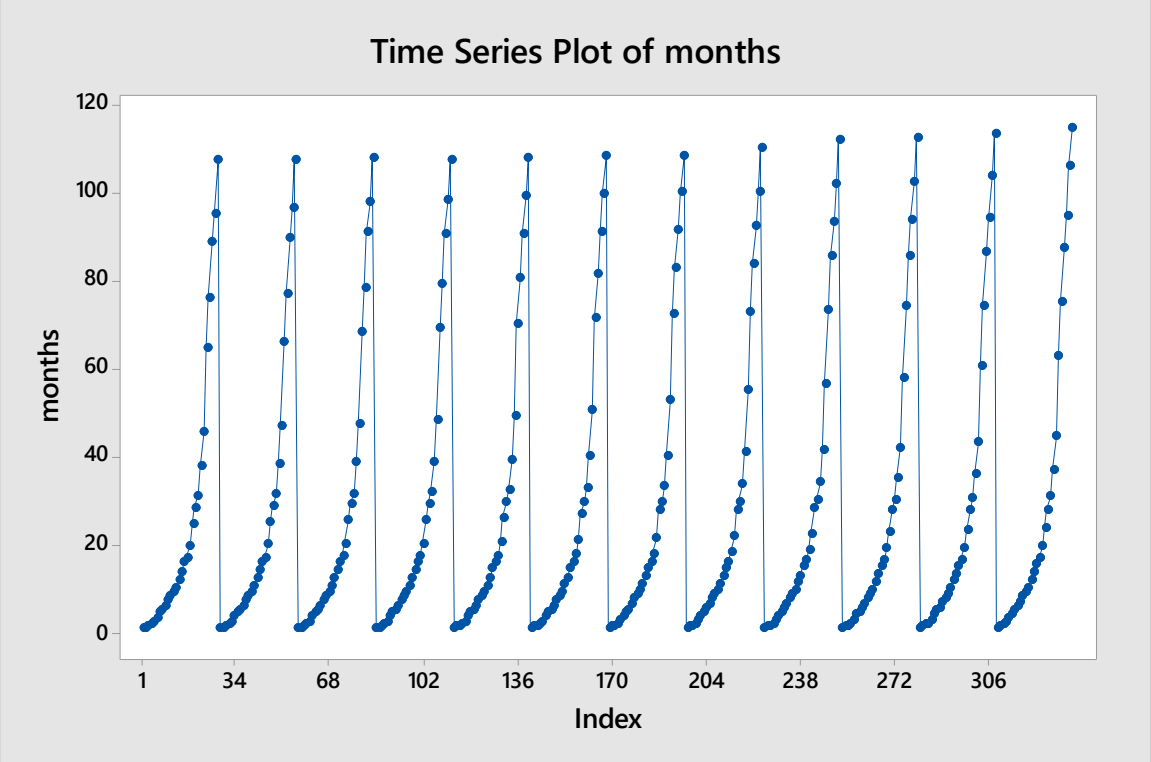
** نکته : مینی تب به طور پیش فرض جمله ثابت را در مدل لحاظ می کند اما اگر پیش بینی ما در Session مینی تب بدین گونه بود که در قسمت Final estimates of parameters در ردیف Constant مقدار پی ول یو بیشتر از ۰,۰۵ بود فرضیه صفر بودن جمله ثابت را می پذیریم در غیر این صورت وجود جمله ثابت در مدل ضروری ست و در صورت پذیرفتن فرض باید در ادامه حالت پیش فرض Include constant term in model برداشته شود تا مدلی بدون جمله ثابت داشته باشیم.

مثال : برای داده های شاخص کل بهای کالاها و خدمات مصرفی در مناطق شهری ایران که از سایت بانک مرکزی گرفته شده است میخواهیم تحلیل سری زمانی با نرم افزار مینی تب را انجام دهیم که داده ها ماهانه از سال ۱۳۶۹ تا سال ۱۳۹۶ به همراه متوسط سال داده شده است .

اولین گام در تجزیه و تحلیل یک سری زمانی رسم نمودار آن است در نتیجه داریم :

Time Series Plot of months

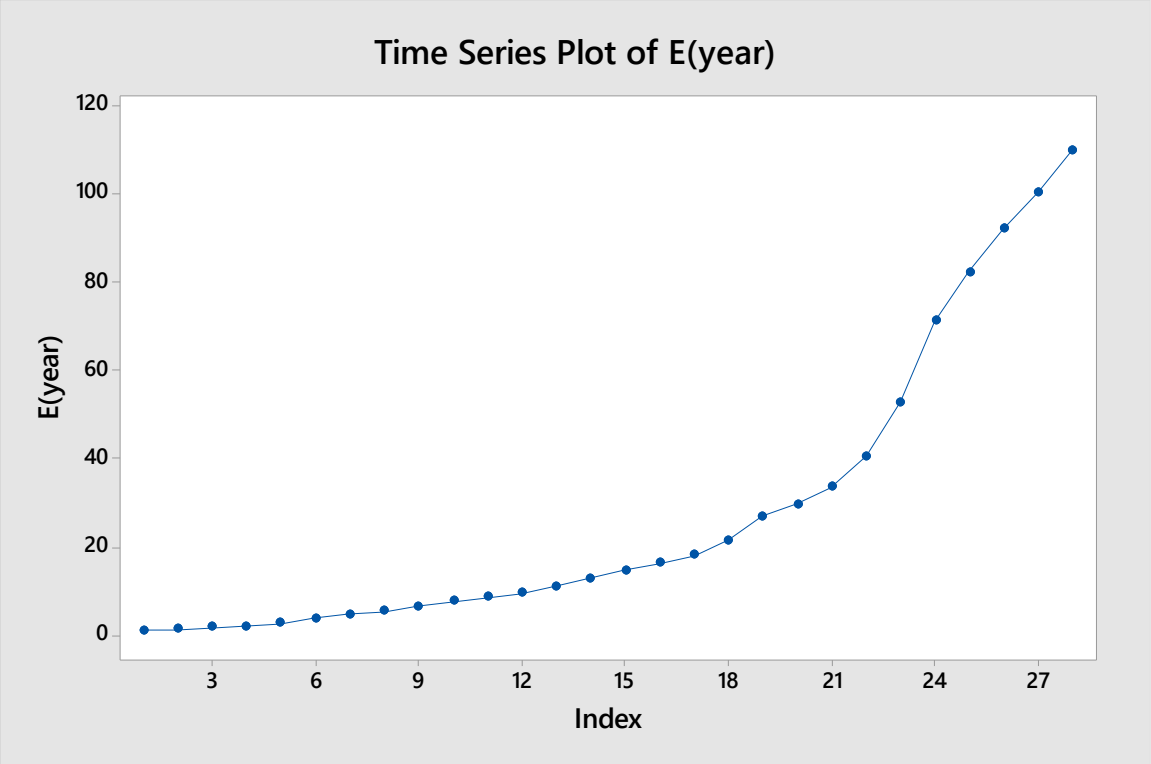
Time Series Plot of months



همانطور که مشاهده می شود این سری زمانی در میانگین ناایستا است میتوان به این دلیل گفت هم شاهد عامل روند و عامل فصلی در این داده ها هستیم از طرفی با استفاده از نمودار متوسط سال ملاحظه می کنیم که

Time Series Plot of E(year)

Time Series Plot of E(year)

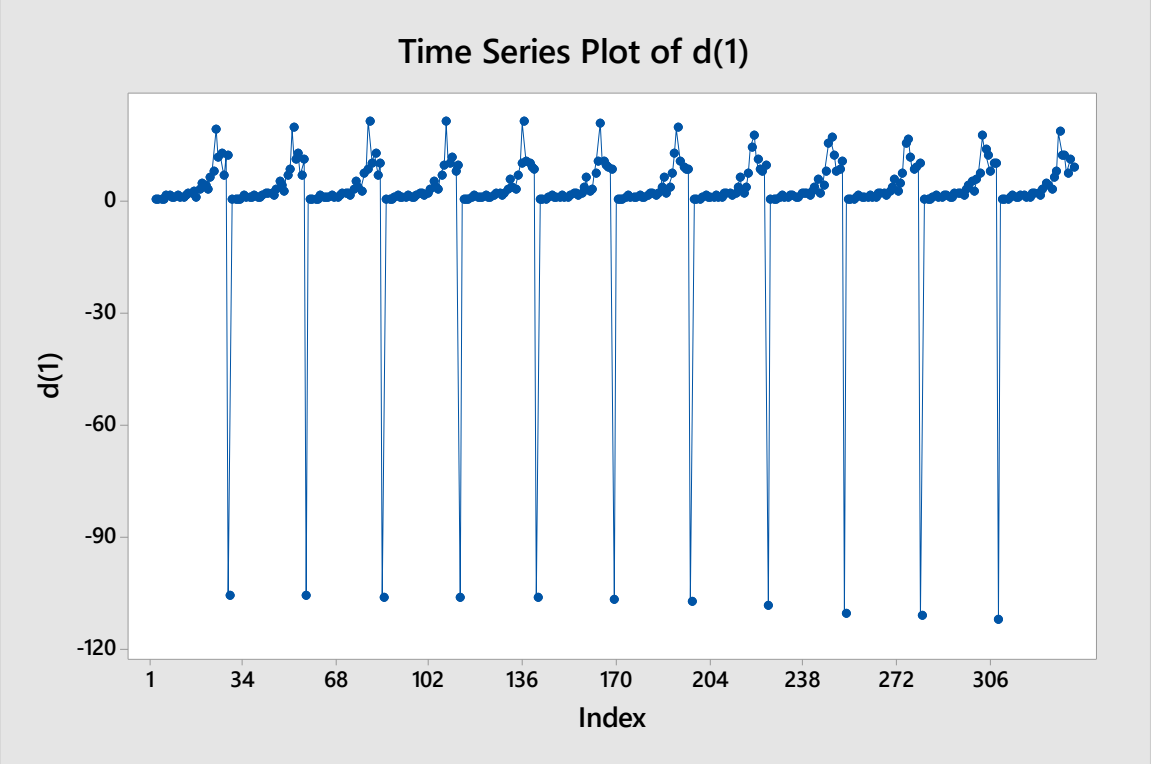


میبینیم که این سری میانگین آن ثابت نیست و همراه با زمان در حال افزایش یا کاهش است.

در نتیجه به سراغ ایستا کردن میانگین از راه تفاضلی کردن می رویم که از طریق Stat, Time series گزینه Differences را انتخاب کرده و در قسمت Lag مرتبه تفاضلی را تعیین می کنیم که در ستونی نتایج را ریخته و در مرحله بعد نمودار سری آن را رسم کرده و داریم

Time Series Plot of d(1)

Time Series Plot of d(1)



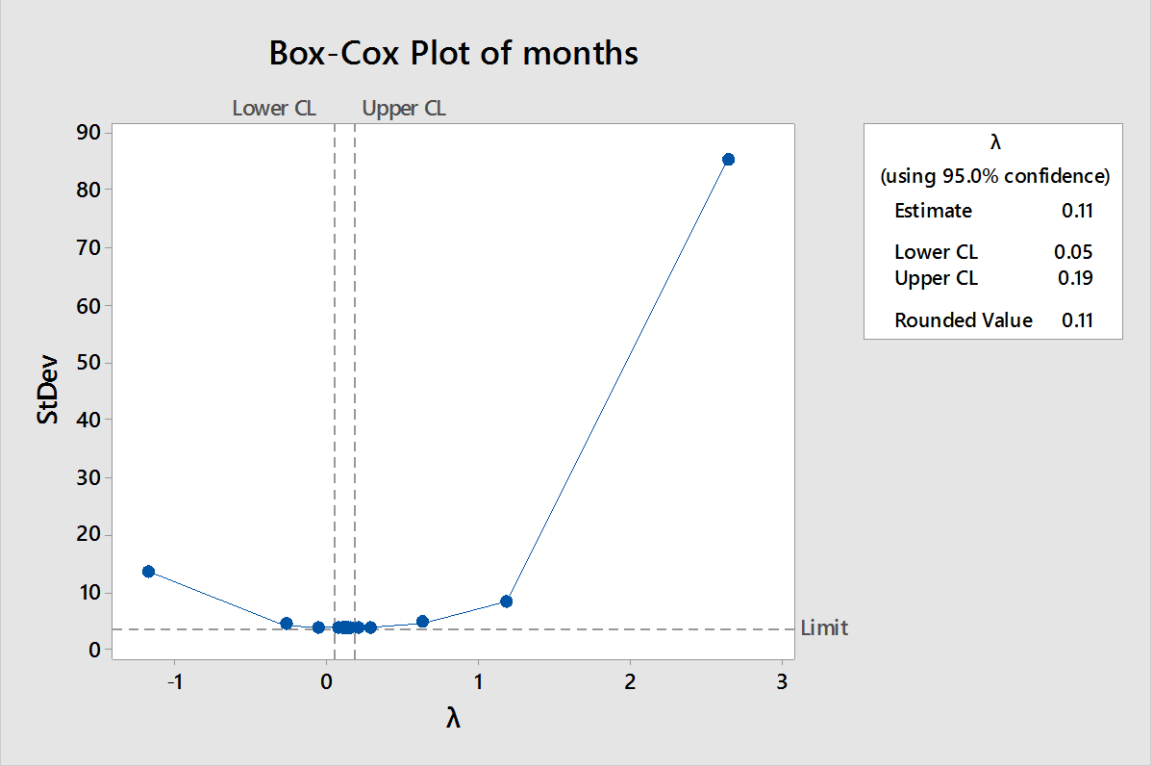
می توان گفت سری فوق یک سری ایستا است زیرا بیشتر مقادیر حول یک میانگین متمرکز شده اند و نیازی به تفاضلی کردن بیشتر نیست.

حال به سراغ بررسی ایستایی در واریانس میرویم ، با توجه به نمودار اول به نظر می رسد که در واریانس ناایستا است یعنی تغییرپذیری سری زمانی با گذشت زمان تقریبا تغییر پیدا کرده اما برای بررسی بیشتر از تبدیل باکس-کاکس استفاده می کنیم از طریق منوی Stat گزینه Control charts و Box-Cox را انتخاب کرده

در قسمت Subgroup sizes عدد یک را وارد می کنیم (زیرا در مبحث کنترل کیفیت حجم زیر گروه ها معمولا ۴ یا ۵ می باشد اما در سری زمانی برای استفاده از باکس-کاکس حجم زیر گروه را ۱ در نظر می گیریم.) در نتیجه داریم

Box-Cox Plot of months

Box-Cox Plot of months



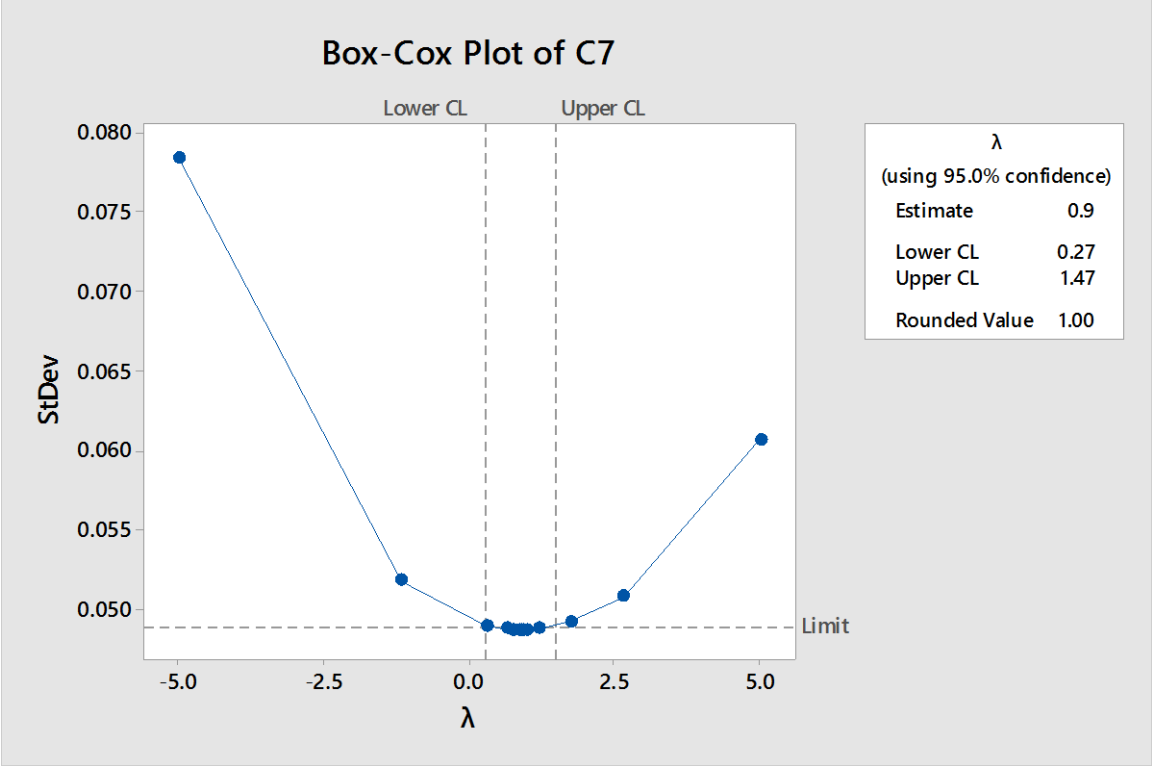
با توجه به کادر سمت راست نمودار برآورد لاندا (در قسمت های قبل این پارامتر را y نامیدیم) و حدود اطمینان ۰,۹۵ و مقدار پیشنهادی برای لاندا داده شده با توجه به اینکه مقدار پیشنهادی برابر ۰,۱۱ است و از طرفی در حدود اطمینان داده شده در نمودار نیز عدد یک وجود ندارد در نتیجه نیاز به تبدیل با توجه به جدول گفته شده داریم.

برای تبدیل داده ها از منوی Calc,Calculator را انتخاب می کنیم و در قسمت Function تابع مورد نظرمان را انتخاب می کنیم و بعد در قسمت Expression ستون داده های مورد نظرمان را وارد می کنیم و نتایج را در یکی از ستون های مینی تب وارد می کنیم

برای داده های مثال خود نتایج را در ستون C7 می ریزیم و در قسمت Function تابع جذر دوم را انتخاب کرده بر روی داده ها که در ستون months وارد کردیم و OK می کنیم و در ادامه برای رسم نمودار باکس-کاکس نتایج ستون C7 از طریق منوی Stat گزینه Control charts و Box-Cox را انتخاب کرده که داریم

Box-Cox Plot of C7

Box-Cox Plot of C7

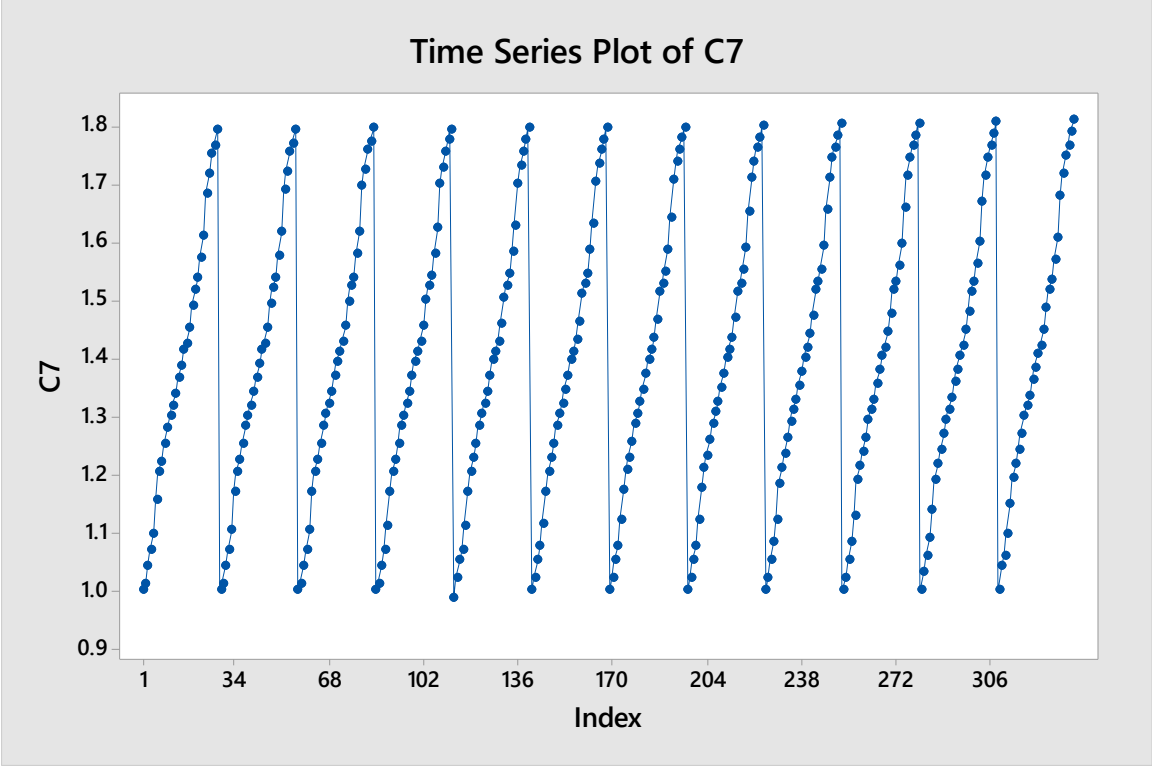


با توجه به کادر سمت راست نمودار برآورد لاندا (در قسمت های قبل این پارامتر را y نامیدیم) و حدود اطمینان ۰,۹۵ و مقدار پیشنهادی برای لاندا داده شده با توجه به اینکه مقدار پیشنهادی برابر ۱ است و از طرفی در حدود اطمینان داده شده در نمودار نیز عدد یک نیز وجود دارد در نتیجه دیگر نیاز به تبدیل برای پایایی واریانس نداریم.

از طرفی می توانیم نمودار سری این مقادیر را از طریق Time series رسم کرده که در نمودار مشاهده می کنیم مقادیر تقریبا حول یک مقادیر ثابت نوسان دارد و تغییرپذیری سری زمانی با گذشت زمان تقریبا تغییر پیدا نکرده.

Time Series Plot of C7

Time Series Plot of C7



حال به سراغ بررسی خود همبستگی سری زمانی مثال مان می رویم از منوی Stat گزینه Time series , Autocorrelation را انتخاب کرده و ستون داده هایمان را وارد کرده و داریم

Autocorrelation Function: months

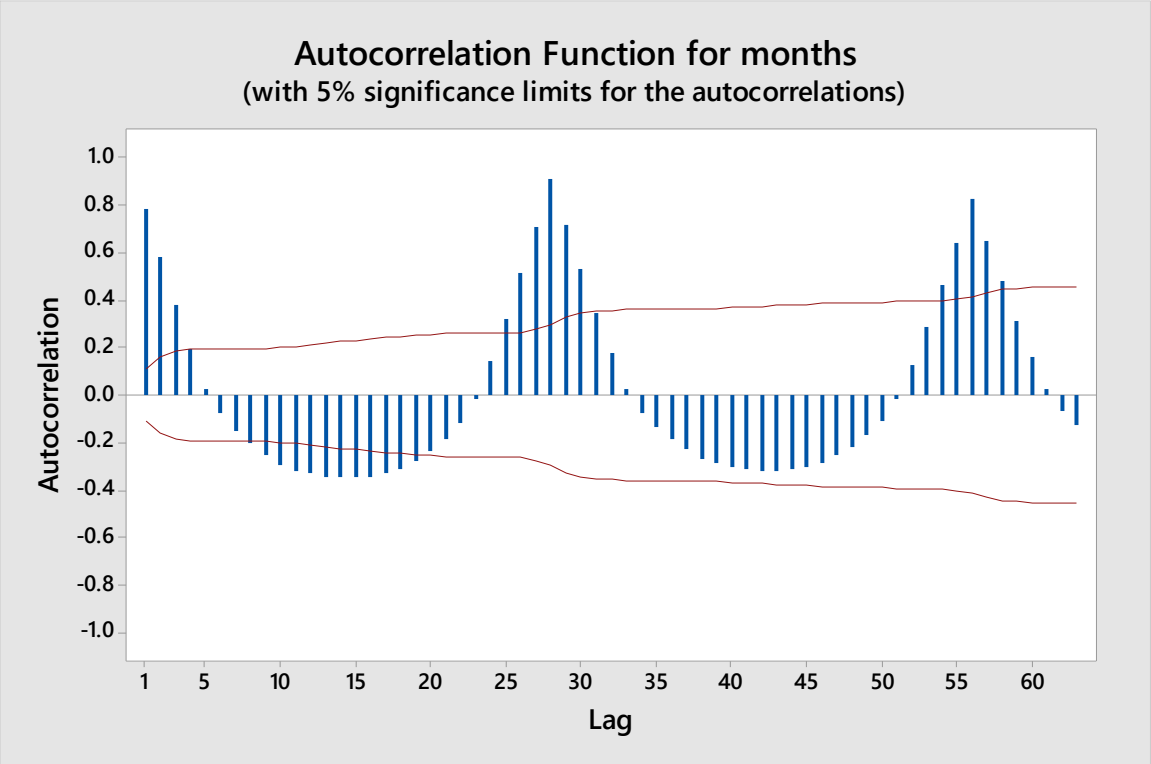
Autocorrelations

Lag	ACF	T	LBQ
1	0.787508	14.44	210.24
2	0.581278	7.12	325.13
3	0.379759	4.08	374.32
4	0.193203	1.98	387.08
5	0.028982	0.29	387.37
6	-0.079768	-0.81	389.56
7	-0.152691	-1.54	397.61
8	-0.206439	-2.07	412.37
9	-0.252464	-2.50	434.50
10	-0.292389	-2.84	464.29
11	-0.317280	-3.01	499.46

12	-0.333693	-3.09	538.49
13	-0.344399	-3.10	580.20
14	-0.350424	-3.07	623.51
15	-0.349332	-2.98	666.68
16	-0.343624	-2.85	708.59
17	-0.331966	-2.69	747.82
18	-0.311944	-2.48	782.57
19	-0.277321	-2.16	810.13
20	-0.236696	-1.82	830.26
21	-0.187917	-1.43	842.99
22	-0.120416	-0.91	848.24
23	-0.016787	-0.13	848.34
24	0.139771	1.06	855.45
25	0.318510	2.40	892.50
26	0.511477	3.79	988.34
27	0.710480	5.05	1173.86
28	0.915571	6.06	1482.95
29	0.720313	4.32	1674.89
30	0.531284	3.02	1779.65
31	0.346061	1.92	1824.24
32	0.175466	0.96	1835.74
33	0.025123	0.14	1835.98
34	-0.072315	-0.40	1837.95
35	-0.138686	-0.76	1845.20
36	-0.187459	-1.02	1858.51
37	-0.229908	-1.25	1878.58
38	-0.266755	-1.44	1905.70
39	-0.289395	-1.56	1937.73
40	-0.304524	-1.63	1973.31
41	-0.314606	-1.67	2011.41
42	-0.320348	-1.68	2051.05
43	-0.319578	-1.67	2090.64
44	-0.314740	-1.63	2129.17
45	-0.304270	-1.56	2165.30
46	-0.286093	-1.46	2197.35
47	-0.254492	-1.29	2222.80

48	-0.217730	-1.10	2241.49
49	-0.173254	-0.87	2253.37
50	-0.111783	-0.56	2258.33
51	-0.016236	-0.08	2258.44
52	0.126924	0.63	2264.88
53	0.290157	1.45	2298.67
54	0.465241	2.31	2385.84
55	0.646180	3.16	2554.59
56	0.831715	3.95	2835.16
57	0.653902	2.97	3009.22
58	0.481941	2.13	3104.10
59	0.313267	1.37	3144.34
60	0.158679	0.69	3154.70
61	0.022277	0.10	3154.90
62	-0.064428	-0.28	3156.62
63	-0.124480	-0.54	3163.07

Autocorrelation for months



همانطور که در قسمت های قبل توضیح دادیم از آن جا که پیش فرض Default number of lags در پنجره Autocorrelation را قبول کردیم و طبق فرمول جدول گفته شده ، تعداد مشاهدات (n) برابر ۳۳۶ است در نتیجه با توجه به فرمول تعداد تاخیرهای ما در قسمت نتایج برابر ۶۳ شده که برای هر کدام تابع خود همبستگی و آماره LBQ (که آن را آماره Q اصلاح شده نیز می نامند و برای آزمون این فرض که آیا همه خود همبستگی ها تا تاخیر K مساوی با صفر هستند یا خیر استفاده می شود) را مشخص می کنند.

در ادامه با توجه به نمودار acf می بینیم که تابع خودهمبستگی به کندی به صفر میل می کند همچنین دارای نوساناتی است پس خود این دلیلی بر ناایستایی داده هایمان است.

حال به سراغ انجام همین عملیات برای داده های که عمل تفاضلی کردن را بر روی شان انجام داده ایم تا ایستا شوند می رویم ، در قسمت های قبل عمل تفاضلی کردن را بر روی داده ها انجام داده ایم و نتایج آن را که در ستون مینی تب ریخته شده بود به نام d(1) نام گذاری کردیم ، مراحل فوق را برای داده های ستون d(1) انجام می دهیم حال داریم

Autocorrelation Function: d(1)

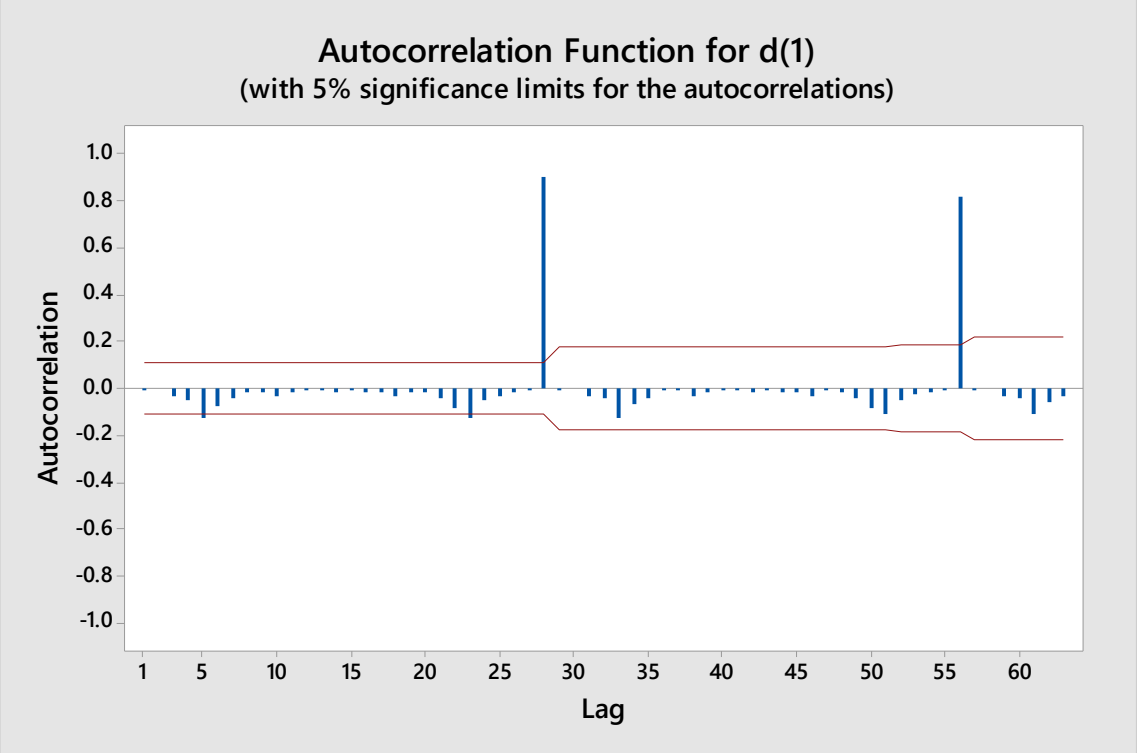
Autocorrelations

Lag	ACF	T	LBQ
1	-0.010425	-0.19	0.04
2	-0.005025	-0.09	0.05
3	-0.032940	-0.60	0.41
4	-0.048247	-0.88	1.21
5	-0.130567	-2.38	7.04
6	-0.077821	-1.40	9.12
7	-0.043064	-0.77	9.76
8	-0.015369	-0.27	9.84
9	-0.013244	-0.24	9.90
10	-0.034743	-0.62	10.32
11	-0.018322	-0.33	10.44
12	-0.012373	-0.22	10.49
13	-0.010600	-0.19	10.53

14	-0.016127	-0.29	10.62
15	-0.010078	-0.18	10.66
16	-0.013540	-0.24	10.72
17	-0.019677	-0.35	10.86
18	-0.034897	-0.62	11.29
19	-0.013393	-0.24	11.36
20	-0.019038	-0.34	11.49
21	-0.044805	-0.79	12.21
22	-0.087960	-1.56	15.00
23	-0.127384	-2.24	20.87
24	-0.052387	-0.91	21.87
25	-0.032564	-0.56	22.25
26	-0.013464	-0.23	22.32
27	-0.012380	-0.21	22.37
28	0.907727	15.65	325.38
29	-0.009811	-0.11	325.41
30	-0.003711	-0.04	325.42
31	-0.031880	-0.35	325.80
32	-0.043182	-0.47	326.49
33	-0.123231	-1.35	332.17
34	-0.066653	-0.73	333.83
35	-0.039262	-0.43	334.41
36	-0.012353	-0.13	334.47
37	-0.012249	-0.13	334.53
38	-0.032358	-0.35	334.93
39	-0.016241	-0.18	335.03
40	-0.010713	-0.12	335.07
41	-0.009731	-0.11	335.11
42	-0.014680	-0.16	335.19
43	-0.008797	-0.10	335.22
44	-0.012731	-0.14	335.28
45	-0.018059	-0.20	335.41
46	-0.031934	-0.35	335.81
47	-0.011378	-0.12	335.86
48	-0.018024	-0.20	335.99
49	-0.040487	-0.44	336.63

50	-0.082711	-0.90	339.34
51	-0.114064	-1.24	344.52
52	-0.047291	-0.51	345.41
53	-0.026723	-0.29	345.69
54	-0.012900	-0.14	345.76
55	-0.008993	-0.10	345.79
56	0.816143	8.80	615.32
57	-0.008766	-0.08	615.35
58	-0.003056	-0.03	615.36
59	-0.030245	-0.27	615.73
60	-0.038303	-0.34	616.33
61	-0.113598	-1.01	621.65
62	-0.056842	-0.51	622.98
63	-0.035513	-0.32	623.51

Autocorrelation for d(1)



تعدادتاخیرها برابر ۶۳ شده که برای هر کدام تابع خود همبستگی و آماره LBQ را محاسبه کرده و در نمودار acf فوق مشاهده می کنیم که آن نوسانات دیگر نیست و همچنین تابع خودهمبستگی به صفر میل می کند و این خود می تواند دلیلی بر ایستاشدن داده ها باشد.

برای رسم تابع خودهمبستگی جزیی در مینی تب برای سری زمانی خود از منوی Stat,Time series,Partial autocorrelation را انتخاب کرده و اول برای خود داده های اصلی که در ستون months ریختیم انجام می دهیم داریم

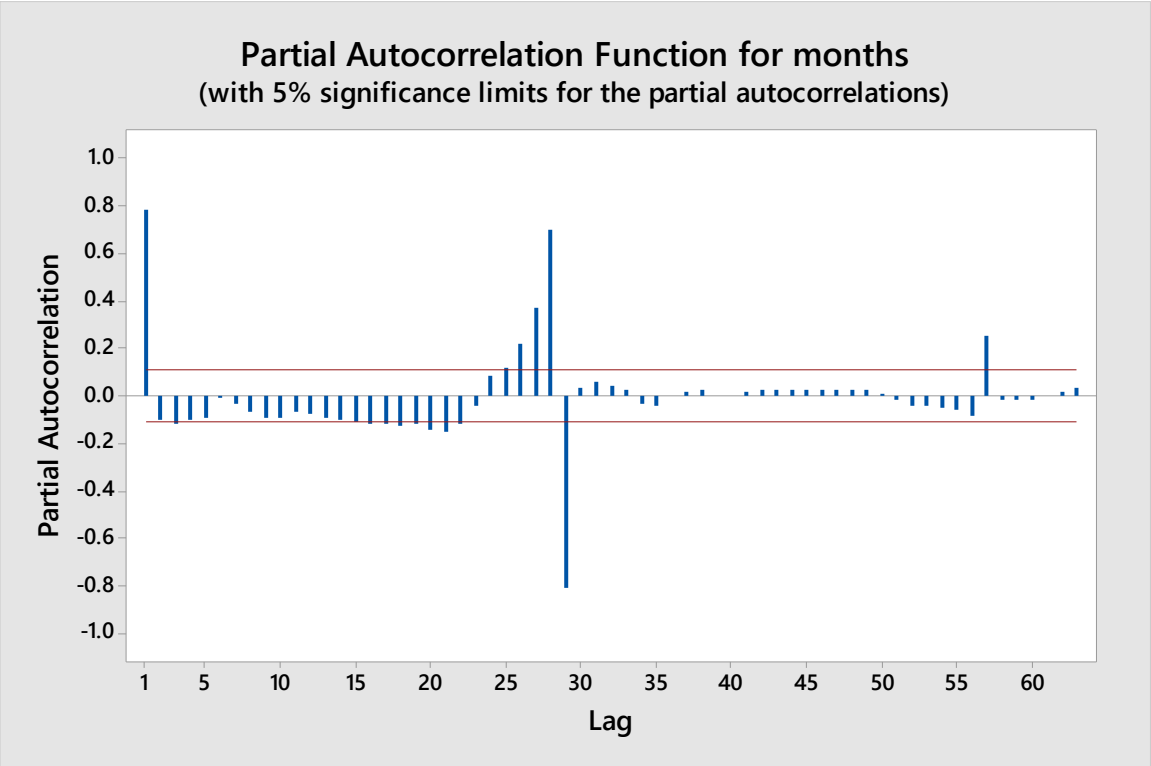
Partial Autocorrelation Function: months

Partial Autocorrelations

Lag	PACF	T
1	0.787508	14.44
2	-0.102391	-1.88
3	-0.117703	-2.16
4	-0.105471	-1.93
5	-0.095327	-1.75
6	-0.009175	-0.17
7	-0.036111	-0.66
8	-0.067108	-1.23
9	-0.091070	-1.67
10	-0.095198	-1.75
11	-0.066145	-1.21
12	-0.078701	-1.44
13	-0.092212	-1.69
14	-0.104796	-1.92
15	-0.106792	-1.96
16	-0.115553	-2.12
17	-0.121070	-2.22
18	-0.125680	-2.30
19	-0.117622	-2.16
20	-0.143112	-2.62
21	-0.148257	-2.72
22	-0.122039	-2.24
23	-0.046448	-0.85
24	0.083595	1.53
25	0.122553	2.25
26	0.221225	4.06
27	0.371223	6.80
28	0.703929	12.90

29	-0.807938	-14.81
30	0.035476	0.65
31	0.055827	1.02
32	0.041861	0.77
33	0.021301	0.39
34	-0.035785	-0.66
35	-0.043015	-0.79
36	0.001122	0.02
37	0.018570	0.34
38	0.023163	0.42
39	0.004403	0.08
40	0.000587	0.01
41	0.017111	0.31
42	0.024485	0.45
43	0.023925	0.44
44	0.025808	0.47
45	0.024753	0.45
46	0.027048	0.50
47	0.025754	0.47
48	0.026385	0.48
49	0.024571	0.45
50	0.011253	0.21
51	-0.016174	-0.30
52	-0.044477	-0.82
53	-0.039403	-0.72
54	-0.049526	-0.91
55	-0.062273	-1.14
56	-0.085813	-1.57
57	0.253009	4.64
58	-0.014351	-0.26
59	-0.016889	-0.31
60	-0.016611	-0.30
61	-0.005298	-0.10
62	0.020082	0.37
63	0.035314	0.65

Partial Autocorrelation for months



با توجه به نمودار تابع خودهمبستگی جزیی داده ها مشاهده می کنیم که وضعیت بهتری نسبت به نمودار تابع خودهمبستگی دارد اما در اکثر جاها به کندی به صفر میل می کند و خط افقی را قطع کرده اند پس به سراغ رسم نمودار pacf داده های ستون d(1) می رویم

Partial Autocorrelation Function: d(1)

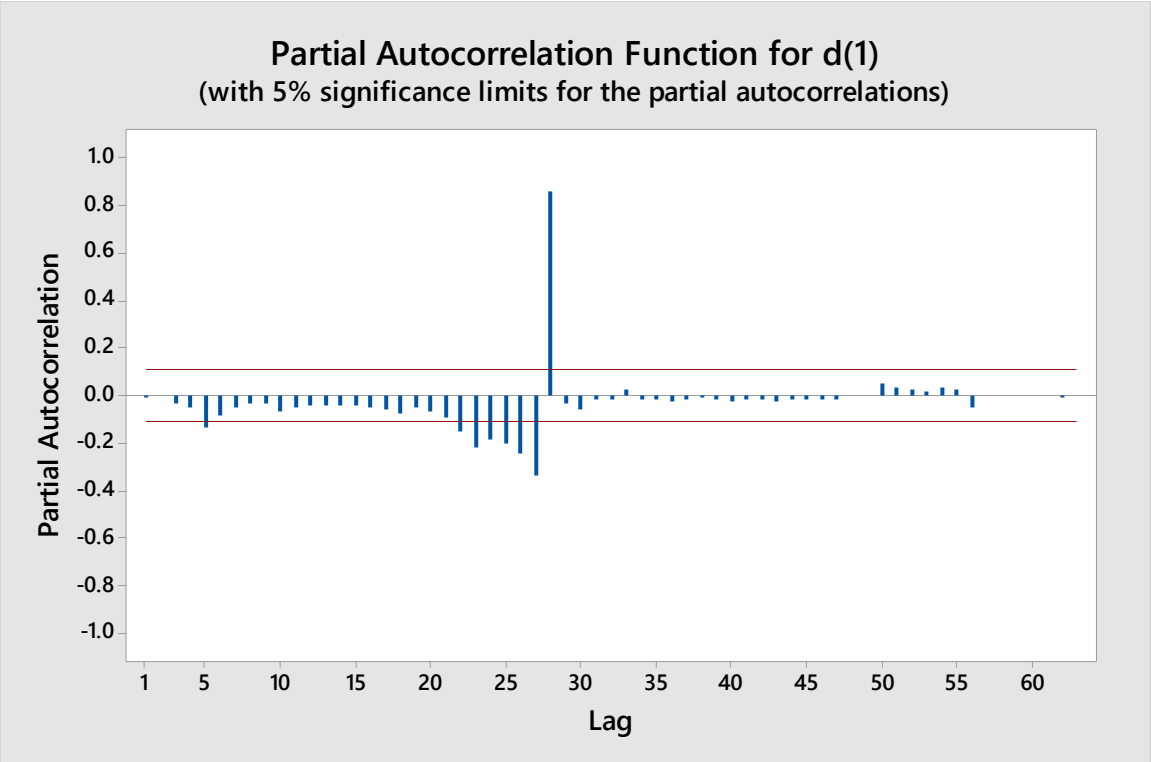
Partial Autocorrelations

Lag	PACF	T
1	-0.010425	-0.19
2	-0.005134	-0.09
3	-0.033051	-0.60
4	-0.049030	-0.90
5	-0.132509	-2.43
6	-0.085333	-1.56
7	-0.053881	-0.99
8	-0.032827	-0.60
9	-0.036803	-0.67
10	-0.069476	-1.27
11	-0.054516	-1.00
12	-0.044724	-0.82

13	-0.039481	-0.72
14	-0.045623	-0.84
15	-0.045663	-0.84
16	-0.049799	-0.91
17	-0.055812	-1.02
18	-0.072173	-1.32
19	-0.054673	-1.00
20	-0.063810	-1.17
21	-0.097012	-1.78
22	-0.154521	-2.83
23	-0.223628	-4.09
24	-0.184899	-3.38
25	-0.202252	-3.70
26	-0.241983	-4.43
27	-0.336250	-6.15
28	0.861423	15.77
29	-0.029681	-0.54
30	-0.056293	-1.03
31	-0.020324	-0.37
32	-0.018392	-0.34
33	0.024726	0.45
34	-0.012850	-0.24
35	-0.016205	-0.30
36	-0.026295	-0.48
37	-0.020753	-0.38
38	-0.011761	-0.22
39	-0.020460	-0.37
40	-0.023399	-0.43
41	-0.020417	-0.37
42	-0.020263	-0.37
43	-0.023619	-0.43
44	-0.020269	-0.37
45	-0.014717	-0.27
46	-0.013621	-0.25
47	-0.017661	-0.32
48	-0.007567	-0.14

49	0.003064	0.06
50	0.048847	0.89
51	0.033708	0.62
52	0.024712	0.45
53	0.013247	0.24
54	0.030783	0.56
55	0.024426	0.45
56	-0.050339	-0.92
57	-0.004114	-0.08
58	-0.006056	-0.11
59	-0.003413	-0.06
60	-0.004018	-0.07
61	0.000971	0.02
62	-0.008753	-0.16
63	-0.008393	-0.15

Partial Autocorrelation for d(1)



با توجه به نمودار تابع خودهمبستگی جزیی به صفر میل می کند و خط افقی را نیز قطع نکرده اند و این خود می توان دلیلی بر ایستایی داده های تفاضلی شده باشد.

حال به سراغ تجزیه وتحلیل سری زمانی که دارای روند و مولفه فصلی می رویم در بخش Stat, Time series,Decomposition می رویم از آنجا که داده های ما فصلی بر حسب ماهانه است طول ۱۲ را وارد کرده و مدل ضربی به همراه گزینه روند با فصلی را انتخاب می کنیم براساس مطالبی که در قسمت های قبل گفته شد و داریم

Time Series Decomposition for months

Method

Model type	Multiplicative Model	
Data	months	
Length	336	طول داده ها
NMissing	0	هیچ داده گمشده ای نداریم

Fitted Trend Equation

$Yt = 21.87 + 0.0354 \times t$

معادله روند فیت شده که در نمودار زیر مشخص است

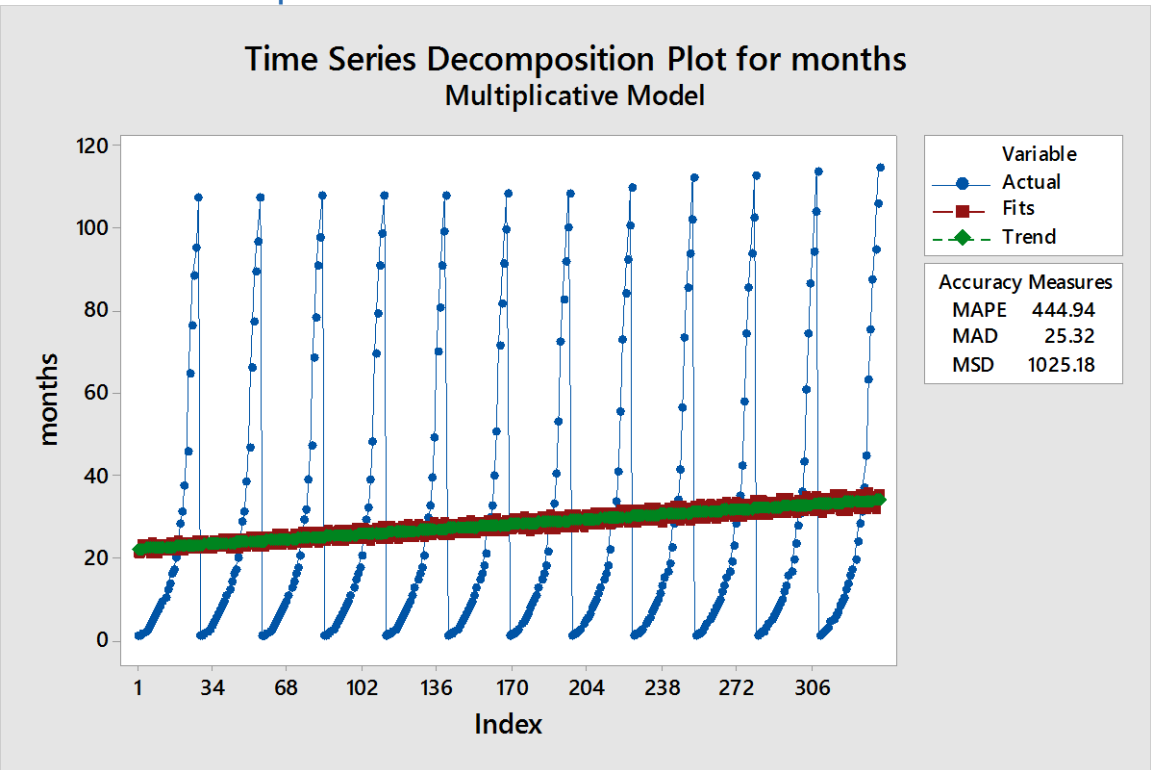
Seasonal Indices

Period	Index
1	0.95253
2	0.94787
3	1.06360
4	1.05481
5	0.96011
6	0.94148
7	1.06301
8	1.03913
9	0.95068
10	0.93395
11	1.04559
12	1.04726

Accuracy Measures معیارهای دقت

MAPE	444.94
MAD	25.32
MSD	1025.18

Time Series Decomposition Plot for months



در خروجی فوق ، قسمت Seasonal indices شاخص های فصلی برای هر ۱۲ ماه را مشخص می کند ، از طرفی در بخش Accuracy measures می بینیم که سه معیار دقت به ترتیب به نام های میانگین قدر مطلق درصد خطا MAPE ، میانگین قدر مطلق انحرافات MAD ، و میانگین مربع انحرافات MSD به ما داده که از این معیارها به منظور ۱- کمک به فرآیند انتخاب مدل پیش بینی ۲- نظارت بر سیستم پیش بینی به منظور دستیابی به اشتباه موردی.

از طرفی طبق نمودار بالا می توانیم هم مولفه روند و هم مولفه فصلی را مشاهده کنیم به علاوه ی معادله ی روند خط فیت شده .

در ادامه می خواهیم به روش باکس-جنکینز مدل سازی انجام دهیم به همراه بررسی مناسبت مدل که در این روش، شامل برازش یک مدل اریما به داده هایمان است که نیاز به تعیین مرتبه هر یک از فرآیندهای AR,MA داریم.

ما ایستایی بودن و مرتبه تفاضلی کردن داده هایمان را ازآنجا که دارای روند و مولفه فصلی بودن بررسی کردیم حال به سراغ تعیین مرتبه های MA,AR از روی داده های تفاضلی شده می رویم زیرا همانطور که گفته شد این فرآیندها برای داده های ایستا مورد استفاده هستند با توجه به نمودار(1)acf d و (1)pacf d می بینیم که به ترتیب بعد از تاخیر q=4 و p=4 قطع می شود و مرتبه تفاضلی کردن نیز برابر یک بود یعنی I=1 در نتیجه از طریق منوی Stat,Time series,ARIMA را انتخاب کرده و در بخش Seasonal مرتبه ها را وارد می کنیم به همراه طول دوره را که برابر ۱۲ است همچنین در بخش گراف این پنجره تیک Four in one را فعال کرده برای بررسی مناسبت مدل از طریق نمودار، درنتیجه داریم

ARIMA Model: months

Estimates at Each Iteration

Iteration	SSE	Parameters								
0	904230	0.100	0.100	0.100	0.100	0.100	0.100	0.100	0.100	1.308
1	641295	-0.050	0.012	0.025	0.033	0.250	0.189	0.175	0.167	0.808
2	620907	0.017	0.135	0.064	-0.081	0.342	0.299	0.173	0.017	0.693
3	594510	0.059	0.285	-0.018	-0.134	0.405	0.447	0.042	-0.041	0.620
4	563635	0.110	0.434	-0.100	-0.197	0.488	0.597	-0.097	-0.108	0.537
5	521379	0.108	0.556	-0.198	-0.238	0.535	0.747	-0.247	-0.136	0.490
6	451313	0.012	0.631	-0.201	-0.329	0.520	0.897	-0.278	-0.224	0.443
7	305662	-0.132	0.568	-0.250	-0.479	0.635	0.994	-0.365	-0.299	0.439
8	227768	-0.282	0.496	-0.298	-0.515	0.617	1.011	-0.466	-0.204	0.554
9	166053	-0.432	0.378	-0.345	-0.575	0.596	1.041	-0.533	-0.151	0.613
10	117841	-0.582	0.230	-0.398	-0.663	0.588	1.089	-0.574	-0.156	0.653
11	93775	-0.688	0.080	-0.454	-0.757	0.609	1.154	-0.586	-0.224	0.585
12	81549	-0.779	-0.070	-0.534	-0.814	0.529	1.212	-0.489	-0.298	0.627
13	73878	-0.859	-0.170	-0.579	-0.842	0.394	1.314	-0.339	-0.419	0.582
14	66668	-0.874	-0.177	-0.573	-0.835	0.244	1.438	-0.194	-0.545	0.549
15	60348	-0.879	-0.170	-0.565	-0.828	0.094	1.562	-0.059	-0.661	0.545
16	57056	-0.860	-0.153	-0.574	-0.840	-0.012	1.702	0.015	-0.769	0.548
17	54200	-0.859	-0.182	-0.628	-0.880	0.019	1.704	-0.014	-0.767	0.582
18	53630	-0.862	-0.206	-0.666	-0.903	-0.002	1.709	0.002	-0.769	0.587
19	53553	-0.861	-0.214	-0.682	-0.914	-0.002	1.711	0.001	-0.771	0.582

20	53553	-0.861	-0.217	-0.688	-0.918	-0.003	1.713	0.002	-0.772	0.581
21	53553	-0.861	-0.217	-0.688	-0.918	-0.003	1.713	0.002	-0.772	0.581

Relative change in each estimate less than 0.001

Final Estimates of Parameters

Type	Coef	SE Coef	T-Value	P-Value
SAR 12	-0.8609	0.0249	-34.59	0.000
SAR 24	-0.2170	0.0343	-6.32	0.000
SAR 36	-0.6881	0.0386	-17.81	0.000
SAR 48	-0.9175	0.0272	-33.75	0.000
SMA 12	-0.0035	0.0711	-0.05	0.961
SMA 24	1.7127	0.0458	37.41	0.000
SMA 36	0.0024	0.0741	0.03	0.975
SMA 48	-0.7721	0.0625	-12.35	0.000
Constant	0.5807	0.0478	12.16	0.000

Differencing: 0 regular, 1 seasonal of order 12
Number of observations: Original series 336, after differencing 324

Residual Sums of Squaresl مجموع مربعات باقی مانده ها

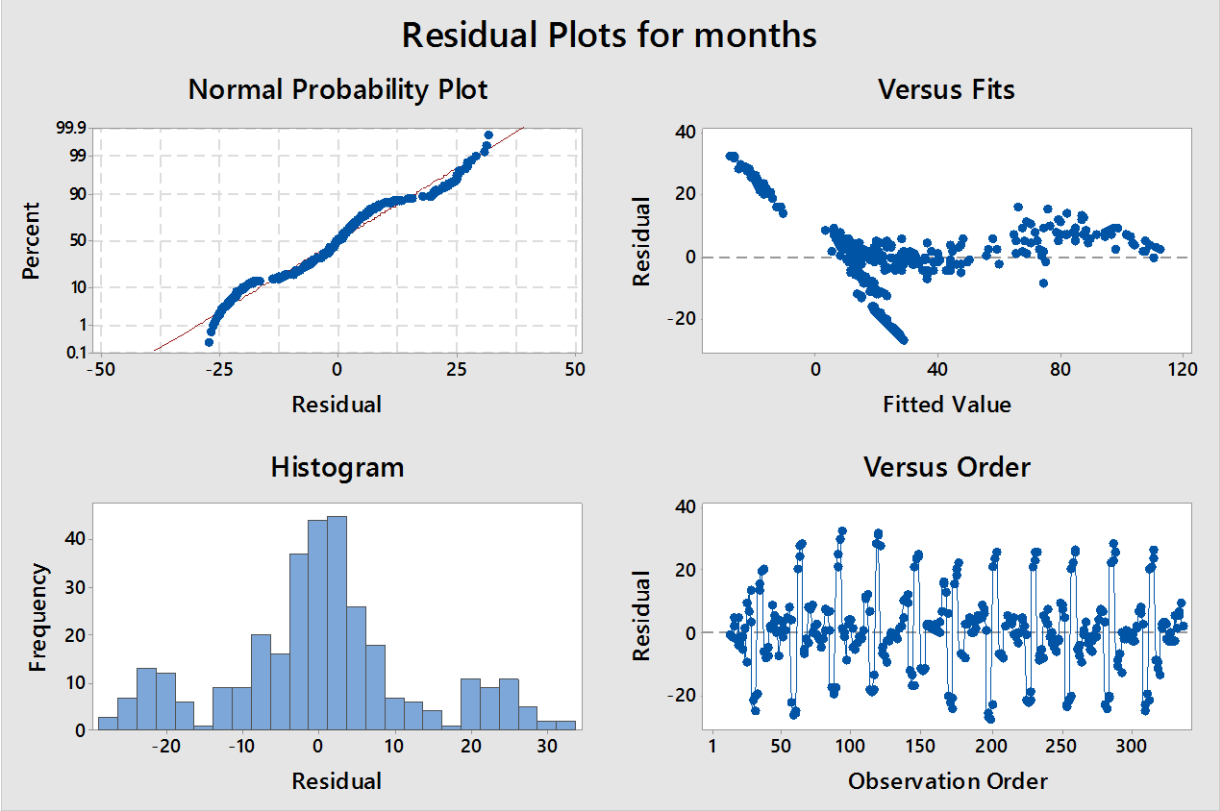
DF	SS	MS
315	52059.0	165.267

Back forecasts excluded

Modified Box-Pierce (Ljung-Box) Chi-Square Statistic آزمون پرت_مانتو

Lag	12	24	36	48
Chi-Square	504.05	820.21	1683.94	1812.26
DF	3	15	27	39
P-Value	0.000	0.000	0.000	0.000

Residual Plots for months



با توجه به نتایج فوق ، بخش Estimates at each iteration برآورد پارامترها و مجموع مربعات خطا را انجام داده و در بخش Final estimates of parameters مقادیر برآورد شده برای پارامترها در کنار انحراف استاندارد هرکدام و همچنین مقدار آماره تی و مقدار پی ول یو ، از طرفی همانطور که قبلا گفته شد در ردیف constant مقدار پی ول یو کمتر از ۰,۰۵ بوده در نتیجه فرضیه صفر بودن جمله ثابت رد می شود پس نیاز نیست که مجددا این عملیات را با برداشتن تیک Include constant term in model انجام دهیم .

بنابراین مدل آزمایشی مرکب ما با توجه به بخش Final estimates of parameters و ضرایب مدل اتو رگرسیو AR با مرتبه ۴ و میانگین متحرک MA با مرتبه ۴ به این صورت است :

$$W_t = \alpha_1 w_{t-1} + ... + \alpha_p w_{t-p} + z_t + \beta_t z_{t-1} + ... + \beta_q z_{t-q}$$

$$W_t = _0.8609 w_{t-1} + _0.2170 w_{t-2} + _0.6881 w_{t-3} + -0.9175_{wt-4} + z_t + _0.0035_{zt-1} + 1.7127_{zt-2} + 0.0024_{zt-3} + _0.7721_{zt-4}$$

حال به سراغ بررسی مناسبت مدل میرویم:

آزمون نموداری : از طریق نمودار ۴گانه ابتدا به بررسی فرض نرمال بودن باقی مانده ها می پردازیم که با توجه به نمودار اول از سمت چپ و نمودار هیستوگرام به نظر می رسد تقریبا باقی مانده ها به صورت نرمال توزیع نشده اند زیرا اکثرا در امتداد یک خط قرار نگرفته اند.

مرحله دوم به بررسی فرض استقلال باقی مانده ها می پردازیم دراین بخش باتوجه به نمودار acf,pacf داده هایمان توجه می کنیم و متوجه می شویم که در هر دو نمودار بخصوص در نمودار acf از حدود استانداردشان تجاوز کرده و این به معنی همبسته بودن باقی مانده ها و عدم تصادفی بودن شان است .

مرحله سوم به بررسی فرض ثابت بودن واریانس باقی مانده ها می پردازیم که از طریق نمودار ۴گانه ، نمودار اول از سمت راست یعنی مقادیر Fitted value ، تقریبا دارای ساختار خاصی هستند (زیرا زمانی که در این نمودار ساختار خاصی نشان داده نشد یا مثلا به شکل قیفی نباشند میتوان نتیجه گرفت واریانس ثابت است) پس می توان شهودی فرض ثابت بودن واریانس باقی مانده ها را رد کرد .

** نکته : نمودار باقی مانده ها در برابر زمان که مربوط به آخرین نمودار در نمودار۴گانه است اگر مدل مناسب باشد انتظار می رود این نمودار در اطراف سطح افقی صفر پراکندگی مستطیلی بدون روند داشته باشند از آنجا که در نمودار بالا تقریبا در اطراف سطح افقی صفر پراکندگی مستطیلی شکل ندارند و با توجه به سه مرحله گفته شده در بالا برای بررسی کفایت مدل می توان نتیجه گرفت بطور شهودی که مدل مناسبی برای پیش بینی نداریم .

آزمون عددی : از طرفی می توان با توجه به آزمون پرت_مانتو که در خروجی مینی تب اریما نیز موجود است کفایت مدل را ار طریق آزمون عددی نیز بررسی کرد و همانطور که مشاهده می کنیم مقادیر پی ول یو برای تمامی تاخیرها کمتر از ۰,۰۵ است در نتیجه فرض H_0 که فرض

صفر بودن همه خود همبستگی ها است رد می شود در نتیجه مدل مناسبی برای پیش بینی نداریم .

**اما در ادامه با توجه به انجام پیش بینی های گفته شده در مطالب قبل میتوانیم پیش بینی های کوتاه مدت یا میان مدت و همچنین بلند مدت نیز انجام داد که در همین قسمت پنجره اریما در بخش Forecasts در پنجره Lead تعداد پیش بینی ها را مشخص کرده و در ادامه پیش بینی خود را انجام می دهیم.

پایان