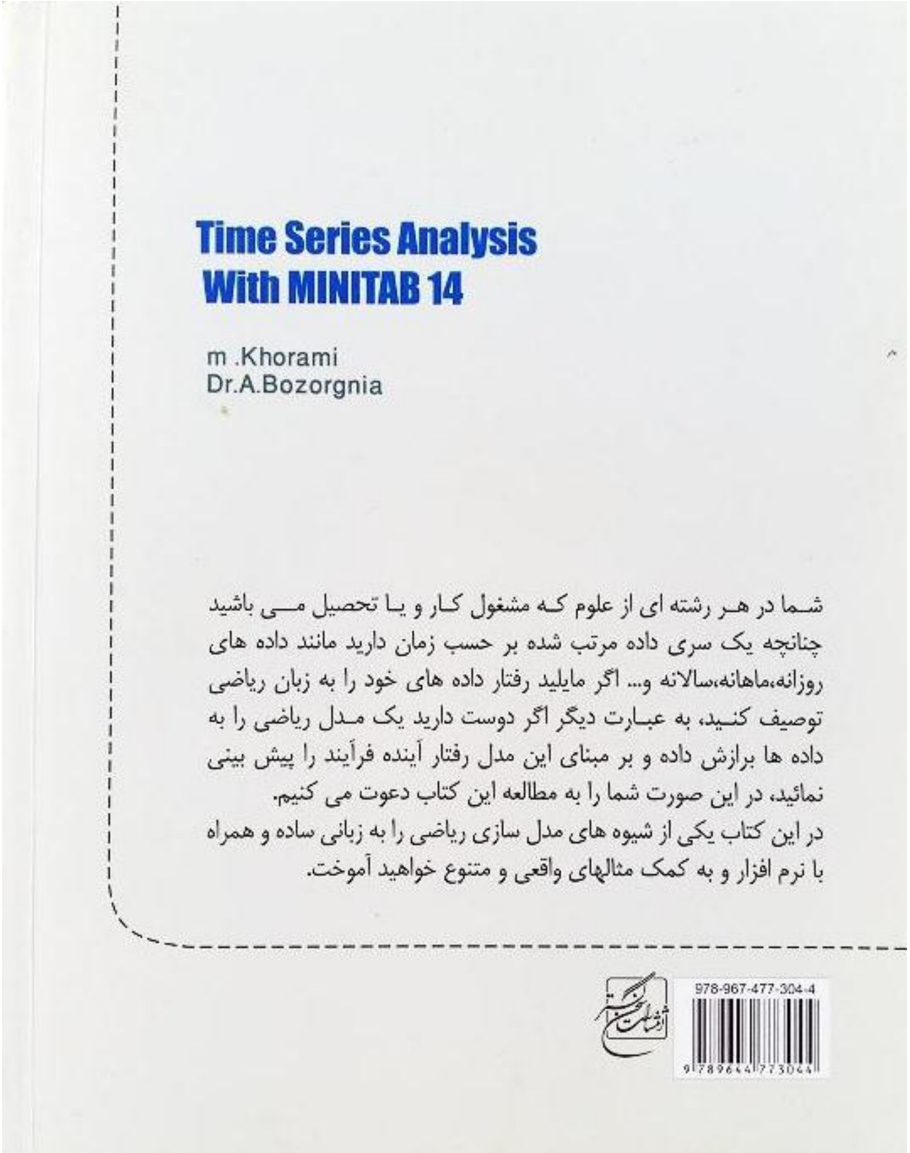

تجزیه و تحلیل
سری های زمانی
با نرم افزار MINITAB 14
Time Series Plot...
Trend Analysis...
Decomposition...
Moving Average...
Single Exp Smoothing...
Double Exp Smoothing...
Winters' Method...
Differences...
Lag...
Autocorrelation...
Partial Autocorrelation...
Cross Correlation...
ARIMA...
همراه با مثالهای واقعی و متنوع
تجزیه و تحلیل ضرایب خود همبستگی
معرفی مدلهای ایستا و ناایستا، فصلی و غیرفصلی
مدل سازی بر اساس روش باکس ـ جنکینز
با تاکید بر مدلهای ARIMA
پیش بینی
معیارهای انتخاب مدل مناسب
Autocorrelation Function for C1
(with 5% significance limits for the autocorrelations)
Autocorrelation
Lag
مولف : مصطفی خرمی ـ دکتر ابوالقاسم بزرگ نیا



خرمی، مصطفی – بزرگ‌نیا، ابوالقاسم
تجزیه و تحلیل سری‌های زمانی با نرم‌افـزار. MINITAB14 / تـألیف: مـصطفی خرمـی، ابوالقاسم بزرگ نیا – مشهد: سخن‌گستر، ۱۳۸۶.
۳۳۶ ص: جدول، نمودار.
ISBN: 978 – 964 –477-304 -4
فهرستنویسی بر اساس اطلاعات فیپا.
۱–نرم‌افزار. ۲– تجزیه و تحلیل ۳– سری‌های زمانی الف– خرمـی، مـصطفی ب– بـزرگ‌نیـا، ابوالقاسم ج– عنوان.
۵ گ ۸ ر/ ۷۹۳ QP   ۹۳۵/۲
کتابخانه ملی ایران   ۸۶-۳۷۶۵۱م

تجزیه و تحلیل سری های زمانی

با نرم افزار ۱۴ MINITAB

دكتر ابو القاسم بزرگنيا

مصطفي خرمي

تقدیم به

قلب مهربانشان

به دستان پینه بسته شان

به چشمان پرامیدشان

تقدیم به عزیزترینهای زندگیم

تقدیم به زیبا ترین واژه های هستی

پدر و مادر

فهرست مطالب

پیشگفتار

کتابهای فارسی موجود در رابطه با سریهای زمانی بیشتر به مباحث تئوری پرداخته اند. اما در عمل برای برازش یک مدل ساده به داده ها و برآورد پارامترهای آن و بررسی شایستگی مدل ناگزیر به استفاده از نرم افزار می باشیم. نرم افزارهای متعددی در این رابطه وجود دارند که به منظور تجزیه و تحلیل سری های زمانی می توان از آنها استفاده کرد. یکی ازاین نرم افزارها که در مبحث سری های زمانی قابلیتهای نسبتا خوبی دارد و بصورت گسترده از آن استفاده می شود نرم افزار MINITAB می باشد.

در تدوین این کتاب از نرم افزار MINITAB ١٤ استفاده شده است. ویرایش ١٤ این نرم افزار به نسبت ویرایشهای قدیمی تر آن از نظر گرافیکی توانایی های بیشتری دارد و نمودار های زیباتری ارائه می دهد.

این کتاب برای کلیه علاقه مندان سری های زمانی که می خواهند این مبحث را به زبانی ساده و همراه با نرم افزار بیاموزند مفید خواهد بود. در این کتاب با پرهیز از بسیاری مباحث تئوری و با ارائه مثالهای واقعی متنوع سعی شده است بیشتر به مدل سازی سری های زمانی یا به عبارتی توصیف رفتار سری به زبان ریاضی توجه شود. در همین راستا پس از مباحث پایه ای که در فصلهای اول تا چهارم مطرح شده است، در فصل پنجم بطور مفصل به مدل سازی سری های زمانی پرداخته شده است. در این فصل، ١٥ مثال متنوع که عمدتا واقعی می باشند مورد تجزیه و تحلیل قرار گرفته است.

روشی که در مدل سازی سری های زمانی در این کتاب مورد توجه قرار گرفته است روش باکس و جنکینز می باشد. این روش در عمل برای مدل سازی بسیاری از سری های زمانی مفید است و بطور گستردهای از آن استفاده می شود. با مطالعه این مثالها انتظار می رود خواننده محترم بتواند با استفاده از نرم افزار یک سری زمانی را مورد تجزیه و تحلیل قرار داده ومدل مناسبی را که به خوبی توصیف کننده رفتار آن باشد تشخیص دهد.

در پایان از زحمات کلیه دوستان و عزیزانی که در تدوین این اثر مرا یاری کرده اند قدر دانی می کنم. به ویژه از استاد ارجمند جناب آقای **دکتر ابوالقاسم بزرگ نیا** که زحمت ویراستاری علمی این اثر را تقبل نموده اند به طور خاص تشکر می کنم.

از خواهر عزیزم خانم **اکرم خرمی** که زحمت تایپ این کتاب را برعهده داشتند، همچنین از دوست صمیمیم جناب آقای **محمد بلبلیان** که در تالیف این کتاب مرا یاری کرده اند صمیمانه قدر دانی می کنم. امیدوارم این اثر که مطمئنا خالی از نقص نخواهد بود برای علاقه مندان سری های زمانی مفید واقع شود و ما را از انتقادات و پیشنهادات سازنده خود بهره مند سازند.

مصطفی خرمی - تیر ۸۵

mkhoram@yahoo.com

۰۹۱۵۸۹۲۸۹۴۸خرمي

ناشر الکترونیکی این کتاب:

شرکت داده پردازی آماری اطمینان شرق- ارائه دهنده خدمات تخصصی تحلیل آماری در کشور

آدرس وب سایت : **www.spss-iran.ir**

تلفن دریافت سفارشات تحلیل و خدمات آماری: ثابت ۰۵۱۳۷٤۱۰۷۳۹ - همراه ۰۹۱۹۸۱۸۰۹۹۱

ایمیل برای دریافت سفارش خدمات تجزیه و تحلیل آماری با نرم افزار:

mojtaba.farshchi@gmail.com

فصل اوّ ل

نمودار سری زمانی

۱-۱ تعریف سری زمانی

یک سری زمانی مجموعه مشاهداتی است که برحسب زمان مرتب شده اند. به عبارت دیگر می توان گفت یک سری زمانی عبارت است از داده هایی که از مشاهده یک پدیده در طول زمان بدست آمده اند.

ما یک سری زمانی را به عنوان دنباله ای از مشاهدات بر روی یک متغییر مورد توجه در نظر می گیریم. متغییر در نقاط گسسته ای از زمان که معمولا فاصله های مساوی دارند، مشاهده می شود. تجزیه وتحلیل سریهای زمانی متضمن توصیف فرآیندیا پدیده ای است که تولید دنباله میکند.

در یک تقسیم بندی کلی سریهای زمانی را به پیوسته و گسسته تقسیم می کنیم. یک سری زمانی را پیوسته گوییم هر گاه مشاهدات بطور پیوسته درزمان ایجاد شده باشند.یک سری زمانی را گسسته گوییم هر گاه مشاهدات فقط در زمانهای معینی که معمولا به فواصل مساوی از یکدیگر قرار دارند اخذ شده باشند.

همچنین یک سری زمانی را تصادفی گوییم اگر پیش بینی کامل آن غیر ممکن باشد و آن را غیر تصادفی گوییم اگر بتوان آن را کاملا پیش بینی کرد.

۱-۲ کاربردهای سری زمانی

سریهای زمانی در رشته های مختلف علوم ازجمله آمار، مهندسی ، اقتصاد ، مدیریت ،هواشناسی ، بازاریابی و... کاربردهای وسیعی دارند. در دنیای واقعی می توان مثالهای متعددی در زمینه های مختلف علوم پیدا کرد که درآنها یک مجموعه از مشاهدات بصورت روزانه،هفتگی،ماهانه, سالانه و... ثبت شده اند. سریهای زمانی ابزار بسیار مهمی در تجزیه و تحلیل چنین مشاهداتی است. کاربرد وسیع سریهای زمانی در علوم مختلف باعث مقبولیت آن در بین متخصصین رشته های مختلف شده است.

۱-۳ هدف از تجزیه و تحلیل سری زمانی

به طور خلاصه می توان دو هدف زیر را برای تجزیه و تحلیل سر ی های زمانی برشمرد:

۱. کشف و شناسایی مدل احتمالی مولد داده ها
۲. پیش بینی مقادیر آینده سری

در یک سری زمانی با بررسی رفتار گذشته سری مدل احتمالی که می تواند مولد داده ها باشد را شناسایی کرده و سپس با فرض اینکه داده ها در آینده نیز رفتاری مشابه خواهند داشت و از مدل برازش داده شده تبعیت خواهند نمود، سعی می کنیم مقادیر آینده سری را پیش بینی کنیم.

جهت پیش بینی سریهای زمانی لازم است که رفتار فرآیند را با یک مدل ریاضی که قابل گسترش برای آینده باشد، توصیف کرد.مدل باید نماینده خوبی ازمشاهدات در هر بخش موضعی از زمان نزدیک به حال باشد. معمولا ما نیاز نداریم که مدل نماینده مشاهدات خیلی قدیمی باشد. به علت اینکه اینگونه مشاهدات از ویژگی های زمان حال برخوردار نیستند. به محض اینکه مدل معتبری برای فرآیند سری های زمانی تثبیت شد، یک تکنیک پیش بینی مناسب را می توان اتخاذ کرد.

۱-٤ مدل سازی یک سری زمانی

مدل سازی یک سری زمانی یا به عبارتی توصیف رفتار یک سری زمانی به زبان ریاضی شامل سه مرحله کلی زیر می باشد:

۱. تشخیص مدلاولیه
۲. برآورد پارامترهای مدل شناسایی شده
۳. بررسی مناسبت مدل

ما در این کتاب مدل سازی سری های زمانی به روش باکس- جنکینز را مورد توجه قرار خواهیم داد و در فصل پنجم به طور مفصل با بررسی مثال های واقعی به تشریح این روش خواهیم پرداخت.

۱-٥ روشهای تجزیه و تحلیل سری های زمانی

برای تجزیه و تحلیل سری های زمانی دو روش کلی را می توان در نظر گرفت :

۱- تجزیه و تحلیل در قلمرو زمان
۲- تجزیه و تحلیل در قلمرو فرکانس

تحلیل سری زمانی براساس ضرایب خود همبستگی را تحلیل در قلمرو زمان می گویند. در این روش از توابع اتوکوواریانس، خود همبستگی و خود همبستگی جزئی برای مطالعه تکامل تدریجی یک سری زمانی با توجه به الگوهای پارامتری استفاده می شود.در این روش سعی می کنیم همه همبستگی موجود در مشاهدات را به مدل درآوریم. بطوریکه باقیمانده های حاصل از برازش مدل, ناهمبسته باشند.

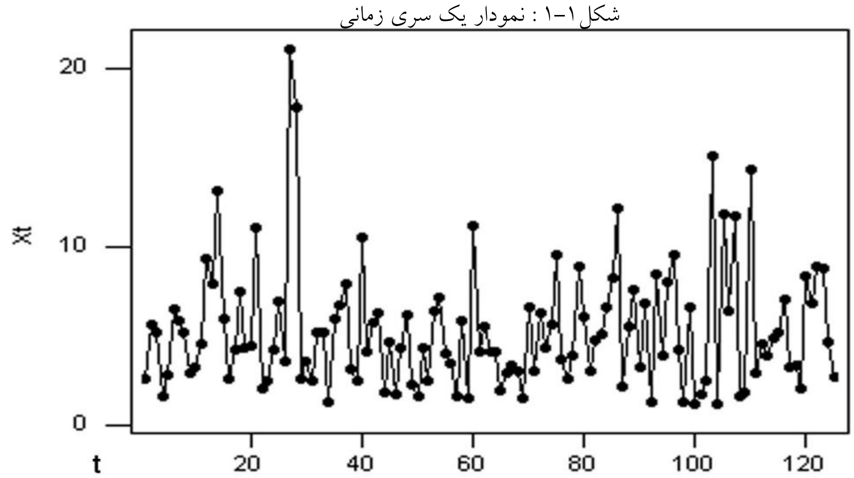
تحلیل در قلمرو فرکانس روش دیگری است که نوسانات سری زمانی را بر حسب رفتار سینوسی در فرکانسهای مختلف بیان می کند. از این روش به طور گسترده در زمینه هایی چون مهندسی برق، ژئوفیزیک و هواشناسی استفاده می شود. در این روش از تابع چگالی طیفی برای تجزیه و تحلیل سری استفاده می شود.

ما در این کتاب فقط تجزیه و تحلیل سری های زمانی تصادفی گسسته در قلمرو زمان را مورد بررسی قرار می دهیم که در عمل نیز کاربرد وسیعی دارند.

۱-۶ نمودار سری زمانی

اولین گام در تجزیه و تحلیل یک سری زمانی رسم نمودار آن سری می باشد. از این نمودار می توان اطلاعات مفیدی در مورد طبیعت داده ها بدست آورد. اگر مشاهدات یک سری زمانی را در برابر زمان وقوع آنها رسم کنیم نموداری به شکل زیر به دست می آید. x_t مشاهده در زمان t است.برای به الگو درآوردن عدم قطعیت در مشاهدات فرض می کنیم x_t برای هر نقطه زمانی،یک متغیر تصادفی باشد. بنابراین می توان رفتار x_t رابوسیله توزیعهای احتمال تعیین کرد.

شکل۱-۱ : نمودار یک سری زمانی

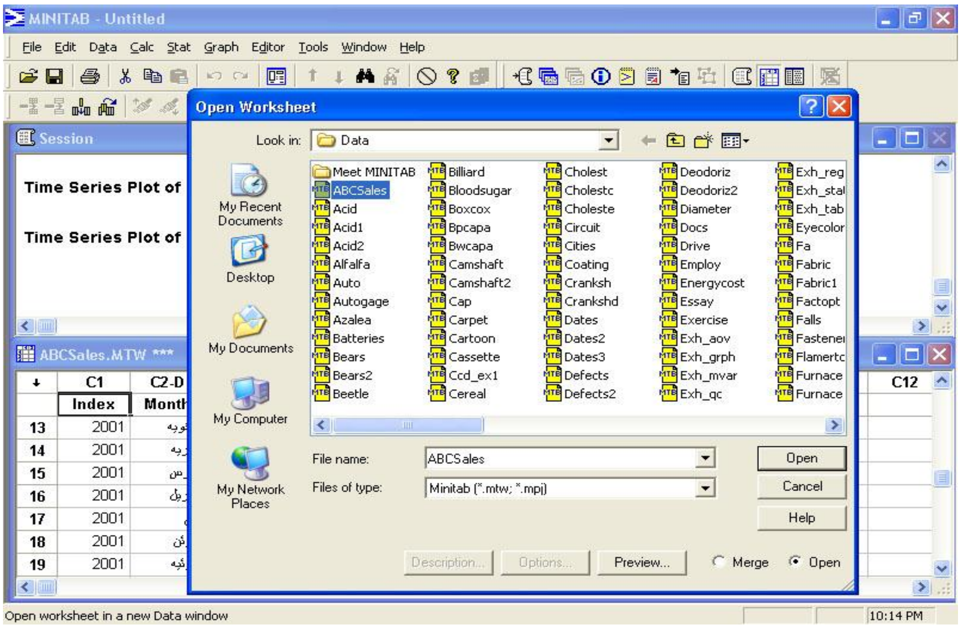


با توجه به اهمیت شیوه های گرافیکی در استنباط مفاهیم نهفته در داده ها،در این قسمت به بررسی شیوه های نمایش داده ها در Minitab می پردازیم.

۱-۷ رسم و ویرایش نمودار ها درMinitab

برای رسم نمودار یکسری زمانی لازم است که ابتدا داده ها را در یکی از ستون های مینی تب وارد نموده وذخیره کنید. چنانچه داده مناسب در اختیار ندارید، می توانید از

داده های موجود در خود مینی تب استفاده کنید. برای این منظور از منوی file گزینه Open worksheet را انتخاب کرده و سپس روی پوشه dataکلیک کرده ویکی از فایلهای آنراانتخاب کنید. ما در اینجا فایل ABCSALES.MTW را انتخاب کرده ایم.



شکل ۱–۲ : استفاده از داده های موجود در مینی تب

↓	C1	C2-D	C3	C4	C5-T	C6
	Index	Month	Sales	Advertis	AdAgency	
1	2000	January	210	30	Alpha	
2	2000	February	205	25	Alpha	
3	2000	March	202	55	Alpha	
4	2000	April	245	43	Alpha	
5	2000	May	237	60	Alpha	
6	2000	June	290	50	Alpha	

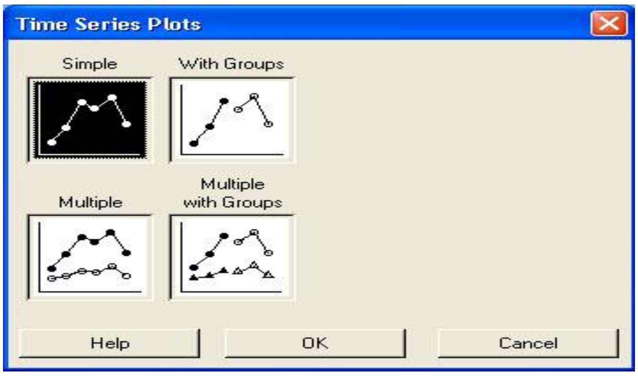
شکل ۱– ۳ : نمایش فایل abcsales.mtw در پنجره worksheet

حال برای رسم نمودار سری زمانی برای متغیر Sales از منوی stat یا Graph مراحل زیر را طی کنید :

Graph > Time Series Plot

Stat > Time Series > Time Series Plot

با طی هر یک از مراحل فوق پنجره ای به شکل زیر باز خواهد شد. در این پنجره چهار شیوه برای نماش داده ها موجود می باشد. در مورد نمودار سری زمانی ما معمولا شیوه Simple را انتخاب می کنیم.

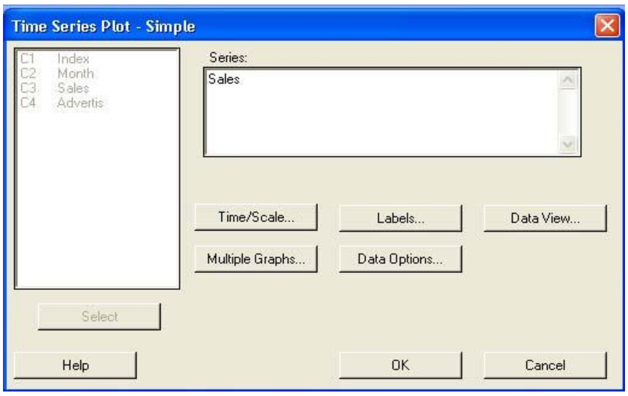


شکل ۱- ٤ : پنجره Time Series Plot

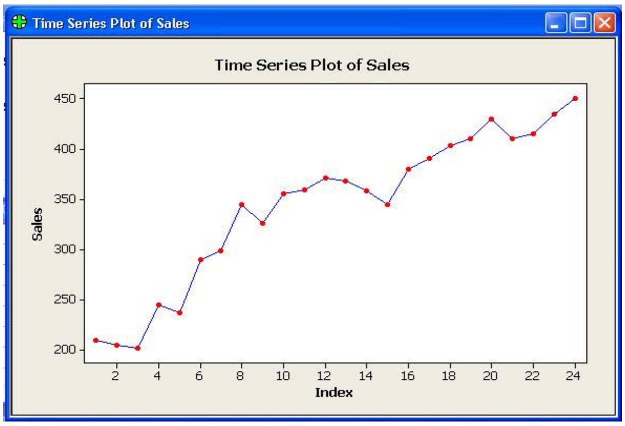
۱-۷-۱ شیوه های نمایش داده ها

روش Simple

از این روش برای نمایش داده ها بصورت ساده استفاده میشود. به عنوان مثال برای نمایش داده های مربوط به متغیر Sales به شیوه Simple پس از انتخاب گزینه Simple پنجره مربوطه را به شکل زیر تکمیل می کنیم. نمودار مربوطه نیز در شکل ۱-٦ آمده است.



شکل ۱– ۵ : رسم نمودار سری زمانی به شیوه Simple



شکل ۱– ٦ : نمودار سری زمانی برای متغیر Sales

روشWith Groups

با انتخاب این شیوه نمایش می توان علاوه بر مشاهده روند کلی متغیر مورد نظر، وضعیت متغیر را در گروههای مختلف نیز به تفکیکمشاهده نمود. برای تشریح بیشتر این روش مثال زیر را در نظر می گیریم.

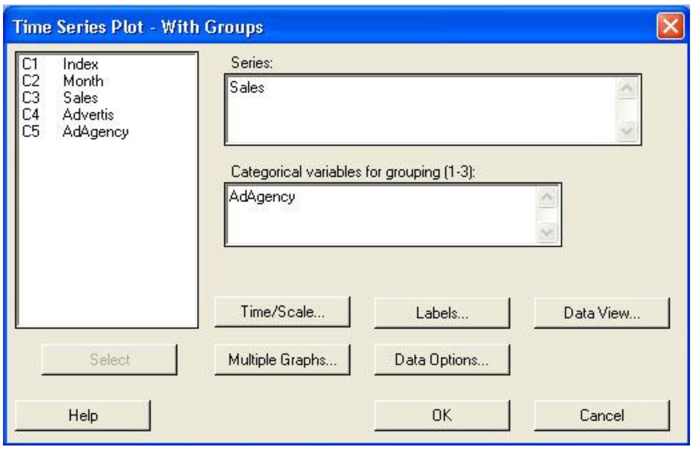
مثال۱–۱

فایل ABCSALES.MTW را که در قسمت قبل باز کردیم در نظر می گیریم. کمپانی abc برای امر تبلیغات خود در سال ۲۰۰۰ از آژانس تبلیغاتی آلفا و در سال ۲۰۰۱ از

آژانس تبلیغاتی امگا استفاده کرده است. برای بررسی فروش کمپانی در دو سال گذشته تصمیم گرفته ایم نمودار داده ها را رسم کنیم.

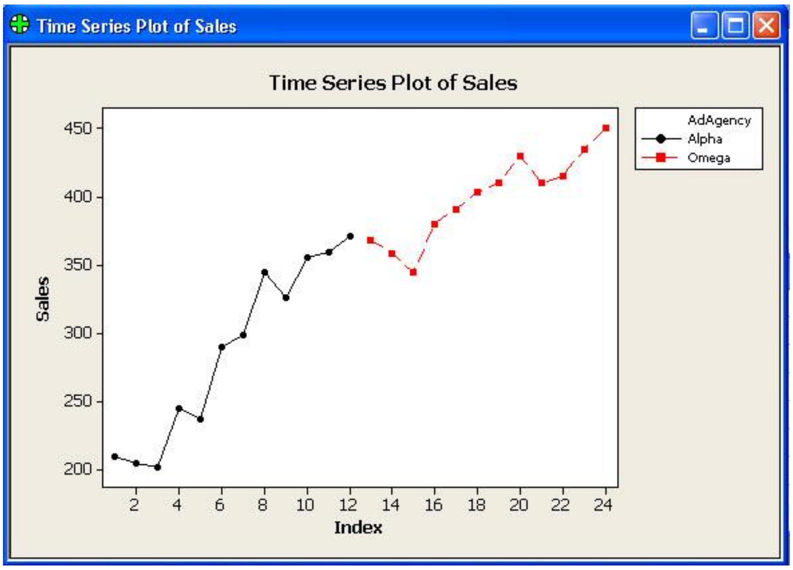
در قسمت قبل با استفاده از شیوه Simple نمودار مقادیر فروش(Sales) را در برابر زمان ملاحظه نمودید. این نمودار فقط یک روند رو به بالا را نشان می دهد و در مورد عملکرد دو شرکت آلفا و امگا چیزی نمیگوید.

اما برای نمایش داده ها بصورت With Groups پنجره مربوطه را بصورت زیر تکمیل می کنیم:



شکل ۱ – ۷ : پنجره With Groups برای رسم نمودار سری زمانی متغیر Sales

همانطور که در شکل زیرملاحظه می شود برای دو شر کت تبلیغاتی ازعلائم ورنگهای متفاوت استفاده شده است که امکان مقایسه بهتری را فراهم می آورد.



شکل ۱ – ۸ : نمودار سری زمانی برای متغیر Sales به شیوهWith Groups

روشMultiple

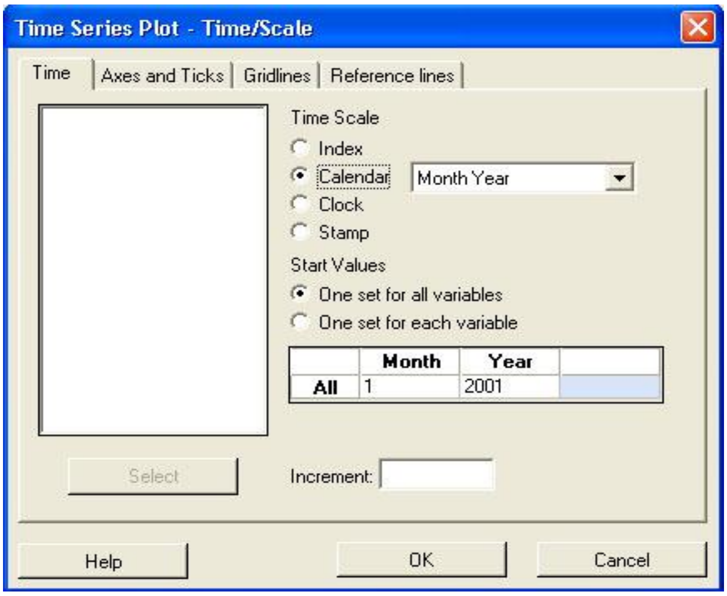
از این نحوه نمایش می توان برای رسم چندین نمودار در یک نمودار وروی یکدیگر(overlaid) استفاده کرد.این نحوه نمایش امکان مقایسه ای خوبی را در اختیار بیننده می گذارد.

مثال۱– ۲

فایلSHAREPRICE.MTWرا در نظر بگیرید.فرض کنید شما در دو شرکت abcوxyz صاحب سهام هستیدو میخواهید عملکرد ماهانه این دوشرکت را برای دو سال گذشته مقایسه کنید. برای این کار بهتر است نمودار مربوط به عملکرد هر دو شرکت را در یک نمودار رسم کنیم تا امکان مقایسه بهتر و آسان تری فراهم شود.

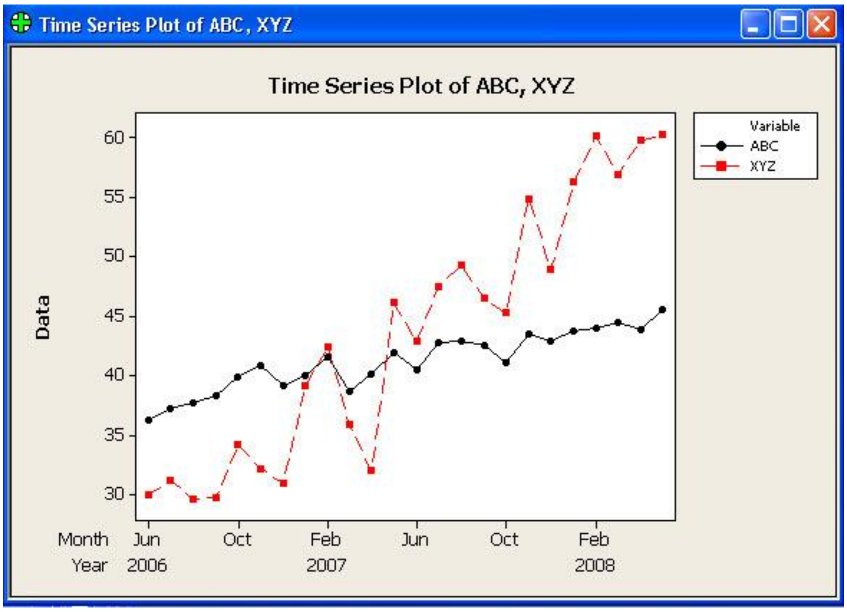
بنابراین از منوی Stat گزینه Time Series و سپس گزینه Time Series Plot را انتخاب کنید، حال از پنجره باز شده نوع نمایش داده ها راMultiple انتخاب کنید. در پنجره باز شده متغیرهای ABCوXYZ را به کادر Series منتقل کنید.

بر روی کادر Time Scale در پایین صفحه کلیک کرده و از پنجره باز شده گزینه Calendar را انتخاب کنید. حال از کادر مقابل آن حالت Month Year را انتخاب کنید. این کار برای این است که مقیاس محور x بصورت ماه و سال باشد. حال باید مقادیر آغازین را برای ماه و سال در کادر مربوطه وارد کنید. برای این کار در کادر مربوطه زیر Monthعدد ۱ و زیر Year عدد ۲۰۰۱ را وارد کنید. جزئیات بیشتر بعدا در قسمت Time Scale توضیح داده خواهد شد.



شکل ۱ – ۹ : نحوه تکمیل کردن پنجره Time/ Scale برای مثال ۱– ۲

نتیجه به شکل زیر خواهد بود :



شکل ۱ – ۱۰ : نمودار سری زمانی به شیوه Multiple

روشMultiple With Groups

از این نحوه نمایش که ترکیبی از دو حالت قبل است برای نمایش چندین نمودار سری زمانی در یک نمودار و به صورت گروه بندی شده استفاده می شود.

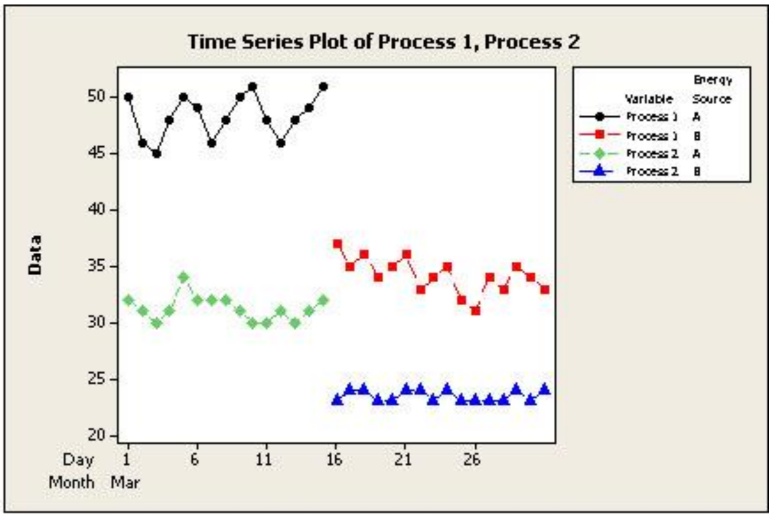
مثال ۱ – ۳

فایل ENERGYCOST.MTW را در نظر بگیرید. برای تولید نوعی پلاستیک از دو فرآیند متفاوت استفاده می شود. هر یک از این دو فرآیند را می توان از دو منبع متفاوت انرژی بنام A وB تغذیه کرد. با توجه به اهمیت هزینه استفاده از منابع مختلف انرژی برای این شرکت مایلیم این مسأله را بیشتر بررسی کنیم. برای این منظور در نیمه اول ماه از منبع انرژی A و در نیمه دوم ماه از منبع انرژی B استفاده شده است. حال با رسم نمودار هزینه استفاده از هر یک از منابع انرژی, مسأله را بررسی می کنیم.

برای این منظور از منوی Stat گزینه Time Seriesو سپس گزینه Time SeriesPlot را انتخاب کنید حال از پنجره باز شده نوع نمایش داده ها راMultiple With Groups انتخاب کنید. متغیرهای Process ۱ و Process۲ را به کادر Series منتقل کنید.

در قسمت (۳-۱) Categorical variables for grouping نام متغیر یا متغیر های گروه بندی را وارد کنید که در اینجا متغیر گروه بندی Energy Source می باشد.

همچنین در پنجره Time Scale گزینه Calendar را انتخاب کنید و سپس از کادر مقابل آن گزینه DayMonth را انتخاب کنید. در کادر مربوطه در قسمت Day عدد ۱ و در قسمت Monthعدد ۳ را وارد کنید. زیرا مشاهدات مربوط به سومین ماه سال بوده اند. نتیجه نمودار زیر خواهد بود:



شکل ۱ – ۱۱ : نمودار سری زمانی به شیوه Multiple With Groups

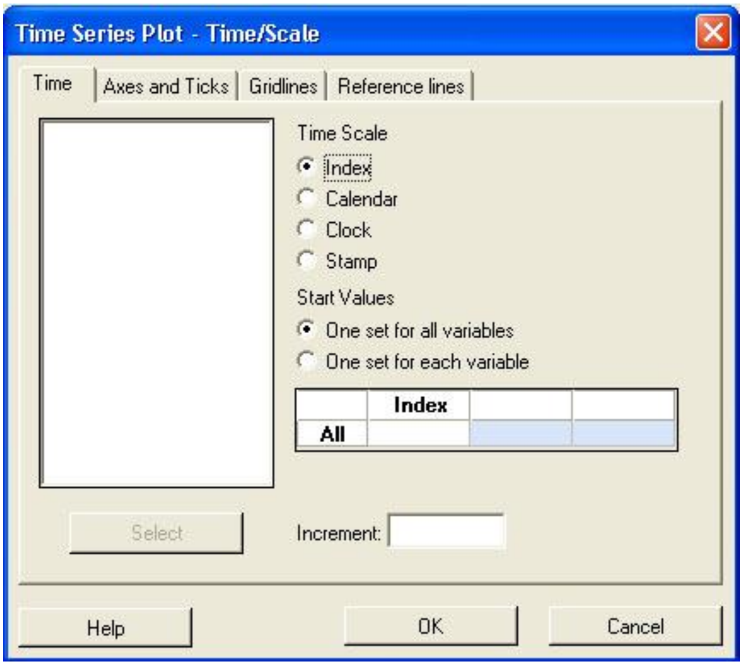
همانطور که ملاحظه می شود هزینه انرژی برای فرآیند۱ در کل بیشتر است. همچنین در صورت استفاده از منبع انرژی نوع B هزینه های هر دو فرآیند کاهش می یابد. بنابراین استفاده از فرآیند ۲ و منبع انرژی نوعB، هزینه های کمتری در بر خواهد داشت.

۱-۷-۲ امکانات پنجره Time Series Plot

برای نمایش داده ها به نحوه دلخواه می توان از امکانات مختلف پنجره Time Series Plot استفاده کرد. در این قسمت به اختصار برخی از این امکانات را معرفی می کنیم.

Time/Scale

با کلیک بر روی این گزینه، پنجره ای به شکل زیر خواهیم داشت :



شکل ۱ – ۱۲ : پنجره Time/ Scale

همانطور که می دانید محور x ,زمان یا تاریخ وقوع مشاهدات را نشان می دهد و محور y مقدار مشاهده شده در طول زمان را نمایش می دهد. در این قسمت در تب Timeمی توان مقیاس محور x را بدلخواه تعیین کرد.

گزینهIndex

حالت پیش فرض برای مقیاس محور x حالت index می باشد. در این حالت محور x با اعداد ۱ تا n شماره بندی می شود. n بیانگر شماره n امین مشاهده می باشد.

گزینهCalendar

با انتخاب این گزینه می توان واحدهای تقویمی مانند روز، ماه، فصل و سال را بر روی محور x نمایش داد. همچنین می توان ترکیبی از این واحدها را مورد استفاده قرار داد. بعنوان مثال محور x می تواند شامل ترکیبی از فصل و سال و یا ترکیبی از روز و ماه و سال باشد.

مثال ۱– ٤

فایل ABCSALES.MTWرا در نظر بگیرید. همانطور که ملاحظه می شود برای متغیر Sales علاوه بر ترتیب زمانی وقوع مشاهدات، ماه و سال مربوط به هر مشاهده نیز در ستونهای C۱ و C۲ مشخص می باشد. ما می خواهیم از این اطلاعات اضافی در نمودار سری زمانی متغیر Sales استفاده کنیم بطوریکه محور x شامل ماه و سال باشد.

ABCSales.MTW ***

↓	C1	C2-D	C3	C4	C5-T	C6
	Index	Month	Sales	Advertis	AdAgency	
1	2000	January	210	30	Alpha	
2	2000	February	205	25	Alpha	
3	2000	March	202	55	Alpha	
4	2000	April	245	43	Alpha	
5	2000	May	237	60	Alpha	
6	2000	June	290	50	Alpha	

شکل ۱– ۱۳ : نمایش فایل abcsales.mtw

برای این منظور مراحل زیر را طی می کنیم :

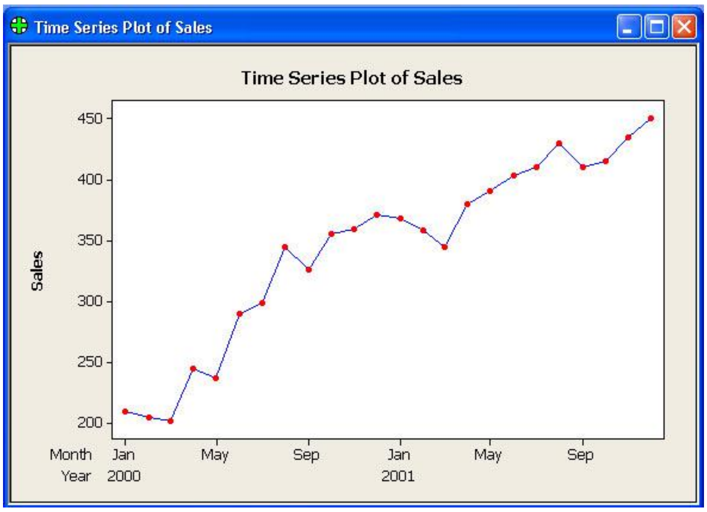
Stat > Time Series > Time Series Plot > Simple

Time ScaleCalendarMonth Year

Start Values One set for all variables ⟵

انتخاب گزینه One set for all variables باعث می شود که برای همه متغیرها مقادیر شروع یکسانی در نظر گرفته شود.

حال باید در کادر مربوطه در پایین صفحه برای هر یک از مؤلفه های مقیاس در نظر گرفته شده، مقادیر شروع آنها را مشخص کنیم. ما در اینجا برای month مقدار ۱ و برای year مقدار ۲۰۰۰ را وارد می نماییم. به این ترتیب نقطه شروع محور x ماه اول (ژانویه) سال ۲۰۰۰ خواهد بود.



شکل ۱– ۱٤ : تغییر مقیاس محور X با استفاده از گزینه Calendar

همانطور که مشاهده می شود محور x شامل دو سطر می باشد. یک سطر برای ماه و یک سطر برای سال.

گزینهClock

با فعال کردن این گزینه می توان واحد های زمانی ثانیه، دقیقه، ساعت ، روز و یا هر ترکیب سه تایی از آنها را بر روی محور x نمایش داد.

گزینه Stamp

این گزینه به شما امکان می دهد که مقیاس محور x را بصورتی که بیانگر تاریخ/زمان مشاهدات باشد، تغییر دهید. با مارک دار کردن دکمه رادیویی این گزینه آن را فعال می کنیم، حال باید نام ستون حاوی date/timeرا که در کادر سمت چپ ظاهر شده است(در صورت وجود) به این قسمت منتقل کنیم. برای روشن شدن مطلب به مثال زیر توجه کنید.

مثال ۱– ۵

فایل marketd.mtwرا بازکنید(file> open worksheet>data>marketd.mtw).

این فایل شامل مشاهداتی برای چهار فصل سالهای ۱۹۹۱و۱۹۹۲ می باشد.فصلها در ستون quarter و با شماره های۱تا٤ مشخص شده اند و ستون year بیانگر سال وقوع مشاهدات می باشد. آخرین ستون این فایل متغیرDateاست. متغیر Date ترکیبی از دو متغیر quarter و year می باشد. بعنوان مثال در ستون Date مقدار ۱Q۹۱ به اولین فصل(سه ماهه اول) سال ۱۹۹۱ اشاره می کند.

شکل ۱– ۱۵ : نمایش فایل marketd.mtw

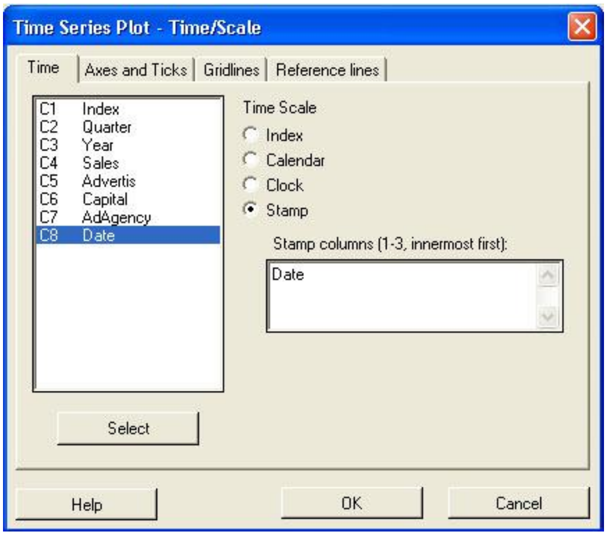
Marketd.MTW ***

↓	C1 Index	C2 Quarter	C3 Year	C4 Sales	C5 Advertis	C6 Capital	C7-T AdAgency	C8-D Date	C9
1	1	1	1991	94	17	8	Omega	1Q91	
2	2	2	1991	99	10	6	Omega	2Q91	
3	3	3	1991	98	9	12	Alpha	3Q91	
4	4	4	1991	92	22	16	Alpha	4Q91	
5	5	1	1992	106	24	29	Alpha	1Q92	
6	6	2	1992	116	18	32	Alpha	2Q92	
7	7	3	1992	113	13	33	Omega	3Q92	
8	8	4	1992	108	14	36	Omega	4Q92	
9									

برای نمایش سری زمانی Sales بگونه ای که محور x حاوی اطلاعات ستون Date باشد، به طریق زیر پنجره مربوطه را باز کنید:

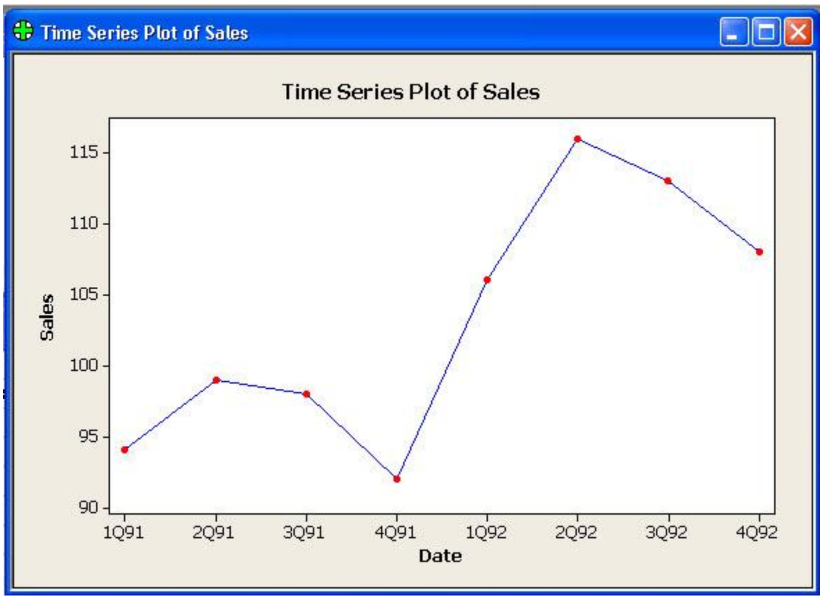
Stat > Time Series > Time Series Plot > Simple

حال در پنجره ظاهر شده متغیر sales را به قسمت series منتقل کنید.بر روی کادر Time/Scale کلیک کرده و در پنجره ظاهر شده در تب Time گزینه Stamp را انتخاب کنید.متغیر Dateرا از کادر سمت چپ به کادر مربوطه در قسمت Stampcolumns(۱- ۳) منتقل کنید و سپس دکمه okرا بفشارید. مراحل تکمیل کردن این پنجره در شکل زیر نشان داده شده است.



شکل ۱– ۱۶: استفاده از گزینه Stamp در تغییر مقیاس محور x

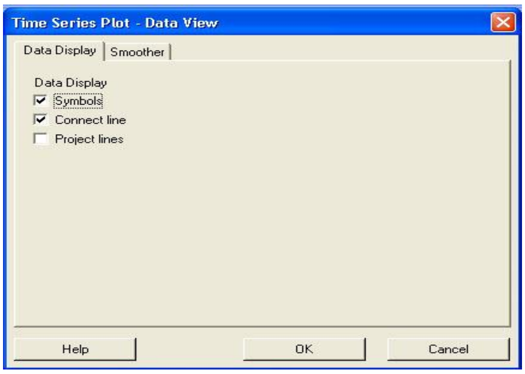
نتیجه استفاده از گزینه Stamp در نمودار سری زمانی متغیر Sales در شکل زیر نشان داده شده است. همانطور که ملاحظه می کنید محور x شامل اطلاعات متغیر Date می باشد.



شکل ۱ – ۱۷ : تغییر مقیاس محور X بصورت Date/Time

Data View

در پنجره اصلی Time Series Plot در قسمت Data View می توان نحوه نمایش داده ها را تعیین کرد. با کلیک بر روی این گزینه پنجره ای باز می شود که دارای دو تب Display Dataو Smoother می باشد.



شکل ۱ – ۱۸ : پنجره Data View

Data Displayتب

در این تب سه گزینه داریم که چگونگی نمایش خطوط و نقاط زمانی را تعیین می کنند.

گزینه Symbols

از این گزینه برای نمایش داده ها بصورت سیمبل(دایره، مربع، مثلث و ...) استفاده می شود.

گزینه Connect line

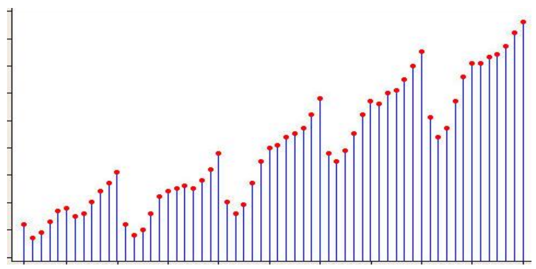
از این گزینه برای اتصال نقاط زمانی به یکدیگر به وسیله یک خط استفاده می شود.

گزینه Project lines

این گزینه هر نقطه را به Base مربوطه ربط می دهد.

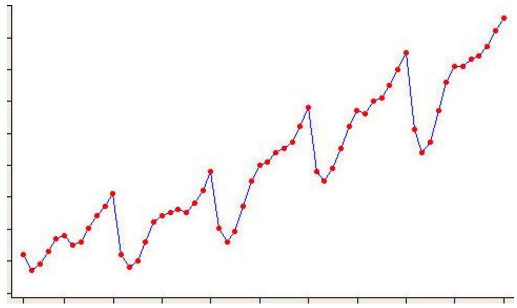
مثال ۱ – ٦

چنانچه گزینه های Symbols وProject lines را مارک دار کنیم، نموداری به شکل زیر خواهیم داشت. با توجه به آنکه سیمبل پیش فرض مینی تب دایره می باشد، در اینجا هر نقطه زمانی با یک دایره نمایش داده شده و سپس این نقاط با خطوط عمودی به محور x (Base) متصل شده اند.



شکل ۱ – ۱۹ : نمایش سری زمانی بصورت Symbols +Project lines

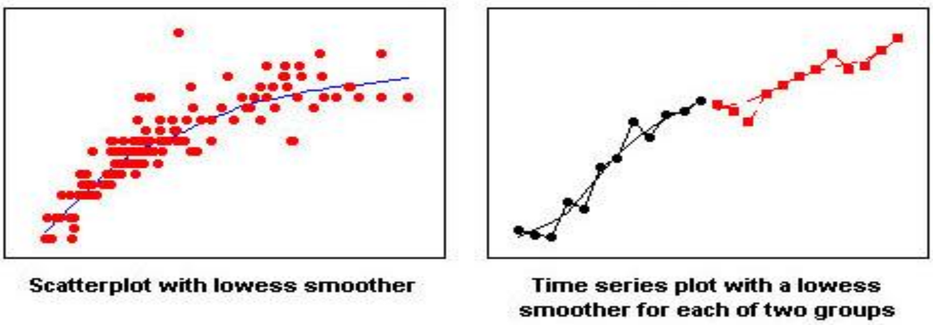
انتخاب گزینه های Symbols و Connect line نموداری به شکل زیر می دهد.



شکل ۱ – ۲۰ : نمایش سری زمانی بصورتSymbols +Connect line

تبSmoother

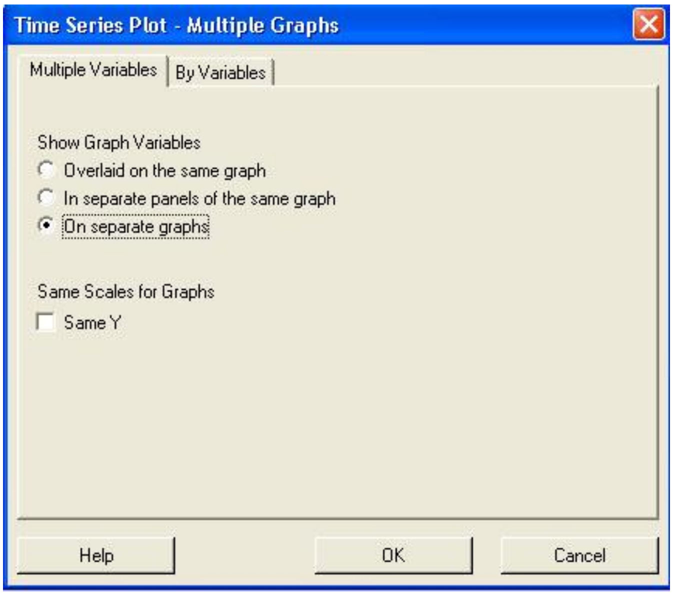
شما می توانید با استفاده از lowess Smoother بدون برازش یک مدل خاص، مانند یک خط رگرسیونی، بینشی از رابطه بین دو متغیر بدست آورید. کلمه lowess مخفف عبارت locally-weighted scatter plot smootherمی باشد. شکل زیر این مسأله را نشان می دهد.



شکل ۱ – ۲۱ : استفاده از lowess Smootherبرای درک رابطه بین دو متغیر

Multiple Graphs

در پنجره اصلی Time Series Plotدر قسمت Multiple Graphs می توان چندین نمودار را در یک صفحه یا در صفحه های جداگانه رسم کرد. این پنجره شامل دو تب Multiple Variables و By Variables می باشد.



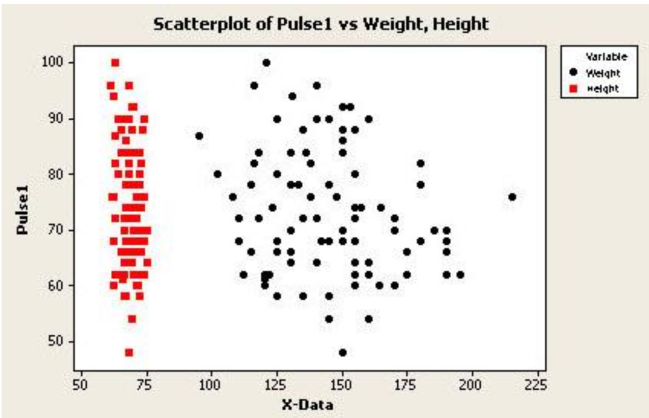
شکل ۱– ۲۲ : پنجره Multiple Graphsتب Multiple Variables

تبMultiple Variables

از این تب برای کنترل موقعیت و مقیاس چندین نمودار که آنها را بطور همزمان در یک پنجره گفتگو لحاظ کرده ایم، استفاده می شود. انتخاب هایی که این شیوه نمایش در اختیار ما می گذارد عبارتند از :

گزینه graphOverlaid on the same

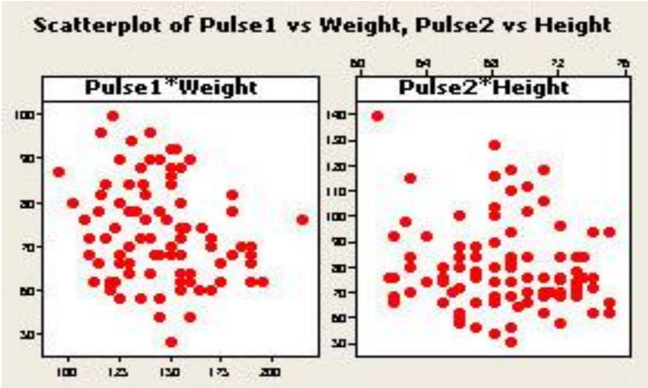
از این گزینه برای رسم نمودار ها در یک data region استفاده می شود. برای تشریح این روش، نمودار زیر را در نظر بگیرید. در نمودار پراکنش زیر که با استفاده از همین گزینه رسم شده است، می توان بطور همزمان و در یک نمودار پراکندگی متغیر pulse۱ را نسبت به متغیر weight که با دایره های سیاه مشخص شده است و متغیر height که با مربع های قرمز نشان داده شده است، مشاهده نمود. در مورد نمودارهای سریهای زمانی نیز قبلا در قسمت Multiple و Multiple With Groups نمونه ای از این نمودارها را دیده ایم.



شکل ۱ – ۲۳ :رسم دو نمودار پراکنش در یک نمودار

گزینه In separate panels of the same graph

با استفاده از این گزینه می توان چندین نمودار را در یک page اما در بخش های جداگانه نمایش داد. نمودار زیر این مطلب را به خوبی نشان می دهد.



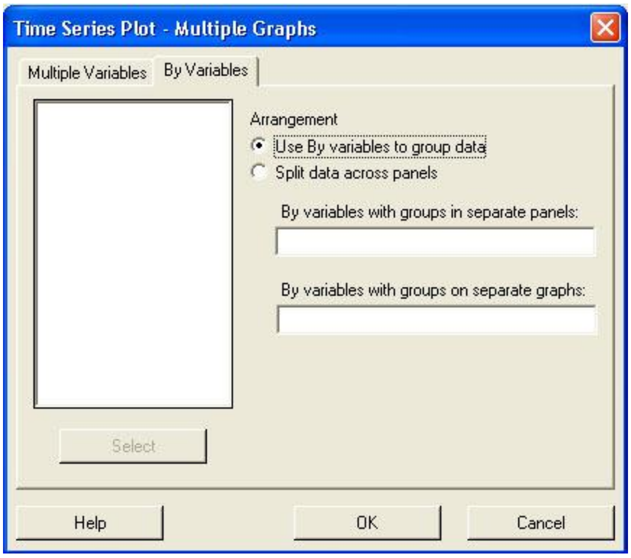
شکل ۱ – ۲٤ : رسم دو نمودار در یک نمودار ولی در بخش های جداگانه

گزینه On separate graphs

از این گزینه که حالت پیش فرض مینی تب می باشد برای نمایش نمودارها بصورت جداگانه استفاده می شود.

تب By Variables

این رویه برای تولید نمودارهای چندگانه برپایه یک متغیر گروه بندی، که ما آن را "By Variables" می نامیم، مورد استفاده قرار می گیرد. همچنین از این رویه برای جدا کردن داده ها به پنل هایی با تعداد مساوی از مشاهدات، استفاده می شود. پنجره مربوط به این تب به شکل زیر است:



شکل ۱ – ۲۵ : تب By Variables

برای توضیح چگونگی استفاده از این رویه دو مثال زیر را بررسی می کنیم.

مثال ۱– ۷

فایل EMPLOY.MTW را باز کنید. برای بررسی روند اشتغال در صنایع غذایی در Wisconsin، اشتغال ماهانه در این صنعت طی پنج سال(۲۰۰۱–۱۹۹۷) مورد بررسی قرار گرفته است. می خواهیم روند اشتغال هر یک از این پنج سال را در نمودار های جداگانه اما در یک صفحه مشاهده کنیم. برای این منظور مراحل زیر را طی می کنیم.

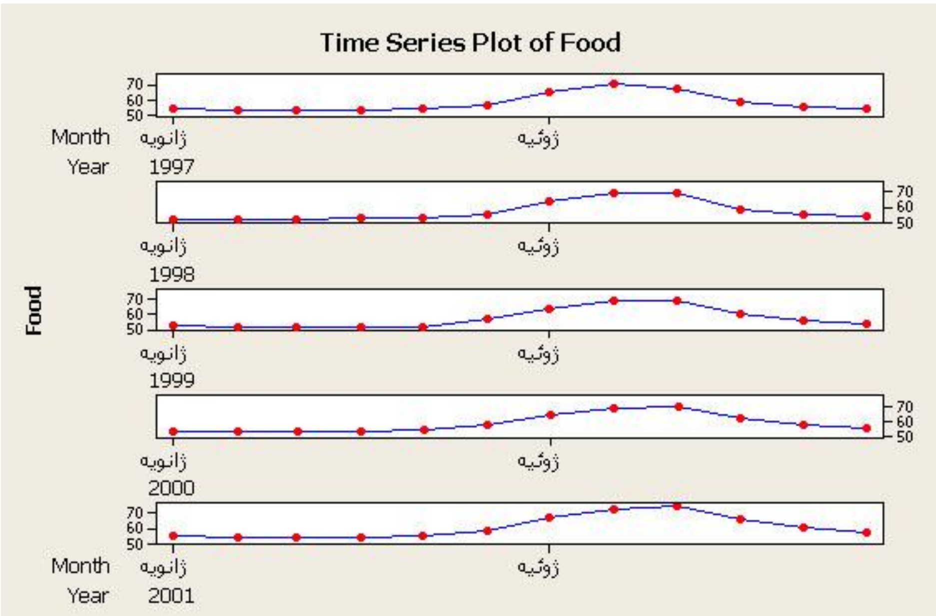
Stat > Time Series > Time Series Plot > Simple

حال در پنجره باز شده متغیر Food را به قسمت Series منتقل کنید. ابتدا برای تغییر مقیاس محور x به گونه ای که نشان دهنده ماه و سال مشاهدات باشد، در قسمت پایین

صفحه بر روی کادر Time/Scale کلیک کرده و از پنجره باز شده گزینهCalendar و از کادر مقابل آن نیز Month Year را انتخاب کنید.

در قسمت Start Valuesنیز گزینه One set for all variables را انتخاب کنید. برای ماه عدد ۱ و برای سال عدد ۱۹۹۷ را به عنوان نقطه شروع در قسمت مربوطه وارد کنید. حال با فشردن okاین پنجره را تأیید می کنیم.

اکنون در پنجره اصلی بر روی گزینه Multiple Graphs کلیک کرده و سپس تب By Variablesرا باز کنید. در این تب گزینه Split data across panels را مارک دار کردهسپس در قسمت Data per panel عدد ۱۲ را وارد کنید. زیرا هر سال شامل ۱۲ ماه است و ما برای هر ماه یک مشاهده داریم. به این ترتیب هر پنل شامل ۱۲ مشاهده خواهد بود. نتیجه به شکل زیر می باشد.



شکل ۱ –۲٦ : رسم چندین نمودار در پنل های جداگانه در یک صفحه

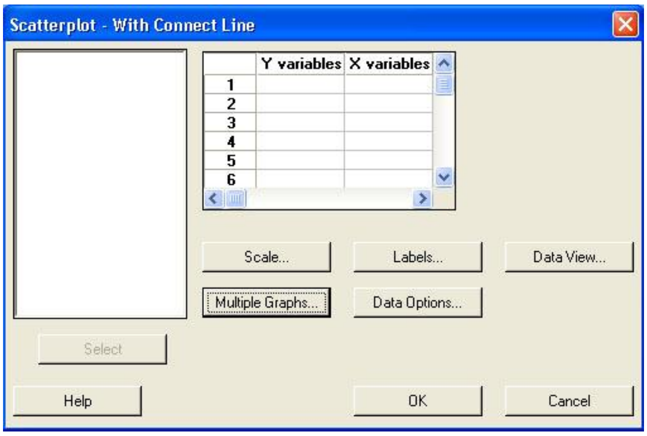
این شیوه نمایش داده ها امکان مقایسه بهتر روند سالانه اشتغال در صنایع غذایی را فراهم می آورد. همانطور که ملاحظه می شود، هر سال اشتغال در صنایع غذایی در ماه

های زمستان کاهش می یابد و از اواسط تابستان رو به افزایش می گذارد و به اوج خود می رسد و سپس دوباره رو به کاهش می نهد. از این شیوه نمایش در مورد نمودارهای پراکنش نیز می توان بخوبی استفاده کرد.

مثال ۱ – ۸

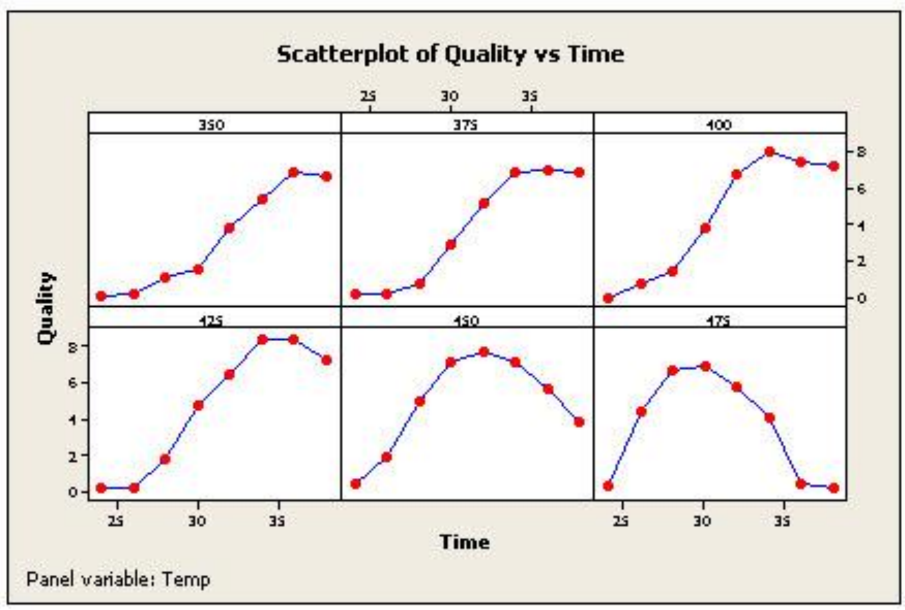
فایل REHEAT.MTW را باز کنید. در این مثال می خواهیم تأثیر دمای پخت و زمان پخت را بر کیفیت غذا بررسی کنیم. ما به دنبال یک شیوه گرافیکی هستیم که بتوان به سادگی تأثیر دماها و زمانهای مختلف را بر کیفیت غذا به طور همزمان مشاهده نمود.

برای این کار کافی است از منوی Graph گزینه Scatterplot و سپس روش WithConnect Line را انتخاب کنید تا پنجره زیر باز شود.



شکل ۱ – ۲۷ : پنجره Scatterplot – With Connect Line

حال در قسمت Y variablesمتغیر Quality و در قسمت X variablesمتغیرTime را وارد کنید. اکنون بر روی کادر Multiple Graphs در پایین صفحه کلیک کرده از پنجره باز شده تب By Variablesرا باز کنید و سپس گزینه By variables with groups in separate panels را مارک دار کنید.



شکل ۱ – ۲۸ : نمودار پراکنش برای یافتن ترکیب بهینه زمان و دما

در اینجا متغیر دما(temp) به عنوان متغیر پنل(panel variable)ظاهر شده است.هر پنل نشان دهنده دمای یکسان اما زمانهای پخت مختلف می باشد و ما به دنبال ترکیب مناسبی از دما و زمان هستیم که به حداکثر کیفیت منجر می شود.

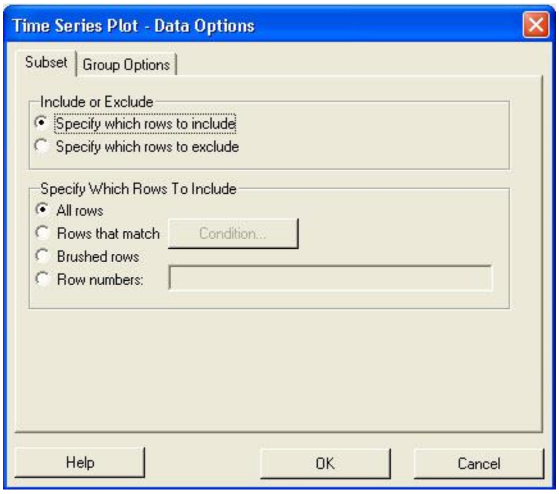
همانطور که ملاحظه می شود در دماهای پایین تر(پنل های بالا) برای رسیدن به حداکثر کیفیت به زمان بیشتری احتیاج داریم. اما در دما های بالاتر(پنل های پایین) زمان پخت بیشتر، باعث کاهش کیفیت غذا می شود.به نظر می رسد زمان و دمای بهینه

که منجر به حداکثر کیفیت می شود عبارت است از دمای ٤٢٥ درجه سانتی گراد و زمان پخت ٣٤ دقیقه.

در پنجره اصلی Time Series Plotدو گزینه دیگر وجود دارند که در رسم و ویرایش نمودارها به ما کمک می کنند. این دو گزینه در زیر به اختصار توضیح داده می شوند.

Data Options

چنانچه بخواهیم یک زیر مجموعه خاص از داده ها را مورد تجزیه و تحلیل قرار دهیم، برای رسم نمودارهای آن زیر مجموعه خاص از امکانات این قسمت استفاده می کنیم.در این قسمت می توان بعضی از سطرها(مشاهدات) را از مجموعه داده ها حذف کرد و یا بلعکس می توان بر روی سطرهای خاصی تمرکز کرد.

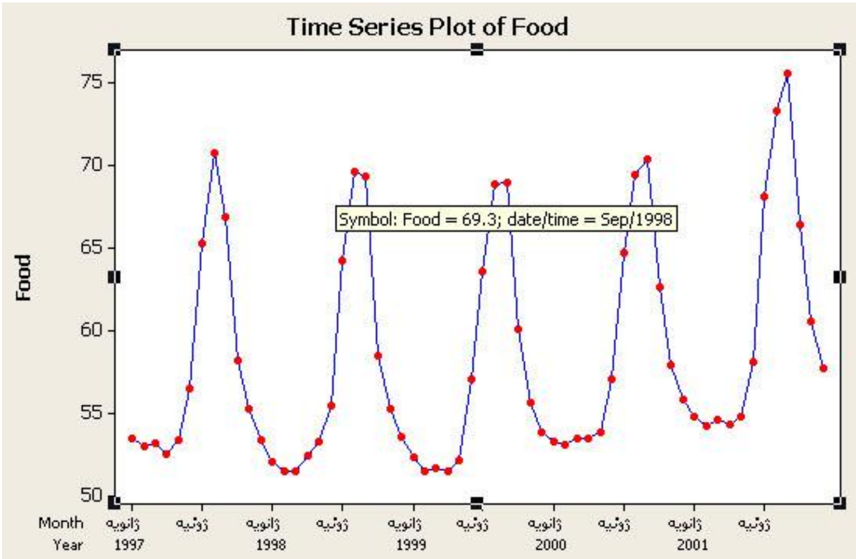


شکل ١ – ٢٩ : پنجره Data options

Labels

برای اضافه کردن عنوان و زیرنویس به نمودار و دادن برچسب به داده ها از این قسمت استفاده می شود.

توجه:با قرار دادن نشانگر موس بر روی هر یک از نقاط نمودار، مختصات آن نقطه در یک کادر زرد رنگ نمایش داده می شود، در نتیجه به راحتی می توان زمان وقوع هر مشاهده و مقدار آن را ملاحظه نمود.



شکل ۱ – ۳۰ : مشاهده مختصات هر نقطه در نمودار سری زمانی

نکته : با دو بار کلیک کردن کردن بر روی هر یک از مؤلفه های یک نمودار مانند Symbols، Connect line، Project lines، data region و... می توان نوع، رنگ، اندازه و سایر ویژگیهای آن مؤلفه را تغییر داد. به این ترتیب می توان یک نمودار را پس از رسم آن نیز ویرایش نمود.

فصل دوم

مفاهیم اساسی

در این فصل به معرفی مفاهیم اساسی در سریهای زمانی می پردازیم. این مفاهیم شامل اجزاء سری زمانی ،رابطه بین این اجزاء، ایستایی و همبستگی می باشد.

۲ – ۱ اجزاء یک سری زمانی

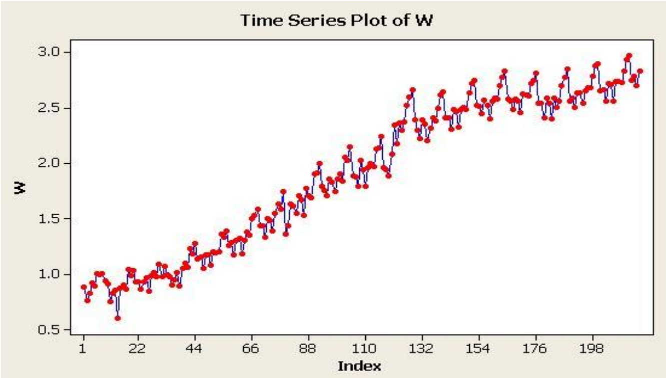
بسیاری از محققین برای توصیف رفتار یک سری زمانی ، اجزاء زیر را برای سری زمانی در نظر می گیرند :

- روند یا تمایل بلند مدت (trend)
- تغییرات دوره ای (cyclical variation)
- تغییرات فصلی (seasonal variation)
- تغییرات نامنظم (irregular variation)

اکنون به اختصار به معرفی این اجزاء می پردازیم.

۱- روند (T)

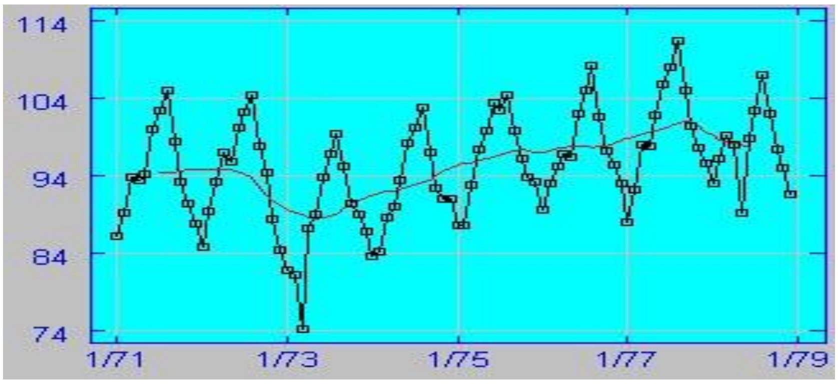
روند یا تمایل بلند مدت عبارت است از تحول متغییر مورد مطالعه در یک دوره طولانی بدون در نظر گرفتن تغییرات دوره ای ، فصلی و نامنظم. به عبارت دیگر می توان گفت روند عبارت است از حرکات رو به بالا و پایین یک سری زمانی که نشان دهنده کاهش یا افزایش بلند مدت یک سری زمانی است. نمودار زیر یک روند افزایشی را نشان می دهد.



شکل ۲ – ۱ : نمودار یک سری زمانی دارای روند رو به بالا

۲- تغییرات دوره ای یا سیکل(C)

تغییرات دورهای عبارت است ازتکرارحرکات رو به بالاوپایین حول سطوح روند. این نوع تغییرات داری دوره نوسان بیشتر از یک سال می باشند. نوسانات دوره ای ممکن است دقیقا از طرح های مشابهی بعد از فواصل زمانی مساوی پیروی کنند ولی همیشه این طور نیست. یک دوره کامل را که معمولا ۷ تا ۹ سال طول می کشد اصطلاحا یک "دوره" می نامند. یکی از معمولی ترین نوسانات سیکلی داده های سری زمانی، سیکل تجاری است. سیکل تجاری وقوع مکرر دوره های رونق و رکود است.

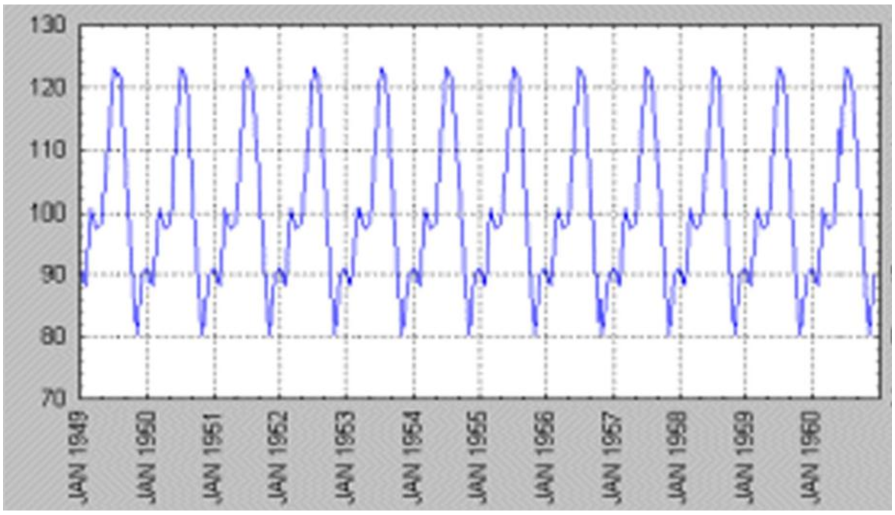


شکل ۲– ۲ : یک سری زمانی دارای تغییرات دوره ای

۳- تغییرات فصلی(S)

تغییراتفصلیتغییراتی هستندکهدردورههای تناوبی کوتاه پیش میآیند. این تغییرات مربوط به عواملی هستند که به طریقی منظم و چرخه ای روی یک دوره کمتر از یک

سال عمر می کنند. در واقع تغییرات فصلیرفتار دوره ای متغیر را نشان می دهد، یعنی رفتاری که معمولا هر سال در همان فصل تقریبا با همان شدت روی می دهد.



شکل ۲ – ۳ : نمودار یک سری زمانی دارای تغییرات فصلی

٤- تغییرات نامنظم(I)

تغییرات نامنظم عبارت است از حرکات پراکنده در یک سری زمانی که از الگوی منظم و مشخصی پیروی نمی کنند. در واقع این حرکات بیان میکنند که پس از محاسبه روند، تغییرات دوره ای و تغییرات فصلی چه چیز دیگری در سری زمانی بجا می ماند.

نوسانات نامنظم معمولا ناشی از وقایع غیرمعمولی هستند که قابل پیش بینی نیستند. مانند زمین لرزه، اعتصاب، طوفان، جنگ، تصادفات و ....

۲-۱-۱ رابطه بین اجزاء تشکیل دهنده سری زمانی

با توجه به ماهیت رابطه بین عناصر سازنده سری دو الگوی کلی را می توان در نظر گرفت :

الف) الگوی جمعی :در الگوی جمعی فرض می شود که چهار عنصر تشکیل دهنده سری از یکدیگر مستقل هستند. چون صحت این شرط در بیشتر موارد مورد تردیدمی باشد, اغلب از این الگو استفاده نمی شود. با توجه به فرض استقلال اجزاء سری این الگو به شکل $Y_i = T_i + C_i + S_i + I_i$ نوشته می شود.

ب) الگوی ضربی :در الگوی ضربی فرض می شود که چهار عنصر تشکیل دهنده سری به یکدیگر وابسته اند. در نتیجه این الگو به شکل $Y_i = T_i * C_i * S_i * I_i$ نوشته می شود.

۲-۲ ایستایی

ایستایی مبحث بسیار مهمی در مدل سازی سری های زمانی می باشد. زیرا بسیاری از مدل های احتمال سری های زمانی بر مبنای ایستایی سری استوار می باشند.

تعریف ۲-۱(ایستایی اکید):یک سری زمانی را اکیدا ایستا گوییم هرگاه به ازای همه مقادیر $t_1, t_2, \dots, t_n$ وτتوزیعتوأ م مانند $x(t_1),\dots,x(t_n)$ توزیع توأم $x(t_1+\tau),\dots,x(t_n+\tau)$ باشد.

تعریف ۲-۲(ایستایی ضعیف):یک سری زمانی را ایستای ضعیف یا ایستای مرتبه دوم می نامند هرگاه میانگینش ثابت باشد و تابع اتوکوواریانس آن فقط به تأخیر k بستگی داشته باشد. به عبارت دیگر میانگین و تابع اتو کوواریانس آن به زمان بستگی نداشته باشد.

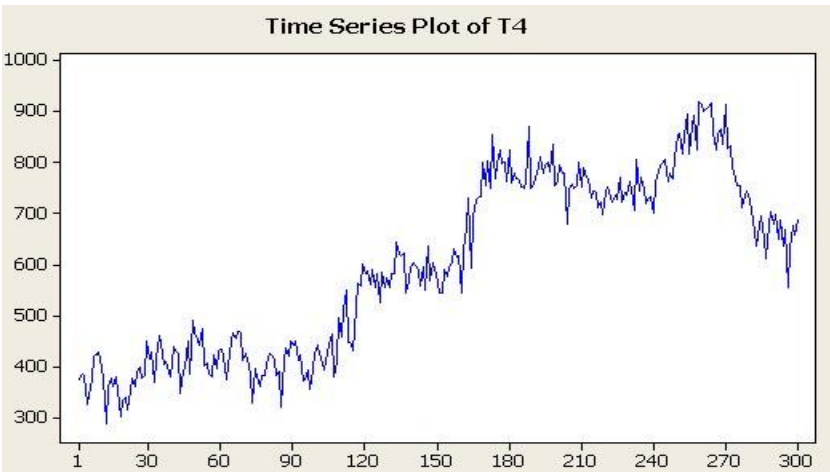
تعریف ۲-۳ (تابع اتوکوواریانس در تأخیر k): این تابع، همبستگی بین مشـاهداتی را کـه k واحد زمانی با یکدیگر اختلاف دارند اندازه می گیرد و بصورت زیر تعریف می شـود :

$$\gamma_k = \mathrm{cov}(x_t, x_{t+k}) = E(x_t - \mu)(x_{t+k} - \mu)$$

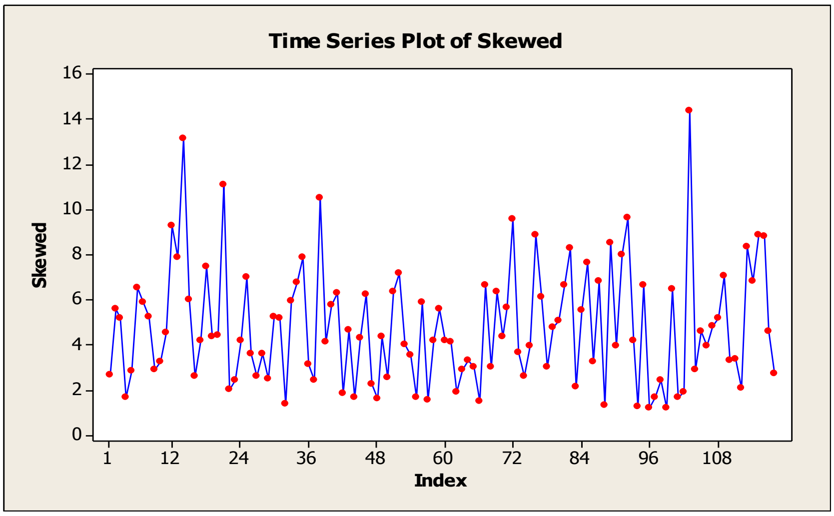
در مبحث همبستگی، تابع خود همبستگی را تعریف می کنیم که تعریفی مشابه خواهد داشت با این تفاوت که این تابع بر خلاف تابع اتوکوواریانس به واحد اندازه گیری مشاهدات بستگی ندارد.

از این پس منظور ما از ایستایی همان ایستایی مرتبه دوم می باشد که درعمل نیز بررسی تحقق آن ساده تر است. به عبارت ساده تر می توان گفت: یک سری زمانی را ایستا یا مانا می نامیم هرگاه مشخصه های آماری آن مانند میانگین و واریانسش در طول زمان ثابت بماند. مفهوم اساسی ایستایی این است که قوانین احتمالی حاکم بر فرآیند با زمان تغییر نمی کند یعنی فرآیند در تعادل آماری باقی می ماند.

اگر یک سری زمانی ایستا باشد، منظور این است که این سری زمانی بطور تصادفی حول یک میانگین ثابت نوسان می کند و اگر سری ناایستا باشد بدین معنی است که این سری هیچ میانگین ثابتی ندارد. یک سری دارای روند، یک سری ناایستا است. به این دلیل که میانگینش ثابت نیست و همراه با زمان در حال افزایش یا کاهش می باشد.



شکل ۲– ٤ : نمودار یک سری زمانی ناایستا



شکل ۲ – ۵ : نمودار یک سری زمانی ایستا

برای بررسی ایستایی یک سری زمانی علاوه بر نمودار آن می توان از همبستگی نگارآنیز استفاده کرد. چنانچه یک سری زمانی ایستا باشد همبستگی نگار آن به سرعت به سمت صفر میل می کند. در مورد سریهای ناایستا نیز همبستگی نگار به کندی به سمت صفر میل می کند. این روش در مبحث همبستگی بیشتر توضیح داده می شود.

همانطور که گفتیم ایستایی مبحث مهمی است. زیرا نظریه احتمال سری های زمانی بیشتر با سری های زمانی ایستا سروکاردارد، بنابراین لازم است که درصورت ناایستا بودن سری ابتدا آن را به یک سری ایستا تبدیل کنیم. برای مثال می توانیم روند و تغییرات فصلی را از مجموعه داده ها حذف کرده و سپس سعی کنیم بوسیله یک فرآیند تصادفی ایستا، تغییر در باقیمانده ها را الگو سازی نماییم.

البته گاهی ممکن است بخواهیم یک سری ناایستا را به همان صورتی که هست بررسی کنیم. مثلا در یک سری دارای روند، می توان جهت پیش بینی های بلند مدت یک منحنی روند را به داده ها برازش داد، اما جهت پیش بینی های کوتاه مدت و دقیقتر ممکن است بخواهیم روند را از داده ها حذف کنیم تا به یک سری ایستا برسیم و آنگاه از یکی از الگوهای احتمال جهت پیش بینی استفاده کنیم.

یک سری زمانی ممکن است در میانگین یا در واریانس یا در هر دو مورد ناایستا باشد. راه حل مناسب برای ایستاسازی یک سری زمانی که درمیانگین ناایستا است، تفاضلی کردن آن سری می باشد. در مورد سری زمانی که در واریانس ناایستا است, یعنی واریانس آن ثابت نیست و همراه با زمان تغییر می کند، راه حل مناسب استفاده از تبدیلات باکس–کاکس می باشد. اینک ایستاسازی یک سری زمانی را در هر یک از دو حالت شرح می دهیم.

ناایستایی در میانگین

چنانچه یک سری زمانی در میانگین ناایستا باشد، مهمترین ابزار تبدیل این سری به یک سری ایستا تفاضلی کردن می باشد که در این بخش به معرفی آن می پردازیم. اما ابتدا عملگر پسرو را معرفی می کنیم که در مبحث سری زمانی از آن زیاد استفاده می شود.

عملگر پسرو

عملگر پسرو که با B نشان داده می شود، روی شاخص زمانی عمل کرده و آنرا به اندازه یک واحد زمانی به عقب می برد. این عملگر به صورت زیر تعریف می شود:

$$Bx_t = x_{t-1}$$

بطور کلی برای هر عدد صحیح m داریم

$$B^m(x_t) = x_{t-m}$$

تفاضلی کردن

تفاضلی کردن مرتبه اول و دوم یک سری زمانی را بصورت زیر تعریف می کنیم :

$$\nabla x_t = x_t - x_{t-1}$$
$$\nabla^2 x_t = \nabla x_t - \nabla x_{t-1} = x_t - x_{t-1} - (x_{t-1} - x_{t-2}) = x_t - 2x_{t-1} + x_{t-2}$$

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

در عمل با یک یا دو بار تفاضلی کردن می توان یک سری ناایستای غیرفصلی را به یک سری ایستا تبدیل کرد. درصورت وجودعامل فصلی درسری ناایستا، تبدیلات پیچیده تری لازم است. به سهولت می توان تفاضلی کردن را برحسب عملگر پسرو به صورت زیر نوشت:

$$\nabla x_t = x_t - x_{t-1} = x_t - B(x_t) = (1-B)x_t$$

بنابراین می توان عملگر تفاضلی ∇ را بصورت زیر نوشت :

$$\nabla = 1 - B$$
$$\nabla^d = (1-B)^d$$

تفاضلی کردن فصلی

یک وسیله مهم در مدل سازی فرآیندهای فصلی ناایستا، تفاضلی کردن فصلی یا دیفرانس گیری فصلی است. اپراتور دیفرانس فصلی را به صورت زیر تعریف می کنیم:

$$\nabla_S = 1 - B^S$$

در نتیجه خواهیم داشت :

$$\nabla_S x_t = (1-B^S)x_t = x_t - x_{t-S}$$

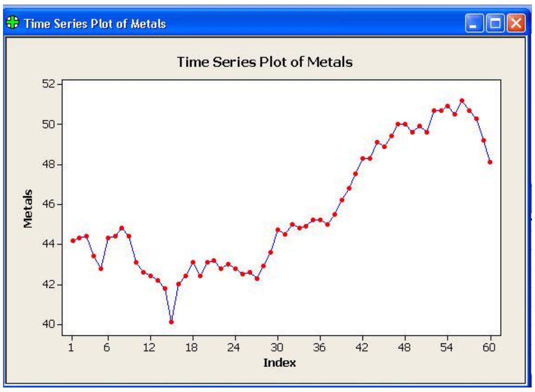
مثلا برای یک سری ماهانه، تغییرات از فروردین تا فروردین، اردیبهشت تا اردیبهشت و غیره را برای سالهای متوالی در نظر می گیریم. برای یک سری بطول n سری تفاضلی شده بطول $n-s$ می باشد، یعنی بواسطه تفاضلی کردن فصلی s داده از دست می رود.

ممکن است D دیفرانس فصلی مورد نیاز باشد تا یک سری ایستا تولید شود. لذا اپراتور دیفرانس فصلی از درجه D بصورت زیر تعریف می شود:

$$\nabla_S^D = (1 - B^S)^D$$

مثال۲-۱ ایستا سازی سری زمانی در Minitab

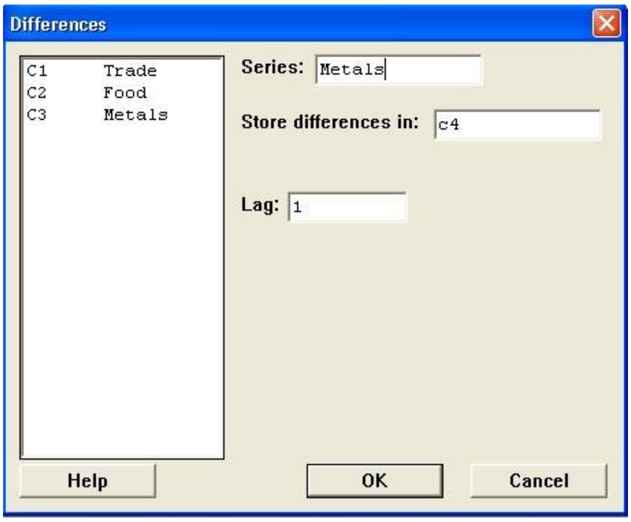
سری ناایستای Metals از فایل Employ را در نظرمیگیریم. نمودار این سری بصورت زیر می باشد.



شکل ۲- ٦ : نمودار سری نا ایستای Metals

همانطور که ملاحظه می شود این سری در میانگین ناایستا است. اکنون می خواهیم با استفاده از نرم افزار این سری را به یک سری ایستا تبدیل کنیم. همانطور که قبلا نیز گفته شد مهمترین ابزار تبدیل یک سری نا ایستا به یک سری ایستا، تفاضلی کردن آن سری می باشد.

در Minitab برای ایستاسازی یک سری زمانی با استفاده از تفاضلی کردن، ابتدا از منوی Stat گزینه time series و سپس گزینه Differencesرا انتخاب می کنیم تا پنجره زیر باز شود.

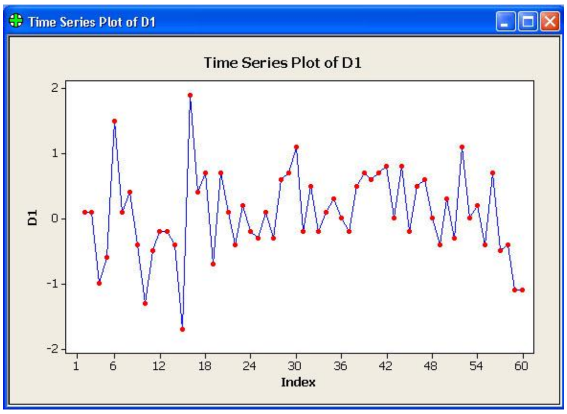


شکل ۲ – ۷ : پنجره Differencesبرای تفاضلی کردن سری زمانی

در پنجره ظاهر شده، درقسمت Lag مرتبه تفاضلی کردن را تعیین می کنیم. در اینجا برای تفاضلی کردن مرتبه اول در کادر مقابل Lag عدد ۱ را وارد می کنیم.

برای ذخیره کردن مقادیر تفاضلی شده نیز نام یکی از ستونهای مینی تب را در قسمت مربوطه وارد می کنیم. با ذخیره کردن این مقادیر، می توان ایستایی سری تفاضلی شده را بررسی کرد.ما در اینجا سری ناایستای Metals را یک بار تفاضلی کرده و مقادیر سری تفاضلی شده را در ستون C٤ ذخیره می کنیم.

برای آنکه ببینیم آیا با یک بار تفاضلی کردن به یک سری ایستا رسیده ایم یا نه می توانیم نمودار سری زمانی را برای سری تفاضلی شده که در اینجا ما آن را C٤ نامیده ایم، رسم کنیم. چنانچه این نمودار هنوز هم ناایستایی در میانگین را نشان دهد، باید تفاضلی کردن مراتب بالاتر را امتحان کرد.



شکل ۲– ۸ : نمودار سری تفاضلی شده Metals

می توان گفت سری فوق یک سری ایستا است. زیرا بیشتر مقادیر حول یک میانگین ثابت متمرکز شده اند. بنابراین تفاضلی کردن مرتبه اول مؤثر واقع شده است و نیازی به تفاضلی کردن بیشتر نمی باشد.

تفاضلی کردن زیاد

تفاضلی کردن یک فرآیند ایستا نیز یک فرآیند ایستا را نتیجه می دهد. با این وجود تفاضلی کردن زیاد ممکن است همبستگی غیر لازمی را در الگو ایجاد کند و الگوی نسبتا ساده ای را پیچیده نماید. بنابراین باید از تفاضلی کردن بیش از حد سری اجتناب کرد.

مفهوم انباشتگی

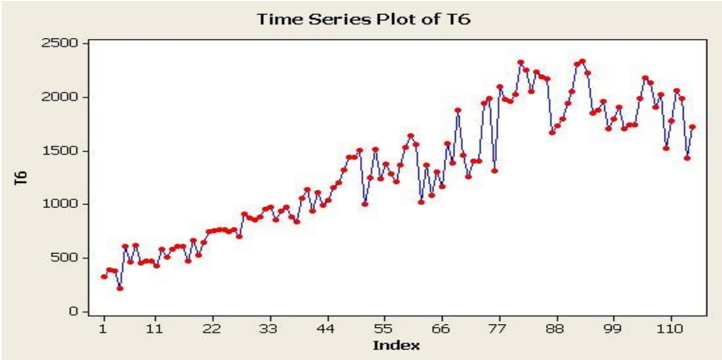
اگر یک سری زمانی یک بار تفاضلی شود و سری تفاضلی شده ایستا باشد، آنگاه سری زمانی اصلی انباشته از مرتبه اول نامیده می شود و بصورت $I(1)$ نشان داده می شود. در صورتی که یک سری زمانی پس از دو بار تفاضلگیری حالت ایستایی پیدا کند، انباشته از مرتبه دوم می باشد و آن را با $I(2)$ نشان می دهند.

بطور کلی اگر از یک سری زمانی d مرتبه تفاضل گرفته شود، آن سری انباشته از مرتبه d یا $I(d)$ است. بدین ترتیب هرگاه یک سری زمانی انباشته از مرتبه یک یا بالاتر باشد، آن سری نا ایستا است. بنابراین فرآیند $I(0)$ نشان دهنده یک فرآیند ایستا می باشد.

ناایستایی در واریانس

اگر به مرور زمان تغییر پذیری یک سری زمانی افزایش یابد بدین معنی است که سری مزبور نسبت به واریانسش ناایستا است. چنانچه یک سری زمانی در میانگین ایستا باشد ولی در واریانس ناایستا باشد باید بااستفاده از تبدیل مناسب اقدام به ایستا سازی واریانس نمود. ممکن است یک سری زمانی هم در میانگین و هم در واریانس ناایستا باشد، در این صورت ابتدا باید واریانس آن را ایستا نمود.

نمودار زیر یک سری زمانی را نشان می دهد که هم در میانگین و هم در واریانس ناایستا می باشد.برای ایستاسازی این سری همانطورکه قبلا گفته شد، ابتدا باید واریانس آنرا ایستا نمود. یعنی با استفاده از یک تبدیل مناسب داده ها را طوری تغییردادکه دارای واریانس ثابت باشند. بعد از تثبیت واریانس، با استفاده از تفاضلی کردن می توان داده ها را نسبت به میانگین نیز ایستا نمود.



شکل ۲– ۹ : نمودار یک سری زمانی ناایستا در میانگین و واریانس

تغییر واریانس یک فرآیند ناایستا، وقتی سطح آن تغییر می کند، بسیار متداول است. بنابراین برای ثابت c و تابع f داریم :

$$\operatorname{var}(x_t) = cf(\mu_t)$$

یعنی واریانس فرآیند تابعی از میانگین آن فرآیند می باشد.

جهت ایستاسازی واریانس باید یک تابع T را بگونه ای پیدا کنیم که سری تبدیل شده $T(x_t)$ یک واریانس ثابت داشته باشد. برای تشریح روش، تابع مورد نظر را با یک سری تیلور مرتبه اول در حول نقطه μ_t تقریب می زنیم.

فرض کنید:

$$T(x_t) \approx T(\mu_t) + T'(\mu_t)(x_t - \mu_t)$$

که $T'(\mu_t)$ مشتق اول $T(x_t)$ در μ_t است. اکنون داریم:

$$\operatorname{var}(T(x_t)) = [T'(\mu_t)]^2 \operatorname{var}(x_t) = c[T'(\mu_t)]^2 f(\mu_t)$$

بنابراین برای اینکه واریانس $T(x_t)$ ثابت شود، باید تبدیل پایداری واریانس $T(x_t)$ را چنان انتخاب کنیم که:

$$T'(\mu_t) = \frac{1}{\sqrt{f(\mu_t)}}$$

از معادله فوق نتیجه می شود:

$$T(\mu_t) = \int \frac{1}{\sqrt{f(\mu_t)}} d\mu_t$$

به عنوان مثال اگر انحراف معیار یک سری متناسب با سطح آن باشد بطوریکه $\mathrm{var}(x_t) = c^2 \mu_t^2$ آنگاه داریم :

$$T(\mu_t) = \int \frac{1}{\sqrt{\mu_t^2}} d\mu_t = \ln(\mu_t)$$

بنابراین یک تبدیل لگاریتمی(مبنای آن مهم نیست) بشکل $\ln(x_t)$ باعث ثابت شدن واریانس خواهد شد.

به عنوان مثال دیگرفرضکنید واریانس سری متناسب باسطح سری باشد، یعنی $\mathrm{var}(x_t) = c\mu_t$، آنگاه داریم:

$$T(\mu_t) = \int \frac{1}{\sqrt{\mu_t}} d\mu_t = 2\sqrt{\mu_t}$$

بنابراین یک تبدبل ریشه دوم سری بشکل $\sqrt{x_t}$ باعث ثابت شدن واریانس می شود.

تبدیلات باکس-کاکس

بطورکلی برای تبدیل واریانس،از تبدیل توانی زیرکه بوسیله باکس و کاکس (۱۹٦٤) معرفی شده است، استفاده می کنیم :

$$T(x_t) = x_t^{(\lambda)} = \frac{x_t^{\lambda} - 1}{\lambda}$$

λ را پارامتر تبدیل می نامند.

تبدیلات مربوط به چند مقدار λ که معمولا مورد استفاده قرار می گیرد، بصورت زیر می باشد. لازم به ذکر است که چنانچه مقدار λ برابر یک شود، نیازی به تبدیل نیست.

مقدار λ	تبدیل مناسب
−۱	$1/x_t$
−۰.۵	$1/\sqrt{x_t}$
۰	$\ln(x_t)$
۰.۵	$\sqrt{x_t}$
۱	نیازی به تبدیل نیست.

جدول (۲−۱) : جدول تبدیلات توانی باکس− کاکس

چند نکته:

۱− تبدیلات پایداری واریانس، فقط برای سریهای مثبت بکار می رود. با وجود این آنطور که به نظر می رسد، محدودیتی وجود ندارد. زیرا همیشه می توان مقدار ثابتی را به سری افزود، بدون اینکه ساختار همبستگی سری تغییر کند.

۲− اگر یک تبدیل پایداری واریانس لازم باشد، باید قبل از هر گونه تحلیلیمانند تفاضلی کردن، به اجرا درآید.

۳− غالبا تبدیل فقط برای پایداری واریانس نیست بلکه تقریب برای نرمال بودن را نیز بهتر می کند.

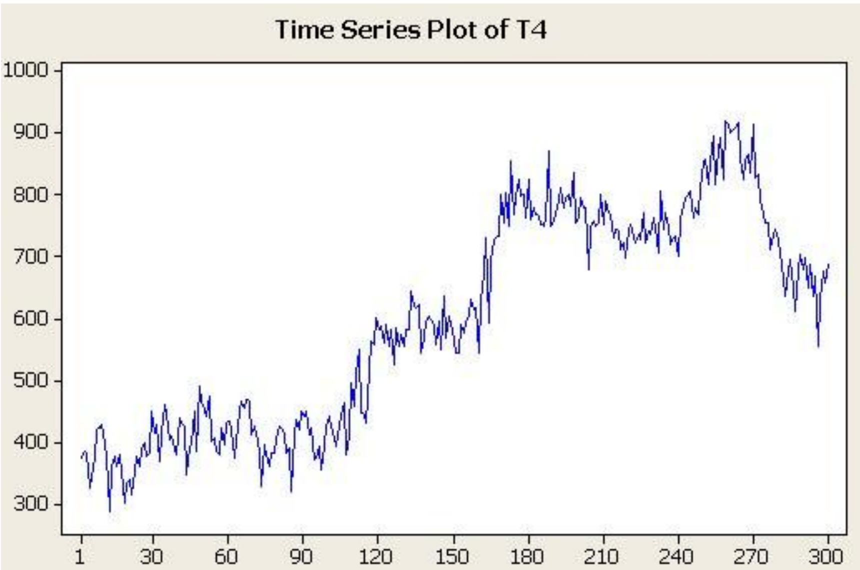
٤− در تبدیل توانی می توان λ را به عنوان یک پارامتر که از سری مشاهده شده برآورد می گردد، در نظر گرفت.برآورد درست نمائی ماکزیمم λ آن است که مجموع مربعات

باقیمانده را می نیمم می کند. برای هر مقدار λ مجموع مربعات از الگوی برازش شده، محاسبه می شود. برآوردرستنمائی ماکزیمم λ آناست که در بین مقادیر دیگر کوچکترین مجموع مربعات باقیمانده را می دهد.

مثال ۲-۲

برای روشن شدن مطلب، سری زمانی T٤ را که در پیوست این کتاب آمده است در نظر می گیریم و لزوم استفاده از تبدیلات توانی باکس-کاکس را برای تثبیت واریانس بررسی می کنیم. نمودار این سری به شکل زیر می باشد.

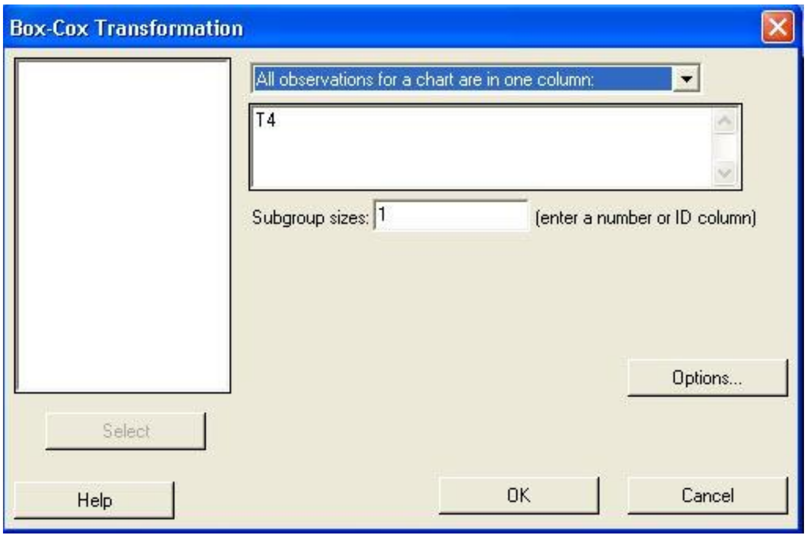
همانطور که ملاحظه می شود این سری در میانگین ناایستا است.یعنی سطح سری با گذشت زمان تغییر می کند. اما با توجه به نمودار سری می توان گفت که واریانس این سری تقریبا ثابت است و نیازی به تبدیل ندارد، یعنی تغییر پذیری سری با گذشت زمان تقریبا ثابت باقی می ماند.



شکل ۲ – ۱۰: نمودار سری زمانی T٤

برای بررسی بیشتر این مطلب از تبدیل باکس-کاکس استفاده می کنیم. چنانچه بهترین مقدار پیشنهادی برای λ (پارامتر تبدیل) عدد یک باشد، با توجه به جدول تبدیلات توانی، می توان گفت واریانس داده ها ثابت است و نیازی به تبدیل داده ها نداریم.

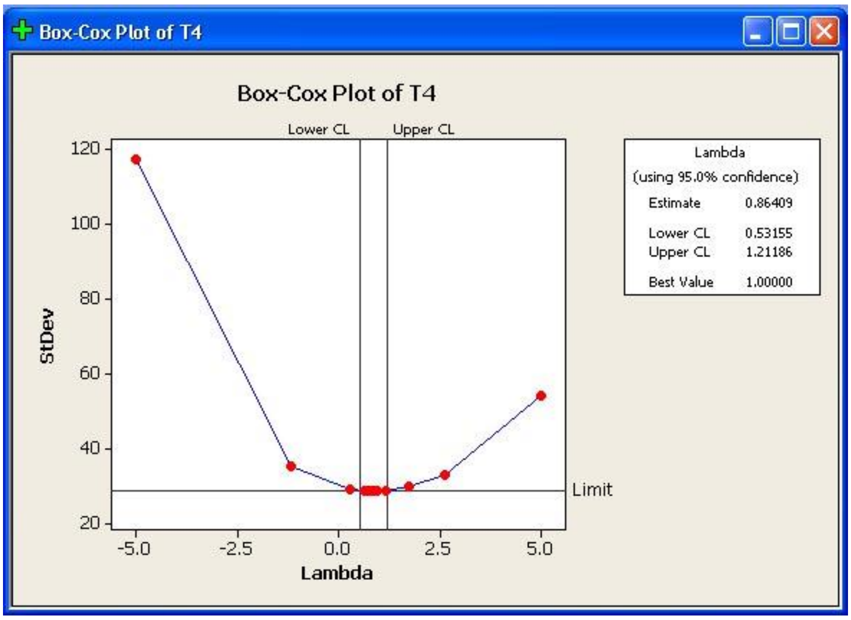
حال تبدیل باکس–کاکس را درمینی تب اجرا می کنیم. برای تثبیت واریانس با استفاده از تبدیلات باکس–کاکس در مینی تب، از منوی Stat گزینه Control Charts را انتخاب نموده و سپس از منوی ظاهر شده گزینه Box-Cox Transformation را انتخاب می کنیم تا پنجره زیر باز شود.



شکل ۲ – ۱۱ : پنجره تبدیلات توانی باکس–کاکس

در پنجره ظاهر شده در کادر بالای صفحه گزینه پیش فرض مینی تب را که می گوید، همه مشاهدات در یک ستون قرار دارند، می پذیریم. سپس با دو بار کلیک کردن برروی سری مورد نظر، آنرا به کادر وسط صفحه منتقل می کنیم.

در قسمت Subgroup sizes نیز عدد یک را وارد می کنیم(در مبحث کنترل کیفیت، حجم زیر گروه ها معمولا ٤ یا ٥ می باشد اما در سری های زمانی برای استفاده از تبدیلات باکس–کاکس، حجم زیر گروه را همیشه ۱ در نظر می گیریم). با فشردن دکمه ok نتیجه به شکل زیر خواهد بود:



شکل ۲ – ۱۲ : نمودار باکس–کاکس برای سری T٤

همانطور که ملاحظه می شود درکادر سمت راست برآورد λ وحدود اطمینان ۰.۹۵ وبهترین مقدارپیشنهادی برای λ، داده شده است. با توجه به اینکه بهترین مقدار پیشنهادی برای λ عدد یک می باشد و با توجه به جدول(۲-۱)،به این نتیجه می رسیم که هیچ تبدیلی برای پایایی واریانس این سری لازم نیست.

توجه:

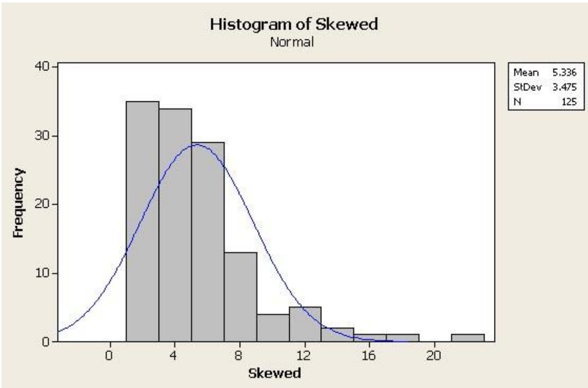
حدود اطمینان ۰.۹۵ شامل همه مقادیر λ می باشد که انحراف استانداردشان کمتر یا مساوی با مقداری است که خط افقی نشان می دهد. بنابراین هرمقداری از λ که انحراف استانداردی نزدیک به خط افقی داشته باشد، نیز یک مقدار قابل قبول و منطقی برای استفاده جهت تبدیل داده ها خواهد بود.

مثال۲- ۳

در کنترل کیفیت از تبدیل باکس–کاکس برای نزدیکتر کردن توزیع داده ها به توزیع نرمال یا به عبارتی نرمال تر کردن داده ها استفاده می شود. مثلا اگر توزیع داده ها بسیار

چوله باشدیاهنگامی که نوسانات وتغییرات داخل زیرگروهها ناپایدار باشد، از این رویه استفاده می شود.

برای روشن شدن موضوع، فایل BOXCOX.MTW را باز کنید.این فایل شامل۱۲۵ مشاهده می باشد که درستونی بانام Skewed ذخیره شده اند. همانطور که از هیستوگرام داده ها مشخص است، این مشاهدات بسیار چوله به راست می باشند. بنا براین تصمیم گرفته ایم برای نرمال تر کردن این مشاهدات از تبدیل استفاده کنیم.

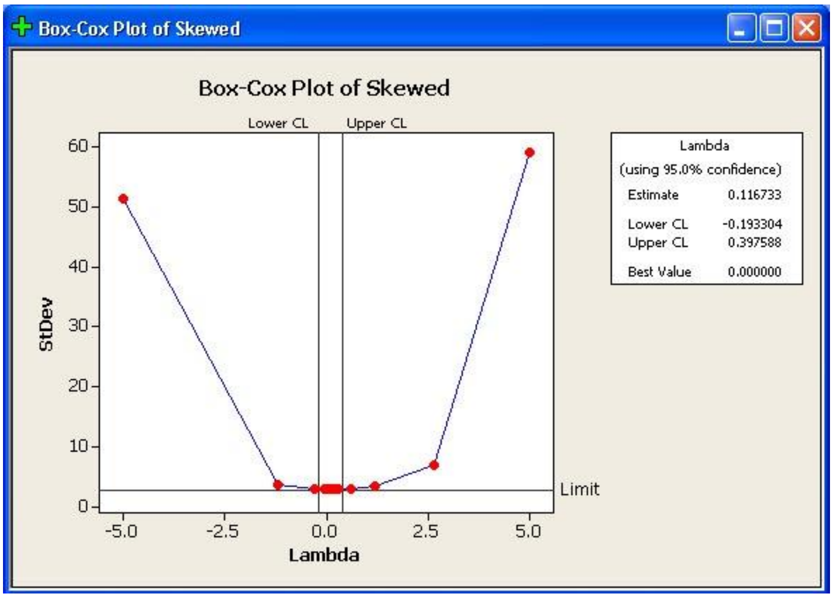


شکل ۲– ۱۳ : هیستوگرام داده ها ی مثال ۲– ۳ قبل از تبدیل

حال تبدیل باکس–کاکس را در مینی تب اجرا می کنیم. برای این کار با طی مراحل زیر پنجره باکس–کاکس را باز می کنیم.

Stat > Control Charts > Box-Cox Transformation

در پنجره ظاهر شده کادر اول را بدون تغییر رها می کنیم. سپس متغیرSkewed را به کادر وسط منتقل می کنیم.در قسمتSubgroup sizesنیز عدد ۵ را وارد می کنیم. زیرا در این طرح حجم زیر گروه ۵ بوده است.با فشردن ok نتیجه به شکل زیر خواهد بود:

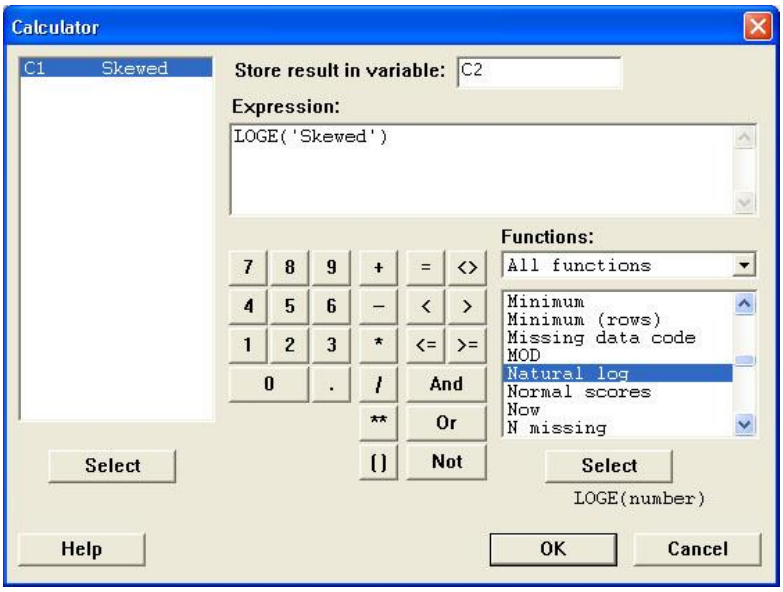


شکل ۲– ۱٤ : نمودار باکس–کاکس برایمتغییر Skewed

همانطور که ملاحظه می شود بهترین مقدار پیشنهادی برای λ عدد صفر می باشد و این به معنی این است که باید از یک تبدیل لگاریتمی استفاده نمود.

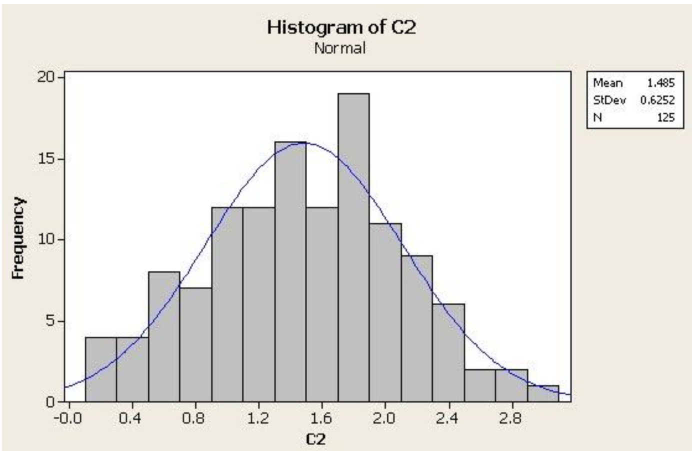
تبدیل لگاریتمی در MINITAB

برای تبدیل داده ها از منوی Calc گزینه Calculator را انتخاب می کنیم. سپس در قسمتFunctions تابع مورد نظر را که دراینجا تابع لگاریتم طبیعی(Natural log) می باشد برمی گزینیم. با دوبارکلیک کردن برروی تابع موردنظر،این تابع بهکادر Expression منتقل می شود. حال از کادر سمت چپ، بر روی متغیر Skewed که قرار است تبدیل شود، دو بار کلیک می کنیم تا تبدیل مورد نظر برروی این سری انجام شود. همچنین در کادر Store result in variable می توان یکی از ستونهای مینی تب را برای ذخیره کردن مقادیر تبدیل شده در نظر گرفت. ما در اینجا مقادیر تبدیل شده را در ستون C۲ ذخیره می کنیم.



شکل ۲ – ۱۵ : تبدیل لگاریتمی متغیر Skewed

حال هیستوگرام داده های تبدیل شده رارسم می کنیم تا ببینیم تبدیل انجام شده چه اثری بر روی توزیع داده ها داشته است و آیا توانسته است توزیع داده ها را به توزیع نرمال نزدیک کند یا نه.



شکل ۲ –۱٦ : هیستوگرام داده های مثال ۲–۳ پس از تبدیل

همانطور که ملاحظه می شود، تبدیل باکس-کاکس تأثیر زیادی بر نزدیک کردن توزیع داده ها به توزیع نرمال داشته است. بنابراین همانطور که قبلا هم گفته شد تبدیل علاوه بر پایدار کردن واریانس، تقریب برای نرمال بودن را نیز بهتر می کند.

۲-۳ همبستگی بین مشاهدات سری زمانی

نظریه آمار بیشتر در مورد نمونه های تصادفی که از مشاهدات مستقل ناشی شده اند بحث می کند. اما در سریهای زمانی ویژگی مهم این است که معمولا مشاهدات متوالی مستقل نیستند و دقیقا این وابستگی است که می خواهیم آن را بررسی کنیم و به مدل درآوریم.برای بررسی این وابستگی از تابع خود همبستگی و تابع خود همبستگی جزئی استفاده می کنیم.

تعریف ۲- ٤ (خود همبستگی در تأخیرK): عبارت است ازهمبستگی بین مشاهداتی که k واحد زمانی با یکدیگر فاصله دارند.تابع خود همبستگی نظری که آن را با ρ_k نشان می دهیم، به شکل زیر تعریف می شود :

$$\rho_k = \frac{\operatorname{cov}(x_t, x_{t+k})}{\operatorname{var}(x_t)} = \frac{\gamma(k)}{\gamma(0)}$$

γ_k را ضریب اتوکوواریانس در تأخیر k می نامیم. اندازه های ضریب اتوکوواریانس به واحد اندازه گیری x_t بستگی دارد.

ρ_k و γ_k هر دو وابستگی(خطی) بین متغیرهای تصادفی را اندازه می گیرند ولی تعبیر و تفسیر همبستگی بدون واحد تا اندازه ای آسان تر است.

برآورد ρ_k را که از یک نمونه n تایی بدست می آید، با r_k نشان می دهیم.ازضرایب خود همبستگی نمونه ای جهت تشخیص الگوی احتمالی مولد داده ها استفاده می شود.

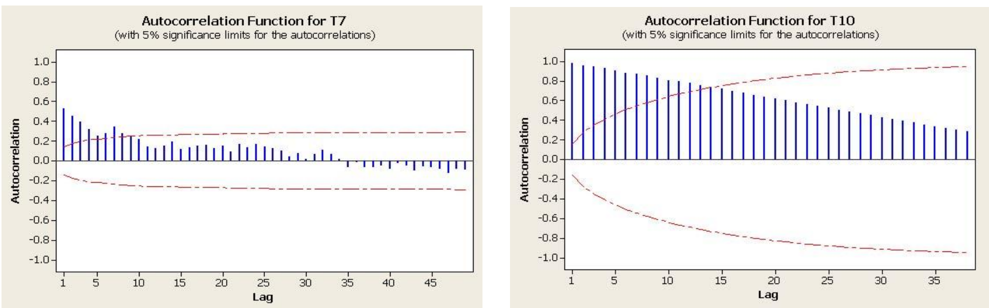
معمولاتابعخودهمبستگی رابا acf نشان می دهند که مخفف عبارت Autocorrelation function می باشد.

همبستگی نگار

نمودار r_k در مقابل تأخیر k را همبستگی نگار می نامند. از این نمودار می توان برای تشخیص الگوی احتمالی مولد داده ها استفاده کرد.

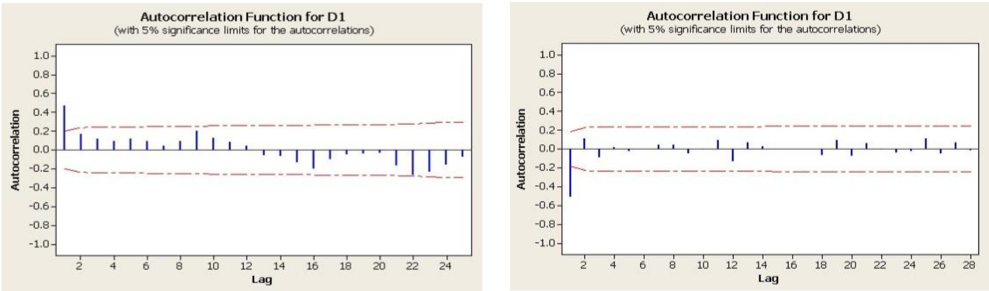
چند نکته:

- اگر یک سری زمانی کاملا تصادفی باشد، به ازای مقادیر بزرگ n، r_k تقریبا صفر خواهد بود.
- برای یک سری زمانی،متغیر تصادفی r_k تقریبا دارای توزیع نرمال بامیانگین صفروواریانس $1/n$می باشد.درنتیجه فاصله اطمینان۹۵درصدبرای r_k تقریبا بصورت$\left(\frac{-2}{\sqrt{n}},\frac{2}{\sqrt{n}}\right)$ می باشد.اگرمقدار مشاهده شده r_k درخارج ازاین حدود واقع شود،می گوییم این مقدار در سطح پنج درصد به طور معنی داری با صفر اختلاف دارد.
- همبستگی نگاری که درآن مقادیر r_k با سرعت معقولی به صفرنزدیک نمی شود، ناایستایی را نشان می دهد.



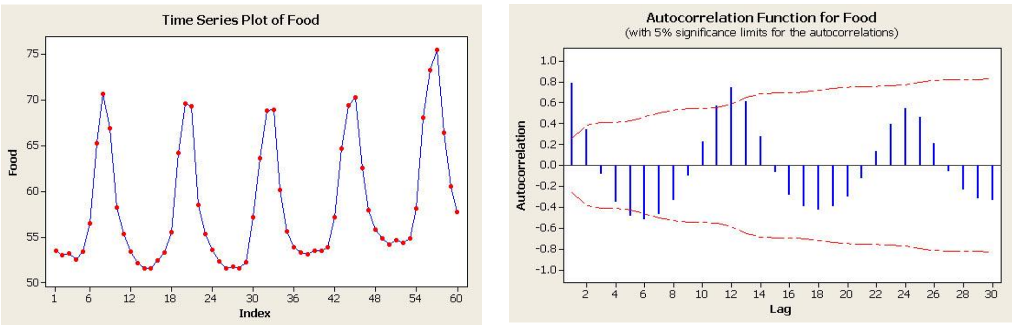
شکل ۲– ۱۷ : نمودار acfدو سری نا ایستا

- همبستگی نگاری که درآن مقادیر r_k نسبتاسریع قطع شوندیانسبتا سریع افول کنند، ایستایی رانشان می دهد.سری های ایستااغلب همبستگی های کوتاه مدتی را نشان می دهند.



شکل ۲ – ۱۸ : نمودار acfدو سری ایستا

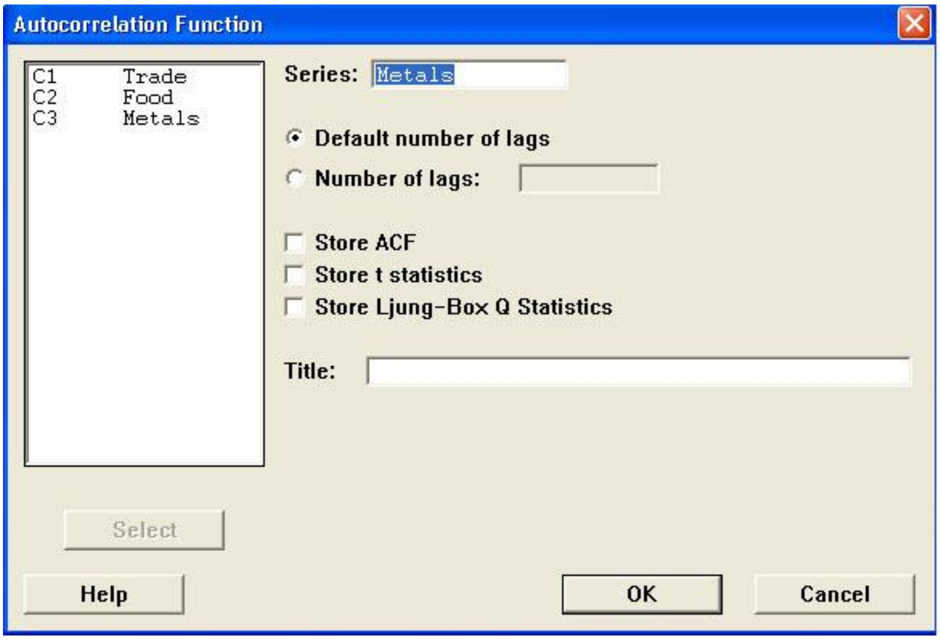
- اگر یک سری زمانی تغییرات فصلیداشته باشد،همبستگی نگارآن نیز نوساناتی درهمان فرکانسهارانشان می دهد، بویژه اگر x_t از یک طرح سینوسی پیروی نماید در آن صورت r_k نیز سینوسی می باشد.



شکل ۲ –۱۹ : نمودار یک سری زمانی با تغییرا ت فصلی وacfآن

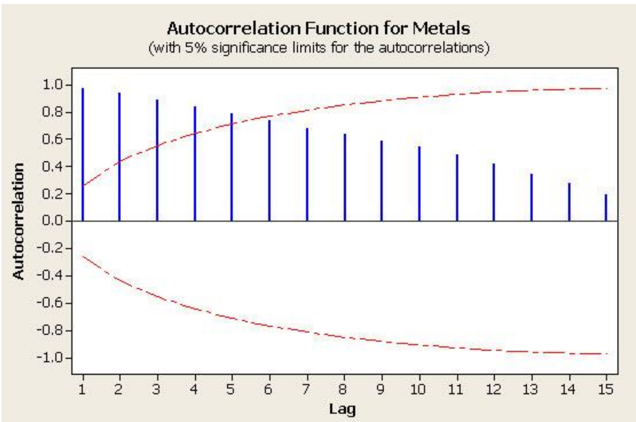
رسم همبستگی نگار درMinitab

در مینی تب برای رسمهمبستگی نگار یک سری زمانی،مانند سری Metals از فایـل EMPLOY.MTW،کافی استازمنوی StatگزینـهTimeSeriesوسپسـگزینه Autocorrelation راانتخاب کنیمتاپنجره ای به شکل زیر باز شود.برای رسم همبستگی نگـار سـری مـورد نظر پنجره باز شده را به شکل زیر تکمیل می کنیم.



شکل ۲– ۲۰ : پنجره مربوط به رسم تابع خود همبستگی

با رها کردن سایر گزینه ها و فشردن ok نمودار حاصل به شکل زیر خواهد بود :



شکل ۱ –۲۱ : همبستگی نگار سری زمانی Metals

همانطور که ملاحظه می شود مقادیر تابع خود همبستگی بسیار به کندی به صفر میل می کنند که مؤید ناایستایی سری مربوطه می باشد.

در حقیقت باید تابع خود همبستگی نمونه ای را برای سری زمانی ایستا محاسبه کنیم. بنابراین قبل از محاسبه acfباید هر گونه روند را حذف کرد.

اینک به اختصار گزینه های پنجره Autocorrelation Functionرا توضیح می دهیم.

گزینهDefault number of lags

چنانچه گزینهDefault number of lags را که انتخاب پیش فرض مینی تب نیز می باشد انتخاب کنید، تعداد تأخیرهایی که تابع خود همبستگی برای آنها محاسبه می شود بصورت زیر خواهد بود:

تعداد مشاهدات (n)	تعداد تأخیرهایی که تابع خود همبستگی برای آنها محاسبه خواهد شد. (k)
کمتر مساوی ۲٤٠	$n/4$
بیشتر از ۲٤٠	$\sqrt{n}+45$

جدول (۲-۲)

گزینهNumber of lags

با انتخاب این گزینه می توان acf را برای تعداد دلخواهی از تأخیرات رسم کرد. مثلا چنانچه عدد۱٠ را در کادر مربوط به این گزینه وارد کنیم، همبستگی بین مشاهداتی که حداکثر ده واحد زمانی با یکدیگر فاصله دارند،یعنی همبستگی بین مشاهدات متوالی، مشاهداتی که دو واحد زمان بایکدیگر فاصله دارند،... و نهایتا همبستگی بین مشاهداتی که ده واحد زمانی با یکدیگر فاصله دارند، محاسبه خواهد شد.

در نمودار acf علاوه بر محاسبه و رسم مقادیر تابع خود همبستگی، حدود اطمینان ۹۵ درصد نیز رسم خواهند شد.

گزینهStore t-statistics

این گزینه باعث می شود که مقادیر آماره tدر یکی از ستونهای مینی تب ذخیره شود. از آماره tبرای آزمون این فرض که آیا خود همبستگی مشاهدات در یک تأخیر دلخواه صفرهست یا نه استفاده می شود. این آزمون را آزمون tمی نامند. نحوه قضاوت در مورد این آزمون بصورت زیر است:

برای تأخیرهای یک تا سه، چنانچه مقدارآماره tبزرگتر از قدر مطلق(۱.۲۵)باشد، مقدار تابع خود همبستگی در آن تأخیر مخالف صفر در نظر گرفته می شود.

برای تأخیرهای بزرگتر ازچهار،چنانچه مقدارآماره tبزرگتر ازقدر مطلق (۲) باشد، مقدار تابع خود همبستگی درآن تأخیر مخالف صفردرنظرگرفته می شود.

ازاین آماره درتشخیص وارزیابی مدلهایARIMA وهمچنین دربررسی خود همبستگی باقیمانده ها استفاده می شود.

گزینهStore Ljung-Box Q statistics

با انتخاب این گزینه می توان مقادیر آماره(Ljung-BoxQ (LBQرا در یکی از ستونهای مینی تب ذخیره کرد.ازاین آماره برای آزمون این فرض که آیا همه خود همبستگی ها تا تأخیر k مساوی با صفر هستند یا نه، استفاده می شود.

فرضیه H_0 این آزمون بصورت زیر می باشد :

$$H_0 : \rho_1 = \rho_2 = ... = \rho_k = 0$$

آماره LBQ، که آن را **آماره Q اصلاح شده**نیز می نامند، به شکل زیر است :

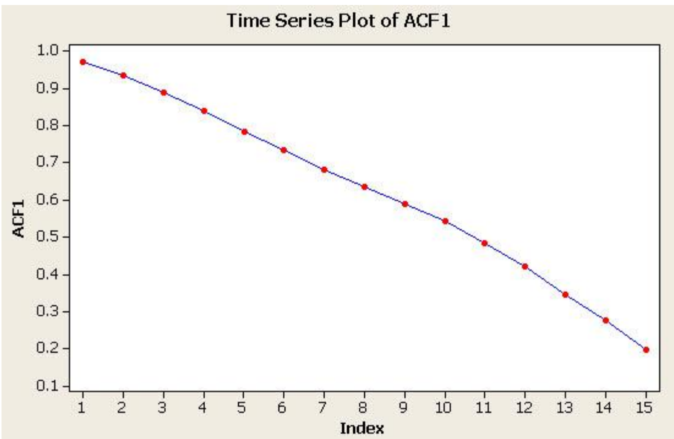
$$Q = n(n+2)\sum_{h=1}^{k}(n-h)^{-1}\hat{\rho}_h^2$$

که در آن n تعداد مشاهدات می باشد. این آماره تحت فرض H_0 تقریبا دارای توزیع χ^2_{k-m} است که m تعداد پارامترهای برآوردشده در مدل می باشدوفرضیه H_0 رد می شود هرگاه مقدار آماره Q از مقدار متناظر جدول کی دو بیشتر باشد.

گزینهStore ACF

با انتخاب این گزینه مقادیرتابع خود همبستگی در یکی از ستونهای مینی تب ذخیره می شود. نام این ستون به طور پیش فرضACF۱ خواهد بود.می توان نمودار سری زمانی را برای این متغیر نیز رسم کرد.

مثلادرسریMetals چنانچه گزینهStoreACF رامارکدارکنیم،مقادیرتابع خود همبستگی درستونی به نامACF۱ ذخیره خواهند شد.حال چنانچه مجددا نمودارسری زمانی را برای متغیر ACF۱رسم کنیم، نموداری بصورت زیر خواهیم داشت:



شکل ۲-۲۲ : نمودار سری زمانی برای مقادیر تابع خود همبستگی

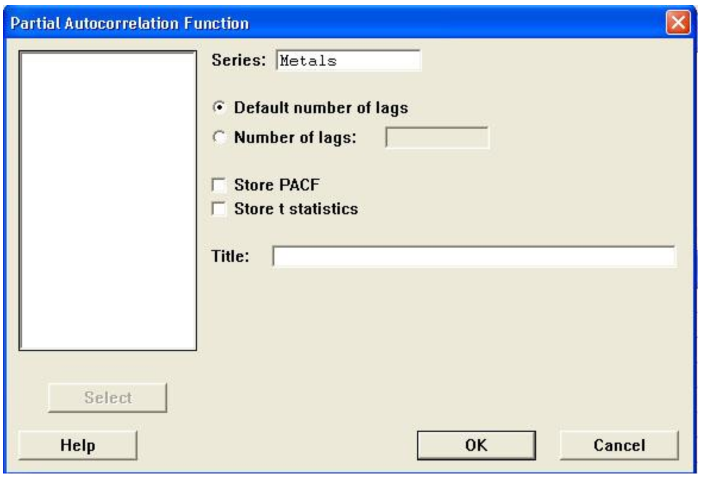
همانطور که از این نمودار نیز مشخص است مقادیر تابع خود همبستگی به کندی به صفر میل می کنند که این دلیلی بر ناایستایی سری Metals می باشد.

خود همبستگی جزئی

تعریف ۲-۵ (ضریب خود همبستگی جزئی): همبستگی بین x_t و x_{t+k} بعد از حذف اثر متغیرهای $x_{t+1}, x_{t+2}, \ldots, x_{t+k-1}$ را ضریب خود همبستگی جزئی می نامند. این خود همبستگی را با φ_{kk} نشان می دهیم. از این تابع که آن را با pacf نیز نشان می دهند در تشخیص الگوی احتمالی مولد داده ها استفاده می شود.

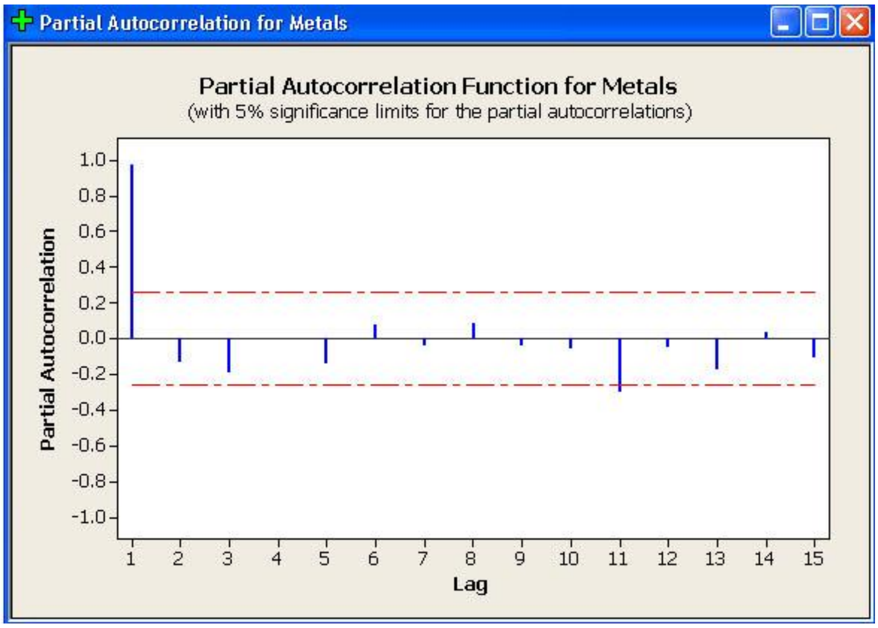
رسم تابع خود همبستگی جزئی در Minitab

در مینی تب برای رسم تابع خود همبستگی جزئی یک سری زمانی مانند سری Metals از فایل EMPLOY.MTW، کافی است ازمنوی Stat گزینه Time Series و سپس گزینه Partial Autocorrelation را انتخاب کنیم تا پنجره ای به شکل زیر بازشود. پنجره باز شده را به شکل زیر تکمیل می کنیم.



شکل ۲- ۲۳ : پنجره pacf

با فشردن دکمه ok نتیجه به شکل زیر خواهد بود:



شکل ۲–۲٤ : pacf برای سری Metals

فصل سوم

مدل های احتمال برای سری های زمانی

آمار : علم مدل سازی پدیده های تصادفی.

مقدمه

همانطور که قبلا هم گفته شد در سری های زمانی سعی می کنیم با بررسی گذشته سری، الگوی احتمالی مولد داده ها را شناسایی کرده و بر مبنای این الگودرباره رفتار آینده سری اظهار نظر نماییم.

مدل احتمالی ای که به سری برازش داده می شود باید بتواند به نحو مناسبی مشاهدات سری را مدل سازی کند. برای بررسی اینکه آیا یک مدل احتمالی واقعا توصیف کننده داده ها است یا خیر، می توان خطاهای پیش بینی را موردتجزیه و تحلیل قرار داد. اگر مدلی به نحو رضایتبخشی نماینده فرآیند باشد، آنگاه انتظار می رودمقدار متوسط خطاهای پیش بینی نزدیک صفر باشد.

مدل های احتمال برای سری های زمانی در واقع **فرآیندهای تصادفی** هستند. یک فرآیند تصادفی یک پدیده آماری است که در طول زمان تکامل پیدا می کند. به عبارت دیگریک فرآیندتصادفی مجموعه متغییرهای تصادفی است که بر حسب زمان مرتب شده باشند.

اغلب در تجزیه و تحلیل سریهای زمانی، ساختن بیش از یک مشاهده در زمـان داده شده امکان پذیر نیست. معهذا می توانیم سری مشاهده شده را دقیقا بصورت مثالی

از مجموعه نامتناهی سری های زمانی که ممکن است مشاهده شونددر نظر بگیـریم. هر عضو این مجموعه یک مصداق بخصوص است.

تجزیه و تحلیل سریهای زمانی اصولا با ارزیابی کردن خواص مدل احتمالی که سری مشاهده شده را تولید می نماید سروکاردارد. بنابراین ابتدا بایداین مدل های احتمالی را بخوبی شناخت.اکنون مهمترین فرآیندهای تصادفی که در مبحث سری های زمانی کاربرد دارندرابه اختصار معرفی می کنیم.

۳-۱فرآیند تصادفی محض

فرآیندگسسته $\{z_t\}$ یک فرآیند تصادفی محض یا اغتشاش خالص نامیده می شود، اگر متغیرهای تصادفی $\{z_t\}$ بصورت دنباله ای از متغیرهای دو به دو مستقل و هم توزیع باشند.

این فرآیند که مهندسین گاهی آنرا white noiseیا سروصدای سفید می نامند، یک فرآیند ایستا است. زیرا میانگین و تابع اتوکوواریانس آن به زمان بستگی ندارد. فرآیند تصادفی محض به تنهایی کاربرد ندارد و از آن بعنوان جزء تشکیل دهنده فرآیندهای پیچیده تر استفاده می شود.

۳-۲ فرآیند خطی کلی

یک فرآیند خطی کلی را می توان بصورت ترکیب خطی وزن داری از مشاهدات حال و گذشته فرآیند تصادفی محض نشان داد.

$$x_t - \mu = \sum_{j=0}^{\infty} \beta_j z_{t-j} \tag{۱-۳}$$

در این معادله μ بیانگر سطح فرآیند و $\{z_t\}$بیانگر فرآیند تصادفی محض می باشد.

معادله (۳-۱) را یک فیلتر خطی می نامند. در واقع فرآیند خطی کلی با عبور دادن یک فرآیند تصادفی محض از یک فیلتر خطی بدست می آید.

واضح است که مشاهدات متوالی در سری زمانی $\{x_t\}$ وابسته اند. زیرا آنها از دریافتهای قبلی یکسان $\{z_t\}$ نتیجه می شوند.

مدلهای سری زمانی که از فیلتر خطی ناشی شده اند، مدلهای باکس-جنکینزنامیده می شوند. این مدلها توانایی معرفی سری های زمانی اعم از ایستا و ناایستا را دارند. از جمله مهمترین مدلهای مشتق شده از فرآیند خطی کلی مدلهای اتورگرسیو و میانگین متحرک می باشند.

در یک تقسیم بندی کلی سری های زمانی را به دو رده فصلی و غیر فصلی تقسیم می کنیم. ما در این قسمت ابتدا مدلهای غیرفصلی(ایستا و ناایستا)را مطالعه می کنیم و سپس مدلهای فصلی را بررسی خواهیم کرد.

۳-۳ مدلهای غیر فصلی

۳-۳-۱ مدلهای غیر فصلی ایستا

مدلهایی که در زیر مورد بحث قرار می گیرند، برای مدل سازی سری های غیر فصلی ایستا استفاده می شوند. این مدلها عبارتند از مدل اتو رگرسیو، مدل میانگین متحرک و مدل مرکب اتورگرسیو- میانگین متحرک.

۱- مدل اتورگرسیومرتبهP یا AR(P)

این مدل بصورت زیر تعریف می شود:

$$x_t = \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + ... + \alpha_p x_{t-p} + z_t \qquad (۲-۳)$$

که $\{z_t\}$فرآیند تصادفی محض بامیانگین صفر و واریانس σ_z^2 است.

همانطور که ملاحظه می شود x_t ترکیبی خطی از جدیدترین p مقدار گذشته خودش بعلاوه یک جمله اغتشاش z_tاست که هر چیز تازه ای در زمان t که بوسیله مقادیر گذشته بیان نشده است را در سری منظور می کند.بنابراین فرض می کنیم z_t مستقل از $x_{t-1}, x_{t-2},...$ است.

فرآیند AR(P) را می توان بااستفاده از عملگر پسرو به شکل ساده تر زیر نوشت:

$$\varphi(B)x_t = z_t \quad (۳-۳)$$

$$\varphi(B) = 1 - \alpha_1 B - \alpha_2 B^2 - .... - \alpha_P B^P$$

چنانچه میانگین فرآیند مخالف صفر باشد، آن را به شکل زیر می نویسیم:

$$\varphi(B)(x_t - \mu) = z_t$$

معمولا فرآیند AR با میانگین غیرصفر را بصورت زیر می نویسیم:

$$\varphi(B)x_t = \theta_0 + z_t \qquad (٤-۳)$$

که θ_0 را می توان به شکل زیر تعریف کرد:

$$\theta_0 = (1-\alpha_1 B - ... - \alpha_p B^p)\mu = (1-\alpha_1 - ... - \alpha_p)\mu$$

چند نکته

۱- به لحاظ نظری شرط ایستایی فرآیند AR(p)آن است که ریشه های معادله مشخصه آن یعنی ریشه های چند جمله ای $\varphi(B)$ از لحاظ قدر مطلق بزرگتر از یک باشند(خارج از دایره واحد قرار داشته باشند). بنابراین ایستایی این فرآیند بستگی به ضرایب α دارد.

۲- فرآیند اتورگرسیو تقریبا مانند مدل رگرسیون چندگانه است، با این تفاوت که x_t روی متغیرهای مستقل رگرسیون نشده استبلکه روی مقادیر گذشته خودش رگرسیون شده است. به همین جهت این فرآیند را اتورگرسیو می نامند.

۳- یکی از ویژگیهای فرآیند AR(P) این است که pacf آن بعداز فاصله p قطع می شود. یعنی تابع خود همبستگی جزیی آن بلافاصله پس از تأخیر p صفر می شود.

٤- به طورکلی در فرآیندهای AR فرض بر این است که مقدار حال یک سری زمانی به مقادیر بلافاصله گذشته آن همراه با یک خطای تصادفی بستگی دارد.

مدل اتورگرسیو مرتبه اول (فرآیند مارکف)

فرآیند اتورگرسیو مرتبه اول که به شکل $x_t = \alpha x_{t-1} + z_t$ است را فرآیند مارکف می نامند. این فرآیند ایستا است اگر $|\alpha| < 1$؛ این شرط هم ارز است با اینکه بگوییم ریشه معادله $(1-\alpha B) = 0$ خارج از دایره واحد واقع شود.

فرآیند اتورگرسیو مرتبه اول را می توانیم با جایگزینی های متوالی بصورت زیر بنویسیم :

$$x_t = \alpha(\alpha x_{t-2} + z_{t-1}) + z_t = \alpha^2(\alpha x_{t-3} + z_{t-2}) + \alpha z_{t-1} + z_t$$

با ادامه این روند، می توان x_t را بصورت یک فرآیند MA از مرتبه نامتناهی به شکل زیر بیان کرد. فرآیند MAدر ادامه همین فصل معرفی خواهد شد.

$$x_t = z_t + \alpha z_{t-1} + \alpha^2 z_{t-2} + \ldots$$

با توجه به آنکه $\{z_t\}$ یک فرآیند تصادفی محض با میانگین صفر و واریانس σ_z^2 می باشد، لذا میانگین و واریانس فرآیند AR(۱)بصورت زیر خواهد بود:

$$E(x_t) = 0$$

$$\mathrm{var}(x_t) = \sigma_z^2(1 + \alpha^2 + \alpha^4 + \ldots)$$

بنابراین اگر $|\alpha| < 1$، واریانس متناهی خواهد بود و داریم :

$$\mathrm{var}(x_t) = \frac{\sigma_z^2}{(1-\alpha^2)} = \sigma_x^2$$

acf این فرآیند بصورت زیر تعریف می شود :

$$\rho(k) = \alpha^k$$

$$k = 0,1,2,..$$

در این فرآیند چنانچه $0 < \alpha < 1$، تمام همبستگی ها مثبت خواهند بود. اما چنانچه $-1 < \alpha < 0$ همبستگی در تأخیر یک منفی است و سایر خود همبستگیها به تناوب از مثبت به منفی تغییر می کنند.

چنانچه مقدار α به ± 1 نزدیک باشد تنزل نمائی تابع خود همبستگی کاملا کند خواهد بود. ولی برای αهای کوچک این تنزل سریع است. وقتی α به ± 1 نزدیک باشد، شدت همبستگی در طول تأخیرهای زیادی ادامه پیدا می کند. به ازای α های مثبت

سری نسبتا هموار خواهد بود ولی برای αهای منفی یک سری دندانه دار خواهیم داشت.

۲- مدل گام برداری تصادفی

در فرآیند اتورگرسیو مرتبه اول به ازای $\alpha = 1$ فرآیند جدیدی بدست می آید که آن را فرآیند گام برداری تصادفی می نامند. این فرآیند که به شکل $x_t = x_{t-1} + z_t$ است بیان می کند که مشاهده در هر مرحله فقط به مشاهده در یک مرحله قبل بعلاوه یک خطای تصادفی بستگی دارد.

z_t را که همان فرآیند اغتشاش خالص می باشد، به عنوان اندازه گام هایی که در زمان t به عقب یا جلو برداشته ایم تعبیر می کنیم. x_t نیز موقعیت گام بردارنده تصادفی در زمان t می باشد. قیمت سهام، مثالی از سری های زمانی است که رفتارش بسیار شبیه گام برداری تصادفی می باشد.

خطای تصادفی+ قیمت آن سهم در روز (t-۱) = قیمت یک سهم در روز t

فرآیندگام برداری تصادفی یک فرآیند ناایستا است. اما تفاضلهای مرتبه اول آن یک فرآیند تصادفی محض به شکل $z_t = x_t - x_{t-1}$ را می سازد که ایستا می باشد.

۳- مدل میانگین متحرک مرتبهq یا MA(q)

این مدل بصورت زیر تعریف می شود:

$$x_t = z_t + \beta_1 z_{t-1} + ... + \beta_q z_{t-q} \qquad (۵-۳)$$

که $\{z_t\}$ فرآیند تصادفی محض با میانگین صفر و واریانس σ_z^2 است.

این فرآیند را بااستفاده از عملگر پسرو می توان به شکل ساده تر زیر نوشت:

$$x_t = \theta(B) z_t \qquad (۶-۳)$$

$$\theta(B) = 1 + \beta_1 B + \beta_2 B^2 + ... + \beta_q B^q$$

چنانچه میانگین این فرآیندمخالف صفر باشدمی توان جمله ثابت μ را به طرف راست معادلات فوق اضافه کرد.

$$x_t = \mu + \theta(B) z_t$$

یا

$$x_t = \theta_0 + \theta(B) z_t \qquad (۷-۳)$$

که $\theta_0 = \mu$

چند نکته

۱- این فرآیند بدون توجه به مقادیر وزنهای { β_i }همواره ایستا است.

۲- تابع خود همبستگی فرآیند MA(q) بعد از تأخیر qقطع می شود. یعنی تابع خودهمبستگی آن برای مقادیر بزرگتر از qصفر خواهد بود.

۳- یک فرآیند MA از مرتبه متناهی را می توان به یک فرآیند AR از مرتبه نامتناهی تبدیل کرد، در صورتی که ریشه های چندجمله ای $\theta(B)$ خارج از دایره واحد باشند. همینطور می توان یک فرآیند AR از مرتبه متناهی را بدون هیچ شرطی به یک فرآیند MA از مرتبه نامتناهی تبدیل کرد.

٤- شرط تبدیل پذیری (وارون پذیری) برای فرآیند MA(q) معادل است با شرط ایستایی برای فرآیند AR(p) . بنابراین فرآیند MA(q) وارون پذیر است اگر و فقط اگر ریشه های معادله مشخصهآن از لحاظ قدر مطلق بیشتر از یک باشند.

۵- یک فرآیند MA را نمی توانیم بطور منحصر بفردی از روی acf آن مشخص کنیم، مگر اینکه آن فرآیند وارون پذیر باشد. در واقع گنجاندن شرط وارون پذیری ما را مطمئن می سازد که برای هر acf معلومی یک فرآیند میانگین متحرک یکتا وجود دارد.

٦- فرآیندهای میانگین متحرک در توصیف پدیده هایی مفید هستند که در آنها، پیشامدها یک اثر آنی را تولید می کنند که برای دوره های زمانی کوتاه باقی می ماند.

۷- فرآیندهای اتورگرسیو و میانگین متحرک تاحدودی معادلند و انتظار می رود که وقتی مدلی با مرتبه پایین از یک نوع به حد کافی یک سری را تشریح می کند، مدل نوع دیگر با مرتبه بالاتر نیز چنین باشد. با وجود این، ما مایلیم با توجه به اصل امساک(صرفه جویی در پارامترها) مدل دارای پارامترهای کمتر را انتخاب کنیم.

٤- مدل مرکب یا ARMA(p,q)

فرآیندهای اتورگرسیو- میانگین متحرک یا ARMA(p,q)که آنها را فرآیندهای مرکب نیز می نامیم، از ترکیب دو فرآیند پیشین بدست می آید. این فرآیند شامل p جمله AR و q جمله MA می باشد و بصورت زیر نوشته می شود:

$$(1-\alpha_1 B-......-\alpha_p B^P)(x_t-\mu)=(1+\beta_1 B+.......+\beta_q B^q)z_t \qquad (۸-۳)$$

یا

$$\varphi(B)(x_t-\mu)=\theta(B)z_t \qquad (۹-۳)$$

که $\varphi(B)$ و $\theta(B)$ بصورت زیر تعریف می شوند :

$$\varphi(B)=1-\alpha_1 B-\alpha_2 B^2-....-\alpha_P B^P$$

$$\theta(B)=1+\beta_1 B+\beta_2 B^2+....+\beta_q B^q$$

مدل ARMA(p,q) را بصورت زیر نیز می توان نوشت:

$$\varphi(B)x_t = \theta_0 + \theta(B)z_t \qquad (۱۰-۳)$$

$$x_t = \theta_0 + \alpha_1 x_{t-1} + ... + \alpha_p x_{t-p} + z_t + \beta_1 z_{t-1} + ... + \beta_q z_{t-q}$$

که در آن :

$$\theta_0 = (1-\alpha_1 B - ... - \alpha_p B^p)\mu = (1-\alpha_1 - ... - \alpha_p)\mu$$

به منظور شناسائی آزمایشی این مدل باید موارد زیر را تعیین کرد:

۱- دخالت یا عدم دخالت جمله ثابت θ_0 در مدل.

۲- کدام یک از عمل کننده های $\varphi(B)$ و $\theta(B)$ باید در مدل دخالت داده شوند. یعنی در یک مدل خاص باید عمل کننده میانگین متحرک غیر فصلی از مرتبه q را به کار برد یا عمل کننده اتورگرسیو غیر فصلی از مرتبه p، یا هر دو و یا هیچ کدام.

۳- تعیین مرتبه هر یک از عمل کننده هایی که در مدل دخالت داده می شوند، یعنی تعیین p و q.

جمله ثابت θ_0

عبارت ثابت θ_0 ممکن است صفر یا مخالف صفر باشد. با توجه به آنکه $\theta_0 = \mu\varphi(B)$، با فرض آنکه μ میانگین واقعی سری زمانی مورد نظر باشد و $\overline{X}$ برآورد نقطه ای μ باشد آنگاه می توان گفت:

- اگر $\overline{X}$ از نظر آماری مخالف صفر باشد آنگاه فرض می کنیم μ نیز مخالف صفر است. در نتیجه θ_0 مخالف صفر خواهد بود و باید آنرا در مدل لحاظ کرد.
- اگر $\overline{X}$ از نظر آماری تفاوت معنی داری با صفر نداشته باشد آنگاه فرض می کنیم μ نیز صفر است. در نتیجه θ_0 نیز صفر خواهد بود.

چند نکته

۱- شرط ایستایی فرآیند مرکب ARMA آن است که مقادیر $\{\alpha_i\}$ به گونه ای باشند که ریشه های $\varphi(B)=0$ خارج دایره واحد واقع شوند.

۲- شرط وارون پذیری این فرآیند آن است که مقادیر $\{\beta_i\}$ به گونه ای باشند که ریشه های $\theta(B)=0$ خارج دایره واحد واقع شوند.

۳- اغلب می توان یک سری زمانی ایستا را با یک مدلARMA بیان نمود که نسبت به مدلهای MA یا AR(به تنهایی)، پارامترهای کمتری دارد.

٤- برای فرآیند ARMA(p,q) ، q تأخیر اول تابع خود همبستگی تحت تأثیر پارامترهای میانگین متحرک و تأخیرهای بزرگتراز q تنها تحت تأثیر پارامترهای اتورگرسیو خواهند بود. بعلاوه تابع خود همبستگی یک فرآیند مرکب بعد از نخستین $q-p$ فاصله، بصورت مخلوطی از جملات نمائی و جملات با میرائی سینوسی می باشد. بالعکس در تابع خود همبستگی جزئی بعد از نخستین $p-q$ فاصله غلبه با مخلوطی از جملات نمائی و جملات با میرائی سینوسی است.

۵- اگر $q-p<0$، تابع خود همبستگی شامل مخلوطی از میرائی های نمائی و یا میرائی های سینوسی خواهد بود که طبیعت آن به وسیله چند جمله ای $\varphi(B)$ و مقادیر آغازین تعیین می شود.

اگر $q-p \geq 0$ به تعداد $q-p+1$ مقدار اولیه تابع خود همبستگی وجود خواهند داشت که از این مدل عمومی پیروی نمی کنند. این حقایق در تعیین سری های مرکب مفید خواهد بود.

۶- فرآیند ایستا و وارون پذیر ARMA را می توان بصورت یک فرآیند اتورگرسیو محض به شکل زیر بیان کرد:

$$\pi(B)x_t = z_t$$

که در آن

$$\pi(B) = \frac{\varphi_p(B)}{\theta_q(B)} = (1-\pi_1 B - \pi_2 B^2 - ....).$$

۷- همچنین این فرآیند را می توان بصورت یک فرآیند میانگین متحرک محض به شکل زیر بیان کرد:

$$x_t = \psi(B)z_t$$

که در آن

$$\psi(B) = \frac{\theta_q(B)}{\varphi_p(B)} = (1+\psi_1 B + \psi_2 B^2 + ...).$$

در جدول (۳-۱) خلاصه ای از ویژگیهای فرآیند هایی که تاکنون مورد بررسی قرار داده ایم آمده است.

جدول (۳-۱) خلاصه ای از خواص فرآیندهای AR ، MAوARMA

فرآیند مرکب	فرآیند میانگین متحرک	فرآیند اتورگرسیو	
$\theta^{-1}(B)\varphi(B)x_t = z_t$	$\theta^{-1}(B)x_t = z_t$	$\varphi(B)x_t = z_t$	مدل بر حسب x های پیشین
$x_t = \varphi^{-1}(B)\theta(B)z_t$	$x_t = \theta(B)z_t$	$x_t = \varphi^{-1}(B)z_t$	مدل بر حسب z های پیشین
ریشه های $\varphi(B)$ خارج دایره واحد واقع شوند.	همواره ایستا است.	ریشههای $\varphi(B)$ خارج دایره واحد واقع شوند.	شرط ایستایی
ریشه های $\theta(B)$ خارج دایره واحد واقع شوند.	ریشه های $\theta(B)$ خارج دایره واحد واقع شوند.	همواره وارون پذیر است.	شرط وارون پذیری
نامحدود (با میرائی نمائی ویا میرائی سینوسی بعد از q-p فاصله اول)	محدود (بعد از تأخیر q قطع می شود)	نامحدود (میرائی نمائی یا میرائی سینوسی)	تابع خود همبستگی

نامحدود (غلبه با میرائی نمائی ویا میرائی سینوسی بعد از p-q تأخیر اول)	نامحدود (غلبه با میرائی نمائیویا میرائی سینوسی)	محدود (بعد از تأخیر p قطع می شود.)	تابع خود همبستگی جزئی

۳-۳-۲مدل غیر فصلی نااایستا

بسیاری از سری های زمانی که در عمل مشاهده می شوند دارای رفتاری هستند که گوئی میانگین ثابتی ندارند. حتی در این موارد صرفنظر از سطح محلی سری یا سطح محلی و روند آن، ممکن است نوعی همگنی در آنها مشاهده شود. به این معنی که یک قسمت از سری مانند قسمتهای دیگر آن رفتار کند. مدلهایی که چنین رفتار همگن ناایستایی را نشان می دهند، به این ترتیب ممکن است بدست آیند که فرض می شود پس از تفاضل گیری تا مرتبه مناسبی، سری ایستا می شود.

در این قسمت ابتدا مفهوم روند قطعی و روند تصادفی را بررسی می کنیم و سپس به معرفی یک کلاس مهم از مدل های احتمال برای سری های ناایستا می پردازیم. این رده از مدلها که تفاضل مرتبه dام آنها به یک فرآیند ایستا از نوع ARMA منجر می شود را فرآیندهای میانگین متحرک تلفیق شده با اتورگرسیو یا ARIMA می نامند.

روند قطعی و روند تصادفی

روند قطعی و روند تصادفی هر دو در به مدل در آوردن سری های زمانی ناایستا در میانگین مفیدند. بطور کلی روند در صورتی که کاملا قابل پیش بینی بوده و متغیر نباشد، قطعی است.

روندهای تصادفی به وسیله تغییرات تصادفی که در سطح و شیب سری رخ می دهد مشخص می شوند. ولی بطور کلی ممکن است بخواهیم یک تابع قطعی از زمان

مانند $f(t)$ را در مدل بگنجانیم. بخصوص منظور کردن یک روند چند جمله ای قطعی از درجه d با غیر صفر گذاشتن θ_0 بخودی خود می تواند انجام شود.

تابع میانگین یک فرآیند ناایستا را با یک روند قطعی از زمان می توان نشان داد. در چنین حالتی از یک مدل رگرسیون استاندارد برای بیان مطلب می توان استفاده کرد. برای مثال اگر تابع میانگین μ_t از یک روند خطی $\mu_t = \alpha + \beta t$ پیروی کند، می توان ازمدل روند خطی قطعی $x_t = \alpha + \beta t + z_t$ استفاده کرد که در آن z_t فرآیند تصادفی محض با میانگین صفر می باشد. بطور کلی یک روند قطعی هر شکل تابعی از زمان می تواند باشد.

در بسیاری از کاربردها که دلیل فیزیکی برای وجود مؤلفه قطعی وجود ندارد، می توان میانگین را صفر فرض کرد مگر آنکه چنین فرضی متناقض با حقایقی باشد که داده ها نشان میدهند. برای بسیاری از کاربردها فرض روند تصادفی واقع بینانه تر از فرض روند قطعی است. این مسئله در پیش بینی یک سری زمانی اهمیت خاصی دارد. زیرا روند تصادفی، سری را به تبعیت از مدل یکسانی که در گذشته شکل گرفته است ملزم نمی کند. در آینده هرگاه $d > 0$ باشد اغلب فرض خواهیم کرد که $\theta_0 = 0$، مگر آنکه از داده ها یا از ماهیت مسأله چنین برآید که یک میانگین غیر صفر یا بطور عمومی تر یک مؤلفه قطعی با شکل مشخص، مورد نیاز می باشد.

یک سری زمانی مانا را می توان به عنوان یک فرآیند TS(فرآیند روند–مانا، یا فرآیندی که مانا حول یک روند است) و یک سری زمانی نامانا را بصورت یک فرآیند DS(فرآیند تفاضل–مانا، یا فرآیندی که مانا در اولین تفاضل است) مدل سازی نمود.

از نقطه نظر پیش بینی، پیش بینی های مبتنی بر فرآیندهای TS معتبرتر می باشند در حالیکه پیش بینی های مبتنی بر فرآیندهای DS غیرقابل اعتماد و گاه خیلی خطرناک هستند.

مدل ARIMA(p,d,q)

یک مدل کلی که توانایی نمایندگی طبقه گسترده ای از سری های زمانی ناایستا را دارد فرآیند تلفیقی اتورگرسیو-میانگین متحرک با درجه (p,d,q)است. با توجه به اینکه در عمل بیشتر سری های زمانی ناایستا هستند، لذا این رده از فرآیندها کاربرد گسترده ای دارند. این فرآیند را به شکل زیر تعریف می کنیم:

$$\varphi(B)\nabla^d x_t = \theta_0 + \theta(B)z_t \qquad (۱۱-۳)$$

یا

$$\varphi(B)w_t = \theta_0 + \theta(B)z_t \,;\, w_t = \nabla^d x_t \qquad (۱۲-۳)$$

اگر معادله فوق را باز کنیم خواهیم داشت:

$$w_t = \alpha_1 w_{t-1} + ... + \alpha_p w_{t-p} + z_t + \beta_1 z_{t-1} + ... + \beta_q z_{t-q}$$

سری $\{w_t\}$ با dمرتبه تفاضلی کردن سری اصلی $\{x_t\}$ بدست آمده است.

پارامتر θ_0 برای $d=0$و $d>0$نقشهای متفاوتی را بازی می کند. اگر $d=0$ فرآیند اولیه ایستا است و θ_0 به میانگین فرآیند وابسته است، یعنی $\theta_0 = \mu\varphi(B)$. با این وجود وقتی $d \geq 1$، θ_0 را جمله روند قطعی می نامند. معمولا فرض می کنیم $\theta_0 = 0$باشد. مگر اینکه داده ها خلاف آن را نشان دهند.

مدل $\varphi(B)\nabla^d x_t = \theta(B)z_t$ معادل است با اینکه سری x_t را به عنوان خروجی یک صافی خطی که ورودی آن فرآیند تصادفی محض z_t باشد، در نظر بگیریم. همچنین می توان مدل مذکور را بعنوان ابزاری برای تغییر شکل دادن فرآیند بشدت همبسته و ناایستای x_t بصورت یک دنباله از متغیرهای تصادفی مستقل z_t در نظر گرفت.

برای یک مدل ARIMA(p,d,q)با $w_t = \nabla^d x_t$ فرآیند ARMA(p,q) ایستا می باشد.

مدل IMA(d,q)

یک فرآیندARIMA که جملات اتو رگرسیو نداشته باشد فرآیندIMA(d,q)نامیده می شود که بیانگر مدل میانگین متحرک تلفیق شده است.

مدل ساده(۱,۱)IMAمعرف بسیاری از سری های زمانی رضایتبخش می باشد. بخصوص سریهایی که در اقتصاد و کسب و کار رخ می دهند. این مدل به شکل زیر است:

$$x_t = x_{t-1} + z_t + \beta z_{t-1}$$

یا

$$(1-B)x_t = (1+\beta B)z_t$$

مدل(۱,۱)IMA پس از یک بار تفاضلی کردن به صورت یک فرآیند (۱)MA ایستا در می آید. مدل گام برداری تصادفی، حالت ویژه این مدل با $\beta = 0$ است. بنابراین پدیده اصلی مدل(۱,۱)IMA با این حقیقت مشخص می شود که acf نمونه ای سری اولیه به سمت صفر میل نکند و acf سری تفاضلی شده طرح یک پدیده میانگین متحرک مرتبه اول را نشان دهد.

مدل ARI(p,d)

یک فرآیندARIMA که جملات میانگین متحرک نداشته باشد فرآیند ARI(p,d) نامیده می شود. فرآیند (۱,۱)ARI بصورت زیر نوشته می شود:

$$\begin{aligned}
&(1-\alpha B)\nabla x_t = z_t \\
&(1-\alpha B)(x_t - x_{t-1}) = z_t \\
&x_t - x_{t-1} = \alpha(x_{t-1} - x_{t-2}) + z_t \\
&x_t = (1+\alpha)x_{t-1} - \alpha x_{t-2} + z_t
\end{aligned}$$

که در آن $|\alpha| < 1$.

فرآیندهای اتو رگرسیو و میانگین متحرک مرتبه اول و دوم و فرآیند مرکب بصورت ساده(۱,d,۱) اهمیت خاصی دارند. به همین جهت خواص توابع خود همبستگی و خود همبستگی جزئی این فرآیندها را در جدول (۳-۲) خلاصه کرده ایم.

جدول (۳-۲) برخی ویژگیهای تفاضل d ام یک فرآیند ARIMA از مرتبه (p,d,q) .

	(۱,d,۱)	(۲,d,۰)	(۰,d,۲)	(۱,d,۰)	(۰,d,۱)
رفتار ρ_k	بعد از تأخیر یک بطور نمائی میرا می شود .	مخلوطی از میرائی نمائی یا سینوسی.	فقط ρ_2,ρ_1 غیر صفر هستند.	بطورنمائی میرا می شود.	فقط ρ_1 غیر صفر است.
رفتار φ_{kk}	غلبه با میرائی نمائی از تأخیر اول به بعد..	فقط $\varphi_{22},\varphi_{11}$ غیر صفر می باشند.	غلبه با مخلوطی از جملات نمائی یا میرائی سینوسی.	فقط φ_{11} غیر صفر است.	میرائی با غلبه نما ئی.
برآورد اولیه از ...	$\rho_1=\frac{(1+\theta_1\varphi_1)(\varphi_1+\theta_1)}{1+\theta_1^2+2\varphi_1\theta_1}$ $\rho_2=\rho_1\varphi_1$	$\varphi_1=\frac{\rho_1(1-\rho_2)}{1-\rho_1^2}$ $\varphi_2=\frac{\rho_2-\rho_1^2}{1-\rho_1^2}$	$\rho_1=\frac{\theta_1(1+\theta_2)}{1+\theta_1^2+\theta_2^2}$ $\rho_2=\frac{\theta_2}{1+\theta_1^2+\theta_2^2}$	$\varphi_1=\rho_1$	$\rho_1=\frac{\beta_1}{1+\beta_1^2}$
ناحیه قابل قبول	$-1<\theta_1<1$ $-1<\varphi_1<1$	$-1<\varphi_2<1$ $\varphi_2+\varphi_1<1$ $\varphi_2-\varphi_1<1$	$-1<\theta_2<1$ $\theta_2+\theta_1<1$ $\theta_2-\theta_1<1$	$-1<\varphi_1<1$	$-1<\beta_1<1$

حالت ریشه واحد

همانطور که گفته شد یک فرآیند ARMA ایستا است، اگر ریشه های $\varphi(B)=0$ "خارج" از دایره واحد واقع شوند. اگر این ریشه ها "داخل" دایره واحد واقع شوند، این فرآیند ناایستا و دارای تغییرات ناگهانی خواهد بود. حالت سوم این است که ریشه های $\varphi(B)=0$، "روی" دایره واحد قرار داشته باشند. این گونه مدلها از لحاظ نمایش دادن سری های زمانی ناایستای همگن، ارزش خاصی دارند و منجر به مسائل خاصی می شوند که به مسائل ریشه واحد مشهورند و در اقتصاد کاربرد دارند. بخصوص در سری های غیر فصلی اغلب مدلهایی که در آنها یک یا چند ریشه واحد وجود دارند بخوبی نمایش دهنده رفتار سری می باشند.

مشکل ریشه واحد در سری های زمانی وقتی نمایان می شود که هر دو چند جمله ای اتورگرسیو و میانگین متحرک در مدلARMA دارای یک ریشه روی دایره واحد یا نزدیک آن باشند. یک ریشه واحد در هر یک از این چند جمله ای ها مفاهیم مهمی برای مدل دارد. به عنوان مثال یک ریشه نزدیک به یک چند جمله ای اتورگرسیو پیشنهاد می کند که داده ها قبل از برازش یک مدلARMA بایستی تفاضلی شوند. در حالی که یک ریشه نزدیک به یک چند جمله ای میانگین متحرک پیشنهاد می کند که داده ها بیش از حد تفاضلی شده اند.

برای توضیح بیشتر مدل اتورگرسیو مرتبه اول $x_t=\alpha x_{t-1}+z_t$ را در نظر می گیریم. چنانچه α یعنی ضریب x_{t-1} برابر یک شود، با مسئله ریشه واحد مواجه هستیم که بیانگر ایستا نبودن سری زمانی x_t می باشد. بنابراین اگر رگرسیون $x_t=\alpha x_{t-1}+z_t$ را اجرا کنیم و تشخیص دهیم که $\alpha=1$، گفته می شود x_t دارای یک ریشه واحد است.

در اقتصاد سنجی سریهای زمانی، سری زمانی که دارای ریشه واحد باشد فرآیند گام برداری تصادفی نامیده می شود که نمونه ای از یک سری زمانی ناایستا است.

بیشتر متغیرهای اقتصاد کلان توسط مدلهای ARIMA با یک ریشه واحد به خوبی توصیف می شوند. در یک مدلARIMA مسأله ریشه واحد هم ارز تعیین پارامتر dدر مدل ARIMA(p,d,q) است.

۳-٤ مدل های فصلی

درعمل بیشتر سری های زمانی شامل یک مؤلفه فصلی دوره ای هستند که هر S مشاهده را تکرار می کنند. Sنشان دهنده طول دوره فصلی است. کمترین دوره زمانی برای این پدیده تکراری را **دوره فصلی**می نامند. برای داده های ماهانه $S=12$و برای داده های فصلی(سه ماهه) $S=4$ می باشد.

برای مثال سری سه ماهه فروش بستنی در هر تابستان بالا است و سری ای که این پدیده را در هر سال تکرار می کند، دارای دوره فصلی بطول٤ می باشد. بطور مشابه فروش ماهانه اتومبیل و فروش ماهانه اسباب بازی در ماه های خاصی از سال، بالا می رود و دوره فصلی در این مثال ۱۲ می باشد.

منشاء پدیده فصلی ممکن است عواملی مانند آب و هوا باشد. زیرا بسیاری از فعالیتهای اقتصادی و کسب و کارها مانند صنعت توریسم، خانه سازی و فروش بسیاری از فرآورده ها و محصولات به وضعیت جوی ارتباط دارد. همچنین برخی رسوم مانند عید نوروز، روز مادر، روز معلم، کریسمس و ... ارتباط نزدیکی با فروش جواهر آلات، اسباب بازیها، کارتها و تمبرها و ... دارد.

۳-٤-۱ مدلهای فصلی باکس- جنکینز

در این قسمت یک خانواده مهم از مدل های فصلی به نام خانواده مدلهای فصلی باکس–جنکینز را معرفی می کنیم.

۱- فرآیند اتورگرسیو فصلی از درجهP

این فرآیند بصورت زیر تعریف می شود:

$$x_t = \lambda_1 x_{t-S} + \lambda_2 x_{t-2S} + ... + \lambda_p x_{t-PS} + z_t \quad (۱۳-۳)$$

یا

$$\Phi(B^S) x_t = z_t \quad (۱٤-۳)$$

$$\Phi(B^S) = 1 - \lambda_1 B^S - \lambda_2 B^{2S} - .... - \lambda_P B^{PS}$$

مدل اتورگرسیو فصلی را می توان به عنوان حالت خاصی از مدل اتورگرسیو غیرفصلی معمولی از درجه $p = PS$ تلقی کرد که در آن ضرایب λ فقط در تأخیرهای فصلی $S, 2S, ..., PS$ مخالف صفرند.

۲- فرآیند میانگین متحرک فصلی از درجه Q

این فرآیند بصورت زیر تعریف می شود:

$$x_t = z_t + \gamma_1 z_{t-S} + \gamma_2 z_{t-2S} .... + \gamma_Q z_{t-QS} \quad (۱۵-۳)$$

یا

$$x_t = \Theta(B^S) z_t \quad (۱٦-۳)$$

$$\Theta(B^S) = 1 + \gamma_1 B^S + \gamma_2 B^{2S} + ... + \gamma_Q B^{QS}$$

مدل فصلی MA(Q) رامی توان به عنوان حالت خاص یک مدل غیر فصلی معمولی MA از مرتبه $q = QS$ تلقی نمود که در آن تمام مقادیر γ بجز در تأخیرهای فصلی $S, 2S, ..., QS$ صفرند.

این سری همواره ایستا است و تابع خود همبستگی آن فقط در تأخیرهای فصلی مخالف صفر است.

۳- فرآیند اتورگرسیو- میانگین متحرک فصلی از درجه P وQ

این مدل به شکل زیر می باشد:

$$x_t = \lambda_1 x_{t-S} + \lambda_2 x_{t-2S} + ... + \lambda_P x_{t-PS} + z_t + \gamma_1 z_{t-S} + \gamma_2 z_{t-2S} + ... + \gamma_Q z_{t-QS}$$

(۱۷-۳)

به ندرت مدلهائی که فقط درتأخیرهای فصلی دارای خودهمبستگی مخالف صفر هستند مورد نیاز است.

٤-ARIMA فصلی ضربی

باکس و جنکینز برای بررسی فصلی بودن، مدلARIMA را تعمیم می دهند و مدل فصلی ضربی کلی را به شکل زیر تعریف می کنند :

$$\phi_p(B)\Phi_P(B^S)w_t = \theta_q(B)\Theta_Q(B^S)a_t \quad (۱۸-۳)$$

مدل فوق را یک مدل فصلی ضرب پذیر از درجه $(p,d,q)\times(P,D,Q)_S$ می نامند. θ_q,Φ_P,ϕ_p و Θ_Q به ترتیب چند جمله ای هایی از مرتبه p,P,q,Q هستند و $\{a_t\}$ که بجای $\{z_t\}$ بکار رفته، نمادی است که باکس و جنکینز برای یک فرآیند تصادفی محض با میانگین صفر و واریانس σ_a^2 بکار می برند. معمولا مقادیر d و D از یک تجاوز نمی کنند.

متغیرهای $\{w_t\}$ که بصورت $w_t=\nabla^d\nabla_S^D x_t$ تعریف شده است با تفاضلی کردن سری اولیه $\{x_t\}$ برای از بین بردن روند و فصلی بودن، بوجود می آیند. به عنوان مثال برای داده های ماهانه $(S=12)$ اگر $d=D=1$ آنگاه داریم:

$$w_t=\nabla\nabla_{12}x_t=\nabla_{12}x_t-\nabla_{12}x_{t-1}=(x_t-x_{t-12})-(x_{t-1}-x_{t-13})$$

اکنون به بررسی بیشتر مدل ضرب پذیر می پردازیم.

مدل ضرب پذیر

وقتی یک سری رفتار فصلی با دوره تناوب S را نشان می دهد، قرار دادن داده ها در یک فرم جدولی شامل S ستون مفید خواهد بود. مانند جدول مربوط به داده های خطوط هوایی. (سری تعداد مسافران خطوط هوایی مشابه سری کل مسافت پیموده شده توسط مسافران خطوط هوایی می باشد که در پیوست با نام T۹ آمده است).

جدول (۳-۳)

ماه / سال	Jan	Feb	Mar	Apr	May	June	July	Aug	Sept	Oct	Now	Dec
۱۹٦۰												
۱۹٦۱												
۱۹٦۲												
....												

۱۹۷۱												

ترتیب آرایش این جدول تأکید می کند که در داده ها ی متناوب یک فاصله زمانی مهم وجود ندارد، بلکه دو فاصله زمانی است که واجد اهمیت می باشد.

برای این مثال، این فاصله ها متناظر ماه ها و سالها می باشد. بخصوص انتظار ظهور روابطی را (الف) بین مشاهدات مربوط به ماه های متوالی در یک سال خاص و (ب) بین مشاهدات مربوط به یک ماه در سالهای متوالی داریم. این وضعیت تا اندازه ای شبیه به مدل تجزیه واریانس دوطرفه است که در آن شباهتهایی را بین مشاهدات موجود در یک ستون و بین مشاهدات موجود در یک سطر می توان انتظار داشت.

با مراجعه به داده های خطوط هوایی، اثر فصلی دلالت براین دارد که مشاهدات برای یک ماه خاص مثلا آوریل با مشاهدات مربوط به ماههای مشابه از سالهای پیشین ارتباط دارد. فرض می کنیم x_t مشاهده t ام و مربوط به ماه آوریل باشد. ممکن است این مشاهده را به مشاهدات در آوریل های قبلی به وسیله مدل زیر ارتباط دهیم:

$$\Phi(B^S)\nabla_S^D x_t = \Theta(B^S)\alpha_t \qquad \text{(I)}$$

(I)

که در آن S=۱۲، $\nabla_S = 1 - B^S$ و $\Phi(B^S)$ و $\Theta(B^S)$ چند جمله ای هائی برحسب B^S بترتیب از درجه P و Q می باشندکه در شرط ایستایی و وارون پذیری صدق می کنند.

بطور مشابه برای ارتباط دادن رفتار فعلی ماه مارس با رفتار مشاهدات پیشین مربوط به ماه مارس مدل زیر را می توان تعریف کرد:

$$\Phi(B^S)\nabla_S^D x_{t-1} = \Theta(B^S)\alpha_{t-1} \qquad \text{(II)}$$

و به همین ترتیب برای هر یک از دوازده ماه می توان مدلهای مشابهی تعریف کرد.

مؤلفه های خطا،... $\alpha_t , \alpha_{t-1},$ در این مدلها بطور کلی ناهمبسته نخواهند بود. مثلا کل مسافران خطوط هوایی در آوریل ۱۹۶۰ در حالیکه به کل مسافران آوریل های قبلی ارتباط دارد، به کل مسافران مارس ۱۹۶۰، فوریه ۱۹۶۰، ژانویه ۱۹۶۰ و الی آخر نیز بستگی خواهد داشت. پس انتظار داریم که α_t در مدل(I) به α_{t-1} در مدل(II)و به α_{t-2} و غیره بستگی داشته باشد. بنابراین با در نظر گرفتن این روابط مدل دومی بصورت زیر خواهیم داشت:

$$\phi(B)\nabla^d \alpha_t = \theta(B)a_t$$

که درآن a_t بیانگر فرآیند تصادفی محض می باشد و $\phi(B)$و $\theta(B)$ چند جمله ای هایی بر حسب B به ترتیب از درجه p و q بوده و در شرط ایستایی و وارون پذیری صدق می کنند.

با جایگزین کردن معادله اخیر در (I)بالاخره یک مدل "ضرب پذیر" عمومی بشکل زیر به دست می آوریم:

$$\phi_p(B)\Phi_P(B^S)\nabla^d \nabla_S^D x_t = \theta_q(B)\Theta_Q(B^S)a_t$$

که برای این مثال خاص $S=12$است. این فرآیند یک مدل ضرب پذیر از مرتبه $(p,d,q)\times(P,D,Q)_S$ نامیده می شود.

فصل چهارم

پیش بینی

آمار: علم تصمیم گیری در شرایط عدم قطعیت.

مقدمه

پیش بینی مقادیر آینده یک سری زمانی مسا لهای مهم وکاربردی و مورد توجه محققین می باشد. سری زمانی مشاهده شده $x_1, x_2, ... x_N$ را درنظر بگیرید. می خواهیم x_{N+q} را برآورد کنیم. به عبارتی می خواهیم در زمان N مقدار سری زمانی را برای q مرحله بعد پیش بینی کنیم. عدد صحیح q را زمان تقدم می نامند. مقدار پیش بینی q مرحله بعد را در زمان N با $\hat{x}(N+q)$ نشان می دهیم.

٤-١ اهمیت پیش بینی

پیش بینی یک عنصر مهم در تصمیم گیری مدیریت است. زیرا کارایی نهایی هر تصمیمی بستگی به طبیعت یک دنباله از حوادث دارد که متعاقب آن تصمیم می آید. توانایی حدس زدن جنبه های غیر قابل کنترل این حوادث قبل از تصمیم گیری، باید به امکان انتخاب بهتری نسبت به مواردی که این توانایی در دسترس نباشد، بیانجامد. به این دلیل سیستم های مدیریت برای طرح ریزی و کنترل عملیات یک سازمان نوعا از یک تابع پیش بینی برخوردارند.

به علت اینکه پیش بینی هیچ وقت کاملا ریسک را حذف نخواهد کرد، لازم است که فرآیند تصمیم سریعا عدم قطعیت به جا مانده که متعاقب پیش بینی است را در نظر بگیرد. اغلب از لحاظ مفهومی تصمیم و پیش بینی توسط رابطه زیر به هم مربوط می شوند:

مقدار مجازبرای خطای پیش بینی + تصمیم با فرض اینکه پیش بینی صحیح است = تصمیم واقعی

این رابطه متضمن آن است که سیستم پیش بینی باید توصیفی از خطای پیش بینی و همچنین خود پیش بینی به عمل آورد. ایده آل خواهد بود که فرآیند پیش بینی به تخمینی از توزیع احتمالی متغییری که حدس زده می شود بیانجامد.این امر اجازه می دهد که ریسک رابه گونه ای عینی وارد فرآیندتصمیم گیری کنیم.

پیش بینی به خودی خود یک نقطه پایان نیست، بلکه وسیله ای برای رسیدن به نقطه انتهاست. سیستم پیش بینی جزئی از یک سیستم بزرگتر مدیریت است و به عنوان یک زیر سیستم با سایر مؤلفه های سیستم کل دارای واکنش داخلی است تا کارایی کل را مشخص کند.

مقیاس هایی وجود دارند که می توان از آنها برای ارزیابی قدرت تأثیر یک سیستم پیش بینی استفاده کرد.از مهمتریناینمقیاسها، دقت پیش بینی، هزینه سیستم، بهره وری ستاده و خواص پایداری و قدرت جواب به تغییرات می باشد.

نمونه ای از مواردی که پیش بینی برای آنها مفید خواهد بود عبارت است از:

مدیریت موجودی، طرح ریزی تولید، برنامه ریزی مالی، برنامه ریزی امور پرسنلی، برنامه ریزی امکانات و کنترل فرآیند.

٤-٢ زمان تقدم

پیش بینی ها برای مقاطع آینده زمانی مانند روزانه، هفتگی، ماهانه، فصلی و سالانه صورت می گیرد. این مقاطع زمانی را چهارچوب زمانی، افق زمان یا زمان تقدم و یا لید تایم می نامند.

معمولا طول چارچوب زمانی بصورت زیر طبقه بندی می شود:

- فوری : کمتر از یک ماه.
- کوتاه مدت : از یک تا سه ماه.
- میان مدت : بیش از سه ماه و کمتر از دو سال.
- بلند مدت : دو سال و بیشتر.

روشهای پیش بینی ممکن است تا اندازه ای به زمان تقدم لازم بستگی داشته باشند. هرچه زمان تقدم طولانی ترباشد، تغییرالگوی موردنظرمحتملتربودهوبنابراین پیش بینی غیرقابل اعتمادترخواهد بود.

٤-٣ خطای پیش بینی

دقت یک روش پیش بینی با تجزیه وتحلیل خطاهای پیش بینی تجربه شده تعیین می شود. اگر x_t مشاهده واقعی در پریود t و $\hat{x}_t$ پیش بینی برای آن پریود باشد که یک یا چند پریود قبل انجام شده است، خطای پیش بینی برای پریود t به صورت $e_t = x_t - \hat{x}_t$ تعریف می شود.

برای یک فرآیند و روش پیش بینی داده شده، خطای پیش بینی به عنوان یک متغیر تصادفی در نظر گرفته می شود که دارای میانگین $E(e)$ و واریانس σ_e^2 است. اگر پیش بینی نااریب باشد، $E(e) = 0$ خواهد بود. درحالیکه پیش بینی نااریب مطلوب است، معمولا مهمتر این است که خطاهای پیش بینی بزرگ به ندرت پیش آیند، از این رو کمیتی مانندارزش انتظاری قدر مطلق خطا که به صورت $E(|e_t|) = E|x_t - \hat{x}_t|$ تعریف می شود یاارزش انتظاری مجذور خطا که عبارت است از

$E(e_t^2)=E(x_t-\hat{x}_t)^2$؛ معمولا به عنوان مقیاسی از دقت پیش بینی به کار می رود. توجه دارید که ارزش انتظاری مربع خطا که معمولا میانگین مربع خطا نامیده می شود، مساوی σ_e^2 است، اگر پیش بینی نااریب باشد.

روش معمول برای بررسی اینکه آیا یک مدل احتمالی واقعا توصیف کننده داده ها است یا خیر، این است که خطای پیش بینی یا همان باقیمانده ها را مورد بررسی قرار دهیم. اگر مدلی به نحو رضایتبخشی نماینده فرآیند باشد، آنگاه انتظار می رود که مقدار متوسط خطاهای پیش بینی نزدیک صفر باشد. بطور کلی اینکه پیش بینی ها نااریب باشند یعنی ارزش انتظاری خطای پیش بینی صفر باشد، موضوع مورد توجه عمده است.

با بررسی خطاهای پیش بینی در طی زمان، غالبا می توان به این نکته پی برد که آیا روش پیش بینی به کار رفته منطبق بر الگوی داده ها بوده است یا خیر. اگر یک روش پیش بینی توانسته باشد اجزاء روند، نوسانات سیکلی و نوسانات فصلی را که در سری زمانی وجود دارد به درستی پیش بینی نماید، در این صورت خطای پیش بینی منعکس کننده جزء بی نظم خواهد بود.

٤-٤ اندازه گیری خطای پیش بینی

در مینی تب برای اندازه گیری دقت مدل برازش داده شده به یک سری زمانی از سه معیار به نامهای MAD، MSD و MAPE استفاده می شود که آنها را معیارهای دقت یا Measures of Accuracy می نامند. در این قسمت به اختصار این معیارها را معرفی می کنیم.

۱- میانگین قدر مطلق درصد خطا

$$MAPE = \frac{\sum_{t=1}^{n} \left|(x_t - \hat{x}_t)/x_t\right|}{n} \times 100 \qquad x_t \neq 0$$

x_t مقدار واقعی سری زمانی و $\hat{x}_t$ مقدار پیش بینی شده متناظر با آن می باشد و n تعداد پیش بینی ها است.

۲-میانگین قدر مطلق انحرافات

$$MAD = \frac{\sum_{t=1}^{n} \left|x_t - \hat{x}_t\right|}{n}$$

۳- میانگین مربع انحرافات

$$MSD = \frac{\sum_{t=1}^{n} (x_t - \hat{x}_t)^2}{n}$$

از این معیارها به دو منظور می توان استفاده کرد:

۱- جهت کمک به فرآیند انتخاب مدل پیش بینی.
۲- جهت نظارت بر سیستم پیش بینی به منظور دستیابی به "اشتباه موردی" که در سیستم پیش بینی رخ داده است.

در مواقعی که خطاها رو به بزرگ شدن می نهند این معیارها به ما علامت داده و ما را آگاه می سازند، از اینکه ممکن است الگوی داده ها تغییر کرده باشد.

٤-٥ ارزیابی پیش بینیها

برای ارزیابی یک روش پیش بینی می توان از روشهای زیر استفاده نمود.

۱- بررسی خود همبستگیهای خطاهای پیش بینی

چون فرض می شود خطاهای پیش بینی فاقد هر گونه ساختار باشند، انتظار داریم این خود همبستگی ها را نزدیک صفر بیابیم. هر خود همبستگی که بزرگ بنظر برسد را می توان با خطای استاندارد آن مقایسه کرد. با توجه به اینکه فاصله اطمینان ۹۵ درصد برای r_k تقریبا بصورت $\left(\frac{-2}{\sqrt{N}},\frac{2}{\sqrt{N}}\right)$ می باشد، اگر مقدار مشاهده شده r_k (برای باقیمانده ها)در خارج از این حدود واقع شود، میگوییم این مقدار در سطح پنج درصد به طور معنی داری با صفر اختلاف دارد.

۲- رسم مشاهده واقعی x_t در مقابل پیش بینی $\hat{x}_t$

اگر پیش بینی ها و مشاهدات یکسان باشند همه نقاط روی خط ٤٥ درجه ای که از مبداء عبور می کند قرار خواهند گرفت. انحرافات معنی دار از این خط عدم کفایت مدل را مشخص می کند. برای مثال اگر بیشتر نقاط بالا یا پایین این خط واقع شوند، پیش بینی ها اریب خواهند بود.

۳- رسم نمودار احتمال نرمال خطاهای پیش بینی

از این نمودار برای بررسی اینکه آیا خطاهای پیش بینی بطور نرمال توزیع شده اند یا نه استفاده می شود. چنانچه خطاها بطور نرمال توزیع شده باشند این نمودار تقریبا بصورت یک خط راست خواهد بود.

٤- بررسی معیارهای دقت

با استفاده از معیارهای دقت یا Measures of Accuracy که مینی تب با هر روش پیش بینی در اختیار ما می گذارد، می توان روشهای مختلف پیش بینی را با یکدیگر مقایسه نمود و مدلی را که برازش بهتری را به داده ها فراهم می آورد، انتخاب نمود.

ممکن است بخواهیم روش های پیش بینی را بر مبنای جوابشان به تغییرات پایدار در فرآیند سری های زمانی و ثباتشان در صورت وجود تغییرات تصادفی و تغییرات موقت(ناپایدار) مقایسه کنیم. این کار از طریق شبیه سازی و برای بعضی روش های آماری خاص، به وسیله تجزیه و تحلیل ریاضی امکان پذیر است.

بهره وری پیش بینی در تکمیل تصمیمات مدیریت، بستگی به وقوع سر وقت، شکل پیش بینی و همچنین دقت آن دارد. منافع به طور کلی باید با توجه به سیستم مدیریت اندازه گیری شود. پیش بینی تنها یک مؤلفه از این سیستم کلی است. هدف دسترسی به تصمیمات خوب است و معمولا این امر با چیزی کمتر از پیش بینی های کامل قابل دسترسی است.

٤-٦ روشهای پیش بینی

به طور کلی روشهای پیش بینی را می توان به روشهای کیفی وروشهای کمی تقسیم کرد.

٤-٦-١ روشهای کیفی پیش بینی

روشهای کیفی پیش بینی مبتنی بر دانش تجربی و درک مستقیم و سایر اطلاعات مرتبط با موضوع می باشد و بطور کلی در این روشها از نظرات و عقاید متخصصین مربوطه استفاده می شود.

برای مثال فرض کنید که قراراست محصولی کاملا جدید معرفی شود. در چنین شرایطی هیچ داده ای در مورد فروش گذشته این محصول وجود ندارد و برای پیش

بینی فروش باید از روشهای کیفی(روشهای مبتنی بر نظرات و عقاید متخصصین) استفاده کرد.

کشف و ورود یک تکنولوژی جدید در عرصه تولید از دیگر مـواردی اسـت کـه در آن داده های زمانی مربوط به گذشته در دسترس نیست. در نتیجه در این موارد هم باید از روشهای کیفی استفاده کرد.

از آنجا که استفاده از داده های مربوط به گذشته به منظور پیش بینی مقادیر آینده بـر این فرض استوار است که الگوی داده های زمـانی ثابـت مـی مانـد، بنـابرین تغییـرات الگوی داده ها را نمی توان بر اساس داده های گذشته پیش بینی کرد. بـرای پـیش بینـی تغییرات الگوی داده های زمانی نیز می توان از روشهای کیفی استفاده کرد.

برخی از مهمترین روشهای کیفی پیش بینی که بـه ذکـر نـام آنهـا بسـنده مـی کنـیم عبارتنـد از: روش بـرازش منحنـی ذهنـی، روش دلفـی، روش تقـابلی، روش درخـت مناسب، روش مقایسات تکنولوژیکی، روش تحقیق مرفولوژیکی و ....

٤-٦-٢ روشهایکمیپیش بینی

در یک تقسیم بندی کلی می توان روشهای کمی پیش بینی را به دو رده روشهای یک متغیره و روشهای چند متغیره تقسیم کرد. آنچه ما در این کتاب بررسی خواهیم کرد روشهای پیش بینی یک متغیره خواهد بود.

- روشهای کمی
  - روشهای یک متغیره
    - برون یابی منحنی روند
    - هموارکردن نمائی
    - روش هلت- وینترز
    - روش باکس-جنکینز
  - روشهای چند متغیره

در روشهای یک متغیره داده های مربوط به گذشته مورد تجزیه و تحلیل قرار می گیرد و سپس با این فرض که الگوی بدست آمده تا آینده ادامه خواهد داشت، پیش

بینی ها انجام می شود. مدلهای تک متغیره برای وضعیتهایی که انتظار می رود همچنان ادامه داشته باشند کاربرد دارد.

در روشهای 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. الگوهای رگرسیون و اقتصادسنجی از این نوع هستند. به روشهای چند متغیره که گاهی آنها را مدلهای علی می نامند چند اشکال زیر وارد است:

۱- بسط و توسعه آنها بسیار مشکل است.
۲- تمامی متغیرهای مورد استفاده در این مدل مستلزم این می باشند که از داده های مربوط به گذشته برخوردار باشند.
۳- توانایی در پیش بینی متغیر وابسته منوط به توانایی پیش بینی کننده در پیش بینی متغیرهای مستقل می باشد.

روشهای یک متغیره پیش بینی

همانطور که گفته شد در این روشها با تجزیه و تحلیل داده های گذشته مدل مناسبی را شناسایی کرده و با فرض بدون تغییر باقی ماندن این مدل، آن را مبنای پیش بینی رفتار آینده سری قرار می دهیم. اکنون مهمترین روشهای پیش بینی یک متغیره را به اختصار بررسی می کنیم.

۱- برون یابی منحنی روند

یک سری دارای روند در واقع یک سری ناایستا است. زیرا میانگین این سری ثابت نیست و دستخوش تغییرات دراز مدت می باشد. برای پیش بینی های دراز مدت، اغلب برازش یک منحنی روند به مشاهدات سالانه و برون یابی آن مفید است. برای این کار دست کم داده های تاریخی مربوط به ۷ تا ۱۰ سال لازم است. نبایستی پیش بینی ها برای دوره ای طولانی تر از حدود نصف تعداد سالهای گذشته ای که برای آنها داده ها در دسترس هستند، در نظر گرفته شود.

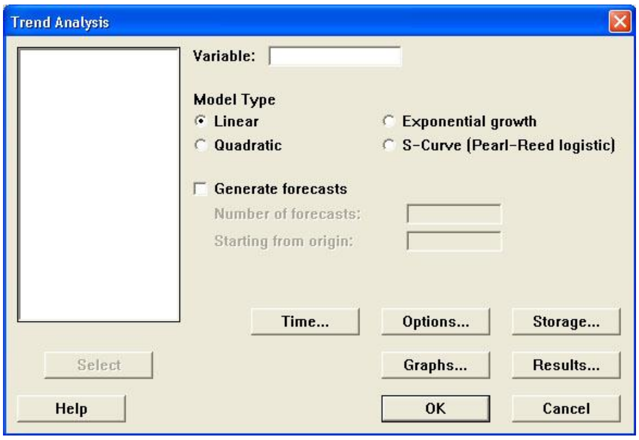
تجزیه و تحلیل سری زمانی که تغییرات دراز مدت در میانگین را نشان می دهد،بستگی دارد به اینکه بخواهیم روند را اندازه گیری کنیم یا اینکه بخواهیم برای تجزیه و تحلیل نوسانات موضعی روند را حذف کنیم.

در جایی که برازش مدل پیچیده ای به داده های گذشته با ارزش به نظر نمی رسد در نظر گرفتن پیش بینی ها برای مدت طولانی با ارزش است. زیرا ممکن است مدل در آینده تغییر کند.

متأسفانه اغلب می توان چندین منحنی پیدا کرد که تقریبا به یک اندازه به مجموعه داده ها برازنده اند. لیکن وقتی پیش بینی می شوند، پیش بینی های کاملا متفاوتی را تولید می کنند.

برازش منحنی روند در Minitab

برای تجزیه و تحلیل یک سری زمانی دارای روند در مینی تب از منوی Statگزینه Time Seriesو سپس گزینه Trend Analysis را انتخاب می کنیم تا پنجره ای به شکل زیر باز شود. از Trend Analysis برای تجزیه و تحلیل داده هایی که دارای یک روند ثابت هستند و فاقد الگوی فصلی نیز می باشند، استفاده می شود.



شکل ٤-١ : پنجره Trend Analysis

در پنجره ظاهر شده در قسمت Model Type چهار منحنی روند جهت برازش به داده ها دیده می شود. برای تمام این منحنی ها، تابع برازنده شده یک اندازه گیری از روند را فراهم می نماید و باقیمانده ها که تفاوت بین مشاهدات واقعی و مقادیر متناظر منحنی برازنده شده می باشند، برآوردی از نوسانات موضعی را نشان می دهد.

گزینهModel Type

در این قسمت می توان یکی از چهار مدل زیر را به داده ها برازش داد.

۱- Linear

این گزینه یک مدل روند خطی به شکل $y_t = \beta_0 + \beta_1 t + e_t$ را به داده ها برازش می دهد.

۲-Quadratic

این گزینه یک مدل درجه دوم به شکل $y_t = \beta_0 + \beta_1 t + \beta_2 t^2 + e_t$ را به داده ها برازش می دهد که در آن β_1 متوسط تغییر در هر پریود را نشان می دهد.

۳-Exponential growth

این گزینه یک منحنی نمائی به شکل $y_t = \beta_0 \beta_1^t + e_t$ را به داده ها برازش می دهد.

٤- S-Curve (Pearl-Reed logistic)

این گزینه یک منحنی S به معادله $y_t = \dfrac{10^a}{\beta_0 + \beta_1 \beta_2^{t-1}}$ را به داده ها برازش می دهد. این منحنی وقتی مناسب است که نمودار داده ها به شکل S باشد.

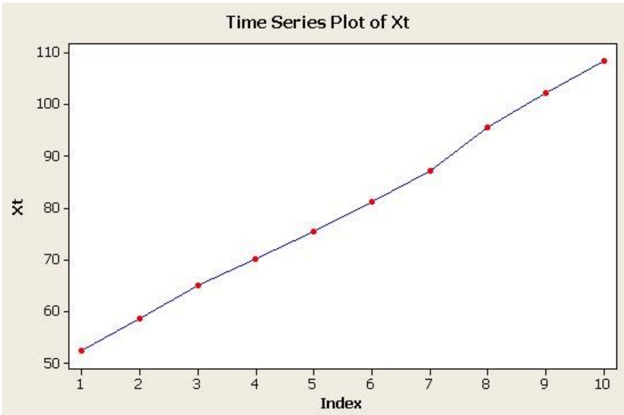
قبل از انتخاب این مدل ابتدا باید داده های گمشده را از مجموعه داده ها حذف کرد. اگر یکی از سه مدل دیگر را انتخاب کنیم مینی تب به طور خودکار داده های گمشده را در نظر نخواهد گرفت.

مثال ٤-١

داده های مربوط به وزن گوساله ها در ده هفته متوالی در جدول زیر آمده است. می خواهیم یک خط روند مناسب را به داده ها برازش دهیم و متوسط نرخ رشد در هفته را محاسبه کنیم و همچنین وزن گوساله ها را برای هفته یازدهم و دوازدهم پیش بینی کنیم.

(سن) t	١	٢	٣	٤	٥	٦	٧	٨	٩	١٠
(وزن) x_t	٥٢.٥	٥٨.٧	٦٥	٧٠.٢	٧٥.٤	٨١.١	٨٧.٢	٩٥.٥	١٠٢.٢	١٠٨.٤

ابتدا داده ها را وارد مینی تب نموده و سپس نمودار سری زمانی را برای این داده ها رسم می کنیم.

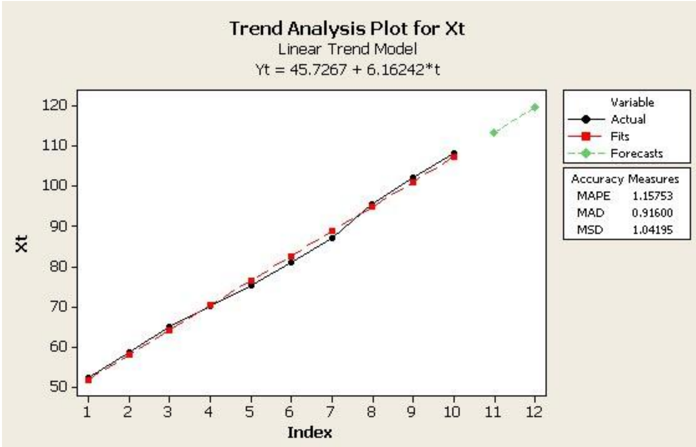


شکل ٤-٢ : نمودار سری زمانی وزن گوساله ها

این نمودار به صراحت یک مدل روند خطی را پیشنهاد می کند. بنابراین با باز کردن پنجره Trend Analysis و با انتخاب گزینه Linear یک مدل روند خطی را به داده ها برازش می دهیم.

برای انجام این کار نام متغیر مورد نظر(نام ستونی که داده ها را در آن ذخیره کرده ایم) را به کادر Variable منتقل می کنیم و سپس از قسمت Model Type گزینه Linear را انتخاب می کنیم. همچنین برای تولید پیش بینی های دو پریود بعد، با انتخاب گزینه Generate forecasts در کادر مقابل Number of forecasts عدد ٢ را وارد می کنیم.

با فشردن دکمه ok خروجی در پنجره sessionو نمودار مربوطه بصورت زیر خواهد بود:



شکل ٤– ٣ : خط روند برازش داده شده به داده های مثال ٤–١

```
Trend Analysis for Xt

Data      Xt
Length    10
NMissing  0


Fitted Trend Equation

Yt = 45.7267 + 6.16242*t


Accuracy Measures

MAPE  1.15753
MAD   0.91600
MSD   1.04195


Forecasts

Period  Forecast
11       113.513
12       119.676
```

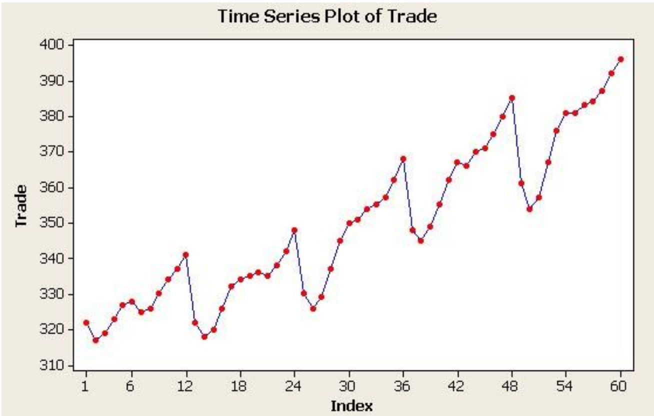
همانطور که ملاحظه می شود معادله خـط رونـد بشـکل $y_t = 45.72 + 6.16*t$ مـی باشد. بنابراین وزنگوساله ها با توجه به خط کمترین مربعات بعد از ۳،۲،۱ و... هفتـه بـه ترتیب عبارت است از $A+B$، $A+2B$، $A+3B$ و ... در نتیجه نـرخ متوسـط رشـد برابر ٦.١٦ واحد می باشد.

مثال ٤–٢

فایل EMPLOY.MTW را باز می کنیم. این فایل شامل داده های مربوط به اشتغال در بخش بازرگانی(trade business) برای ٦٠ ماه می باشد. می خواهیم یک منحنی روند

مناسب را به این داده ها برازش دهیم و سپس میزان اشتغالزایی در این بخش را برای ۱۲ ماه آینده پیش بینی کنیم.

ابتدا نمودار داده ها را رسم کرده ایم. این نمودار یک منحنی درجه دوم را پیشنهاد می کند.

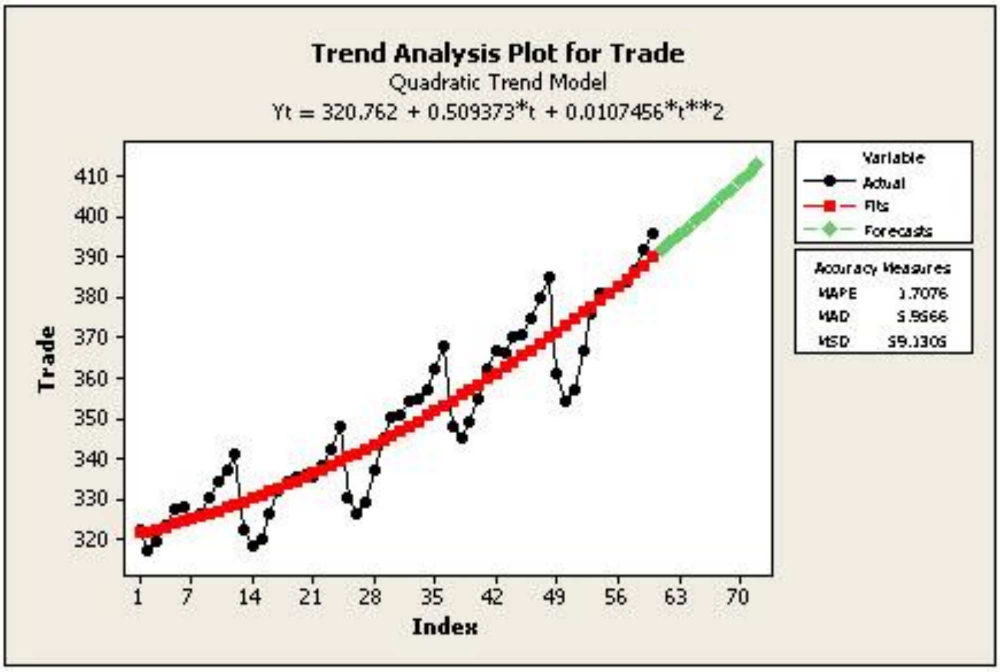


شکل ٤–٤ : نمودار سری زمانی برای متغیرTrade

برای برازش یک منحنی درجه دوم به این داده ها مانند مثال قبل پنجرهTrendAnalysis را باز کرده و سپس متغیرtrade را بهکادرVariable منتقل می کنیم. در قسمت Model Type نیز گزینه Quadratic را انتخاب می کنیم.

برای تولید پیش بینی های ۱۲ ماه آینده با انتخاب گزینهGenerate forecasts در کادر مقابل Number of forecasts عدد ۱۲ را وارد می کنیم.

با فشردن دکمه ok نمودار مربوطه و خروجی در پنجره session بصورت زیر خواهد بود :



شکل ٤– ٥ : منحنی روند برازش داده شده به سری زمانی trade

Trend Analysis for Trade

```
Data      Trade
Length    60
NMissing  0


Fitted Trend Equation

Yt = 320.762 + 0.509373*t + 0.0107456*t**2


Accuracy Measures

MAPE   1.7076
MAD    5.9566
MSD   59.1305


Forecasts

Period  Forecast
61       391.818
62       393.649
63       395.502
64       397.376
65       399.271
66       401.188
67       403.127
68       405.087
69       407.068
70       409.071
71       411.096
72       413.142
```

توجه :

چنانچه نتوانستیم تشخیص دهیم که کدام یک از مدلهای روند برازش مناسبی را فراهم می آورد، بهتر است هر چهار مدل را امتحان کنیم و سپس باتوجه به معیارهای دقت، مدل مناسب را انتخاب کنیم. کمتر بودن مقادیر مربوط به این معیارها حاکی از بهتر بودن مدل انتخاب شده می باشد.

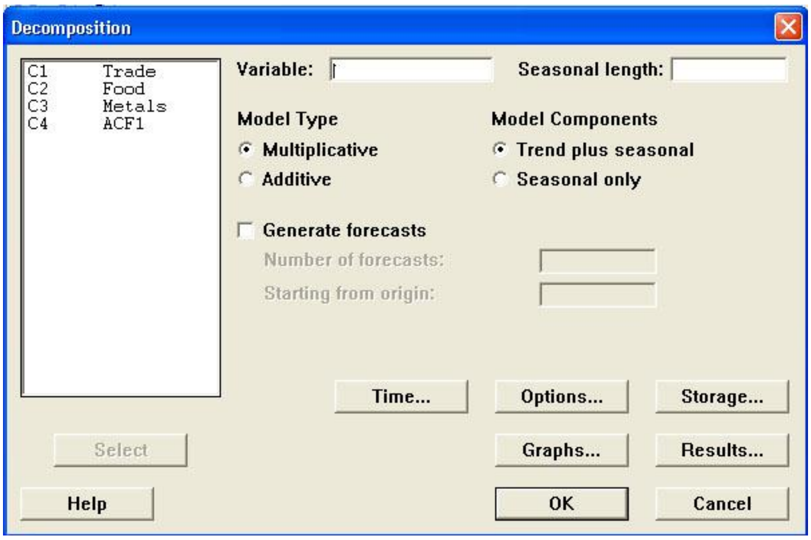
تجزیه و تحلیل سری زمانی دارای روند و مؤلفه فصلی

چنانچه علاوه بر روند، داده ها یک رفتار فصلی را نیز نشان دهند می توان برای تجزیه سری به روند خطی و عامل فصلی از رویهDecomposition استفاده کرد.

این رویه در مینی تب شامل مراحل زیر می باشد:

۱- با استفاده از رگرسیون کمترین مربعات یک خط روند به داده ها برازش می یابد.
۲- در مدل ضربی با تقسیم داده ها بر عامل روند و در مدل جمعی با تفریق روند از داده ها، داده ها فاقد روند می شوند.
۳- با استفاده از میانگین متحرک مرکزی(با طولی برابر با طول دوره فصلی)، داده های فاقد روند هموار می شوند.
٤- پس از محاسبه میانگین متحرک، در مدل ضربی با تقسیم داده ها بر آن و در مدل جمعی با تفریق داده ها از آن، مقادیری با عنوان raw seasonalsحاصل می شود.
٥- در هر پریود فصلی میانه مقادیر raw seasonals محاسبه شده و سپس تعدیل می شود. این میانه های تعدیل شده را شاخص های فصلی(seasonalindices) می نامند.
٦- از این شاخص های فصلی برای تبدیل داده ها به داده های فصلی تعدیل شده(seasonally adjusted data) استفاده می شود.

برای اجرای این رویه در مینی تب کافی است از منوی Stat گزینه Time Series و سپس از منوی ظاهر شده گزینه Decomposition را انتخاب کنیم تا پنجره ای به شکل زیر باز شود.



شکل ٤–٦ : پنجره Decomposition

اکنون به اختصار به معرفی گزینه های این پنجره می پردازیم.

گزینهSeasonal length

در این قسمت باید یک عدد صحیح مثبت بزرگتر مساوی ۲ را وارد کنیم. این عدد طول عامل فصلی است. برای مثال چنانچه داده ها ماهانه باشند، باید عدد ۱۲ را وارد کرد.

گزینهModel Type

در این قسمت باید با توجه به ماهیت داده ها یکی از دو مدل جمعی یا ضربی را انتخاب کرد.

(Multiplicative)مدل ضربی

چنانچه اندازه الگوی فصلی در داده ها بستگی به سطح داده ها داشته باشد از مدل ضربی استفاده می شود. در این مدل فرض بر این است که با افزایش سطح داده ها اندازه مؤلفه فصلی نیز افزایش می یابد. بیشتر سری های زمانی چنین رفتاری دارند. الگوی ضربی بشکل $y_t = trend * seasonal + error$ می باشد.

(Additive)مدل جمعی

مدل جمعی به شکل $y_t = trend + seasonal + error$ می باشد. این مدل فرض می کند که عامل روند و عامل فصلی مستقل از یکدیگر می باشند.

Model Componentsگزینه

در این قسمت باید مؤلفه های مدل را برای نرم افزار تعیین کرد. این مؤلفه ها می تواند شامل روند و مؤلفه فصلی و یا شامل مؤلفه فصلی به تنهایی باشد که در این صورت باید از عامل روند صرفنظر کرد.

بنابراین برای لحاظ کردن عامل روند در مدل گزینه Trend plus seasonalرا انتخاب می کنیم و چنانچه بخواهیم عامل روند را نادیده بگیریم گزینه Seasonal onlyرا انتخاب می کنیم. چنانچه قبلا در قسمت AnalysisTrend روند را از داده ها حذف کرده باشیم، ممکن است بخواهیم از این گزینه استفاده کنیم.

اگر داده ها شامل روند باشند و ما آنرا نادیده گرفته باشیم، این موضوع می تواند برآورد شاخص های فصلی را تحت تأثیر قرار دهد.

Generate forecastsگزینه

جهت تولید پیش بینی برای پریودهای آینده این گزینه را مارک دار می کنیم. همچنین باید تعداد پیش بینی ها و نقطه شروع پیش بینی ها را مشخص کرد.

مثال٤ – ٣ اجرای رویه Decompositionدر مینی تب

مجددا فایل EMPLOY.MTWرا باز می کنیم. سری مشاهدات مربوط به متغیر Trade را در نظر می گیریم. وجود روند در داده ها کاملا واضح است. اما علاوه بر روند یک عامل فصلی نیز به چشم می خورد که تقریبا هر ١٢ مشاهده را تکرار می کند. بنابراین ما با یک سری زمانی که ترکیبی از روند و عامل فصلی می باشد سر و کار داریم. با

استفاده از رویه Decomposition می توان عامل روند و عامل فصلی را از یکدیگر تفکیک نمود تا برازش بهتری حاصل شود.

رویهDecomposition را به دو صورت می توان اجرا کرد:

۱- می توان ابتدا روند را از داده ها حذف نمود و سپس رویه Decomposition را بر روی داده های فاقد روند اجرا کرد. در این صورت لازم است که ابتدا با استفاده از Trend Analysis داده ها را فاقد روند کرده و سپس در پنجره Decomposition گزینه Seasonal only را انتخاب نمود.

۲- همچنین می توان این رویه را از همان ابتدا بر روی داده های اصلی اجرا کرد. در این صورت باید در پنجره اصلی Decomposition گزینه Trend plus seasonal را انتخاب کرد.

ما در اینجا رویه Decomposition را بر روی داده های فاقد روند اجرا می کنیم. برای داشتن داده های فاقد روند(detrended data) در پنجره Trend Analysis در قسمت Storage گزینه(Residuals (detrended data را مارک دار کنید. با مارک دار کردن این گزینه باقیمانده ها که تفاضل مقادیر واقعی از مقادیر برازش داده شده توسط منحنی روند می باشند، در ستونی جداگانه ذخیره خواهند شد.

نام این ستون بطور پیش فرض RESI۱ خواهد بود. حال از منوی Time series گزینه Decomposition را انتخاب می کنیم. در پنجره باز شده متغیر RESI۱ را به قسمت Variable منتقل می نماییم.

با توجه به اینکه داده ها ماهانه هستند طول دوره فصلی ۱۲ می باشد. بنابراین در قسمت Seasonal length عدد ۱۲ را وارد می نماییم.

نوع مدل را جمعی در نظر می گیریم. زیرا به نظر می رسد که سطح سری(مقدار متوسط سری) و اندازه مؤلفه فصلی مستقل از یکدیگر می باشند.

در قسمت Model Components گزینه Seasonal only را انتخاب می کنیم زیرا ما روند را از داده ها حذف نموده ایم و اکنون فقط عامل فصلی وجود دارد.

همچنین برای تولید پیش بینی هایی برای یک سال آینده با توجه به آنکه داده ها ماهانه می باشند، در قسمتNumber of forecasts عدد ۱۲ را وارد می کنیم.

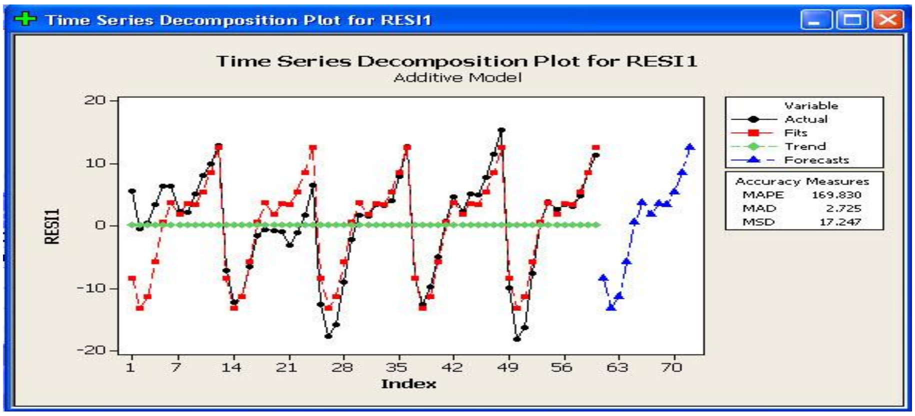
نتیجه اجرای این رویه درپنجره Sessionبصورت زیر می باشد:

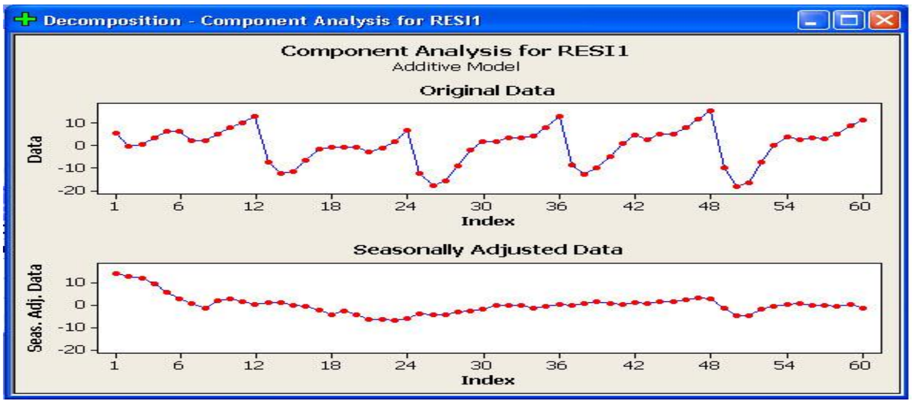
```
Time Series Decomposition for RESI1

Additive Model


Data      RESI1
Length    60
NMissing  0


Seasonal Indices

Period     Index
     1   -8.4824
     2  -13.3365
     3  -11.4415
     4   -5.8155
     5    0.5587
     6    3.5597
     7    1.7676
     8    3.4760
     9    3.2669
    10    5.3919
    11    8.4962
    12   12.5588
```

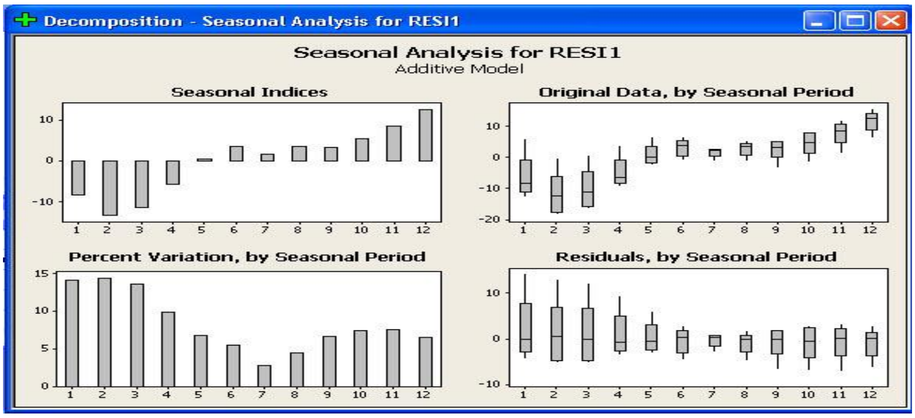
```
Accuracy Measures

MAPE  169.830
MAD     2.725
MSD    17.247


Forecasts

Period  Forecast
61       -8.3863
62      -13.2404
63      -11.3454
64       -5.7194
65        0.6548
66        3.6558
67        1.8637
68        3.5721
69        3.3630
70        5.4880
71        8.5923
72       12.6549
```

در قسمت seasonal indicesشاخص های فصلی مشاهده می شود. این ضرایب یا شاخص ها اثر هر یک از فصول را بر متغیر تعیین می کند. با استفاده از این ضرایب می توان داده ها را از فصلی بودن درآورد. یعنی اثر هر یک از فصول را بر متغیر از بین برد. ارزشهای غیر فصلی شده در مورد تمایل بلند مدت متأثر از تغییر دوره ای(در صورتی که وجود داشته باشد) اطلاعاتی در اختیار ما می گذارد.

اجرای این رویه باعث تولید سه سری نمودار می شود.مدل برازش شده توسط رویه Decomposition و مقادیر پیش بینی بر پایه این مدل در نمودار اول نشان داده شده است. همچنین معیارهای دقت یاMeasures of Accuracy نیز آورده شده اند.







شکل ٤–٧ : نمودارهای تولید شده توسط رویه Decomposition

۲- هموار سازی داده ها

هموار کردن داده ها به منظور کاهش نوسان موجود در آنها بکار می رود. در این قسمت چند شیوه هموارسازی مانند میانگین متحرک ساده و هموار سازی نمایی را مورد بررسی قرار می دهیم.

الف) میانگین متحرک ساده

میانگین متحرک یک شیوه هموار سازی است. این رویه با متوسط گیری از مشاهدات متوالی در هر لحظه، داده ها را هموار می کند. به این معنی که Noise یا اختلال در مشاهدات را کاهش می دهد. میانگینهای متحرک برای جست جوی الگوی داده های زمانی که سیگنال اساسی به وسیله سروصدا (Noise) پنهان مانده است، مفیدند.

در یک فرآیند ثابت به شکل $x_t = b + \varepsilon_t$ که $\{\varepsilon_t\}$ متغیرهای تصادفی غیرهمبسته با میانگین صفر و واریانس σ_ε^2 می باشند. میانگین متحرک معادل تجزیه و تحلیل n مشاهده اخیر است. b یک پارامتر نامعلوم است که باید برآورد شود. اگر تمامی مشاهدات در برآورد b از اهمیت یکسانی برخوردار باشند، در این صورت میانگین نمونه ای سری برآورد مناسبی برای b خواهد بود.

$$\hat{b} = \frac{1}{n}\sum_{t=1}^{n} x_t$$

میانگین متحرک ساده n پریودی (n مقداری) به شکل زیر تعریف می شود:

$$m_t = \frac{x_t + x_{t-1} + x_{t-2} + \ldots + x_{t-n+1}}{n}$$

همچنین می توان نوشت:

$$m_t = m_{t-1} + \frac{x_t - x_{t-n}}{n}$$

n دامنه میانگین متحرک است. دامنه تعداد مشاهدات گذشته است که از آنها میانگین گرفته می شود. همانطور که ملاحظه می شود m_t در واقع میانگین n مشاهده اخیر است. در هر پریود قدیمی ترین مشاهده حذف شده و جدیدترین مشاهده به مجموعه اضافه می شود.

به عنوان مثال چنانچه از میانگین متحرک سه مقداری استفاده کنیم دنباله مشاهدات واقعی با دنباله زیر جایگزین خواهد شد:

$$\frac{x_1+x_2+x_3}{3}, \frac{x_2+x_3+x_4}{3}, \frac{x_3+x_4+x_5}{3}, \frac{x_4+x_5+x_6}{3}, ...$$

سری جدید نوسان کمتری خواهد داشت. اما در عوض برخی از مشاهدات از دست خواهند رفت.

در انتهای پریود t،پیش بینی برای پریود آینده $t+q$بصورت $\hat{x}_{t+q}(t) = m_t$ خواهد بود.

خواص روش میانگین متحرک ساده بستگی به دامنه آن دارد. دامنه بزرگ باعث می شود که میانگین متحرک در برابر تغییرات در پارامتر b به کندی واکنش نشان دهد و دامنه کوچک نتیجه معکوس می دهد.

اگر فرآیند سری های زمانی دارای ارزش $b = b_1$ باشد و ناگهان به ارزش جدید $b = b_2$ منتقل شود، میانگین متحرک nپریودی وقت لازم را خواهد داشت تا پیش بینی هایی تولید کند که با مقدار جدید b سازگار باشند.

اگر خطاهای تصادفی متغیرهای مستقل باشند، واریانس m_t بصورت $\frac{\sigma_\varepsilon^2}{n}$ می باشد.
در نتیجه برای n های کوچک، واریانس m_t نسبتا بزرگ خواهد بود.

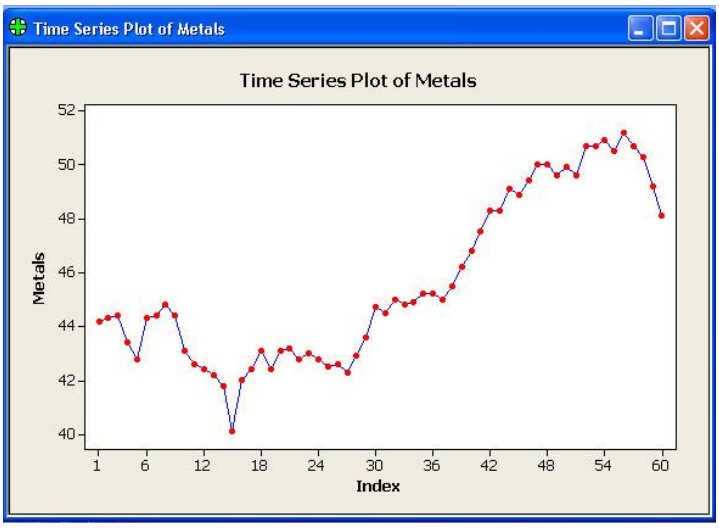
بنابراین زمانی که فرآیند واقعا ثابت است تمایل داریم از nبزرگ استفاده کنیم تا بتوان مقدار b را دقیقتر تخمین زد و موقعی که فرآیند در حال تغییر است تمایل داریم از n های کوچک استفاده کنیم بنحوی که m_t در مقابل تغییرات b جواب دهنده باشد.

در یک فرآیند دارای روند خطی بصورت $x_t = b_1 + b_2 t + \varepsilon_t$ نیز می توان از روش میانگین های متحرک برای هموار کردن داده ها استفاده کرد. در این حالت از میانگین متحرک دوگانه استفاده می شود.

روشهای میانگین متحرک که تا کنون بررسی کرده ایم، در رده هموار کننده های داده های خطی یا فیلترهای خطی قرار دارند. یعنی آنها هر مشاهده x_t را با یک میانگین(احتمالا وزنی) از نقاطی که از لحاظ زمان نزدیک آن هستند جایگزین می کنند.

مثال ٤-٤ میانگینهای متحرک در Minitab

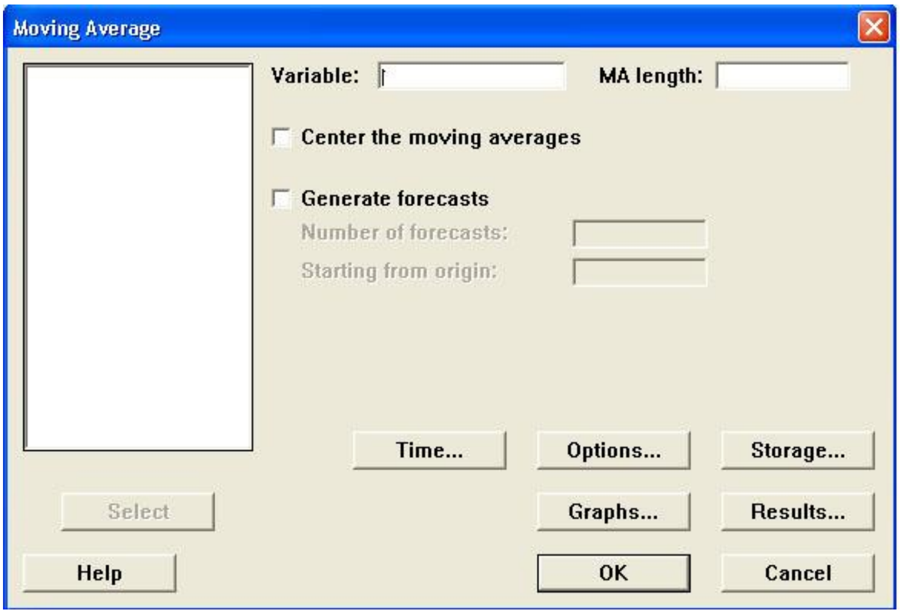
فایل EMPLOY.MTW را در نظر می گیریم. این فایل شامل سه متغیر Trade، FoodوMetalsاست. متغیر Trade دارای روندو عامل فصلی می باشد. متغیر Food نیز دارای عامل فصلی واضحی می باشد. اما متغیر Metals فاقد روند یا الگوی فصلی مشخصی می باشد.



شکل ٤- ٨ : نمودار سری زمانی Metals

به طور کلی از رویه میانگین متحرک برای تولید پیش بینی های کوتاه مدت در صورتی که داده ها فاقد روند یا الگوی فصلی مشخصی باشند استفاده می شود. بنابراین می توان از این روش برای پیش بینی های کوتاه مدت سری Metals استفاده کرد.

در مینی تب برای اجرای این رویه از منوی Time Series گزینه Moving Average را انتخاب می کنیم. با انتخاب این گزینه پنجره زیر باز می شود:



شکل ٤- ٩ : پنجره میانگین متحرک

گزینه Variable

در کادر مقابل این گزینه باید نام متغیر مورد نظر را که در اینجا متغیر Metals می باشد، وارد کرد.

گزینه MA length

در این قسمت باید دامنه یا طول میانگین متحرک را مشخص کرد. برای یک سری زمانی غیر فصلی معمولا از یک میانگین متحرک کوتاه استفاده می شود. ما در این مثال طول میانگین متحرک را ۳ در نظر می گیریم.

یک میانگین متحرک طولانی تر(با دامنه بزرگتر) قادر است noise بیشتری را از مشاهدات حذف کند. اما در عوض حساسیت آن نسبت به تغییرات سری کمتر خواهد بود.

در مورد سری های فصلی معمولا طول میانگین متحرک را برابر با طول یک دوره سالیانه در نظر می گیرند.

گزینه Center the moving averages

این گزینه مقادیر میانگین متحرک را در مرکز رنج قرار می دهد. در حالت پیش فرض این مقادیر در انتهای رنج قرار می گیرند.

اگر دامنه میانگین متحرک یک عدد فرد مانند ۳ باشد، اولین مقدار محاسبه شده در پریود دوم قرار می گیرد و دومین مقدار در پریود سوم والی آخر. در این حالت مقدار اولین و آخرین پریود missing خواهد بود.

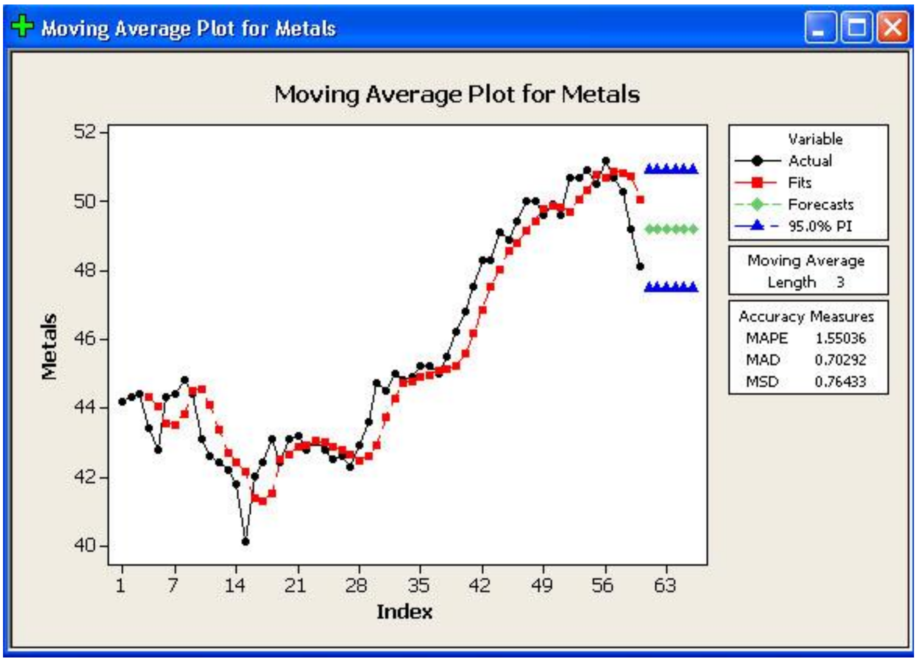
اگر دامنه میانگین متحرک یک عدد زوج مانند ٤ باشد، مینی تب میانگین متحرک اولین چهار مقدار را محاسبه کرده و آن را ma۱ می نامد. سپس میانگین متحرک برای دومین چهار مقدار را محاسبه کرده و آن را ma۲ می نامد. حال در ستون سوم از ma۱ و ma۲ میانگین گرفته و آن را در پریود سوم قرار می دهد و همین روال را برای سایر مشاهدات ادامه می دهد. در این حالت مقدار میانگین متحرک برای اولین دو پریود و آخرین دو پریود missingخواهد بود.

گزینه Generate forecasts

معمولا برای تولید پیش بینی های کوتاه مدت از روشهای هموار سازی مانند میانگین متحرک استفاده می شود.

در اینجا ما با مارک دار کردن این گزینه می خواهیم مقدار سری را برای شش ماه آینده پیش بینی نماییم. بنابراین در کادر مقابل Number of forecasts عدد شش را وارد می کنیم.با تکمیل پنجره فوق نتیجه در پنجره session به صورت زیر خواهد بود.

```
Moving average

Data        Metals
Length      60.0000
NMissing    0

Moving Average
Length: 3

Accuracy Measures
MAPE:  1.55036
MAD:   0.70292
MSD:   0.76433


 Row  Period  Forecast    Lower    Upper

   1      61      49.2  47.4865  50.9135
   2      62      49.2  47.4865  50.9135
   3      63      49.2  47.4865  50.9135
   4      64      49.2  47.4865  50.9135
   5      65      49.2  47.4865  50.9135
   6      66      49.2  47.4865  50.9135
```



شکل ٤– ١٠ : پیش بینی سری Metalsبا استفاده از روش میانگین متحرک ساده

همانطور که ملاحظه می شود مقادیر پیش بینی برای شش ماه آینده و حدود اطمینان ۹۵ درصد برای این پیش بینی ها در پنجره session و در نمودار مربوطه داده شده است.

ب) هموار سازی نمائی

پیش بینی نمائی احتمالا گسترده ترین کاربرد را در میان روشهای مختلف پیش بینی جهت سریهای زمانی گسسته دارد که آینده فوری و نزدیک را پیش بینی می کنند.

هموار سازی نمائی تکنیکی کارا برای تخمین ضرایب در یک مدل سری زمانی چند جمله ای است. چند جمله ای برازش شده را می توان برای پیش بینی مقادیر آینده سری به کار برد. با وجود این سری های بسیاری وجود دارند که نمی توانند بوسیله یک مدل چند جمله ای به طور مناسب مدل بندی شوند. مانند یک سری زمانی با تغییرات سیکلی یا فصلی. البته برای هموار سازی سری های زمانی فصلی دارای روند می توان از روش وینترز استفاده کرد که بعدا توضیح داده می شود.

اغلب سری های زمانی شامل یک روند یا الگوی فصلی هستند. لیکن برای بدست آوردن یک سری ایستا می توان این اثرات را اندازه گرفت و حذف کرد.

در یک سری زمانی غیر فصلی ایستای $x_1, x_2, ..., x_n$، در نظر گرفتن مجموع وزن داری از مشاهدات گذشته به عنوان برآورد x_{n+1} امری طبیعی است. یعنی

$$\hat{x}(n,1) = c_0 x_n + c_1 x_{n-1} + c_2 x_{n-2} + ... \quad (\text{I})$$

معقول به نظر می رسد که به مشاهدات جدید وزن بیشتر و به مشاهدات گذشته دور وزن کمتری نسبت دهیم.

یک مجموعه شهودی از وزنها، وزنهای هندسی هستند که با یک نسبت ثابت تنزل پیدا می کنند. برای آنکه مجموع وزنها یک شود آنها را بصورت زیر در نظر می گیریم:

$$c_i = \alpha(1-\alpha)^i \qquad i = 0,1,2,...$$

α عددی است بین صفر و یک که آن را ثابت هموار سازی می نامند. در صورت استفاده از وزنهای هندسی فوق در معادله(I) خواهیم داشت:

$$\hat{x}(n,1) = \alpha x_n + \alpha(1-\alpha)x_{n-1} + \alpha(1-\alpha)^2 x_{n-2} + ... \qquad \text{(II)}$$

روشی که با معادله(II) تعریف شده است، هموار کردن نمائی نامیده می شود. صفت نمائی از این حقیقت ناشی می شود که وزنهای هندسی روی یک منحنی نمائی قرار دارند.

مقدار ثابت هموار سازی α به خواص سری زمانی داده شده بستگی دارد. مقادیر بین ۰.۱ و ۰.۳ برای بدست آوردن پیش بینی هایی که به تعداد زیادی از مشاهدات گذشته بستگی دارد، بکار می رود.

مقادیر نزدیک به یک اغلب کمتر مورد استفاده واقع می شوند و پیش بینی هایی را می دهند که تا حد زیادی به مشاهدات جدید بستگی دارند.

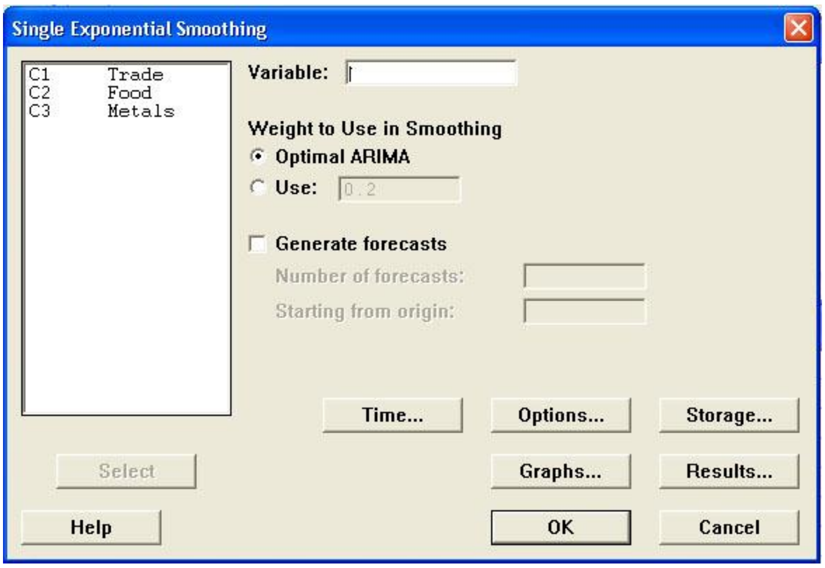
در مینی تب برای هموار سازی نمائی دو گزینه داریم Single Exp Smoothingیا هموار سازی نمائی یگانه و Double Exp Smoothing یا هموار سازی نمائی دوگانه که هر دو روش در مورد سری های زمانی غیر فصلی بکار می رود و از هر دو روش می توان جهت تولید پیش بینی های کوتاه مدت استفاده کرد.

چنانچه سری غیر فصلی فاقد روند باشد از هموار سازی نمائی یگانه استفاده می شود اما چنانچه سطح سری با نرخ ثابتی رو به افزایش باشد یعنی سری دارای روند باشد می توان از هموار سازی نمائی دوگانه استفاده کرد.

مثال ٤- ٥ هموار سازی نمائی یگانه (Single Exp Smoothing) در مینی تب

همانطور که گفته شد، این روش فقط برای سری های زمانی غیرفصلی که روندی را نشان نمی دهند بکار برده می شود. برای بررسی بیشتر این رویه مجددا فایل

EMPLOY را در نظر می گیریم. از آنجا که سری Metals در این فایل فاقد روند و الگوی فصلی مشخصی می باشد، می توان از رویه Single Exp Smoothing برای هموار کردن داده ها استفاده کرد. برای این منظور از منوی Time Series گزینه Single Exp Smoothing را انتخاب می کنیم تا پنجره ای به شکل زیر باز شود.



شکل ٤-١١ : پنجره هموار سازی نمایی یگانه

ابتدا متغیر Metalsرا به کادر Variable منتقل می کنیم. در قسمت Weight to Use in Smoothingمی توان مقدار ثابت هموار سازی را تعیین کرد. انتخاب پیش فرض مینی تب حالت Optimal ARIMA است که با مینیمم کردن مجموع مربعات خطا عدد مناسبی را به عنوان ثابت هموار سازی در نظر می گیرد. اما می توان با انتخاب گزینه Use عدد دلخواه خود را که باید عددی بین صفر و یک باشد وارد کرد.

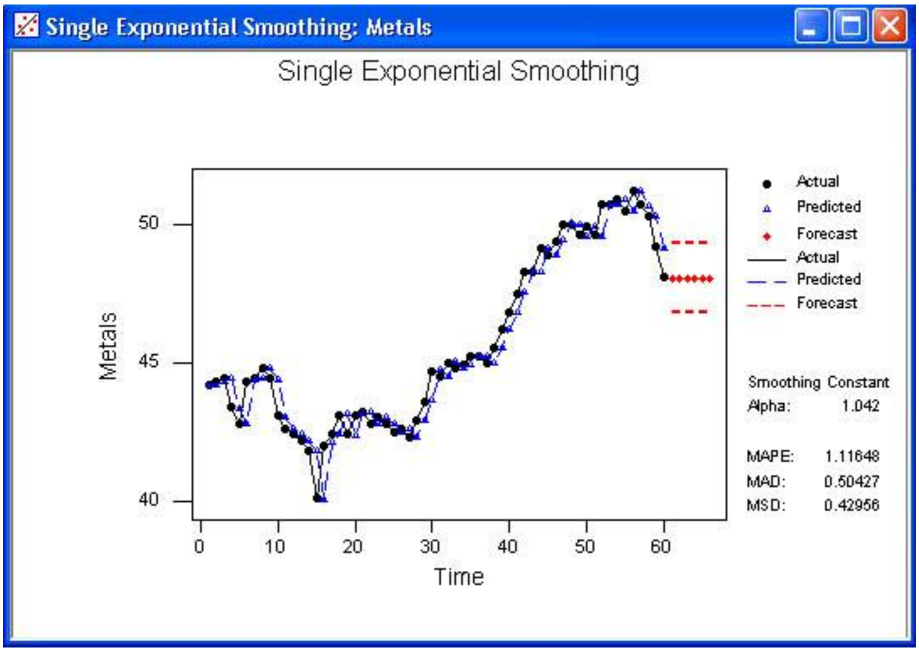
همچنین برای تولید پیش بینی های شش ماه آینده گزینه Generate forecasts را مارک دار نموده و در کادر مقابل Number of forecasts عدد ٦ را وارد می نماییم.

خروجی در پنجره session ونمودار مربوطه به صورت زیر می باشد. همانطور که ملاحظه می شود مقادیر پیش بینی برای شش ماه آینده و حدود بالا و پایین پیش بینی در انتهای خروجی آمده است.

Single Exponential Smoothing

```
Data       Metals
Length     60.0000
NMissing   0

Smoothing Constant
Alpha: 1.04170

Accuracy Measures
MAPE: 1.11648
MAD:  0.50427
MSD:  0.42956
```

Row	Period	Forecast	Lower	Upper
1	61	48.0560	46.8206	49.2915
2	62	48.0560	46.8206	49.2915
3	63	48.0560	46.8206	49.2915
4	64	48.0560	46.8206	49.2915
5	65	48.0560	46.8206	49.2915
6	66	48.0560	46.8206	49.2915



شکل ٤–١٢ : پیش بینی با استفاده از هموار سازی نمایی یگانه

توجه:

با مقایسه مقادیر Accuracy Measuresدر دو مدل moving averageو singleexponential smoothing به این نتیجه می رسیم که مدل اخیر یعنی single exponential smoothing modelبرازش بهتری را به داده ها فراهم می آورد.

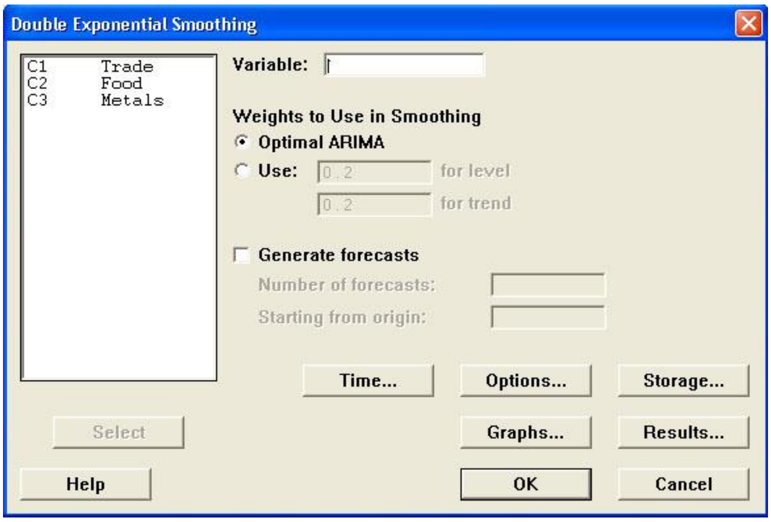
Accuracy Measures	MAPE	MAD	MSD
Single Exponential Smoothing	١.١١٦٤٨	٠.٥٠٤٢٧	٠.٤٢٩٥٦
Moving Average	١.٥٥٠٣٦	٠.٧٠٢٩٢	٠.٧٦٤٣٣

جدول (٤–١)

مثال ٤- ٦ هموار سازی نمائی دوگانه (Double Exp Smoothing) در مینی تب

هنگامی که داده ها فقط شامل روند باشند می توان از هموار سازی نمائی دوگانه برای هموار کردن داده ها استفاده کرد. در این رویه برآوردهای پویایی برای هر دو مؤلفه levelو trend یعنی سطح سری و روند موجود در سری محاسبه می شود. منظور از سطح سری همان مقدار متوسط سری(μ) است.

برای اجرای این رویه در مینی تب کافی است از منویTime Series گزینه DoubleExp Smoothing را انتخاب کنیم.



شکل ٤- ١٣ : پنجره هموار سازی نمایی دوگانه

همانطور که ملاحظه می شود گزینه های این پنجره نیز مانند پنجره هموار سازی نمایی یگانه می باشد.

برای بررسی بیشتر این روش پیش بینی، فایل EMPLOY.MTW را در نظر می گیریم. با وجود اینکه در سری غیر فصلی Metals روند واضحی به چشم نمی خورد، اما مایلیم مدل حاصل از هموار سازی نمائی دوگانه را به آن برازش دهیم و با توجه به این مدل مقدار سری را برای شش ماه آینده پیش بینی کنیم.

البته می توان با توجه به معیار های دقت(Accuracy Measures) مدلی که برازش بهتری را فراهم می آورد شناسایی کرد.

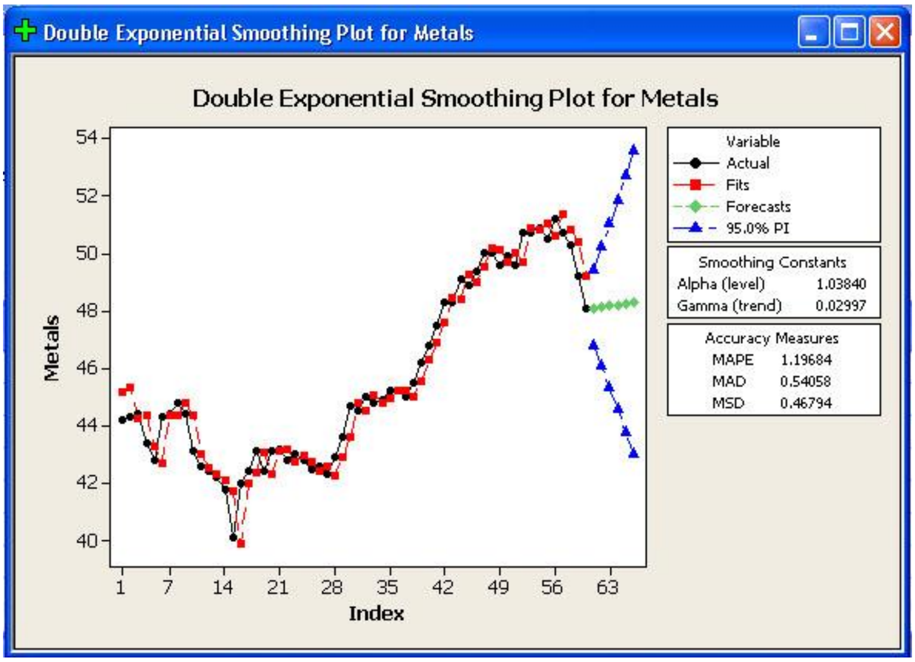
برای استفاده از روش هموار سازی نمایی دوگانه، متغیرMetals را به کادر Variable منتقل می کنیم. در قسمت Weight to Use in Smoothing نیز می توان مقدار ثابت هموار سازی را تعیین کرد. ما در اینجا انتخاب پیش فرض مینی تب را که حالت Optimal ARIMA می باشد می پذیریم که با مینیمم کردن مجموع مربعات خطا عدد مناسبی را به عنوان ثابت هموار سازی در نظر می گیرد. همچنین برای تولید پیش بینی های شش ماه آینده گزینه Generate forecasts را مارک دار نموده و در کادر مقابل Number of forecasts عدد ٦ را وارد می نماییم.

با تکمیل پنجره فوق و ok کردن آن خروجی در پنجره Session بصورت زیر می باشد.

Double Exponential Smoothing

```
Data        Metals
Length      60.0000
NMissing    0

Smoothing Constants
Alpha (level): 1.03840
Gamma (trend): 0.02997

Accuracy Measures
MAPE: 1.19684
MAD:  0.54058
MSD:  0.46794


 Row  Period  Forecast    Lower    Upper

   1      61   48.0961  46.7717  49.4206
   2      62   48.1357  46.0599  50.2114
   3      63   48.1752  45.3134  51.0369
   4      64   48.2147  44.5545  51.8748
   5      65   48.2542  43.7898  52.7185
   6      66   48.2937  43.0220  53.5653
```

در این پنجره در قسمت Smoothing Constantsبرآوردهای روند و سطح سری با Alpha و Gamma نشان داده شده اند.



شکل ٤– ١٤ : پیش بینی با استفاده از روش هموار سازی نمایی دوگانه

با مقایسه معیار های دقت به این نتیجه می رسیم که در مورد سری Metals رویه Single Exp Smoothingبرازش بهتری را فراهم می کند.

از آنجا که در این مثال اختلاف معیارهای سه گانه دقت در دو روش هموارسازی نمائی ناچیز است، ممکن است بخواهیم ملاک انتخاب مدل بهتر را نوع پیش بینی قرار دهیم. درروش هموارسازی نمائی یگانه پیش بینی ها برروی یک خط مستقیم قرار می گیرند. در حالی که در روش هموار سازی نمائی دوگانه پیش بینی ها روی یک خط شیب دار قرار دارند.

همان طور که ملاحظه می شود در نمودار Double Exp Smoothing برای متغیر Metals مقادیر پیش بینی شده به کندی در حال افزایش هستند و این در حالی است که چهار مشاهده اخیر این سری زمانی در حال کاهش هستند.

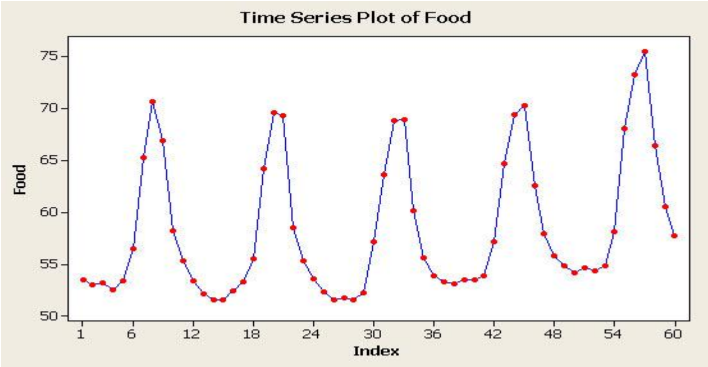
اگر وزن بیشتری را برای مؤلفه روند در نظر بگیریم باعث می شود که برآوردها در همان جهت داده ها باشند که این به واقعیت نزدیک تر است. برای آنکه بتوانیم برای مؤلفه روند وزن دلخواهی در نظر بگیریم باید در قسمت Weight to Use inSmoothing بجای گزینه Optimal ARIMA گزینه Use را انتخاب کرده و عدد دلخواه خود را برای هر یک از دو مؤلفهTrend و Level وارد کنیم. حال می توانیم با مقایسه معیارهای دقت برای وزنهای مختلف، آنکه برازش بهتری را فراهم می آورد انتخاب کنیم.

۳ - روش پیش بینی هلت- وینترز

یکی دیگر از روش های پیش بینی، روش هلت-وینترز است. با استفاده از این روش به آسانی می توان هموار کردن نمائی را به سری هایی که شامل روند و تغییرات فصلی می باشند تعمیم داد. از روش وینترز برای پیش بینی های کوتاه مدت و همچنین پیش بینی های میان مدت استفاده می شود. این رویه برآوردهای پویایی از مؤلفه های روند، سطح و مؤلفه فصلی فراهم می آورد. برای بررسی بیشتر این روش به مثال٤-٧ توجه کنید.

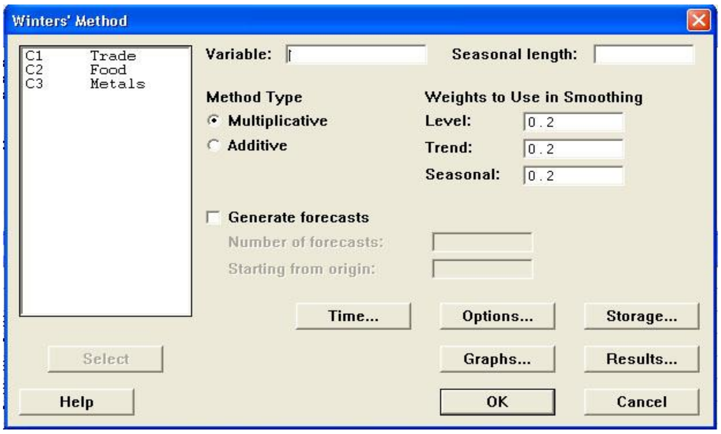
مثال ٤- ٧ روش پیش بینی هلت-وینترز در MINITAB

سری زمانیFood **از** فایلEMPLOY.MTWرا در نظر می گیریم. همانطور که از نمودار سری مشخص است داده ها شامل یک روند ویک الگوی فصلی هستند که به نظر می رسد مؤلفه فصلی متناسب با تغییر سطح سری تغییر می کند.



شکل ٤– ١٥ : نمودار سری زمانی Food

برای انجام پیش بینی با استفاده از روش هلت–وینترز، از منوی Time Series گزینه Winters' Method را انتخاب می کنیم تا پنجره مربوطه باز شود.



شکل ٤ – ١٦ : پنجره هلت– وینترز

در پنجره ظاهر شده با توجه به آنکه داده ها ماهانه می باشند، طول عامل فصلی را ١٢ در نظر می گیریم. نوع مدل را نیز ضربی در نظر می گیریم.

در قسمت Weights to Use in Smoothing باید وزنهایی برای مؤلفه روند، مؤلفه سطح و مؤلفه فصلی در نظر بگیریم. انتخاب پیش فرض نرم افزار برای هر سه مؤلفه عدد ٠.٢ می باشد. ما نیز همین انتخاب پیش فرض را می پذیریم.

در اینجا به علت وجود محدودیتهایی مانند هموار سازی نمائی یگانه و دوگانه حالت Optimal ARIMA نداریم.

همچنین برای تولید پیش بینی هایی برای شش ماه آینده عدد ٦ را در کادر Number of forecasts وارد می کنیم.

در قسمت Starting from origin می توانیم نقطه شروع تولید پیش بینی ها را برای نرم افزار مشخص کنیم. به عنوان مثال در سری فوق که شامل ٦٠ مشاهده می باشد، چنانچه در این قسمت هیچ عددی وارد نکنیم، نرم افزار نقطه شروع پیش بینی ها را بلافاصله بعد از آخرین مشاهده در نظر می گیرد. اما چنانچه مثلا عدد ٥٤ را وارد کنیم و بخواهیم سه پیش بینی داشته باشیم، مینی تب پیش بینی هایی برای پریودهای ٥٥، ٥٦ و ٥٧ می دهد. که این پیش بینی ها بر پایه وضعیت مؤلفه های سه گانه هموار سازی(روند، سطح و عامل فصلی) در پریود ٥٤ تولید می شوند. با تکمیل پنجره فوق خروجی بصورت زیر می باشد:

Winters' Method for Food

Smoothing Constants

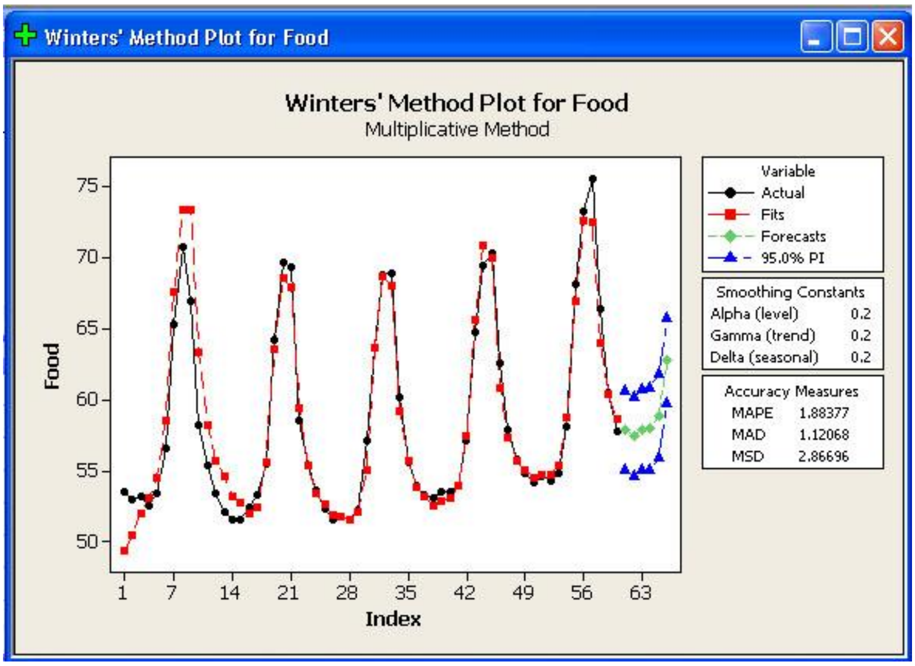
Alpha (level)	0.2
Gamma (trend)	0.2
Delta (seasonal)	0.2

Accuracy Measures

MAPE	1.88377
MAD	1.12068
MSD	2.86696

Forecasts

Period	Forecast	Lower	Upper
61	57.8102	55.0646	60.5558
62	57.3892	54.6006	60.1778
63	57.8332	54.9966	60.6698
64	57.9307	55.0414	60.8199
65	58.8311	55.8847	61.7775
66	62.7415	59.7339	65.7492



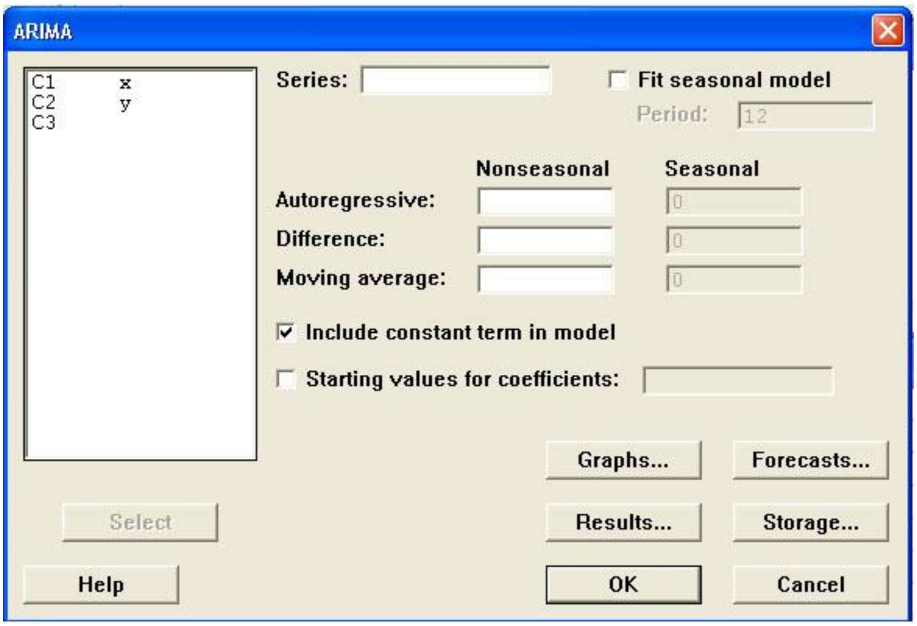
شکل ٤ – ١٧ : پیش بینی با استفاده از روش هلت–وینترز

٤ – روش پیش بینی باکس– جنکینز

این روش اساسا شامل برازش یک مدلARIMA به داده ها می باشد. برای تولید پیش بینی بر پایه یک مدلARIMA، پس از مشخص شدن مدل نهایی یعنی تعیین مرتبه تفاضلی کردن و تعیین مرتبه هر یک از فرآیندهای ARو MA، بر روی گزینه

.... Forecasts در پنجره اصلی ARIMA کلیک کرده و در پنجره باز شده تعداد پیش بینی ها و نقطه شروع آن را مشخص می کنیم.

برای این کار از منوی Stat گزینه Time Series و سپس گزینهARIMA را انتخاب می کنیم.



شکل ٤– ١٨ : پنجره ARIMA

در این پنجره با تعیین نوع و مرتبه هر یک از مؤلفه های فرآیند, مدل آزمایشی شناسایی شده را به داده ها برازش می دهیم.

چنانچه مدل شناسایی شده از نوع فصلی باشد، با مارک دار کردن گزینه Fit seasonal modelمؤلفه ها ی فصلی مدل را مشخص می کنیم. به این ترتیب که طول دوره فصلی را در قسمت Periodتعیین می کنیم و مرتبه اتورگرسیو فصلی و میانگین متحرک فصلی و همچنین درجه تفاضلات فصلی را در ردیف مربوطه در ستون Seasonalمی نویسیم.

چنانچه مدل غیر فصلی باشد در ستون Nonseasonalمرتبه هر یک از فرآیندهای اتورگرسیو و میانگین متحرک و همچنین درجه تفاضلی کردن را مشخص می کنیم.

مینی تب بطور پیش فرض جمله ثابت θ_0 را در مدل لحاظ می کند و مدل را با جمله ثابت برازش می دهد. در صورتی که با بررسی مدل برازش شده به این نتیجه برسیم که نیازی به حضور جمله ثابت در مدل نمی باشد، کافی است چک مارک عبارت Include constant term in model را برداریم تا مدلی بدون جمله ثابت داشته باشیم.

در قسمت Starting values for coefficientsچنانچه لازم باشد مقادیر اولیه را برای برآورد پارامترهای مدل تعیین می کنیم. به این ترتیب که مقادیر اولیه مورد نظر را در یکی از ستونهای مینی تب ذخیره کرده و سپس نام این ستون را در کادر مقابل عبارت فوق وارد می کنیم. لازم به ذکر است که الگوریتم ARIMA تا ۲۵ تکرار انجام می دهد تا مدل داده شده را برازش دهد. اگر جواب همگرا نباشد، می توان پارامترهای برآورد شده را ذخیره کرده و از آنها به عنوان مقادیر اولیه برای برازش دوم استفاده کرد.

جهت تعیین مقادیر اولیه برای تخمین پارامترها، باید این مقادیر را به ترتیب زیر در یک از ستونهای worksheet وارد کنیم:

مقادیراتورگرسیو- مقادیر اتو رگرسیو فصلی- مقادیر میانگین متحرک- مقادیر میانگین متحرک فصلی و مقدار ثابت.

سپس در قسمتStarting values for coefficients نام ستونی که شامل مقادیر اولیه برای هر پارامتر می باشد را وارد می کنیم.

فصل بعد بطور ویژه به مدلسازی سریهای زمانی با روش باکس-جنکینز می پردازد. در این فصل با ارائه مثالهای متنوع، مدل سازی سری های زمانی را با روش باکس-جنکینز تمرین خواهیم کرد. با مطالعه این فصل انتظار می رود خواننده بتواند مدل مناسب برای چند سری را که به عنوان تمرین در پایان فصل آمده است به راحتی تشخیص دهد.

فصل پنجم

مدل سازی سری های زمانی به روش باکس – جنکینز

مقدمه

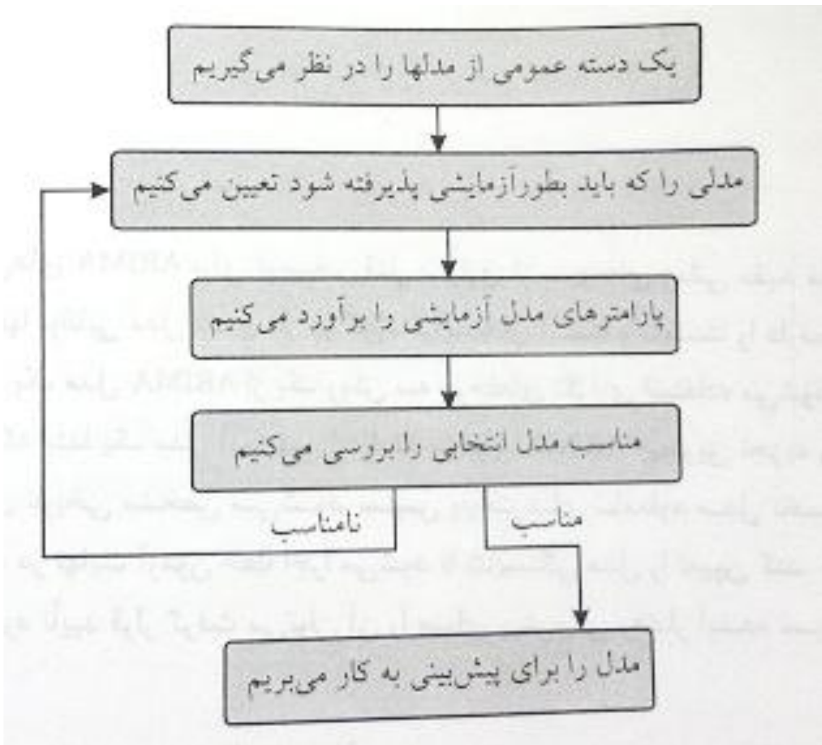
مدل های ARIMA برای توصیف رفتار بسیاری از سری های زمانی مفید می باشند. این مدلها توانایی مدل بندی هر دو گروه فرآیندهای ایستا و ناایستا را دارند. برای ساختن یک مدل ARIMA از یک روش سه مرحله ای تکراری استفاده می شود. به این ترتیب که ابتدا یک مدل آزمایشی از طبقه مدلهای ARIMA از طریق تجزیه و تحلیل داده های تاریخی مشخص می شود. سپس پارامترهای نامعلوم مدل تخمین زده می شود. در نهایت آزمون خطا اجرا می شود تا شایستگی مدل را تعیین کند. چنانچه مدل مورد تأیید قرار گرفت می توان آن را مبنای پیش بینی رفتار آینده سری قرار داد.

۵- ۱ استراتژی مدل سازی

مدل سازی یک سری زمانی به طور کلی شامل سه مرحله می باشد:

۱. تشخیص مدل آزمایشی
۲. تخمین پارامترهای مدل(برازش مدل)
۳. بررسی مناسب بودن مدل

در این قسمت چگونگی انجام هر یک از مراحل فوق را بررسی می کنیم. در فلوچارت زیر نیز می توان مراحل ساخت مدل بطریق تکراری را مشاهده نمود.



شکل ۵ – ۱ : فلوچات مربوط به مراحل مدل سازی یک سری زمانی

۵-۱-۱ تشخیص مدل آزمایشی

جهت تشخیص یک مدل آزمایشی باید حداقل۵۰ مشاهده از سری مورد نظر دراختیار داشته باشیم. اکنون مراحل تشخیص مدل آزمایشی را بررسی می کنیم تا بتوانیم یک مدل مناسب از خانواده مدلهای ARIMA بیابیم.

مرحله اول : بررسی ایستایی در واریانس

اولین گام در مدل سازی یک سری زمانی رسم نمودار آن می باشد. نمودار سری زمانی به شناسایی روند، ناایستایی در واریانس، فصلی بودن و شناسایی داده های پرت کمک شایانی می کند. با توجه به آنکه مدلهای احتمال سریهای زمانی برای سری های ایستا در میانگین و واریانس تعریف شده اند، لازم است که ابتدا ایستایی سری را بررسی کرد و در صورت ناایستا بودن سری، با انجام تبدیلات مناسب آن را به یک سری ایستا تبدیل کرد. مهمترین تبدیلات مورد نیاز سری زمانی تبدیلات تثبیت کننده واریانس و تبدیلات تفاضلی می باشد. اگر قرار باشد هر دوی این تبدیلات را انجام دهیم، ابتدا باید تبدیلات تثبیت کننده واریانس را انجام دهیم. به همین جهت ابتدا پایایی واریانس سری را بررسی می کنیم. مهمترین ابزار بررسی پایایی واریانس همانطور که در فصل دوم توضیح داده شد استفاده از رویه COX-BOXمی باشد.

مرحله دوم : بررسی ایستایی در میانگین

همانطور که در فصل دوم دیدیم برای بررسی ایستایی سری در میانگین می توان از نمودار سری و همبستگی نگار آن استفاده کرد. چنانچه acf نمونه ای بسیار کند تنزل کند و pacf بعد از تأخیر یک قطع شود، لزوم تفاضلی کردن را می رساند. ناایستایی در میانگین از نمودار سری زمانی نیز مشخص است. ممکن است برای رفع ناایستایی لازم باشد داده های اولیه را dبار تفاضلی کنیم. البته تجربه نشان داده است که dمعمولا از ۲ تجاوز نمی کند.

مرحله سوم : رسم acfو pacf نمونه ای

وسیله مهم در تشخیص مدل، تابع خود همبستگی و تابع خود همبستگی جزئی می باشد. رسم این نمودارها در تعیین نوع و مرتبه فرآیند مفید است. برای تشخیص بهتر مدل توصیه می شود که حجم نمونه حداقل ۵۰ باشد وacf و pacf حداقل تا تأخیر $k=\frac{n}{4}$محاسبه و رسم شوند.

ما در این مرحله ابتدا رفتار acf و pacf نظری را بررسی می کنیم و سپس با توجه به اینکه رفتار acf و pacf نمونه ای تا اندازه ای مشابه رفتار نوع نظری آن می باشد، رفتار این توابع را با نوع نظری آنها مقایسه می کنیم. این مقایسه در تشخیص نوع و مرتبه فرآیند مفید است. رفتار توابع خود همبستگی و خود همبستگی جزئی نظری برای مدلهای ایستا در جدول زیر به اختصار آمده است.

PACF	ACF	
بعد از تأخیر p قطع می شود.	**به صورت یک تنزل نمائی یا موج سینوسی میرا به سمت صفر میل می کند.**	**AR(p)**
به صورت یک تنزل نمائی یا موج سینوسی به سمت صفر میل می کند.	**بعد از تأخیر q قطع می شود.**	**MA(q)**
بعد از تأخیر (p-q) به سمت صفر میل می کند.	**بعد از تأخیر (q-p) به سمت صفر میل می کند.**	**ARMA(p,q)**

جدول (۵-۱)

می توان از طریق مقایسه مدل مشاهده شده از توابع نمونه ای acf وpacf با همتای نظری آنها، یک مدل آزمایشی را مشخص کرد.

به عنوان یک قاعده کلی ما فرض خواهیم کرد که یک ضریب خود همبستگی یا ضریب خود همبستگی جزئی صفر است اگر قدر مطلق تخمین آن کمتر از ۲ برابر خطای استاندارد آن باشد.

معمولا مرتبه های p و q از ۳ تجاوز نمی کنند. در مورد فرآیندهای مرکب تشخیص مرتبه های درست p و qبسیار دشوار است و گاهی اوقات به تجربه و مهارت قابل ملاحظه ای نیاز دارد. معمولا اگر acf و pacf هر دو به سمت صفر میل کنند یک فرآیند مرکب شناسایی می شود.

مرحله چهارم : آزمون وجود روند قطعی در مدل

ممکن است در یک سری زمانی روند قطعی وجود داشته باشد. یک مدلARIMA با روند قطعی را بصورت زیر نمایش می دهیم :

$$\varphi(B)w_t = \theta_0 + \theta(B)z_t \quad ; w_t = \nabla^d x_t = (1-B)^d x_t$$

وجود روند قطعی در مدل معادل است با وجود جمله ثابت θ_0 در مدل. بنابراین برای آزمون وجود روند قطعی در مدل باید فرضیه صفر $\theta_0 = 0$ را آزمون کرد.

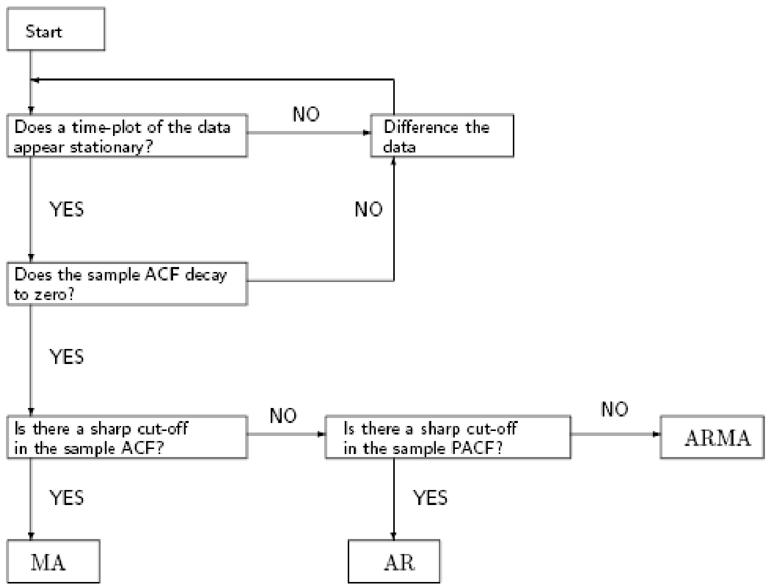
برای یک مدل ناایستا معمولا پارامتر θ_0 را حذف می کنند، بطوری که توانایی نمایش دادن سری با تغییرات تصادفی در سطح، ضریب زاویه یا روند را داشته باشد.

پس از اینکه تبدیلات لازم برای ایستایی واریانس و میانگین انجام شد و سری ایستای w_t بدست آمد، برای آزمون اینکه آیا به وجود جمله ثابت θ_0 در مدل نیازی هست یا

نه، میانگین و خطای معیار سری جدید را محاسبه می کنیم و سپس آماره $T = \frac{\overline{w}}{s_{\overline{w}}}$ را بدست می آوریم و آزمون زیر را انجام می دهیم :

$$\left|\frac{\overline{w}}{s_{\overline{w}}}\right| > t_{n-1,\alpha/2} \Longrightarrow \text{روند قطعی وجود دارد}$$

معمولا در عمل اگر مقدار آماره T بیشتر از ۲ باشد، وجود روند قطعی در مدل را می پذیریم. در مینی تب با توجه به p-value برای جمله ثابت نیز می توان در مورد حضور یا عدم حضور آن در مدل تصمیم گرفت. بطوری که اگر p-value>۰.۰۵ فرضیه $H_0 : \theta_0 = 0$ را می پذیریم در غیر این صورت این فرضیه در سطح معنی داری ۰/۰۵ رد می شود و باید جمله ثابت را در مدل لحاظ کرد.در پایان این بخش توجه شما را به فلوچارت زیر جلب می کنیم. این فلوچارت چگونگی تشخیص نوع فرآیند را نشان می دهد.



شکل ۵-۲ فلوچارت تشخیص مدل آزمایشی

۵-۱-۲ تخمین پارامترهای مدل(برازش مدل)

برازش مدل به معنی برآورد پارامترهای مجهول مدل می باشد. پس از شناسائی مدل آزمایشی، با باز کردن پنجره اصلی ARIMA ،مدل مورد نظر خود را به نرم افزار معرفی

می کنیم و نرم افزار با استفاده از یک روش تکراری پارامترها را برآورد نموده و در قسمت Final Estimates of Parametersآنها را لیست می کند. لازم به ذکر است که در مینی تب بیشترین تعداد پارامتری که می توان برآورد کرد ۱۰ پارامتر می باشد.

همانطور که در فصل قبل نیز گفته شد الگوریتم ARIMA تا ۲۵ تکرار انجام می دهد تا مدل داده شده را برازش دهد.اگر جواب همگرا نباشد، می توان پارامترهای برآورد شده را ذخیره کرده و از آنها به عنوان مقادیر اولیه برای برازش دوم استفاده کرد.

۵-۱-۳ بررسی مناسبت مدل

پس از تشخیص یک مدل مناسب و برآورد پارامترهای آن، سؤالی که باقی می ماند این است که آیا این مدل رسا است یا نه؟ اگر دلایلی از نارسایی شدید وجود داشته باشد، می خواهیم بدانیم مدل در سیکل تکراری بعدی چگونه باید تغییر داده شود.

البته هیچ فرمی از مدل، هرگز حقیقت را بطور مطلق نشان نمی دهد. درنتیجه با داشتن داده های کافی, آزمونهای آماری می توانند مدلهایی را که برای منظور معلومی کاملا رسا هستند، بی اعتبار سازند. برعکس آزمونها می توانند در نشان دادن انحرافات شدید از فرضها ناموفق باشند. زیرا این آزمونها برای انواع انحرافاتی که رخ می دهد حساس نیستند.

بررسی میزان مناسبت مدل باید چنان باشند که مدل را به مخاطره بیاندازد. یعنی بایستی نسبت به انحرافاتی که احتمالا رخ خواهند داد، حساس باشد.

در بررسی مناسبت مدل ما از دو روش که مکمل یکدیگرند استفاده می کنیم :

۱- تجزیه و تحلیل باقیمانده های مدل برازش داده شده.

۲- تجزیه و تحلیل مدل هایی که پارامتر بیشتری دارند. یعنی مدلی که کلی تر از مدل مشخص شده است و این مدل را به عنوان یک حالت خاص در بر می گیرد.

یک حالت جالب توجه از نارسایی مدل که ممکن است تصور کرد وقتی رخ می دهد که "شکل" مدل به همان حال باقی می ماند ولی پارامترها در یک دوره ممتدی از زمان تغییر می کند.

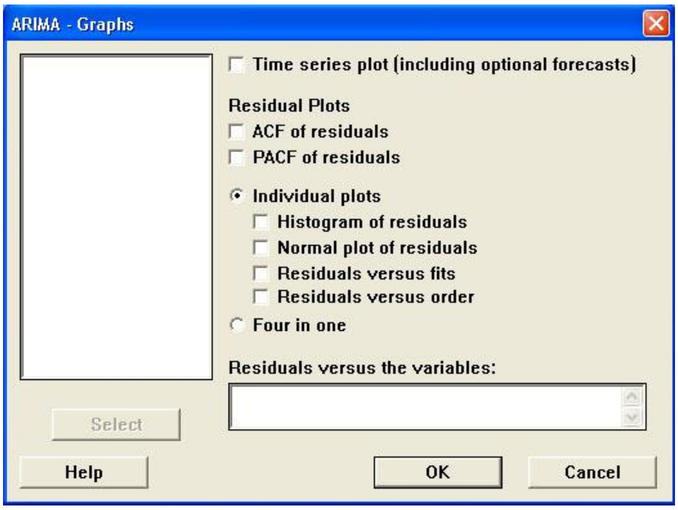
در ادامه این بخش چگونگی بررسی مناسبت مدل را توضیح می دهیم.

۱- تجزیه و تحلیل با قیمانده ها

اگر یک مدل درست تشخیص داده شده باشد، در این صورت باقیمانده های حاصل از برازش آن مدل باید تقریبا دارای خواص متغیرهای تصادفی نرمال مستقل هم توزیع با میانگین صفر و واریانس ثابت باشند.

چنانچه مدل مناسب باشد باید باقیمانده ها فاقد ساختار باشند. یعنی باید با هر متغیر دیگری مانند زمان جمع آوری داده ها یا مقادیر برآورد شده بی ارتباط باشند. وجود هر گونه ساختاری در این قبیل نمودارها حاکی از اثر متغیر مربوطه بر پاسخ است.

یکی از روشهای تجزیه و تحلیل باقیمانده ها بررسی نمودارهای مربوط به باقیمانده ها می باشد. برای این کار در پنجره اصلی ARIMA با انتخاب گزینه Graphsنمودارهای دلخواه را مشخص می کنیم.



شکل ٥ – ٣ : پنجره Graphs

الف) بررسی فرض نرمال بودن باقیمانده ها

در صورتی که مدل انتخابی درست تشخیص داده شود باید خطاها دارای توزیع نرمال، مستقل و همتوزیع باشند. برای بررسی این فرض می توانیم هیستوگرام باقیمانده ها یا نمودار احتمال نرمال آنها را رسم کنیم.

در صورتی که توزیع خطا نرمال باشد، باید در نمودار احتمال نرمال، نقاط در امتداد یک خط مستقیم قرار بگیرند. البته در تأیید خط مستقیم، روی مقادیر مرکزی نسبت به کرانها بیشتر تأکید داریم.

ب) بررسی فرض استقلال باقیمانده ها

برای بررسی تصادفی بودن باقیمانده ها می توان ازacf و pacf باقیمانده ها کمک گرفت. چنانچه این نمودارها روند خاصی را نشان ندهند و از حدود مجاز خود تجاوز نکنند می توان استقلال باقیمانده ها را پذیرفت.

ج) بررسی فرض ثابت بودن واریانس باقیمانده ها

برای بررسی ثابت بودن واریانس باقیمانده ها می توان نمودار باقیمانده ها در مقابل مقادیر برازش شده(Residuals versus fits) و نمودار باقیمانده ها در برابر زمان(Residuals versus order) را رسم کرد. در صورت ثابت بودن واریانس باقیمانده ها این نمودارها باید فاقد ساختار باشند. شکل قیفی در این نمودار ها حاکی از ثابت نبودن واریانس می باشد. همچنین می توان اثر مقادیر مختلف پارامتر تبدیل λ را از طریق روش باکس–کاکس ارزیابی کرد.

د) رسم نمودار باقیمانده ها در طول زمان

این نمودار که تحت عنوان Residuals versus order آمده است برای بررسی این مطلب که آیا باقیمانده ها نمایشگر یک فرآیند تصادفی محض می باشند یا نه مفید است.اگر مدل مناسب باشد انتظار می رود این نمودار در اطراف سطح افقی صفر پراکندگی مستطیلی بدون روندی را نشان دهد.

ر) آزمون پرت–مانتو

در کنار روشهای نموداری یک آزمون مفید برای بررسی کفایت مدل آزمون پرت–مانتو است. این آزمون از خود همبستگی های باقیمانده ها برای بررسی فرضیه صفر توأم $H_0 : \rho_1 = \rho_2 = ... = \rho_k = 0$ با آماره آزمون زیر استفاده می کند.

$$Q = n(n+2)\sum_{h=1}^{k}(n-h)^{-1}\hat{\rho}_h^2$$

که در آن n تعداد مشاهدات می باشد. این آماره آزمون، **آماره Q اصلاح شده**یا همان آماره LBQ است و تحت فرض H_0 تقریبا دارای توزیع χ^2_{k-m} است. m تعداد پارامترهای برآورد شده در مدل می باشد.

هر گاه مقدار آماره Q از مقدار متناظر جدول کی دو بیشتر باشدفرضیه H_0 رد می شود. گاهی فرضیه H_0را **فرضیه کفایت مدل**نیز می نامند.

۲- برازش جامع تر (overfitting)

یک تکنیک که برای بررسی میزان مناسبت مدل می تواند مورد استفاده قرار بگیرد، برازاندن بیش از حد یا برازش جامع تر است. به این ترتیب که پس از تشخیص یک مدل مناسب مدلی عمومی تر را به داده ها برازش می دهیم. این کار مدل تشخیص داده شده را به مخاطره می اندازد. زیرا مدل عمومی تر شامل پارامترهای اضافی است که جهت هایی را که بیم آن می رود انحراف در آن جهت ها باشد در بر می گیرد. در این روش فرض می شود که می توانیم جهتی را که مدل احتمالا در آن جهت نارسا خواهد بود حدس بزنیم.

بنابراین اگر یک ARMA(p,q)به عنوان یک مدل مناسب انتخاب شود، ما مدلهای بزرگتر مانندARMA(p+۱,q)وARMA(p,q+۱) را که مدل اصلی را به عنوان یک حالت خاص شامل می شوند، برازش می دهیم. توجه داشته باشید که نباید بطور همزمان p و qرا افزایش داد. در این صورت مدل اصلی مورد تأیید قرار خواهد گرفت اگر :

۱- برآورد پارامترهای اضافی تفاوت معنی داری با صفر نداشته باشد.

۲- برآورد پارامترهای مشترک با برآوردپارامترهای اولیه آنها اختلاف معنی داری نداشته باشد.

اگر تجزیه و تحلیل قادر به نشان دادن اینکه پارامترهای اضافی مورد نیاز است نباشد، الزاما ثابت نخواهد شد که مدل ما صحیح است. یک مدل فقط این استعداد را دارد که پس از مورد آزمون قرار گرفتن ثابت شود که خوب است یا نه.

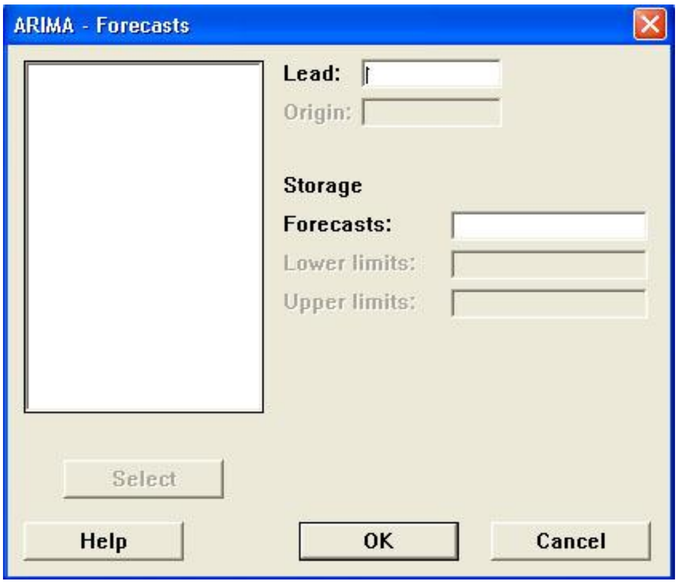
در روش برازاندن بیش از حد به وسیله بسط مدل در یک جهت خاص، فرض می شود که می دانیم باید از چه نوع انحرافاتی بیم داشته باشیم. روشهایی که به چنین اطلاعاتی کمتر بستگی دارد مبتنی بر تجزیه و تحلیل باقیمانده ها است.

بر پایه نتایج این تحلیلها چنانچه مدل پیشنهادی نامناسب باشد، باید مدل دیگری را در نظر بگیریم. اما چنانچه بعد از مراحل فوق به دو یا چند مدل مناسب دست یافتیم و تحلیل باقیمانده ها برای تشخیص اینکه کدام مدل بهتر است کافی نبود، به محکهای دیگری برای شناسایی بهترین مدل نیاز داریم. یکی از این محکها معیار اطلاعاتی آکائیک است که در ادامه این بخش توضیح داده خواهد شد.

٥-١-٤ پیش بینی

پس از مشخص شدن مدل مناسب می توان از آن برای پیش بینی رفتار آینده سری استفاده کرد. برای این امر در پنجره اصلی ARIMA برروی گزینه Forecasts کلیک می کنیم.در پنجره ظاهر شده درقسمت Lead باید تعداد پیش بینی ها را مشخص کرد.

در قسمتOrigin نقطه آغاز پیش بینی را مشخص می کنیم. چنانچه در این قسمت چیزی ننویسم، مینی تب پیش بینی ها را از انتهای مشاهدات تولید میکند.



شکل ٥ – ٤ : پنجره پیش بینی ARIMA

۵– ۲ معیار اطلاعاتی آکائیک (AIC)

اگر برای یک مجموعه داده چند مدل قابل قبول وجود داشته باشد، معمولا ملاک انتخاب مدل بهتر برمبنای آماره های خلاصه شده ای است که از باقیمانده های مدل های برازش شده محاسبه می شوند.

روش دیگر بر پایه خطاهای پیش بینی است که از پیش بینی های خارج نمونه محاسبه می گردد. در این روش اغلب با استفاده از بخش اول سری یک مدل مناسب تشخیص داده می شود و سپس از بخش باقیمانده سری به عنوان یک دوره قابل دسترس برای ارزیابی پیش بینی استفاده می شود.

در اینجا ما با استفاده از معیار آکائیک که برمبنای باقیمانده های مدل برازش شده می باشد, سعی می کنیم مدل مناسب را از بین چند مدل انتخاب کنیم.

فرض کنید یک مدل آماری M پارامتری به داده ها برازش نموده ایم. در این صورت معیار اطلاعاتی آکائیک بصورت زیر تعریف می شود:

AIC(M) =–۲(لگاریتم درستنمائیماکزیمم) + ۲M

این معیار برای مدلARMA با تعداد مشاهدات nبصورت زیر خواهد بود:

$$AIC(M) = n \ln \hat{\sigma}_e^2 + 2M$$

که σ_e^2 واریانس مانده های مدل و $\hat{\sigma}_e^2$ برآورد درستنمائی ماکزیمم σ_e^2 است. مدل بهترمدلی است که برای آن مقدار AIC مینیمم باشد.

در مینی تب این معیار مستقیما محاسبه نمی شود. اما از آنجا که میانگین مجموع مربعات خطا یک برآورد نااریب برای واریانس خطا می باشد، برای هر مدل می توان

معیار آکائیک را بصورت زیر تعریف کرد و سپس مدلی را که برای آن این معیار کوچکتر باشد انتخاب نمود.

$$AIC = n\ln(MSE) + 2(p+q)$$

۵- ۳ استفاده از باقیمانده ها برای اصلاح مدل

وقتی تابع خود همبستگی باقیمانده های مدل برازش شده نارسایی مدل را نشان دهد، می خواهیم بدانیم که چگونه می توان این مدل را اصلاح کرد.فرض کنید در مدل $\varphi_0(B)\nabla^{d_0}x_t = \theta_0(B)l_t$بنظر می رسد که باقیمانده های l_tغیر تصادفی باشند. با استفاده از تابع خود همبستگی اکنون ممکن است یک مدل بصورت زیر را برای سری l_t بکار ببریم.

$$\bar{\varphi}(B)\nabla^{\bar{d}}l_t = \bar{\theta}(B)a_t$$

پس از حذف l_tبین دو معادله فوق به یک مدل جدید می رسیم :

$$\varphi_0(B)\bar{\varphi}(B)\nabla^{d_0}\nabla^{\bar{d}}x_t = \theta_0(B)\bar{\theta}(B)a_t$$

اکنون این مدل جدید را به داده ها برازش داده و مناسبت آن را بررسی می کنیم.

مثال ۵-۱

فرض کنیدکه یک سری به غلط بصورت یک فرآیندIMA(۰,۱,۱)بصورت $\nabla x_t = (1+0.6B)l_t$ تشخیص داده شده است. همچنین فرض کنید که یک مدل بصورت $\nabla l_t = (1-0.8B)a_t$ برای سری با قیمانده ها تشخیص داده شده است. در اینصورت پس از حذف l_tبین دو معادله فوق خواهیم داشت :

$$\nabla^2 x_t = (1-0.2B-0.48B^2)a_t$$

که پیشنهاد می کند اکنون یک فرآیند IMA(۰,۲,۲)بایستی مورد توجه قرار گیرد.

مثال ۵-۲

فرض کنید یک مدل AR(۱) بصورت زیر به داده ها برازش داده ایم:

$$(1-\varphi_1 B)(z_t-\mu)=b_t$$

حال با بررسی باقیمانده های حاصل از برازش این مدل متوجه شده ایم که باقیمانده ها بجای آنکه رفتاری مانند یک فرآیند تصادفی محض داشته باشند، مانند یک مدل MA(۱) به شکل $b_t=(1-\theta_1 B)a_t$ رفتار می کنند.

بنابراین ما در مرحله بعد یک مدل ARMA(۱,۱) را به داده ها برازش می دهیم و مراحل تکراری ساختن مدل را تا حصول یک مدلی رضایت بخش ادامه می دهیم.

۵- ٤ مقادیر گمشده در سری های زمانی

داشتن مقادیر گمشده در کلیه سری های زمانی امری متعارف است. بسیاری از سری ها داده هایی را شامل می شوند که شناخته شده یا مظنون به داشتن مشاهدات اشتباه می باشند که ممکن است کنار گذاشتن آنها ارجح تر از کوشش برای کم کردن تأثیرشان بر روی میانگین سایر مقادیر باشد.

یک راه ساده برای رفتار کردن با مقادیر گمشده در سری های زمانی ایستا، جانشین کردن میانگین مقادیر گمشده به جای آنها است. بعضی روشها در سری های زمانی بر اساس کم توجهی به مقادیر گمشده متمرکز است که به تبحر تشخیص فرد بستگی دارد.

درون یابی سری های زمانی

عبارت درون یابی را برای فرآیند برآورد مشاهدات گمشده در سری زمانی بکار می بریم. برای درون یابی باید یک سری زمانی گسسته که معمولا در یک دنباله از نقاط با فاصله گذاری یکسان مشاهده شده اند را در نظر گرفت.

معمول ترین وضعیت این است که یک تک وقفه از یک یا چند مشاهده را داشته باشیم که در این صورت یک دنباله غیر منفصل از مشاهدات را که بلافاصله قبل و بعد از وقفه قرار دارند را خواهیم داشت. ممکن است بیش از یک وقفه موجود باشد که هر وقفه می تواند به طور جداگانه پر شده یا درون یابی شود.

در موارد زیر درون یابی مفید واقع می شود:

۱- وقتی داده ها بصورت اشتباه ثبت شوند و یا اینکه در داده ها نقاط پرت وجود داشته باشد مانند سری های اقتصادی و صنعتی. در این صورت برای دوری از خطاهای بکارگیری داده ها، تعویض نقاط پرت با مقادیر درون یابی شده مناسب، ضروری به نظر می رسد.

۲- وقتی یک پیشامد غیر عادی یک سری زمانی را تحت تأثیر قرار می دهد. مثلا بعضی عوامل ممکن است تولید را تحت تأثیر قرار دهند و یا تعطیلات ملی یک روزه ممکن است در میزان مصرف برق تأثیر داشته باشد.

برای بدست آوردن درون یابهای مناسب، ابتدا باید یک مدل برای نمایشسری ساخته و مدل را به داده هایی که موجودند برازش داد(یعنی پارامترهای آن را برآورد کرد) و سپس چک کنیم که آیا مدل بطور مؤثری نماینگر داده ها می باشد یا خیر؟ در نهایت این مدل برای بدست آوردن درون یابها بکار گرفته می شود.

۵- ۵ مثالهایی از مدلسازی سری های زمانی

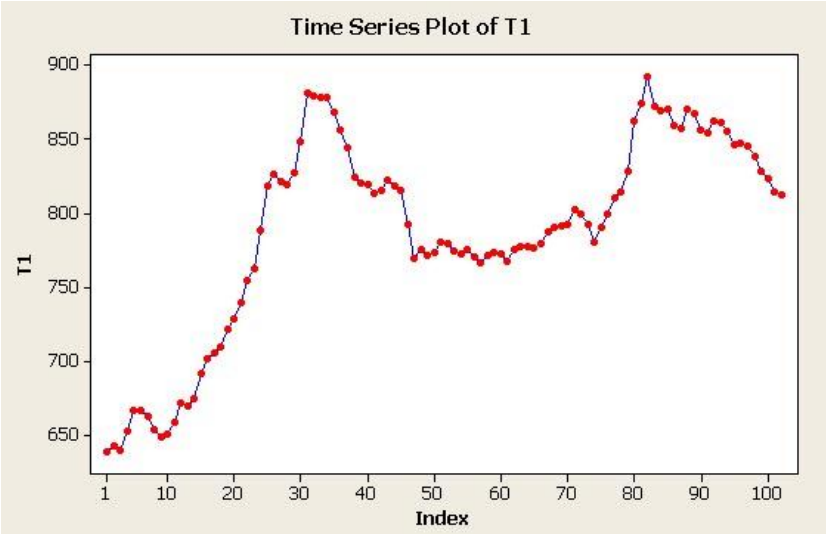
این قسمت شامل ۱۵ مثال جالب از مدل سازی سری های زمانی می باشد. تنوع مثالها کابرد وسیع سری های زمانی را نشان می دهد. تکنیک مورد استفاده در مدل سازی این سری ها همان روش تکراری باکس–جنکینز می باشد. در انتهای فصل نیز چند تمرین آمده است که از خواننده انتظار می رود با توجه به مثالهای متعددی که در این فصل آمده است به راحتی بتواند مدلمناسب را برای هر سری تشخیص دهد و سپس مناسبت آن را بررسی کند.

مثال ۱

داده های مربوط به سری مصالح خط راه آهن آمریکااز ژانویه ۱۹٦۸ تا ژوئن ۱۹۷٦ که در پیوست این کتاب با نام T۱ آمده است را در نظر می گیریم. می خواهیم با استفاده از روش تکراری باکس- جنکینز مدلی مناسب از خانواده مدلهای ARIMA را برای این سری شناسایی کرده و سپس پارامترهای آن مدل را برآورد کنیم و در نهایت مناسبت و شایستگی مدل را مورد بررسی قرار دهیم.

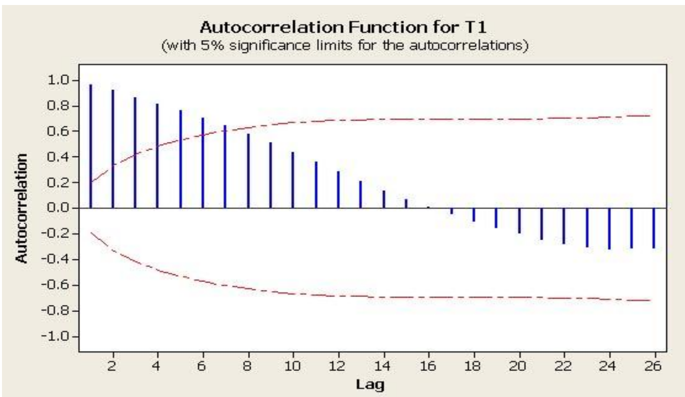
تشخیص مدل آزمایشی

در اولین گام با رسم نمودار سری زمانی و نمودار acf برای این مشاهدات به بررسی ایستایی سری می پردازیم.



شکل ۵– ۵ :نمودار سری زمانی مصالح خط راه آهن

این نمودار ناایستایی سری در میانگین را نشان می دهد. نمودار acf نیز این مطلب را تأیید می کند.



شکل ۵– ٦ : همبستگی نگار سری مصالح خط راه آهن

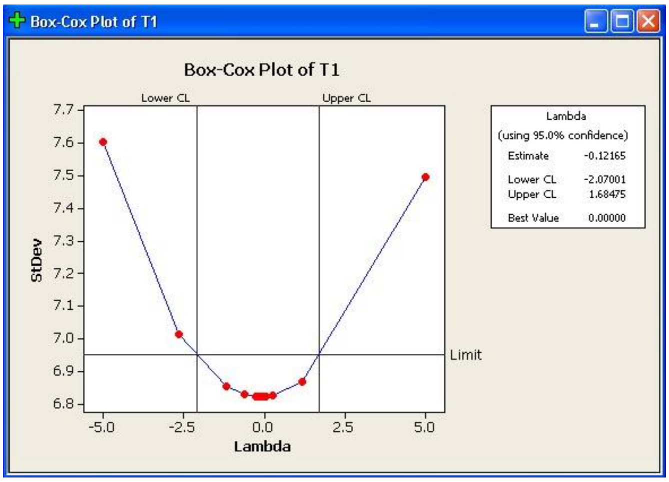
همانطور که ملاحظه می شود مقادیر تابع خود همبستگی بسیار کند به سمت صفر میل می کنند. این نوع رفتار در تابع خود همبستگی دال بر ناایستایی میانگین سری مربوطه است. بنابراین لازم است برای حصول ایستایی سری اولیه را یک بار تفاضلی کنیم. البته ابتدا باید پایایی واریانس سری را مورد بررسی قرار داد.

بررسی پایایی واریانس

قبل از هر گونه تبدیلی به منظور پایا کردن میانگین سری باید از پایایی واریانس آن مطمئن شویم. برای انجام این کار در مینی تب ابتدا از منوی Editor گزینه Enablecommands را انتخاب می کنیم تا خط فرمان در منوی session فعال شود. سپس فرمان مربوط به اجرای رویه باکس–کاکس را بصورت زیر تایپ کرده و کلید اینتر را برای تأیید فرمان می فشاریم.

MTB > BOXCOX T۱ ۱

نتیجه اجرای این رویه در مینی تب به شکل زیر می باشد :



شکل ۵– ۷ : نمودار تبدیلات باکس– کاکس برای سری T۱

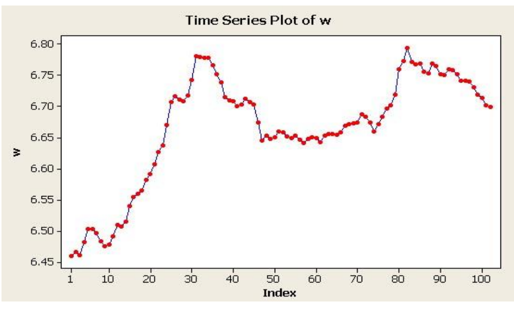
همانطور که ملاحظه می شود در کادر سمت راست برای پارامتر تبدیل λ، برآورد حدود اطمینان ۹۵ درصد و بهترین مقدار پیشنهادی ارائه شده است.

در مورد این داده ها مینی تب بهترین مقدار پارامتر تبدیل λ را صفر پیشنهاد می کند. مقدار صفر برای پارامتر λ اشاره بر لزوم تبدیل لگاریتمی داده ها دارد.

توجه

حدود اطمینان ۹۵ درصد شامل همه مقادیر λ می باشد که انحراف استانداردشان کمتر یا مساوی با مقداری است که خط افقی نشان می دهد. بنابراین هر مقداری از λ که انحراف استانداردی نزدیک به خط افقی داشته باشد، نیز یک مقدار قابل قبول و منطقی برای استفاده جهت تبدیل داده ها خواهد بود.

در مورد مثال فوق چون عدد یک داخل حدود اطمینان ۹۵ درصد قرار دارد، می توان آن را به عنوان یک مقدار قابل قبول پارامتر تبدیل پذیرفت. بنابراین می توان از تبدیل داده ها صرف نظر کرد. زیرا با توجه به جدول (۲-۱) عدد یک برای پارامتر تبدیل λ به معنی عدم نیاز به تبدیل داده ها است. ما نیز در این مثال از تبدیل داده ها صرف نظر می کنیم. البته در صورت تبدیل داده ها، نمودار سری زمانی برای داده های تبدیل شده تفاوت چندانی با نمودار داده های اولیه نخواهد داشت. نمودار سری تبدیل شده که آن را w نامیده ایم در زیر رسم شده است.

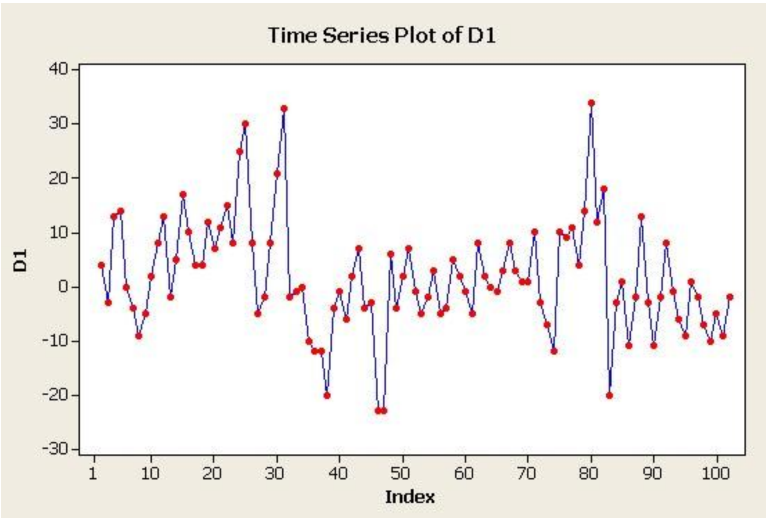


شکل ۵- ۸ : نمودار سری لگاریتمی

تفاضلی کردن

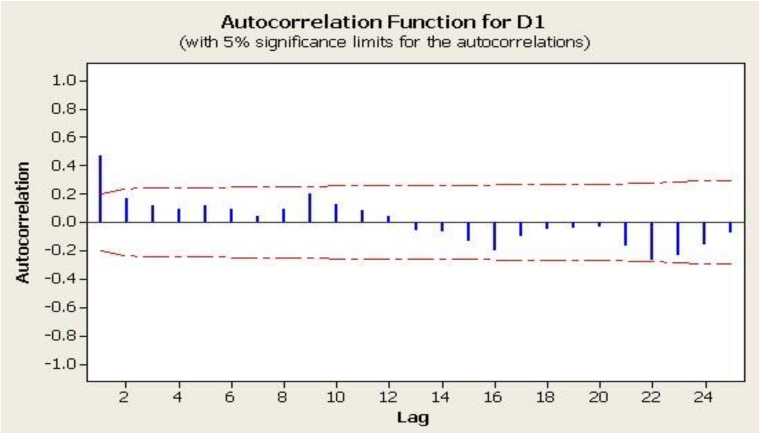
برای ایستایی سری در میانگین لازم است که آن را یک بار تفاضلی کنیم. برای انجام این کار در مینی تب از منوی Time Seriesگزینه Differences را انتخاب می کنیم. در قسمت Lag عدد یک را وارد می کنیم که مرتبه تفاضلی کردن را تعیین می کند.

سری تفاضلی شده راD۱۱نامیده ایم. حال با رسم نمودار سری زمانی برای سری تفاضلیD۱ ایستایی آن را بررسی می کنیم.



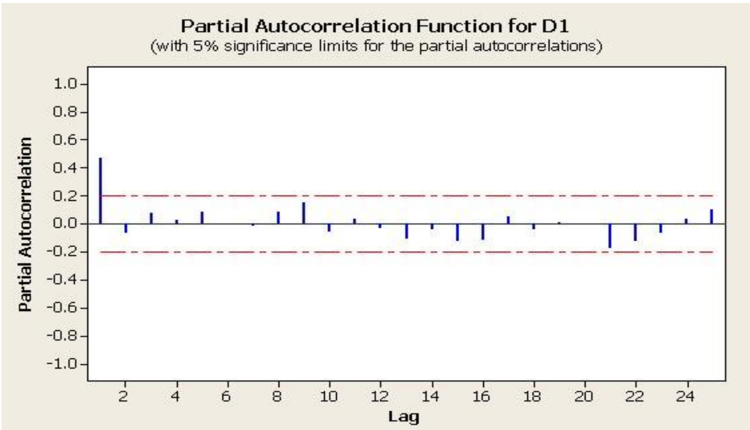
شکل ۵– ۹ : نمودار سری تفاضلی شده مصالح خط راه آهن

نمودار سری تفاضلی شده، ایستایی در میانگین را نشان می دهد. برای بررسی بیشتر acf سری تفاضلی شده را رسم می کنیم. همانطور که ملاحظه می شود acf بصورت نمائی و به سرعت تنزل پیدا می کند و این خود دلیلی بر ایستایی سری تفاضلی شده می باشد.



شکل ۵– ۱۰ : نمودار acf سری تفاضلی شده مصالح خط راه آهن

برای کمک به شناسایی مدل آزمایشی pacf سری تفاضلی شده را نیز رسم می کنیم.



شکل ۵– ۱۱ : نمودار pacf سری تفاضلی شده مصالح خط راه آهن

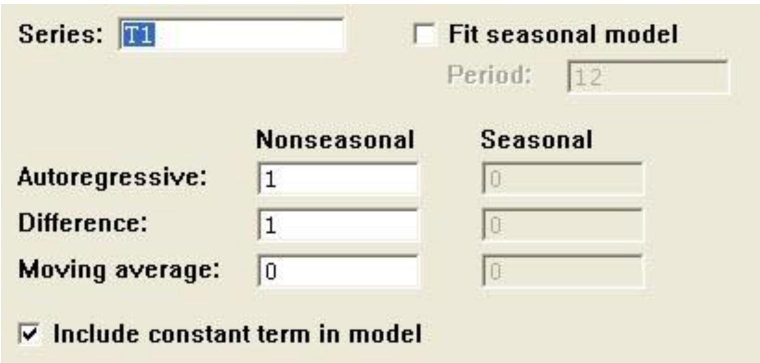
همانطور که ملاحظه می شود بجز اولین مقدار، سایر مقادیر تابع خود همبستگی جزئی داخل حدود استاندارد قرار گرفته اند. بنابراین می توان گفت مقادیر تابع خود همبستگی جزئی از تأخیر اول به بعد همگی صفر هستند(اختلاف معنی داری با صفر ندارند). این مطلب دلالت بر یک فرآیند AR(۱) برای سری تفاضلی شده دارد.

برازش مدل(۱)ARبه سری تفاضلی شده معادل برازش مدل(۱,۱,۰)ARIMA به سری اصلی می باشد. همانطور که در فصل سوم گفته شد، با توجه به این که در مدل (۱,۱,۰)ARIMAجمله MAنداریم، این مدل را بصورت (۱,۱)ARI نشان می دهیم.

بنابراین ما با بررسی نمودار سری زمانی، acfو pacf آن مدل آزمایشی(۱,۱)ARI را شناسایی کرده ایم. در مرحله بعد این مدل را به داده ها برازش می دهیم. برازش مدل به معنی برآورد پارامترهای مجهول مدل می باشد.

برازش مدل(۱،۱)ARI

برای انجام این کار در مینی تب پنجره اصلی ARIMA را باز کرده و آن را به شکل زیر تکمیل می کنیم :



شکل ۵– ۱۲ : تکمیل پنجره ARIMA برای برازش مدل(۱,۱)ARI

توجه:

برازش مدل با جمله ثابت، حالت پیش فرض نرم افزار می باشد و همانطور که ملاحظه می شود در پنجره اصلی ARIMA همیشه عبارت Include constant term in model دارای چک مارک می باشد. ما نیز پیش فرض مینی تب را می پذیریم و مدل را با جمله ثابت برازش می دهیم. در مرحله بعد با توجه به آماره t و مقدار $p-value$ در مورد حضور یا عدم حضور جمله ثابت در مدل تصمیم می گیریم. به این ترتیب که چنانچه

در قسمت Final Estimates of Parametersدر ردیف Constant که مربوط به جمله ثابت می باشد، مقدار p-value بیشتر از ۰.۰۵ باشد، فرضیه صفر بودن جمله ثابت را با ۹۵ درصد اطمینان می پذیریم. در غیر این صورت وجود جمله ثابت را در مدل لازم تشخیص می دهیم.

همچنین می توان از آماره tبرای آزمون این فرض استفاده کرد. بدین ترتیب که مقدار مشاهده ای این آماره را با مقدار متناظر آن از جدول توزیع t–استودنت مقایسه می کنیم، چنانچه مقدار مشاهده شده از مقدار متناظر جدول در سطح ۰.۰۵ و با $n-1$ درجه آزادی کمتر بود، می توان گفت که پارامتر مورد نظر اختلاف معنی داری با صفر ندارد و می توان آن پارامتر را در مدل نادیده گرفت.البته به طور تقریبی می توان گفت چنانچه مقدار مشاهده شده آماره tبرای یک پارامتر بیشتر از۲باشد، آنگاه آن پارامتر اختلاف معنی داری با صفر دارد و در غیر این صورت، یعنی چنانچه مقدار این آماره کمتر از ۲ باشد، می توان آن پارامتر را نادیده گرفت.

در صورت پذیرفتن فرضیه $H_0 : \theta_0 = 0$، یک بار دیگر پنجره ARIMA را باز کرده و مدل مورد نظر را برازش می دهیم اما این بار چک مارک عبارت Final Estimates of Parameters را برمی داریم تا مدلی بدون جمله ثابت داشته باشیم.

خروجی اجرای رویه فوق در پنجره session بصورت زیر می باشد :

ARIMA Model: T۱

```
Estimates at each iteration

Iteration       SSE   Parameters
        ۰   ۹۹۱٤.۰٦   ۰.۱۰۰   ۱.٦۳۲
        ۱   ۸۹٦۳.۲٥   ۰.۲٥۰   ۱.۳۳٦
        ۲   ۸٤۹۸.٦۰   ۰.٤۰۰   ۱.۰٤۱
```

٣ ٨٤٤٨.٢٤ ٠.٤٦٥ ٠.٩١١

٤ ٨٤٤٨.١١ ٠.٤٦٨ ٠.٩٠٤

٥ ٨٤٤٨.١١ ٠.٤٦٨ ٠.٩٠٤

Relative change in each estimate less than ٠.٠٠١٠

Final Estimates of Parameters

Type	Coef	SE Coef	T	P
AR ١	٠.٤٦٨٣	٠.٠٨٨٩	٥.٢٧	٠.٠٠٠
Constant	٠.٩٠٤١	٠.٩١٩١	٠.٩٨	٠.٣٢٨

Differencing: ١ regular difference

Number of observations: Original series ١٠٢, after differencing ١٠١

Residuals: SS = ٨٤٤٧.٢٠ (backforecasts excluded)

MS = ٨٥.٣٣ DF = ٩٩

Modified Box-Pierce (Ljung-Box) Chi-Square statistic

Lag	١٢	٢٤	٣٦	٤٨
Chi-Square	٦.٤	١٩.٣	٢٨.١	٣٨.٢
DF	١٠	٢٢	٣٤	٤٦
P-Value	٠.٧٨٢	٠.٦٣٠	٠.٧٥١	٠.٧٨٦

همانطور که ملاحظه می شود مقدار آماره tبرای جمله ثابت۹۸ درصد می باشد که از ۲ کمتر است. بنابراین نیازی به لحاظ کردن جمله ثابت در مدل نمی باشد.

همچنین مقدار $p-value$ برای جمله ثابت برابر ۰.۳۲۸ می باشد که بیشتر از ۰.۰۵ است. بنابراین فرضیه $H_0:\theta_0=0$ را نمی توان رد کرد.

برازش مدل(۱،۱)ARIبدون جمله ثابت

اکنون مدل فوق را یک بار دیگر به داده ها برازش می دهیم. اما این بار با برداشتن چک مارک عبارت Include constant term in modelدر پنجره ARIMA جمله ثابت را از مدل حذف می کنیم. نتیجه در پنجره session بصورت زیر می باشد :

ARIMA Model: T۱

```
Estimates at each iteration

Iteration        SSE   Parameters
        ۰    ۱۰۱۵۳.۳        ۰.۱۰۰
        ۱     ۹۱۲۹.۸        ۰.۲۵۰
        ۲     ۸۶۰۵.۵        ۰.۴۰۰
        ۳     ۸۵۳۰.۱        ۰.۴۷۸
        ٤     ۸۵۲۹.۹        ۰.٤۸۲
        ٥     ۸۵۲۹.۹        ۰.٤۸۳

Relative change in each estimate less than ۰.۰۰۱۰

Final Estimates of Parameters
```

```
Type       Coef  SE Coef      T       P
AR   ١  ٠.٤٨٢٥   ٠.٠٨٧٦  ٥.٥١  ٠.٠٠٠

Differencing: ١ regular difference
Number of observations:  Original series ١٠٢, after
differencing ١٠١
Residuals:    SS =  ٨٥٢٧.٠٢ (backforecasts excluded)
MS =  ٨٥.٢٧   DF = ١٠٠

Modified Box-Pierce (Ljung-Box) Chi-Square statistic

Lag            ١٢     ٢٤     ٣٦     ٤٨
Chi-Square    ٦.٤   ١٨.٩   ٢٧.٧   ٣٧.٨
DF             ١١     ٢٣     ٣٥     ٤٧
P-Value     ٠.٨٤٦  ٠.٧٠٩  ٠.٨٠٦  ٠.٨٢٩
```

همانطور که ملاحظه می شود مقدار برآورد شده برای تنها پارامتر مدل عبارت است از ٠.٤٨٢٥ با انحراف استاندارد ٠.٠٨٧، همچنین مقدار آماره tو مقدار $p-value$ نیز حاکی از معنی دار بودن این پارامتر می باشد.بنابراین مدل آزمایشی ما همانطور که در فصل سوم هم دیدیم بصورت زیر خواهد بود :

$$x_t = (1+\alpha_1)x_{t-1} - \alpha_1 x_{t-2} + z_t$$

باتوجه به آنکه برآورد پارامتر α برابر ۰.٤٨٢٥ می باشد ، خواهیم داشت :

$$x_t = 1.48x_{t-1} - 0.48x_{t-2} + z_t$$

توجه: در صورت برازش مدل شناسایی شده به داده های لگاریتمی نیز به نتایج مشابهی خواهیم رسید.

ARIMA Model: log T۱

```
Final Estimates of Parameters

Type        Coef  SE Coef      T       P
AR   ۱  ۰.٤٩٢٨    ۰.۰٨٧۰   ٥.٦٦   ۰.۰۰۰

Differencing: ۱ regular difference
Number of observations:  Original series ۱۰۲, after
                         differencing ۱۰۱
Residuals:    SS =  ۰.۰۱۳۲۰۳٨ (backforecasts excluded)
              MS =  ۰.۰۰۰۱۳۲۰   DF = ۱۰۰

Modified Box-Pierce (Ljung-Box) Chi-Square statistic

Lag               ۱۲      ۲٤      ۳٦      ٤٨
```

Chi-Square	٦.٦	٢٠.٣	٢٩.٩	٤٠.١
DF	١١	٢٣	٣٥	٤٧
P-Value	٠.٨٢٩	٠.٦٢٣	٠.٧١٢	٠.٧٥٣

بررسی مناسبت مدل

همانطور که گفته شد برای بررسی مناسبت مدل از دو روش که مکمل یکدیگرند استفاده می کنیم :

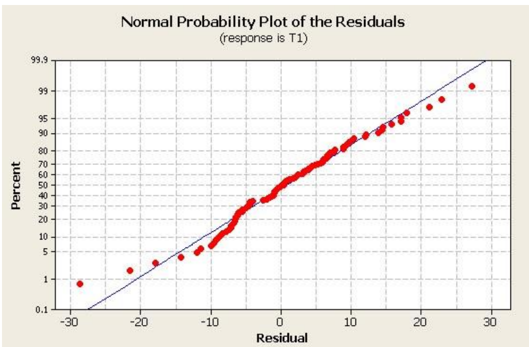
۱- تجزیه و تحلیل باقیمانده های مدل برازش شده.
۲- تجزیه و تحلیل مدل هایی که پارامتر بیشتری دارند(overfitting).

تجزیه و تحلیل باقیمانده ها

همانطور که گفته شد تجزیه و تحلیل باقیمانده ها به کمک نمودارهای مربوط به باقیمانده ها و همچنین آزمون پرت- مانتو انجام می شود.

الف) بررسی فرض نرمال بودن باقیمانده ها

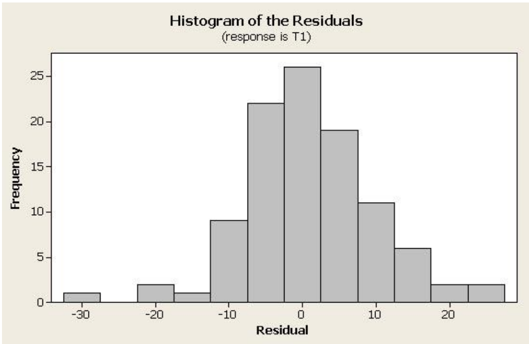
برای بررسی این فرض نمودار احتمال نرمال باقیمانده ها و هیستوگرام باقیمانده ها را رسم می کنیم.



شکل ۵- ۱۳ : نمودار هیستوگرام و نمودار

احتمال نرمال باقیمانده ها

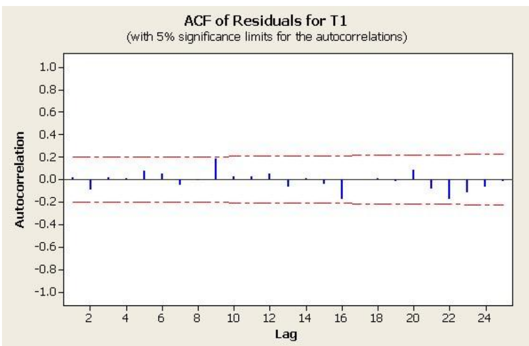
.



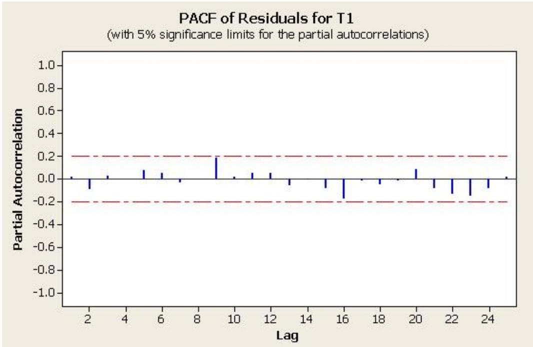
همانطور که ملاحظه می شود در نمودار احتمال نرمال، نقاط تقریبا در امتداد یک خط راست قرار گرفته اند و این نشان می دهد که باقیمانده های حاصل از برازش مدل ARI(۱,۱) بصورت نرمال توزیع شده اند. نمودار هیستوگرام نیز نشان می دهدکه با قیمانده ها تقریبا بصورت نرمال توزیع شده اند.

ب) بررسی فرض استقلال باقیمانده ها

برای بررسی این فرض acf و pacf باقیمانده ها را بررسی می کنیم.



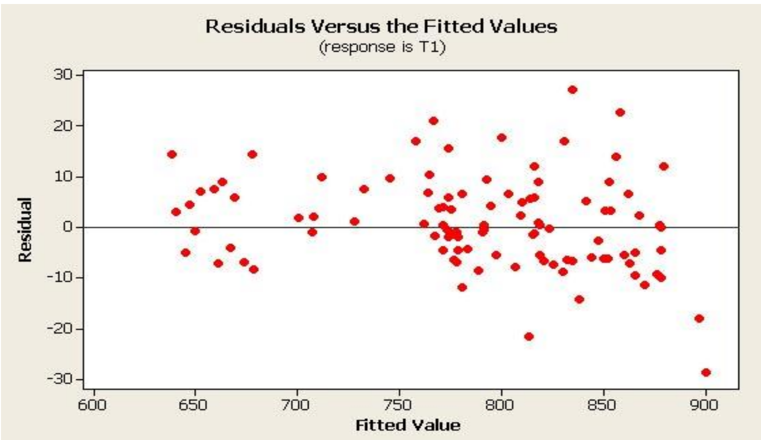
شکل ۵– ۱٤ : acf و pacf باقیمانده های سری مصالح خط راه آهن



همانطور که ملاحظه می شود هیچ یک از خود همبستگی ها معنی دار نیستند. زیرا از حدود استانداردشان تجاوز نکرده اند. این به معنی ناهمبسته بودن باقیمانده ها و تصادفی بودن آنها می باشد.

ج)بررسی فرض ثابت بودن واریانس باقیمانده ها

بررسی این فرض به کمک نمودار باقیمانده ها در برابر مقادیر برازش داده شده صورت می گیرد. چنانچه این نمودار ساختار خاصی را نشان ندهد مثلا به شکل قیفی نباشد، می توان فرض ثابت بودن واریانس باقیمانده ها را پذیرفت.

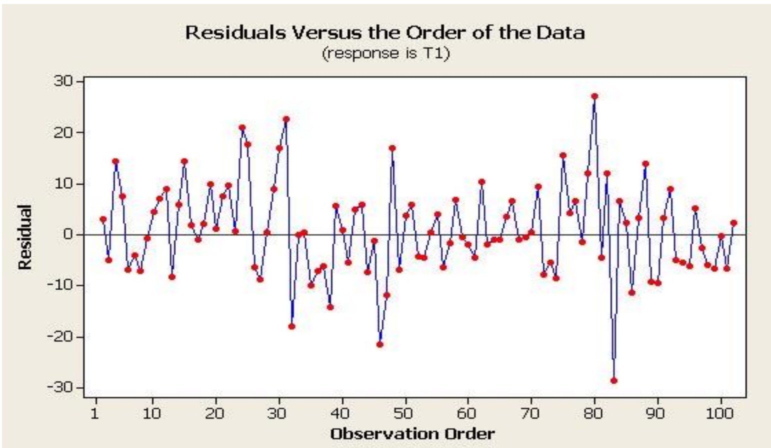


شکل ٥ – ١٥ :نمودار باقیمانده ها در مقابل مقادیر برازش شده

همانطور که ملاحظه می شود این نمودار ساختار خاصی را نشان نمی دهد. بنابراین می توان فرض ثابت بودن واریانس باقیمانده ها را پذیرفت.

د) رسم نمودار باقیمانده ها در برابر زمان

همانطور که گفته شد اگر مدل مناسب باشد انتظار می رود این نمودار در اطراف سطح افقی صفر پراکندگی مستطیلی بدون روندی را نشان دهد. چنانچه رفتار این نمودار شبیه رفتار یک فرآیند تصادفی محض با میانگین صفر و واریانس ثابت باشد، آنگاه می توان مدل برازش داده شده را تأیید نمود.



شکل ٥ – ١٦ : نمودار با قیمانده ها در طول زمان

با توجه به نمودار فوق می توان صحت مدل برازش داده شده را تأیید کرد.

ر) آزمون پرت- مانتو

روش رسمی تر برای بررسی مناسبت مدل که بر مبنای خود همبستگی های باقیمانده ها می باشد, آزمون پرت–مانتو است.نتایج مربوط به این آزمون در انتهای خروجی مینی تب آمده است :

Modified Box-Pierce (Ljung-Box) Chi-Square statistic

Lag	١٢	٢٤	٣٦	٤٨
Chi-Square	٦.٤	١٨.٩	٢٧.٧	٣٧.٨
DF	١١	٢٣	٣٥	٤٧
P-Value	٠.٨٤٦	٠.٧٠٩	٠.٨٠٦	٠.٨٢٩

همانطور که ملاحظه می شود مقدار $p-value$ برای تمامی تأخیر های فوق بیشتر از ٠.٠٥ می باشد. به عنوان مثال در تأخیر ٢٤ مقدار $p-value$ برابر است با ٠.٧٠٩ که بیشتر از ٠.٠٥ می باشد. در نتیجه می توان فرض صفر بودن همه خود همبستگی ها تا تأخیر ٢٤ را پذیرفت($H_0 : \rho_1 = \rho_2 = ... = \rho_{24} = 0$).

بنابراین می توان گفت باقیمانده های حاصل از برازش مدلی ARI(١,١) ناهمبسته هستند.

روش دیگر برای قضاوت در مورد فرضیه H_0 مقایسه مقدار مشاهده شده برای آماره کی دو با مقدار مناسب آن در جدول کی دو می باشد. چنانچه مقدار مشاهده شده آماره از مقدار جدول کمتر باشد فرضیه H_0 پذیرفته می شود. به عنوان مثال مقدار آماره برای تأخیر ٢٤ برابر ١٨.٩ می باشد. این مقدار را با مقدار متناظر از توزیع کی دو با ٢٣ درجه آزادی در سطح ٠.٠٥ درصد مقایسه می کنیم. با توجه به آنکه $\chi^2_{0.05,23} = 35.17$ و چون ١٨.٩>٣٥.١٧ بنابراین فرضیه H_0 با ٩٥ درصد اطمینان پذیرفته می شود.

برازش جامع تر

در این قسمت با برازش دو مدل کلیتر ARI(۲,۱)و ARIMA(۱,۱,۱)به کنکاش بیشتر در مورد شایستگی مدل اولیه ARI(۱,۱) می پردازیم.

الف) برازش مدل ARI(۲،۱)

در صورت برازش این مدل به داده ها خواهیم داشت :

ARIMA Model: T۱

Estimates at each iteration

Iteration	SSE	Parameters	
۰	۹۹٤۲.۱۰	۰.۱۰۰	۰.۱۰۰
۱	۹۰۷۹.۸٤	۰.۲٥۰	۰.۰٤٥
۲	۸٦۰٦.۸٦	۰.٤۰۰	−۰.۰۱۱
۳	۸٥۰۷.۲۹	۰.٥۰۲	−۰.۰٤۹
٤	۸٥۰٦.۹۸	۰.٥۰۷	−۰.۰٥۲
٥	۸٥۰٦.۹۸	۰.٥۰۸	−۰.۰٥۲
٦	۸٥۰٦.۹۸	۰.٥۰۸	−۰.۰٥۲

Relative change in each estimate less than ۰.۰۰۱۰

```
Final Estimates of Parameters

Type        Coef   SE Coef       T       P
        AR   ١    ٠.٥٠٧٧    ٠.١٠٠٤    ٥.٠٦  ٠.٠٠٠
        AR   ٢  -٠.٠٥٢١    ٠.١٠٠٧  -٠.٥٢  ٠.٦٠٦

Differencing: ١ regular difference
Number of observations:  Original series ١٠٢, after
differencing ١٠١
Residuals:    SS =  ٨٥٠٣.٢٩ (backforecasts excluded)
MS =  ٨٥.٨٩  DF = ٩٩

Modified Box-Pierce (Ljung-Box) Chi-Square statistic

Lag              ١٢      ٢٤      ٣٦      ٤٨
Chi-Square     ٦.٠    ١٨.٣    ٢٧.٣    ٣٦.٧
DF               ١٠      ٢٢      ٣٤      ٤٦
P-Value       ٠.٨١٤   ٠.٦٨٨   ٠.٧٨٤   ٠.٨٣٤
```

همانطور که ملاحظه می شود برآورد پارامتر اضافی α_2 که مینی تب آن را با AR٢نشان می دهد اختلاف معنی داری با صفر ندارد. زیرا مقدار آماره tبرای این پارامتر اضافی ٠.٥٢- است که کمتر از ٢ می باشد. همچنین مقدار $p-value$ برابر ٠.٦٠٦می باشد که بیشتر از ٠.٠٥ است و فرضیه $H_0: \alpha_2 = 0$ را در سطح ٠.٠٥ تأیید می کند. بنابراین

نیازی به لحاظ کردن پارامتر اضافی α_2 در مدل نمی باشد و مدل اولیه مورد تأیید قرار می گیرد.

ب) برازش مدل ARIMA(۱،۱،۱)

ممکن است لحاظ کردن یک جمله MA در مدل باعث بهبود مدل شود اما ما در مدل اولیه آن را نادیده گرفته باشیم. بنابراین مدل را از طرف راست توسعه می دهیم و یک پارامتر میانگین متحرک را در مدل لحاظ می کنیم. ما در واقع با این کار مدل اولیه را به مخاطره می اندازیم. در صورتی که مدل جدید مورد تأیید قرار نگیرد مدل اولیه را با اطمینان بیشتری می پذیریم. نتیجه برازش مدل جدید بصورت زیر است :

ARIMA Model: T۱

Estimates at each iteration

Iteration	SSE	Parameters	
۰	۱۱۱۱۳.۰	۰.۱۰۰	۰.۱۰۰
۱	۸۹۱۷.۱	۰.۲۵۰	−۰.۰۵۰
۲	۸٤۹۸.۹	۰.۳۳٦	−۰.۱٦٤
۳	۸٤۹٥.۱	۰.۳٥۷	−۰.۱٦۱
٤	۸٤۹٥.۱	۰.۳٥۹	−۰.۱٦۰
٥	۸٤۹٥.۱	۰.۳٥۹	−۰.۱٦۰
٦	۸٤۹٥.۱	۰.۳٥۹	−۰.۱٦۰

Relative change in each estimate less than ۰.۰۰۱۰

Final Estimates of Parameters

```
Type        Coef  SE Coef      T      P
AR    ١   ٠.٣٥٩٢   ٠.١٩١٧   ١.٨٧  ٠.٠٦٤
MA    ١  -٠.١٥٩٥   ٠.٢٠٣٠  -٠.٧٩  ٠.٤٣٤


Differencing: ١ regular difference
Number of observations:  Original series ١٠٢, after
                         differencing ١٠١
Residuals:    SS =  ٨٤٩٠.٣٦ (backforecasts excluded)
              MS =  ٨٥.٧٦  DF = ٩٩


Modified Box-Pierce (Ljung-Box) Chi-Square statistic

Lag              ١٢     ٢٤     ٣٦     ٤٨
Chi-Square      ٥.٨   ١٨.١   ٢٧.٢   ٣٦.٣
DF               ١٠     ٢٢     ٣٤     ٤٦
P-Value       ٠.٨٢٩  ٠.٧٠١  ٠.٧٨٩  ٠.٨٤٧
```

در این مورد نیز برآورد پارامتر اضافی β_1 که مینی تب آن را با MA١ نشان می دهد معنی دار نیست. بنابراین مدل اولیه ARI(١,١)مورد تأیید قرار می گیرد.

پیش بینی

اکنون که با استفاده از استراتژی مدل سازی باکس–جنکینز مدل نهایی را شناسایی کرده ایم، می توانیم آن را مبنای پیش بینی رفتار آینده سری قرار دهیم. ما در این مثال می خواهیم مقادیر سری مصالح خط راه آهن را برای ١٢ ماه آینده پیش بینی کنیم.

برای انجام این کار مجددا پنجره ARIMA را باز کرده و مدل(١,١)ARI را برازش می دهیم، سپس بر روی گزینه Forecasts در پایین پنجره کلیک می کنیم و در پنجره باز شده در قسمت lead عدد ١٢ را وارد می کنیم.

در کادر Origin باید نقطه شروع پیش بینی ها را مشخص کنیم. چنانچه دراین قسمت چیزی ننویسیم، پیش بینی ها از انتهای سری تولید خواهند شد.

در قسمت Storage می توان مقادیر پیش بینی شده و حدود پیش بینی را ذخیره کرد.

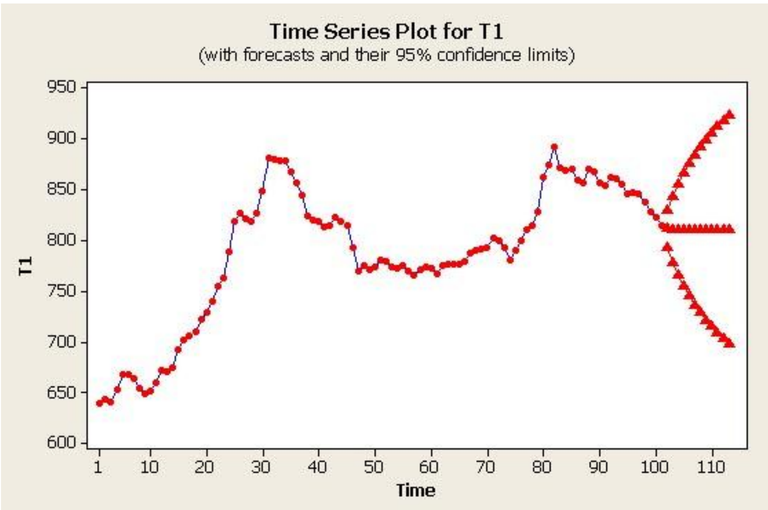
خروجی بصورت زیر خواهد بود :

```
Forecasts from period ١٠٢

                         ٩٥ Percent
                             Limits
Period  Forecast    Lower     Upper  Actual
```

Period	Forecast	Lower	Upper	Actual
١٠٣	٨١١.٠٣٥	٧٩٢.٩٣٢	٨٢٩.١٣٨	
١٠٤	٨١٠.٥٦٩	٧٧٨.١٩٧	٨٤٢.٩٤٢	
١٠٥	٨١٠.٣٤٤	٧٦٥.٤٨٦	٨٥٥.٢٠٣	
١٠٦	٨١٠.٢٣٦	٧٥٤.٤٩٥	٨٦٥.٩٧٧	
١٠٧	٨١٠.١٨٤	٧٤٤.٨٥٦	٨٧٥.٥١١	
١٠٨	٨١٠.١٥٨	٧٣٦.٢٦١	٨٨٤.٠٥٦	
١٠٩	٨١٠.١٤٦	٧٢٨.٤٧٧	٨٩١.٨١٦	
١١٠	٨١٠.١٤٠	٧٢١.٣٣٤	٨٩٨.٩٤٧	
١١١	٨١٠.١٣٨	٧١٤.٧٠٧	٩٠٥.٥٦٨	
١١٢	٨١٠.١٣٦	٧٠٨.٥٠٤	٩١١.٧٦٩	
١١٣	٨١٠.١٣٦	٧٠٢.٦٥٤	٩١٧.٦١٧	
١١٤	٨١٠.١٣٥	٦٩٧.١٠٥	٩٢٣.١٦٥	

توجه

برای مشاهده مقادیر پیش بینی شده در نمودار سری زمانی کافی است در پنجره اصلی ARIMA بر روی گزینه Graphs کلیک کرده و از پنجره باز شدهاولین گزینه که عبارت است از Time series plot (including optional forecasts)را انتخاب کنیم. این نمودار برای مثال فوق بشکل زیر است :

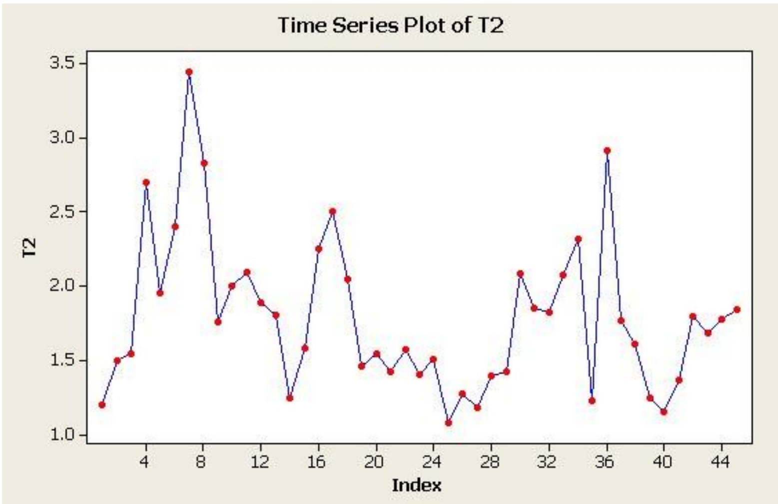


شکل ۵ – ۱۷ : پیش بینی سری مصالح خط راه آهن

همانطور که ملاحظه می شود حدود پیش بینی با افزایش زمان تقدم توسعه پیدا می کند.

مثال ۲

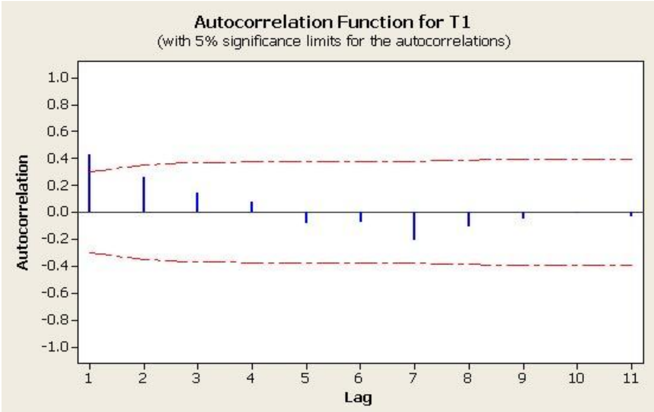
نمودار زیر مربوط به سری T۲ می باشد که متوسط تعداد نقایص مشاهده شده در هر کامیون را در بازرسی نهایی، در انتهای خط تولید کارخانه نشان می دهد. داده ها شامل ٤٥ مشاهده روزانه متوالی می باشند.

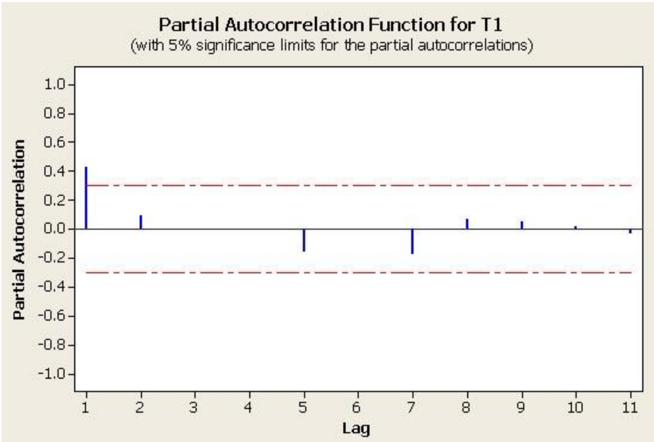


شکل ٥ – ١٨ : نمودار سری زمانی تعداد نقایص کامیون

تشخیص مدل آزمایشی

نمودار فوق یک فرآیندتقریباایستا با میانگین و واریانس ثابت را پیشنهاد می کند. acfوpacf نمونه ای را رسم می کنیم.همانطور که مشاهده می شود acf بطور نمائی تنزل می کند و pacf نمونه تنها در تأخیر ۱ معنی دار می باشد. در نتیجه می توان گفت که این سری احتمالا بوسیله یک فرآیند(AR(۱تولیدشده است.





شکل ۵– ۱۹ : acf و pacf سری تعداد نقایص کامیون

توجه:

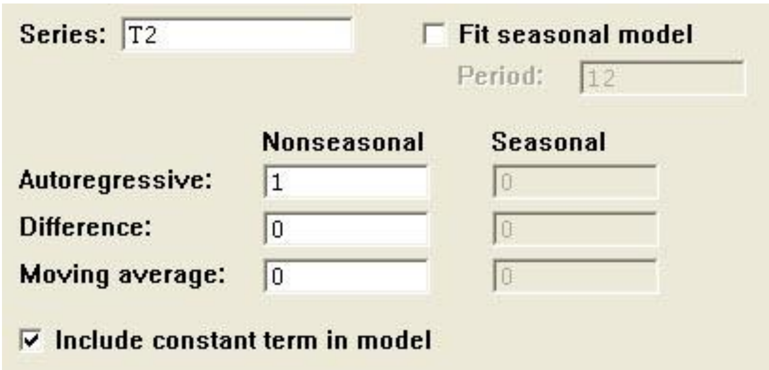
همانطور که در نمودارهای فوق مشاهده می کنید تابع خود همبستگی و تابع خود همبستگی جزیی نمونه، هر دو بعد از تأخیر یک قطع می شوند. بدین معنی که مقادیر توابع مربوطه بعد از تأخیر یک معنی دار نیستند و از حدود دو برابر خطای استاندارد تجاوز نکرده اند. اما آنچه حایز اهمیت است این است که مقادیر تابع خود همبستگی جزئی بعد از تأخیر یک به سرعت به صفر میل می کنند، در حالی که مقادیر تابع خود

همبستگی به کندی به صفر میل می کنند. در نتیجه بهتر است مدل را AR(۱) در نظر بگیریم تا MA(۱).

نکته : بطور کلی در مورد داده های غیر فصلی،acf و pacf هر دو ممکن است بطور نسبتا سریع قطع گردند و یا هر دو نسبتا سریع افول کنند. ولی در اغلب موارد اگر یکی از این دو تابع نسبتا سریع قطع شود، دیگری نسبتا سریع افول خواهد کرد. بدین ترتیب اگر چنین تعبیر کنیم که هر دوی این توابع قطع می گردند در این صورت باید تعیین نمود که کدام یک سریع تر قطع می شود که به معنی افول تابع دیگر می باشد.

برازش مدلAR(۱)

برازش مدل به معنی برآورد پارمترهای مجهول مدل می باشد. برای انجام این کار در مینی تب پنجره اصلی ARIMA را باز کرده و آن را به شکل زیر تکمیل می کنیم :



شکل ۵ – ۲۰ : تکمیل کردن پنجره آریما برای برازش مدلAR(۱)

پس از تکمیل پنجره ARIMA با فشردن دکمه okمدل مورد نظر به داده ها برازش می یابد. خروجی اجرای این رویه در پنجره session بصورت زیر می باشد.

ARIMA Model: T٢

Estimates at each iteration

Iteration	SSE	Parameters	
٠	١١.٢٤١٩	٠.١٠٠	١.٧٠٠
١	١٠.٠٨٥٨	٠.٢٥٠	١.٣٩٣
٢	٩.٥٦٤٩	٠.٤٠٠	١.٠٨٦
٣	٩.٥٣١٦	٠.٤٣٦	١.٠٠٦
٤	٩.٥٣٠٩	٠.٤٤١	٠.٩٩٥
٥	٩.٥٣٠٩	٠.٤٤٢	٠.٩٩٣
٦	٩.٥٣٠٩	٠.٤٤٢	٠.٩٩٣

Relative change in each estimate less than ٠.٠٠١٠

Final Estimates of Parameters

Type	Coef	SE Coef	T	P
AR ١	٠.٤٤٢١	٠.١٣٦٥	٣.٢٤	٠.٠٠٢
Constant	٠.٩٩٢٨٠	٠.٠٦٩٩٩	١٤.١٩	٠.٠٠٠
Mean	١.٧٧٩٥	٠.١٢٥٤		

Number of observations: ٤٥

```
Residuals:    SS =  ٩.٤٧٨١١ (backforecasts excluded)
                          MS =  ٠.٢٢٠٤٢  DF = ٤٣

Modified Box-Pierce (Ljung-Box) Chi-Square statistic

Lag              ١٢      ٢٤      ٣٦  ٤٨
  Chi-Square    ٤.٩     ٨.٩    ٣٠.٩    *
DF              ١٠      ٢٢      ٣٤    *
    P-Value    ٠.٨٩٩  ٠.٩٩٤  ٠.٦٢٠    *
```

همانطور که ملاحظه می شود مقدار $p-value$ برای جمله ثابت ۰.۰۰۰ است که کمتر از ۰.۰۵ می باشد. بنابراین فرضیه $H_0:\theta_0=0$ رد می شود. این به معنی حضور جمله ثابت و یا به عبارتی وجود روند قطعی در مدل می باشد. بنابراین نیازی به برازش مجدد مدل بدون جمله ثابت نیست. در نتیجه مدل آزمایشی بصورت زیر خواهد بود :

$$x_t=\theta_0+\alpha x_{t-1}+z_t$$

$$x_t=0.99+0.44x_{t-1}+z_t$$

توجه : همانطور که در فصل سوم گفته شد فرآیند اتورگرسیو مرتبه اول را فرآیند مارکف می نامند. این فرآیند ایستا است اگر $|\alpha|<1$. این شرط هم ارز است با اینکه بگوییم ریشه معادله $(1-\alpha B)=0$ باید خارج دایره واحد واقع شود. بنابراین در این

مثال چون مقدار برآورد شده برای α کمتر از یک می باشد، می توان گفت مدل فوق ایستا است.فرآیند اتورگرسیو مرتبه اول $x_t = \alpha x_{t-1} + z_t$ را می توانیم با جایگزینی های متوالی بصورت زیر بنویسیم :

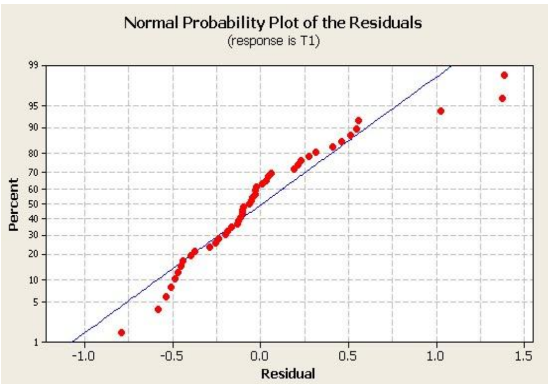
$$x_t = \alpha(\alpha x_{t-2} + z_{t-1}) + z_t = \alpha^2(\alpha x_{t-3} + z_{t-2}) + \alpha z_{t-1} + z_t$$

با ادامه این روند، می توان x_t را بصورت یک فرآیند MA از مرتبه نامتناهی به شکل زیر بیان کرد.

$$x_t = z_t + \alpha z_{t-1} + \alpha^2 z_{t-2} + \ldots$$

بررسی مناسبت مدل

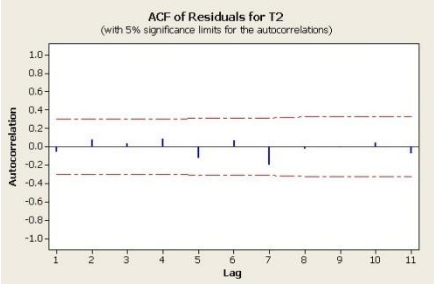
الف) بررسی فرض نرمال بودن باقیمانده ها

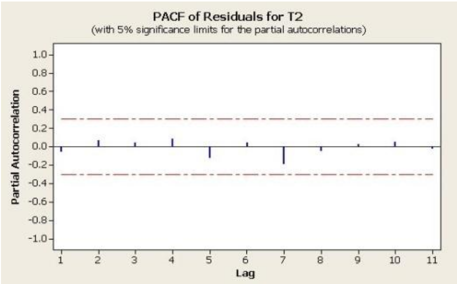


شکل ۵ – ۲۱ : نمودار احتمال نرمال باقیماندهای مدل برازش شده برای سری تعداد نقایص کامیون

همانطور که ملاحظه می شود نقاط تقریبا در امتداد یک خط راست قرار گرفته اند و این نشان می دهد که باقیمانده های حاصل از برازش مدل AR(۱) تقریبا بصورت نرمال توزیع شده اند.

ب) بررسی فرض استقلال باقیمانده ها

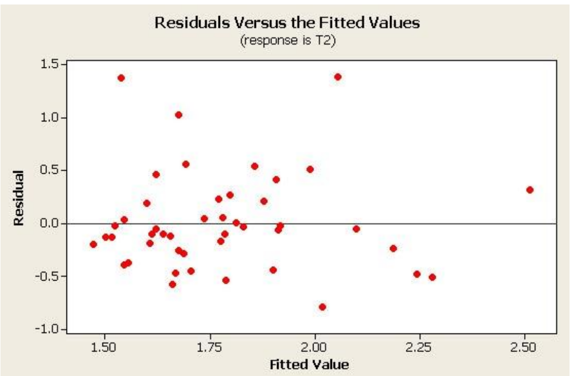




شکل ۵ –۲۲ : acf و pacf باقیمانده های مدل برازش شده برای سری تعداد نقایص کامیون

همانطور که ملاحظه می شود هیچ یک از خود همبستگی ها معنی دار نیستند. بنابراین می توان فرض استقلال باقیمانده ها را پذیرفت.

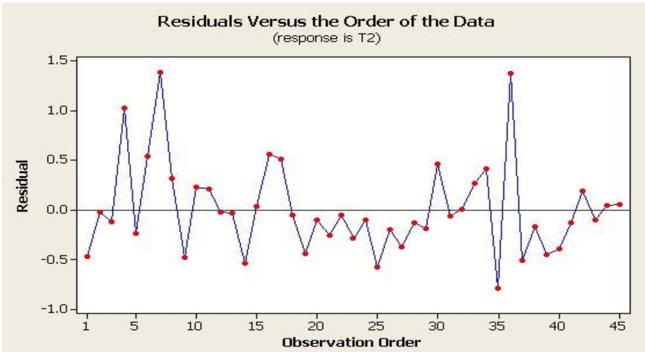
ج) بررسی فرض ثابت بودن واریانس باقیمانده ها



شکل ۵ – ۲۳ : نمودار باقیمانده هادر برابر مقادیر برازش شده برای سری تعداد نقایص کامیون

همانطور که ملاحظه می شود این نمودار ساختار خاصی را نشان نمی دهد. بنابراین می توان فرض ثابت بودن واریانس باقیمانده ها را پذیرفت.

د)نمودار باقیمانده ها در برابر زمان



شکل ۵- ۲٤ : نمودار باقیمانده ها در برابر زمانبرای سری تعداد نقایص کامیون

این نمودار مشابه ضربات ناشی از یک فرآیند اغتشاش خالص می باشد که بر صحت مدل برازش داده شده تأکید می کند.

ر) آزمون پرت- مانتو

نتایج مربوط به این آزمون در قسمت انتهای خروجی مینی تب به شکل زیر است :

Modified Box-Pierce (Ljung-Box) Chi-Square statistic

Lag	۱۲	۲٤	۳٦	٤٨
Chi-Square	٤.۹	۸.۹	۳۰.۹	*
DF	۱۰	۲۲	۳٤	*
P-Value	۰.۸۹۹	۰.۹۹٤	۰.٦۲۰	*

همانطور که ملاحظه می شود مقدار $p-value$ برای تمام تأخیرها بزرگتر از ۰.۰۵ است. به عنوان مثال در تأخیر۱۲ مقدار $p-value$ برابر ۰.۸۹ است که فرضیه صفر بودن همه خود همبستگی ها تا تأخیر ۱۲ را تأیید می کند.

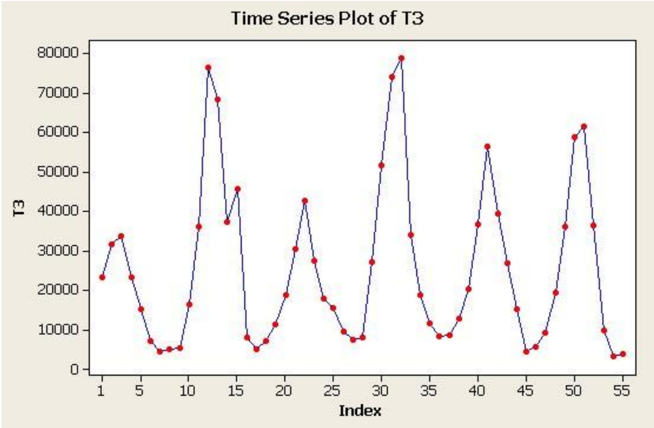
بنابراین بررسی باقیمانده ها مناسبت مدل را مورد تأیید قرار می دهد. خواننده محترم می تواند با استفاده از روش برازش جامع تر نیز مناسبت مدل آزمایشی را مورد بررسی قرار دهد.

مثال ۳

داده های مربوط به فروش تعداد تخته پوستهای خز به وسیله شرکت HUDSONBAY در کانادا بین سالهای ۱۸۵۷ تا ۱۹۱۱ در پیوست با نام T۳ آمده است. با رسم نمودار این سری و بررسی رفتار acf و pacf آن می خواهیم یک مدل مناسب را تشخیص داده و سپس شایستگی آن را بررسی کنیم.

شناسایی مدل آزمایشی

اولین گام در تجزیه و تحلیل سری زمانی رسم نمودار آن سری می باشد. برای سری T۳ نمودار سری زمانی به شکل زیر است.



شکل ۵– ۲۵ : نمودار سری زمانی فروش تعداد تخته پوستهای خز

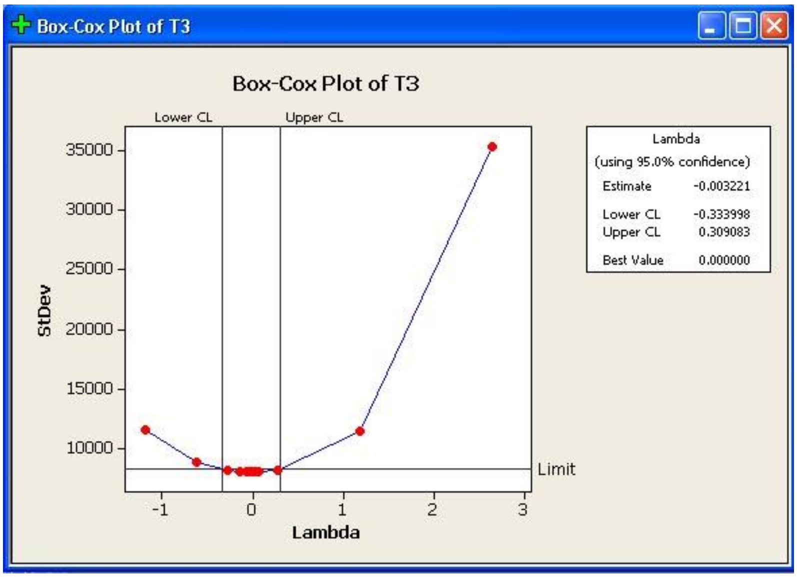
این نمودار نشان می دهد که سری مربوطه در میانگین ایستا می باشد. بنابراین نیازی به تفاضلی کردن سری به منظور ایستا کردن آن نداریم. اکنون به بررسی پایایی واریانس سری می پردازیم. البته لازم به ذکر است که چنانچه این سری نیاز به تفاضلی کردن می داشت این کار باید بعد از تبدیلات پایایی واریانس انجام می شد.

بررسی پایایی واریانس سری

برای اینکه بدانیم آیا برای پایا کردن واریانس سری احتیاج به تبدیل داده ها داریم یا نه، رویه باکس–کاکس را با تایپ کردن خط فرمان زیر در پنجره session اجرا می کنیم. البته قبلا باید از منوی Editor گزینه Enable Commands را انتخاب کنیم تا خط فرمان فعال شود.

```
MTB > boxcox T۳ ۱
```

نتیجه اجرای این فرمان نمودار زیر می باشد:

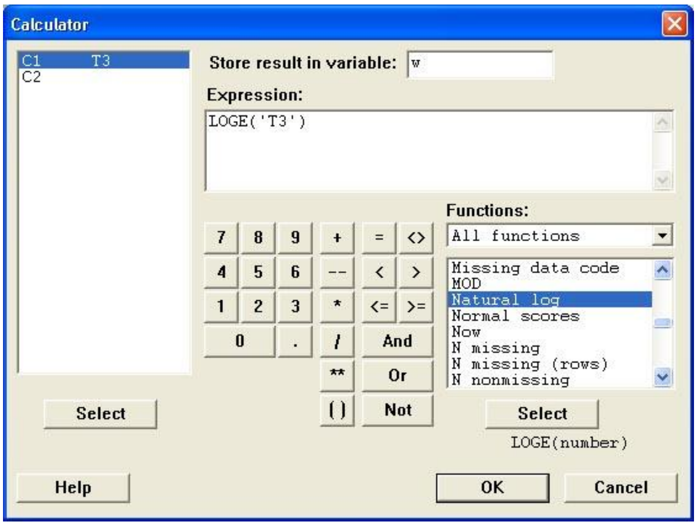


شکل ٥ – ٢٦ : نمودار تبدیلات باکس- کاکس برای سری فروش تعداد تخته پوستهای خز

همانطور که ملاحظه می شود بهترین مقدار پیشنهادی برای پارامتر تبدیل λ، صفر است که بر لزوم یک تبدیل لگاریتمی تأکید می کند.

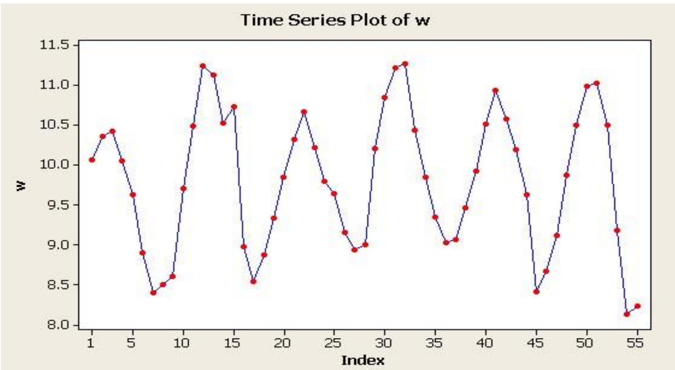
تبدیل لگاریتمی

برای تبدیل داده ها از منوی Calc گزینه Calculator را انتخاب می کنیم. سپس در قسمت Functionsتابع مورد نظر را که در اینجا تابع لگاریتم طبیعی(Natural log) می باشد برمی گزینیم. با دو بار کلیک کردن بر روی تابع مورد نظر، این تابع به کادر Expression منتقل می شود. حال از کادر سمت چپ، بر روی سری مورد نظر که قرار است تبدیل شود دو بار کلیک می کنیم تا تبدیل مورد نظر بر روی این سری انجام شود. همچنین در کادر Store result in variable یکی از ستونهای مینی تب را برای ذخیره کردن مقادیر تبدیل شده در نظر می گیریم. مراحل فوق در شکل زیر به وضوح دیده می شود.

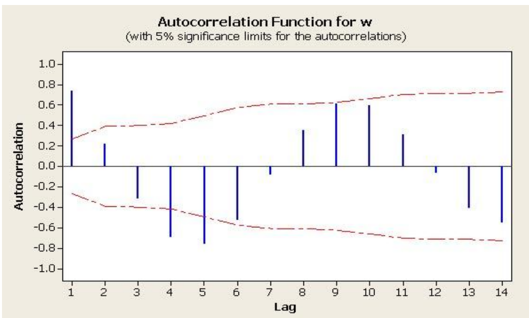


شکل ۵– ۲۷ : تبدیل لگاریتمی سری T۳

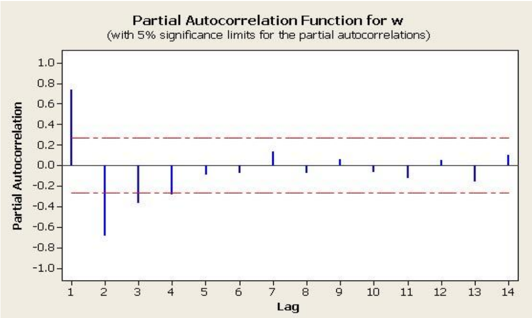
سری تبدیل شده را که در ستون C۲ ذخیره کرده ایم w_t می نامیم. حال نمودار سری زمانی را برای سری تبدیل شده $w_t = \ln x_t$ رسم می کنیم.این نمودار بیانگر یک سری زمانی است که در میانگین و واریانس ایستا می باشد. اکنون برای تشخیص نوع و مرتبه مدل، acf و pacf سری لگاریتمی را رسم می کنیم.



شکل ۵– ۲۸ : نمودار سری لگاریتمی w_t



شکل ۵ – ۲۹ :acf و pacfسری لگاریتمی



acf یک پدیده سینوسی-کسینوسی واضح را نشان می دهد که یک مدل AR(p) را مشخص می کند. سه pacf معنی دار، p=۳ را قویا پیشنهاد می کند. بنابراین مدل پیشنهادی ما عبارت است از :

$$(1-\alpha_1 B-\alpha_2 B^2-\alpha_3 B^3)\ln x_t = z_t$$

برازش مدل AR(۳)

برای برازش این مدل پنجره اصلی ARIMA را بصورت زیر تکمیل می کنیم.

شکل ۵ – ۳۰ : تکمیل پنجره آریما برای برازش مدل AR(۳)

نتیجه برازش این مدل در پنجره session بصورت زیر می باشد :

ARIMA Model: w

Estimates at each iteration

Iteration	SSE	Parameters			
۰	۳۹.٤۰۲۱	۰.۱۰۰	۰.۱۰۰	۰.۱۰۰	٦.۹۳۲
۱	۲۹.۳۰۵۳	۰.۲۵۰	۰.۰۵۸	۰.۰۰۳	٦.۸۱۵
۲	۲۱.۰٤۹۷	۰.٤۰۰	۰.۰۱۷	−۰.۰۹۵	٦.۷۰۱
۳	۱٤.٦۱۱٤	۰.۵۵۰	−۰.۰۲۵	−۰.۱۹٤	٦.۵۹۵
٤	۹.۹٦۲۵	۰.۷۰۰	−۰.۰٦۳	−۰.۲۹۸	٦.۵۱٤
۵	۷.۲۷۷٦	۰.۸۵۰	−۰.۱۰۲	−۰.٤۰۲	٦.٤۳۱
٦	٦.۵۲٦۹	۰.۹٤۷	−۰.۰۹۸	−۰.۵۰۸	٦.٤٦٦
۷	٦.۵۱۸۷	۰.۹۷۱	−۰.۱۲٤	−۰.۵۰٤	٦.٤٤٦

Unable to reduce sum of squares any further

Final Estimates of Parameters

Type		Coef	SE Coef	T	P
AR	۱	۰.۹۷۰۷	۰.۱۲۳۳	۷.۸۷	۰.۰۰۰
AR	۲	−۰.۱۲٤۲	۰.۱۸۵۷	−۰.٦۷	۰.۵۰۷
AR	۳	−۰.۵۰۳٦	۰.۱۲۸۷	−۳.۹۱	۰.۰۰۰
Constant		٦.٤٤٦٤۳	۰.۰٤۸۱٦	۱۳۳.۸٦	۰.۰۰۰

Mean ۹.۸۱۰٤۸ ۰.۰۷۳۲۹

Number of observations: ٥٥

Residuals: SS = ٦.٤٨٥٠٠ (backforecasts excluded)

MS = ۰.۱۲۷۱٦ DF = ٥۱

Modified Box-Pierce (Ljung-Box) Chi-Square statistic

Lag	۱۲	۲٤	۳٦	٤٨
Chi-Square	۱۲.۷	۲۷.۹	۳۹.٥	٤٤.۳
DF	۸	۲۰	۳۲	٤٤
P-Value	۰.۱۲٤	۰.۱۱۱	۰.۱٦۹	۰.٤٦۰

همانطور که ملاحظه می شود جمله ثابت اختلاف معنی داری باصفر دارد. زیرا آماره t مربوط به آن برابر ۱۳۳.۸٦ می باشد که بسیار بزرگتر از ۲ است. همچنین مقدار $p-value$ مربوطه کمتر از ۰.۰٥ می باشد که فرضیه معنی دار نبودن جمله ثابت یا به عبارتی $H_0 : \theta_0 = 0$ را رد می کند. بنابراین نمی توان وجود یک جمله ثابت مثبت را در مدل که بیانگر یک روند قطعی روبه بالا می باشد، نادیده گرفت.

همچنین در بین پارامترهای اتورگرسیو α_2 که مینی تب آن را با AR۲ نشان می دهد، معنی دار نیست بنابراین می توان آن را از مدل حذف کرد. در صورت نگه داشتن این پارامتر مدل آزمایشی بصورت زیر خواهد بود:

$$(1-0.97B+0.12B^2+0.5B^3)(\ln x_t-9.8)=z_t$$

این مدل را می توان با استفاده از θ_0 بجای μ به شکل زیر نوشت :

$$(1-0.97B+0.12B^2+0.5B^3)\ln x_t=6.44+z_t$$

زیرا همانطور که گفته شد $\theta_0=\varphi(B)\mu=(1-\alpha_1-\alpha_2-\alpha_3)\mu$.

چنانچه یک مدلی AR(۲)را به داده ها برازش دهیم، با وجود اینکه همه پارامترها معنی دار می باشند، اما باقیمانده های حاصل ناهمبسته نخواهند بود و با توجه به آزمون پرت-مانتو فرضیه ناهمبسته بودن باقیمانده ها را در هیچ یک از تأخیرها نمی توان پذیرفت. خروجی برازش مدلیAR(۲) بصورت زیر است :

ARIMA Model: w

Final Estimates of Parameters

Type		Coef	SE Coef	T	P
AR	۱	۱.۳۲۰۲	۰.۰۹٤۷	۱۳.۹٤	۰.۰۰۰
AR	۲	-۰.۷٦۲۱	۰.۰۹۸۲	-۷.۷٦	۰.۰۰۰
Constant		٤.۳۳۳٥۹	۰.۰٥٤۳۳	۷۹.۷۷	۰.۰۰۰
Mean			۹.۸۰۷۸	۰.۱۲۲۹	

```
Number of observations:  ٥٥

Residuals:    SS =  ٨.٤٠٧٩٩ (backforecasts excluded)

              MS =  ٠.١٦١٦٩  DF = ٥٢
```

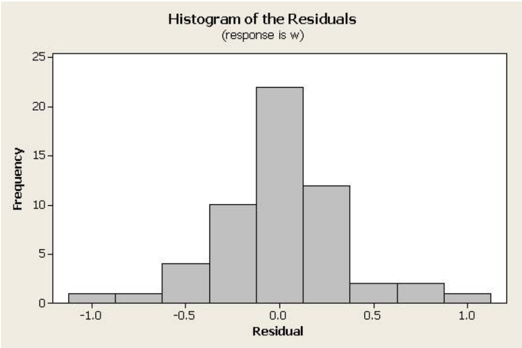
```
Modified Box-Pierce (Ljung-Box) Chi-Square statistic

Lag            ١٢     ٢٤     ٣٦     ٤٨
Chi-Square    ٢٤.٠   ٤٧.٧   ٦٤.٩   ٦٩.٥
DF             ٩      ٢١     ٣٣     ٤٥
P-Value      ٠.٠٠٤  ٠.٠٠١  ٠.٠٠١  ٠.٠١١
```

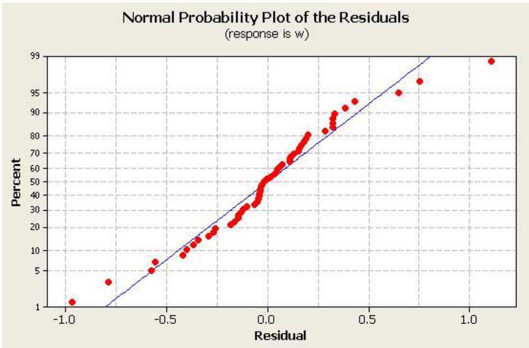
همانطور که ملاحظه می شود فرضیه ناهمبسته بودن باقیمانده ها به شدت رد شده است. بنابراین ما ترجیح می دهیم همان مدل AR(٣)را به عنوان مدل آزمایشی بپذیریم.

بررسی مناسبت مدل

بررسی مناسبت مدل را با استفاده از تجزیه و تحلیل باقیمانده ها شروع می کنیم.

الف) بررسی فرض نرمال بودن باقیمانده ها

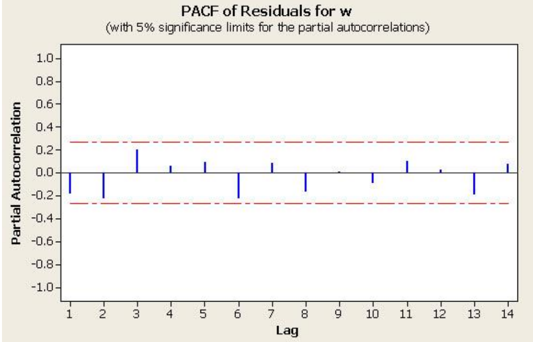




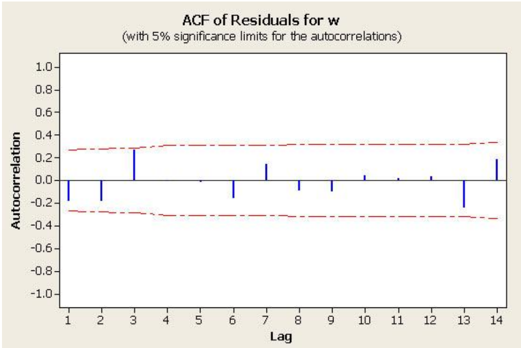
شکل ۵ – ۳۱ : هیستوگرام و نمودار احتمال نرمال با قیمانده های مدل(AR(۳برای سری لگاریتمی فروش تعداد تخته پوستهای خز

همانطور که ملاحظه می شود در نمودار احتمال نرمال نقاط تقریبا در امتداد یک خط راست قرار دارند. در نمودار هیستوگرام نیز مشاهده می شود که باقیمانده ها تقریبا بصورت نرمال توزیع شده اند. بنابراین می توان فرض نرمال بودن توزیع باقیمانده های حاصل از برازش مدل(AR(۳ را پذیرفت.

ب) بررسی فرض استقلال باقیمانده ها

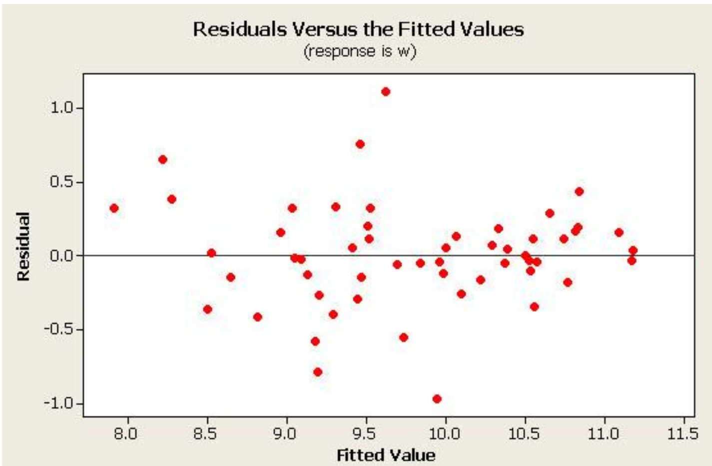


شکل ۵ – ۳۲ : acf و pacf باقیمانده های سری لگاریتمی فروش تعداد تخته پوستهای خز



با توجه به آنکه همه خود همبستگی ها داخل حدود استاندارد ترسیم شده قرار گرفته اند، بنابراین می توان فرضیه ناهمبسته بودن خود همبستگی های باقیمانده ها را پذیرفت.

ج) بررسی فرض ثابت بودن واریانس باقیمانده ها

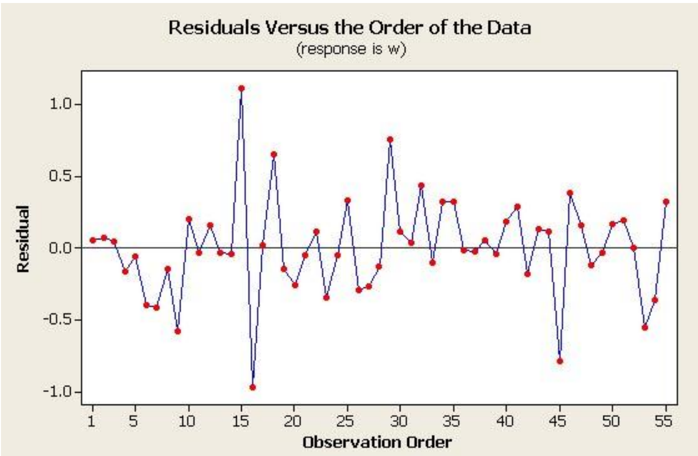


شکل ۵– ۳۳ : نمودار باقیمانده های مدل(AR(۳ در برابر مقادیر برازش شده برای سری لگاریتمی فروش تعداد تخته پوستهای خز

با توجه به آنکه این نمودار طرح خاصی را نشان نمی دهد بنابراین می توان فرضیه ثابت بودن واریانس باقیمانده ها را پذیرفت.

د) رسم نمودار باقیمانده ها در برابر زمان

این نمودار نیز طرح خاصی را نشان نمی دهد و دارای پراکندگی افقی حول صفر با واریانس ثابت می باشد. بنابراین به نظر می رسد باقیمانده های حاصل از برازش مدل AR(۳) مشابه یک فرآیند اغتشاش خالص رفتار می کنند که نشان دهنده مناسبت مدل برازنده شده است.



شکل ۵– ۳٤ : نمودار باقیمانده های مدل AR(۳) در برابر زمان برای سری لگاریتمی فروش تعداد تخته پوستهای خز

ر) آزمون پرت– مانتو

به عنوان یک روش رسمی تر برای آزمون فرضیه ناهمبسته بودن باقیمانده ها از آزمون پرت–مانتو که بر مبنای آماره اصلاح شده باکس–پیرسن می باشد استفاده می کنیم. خروجی مربوط به این آزمون بصورت زیر است :

```
Modified Box-Pierce (Ljung-Box) Chi-Square statistic

Lag            ۱۲     ۲٤     ۳٦     ٤۸
Chi-Square     ۱۲.۷   ۲۷.۹   ۳۹.۵   ٤٤.۳
```

DF	٨	٢٠	٣٢	٤٤
P-Value	٠.١٢٤	٠.١١١	٠.١٦٩	٠.٤٦٠

همانطور که ملاحظه می شود فرضیه ناهمبسته بودن باقیمانده ها در تمامی تأخیرها پذیرفته می شود. به عنوان مثال در تأخیر ٢٤ داریم $p-value$ ٠.١١١=که چون بیشتر از ٠.٠٥ است فرضیه را نمی توان رد کرد.

همچنین مقدار آماره کی دو در تأخیر ٢٤ برابر ٢٧.٩ می باشد که کمتر از مقدار بحرانی متناظر در جدول کی دو است($\chi^2_{(0.05,20)} = 31.41$).

برازش جامع تر

یکی دیگر از روش های بررسی مناسبت مدل برازش جامع تر یا **overfitting** می باشد. ما در این قسمت با برازش دو مدل کلی تر (AR(٤و (ARMA(٢,١، در واقع می خواهیم مدل شناسایی شده را به مخاطره بیندازیم و ببینیم آیا جوانب دیگری وجود دارد که مدل شناسایی شده آن را نادیده گرفته باشد.

الف) برازش مدل(AR(٤

در صورت برازش این مدل خروجی بصورت زیر خواهد بود:

Final Estimates of Parameters

Type		Coef	SE Coef	T	P
AR	١	٠.٧٧٣٦	٠.١٣١٣	٥.٨٩	٠.٠٠٠
AR	٢	−٠.١٤١٦	٠.١٧٢٩	−٠.٨٢	٠.٤١٧
AR	٣	−٠.١٢٧٨	٠.١٧٣٥	−٠.٧٤	٠.٤٦٥

```
AR   ٤     -٠.٤١٩٣   ٠.١٣٨٠    -٣.٠٤  ٠.٠٠٤
Constant  ٨.٩٧٥٤٤  ٠.٠٤٤٨١  ٢٠٠.٢٩  ٠.٠٠٠
Mean       ٩.٨٠٨٢٧  ٠.٠٤٨٩٧

Number of observations:  ٥٥
Residuals:    SS =  ٥.٥٠١٨٨ (backforecasts excluded)
              MS =  ٠.١١٠٠٤  DF = ٥٠

Modified Box-Pierce (Ljung-Box) Chi-Square statistic

Lag            ١٢    ٢٤    ٣٦    ٤٨
Chi-Square     ٨.٨   ٢٣.٠  ٣٥.٨  ٤٣.٦
DF             ٧     ١٩    ٣١    ٤٣
P-Value     ٠.٢٦٦  ٠.٢٣٧  ٠.٢٥٣  ٠.٤٤٧
```

همانطور که ملاحظه می شود هر چند برآورد پارامتر اضافی AR٤ معنی دار می باشد اما وارد کردن این پارامتر اضافی در مدل باعث شده است که AR٣ دیگر معنی دار نباشد.همچنین برآورد پارامتر مشترک AR٣ دستخوش تغییر شده است. برآورد پارامترAR٣ در مدل اول عبارت است از ٠.٥٠- در حالی که برآورد این پارامتر در مدل اخیر عبارت است از ٠.١٢-بنابراین دلیل قانع کننده ای برای در نظر گرفتن یک مدل AR از مرتبه بالاتر نداریم.

ب) برازش مدلARMA(٣،١)

خروجی برازش این مدل بصورت زیر می باشد:

Final Estimates of Parameters

Type		Coef	SE Coef	T	P
AR	١	٠.٣٩٤٠	٠.١٣٥٨	٢.٩٠	٠.٠٠٦
AR	٢	٠.٤٥٠٢	٠.١٥٤٩	٢.٩١	٠.٠٠٥
AR	٣	-٠.٦٩٢٢	٠.١٣٥٦	-٥.١١	٠.٠٠٠
MA	١	-٠.٩٥٤٠	٠.٠٧٧٢	-١٢.٣٦	٠.٠٠٠
Constant		٨.٣١٨٣	٠.١٠٧٦	٧٧.٣٢	٠.٠٠٠
Mean			٩.٨٠٨٩	٠.١٢٦٩	

Number of observations: ٥٥

Residuals: SS = ٨.٣٦١٨٩ (backforecasts excluded)

MS = ٠.١٦٧٢٤ DF = ٥٠

Modified Box-Pierce (Ljung-Box) Chi-Square statistic

Lag	١٢	٢٤	٣٦	٤٨
Chi-Square	٢٢.٥	٤٦.٠	٦٢.٦	٦٧.٢
DF	٧	١٩	٣١	٤٣
P-Value	٠.٠٠٢	٠.٠٠١	٠.٠٠١	٠.٠١٠

در این مدل هرچند همه پارامترها و از جمله پارامتر اضافی میانگین متحرک معنی دار می باشند. اما برآورد پارامترهای مشترک دستخوش تغییرات جدی شده است. بعلاوه همانطور که در قسمت مربوط به خروجی آزمون پرت-مانتو مشاهده می شود، فرض ناهمبسته بودن باقیمانده ها در تمامی تأخیرها به شدت رد شده است.

همچنین مجموع مربعات خطا از٦.٤٨در مدل AR(٣)به ٨.٣٦ در مدل اخیر افزایش یافته است که این خود می تواند دلیلی بر بهتر نشدن مدل باشد.

بنابراین با برازش بیشتر مدلAR(٣)مدل بهتری به دست نیاورده ایم. در نتیجه همان مدلAR(٣)را به عنوان یک مدل آزمایشی مناسب در نظر می گیریم.

ج) برازش مدل ARMA(٢،١)

در پی جستجوی مدلی بهتر یک مدلARMA(٢,١)را نیز در نظر گرفته ایم. لازم به ذکر است که این مدل با استفاده از تجزیه و تحلیل ESACF، که معرف تابع خود همبستگی نمونه ای تعمیم یافته می باشد، پیشنهاد شده است. ازاین تابع برای شناخت مرتبه فرآیندهای مرکب ARMA(p,q)استفاده می شود. این روش بر این حقیقت استوار است که اگر x_t از یک مدل مرکب ARMA(p,q) به شکل زیر پیروی کند :

$$\varphi(B)x_t = \theta_0 + \theta(B)z_t$$

که در آن

$$\theta(B) = 1 + \beta_1 B + \beta_2 B^2 + ... + \beta_q B^q$$

$$\varphi(B) = 1 - \alpha_1 B - \alpha_2 B^2 - ... - \alpha_p B^p$$

آنگاه $y_t = \varphi(B)x_t = (1 - \alpha_1 B - ... - \alpha_p B^p)x_t$ از یک مدلMA(q) به شکل $y_t = \theta(B)z_t$ پیروی می کند که بدون خدشه دار شدن کلیت مسأله فرض می کنیم $\theta_0 = 0$.

حال از acf نمونه ای باقیمانده های برآورد شده

$$\hat{y}_t = (1 - \hat{\alpha}_1 B - ... - \hat{\alpha}_p B^p) x_t$$

از یک کمترین مربعات متداول(OLS) برازش شده به AR جهت تشخیص qاستفاده می کنیم. در نتیجه مرتبه های p و qمدل ARMA(p,q) مشخص می شود.

برای مثال در یک فرآیند MA(۲) باقیمانده های حاصل از برازش یک AR نتیجه اش یک مدل ARMA(۱,۲) است.

با توجه به آنکه مینی تب این معیار را محاسبه نمی کند، ما از توضیح بیشتر این روش خودداری می کنیم. خواننده می تواند برای کسب اطلاعات بیشتر در این زمینه به منبع [۱] مراجعه کند.

نتیجه برازش مدل ARMA(۲,۱)بصورت زیر می باشد :

Final Estimates of Parameters

Type		Coef	SE Coef	T	P
AR	۱	۱.۵۴۱۱	۰.۰۶۷۲	۲۲.۹۵	۰.۰۰۰
AR	۲	-۰.۹۲۱۵	۰.۰۶۰۵	-۱۵.۲۳	۰.۰۰۰
MA	۱	۰.۶۰۱۱	۰.۱۳۷۸	٤.۳٦	۰.۰۰۰
Constant		۳.۷۳۲۲٦	۰.۰۱۸۷۸	۱۹۸.۷۵	۰.۰۰۰
Mean		۹.۸۱۳۱٦	۰.۰٤۹۳۷		

Number of observations: ۵۵

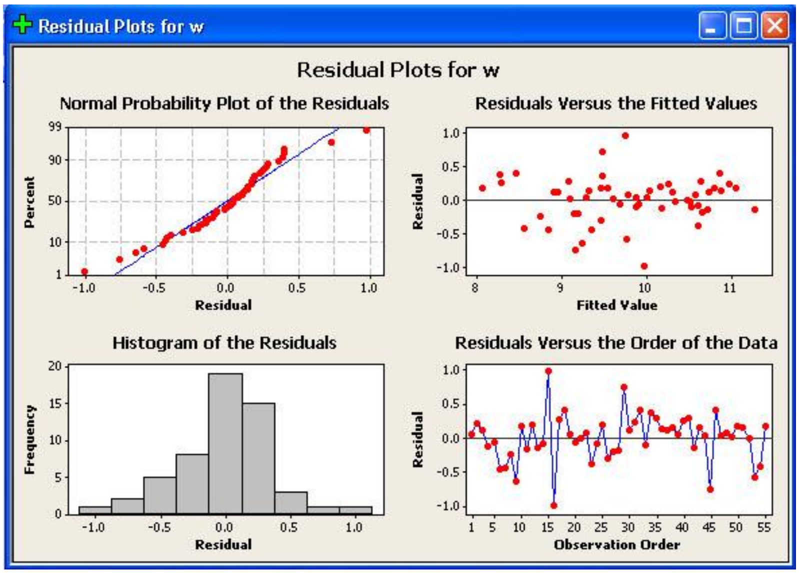
Residuals: SS = ٦.۱۸٦۱۵ (backforecasts excluded)

MS = ۰.۱۲۱۳۰ DF = ۵۱

Modified Box-Pierce (Ljung-Box) Chi-Square statistic

Lag	١٢	٢٤	٣٦	٤٨
Chi-Square	٩.٤	٢٥.٧	٣٧.١	٤٤.١
DF	٨	٢٠	٣٢	٤٤
P-Value	٠.٣١٣	٠.١٧٦	٠.٢٤٤	٠.٤٦٦

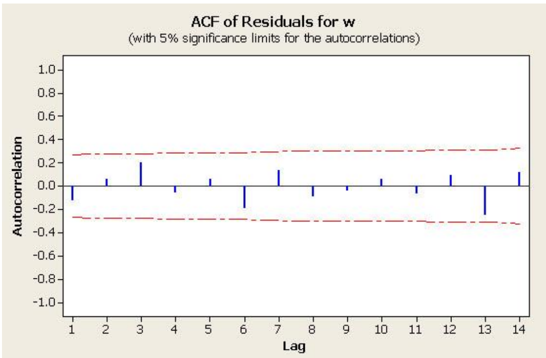
همانطور که ملاحظه می شود تمامی پارامترها معنی دار می باشند. همچنین آزمون پرت–مانتو فرضیه ناهمبسته بودن باقیمانده های حاصل از برازش این مدل را در تمامی تأخیر ها تأیید می کند. نمودارهای زیر نیز نرمال بودن باقیمانده ها و ثابت بودن واریانس آنها را تأیید می کنند.



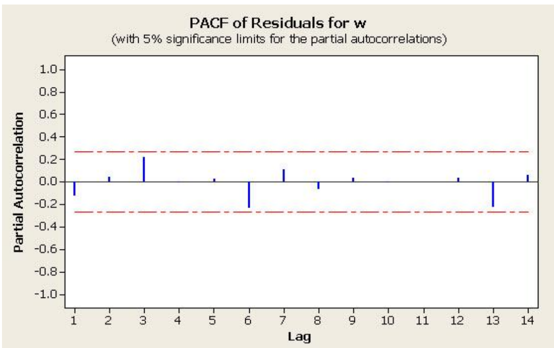
شکل ٥– ٣٥ : نمودارهای مربوط به باقیمانده های حاصل از برازش مدل(٢،١)ARMA

برای رسم چنین نموداری کافی است در پنجره Graphs گزینه Four in oneرا انتخاب کنیم تا هر چهار نمودار در یک پنجره رسم شوند.

همانطور که در شکل(۵- ۳٦) ملاحظه می شودacfو pacf باقیمانده های مدل برازش شده نیز طرح خاصی را نشان نمی دهد.



شکل ۵ – ۳٦ : acfو pacf باقیمانده های مدلARMA(۲,۱)



بنابراین می توان مدلARMA(۲,۱) را نیز به عنوان یک مدل مناسب در نظر گرفت. با توجه به برآورد پارامترها این مدل بصورت زیر خواهد بود :

$$(1-1.54B+0.92B^2)\ln x_t = 3.73+(1-0.6B)z_t$$

استفاده از معیار AIC برای انتخاب مدل

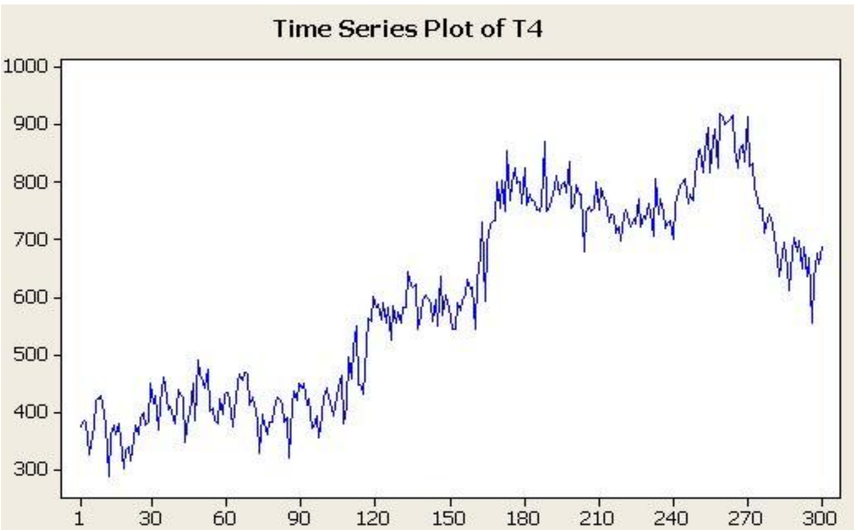
اکنون با توجه به به معیار آکائیک می خواهیم بدانیم از بین دو مدل AR(۳) و ARMA(۲,۱)کدامیک مناسب تر است. مدل اولی راM۱ و مدل دوم را M۲ می نامیم.

$$AIC(M1) = 55\log(0.1271) + 2(3) = -43.27$$
$$AIC(M2) = 55\log(0.1213) + 2(3) = -44.38$$

همانطور که ملاحظه می شود مدل دوم داری معیار آکائیک کمتری می باشد. بنابراین ما مدل ARMA(۲,۱) را با توجه به معیار آکائیک بر مدل AR(۳) ترجیح می دهیم.

مثال ٤

سری ماهانه تعداد زنان بیکار ١٦ تا ١٩ ساله(در هزار نفر) از ژانویه ١٩٦١ تا دسامبر ١٩٨٥ در شکل زیر نشان داده شده است. این سری را در پیوست با T٤مشخص کرده ایم. می خواهیم یک مدل مناسب از خانواده مدلهای ARIMA را به این داده ها برازش دهیم.



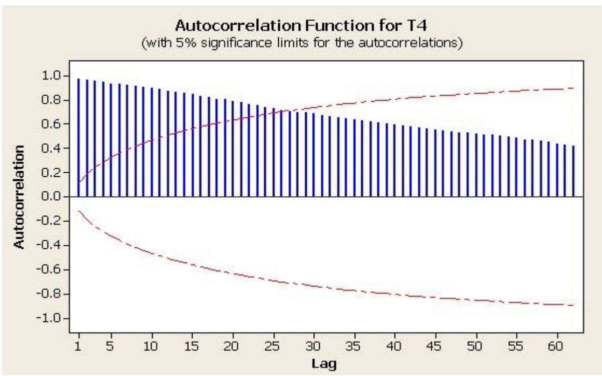
شکل ٥ – ٣٧ : نمودار سری زمانی تعداد زنان بیکار

توجه : در رسم نمودار فوق برای جلوگیری از شلوغ شدن نمودار و درک بهتر فرآیند، در قسمت Data Viewدر تب Data Displayچک مارک مقابلSymbolsرا برداشته ایم تا نحوه نمایش داده ها فقط بصورت Connect line باشد. در نتیجه مشاهدات با یک خط به هم مربوط شده اند و دیگر سیمبلهای دایره شکل را مشاهده نمی کنیم.

تشخیص مدل آزمایشی

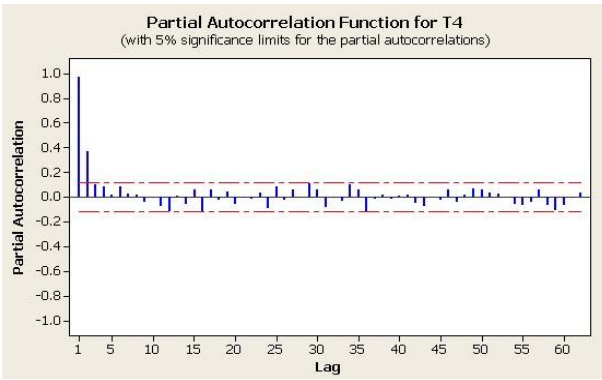
همانطور که از نموداراین سری مشخص است، این سری در میانگین ناایستا می باشد. بررسی acfنمونه ای نیز ناایستایی سری را تأیید می کند. همانطور که ملاحظه می شود مقادیر تابع خود همبستگی بسیار کند به سمت صفر میل می کنند.

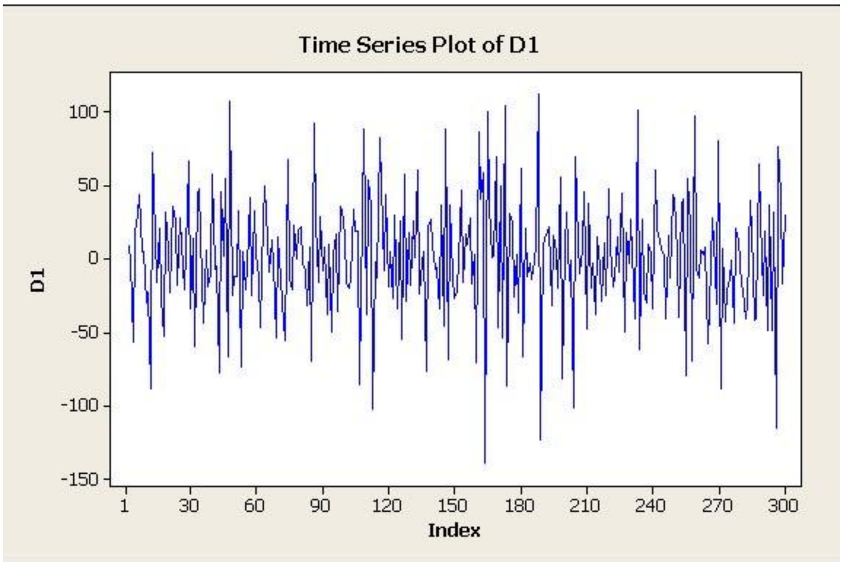
همچنین در مورد پایداری واریانس، با اجرای رویه box-coxبهترین مقدار پیشنهادی برای پارامتر تبدیل عدد یک می باشد که حاکی از ثابت بودن واریانس داده ها و عدم نیاز به تبدیلات پایداری واریانس می باشد.



شکل ۵ – ۳۸ : acf و pacfسری زمانی تعداد زنان بیکار

برای رفع مشکل ناایستایی، سری داده شده را یک بار تفاضلی می کنیم. سری تفاضلی شده را D۱ می نامیم. حال نمودار سری زمانی را برای سری تفاضلی شده رسم می کنیم.

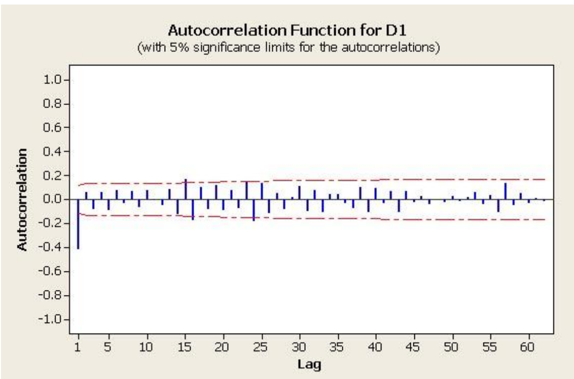


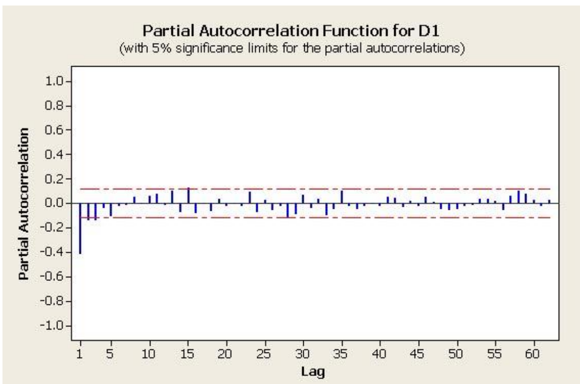


شکل ۵ – ۳۹ : نمودار سری تفاضلی شده D۱

همانطور که ملاحظه می کنید سری حاصل از تفاضلی کردن مرتبه اول سری T٤ یک سری ایستا می باشد. حال برای برازش یک مدل مناسب از خانواده مدلهای ARIMA به سری فوقacf و pacf سری تفاضلی شده را رسم می کنیم.

شکل ۵ – ٤۰ : acf و pacf سری تفاضلی شده





همانطور که ملاحظه می کنید نمودارacfبعد از تأخیر یک قطع می شود در حالی که نمودار pacf به کندی به سمت صفر میل می کند. بنابراین می توانیم یک مدل(۱)MA را برای سری تفاضلی شده و لذا یک مدل(۱,۱) IMA را برای سری اولیه در نظر بگیریم.

برای آزمون لزوم یک پارامتر روند قطعی، نسبت میانگین سری تفاضلی شده به انحراف معیار سری را محاسبه کرده ایم که معنی دار نمی باشد. بنابراین نیازی به وارد کردن جمله ثابت در مدل نداریم. لذا مدل آزمایشی به شکل زیر خواهد بود.

$$(1-B)x_t = (1+\beta B)z_t$$

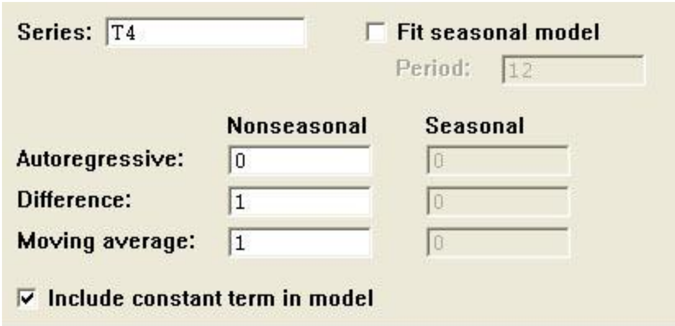
یا

$$x_t = x_{t-1} + z_t + \beta z_{t-1}$$

لازم به ذکر است که x_t همان سری اولیه می باشد که ما در این مثال آن را با T٤ نشان داده ایم. β تنها پارامتر مدل می باشد که می توان به سادگی آن را برآورد نمود. جمله z_t نیز بیانگر یک فرآیند تصادفی محض(اغتشاش خالص) می باشد.

برازش مدل(۱،۱)IMA

برای برازش این مدل پنجره ARIMA را به شکل زیر تکمیل می کنیم.



شکل ۵– ٤١ : تکمیل پنجره ARIMAبرای برازش مدل(۱،۱)IMA

خروجی در قسمت مربوط به برآورد پارامتر ها در پنجره session بصورت زیر است:

```
Final Estimates of Parameters


Type          Coef  SE Coef       T       P
MA   ١      ٠.٥١٦٢    ٠.٠٤٩٨  ١٠.٣٦  ٠.٠٠٠
Constant    ٠.٩٩٦     ١.٠٤٧   ٠.٩٥   ٠.٣٤٢
```

همانطور که ملاحظه می شود مقدار آماره tو مقدار $p-value$ برای جمله ثابت معنی دار نمی باشد بنابراین وجود روند قطعی در مدل مورد تأیید قرار نمی گیرد.

برازش مدل(۱،۱)IMA بدون جمله ثابت

با برداشتن چک مارک عبارت Include constant term in model جمله ثابت را از مدل حذف می کنیم. خروجی در پنجره seesion بصورت زیر می باشد :

ARIMA Model: T٤

```
Estimates at each iteration
```

Iteration	SSE	Parameters
٠	٤٨٥٣٥٤	٠.١٠٠
١	٤٤٥٣٤٠	٠.٢٥٠
٢	٤٢٢٣٣٠	٠.٤٠٠
٣	٤١٦٥٩٥	٠.٥٠٣
٤	٤١٦٥٥٥	٠.٥١١
٥	٤١٦٥٥٤	٠.٥١٢
٦	٤١٦٥٥٤	٠.٥١٢

Relative change in each estimate less than ٠.٠٠١٠

Final Estimates of Parameters

Type		Coef	SE Coef	T	P
MA	١	٠.٥١١٦	٠.٠٤٩٩	١٠.٢٥	٠.٠٠٠

Differencing: ١ regular difference

Number of observations: Original series ٣٠٠, after differencing ٢٩٩

Residuals: SS = ٤١٦٥٥٤ (backforecasts excluded)

MS = ١٣٩٨ DF = ٢٩٨

Modified Box-Pierce (Ljung-Box) Chi-Square statistic

Lag	١٢	٢٤	٣٦	٤٨
Chi-Square	٨.٨	٣٢.٠	٤٨.٧	٥٩.٩
DF	١١	٢٣	٣٥	٤٧
P-Value	٠.٦٣٧	٠.١٠٠	٠.٠٦٢	٠.٠٩٩

همانطور که ملاحظه می شود مقدار برآورد شده برای پارامتر β که مینی تب آن را با MA نشان می دهد برابر ۰.۵۱ می باشد. مقدار آماره tو مقدار $p-value$ نیز حاکی از معنی دار بودن این پارامتر می باشد.

توجه: همانطورکه در فصل سوم دیدیم مدلMA(q)بصورت زیر تعریف شد :

$$x_t = z_t + \beta_1 z_{t-1} + ... + \beta_q z_{t-q}$$

اما در مینی تب مدل فوق بصورت زیر تعریف شده است:

$$X_t = \varepsilon_t - MA1\varepsilon_{t-1} - MA2\varepsilon_{t-2} - ....$$

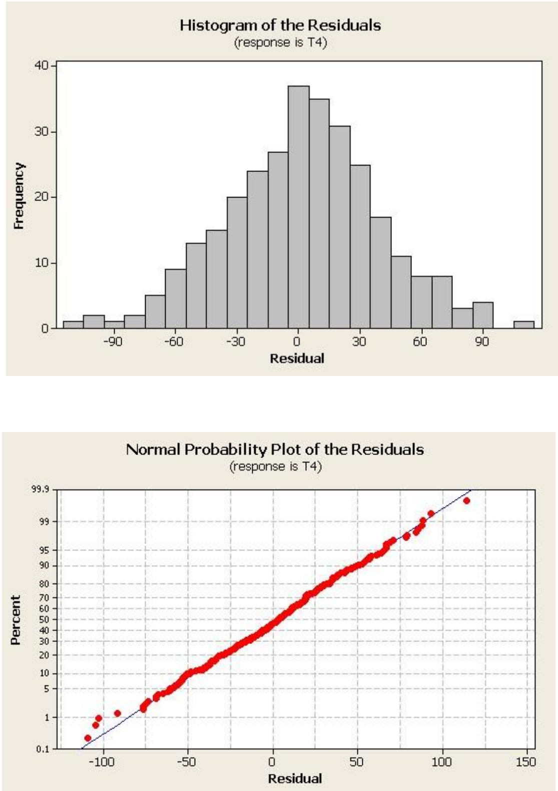
بنابراین ضرایب MA_i محاسبه شده در مینی تب قرینه ضرایبی است که ما تعریف کرده ایم.

یعنی $\beta_i = -MA_i$، به عنوان مثال در مدل فوق مقدار برآورد شده برای پارامتر میانگین متحرک برابر ۰.۵۱۱۶ می باشد. اما ما در مدل آنرا بصورت ۰.۵۱۱۶ − می نویسیم. بنابراین مدل آزمایشی برای این داده ها بصورت زیر می باشد:

$$x_t = x_{t-1} + z_t - 0.51z_{t-1}$$

بررسی مناسبت مدل

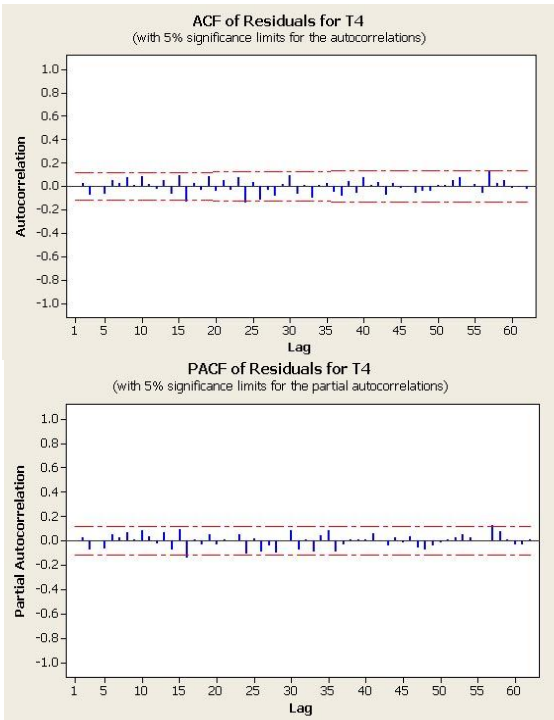
الف) بررسی فرض نرمال بودن باقیمانده ها



شکل ۵ – ۴۲ : هیستوگرام و نمودار احتمال نرمال باقیمانده های مدل IMA(۱,۱)

همانطور که ملاحظه می شود نمودار احتمال نرمال و هیستوگرام باقیمانده ها فرضیه نرمال بودن باقیمانده های حاصل از برازش مدلIMA(۱,۱)را بخوبی تأیید می کند.

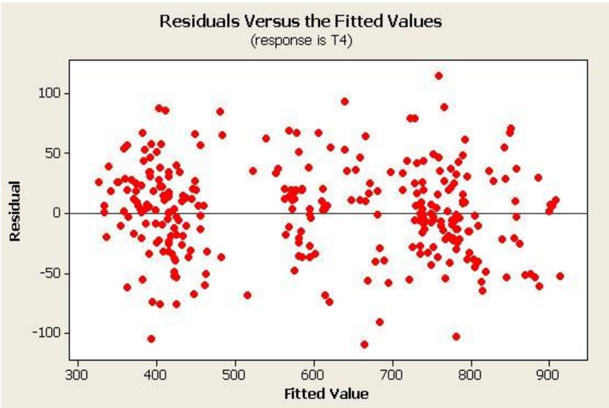
ب) بررسی فرض استقلال باقیمانده ها



شکل ۵ – ۴۳ : acf و pacf باقیمانده های مدل IMA(۱,۱) acf و pacf باقیمانده ها طرح خاصی را نشان نمی دهند و در هیچ یک از تأخیر ها تفاوت معنی داری با صفر ندارند. بنابراینفرضیه استقلال باقیمانده ها پذیرفته می شود.

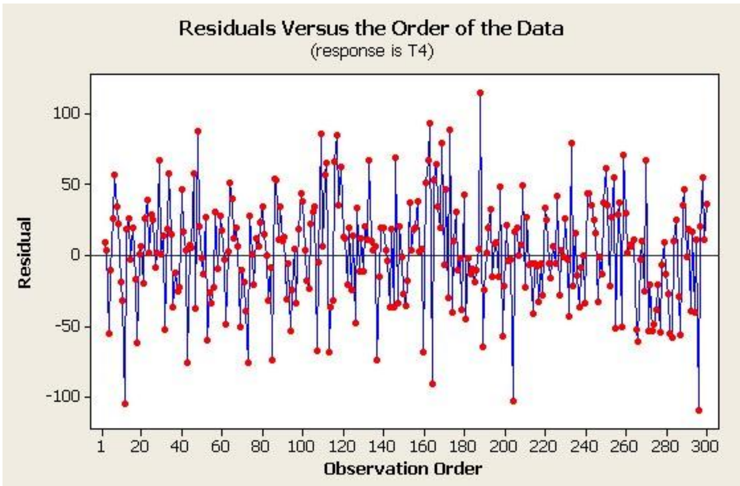
ج) بررسی فرض ثابت بودن واریانس باقیمانده ها

شکل ۵ – ٤٤ : نمودار باقیمانده ها در برابر مقادیر برازش شده برای مدل(۱,۱)IMA



همانطور که ملاحظه می شود این نمودار طرح خاصی را نشان نمی دهد. بنابراین می توان فرض ثابت بودن واریانس باقیمانده ها را پذیرفت.

د) رسم نمودار باقیمانده ها در برابر زمان



شکل ۵ – ٤٥ : نمودار باقیمانده های مدل(۱,۱)IMA در برابر زمان

این نمودار مشابه ضربات تصادفی ناشی از یک فرآیند اغتشاش خالص می باشد. پراکندگی مستطیلی حول صفر و واریانس ثابت از ویژگی های بارز نمودار فوق است. بنابراین نمودارهای مربوط به باقیمانده ها همگی صحت مدلبرازش داده شده را مورد تأیید قرار می دهند. اکنون با استفاده از آزمون پرت–مانتو به عنوان یک روش رسمی تر صحت مدل برازش داده شده را بررسی می کنیم.

ر) آزمون پرت–مانتو

همانطور که ملاحظه می شود خروجی مربوط به آزمون پرت–مانتو بصورت زیر می باشد.

Modified Box-Pierce (Ljung-Box) Chi-Square statistic

Lag	۱۲	۲٤	۳٦	٤۸
Chi-Square	۸.۸	۳۲.۰	٤۸.۷	۵۹.۹
DF	۱۱	۲۳	۳۵	٤۷
P-Value	۰.٦۳۷	۰.۱۰۰	۰.۰٦۲	۰.۰۹۹

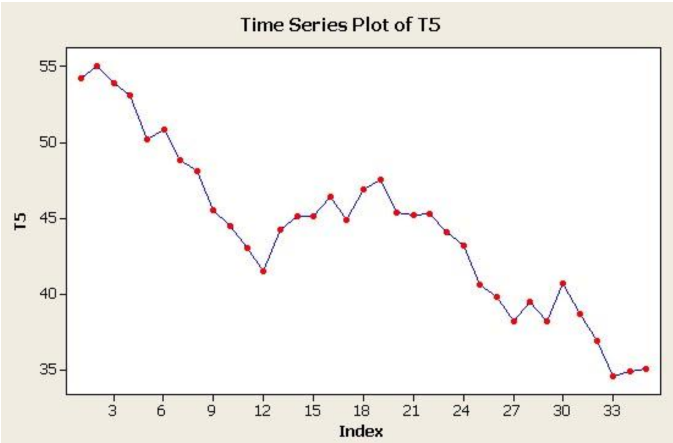
به عنوان نمونه در تأخیر ۲٤ مقدار آماره کی دو ۳۲ می باشد که از مقدار متناظر جدول کی دو با ۲۳ درجه آزادی و در سطح معنی داری ۰.۰۵ کمتر می باشد($\chi^2_{(0.05,23)} = 35.17$).

همچنین مقدار $p-value$ برابر ۰.۱۰ می باشد که بیشتر از ۰.۰۵ است بنابراین می توان فرضیه صفر بودن همه خود همبستگی ها را تا تأخیر ۲٤ پذیرفت. در تأخیر های بالاتر نیز این فرضیه رد نشده است.

مثال ۵

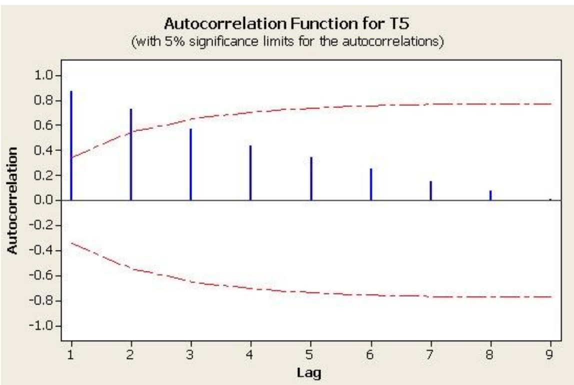
میزان مرگ و میر در اثر تصادف، یک آمار حیاتی مهم برای بسیاری از کشورها می باشد. شکل زیر میزان مرگ و میر ناشی از تصادفات سالانه(بر حسب یکصد میلیون نفر) در ایالت پنسیلوانیا بین سالهای ۱۹۵۰ تا ۱۹۸٤ را نشان می دهد. این داده ها را در پیوست با T۵ نشان داده ایم.

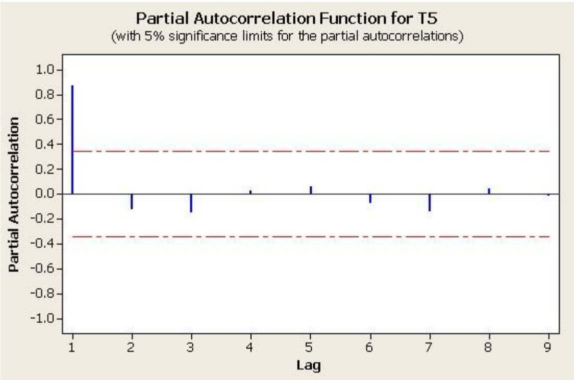
شخیص مدل آزمایشی



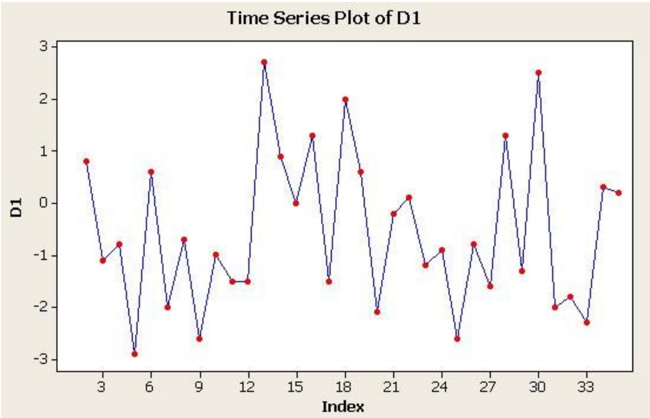
شکل ۵ – ٤٦ : نمودار سری زمانی T۵ (میزان مرگ و میر ناشی از تصادفات در ایالت پنسیلوانیا)

با توجه به نمودار سری زمانی، این سری ناایستا بوده و یک روند نزولی را نشان می دهد. همبستگی نگار این سری نیز به کندی به صفر میل می کند که تأییدی بر ناایستایی سری می باشد.در نتیجه برای به دست آوردن یک سری ایستا، سری فوق را تفاضلی نموده و سپس با رسم نمودارهای مربوطه، ایستایی سری تفاضلی شده را بررسی می کنیم. سری تفاضلی شده را D۱ می نامیم.

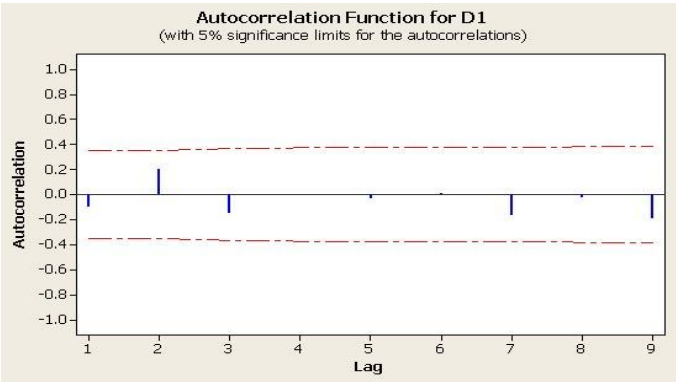




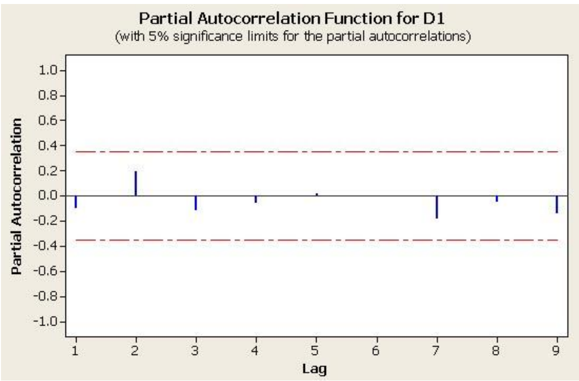
شکل ٥ – ٤٧: acf و pacfسری زمانیT٥



شکل ٥ – ٤٨ : نمودار سری تفاضلی شده D١



شکل ٥ – ٤٩ : acf و pacfسری تفاضلی شده



نمودار سری زمانی، acf و pacf داده های تفاضلی شده یک فرآیند تصادفی محض(اغتشاش خالص) را پیشنهاد می کنند. همانطور که ملاحظه می شود، هیچ یک از خود همبستگی ها معنی دار نیستند. یعنی وابستگی بین مشاهدات تقریبا صفر است و می توانیم این مشاهدات را به عنوان یک نمونه مستقل در نظر بگیریم.

در پنجرهsession می توان مقدار آماره tو آماره LBQ را برای مقادیر تابع خود همبستگی مشاهده نمود.

Autocorrelation Function: d۱

Lag	ACF	T	LBQ
۱	-۰.۰۹۸۲۵۲	-۰.۵۷	۰.۳۶
۲	۰.۱۹۷٦٤۲	۱.۱٤	۱.۸٥
۳	-۰.۱٤۸۳٦٦	-۰.۸۳	۲.۷۲
٤	۰.۰۰۱۳٤۱	۰.۰۱	۲.۷۲
٥	-۰.۰۲۹۸۹۷	-۰.۱٦	۲.۷٦
٦	۰.۰۱۱۱۷٦	۰.۰٦	۲.۷٦
۷	-۰.۱٦٦٥۷۷	-۰.۹۱	٤.۰۲
۸	-۰.۰۲۳۲۰٤	-۰.۱۲	٤.۰٥
۹	-۰.۱۸۹۸٤۷	-۱.۰۱	٥.۸۱

همانطور که مشاهده می شود مقدار آماره t در هیچ یک از تأخیرها معنی دار نیست. همچنین با استفاده از آماره LBQمی توان فرضیه صفر بودن همه خود همبستگی ها را تا تأخیر ۹ آزمون کرد.

چون نمی توان مدلی بصورت فرآیند اغتشاش خالص را به داده ها برازش داد، بنابراین محاسبات مربوطه را خودمان باید انجام دهیم.

بررسی وجود روند قطعی در مدل

نسبت tبرای سری تفاضلی شده یک جمله روند قطعی را توصیه می کند. همانطور که در ابتدای این فصل گفته شد آماره tبشکل $T = \frac{\overline{w}}{s_{\overline{w}}}$ است.

برای محاسبه این نسبت از منوی Stat گزینه Basic Statisticsو سپس گزینه Display Descriptive Statistics را انتخاب می کنیم . نام سری تفاضلی شده D۱ را به کادر Variables منتقل می کنیم. می توان بر روی گزینه Statisticsدر پایین صفحه کلیک کرده و از پنجره باز شده آماره های دلخواه را که در اینجا Mean و SE of mean می باشد را انتخاب می کنیم. نتیجه به شکل زیر خواهد بود.

Descriptive Statistics: d۱

Variable	Mean	SE Mean
d۱	-۰.۵٦٢	۰.۲٥٤

حال برای محاسبه آماره tکه عبارت است از نسبت میانگین سری به انحراف استاندارد میانگین سری، این دو عدد را بر هم تقسیم می کنیم.

$$t = \frac{-0.562}{0.254} = -2.21$$

چون قدر مطلق این مقدار بزرگتر از ۲ می باشد، بنابراین می توان گفت پارامتر مربوطه(در اینجا جمله ثابت) اختلاف معنی داری با صفر دارد. در نتیجه مدل پیشنهادی یک فرآیند اغتشاش خالص بعلاوه یک جمله روند قطعی رو به پایین می باشد.

بنابراین مدل پیشنهادی بصورت زیر خواهد بود :

$$(1-B)x_t = \mu + z_t$$

یا

$$x_t - x_{t-1} = -0.56 + z_t$$

بر پایه داده های تفاضلی نشده اولیه، با توجه به acf و pacf سری اولیه، یک مدلی (۱)AR را نیز می توان پیشنهاد نمود.

$$(1-\alpha_1 B)(x_t - \mu) = z_t$$

آزمون خود همبستگی ها

همانطور که می دانید، می توان با استفاده از آماره LBQ فرضیه صفر بودن همه خود همبستگیها تا تأخیر k را مورد آزمون قرار داد. در صورت پذیرفته شدن این فرضیه می توان گفت داده ها مستقل از یکدیگر هستند.

ما در اینجا می خواهیم فرضیه صفربودن خود همبستگی ها را تا تأخیر ۹ بررسی کنیم. همانطور که ملاحظه می شود، مقدار آماره LBQ برای تأخیر ۹ برابر ۵.۸۱ است.

مرحله اول : محاسبه تابع احتمال تجمعی

برای این منظور مراحل زیر را طی می کنیم.

Calc > Probability Distributions > Chi-Square

Cumulative Probability

Degrees of freedom ۹ (تأخیر مورد آزمون) $\longrightarrow$

Input constant ۵.۸۱ (مقدار آماره LBQ در تأخیر ۹) $\longrightarrow$

Optional storage *Cumprob* $\longrightarrow$

مرحله دوم : محاسبه p-value

Calc > Calculator

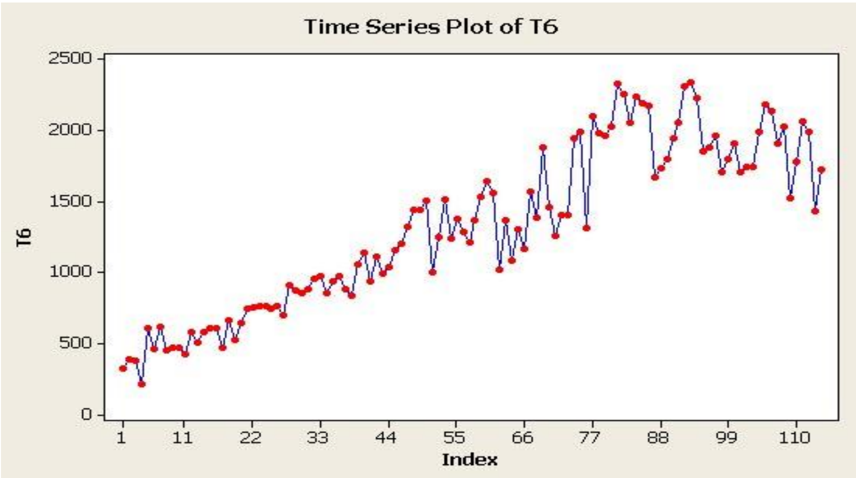
Store result in variable ⟶ *p-value*

Expression ⟶ عبارت- '*Cumprob*' را تایپ کنید

با طی این مراحل در مینی تب در ستونی با نام $p-value$ مقدار ۰.۷۵۸۷۷۳ مشاهده می شود. با توجه به اینکه این $p-value$ بزرگتر از ۰.۰۵ می باشد در نتیجه فرضیه صفر را نمی توان رد کرد. فرضیه صفر می گوید که همه خود همبستگی ها تا تأخیر ۹ صفر می باشند. یعنی با ۹۵ درصد اطمینان می توان گفت همبستگی بین مشاهدات معنی دار نیست و داده ها مستقل از یکدیگر می باشند. بنابراین می توان گفت این مشاهدات به وسیله یک فرآیند تصادفی محض تولید شده اند.

مثال ٦

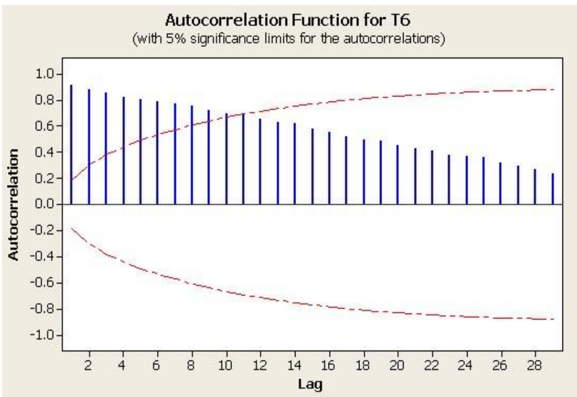
سری تولید سالانه تنباکو از سال ۱۸۷۱ تا ۱۹۸٤ را مورد بررسی قرار می دهیم. این داده ها در سال ۱۹۸۵ به وسیله دپارتمان عایدی ایالات متحده چاپ شده است. این سری را در پیوست با T٦ نشان داده ایم.



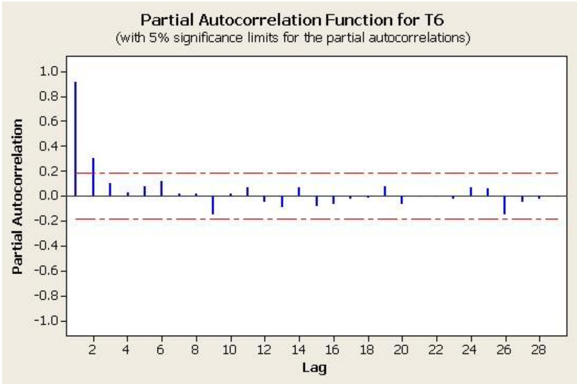
شکل ٥ – ٥٠ : نمودار سری زمانی تولید سالانه تنباکو از ۱۸۷۱ تا ۱۹۸٤

تشخیص مدل آزمایشی

نمودار سری ناایستایی در میانگین و واریانس را نشان می دهد. در حقیقت انحراف معیار سری، تقریبا متناسب با سطح آن است. بنابراین یک تبدیل لگاریتمی پیشنهاد می شود.



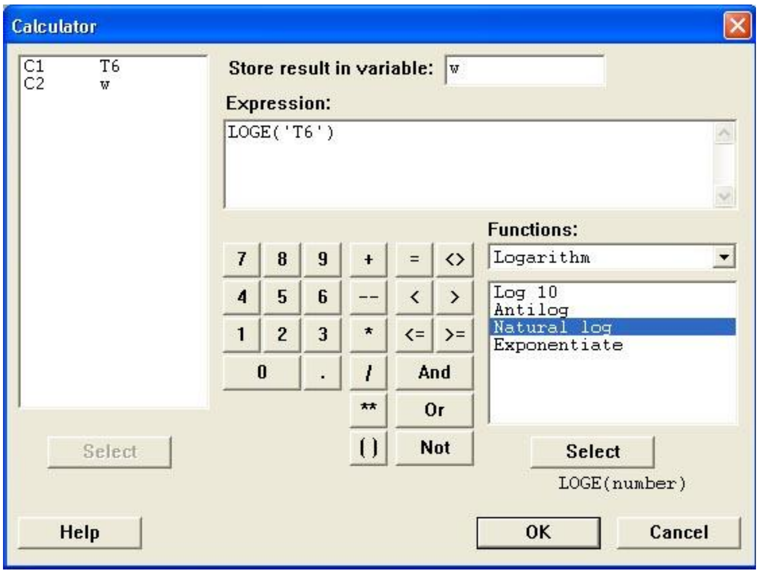
شکل ٥ – ٥١ :acf و pacfسری زمانی T٦



تنزل بسیار کند acf ناایستایی سری را تأیید نموده و بر لزوم تفاضلی کردن سری تأکید می کند. اما قبل از تفاضلی کردن سری ابتدا باید تبدیلات پایداری واریانس را انجام داد.

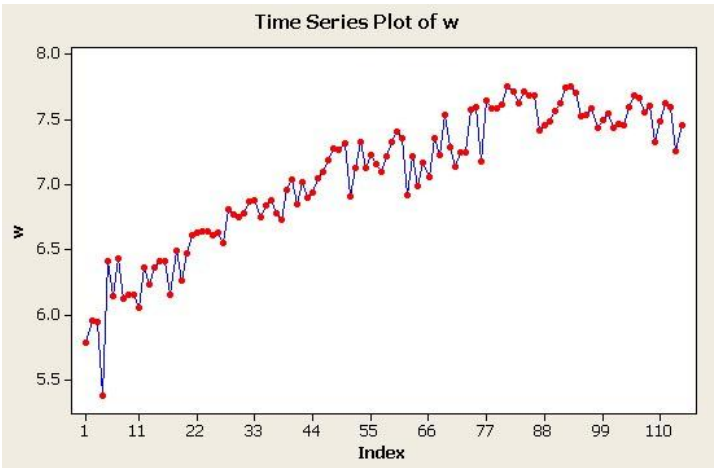
تبدیل لگاریتمی

همانطور که قبلا هم توضیح داده شده است برای تبدیل داده ها از منوی Calc گزینه Calculator را انتخاب می کنیم. سپس در قسمت Functionsتابع مورد نظر را که در اینجا تابع لگاریتم طبیعی(Natural log) می باشد برمی گزینیم. با دو بار کلیک کردن بر روی تابع مورد نظر، این تابع به کادر Expression منتقل می شود.حال از کادر سمت چپ، بر روی سری مورد نظر که قرار است تبدیل شود، دو بار کلیک می کنیم تا تبدیل مورد نظر بر روی این سری انجام شود. همچنین در کادر Store result invariable یکی از ستونهای مینی تب را برای ذخیره کردن مقادیر تبدیل شده در نظر می گیریم. مراحل فوق در شکل زیر به وضوح دیده می شود.



شکل ۵ – ۵۲ : تبدیل لگاریتمی سری T6

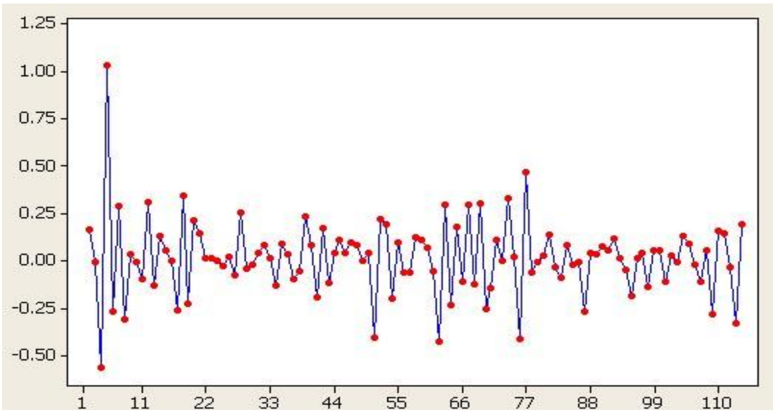
سری تبدیل شده را که در ستون C2 ذخیره کرده ایم w_t می نامیم. حال نمودار سری زمانی را برای سری تبدیل شده $w_t = \ln x_t$ رسم می کنیم و سپس سری تبدیل شده را که تقریبا دارای واریانس ثابت می باشد، تفاضلی می کنیم تا در میانگین نیز ایستا شود.



شکل ۵ – ۵۳ : نمودار سری لگاریتمی w_t

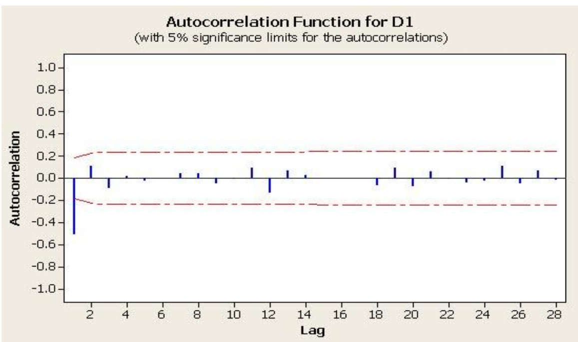
همانطور که قبلا هم گفته شده است برای تفاضلی کردن این سری از منوی TimeSeries گزینه Differences را انتخاب می کنیم. در پنجره ظاهر شده با دوبار کلیک کردن بر روی سری مورد نظر(w_t) در کادر سمت چپ، آن را به قسمت Series منتقل می کنیم.

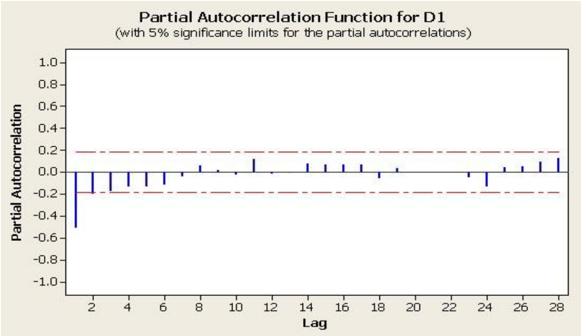
در کادر مقابل Store differences in نام ستونی را که می خواهیم مقادیر تفاضلی شده را در آن ذخیره کنیم وارد می کنیم. ما در اینجا مطابق معمول تفاضلات مرتبه اول را در ستونی با نام D۱ ذخیره می کنیم. در کادر مقابل Lag نیز عدد یک را وارد می کنیم که نشان دهنده مرتبه تفاضل می باشد. حال نمودار سری زمانی را برای سری تفاضلی شده D۱ رسم می کنیم. سری جدید در واقع به شکل $\nabla w_t = (1-B)w_t = (1-B)\ln x_t$ می باشد.



شکل ۵ – ٥٤ : نمودار سری تفاضلی شده D۱

اکنون یک سری ایستا داریم و می توانیم آنرا به مدل درآوریم. برای تشخیص مدل آزمایشی، از تابع خود همبستگی و تابع خود همبستگی جزئی کمک می گیریم.





شکل ۵– ۵۵: acf و pacf سری تفاضلی شده

همانطور که ملاحظه می شود acf بعد از تأخیر یک قطع شده است و pacf بطور نمائی تنزل پیدا می کند. البته، با توجه به نکته ای که در مثال۲ نیز توضیح داده شد، هر چند pacf نیز بعد از تأخیر یک قطع شده استاما در مقایسه با acf که بسیار سریع به سمت صفر میل کرده، pacf به کندی به صفر میل می کند. در نتیجه می گوییم در مقایسه با acf نمودارpacf بطور نمائی تنزل پیدا کرده است.

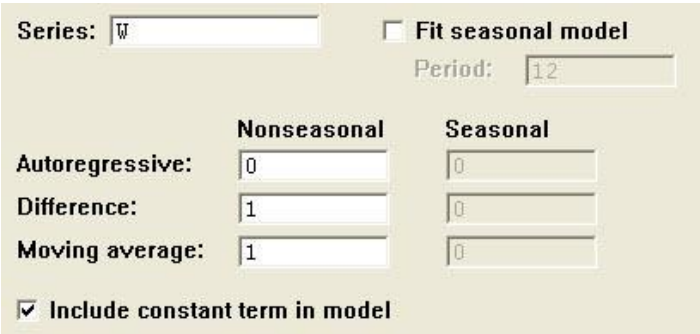
از آنجا که رفتار سری تفاضلی شده مشابه رفتار فرآیند MA(۱)می باشد، بنابراین ما مدل آزمایشی(ARIMA(۰,۱,۱ یا همان (IMA(۱,۱را برای سری اولیه w_t در نظر می گیریم، که این مدل بصورت زیر است :

$$(1-B)w_t = (1+\beta B)z_t$$

$$(1-B)\ln x_t = (1+\beta B)z_t$$

برازش مدل(IMA(۱،۱

برای برازش این مدل پنجره ARIMA را به شکل زیر تکمیل می کنیم.



شکل ۵ – ۵٦ : تکمیل پنجره آریما برای برازش مدل (IMA(۱,۱

خروجی در پنجره session بصورت زیر است:

ARIMA Model: W

```
Estimates at each iteration

Iteration      SSE   Parameters
        ٠   ٥.٤٢٦٠١   ٠.١٠٠   ٠.١١٥
        ١   ٤.١١٦١٩   ٠.٢٥٠   ٠.٠٦٦
        ٢   ٣.٤٠٨٠١   ٠.٤٠٠   ٠.٠٣٦
        ٣   ٣.٠٥٦٣٢   ٠.٥٥٠   ٠.٠٢٠
        ٤   ٢.٩٥٢٢٩   ٠.٦٩٤   ٠.٠١٣
        ٥   ٢.٩٥٠٦٧   ٠.٦٩٨   ٠.٠١٤
        ٦   ٢.٩٥٠٦٧   ٠.٦٩٧   ٠.٠١٤
        ٧   ٢.٩٥٠٦٧   ٠.٦٩٧   ٠.٠١٤

Relative change in each estimate less than ٠.٠٠١٠

Final Estimates of Parameters

Type          Coef   SE Coef      T      P
MA   ١       ٠.٦٩٧٤     ٠.٠٦٩٢   ١٠.٠٧   ٠.٠٠٠
Constant  ٠.٠١٤٢٥٦  ٠.٠٠٤٦٨١    ٣.٠٥   ٠.٠٠٣

Differencing: ١ regular difference
```

```
Number of observations:  Original series ١١٤, after
                                    differencing ١١٣

Residuals:    SS =  ٢.٩٤١٠٣ (backforecasts excluded)

                         MS =  ٠.٠٢٦٥٠  DF = ١١١


Modified Box-Pierce (Ljung-Box) Chi-Square statistic


Lag              ١٢      ٢٤      ٣٦      ٤٨

   Chi-Square     ٥.٥     ٨.٩    ١٧.٨    ٢٨.٢

DF               ١٠      ٢٢      ٣٤      ٤٦

     P-Value    ٠.٨٥٧  ٠.٩٩٤  ٠.٩٩٠  ٠.٩٨٢
```

همانطور که ملاحظه می شود مقدار آماره tبرای جمله ثابت بیشتر از ۲ می باشد. همچنین $p-value$ مربوطه نیز کمتر از ۰.۰۵ می باشد. در نتیجه فرضیه $H_0: \theta_0 = 0$ رد می شود. بنابراین حضور جمله ثابت در مدل ضروری است. به عبارتی مدل دارای یک روند قطعی رو به بالا می باشد. همچنین با توجه به آماره tو $p-value$ ، پارامتر میانگین متحرک نیز معنی دار می باشد. بنابراین مدل آزمایشی بصورت زیرخواهد بود:

$$(1-B)w_t = \theta_0 + (1+\beta B)z_t$$

$$(1-B)w_t = 0.014 + (1-0.697B)z_t$$

$$w_t = 0.014 + w_{t-1} + z_t - 0.697z_{t-1}$$

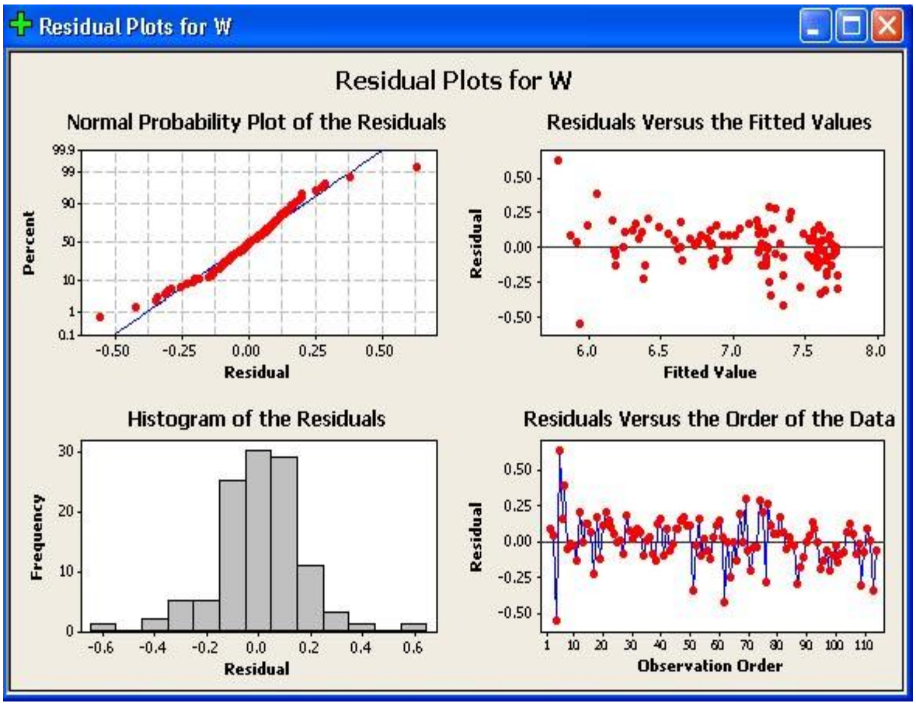
بررسی مناسبت مدل

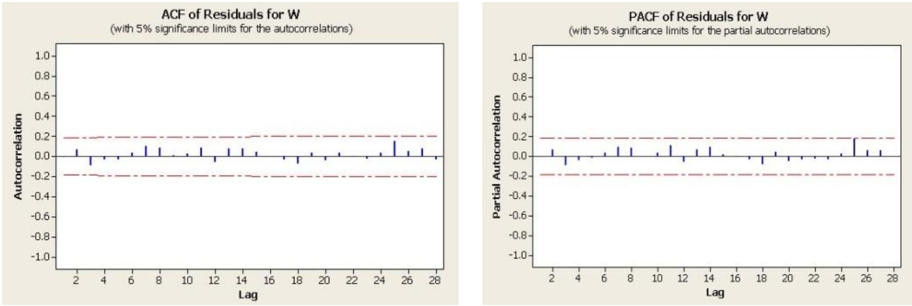
با انتخاب گزینه Four in oneدر قسمت Individual plotsپنجره Graphs همه نمودارهای مربوط به باقیمانده ها بجز نمودار acf و pacf باقیمانده ها، در یک صفحه رسم خواهند شد.

با توجه به نمودارهای شکل(۵–۵۷) فرض نرمال بودن توزیع باقیمانده های حاصل از برازش مدل IMA(۱,۱)و همچنین فرض ثابت بودن واریانس باقیمانده ها پذیرفته می شود.

نمودار باقیمانده ها در برابر زمان نیز طرح خاصی را نشان نمی دهد و رفتار آن مشابه رفتار یک فرآیند اغتشاش خالص با میانگین صفر و واریانس ثابت می باشد.

شکل ۵ – ۵۷ : نمودار های مربوط به باقیمانده های مدلIMA(۱,۱)





شکل ٥- ٥٨ : acf و pacf باقیمانده های مدل(١,١)IMA

نمودارهای فوق حاکی از آن است که باقیمانده های حاصل از برازش مدل(١,١)IMA ناهمبسته هستند. بنابراین، با توجه به آنکه هیچ یک از خود همبستگی ها از حدود استانداردشان تجاوز نکرده اند، می توان فرضیه استقلال باقیمانده ها را پذیرفت.

آزمون پرت-مانتو

خروجی مربوط به این آزمون بصورت زیر است:

Modified Box-Pierce (Ljung-Box) Chi-Square statistic

Lag	١٢	٢٤	٣٦	٤٨
Chi-Square	٥.٥	٨.٩	١٧.٨	٢٨.٢
DF	١٠	٢٢	٣٤	٤٦
P-Value	٠.٨٥٧	٠.٩٩٤	٠.٩٩٠	٠.٩٨٢

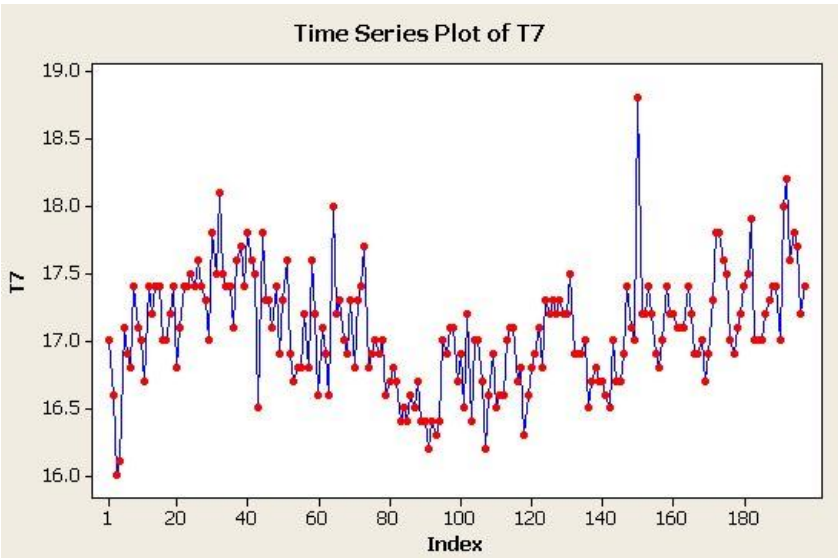
همانطور که ملاحظه می شود مقدار $p-value$ برای تمام تأخیرهای فوق بیشتر از ٠.٠٥ می باشد که فرضیه ناهمبسته بودن باقیمانده ها را مورد تأیید قرار می دهد.

به عنوان مثال در تأخیر ٢٤ مقدار $p-value$ برابر با ٠.٩٩استکه بیشتر از ٠.٠٥ می باشد و فرضیه ناهمبسته بودن باقیمانده ها را تا تأخیر ٢٤، قویا تأیید می کند. همچنین مقدار

آماره کی دو در تأخیر ۲٤ برابر ۸.۹ می باشد، در حالی که $\chi^2_{(0.05,22)} = 33.92$، در نتیجه چون ۸.۹ از مقدار بحرانی متناظر با آن در جدول کی دو کمتر است، فرضیه H_0 رد نمی شود. بنابراین با توجه به بررسی های فوق، مناسبت مدل انتخاب شده مورد تأیید قرار می گیرد.

مثال ۷

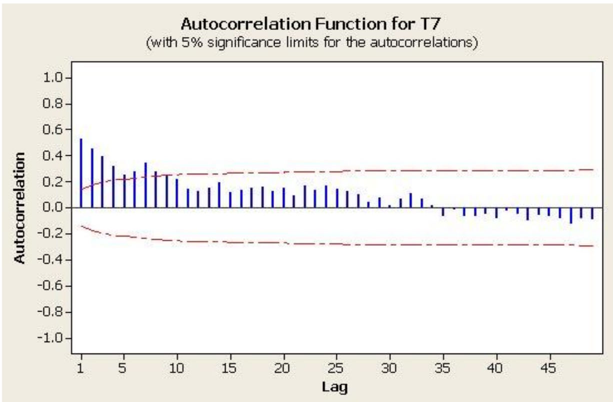
باکس و جنکینز ۱۹۷ غلظت ثبت شده از یک فرآیند شیمیایی را لیست کرده اند(سری T۷ پیوست). برای مدل سازی این مشاهدات ابتدا داده ها را در برابر زمان رسم می کنیم.همانطور که مشاهده می شود داده ها دارای روند می باشند. بنابراین سری زمانی ناایستا است. رسم نمودارهای acf و pacf نیز ناایستایی سری را تأیید می کند.

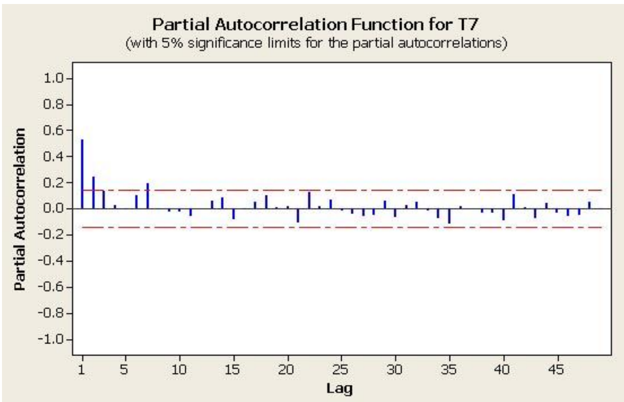


شکل ۵ – ۵۹ : نمودار سری زمانی غلظت یک فرآیند شیمیایی

در حالیکه خود همبستگیهای برآورد شده به کندی به صفر میل می کنند، خود همبستگی های جزئی از مرتبه یک به بعد تقریبا صفر می شوند(تفاوت معنی داری با صفر ندارند) که این خود دلیلی بر ناایستایی سری می باشد.

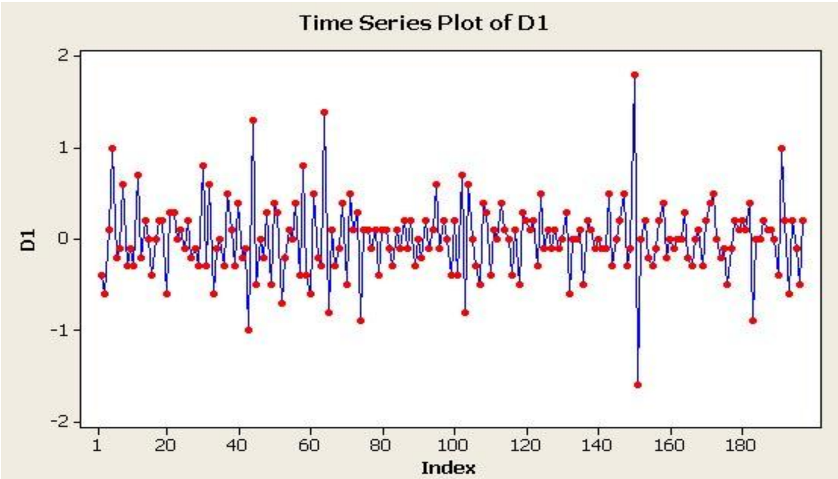
شکل ۵ – ۶۰ : acf و pacfسری زمانیT۷





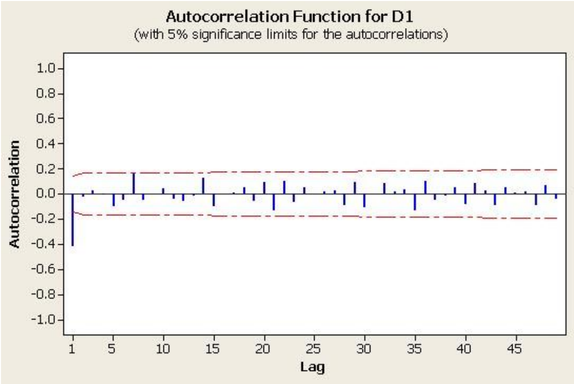
ایستایی در کوواریانس شامل ایستایی در میانگین و ایستایی در واریانس می باشد. بنابراین داده ها باید از لحاظ همگنی واریانس هم بررسی شوند تا اگر تبدیلی لازم باشد، بر روی داده ها صورت گیرد. با توجه به بررسی انجام شده، رویه box-coxتبدیلی را پیشنهاد نمی کند. بنابراین ایستا بودن داده ها در واریانس را می پذیریم و سپس داده ها را از لحاظ ایستایی در میانگین بررسی می کنیم.

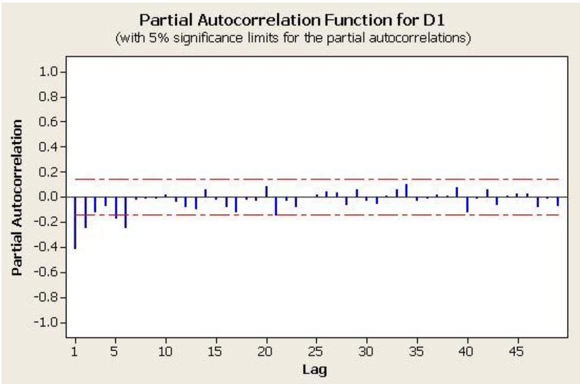
همانطور که گفته شد با توجه به نمودار سری، acfو pacf آن، این سری ناایستا است. بنابراین تفاضلات مرتبه اول سری را محاسبه کرده و سپس ایستایی سری تفاضلی را بررسی می کنیم. سری تفاضلی شده را D۱ می نامیم و نمودار سری زمانی،acf و pacf مربوطه را رسم می کنیم.



شکل ۵ – ٦١ : نمودار سری تفاضلی شده غلظت فرآیند شیمیایی

تفاضل مرتبه اول سری هیچ شواهدی از میانگین و واریانس در حال تغییر، نشان نمی دهد. این مطلب با مشاهده نمودارهایacf و pacf تأیید می شود.





شکل ٥ – ٦٢ : acf و pacf سری تفاضلی شده

اکنون با توجه به نمودارacfوpacfفوق، سه مدل ARIMA(۱,۱,۱) وARIMA(۱,۱,۲)وARIMA(۰,۱,۱)را به داده ها برازش می دهیم. چنانچه مناسبت این مدلها را به روشهای گرافیکی بررسی کنیم می بینیم که تقریبا هر سه مدل برازش خوبی به داده ها دارند و باقیمانده های حاصل از هر یک از این مدلها نیز تقریبا دارای توزیع نرمال می باشند. اما ما می خواهیم با توجه به معیار اطلاعاتی آکائیک، مدلی که برازش بهتری را فراهم می آورد شناسایی کنیم. همانطور که می دانیم، مدل بهتر مدلی است که معیار آکائیک مربوط به آن کمتر باشد.

اطلاعات مربوط به برازش این مدل ها در پنجره session بصورت زیر است:

برازش مدلARIMA(۱،۱،۱)

MTB > ARIMA ۱ ۱ ۱ 'T۷';

Final Estimates of Parameters

Type		Coef	SE Coef	T	P
AR	۱	۰.۲۰۱۷	۰.۰۹۲۸	۲.۱۷	۰.۰۳۱
MA	۱	۰.۸۲۸۷	۰.۰۵۲٦	۱٥.۷٥	۰.۰۰۰

Differencing: ۱ regular difference

Number of observations: Original series ۱۹۷, after differencing ۱۹٦

Residuals: SS = ۲۳.۱۸۱۲ (backforecasts excluded)

MS = ۰.۱۱۹٥ DF = ۱۹٤

Modified Box-Pierce (Ljung-Box) Chi-Square statistic

Lag	۱۲	۲٤	۳٦	٤۸
Chi-Square	۱۱.۸	۲۰.۹	۳۲.۲	٤۰.۳
DF	۱۰	۲۲	۳٤	٤٦
P-Value	۰.۲۹٥	۰.٥۲٥	۰.٥٥۸	۰.۷۰۷

با توجه پارامترهای برآورد شده، مدل ARIMA(۱,۱,۱) بصورت زیر خواهد بود:

$$x_t = 0.2017x_{t-1} + z_t - 0.8287z_{t-1}$$

که در آن z_t فرآیند تصادفی محض با میانگین صفر و واریانس ۰.۱۱۹۵ می باشد. این مدل را M۱ می نامیم.

برازش مدل(۱،۱،۲)ARIMA

MTB > ARIMA ۱ ۱ ۲ 'T۷';

Final Estimates of Parameters

Type		Coef	SE Coef	T	P
AR	۱	۰.٦٦٦٨	۰.۰٦۰۳	۱۱.۰٥	۰.۰۰۰
MA	۱	۱.۳۰۹٤	۰.۰۰٤٤	۲۹۹.۷۹	۰.۰۰۰
MA	۲	-۰.۳٦٤۸	۰.۰۲۲٦	-۱٦.۱٥	۰.۰۰۰

Differencing: ۱ regular difference

Number of observations: Original series ۱۹۷, after differencing ۱۹٦

Residuals: SS = ۲۳.۰۳٦۹ (backforecasts excluded)

MS = ۰.۱۱۹٤ DF = ۱۹۳

Modified Box-Pierce (Ljung-Box) Chi-Square statistic

Lag	۱۲	۲٤	۳٦	٤۸
Chi-Square	۱۱.۷	۱۹.٦	۳۱.۳	۳۸.۸
DF	۹	۲۱	۳۳	٤٥
P-Value	۰.۲۳۳	۰.٥٤۸	۰.٥٥۲	۰.۷۳۰

مدل حاصل بصورت زیر خواهد بود:

$$x_t = 0.6668x_{t-1} + z_t - 1.3094z_{t-1} + 0.3648z_{t-2}$$

که در آن z_t فرآیند تصادفی محض با میانگین صفر و واریانس ۰.۱۱۹٤ می باشد. این مدل را M۲ می نامیم.

برازش مدل(۰،۱،۱)ARIMA

MTB > ARIMA ۰ ۱ ۱ 'T۷';

Final Estimates of Parameters

Type	Coef	SE Coef	T	P
MA ۱	۰.۷۰۸۲	۰.۰۵۰۵	۱٤.۰۳	۰.۰۰۰

Differencing: ۱ regular difference

Number of observations: Original series ۱۹۷, after differencing ۱۹٦

Residuals: SS = ۲۳.۵۱٤٤ (backforecasts excluded)

MS = ۰.۱۲۰٦ DF = ۱۹۵

Modified Box-Pierce (Ljung-Box) Chi-Square statistic

Lag	۱۲	۲٤	۳٦	٤۸
Chi-Square	۱۸.۰	۲۵.٤	۳۷.٦	٤۵.۱
DF	۱۱	۲۳	۳۵	٤۷

P-Value ۰.۰۸۱ ۰.۳۳۱ ۰.۳۵۱ ۰.۵۵۱

مدل حاصل بصورت $x_t = z_t - 0.7082 z_{t-1}$ خواهد بود که در آن z_t فرآیند تصادفی محض با میانگین صفر و واریانس ۰.۱۲۰۶ می باشد. این مدل را M۳ می نامیم.

نکته ۱:

در برازش مدلهای فوق ما جمله ثابت را با برداشتن چک مارک عبارت Includeconstant term in model در پنجره اصلی ARIMA نادیده گرفته ایم. زیرا در هیچ یک از مدل های برازش داده شده جمله ثابت معنی دار نبوده است. ما ابتدا مدل را با جمله ثابت برازش می دهیم و سپس با توجه به $p - value$ مربوطه درباره حضور یا عدم حضور جمله ثابت در مدل تصمیم می گیریم. همانطور که می دانید چنانچه $p - value > \alpha$ باشد، فرض H_0 که می گوید پارامتر مورد نظر(در اینجا جمله ثابت) صفر است، پذیرفته می شود. α سطح معنی داری است که معمولا ۰.۰۵ در نظر گرفته می شود. همچنین با توجه به آماره tنیز می توان فرضیه صفر بودن جمله ثابت را پذیرفت.

به عنوان مثال مدل اول یعنی مدل ARIMA(۱,۱,۱)را با جمله ثابت برازش می دهیم(مدل با جمله ثابت، حالت پیش فرض مینی تب می باشد). در این صورت خواهیم داشت :

```
MTB > ARIMA ۱ ۱ ۱ 'TV';
```

Final Estimates of Parameters

Type		Coef	SE Coef	T	P
AR	۱	۰.۲۰٤٤	۰.۰۹۲۱	۲.۲۲	۰.۰۲۸
MA	۱	۰.۸۳٤۰	۰.۰۵۱٤	۱٦.۲۱	۰.۰۰۰
Constant		۰.۰۰۳۰۰۰	۰.۰۰٤۱٤۳	۰.۷۲	۰.٤۷۰

همانطور که ملاحظه می شود، در ردیف Constantمقدار $p-value$ برابر ۰.٤۷ می باشد که بزرگتر از ۰.۰۵ است. بنابراین می توان فرض صفر بودن جمله ثابت را پذیرفت. همچنین مقدار آماره t برابر ۰.۷۲ است که کمتر از ۲ می باشد.

همانطور که قبلا گفته شد وجود جمله ثابت در مدل به معنای وجود روند قطعی در مدل می باشد.

نکته۲ :

همانطورکه قبلا هم گفته شدمدل(MA(qکه ما آن را بصورت $x_t = z_t + \beta_1 z_{t-1} + ... + \beta_q z_{t-q}$ تعریف کرده ایم، در مینی تب بصورت زیر تعریف شده است:

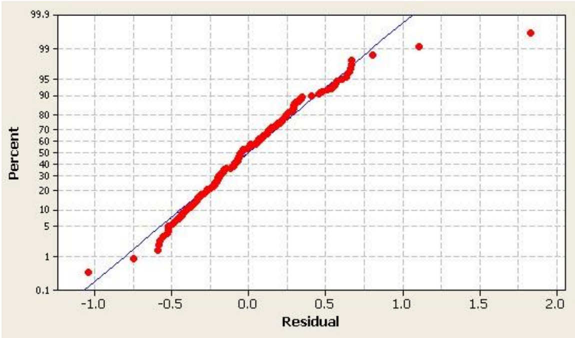
$$X_t = \varepsilon_t - MA1\varepsilon_{t-1} - MA2\varepsilon_{t-2} - ....$$

بنابراین ضرایب MA_i محاسبه شده در مینی تب قرینه ضرایبی است که ما تعریف کرده ایم. یعنی $\beta_i = -MA_i$. به عنوان مثال در مدل اول مقدار برآورد شده برای پارامتر میانگین متحرک برابر ۰.۸۲۸۷ می باشد، اما ما در مدل آنرا بصورت ۰.۸۲۸۷- نوشته ایم.

بررسی مناسبت مدل

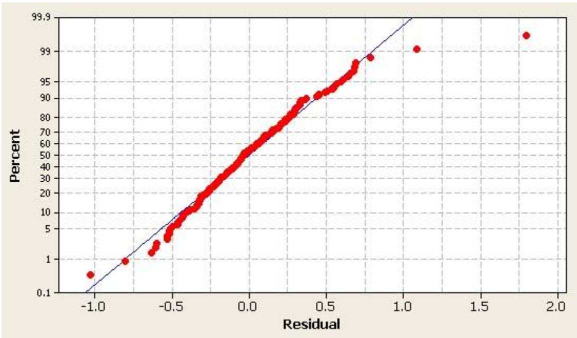
اکنون نمودار احتمال نرمال مربوط به هر یک از این سه مدل را رسم می کنیم تا نرمال بودن باقیمانده های حاصل از هر مدل را ملاحظه کنیم. بررسی سایر فرضهای مربوط به باقیمانده ها(استقلال باقیمانده ها، ثابت بودن واریانس باقیمانده ها و فاقد ساختار بودن نمودار باقیمانده ها در طول زمان) را به خواننده واگذار می کنیم.

۱- نمودار احتمال نرمال برای باقیمانده های مدل (ARIMA(۱,۱,۱



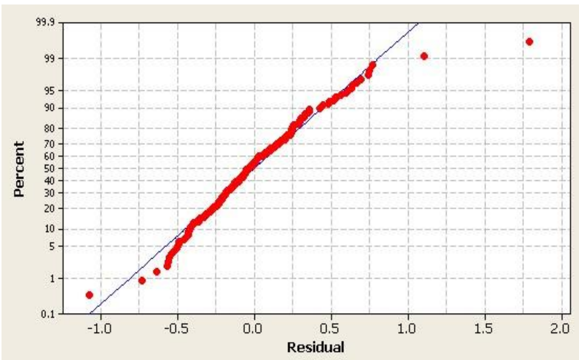
شکل ۵ – ٦٣

۲- نموداراحتمال نرمال برای باقیمانده هایمدل ARIMA(۱,۱,۲)



شکل ۵ – ٦٤

نموداراحتمال نرمال برای باقیمانده های مدل ARIMA(۰,۱,۱)



شکل ۵ – ٦٥

همانطور که ملاحظه می شود نمی توان فرض نرمال بودن باقیمانده ها را برای هیچ یک از سه مدل فوق رد کرد.

انتخاب مدل مناسب براساس معیار آکائیک

اکنون معیار آکائیک را با توجه به رابطه $AIC(M) = n\ln\hat{\sigma}_e^2 + 2M$ برای هر مدل محاسبه می کنیم. σ_e^2 واریانس مانده های مدل و $\hat{\sigma}_e^2$ برآورد درستنمائی ماکزیمم σ_e^2 است. اما همانطور که گفته شد می توان این معیار را در مینی تب بصورت زیر نوشت :

$$AIC = n\ln(ssr) + 2(p+q)$$

که در آن ssr بیانگر مجموع مربعات باقیمانده ها می باشد.می توان بجای آن از msr یا میانگین مجموع مربعات باقیمانده ها نیز استفاده کرد. همانطور که می دانید میانگین مجموع مربعات باقیمانده ها برآوردی برای واریانس باقیمانده ها می باشد.

$$AIC(M1) = 197\log 0.1195 + 4 = -177.75$$
$$AIC(M2) = 197\log 0.1194 + 6 = -175.83$$
$$AIC(M3) = 197\log 0.1206 + 2 = -178.96$$

همانطور که ملاحظه می شود، مدل M۳ نسبت به دو مدل دیگر دارای معیار آکائیک کمتری می باشد. بنابراین ما این مدل را به عنوان مدل مناسب انتخاب می کنیم.

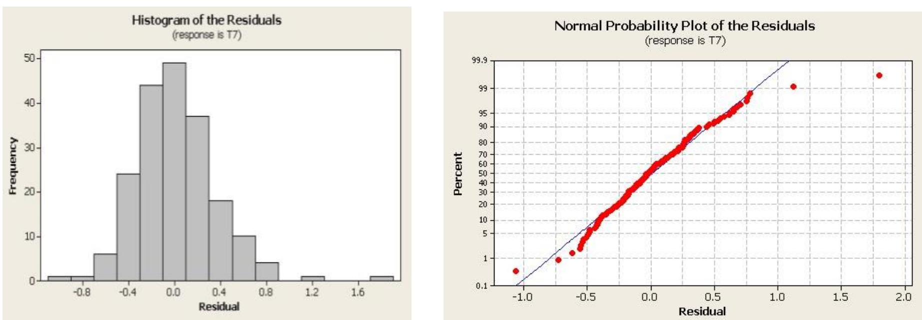
بررسی مناسبت مدل انتخاب شده بر اساس معیار آکائیک

قبلا نمودار احتمال نرمال را برای مدل M۳ دیدیم. این نمودار فرض نرمال بودن باقیمانده ها را تأیید می کند. اکنون برای بررسی دقیق تر، سایر نمودارهای مربوط به باقیمانده های مدل انتخاب شده را بررسی می کنیم. هدف ما از این کار بررسی فرضیات مربوط به باقیمانده ها با استفاده از روشهای نموداری می باشد.

برای این منظور در پنجره اصلی ARIMA مدل ARIMA(۰,۱,۱)را مشخص می کنیم و سپس در قسمت پایین صفحه بر روی گزینه Graphsکلیک کرده و همه نمودارهای مربوط به باقیمانده ها را انتخاب می کنیم. این باقیمانده ها، باقیمانده های حاصل از مدل برازش داده شده است.

الف) بررسی فرض نرمال بودن باقیمانده ها

بررسی این فرض با توجه به نمودارهای Histogram of residualsو Normalplot of residualsانجام می شود. همانطور که در شکل(۵– ٦٦) ملاحظه می شود، دلیلی بر رد فرض نرمال بودن باقیمانده ها نداریم.

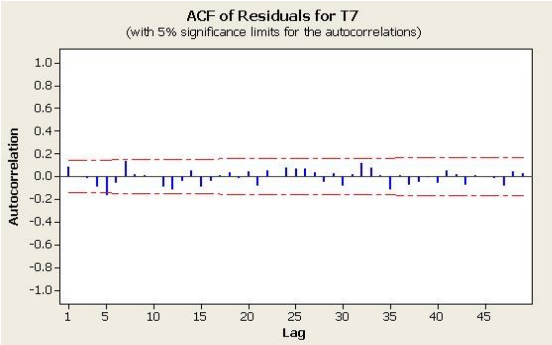


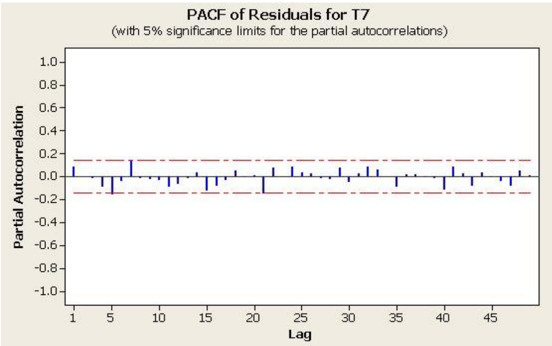
شکل ۵ – ٦٦ : هیستوگرام و نمودار احتمال نرمال باقیمانده های مدل ARIMA(۰,۱,۱)

ب) بررسی فرض استقلال باقیمانده ها

بررسی این فرض با رسم acf و pacf باقیمانده ها انجام می شود. همانطور که ملاحظه می شود، هیچ یک از خود همبستگی ها معنی دار نیستند.

بنابراین می توان فرض ناهمبسته بودن باقیمانده ها را پذیرفت که این خود دلیلی بر کفایت مدل می باشد.



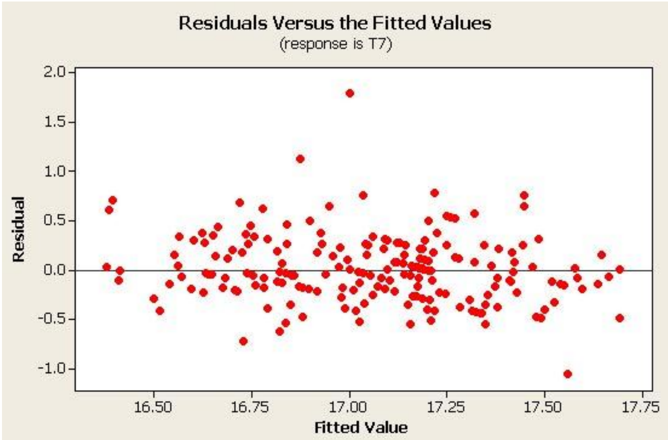


شکل ۵ – ٦٧ :

acf و pacfباقیمانده های مدلARIMA(۰,۱,۱)

ج) بررسی فرض ثابت بودن واریانس باقیمانده ها

بررسی این فرض با رسم نمودار باقیمانده ها در برابر مقادیر برازش شده(Residuals versus fits) انجام می شود. چنانچه این نمودار ساختار خاصی را نشان ندهد، مثلا بصورت قیفی شکل نباشد، می توان فرض ثابت بودن واریانس باقیمانده ها را پذیرفت. همانطور که در نمودار زیر دیده می شود، پراکندگی حول صفر است و نمودار قیفی شکل نمی باشد. لذا فرض ثابت بودن واریانس باقیمانده ها را می پذیریم.



شکل ۵ – ٦٨ : نمودار باقیمانده های مدل ARIMA(۰,۱,۱) در برابر مقادیر برازش شده

ر) آزمون پرت- مانتو

در این قسمت مناسبت مدل انتخاب شده را بر اساس آماره Q اصلاح شده مورد آزمون قرار می دهیم. در قسمت انتهایی خروجی مینی تب می توان مقادیر مختلف این آماره را مشاهده نمود.

```
Modified Box-Pierce (Ljung-Box) Chi-Square statistic

Lag            ۱۲     ۲٤     ۳٦     ٤٨
Chi-Square     ۱۸.۰   ۲٥.٤   ۳۷.٦   ٤٥.۱
DF             ۱۱     ۲۳     ۳٥     ٤۷
P-Value        ۰.۰۸۱  ۰.۳۳۱  ۰.۳٥۱  ۰.٥٥۱
```

به عنوان مثال برای $k=۱۲$ مقدار آماره کی دو برابر ۱۸ می باشد. اما مقدار متناظر از جدول توزیع کی دو عبارت است از $\chi^2_{0.05}(11)=19.67$. بنابراین مقدار مشاهده شده کی دو در سطح معنی داری ۰.۰۵ با ۱۱ درجه آزادی معنی دار نیست. همچنین چون $p-value$ مربوطه (۰.۰۸۱) بیشتر از ۰.۰۵ می باشد می توان فرضیه صفر را پذیرفت.

در نتیجه می توان گفت که برازش مدل(۰,۱,۱)ARIMAبرای داده ها مناسب است.

یادآوری

آماره Q اصلاح شده که در مینی تب آن را آماره LBQ می نامیم، فرضیه صفر بودن همه خود همبستگی ها را تا تأخیر kآزمون می کند. این آزمون را که براساس آماره Q اصلاح شده بنا شده است، آزمون پرت–مانتومی نامند. فرضیه صفر این آزمون به شکل زیر است :

$$H_0 : \rho_1 = \rho_2 = ... = \rho_k = 0$$

چنانچه این فرض پذیرفته شود به این معنی است که باقیمانده ها ناهمبسته هستند و این به معنی تصادفی بودن باقیمانده ها است. تصادفی بودن باقیمانده ها دلیلی بر مناسب بودن مدل انتخاب شده می باشد. یعنی مدلی که این باقیمانده ها را تولید کرده است. همچنین می توان گفت ناهمبسته بودن باقیمانده ها به این معنی است که ما توانسته ایم همبستگی موجود در مشاهدات را به خوبی به مدل درآوریم.

در کنار روشهای نموداری برای بررسی مناسبت مدل، مانند رسم نمودار احتمال نرمال باقیمانده ها و رسم هیستوگرام باقیمانده ها، استفاده از این آزمون نیز به عنوان یک روش غیرنموداری و عددی می تواند مفید باشد.

بررسی مناسبت مدل انتخاب شده با استفاده از شبیه سازی

یکی از راه های بررسی مناسبت مدل، شبیه سازی بر مبنای آن مدل است. چنانچه مدلی مناسب باشد، داده های شبیه سازی شده بر پایه آن مدل، مناسبت مدل را تأیید

خواهند نمود. می توان با شبیه سازی نمونه ای به حجم ۵۰۰ براساس مدل انتخاب شده یعنی ARIMA(۰,۱,۱) و رسم نمودارهای مربوطه مناسبت مدل را بررسی کرد. برنامه شبیه سازی با استفاده از نرم افزار sas بصورت زیر است :

```
data simul;
Y۱=۰.۰۰۱ ;    Z۱=۰.۰۲;
do    t=۱     to    ٥۰۰;
Z = normal (٧٦٤٥٣٢١٦);
Y = Y۱ + Z -۰.٧۰٨٢ * Z۱;
output;
Z۱ = Z ;  Y۱ = Y ;
end ;
proc gplot ;
```

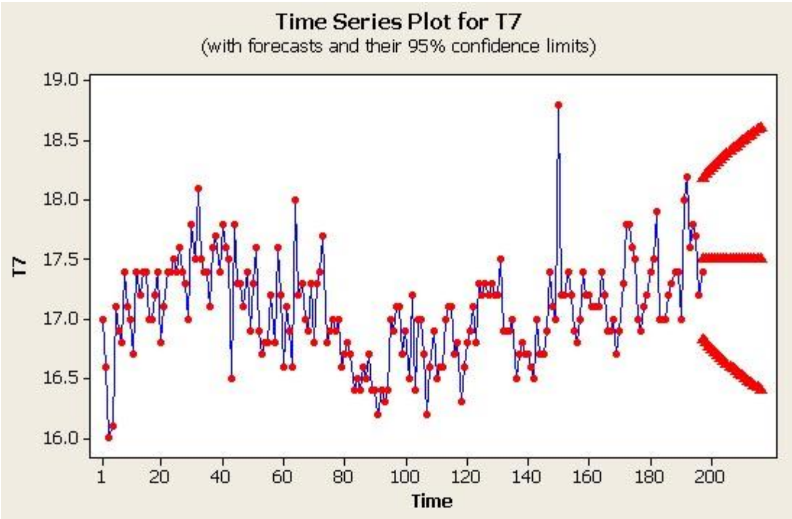
پیش بینی

حال که با روشهای مختلف مناسبت مدل انتخاب شده مورد تأیید قرار گرفته است، می توانیم این مدل را مبنای پیش بینی رفتار آینده سری قرار دهیم. برای این کار در پنجره اصلی ARIMA گزینه Forecasts را انتخاب می کنیم. به عنوان مثال پیش بینی ۲۰ مقدار آینده سری بصورت زیر می باشد.

```
                    Forecasts from period ۱۹٧

                                    ۹٥ Percent
                                        Limits
Period  Forecast    Lower    Upper  Actual
                 ۱۹٨    ۱٧.٥۰٦۱  ۱٦.٨٢٥٤  ۱٨.۱٨٦۹
                 ۱۹۹    ۱٧.٥۰٦۱  ۱٦.٧۹٧۰  ۱٨.٢۱٥٣
```

٢٠٠	١٧.٥٠٦١	١٦.٧٦٩٧	١٨.٢٤٢٦
٢٠١	١٧.٥٠٦١	١٦.٧٤٣٣	١٨.٢٦٨٩
٢٠٢	١٧.٥٠٦١	١٦.٧١٧٩	١٨.٢٩٤٤
٢٠٣	١٧.٥٠٦١	١٦.٦٩٣٢	١٨.٣١٩٠
٢٠٤	١٧.٥٠٦١	١٦.٦٦٩٣	١٨.٣٤٢٩
٢٠٥	١٧.٥٠٦١	١٦.٦٤٦٠	١٨.٣٦٦٢
٢٠٦	١٧.٥٠٦١	١٦.٦٢٣٤	١٨.٣٨٨٨
٢٠٧	١٧.٥٠٦١	١٦.٦٠١٣	١٨.٤١٠٩
٢٠٨	١٧.٥٠٦١	١٦.٥٧٩٨	١٨.٤٣٢٥
٢٠٩	١٧.٥٠٦١	١٦.٥٥٨٧	١٨.٤٥٣٥
٢١٠	١٧.٥٠٦١	١٦.٥٣٨١	١٨.٤٧٤٢
٢١١	١٧.٥٠٦١	١٦.٥١٧٩	١٨.٤٩٤٣
٢١٢	١٧.٥٠٦١	١٦.٤٩٨١	١٨.٥١٤١
٢١٣	١٧.٥٠٦١	١٦.٤٧٨٧	١٨.٥٣٣٥
٢١٤	١٧.٥٠٦١	١٦.٤٥٩٧	١٨.٥٥٢٥
٢١٥	١٧.٥٠٦١	١٦.٤٤١٠	١٨.٥٧١٢
٢١٦	١٧.٥٠٦١	١٦.٤٢٢٦	١٨.٥٨٩٦
٢١٧	١٧.٥٠٦١	١٦.٤٠٤٦	١٨.٦٠٧٧

با توجه به مقادیر پیش بینی، چون پیش بینی ها بین حدود ۰.۹۵ پیش بینی قرار دارند، لذا از این جنبه نیز می توان فهمید که مدل انتخاب شده درست است. نمودار سری زمانی پیش بینی شده نیز بصورت زیر است. برای رسم این نمودار در پنجره Graphsاولین گزینه یعنی(Time series plot (including optional forecasts را مارک دار می کنیم.



شکل ٥ – ٦٩ : پیش بینی سری T۷ بر اساس مدل ARIMA(۰,۱,۱)

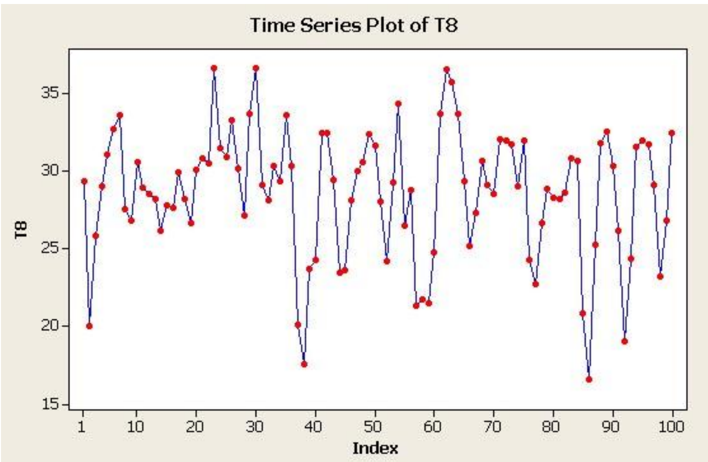
در این نمودار خط وسط مقادیر پیش بینی و خطوط بالایی و پائینی به ترتیب حد بالایی فاصله اطمینان ۹۵ درصد پیش بینی و حد پائینی فاصله اطمینان ۹۵ درصد پیش بینی را نشان می دهند.

مثال ۸

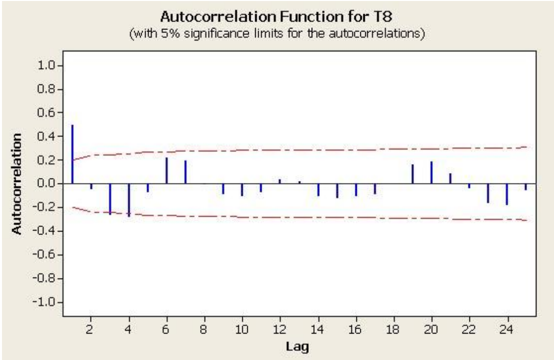
در یک فرآیند شیمیایی چسبندگی محصول یک ویژگی کیفی مهم است. آمار ۱۰۰ رقم اخیر چسبندگی در دسترس بوده و در پیوست با نام T۸ لیست شده اند. ما می خواهیم یک مدل سری زمانی برای این فرآیند توسعه دهیم به نحوی که بتوان یک طرح کنترلی را برای خروجی فرآیند گسترش داد.

این مثال نمونه ای از کاربرد سری های زمانی را در کنترل فرآیند نشان می دهد. در کنترل فرآیند، مسأله این است که با اندازه گیری متغیری که کیفیت فرآیند را نشان می دهد، تغییرات در اجرای فرآیند را جستجو کنیم. این تغییرات را می توان در مقابل زمان رسم کرد. وقتی اندازه ها از مقدار مشخصی خیلی دور باشند، برای کنترل فرآیند باید عمل اصلاح کننده مناسبی در نظر گرفته شود.

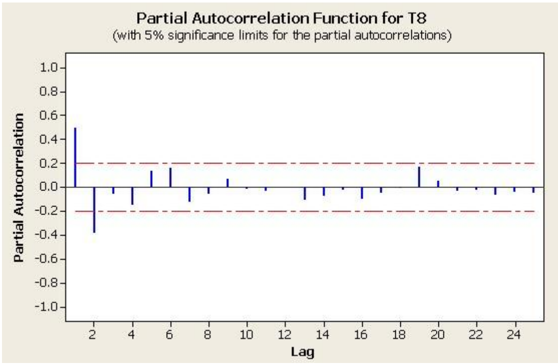
تشخیص مدل آزمایشی



شکل ۵ – ۷۰: نمودار سری زمانی چسبندگی یک فرآیند شیمیایی

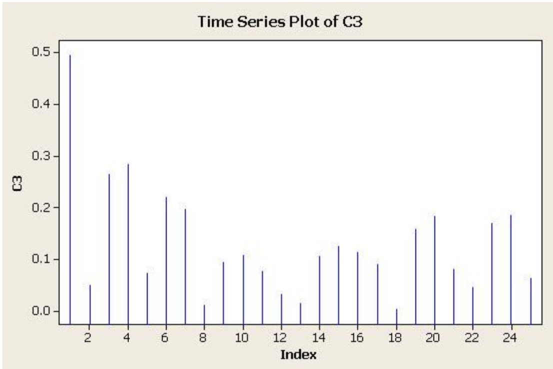


شکل ۵ – ۷۱ : acfو pacfسری زمانی چسبندگی



همانطور که ملاحظه می شود تابع خود همبستگی نمونه با یک کاهش سینوسی قطع می شود در حالی که تابع خود همبستگی جزئی نمونه پس از دو تأخیر قطع می شود. بنابراین ما به طور آزمایشی مدل سری تحت بررسی را به صورت مدل(AR(۲ایستا مشخص می کنیم.

گاهی اوقات در تشخیص مدل مفید خواهد بود که نمودار قدر مطلق مقادیر خود همبستگی و خود همبستگی جزئی را رسم کنیم تا اثر همبستگی منفی موجب سردرگمی تحلیلگر نشود. چنین نمودارهایی باعث می شوند که کاهش تابع خود همبستگی و تابع خود همبستگی جزئی آسانتر مشاهده شود. برای مثال نمودار مقادیر $|r_k|$ را برای این سری زمانی رسم نموده ایم. در این نمودار کاهش سینوسیتابع خود همبستگی واضح تر است.



شکل ۵ – ۷۲: نمودار قدر مطلق مقادیر تابع خود همبستگی سری T۸

برای رسم چنین نموداری در مینی تب ابتدا در پنجره اصلی Autocorrelation با انتخاب گزینه Store ACF مقادیر تابع خود همبستگی را در یکی از ستونهای مینی تب(مثلا C۲) ذخیره می کنیم. سپس از منوی Calc گزینه Calculator را انتخاب می کنیم. از کادر Functions گزینه Arithmetic و سپس گزینه absolute value را انتخاب می کنیم.

با انتخاب گزینه select تابع انتخاب شده به کادر Expression منتقل می شود. حال با دو با کلیک کردن بر روی متغیری که قرار است قدر مطلق مقادیر آن محاسبه شودآنرا به پرانتز تابع قدر مطلق منتقل می کنیم. با *ok* کردن این پنجره قدر مطلق مقادیر تابع خود همبستگی محاسبه می شود. در قسمت Store result in variable می توان C۳ را وارد کرد تا مقادیر تبدیل شده در این ستون ذخیره شوند.

حال کافی است نمودار سری زمانی را برای متغیر C۳ رسم کنیم تا نموداری به شکل فوق داشته باشیم.

برازش مدل(۲)AR

پس از شناسایی مدل آزمایشی، مرحله دوم در مدل سازی یک سری زمانی تخمین پارامترهای آن مدل است که این کار را نرم افزار انجام می دهد. ما مدل مورد نظر را بصورت زیر به نرم افزار معرفی می کنیم.

شکل ۵ – ۷۳ : تکمیل پنجره آریما برای برازش مدل (۲)AR

خروجی در پنجره session به شکل زیر خواهد بود.

ARIMA Model: T۹

Estimates at each iteration

Iteration	SSE	Parameters		
۰	۱۵۹۰.٦۷	۰.۱۰۰	۰.۱۰۰	۲۲.۹۳۵
۱	۱۳٦٦.٦۲	۰.۲۵۰	−۰.۰۲۰	۲۲.۰۷۱
۲	۱۲۰۱.۹۳	۰.٤۰۰	−۰.۱٤٤	۲۱.۲۹۷
۳	۱۰۹۷.۵۱	۰.۵۵۰	−۰.۲۷۱	۲۰.٦۵۱
٤	۱۰۵۵.۱۹	۰.٦۹۸	−۰.٤۰۵	۲۰.۲٦۰
۵	۱۰۵٤.۱۷	۰.۷۱۵	−۰.٤۲۹	۲۰.٤٤۲
٦	۱۰۵٤.۱۳	۰.۷۱۸	−۰.٤۳۳	۲۰.٤۹۱
۷	۱۰۵٤.۱۳	۰.۷۱۹	−۰.٤۳٤	۲۰.۵۰۱

```
٨   ١٠٥٤.١٣   ٠.٧١٩   -٠.٤٣٤   ٢٠.٥٠٣


Relative change in each estimate less than ٠.٠٠١٠

Final Estimates of Parameters


Type          Coef   SE Coef       T       P
AR   ١      ٠.٧١٨٦    ٠.٠٩٢٣    ٧.٧٨   ٠.٠٠٠
AR   ٢     -٠.٤٣٤٣    ٠.٠٩٢٢   -٤.٧١   ٠.٠٠٠
Constant  ٢٠.٥٠٢٨    ٠.٣٢٧٩   ٦٢.٥٢   ٠.٠٠٠
Mean       ٢٨.٦٥٠٦   ٠.٤٥٨٣


Number of observations:  ١٠٠
Residuals:    SS =  ١٠٤٢.٨٢ (backforecasts excluded)
MS =  ١٠.٧٥   DF = ٩٧


Modified Box-Pierce (Ljung-Box) Chi-Square statistic

Lag              ١٢      ٢٤      ٣٦      ٤٨
Chi-Square     ١٢.٢    ٢١.١    ٢٨.٣    ٣٨.٠
DF                ٩      ٢١      ٣٣      ٤٥
P-Value       ٠.٢٠٤   ٠.٤٥١   ٠.٧٠٠   ٠.٧٦٠
```

همانطور که ملاحظه می شود همه پارامترها معنی دار می باشند. بنابراین نیازی به برازش مجدد مدل بدون جمله ثابت نمی باشد.

همانطور که می دانید مدل اتورگرسیو مرتبه دوم در حالتی که میانگین فرآیند مخالف صفر باشد به شکل زیر می باشد.

$$x_t - \mu = \alpha_1(x_{t-1} - \mu) + \alpha_2(x_{t-2} - \mu) + z_t$$

با ساده کردن این مدل خواهیم داشت:

$$x_t = \mu(1 - \alpha_1 - \alpha_2) + \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + z_t$$

برآوردهای نهایی مربوط به پارامترهای α_1, α_2 و جمله ثابت که در اینجا بصورت $\mu(1-\alpha_1-\alpha_2)$ می باشد در قسمت Final Estimates of Parametersدر ستون coef آمده است. همانطور که ملاحظه می شود میانگین فرآیند نیز ۲۸.٦٥ می باشد . بنابراین مدل ما به شکل زیر خواهد بود.

$$x_t = 20.5028 + 0.7186 x_{t-1} - 0.4343 x_{t-2} + z_t$$

می توان مدل آزمایشی را برای تولید مقادیر برازش شده سری زمانی به کار برد. به عنوان مثال اولین مقدار برازش شده بصورت زیر محاسبه می شود:

$$\hat{x}_3 = 20.5028 + 0.7186 x_2 - 0.4343 x_1 =$$
$$20.5028 + 0.7186(19.98) - 0.4343(29.33) = 22.12$$

بنابراین با قیمانده مربوطه عبارت خواهد بود از:

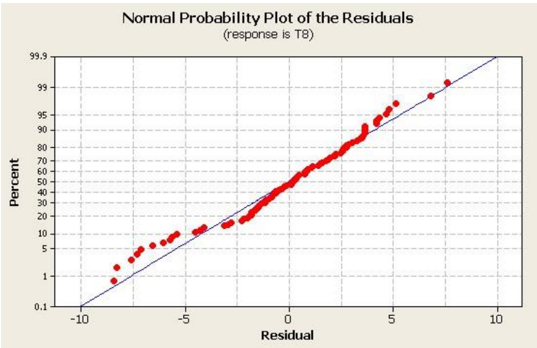
$$x_3 - \hat{x}_3 = 25.76 - 22.12 = 3.64$$

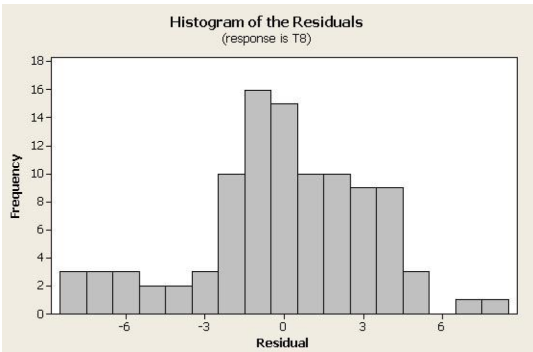
بررسی مناسبت مدل

همانطور که می دانید برای ارزیابی یک مدل باید باقیمانده های حاصل از برازش آن مدل را مورد تجزیه و تحلیل قرار داد. چنانچه مدل مناسب باشد، باید باقیمانده ها فاقد ساختار باشند. یعنی باید هیچگونه طرح مشخصی نداشته باشند. در نتیجه باید خود همبستگی مانده ها نزدیک صفر باشد.

برای تجزیه و تحلیل باقیمانده ها بر اساس نمودارهای مربوطه، پس از باز کردن پنجره ARIMA ابتدا باید نوع مدل را مشخص کنیم. در اینجا مدل مورد نظر، مدل AR(۲) می باشد. سپس بر روی گزینه Graphsدر پایین صفحه کلیک می کنیم و نمودارهای مورد نظر را انتخاب می کنیم.

الف) بررسی فرض نرمال بودن باقیمانده ها





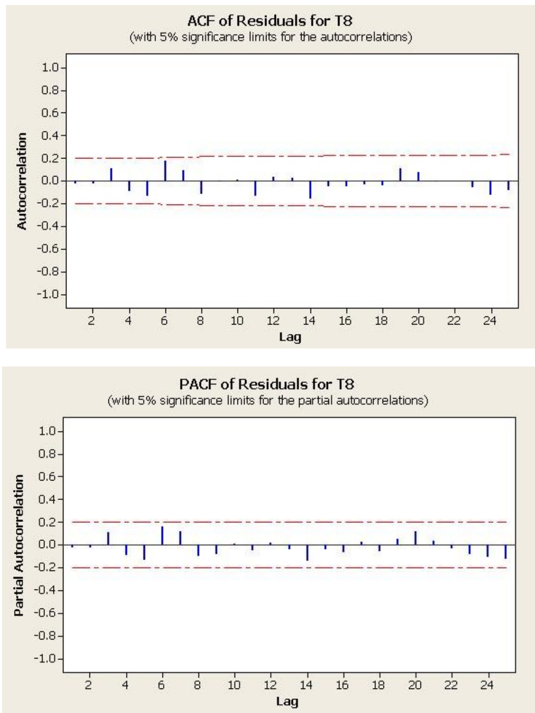
شکل ۵ – ۷۴ : نمودار احتمال نرمال و هیستوگرام باقیمانده های مدل AR(۲)

همانطور که ملاحظه می شود در نمودار احتمال نرمال نقاط تقریبا در امتداد یک خط راست قرار گرفته اند. البته همانطور که قبلا هم گفته شد در تأیید خط راست روی مقادیر مرکزی نسبت به کرانها بیشتر تأکید داریم.

همچنین نمودار هیستوگرام تقریبا مشابه هیستوگرام یک جامعه نرمال می باشد. بنابراین می توان فرضیه نرمال بودن توزیع باقیمانده های حاصل از برازش مدلAR(۲)را پذیرفت.

ب) بررسی فرض استقلال باقیمانده ها

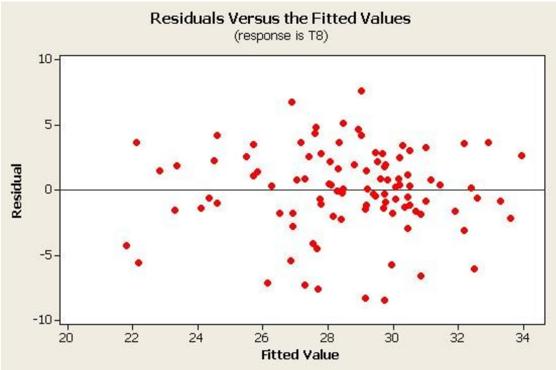
برای بررسی این فرض acf و pacf باقیمانده ها را بررسی می کنیم.



شکل ۵– ۷۵ : acfو pacfباقیمانده های مدل(۲)AR

همانطور که ملاحظه می شود هیچ یک از خود همبستگی ها از حدود دو برابر خطای استانداردشان تجاوز نکرده اند. بنابراین می توان فرضیه استقلال باقیمانده ها را پذیرفت.

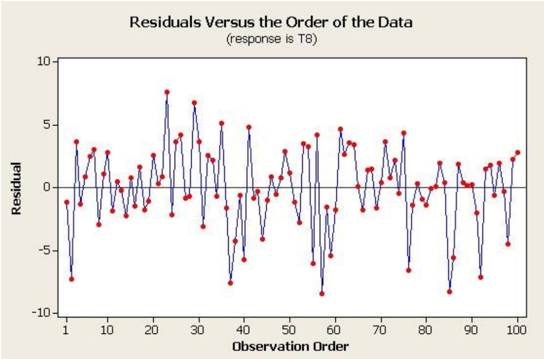
ج) بررسی فرض ثابت بودن واریانس باقیمانده ها



شکل ۵ – ۷۶ : نمودار باقیمانده های مدل AR(۲)در برابر مقادیر برازش شده

این نمودار طرح خاصی را نشان نمی دهد. بنابراین می توان فرضیه ثابت بودن واریانس باقیمانده ها را پذیرفت.

د) رسم نمودار باقیمانده ها در برابر زمان



شکل ۵– ۷۷ : نمودار باقیمانده های مدل AR(۲)در برابر زمان

این نمودار مشابه ضربات تصادفی ناشی از فرآیند تصادفی محض با میانگین صفر و واریانس ثابت می باشد. بنابراین می توان مدل برازش داده شده را به عنوان یک مدل مناسب پذیرفت.

ر) آزمون پرت–مانتو

خروجی این آزمون بصورت زیر می باشد:

```
Modified Box-Pierce (Ljung-Box) Chi-Square statistic
```

Lag	۱۲	۲٤	۳٦	٤۸
Chi-Square	۱۲.۲	۲۱.۱	۲۸.۳	۳۸.۰
DF	۹	۲۱	۳۳	٤٥
P-Value	۰.۲۰٤	۰.٤٥۱	۰.۷۰۰	۰.۷٦۰

همانطور که ملاحظه می شود $p-value$ در تمام تأخیر های فوق بیشتر از ۰.۰۵ می باشد. به عنوان مثال در تأخیر ۲٤ مقدار $p-value$ عبارت است از ۰.٤٥ که بیشتر از

۰.۰۵ می باشد و فرضیه صفر بودن همه خود همبستگی ها تا تأخیر ۲٤ را مورد تأیید قرار می دهد.

همچنین مقدار آماره کی دو برابر با ۲۱.۱ است که از مقدار بحرانی متناظرش در جدول کی دو کمتر است ($\chi^2_{(0.05,21)} = 32.67$).

بنابراین با استفاده از آماره کی دو نیز فرضیه صفر بودن خود همبستگی های باقیمانده ها یا به عبارتی ناهمبسته بودن باقیمانده ها مورد تأیید قرار می گیرد.

برازش جامع تر

برای بررسی بیشتر صحت مدل شناسایی شده مدل جامع تر AR(۳)و ARMA(۲,۱) را به داده ها برازش می دهیم.

الف) برازش مدلAR(۳)

نتایج مربوط به برآورد پارامترها بصورت زیر می باشد:

Final Estimates of Parameters

Type		Coef	SE Coef	T	P
AR	۱	۰.٦۹۷۱	۰.۱۰۲۱	٦.۸۳	۰.۰۰۰
AR	۲	-۰.۳۹٥٥	۰.۱۱۷۹	-۳.۳٦	۰.۰۰۱
AR	۳	-۰.۰٥۳٦	۰.۱۰۲۷	-۰.٥۲	۰.٦۰۳
Constant		۲۱.٥٤۷۱	۰.۳۲۹۳	٦٥.٤۳	۰.۰۰۰
Mean			۲۸.٦٥۲٦	۰.٤۳۷۹	

همانطور که ملاحظه می شود برآورد پارامتر اضافی α_3 که مینی تب آن را با AR۳نشان می دهد، معنی دار نمی باشد. زیرا مقدارآماره t برای این پارامتر اضافی برابر-۰.۵۲ می باشد که کمتر از ۲ است. همچنین $p-value$ مربوطه برابر ۰.٦۰ است که بیشتر از ۰.۰۵ می باشد. در نتیجه فرضیه $H_0 : \alpha_3 = 0$ در سطح ۰.۰۵ رد نمی شود. بنابراین نیازی به لحاظ کردن پارامتر اضافی α_3 در مدل نمی باشد.

همچنین توجه داریم که برآورد پارامترهای مشترک AR۱ و AR۲ چندان دستخوش تغییر نشده اند.

در مدل اول داریم:

Type		Coef
AR	۱	۰.۷۱۸٦
AR	۲	-۰.٤۳٤۳

و در مدل دوم:

Type		Coef
AR	۱	۰.٦۹۷۱
AR	۲	-۰.۳۹۵۵

بنابراین با توجه به برازش جامع تر نیز می توان کفایت مدل انتخاب شده را پذیرفت.

ب) برازش مدل(۲،۱)ARMA

مدل منطقی دیگر برای برازش بیشتر (۲)AR یک مدل(۲,۱)ARMAاست. نتایج مربوط به برازش این مدل در قسمت برآورد پارامتر ها بصورت زیر است:

```
Final Estimates of Parameters

Type          Coef   SE Coef      T      P
AR   ١       ٠.٨٧٦٩    ٠.١٩٧٢    ٤.٤٥   ٠.٠٠٠
AR   ٢      -٠.٥١٠٥    ٠.١١٧٢   -٤.٣٦   ٠.٠٠٠
MA   ١       ٠.١٩٤١    ٠.٢٢٦٣    ٠.٨٦   ٠.٣٩٣
Constant   ١٨.١٥٥٠    ٠.٢٦٥١   ٦٨.٤٧   ٠.٠٠٠
               Mean     ٢٨.٦٥٣٢   ٠.٤١٨٥
```

همانطور که ملاحظه می شود مقدار آماره tو مقدار $p-value$ برای پارامتر میانگین متحرک معنی دار نمی باشد. بنابراین می توان گفت برآورد این پارامتر تفاوت معنی داری با صفر ندارد و نیازی به لحاظ کردن پارامتر اضافی MA۱ در مدل نمی باشد. همچنین برآورد پارامترهای مشترک AR۱ و AR۲ خیلی دستخوش تغییر نشده اند.

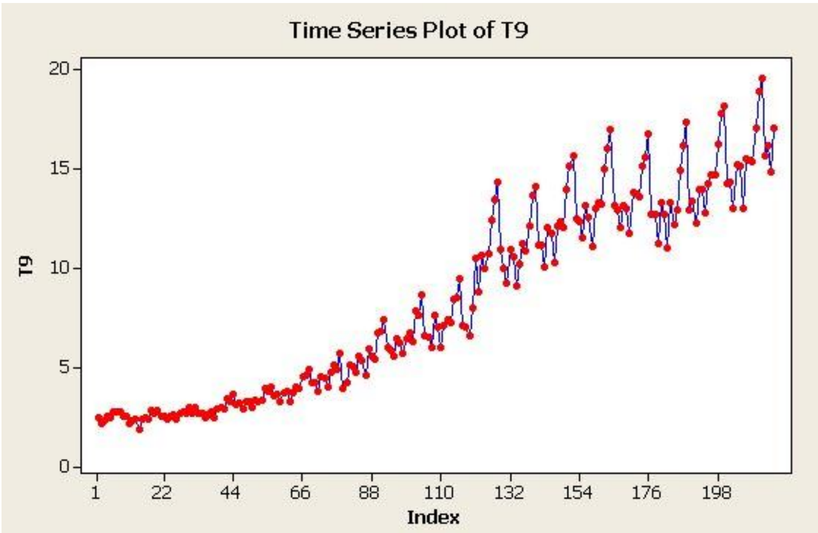
بنابراین شایستگی مدل شناسایی شده (۲)AR هم از طریق تجزیه و تحلیل باقیمانده ها و هم از طریق برازش مدلهای جامع تر مورد تأیید قرار می گیرد.

پیش بینی

اکنون می توان مدل شناسایی شده را مبنای پیش بینی رفتار آینده سری قرار داد. خواننده می تواند در صورت تمایل رفتار سری را برای یک افق زمانی دلخواه پیش بینی نماید.

مثال ۹

سری مربوط به کل مسافت پیموده شده مسافران هوایی ایالات متحده از ژانویه ۱۹۶۰ تا دسامبر ۱۹۷۰ را در نظر می گیریم(سری T۹). نمودار این سری در زیر رسم شده است.



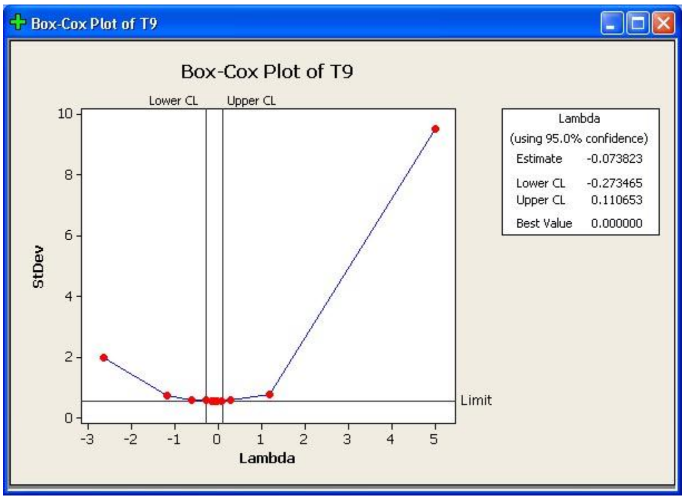
شکل ۵ — ۷۸ : نمودار سری زمانی کل مسافت پیموده شده توسط مسافران خطوط هوایی

تشخیص مدل آزمایشی

نمودار سری ناایستایی در میانگین و واریانس را نشان می دهد. در حقیقت انحراف معیار سری تقریبا متناسب با سطح آن است. بنابراین یک تبدیل لگاریتمی مناسب به نظر می رسد.

تبدیل باکس–کاکس نیز بر لزوم تبدیل لگاریتمی تأکید می کند. برای اجرای این رویه ابتدا از منوی Editor گزینه Enable commandsرا انتخاب می کنیم تا خط فرمان در منوی session فعال شود. سپس فرمان مربوط به اجرای رویه باکس–کاکس را تایپ

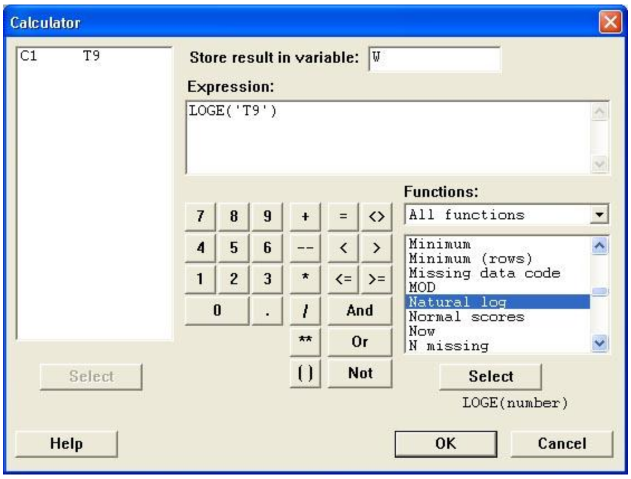
کرده و کلید اینتر را برای تأیید فرمان می فشاریم.نتیجه اجرای فرمان فوق نمودار(۵-۷۹) است. همانطور که ملاحظه می شود در کادر سمت راست بهترین مقدار پیشنهادی برای پارامتر تبدیل λ صفر است که یک تبدیل لگاریتمی را پیشنهاد می کند.



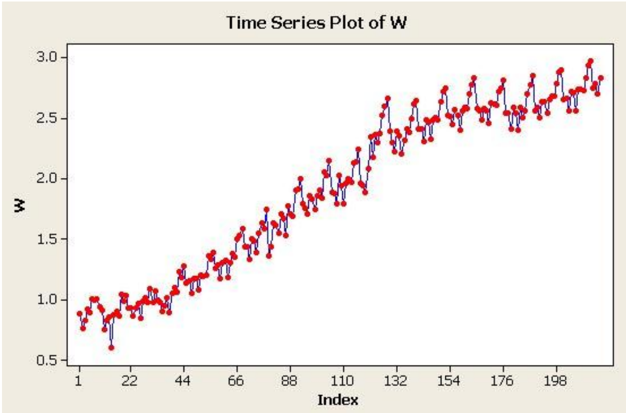
شکل ۵ – ۷۹ : نمودار تبدیلات باکس– کاکس برای سری T۹

تبدیل لگاریتمی

همانطور که قبلا هم توضیح داده شده است برای تبدیل لگاریتمی داده ها پنجره مربوطه را به شکل زیر تکمیل می کنیم.سری تبدیل شده را که در ستون C۲ ذخیره کرده ایم w_t می نامیم. سپس نمودار سری زمانی را برای سری تبدیل شده $w_t = \ln x_t$ رسم می کنیم.



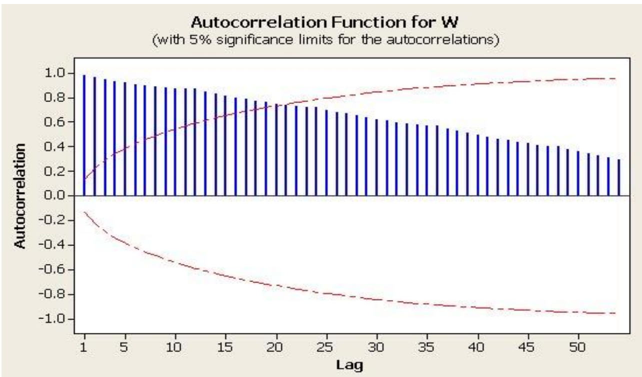
شکل ۵ – ۸۰ : تبدیل لگاریتمی سری T۹



شکل ۵ – ۸۱ : نمودار سری تبدیل شده w_t

در این نمودار ما هنوز فصلی بودن و یک روند افزایشی را مشاهده می کنیم. ولی اکنون تغییر در سطوح مختلف سری قابل مقایسه می باشد. از این پس تمام تجزیه و تحلیل بر مبنای لگاریتمهای سری اولیه خواهد بود.

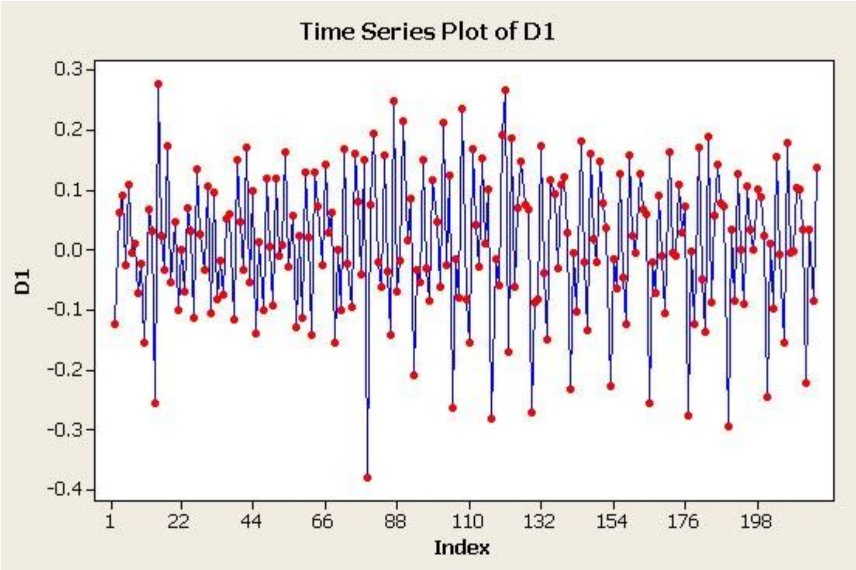
نمودار زیر تابع خود همبستگی نمونه ای را برای سری لگاریتمها می دهد. روند افزایشی در سری به اندازه ای شدید است که فصلی بودن در تابع خود همبستگی نمونه ای تقریبا محو شده است.



شکل ۵ – ۸۲ :acfسری تبدیل شده w_t

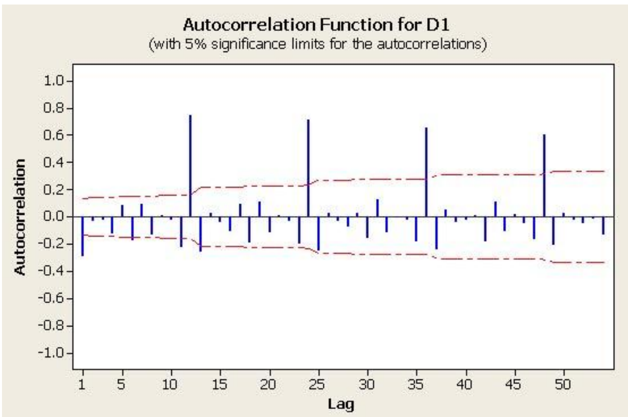
تفاضلی کردن

برای ایستا کردن سری $w_t = \ln x_t$ آن را یکبار تفاضلی می کنیم. سری تفاضلی شده را D۱ می نامیم. نمودار سری تفاضلی شده از مرتبه اول در زیر رسم شده است. در این نمودار فصلی بودن کاملا آشکار است.



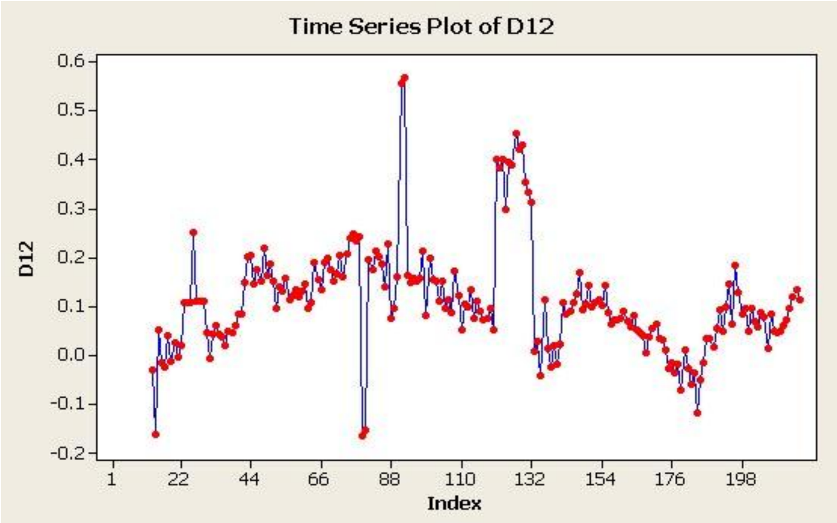
شکل ۵ — ۸۳ : نمودار سری زمانی تفاضلی شده

تابع خود همبستگی برای سری تفاضلی شده نیز فصلی بودن را بطور برجسته ای نشان می دهد. توجه کنید که در تأخیرهای ۱۲، ۲٤ و ۳٦ خود همبستگی مثبت شدیدی که به آرامی تنزل پیدا می کند، وجود دارد.

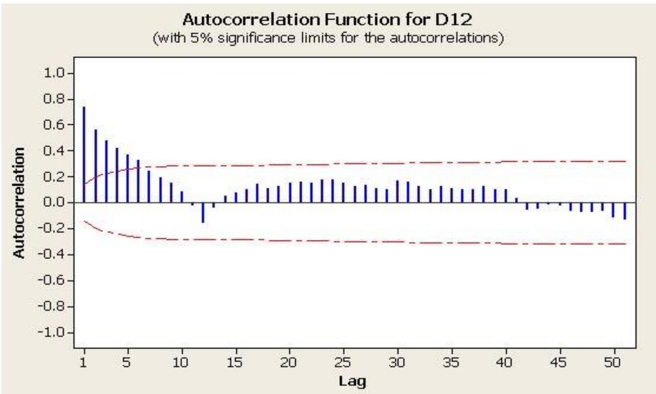


شکل ۵ — ۸٤ : acf سری تفاضلی شده

اگر بجای تفاضلی کردن غیر فصلی یک تفاضل فصلی با دوره ۱۲ را محاسبه کنیم، نمودار های مربوطه به شکل زیر خواهند بود. این سری را D۱۲نامیده ایم.



شکل ٥ – ٨٥ : نمودار سری تفاضلی شده از مرتبه ١٢



شکل ٥ – ٨٦ : acf سری تفاضلی شده از مرتبه ١٢

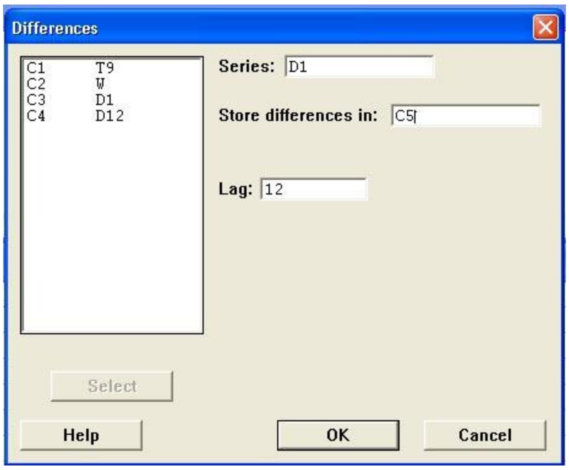
نمودار سری تفاضلی شده از مرتبه ١٢ نشان می دهد که پدیده فصلی تا حدودی حذف شده است ولی ایستایی سری را نمی توان فرض کرد. تابع خود همبستگی نیز فصلی بودن کم وناایستایی شدیدی را نشان می دهد.

اعمال تفاضلات فصلی و غیر فصلی

حال برای حذف روند و مؤلفه فصلی تفاضلات فصلی و غیر فصلی را بطور همزمان بر روی سری w_t اعمال می کنیم. می توان سری جدیدرا بصورت زیرنوشت:

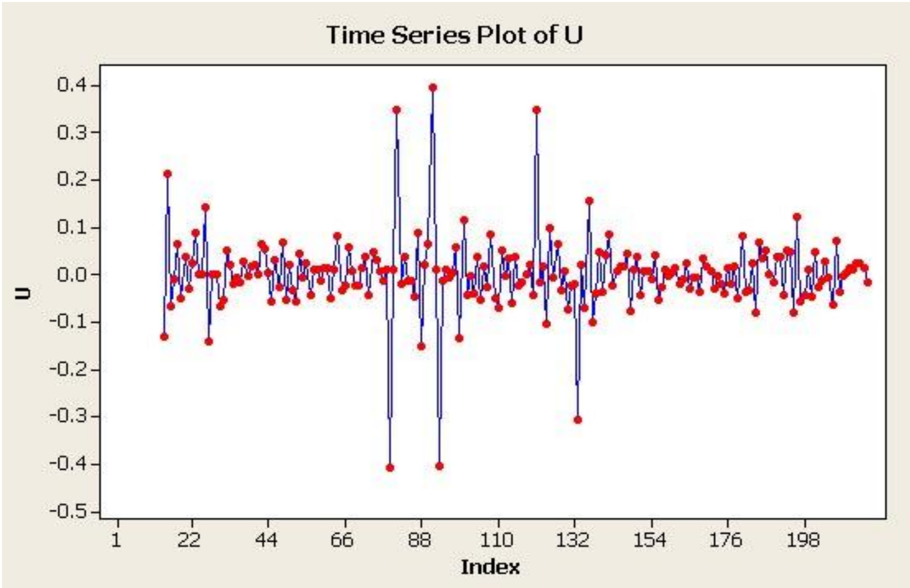
$$u_t = \nabla\nabla_{12} w_t = (1-B)(1-B^{12})w_t$$
$$= (1-B)(1-B^{12})\ln x_t$$

برای انجام این کار در مینی تب، سری تفاضلی شده از مرتبه اول را یکبار دیگر از مرتبه ۱۲ تفاضلی می کنیم. پنجره مربوطه به شکل زیر خواهد بود.



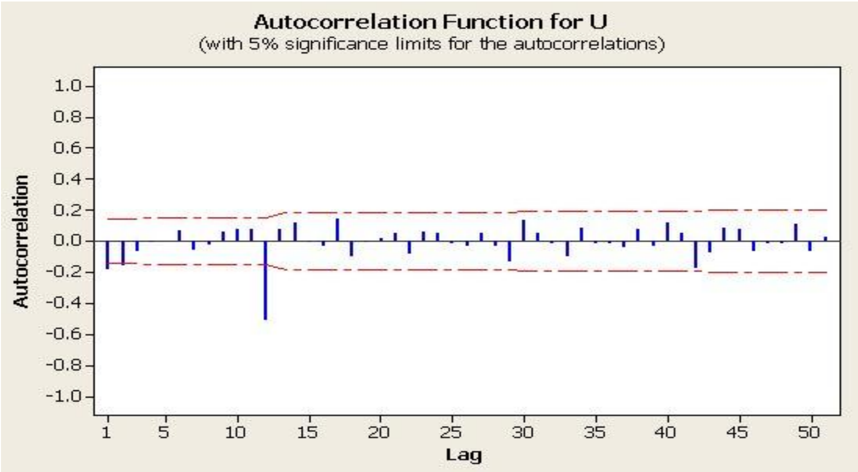
شکل ۵ – ۸۷ : تفاضلی کردن سری D1 از مرتبه ۱۲

نمودار سری جدید u_t بصورت زیر می باشد:

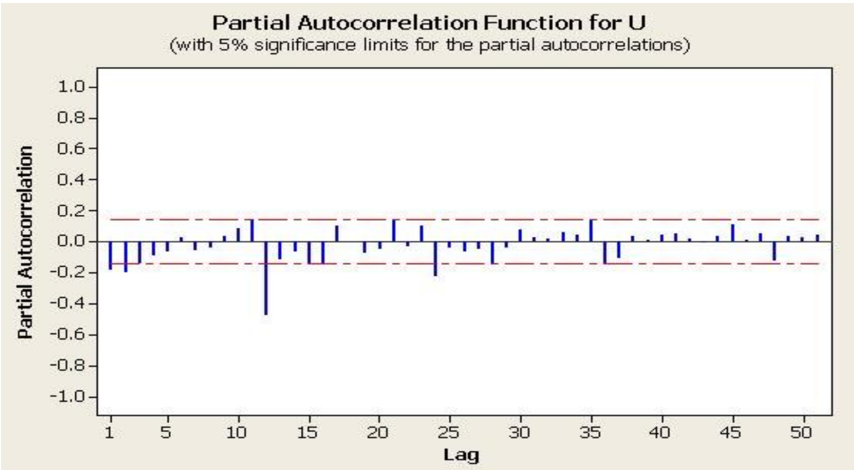


شکل ۵ – ۸۸ : نمودار سری زمانی u_t

نمودار acfو pacf سری u_t نیز در شکلهای (۵–۸۹) و (۵–۹۰) رسم شده است.هر دو نمودار پیشنهاد می کنند که بایستی یک مدل ایستا در نظر گرفته شود. توجه کنید که بعد از تفاضلی کردن همبستگی نسبتا زیادی در تأخیر ۱۲ باقی می ماند ولی در تأخیرهای۲٤ و۳٦ این همبستگی قابل اغماض می باشد. همچنین بخاطر امساک(صرفه جویی در تعداد پارامترها) می توان مقدار تابع خود همبستگی در تأخیر ۲ را نادیده گرفت.



شکل ۵ – ۸۹ : acf سری زمانی u_t



شکل ۵ – ۹۰ : pacf سری زمانی u_t

نظر به اینکه تابع خود همبستگی قطع شده است و تابع خود همبستگی جزئی آهسته تر کاهش می یابد، مدل داری شکل میانگین متحرک خواهد بود. تنها خود همبستگی های نمونه که از دو برابر خطاهای استانداردشان تجاوز کرده اند r_{12}, r_1

هستند(اگر چه r_2 تقریبا مساوی با حدود دو برابر خطای استاندارد است). بنابراین یک مدل آزمایشی بصورت زیر خواهد بود:

$$(1-B)(1-B^{12})w_t = (1+\beta B)(1+\gamma B^{12})z_t$$

توجه کنید که این رابطه یک مدل فصلی ضرب پذیر از درجه ۱۲(۰,۱,۱)(۰,۱,۱) می باشد.بنابراین ما مدل۱۲(۰,۱,۱)*(۰,۱,۱)ARIMA را برای بررسی بیشتر در نظر می گیریم. این مدل را به سری لگاریتمی w_t برازش داده و سپس مناسبت آن را بررسی می کنیم.

همانطور که ملاحظه شد ما در واقع برای تشخیص مدل مناسب,acf را برای سری های زیر بررسی کردیم:

الف) لگاریتم سری اصلی، x_t

ب) لگاریتم سری تفاضل گیری شده فقط نسبت به ما هها، ∇x_t

ج) لگاریتم سری تفاضل گیری شده فقط نسبت به سالها، $\nabla_{12} x_t$

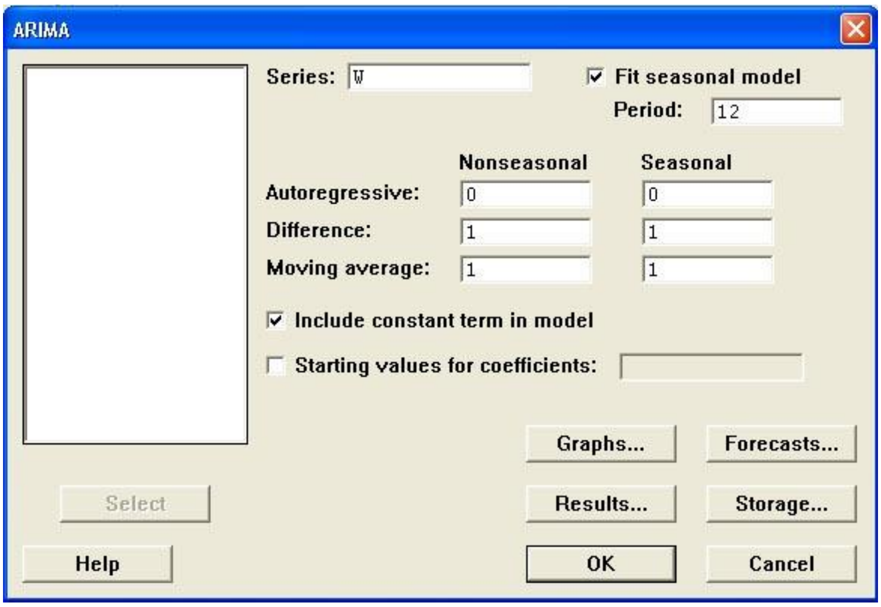
د) لگاریتم سری تفاضل گیری شده نسبت به ماهها و سالها، $\nabla\nabla_{12} x_t$

خود همبستگی ها برای لگاریتم سری اصلی بسیار بزرگ می باشند و در فواصل بزرگتر از میرائی باز می مانند. در صورتیکه تفاضل گیری ساده خود همبستگی ها را بطور کلی کاهش می دهد و فقط یک مؤلفه تناوبی قوی باقی می ماند. این مطلب بخصوص برای همبستگی های بزرگ در فواصل۳۶ ،۱۲،۲٤ و ٤۸ دیده می شود.

تفاضل گیری ساده نسبت به دوره تناوب ۱۲، همبستگی هایی می دهد که در ابتدا دائما مثبت و سپس دائما منفی هستند. بلعکس تفاضل گیری $\nabla\nabla_{12}$ تمام همبستگی ها را بطور قابل توجهی کاهش می دهد.

برازش مدل$_{١٢}$(٠،١،١) *(٠،١،١)ARIMA

پس از تشخیص مدل آزمایشی برای سری زمانی، مبادرت به برآورد پارامترهای مدل می نماییم. همانطور که می دانید مدلARIMA ی فصلی ضربی حالت خاصی از مدلARIMA ی کلی است. بنابراین برای برازش مدلARIMA ی فصلی ضربی پنجره اصلی ARIMA را بصورت زیر تکمیل می کنیم:



شکل ٥ – ٩١ : تکمیل پنجره ARIMA برای برازش مدل $_{١٢}$(٠,١,١) *(٠,١,١)ARIMA

خروجی قسمت مربوط به برآورد پارامترها در پنجره session بصورت زیر است:

ARIMA Model: W

Final Estimates of Parameters

Type	Coef	SE Coef	T	P
MA ١	٠.٢٨٠٦	٠.٠٦٨٤	٤.١٠	٠.٠٠٠
SMA ١٢	٠.٧٤٥٣	٠.٠٥١٠	١٤.٦١	٠.٠٠٠
Constant	٠.٠٠٠٠٩٢٩	٠.٠٠٠٨٧٩٠	٠.١١	٠.٩١٦

با توجه به مقدارآماره t و $p-value$ مربوطه می توان فرض صفر بودن جمله ثابت(Constant) را پذیرفت. همانطور که ملاحظه می شود $p-value=$۰.۹۱۶ که بزرگتر از ۰.۰۵ می باشد، بنابراین با ۰.۹۵درصد اطمینان می توان فرضیه $H_0 : \theta_0 = 0$ را پذیرفت.

همچنین توجه کنید که با توجه به مقدار آماره tو $p-value$ مربوط به پارامترهای میانگین متحرک غیر فصلی(MA) و فصلی(SMA)، می توان گفت که این پارامترها معنی دار هستند و فرض صفر بودن آنها به شدت رد می شود.

در مرحله بعد با برداشتن چک مارک عبارت Include constant term inmodelدر پنجره اصلی ARIMA یک مدل بدون جمله ثابت به سری w_t برازش می دهیم.

برازش مدل$_{۱۲}$(۰،۱،۱) *(۰،۱،۱)ARIMA بدون جمله ثابت

خروجی برازش مدل مورد نظر بدون جمله ثابت در پنجره session بصورت زیر می باشد.

ARIMA Model: W

```
Estimates at each iteration

Iteration       SSE    Parameters
0           1.15980   0.100   0.100
1           1.01506   0.132   0.250
2           0.89961   0.164   0.400
3           0.81675   0.194   0.550
```

٤	٠.٧٦٣٠٥	٠.٢٣٣	٠.٧٠٠
٥	٠.٧٥٤٥٣	٠.٢٥٩	٠.٧٣٩
٦	٠.٧٥٤٥١	٠.٢٧٥	٠.٧٣٢

Unable to reduce sum of squares any further

Final Estimates of Parameters

Type		Coef	SE Coef	T	P
MA	١	٠.٢٧٥٣	٠.٠٦٨٤	٤.٠٣	٠.٠٠٠
SMA	١٢	٠.٧٣١٧	٠.٠٥٢٦	١٣.٩٢	٠.٠٠٠

Differencing: ١ regular, ١ seasonal of order ١٢

Number of observations: Original series ٢١٦, after differencing ٢٠٣

Residuals: SS = ٠.٧٣٧٧٦٧ (backforecasts excluded)

MS = ٠.٠٠٣٦٧٠ DF = ٢٠١

Modified Box-Pierce (Ljung-Box) Chi-Square statistic

Lag	١٢	٢٤	٣٦	٤٨
Chi-Square	١٩.٢	٢٥.٥	٣١.٧	٥٠.٨
DF	١٠	٢٢	٣٤	٤٦
P-Value	٠.٠٣٨	٠.٢٧٥	٠.٥٨١	٠.٢٩١

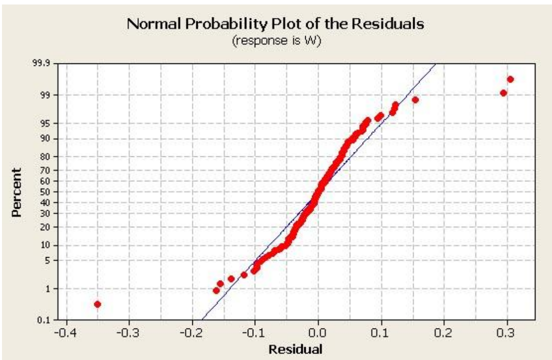
همانطور که در قسمت مربوط به برآورد پارامترها مشاهده می شود، مقدار برآورد شده برای پارامترهای میانگین متحرک غیر فصلی و فصلی به ترتیب عبارت است از ۰.۲۷۵ و ۰.۷۳۱ و هر دو پارامتر اختلاف معنی داری با صفر دارند. بنابراین مدل آزمایشی بصورت زیر می باشد:

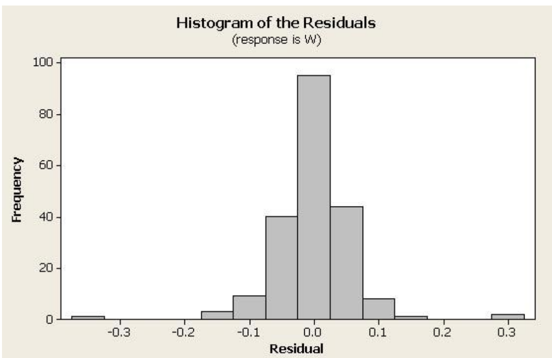
$$(1-B)(1-B^{12})w_t = (1-0.27B)(1-0.73B^{12})z_t$$

بررسی مناسبت مدل

۱- تجزیه و تحلیل باقیمانده ها

الف) بررسی فرض نرمال بودن باقیمانده ها

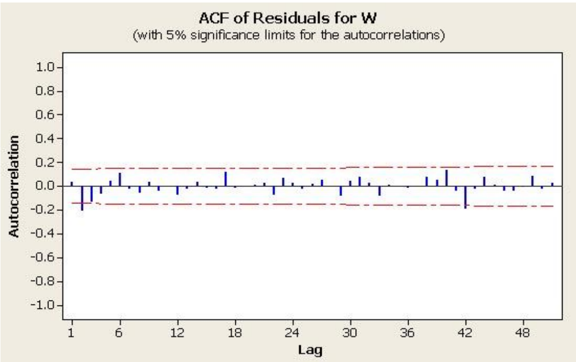




شکل ۵ – ۹۲ : نمودار احتمال نرمال و هیستوگرام باقیمانده های مدل ARIMA(۰,۱,۱)* (۰,۱,۱)$_{۱۲}$

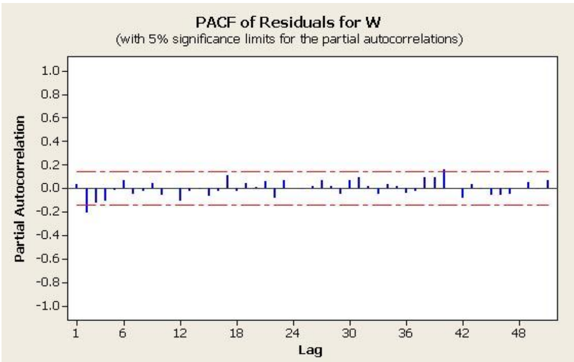
با توجه به نمودار احتمال نرمال و هیستوگرام باقیمانده ها تقریبا می توان فرض نرمال بودن باقیمانده ها را پذیرفت.

ب) بررسی فرض استقلال باقیمانده ها



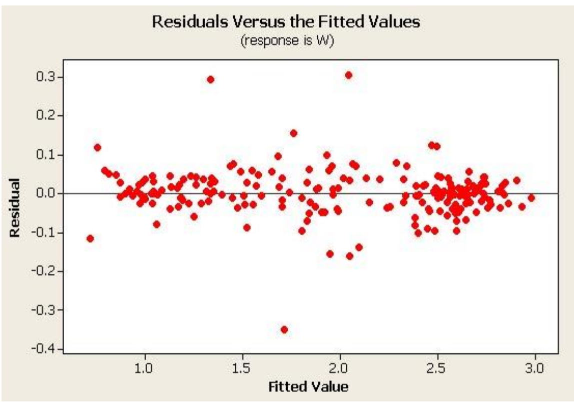
شکل ۵ – ۹۳ : acf و pacf باقیمانده های

مدل ARIMA(۰,۱,۱)* $(۰,۱,۱)_{۱۲}$



مقادیر تابع خود همبستگی بجز در تأخیر های ۲ و ٤٢ از حدود استاندارد تجاوز نکرده اند و تقریبا می توان فرض تصادفی بودن باقیمانده ها را پذیرفت.

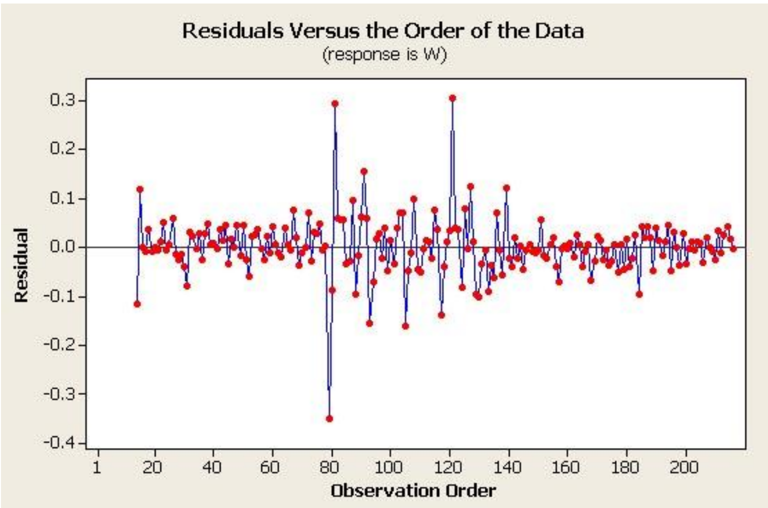
ج) بررسی فرض ثابت بودن واریانس باقیمانده ها



شکل ۵ – ۹۴ : نمودار باقیمانده های مدل ARIMA(۰,۱,۱)*(۰,۱,۱)۱۲ در برابر مقادیر برازش شده

این نمودار ساختار خاصی را نشان نمی دهد. بنابراین می توان فرض ثابت بودن واریانس باقیمانده ها را پذیرفت.

د) رسم نمودار مانده ها در برابر زمان



شکل ۵ – ۹۵ : نمودار باقیمانده های مدل ARIMA(۰,۱,۱)*(۰,۱,۱)۱۲ در برابر زمان

این نمودار نیز طرح خاصی را نشان نمی دهد. بنابراین می توان مناسبت مدل برازش داده شده را پذیرفت.

ر) آزمون پرت-مانتو

اکنون به عنوان یک روش رسمی تر به بررسی مناسبت مدل با استفاده از آماره Q اصلاح شده می پردازیم.

با توجه به قسمت آخر خروجی مینی تب در پنجره session به بررسی مناسبت مدل با استفاده از این آماره می پردازیم.

Modified Box-Pierce (Ljung-Box) Chi-Square statistic

Lag	١٢	٢٤	٣٦	٤٨
Chi-Square	١٩.٢	٢٥.٥	٣١.٧	٥٠.٨
DF	١٠	٢٢	٣٤	٤٦
P-Value	٠.٠٣٨	٠.٢٧٥	٠.٥٨١	٠.٢٩١

توجه: برای مدلهای فصلی باید آماره اصلاح شده باکس-پیرسن با مقادیر بحرانی توزیع کی دو با $k-m$درجه آزادی مقایسه شود که $m=p+q+P+Q$ تعداد پارامترهای برآورد شده است.

در مدلهای فصلی باید ماکزیمم k تا اندازه ای بزرگتر انتخاب شود. معمولا برای داده های ماهانه ٣٦= kیا ٤٨= kانتخاب خوبی است.

در اینجا مقدار آماره Q برای تأخیر ٣٦مساوی ٣١.٧ می باشد. با توجه به مقدار٠.٥٨= $p-value$ که بیشتر از ٠.٠٥ می باشدمی توان با ٠.٩٥اطمینان فرضیه صفر را پذیرفت.

همچنین می توان مقدار آماره Qرا با مقدار جدول کی دو با ٣٤ درجه آزادیمقایسه کرد ($\chi^2_{(0.05,34)}=48.6$). بنابراین با توجه به آنکه مقدار آماره Q اصلاح شده از مقدار متناظر جدول کی دو در سطح ٠.٠٥ با ٣٤ درجه آزادی تجاوز نکرده است، نمی توان فرضیه صفر را رد کرد و با ٠.٩٥ اطمینان آن را می پذیریم.

همانطور که قبلاً هم گفته شد فرضیه صفر که آن را فرضیه کفایت مدل نیز می نامیم بصورت زیر می باشد:

$$H_0 : \rho_1 = \rho_2 = ... = \rho_k = 0$$

پذیرش این فرض به معنی ناهمبسته بودن همه خود همبستگی های مربوط به باقیمانده ها تا تأخیر k می باشد و این خود دلیلی بر شایستگی مدل است.

۲- بررسی مناسبت مدل با روش برازش جامع تر (overfitting)

همانطور که در ابتدای این فصل گفته شد یکی از تکنیکهای بررسی مناسبت مدل، برازش جامع تر می باشد. در این روش ما یک مدل کلی تر را به داده ها برازش می دهیم و سپس به بررسی مناسبت آن می پردازیم. مدل اصلی مورد تأیید قرار خواهد گرفت اگر برآورد پارامترهای اضافی تفاوت معنی داری با صفر نداشته باشد و همچنین برآورد پارامترهای مشترک با برآورد پارامترهای اولیه اختلاف معنی داری نداشته باشد.

برازش مدل $ARIMA(۰،۱،۲)*(۰،۱،۱)_{۱۲}$

در مورد این مثال ما مدل کلی تر $ARIMA(۰,۱,۲)(۰,۱,۱)_{۱۲}$ را به داده ها برازش می دهیم. خروجی در پنجره session بصورت زیر می باشد.

ARIMA Model: W

```
Estimates at each iteration

Iteration       SSE        Parameters
     ۰     ۱.۱۲۶۵۰   ۰.۱۰۰   ۰.۱۰۰   ۰.۱۰۰
     ۱     ۰.۹۷۲۱۸   ۰.۱۴۲   ۰.۱۳۲   ۰.۲۵۰
```

٢	٠.٨٥٧٥٧	٠.١٧٨	٠.١٦٣	٠.٤٠٠
٣	٠.٧٧٣٥٨	٠.٢١٠	٠.١٩٤	٠.٥٥٠
٤	٠.٧٤٤٩٩	٠.٢٥٨	٠.٢٣٣	٠.٦٠٨
٥	٠.٧٢٧٤٣	٠.٢٧٤	٠.٢٤٤	٠.٧٥٨
٦	٠.٧١٤٠٠	٠.٢٨٤	٠.٢٣٢	٠.٧٢٠
٧	٠.٧١١١٣	٠.٢٨٤	٠.٢٣٤	٠.٧٣٧

Unable to reduce sum of squares any further

Final Estimates of Parameters

Type		Coef	SE Coef	T	P
MA	١	٠.٢٨٤٠	٠.٠٦٨٨	٤.١٣	٠.٠٠٠
MA	٢	٠.٢٣٤٣	٠.٠٦٩٠	٣.٣٩	٠.٠٠١
SMA	١٢	٠.٧٣٧١	٠.٠٥٣٣	١٣.٨٣	٠.٠٠٠

Differencing: ١ regular, ١ seasonal of order ١٢

Number of observations: Original series ٢١٦, after differencing ٢٠٣

Residuals: SS = ٠.٦٩٢٨٨٧ (backforecasts excluded)

MS = ٠.٠٠٣٤٦٤ DF = ٢٠٠

Modified Box-Pierce (Ljung-Box) Chi-Square statistic

Lag	١٢	٢٤	٣٦	٤٨

Chi-Square	٦.٦	١٣.٨	١٨.٤	٣٢.٦
DF	٩	٢١	٣٣	٤٥
P-Value	٠.٦٧٩	٠.٨٧٧	٠.٩٨١	٠.٩١٥

همانطور که ملاحظه می شود پارامتر اضافی MA٢ معنی دار می باشد و برآورد پارامترهای مشترک چندان تغییر نکرده است. بعلاوه مجموع مربعات خطا در مدل دوم کمتر شده است(SSR_1=٠.٧٣٧ و SSR_2=٠.٦٩٢).

همچنین مقدار $p-value$ برای آماره اصلاح شده باکس-پیرسن در تمامی تأخیرها افزایش یافته است. بنابراین مدل دوم فرضیه ناهمبسته بودن باقیمانده ها را با قوت بیشتری تأیید می کند. خروجی آزمون پرت-مانتو برای مدل اول و دوم در زیر آمده است:

خروجی آزمون پرت-مانتو در مدل اول:

Modified Box-Pierce (Ljung-Box) Chi-Square statistic

Lag	١٢	٢٤	٣٦	٤٨
Chi-Square	١٩.٢	٢٥.٥	٣١.٧	٥٠.٨
DF	١٠	٢٢	٣٤	٤٦
P-Value	٠.٠٣٨	٠.٢٧٥	٠.٥٨١	٠.٢٩١

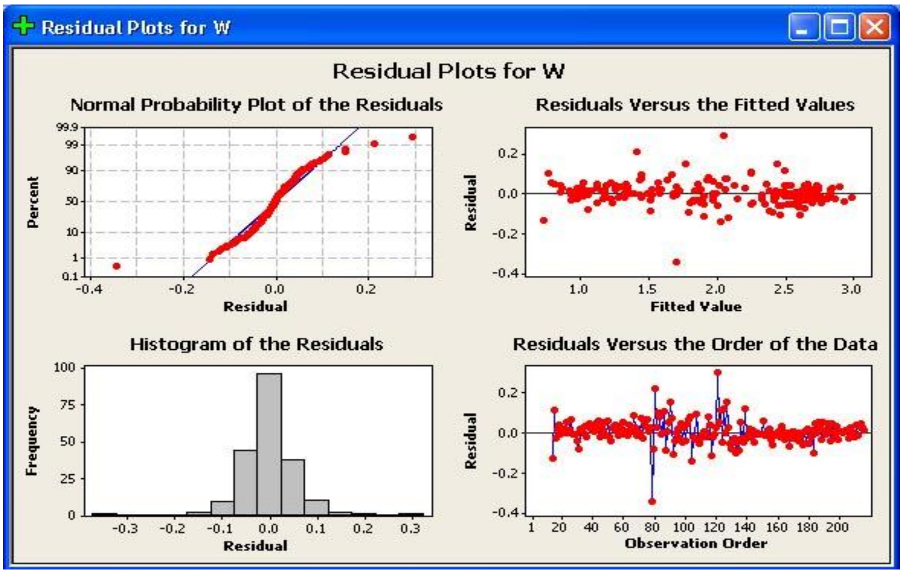
خروجی آزمون پرت-مانتو در مدل دوم:

Modified Box-Pierce (Ljung-Box) Chi-Square statistic

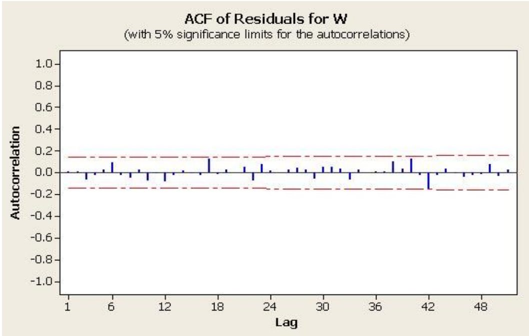
Lag	١٢	٢٤	٣٦	٤٨
Chi-Square	٦.٦	١٣.٨	١٨.٤	٣٢.٦
DF	٩	٢١	٣٣	٤٥
P-Value	٠.٦٧٩	٠.٨٧٧	٠.٩٨١	٠.٩١٥

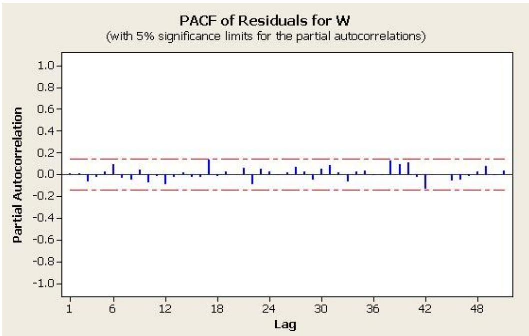
بررسی مناسبت مدل ARIMA(٠،١،٢)(٠،١،١)$_{١٢}$

برای بررسی مناسبت مدل نمودارهای مربوطه را بررسی کرده ایم که همگی کفایت مدل برازش شده را مورد تایید قرار داده اند.



شکل ٥ — ٩٦ : نمودار باقیمانده های مدل ARIMA(٠,١,٢)(٠,١,١)$_{١٢}$





شکل ۵ – ۹۷ : acf و pacf باقیمانده های مدل ARIMA(۰,۱,۲)(۰,۱,۱)$_{۱۲}$

انتخاب مدل بر مبنای معیار آکائیک

اکنون با محاسبه معیار آکائیک برای هر یک از دو مدل فوق که آنها را M۱ و M۲می نامیم، خواهیم داشت:

$$AIC(M1) = 216\log(0.003670) + 2(2) = -522.032$$
$$AIC(M2) = 216\log(0.003464) + 2(3) = -525.045$$

همانطور که ملاحظه می شود معیار آکائیک برای مدل دوم کوچکتر است. بنابراین ما مدل دوم یعنی مدل ARIMA(۰,۱,۲)*(۰,۱,۱)$_{۱۲}$را به عنوان یک مدل مناسب انتخاب می کنیم.

ماهمچنین مدلهای ARIMA از مرتبه های $(0,1,2)*(1,1,1)_{12}$ و $(1,1,2)*(0,1,1)_{12}$، $(0,1,2)*(0,1,2)_{12}$، $(0,1,3)*(0,1,1)_{12}$ را برازش داده ایم اما پارامتر اضافی در هیچ یک از مدلهای فوق معنی دار نبوده است.

یادآوری

بطور کلی وقتی در یک سری زمانی بعد ازS فاصله زمانی پایه، تشابهاتی رخ دهد می گوئیم سری یک رفتار تناوبی با دوره تناوب S را نشان می دهد. در مثال بالا فاصله زمانی پایه یک ماه و دوره تناوب ۱۲= S ماه است. ولی مثالهایی نیز وجود دارند که S می تواند مقادیر دیگری به خود بگیرد. برای مثال S=٤ برای داده های سه ماهه که اثرات فصلی داخل سال را نشان می دهد.

گاهی اتفاق می افتد که بیش از یک تناوب وجود دارد. به عنوان مثال چون تمایل به پرداخت صورت حسابها بصورت ماهیانه است، انتظار داریم که معاملات انجام شده در یک هفته بوسیله بانک تناوبی در حدود ٤ در داخل ماه ها نشان دهد در حالیکه معاملات انجام شده در یک ماه تناوب ۱۲ را نشان می دهد.

مثال ۱۰

۱۵۰ داده از یک مدل$_4$(۰,۱,۱)*(۰,۱,۱)ARIMA با $\beta = 0.8$ و $\gamma = 0.6$ شبیه سازی شده است. این داده ها در پیوست با نام T۱۰ لیست شده اند. همانطور که می دانید شکل این مدل بصورت زیر می باشد:

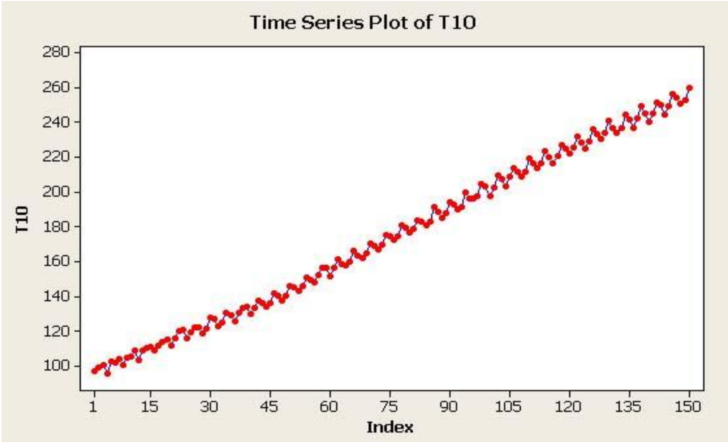
$$(1-B)(1-B^4)x_t = (1+\beta B)(1+\gamma B^4)z_t$$

که z_t یک متغیر تصادفی نرمال با میانگین صفر و واریانس یک می باشد. بنابراین خواهیم داشت:

$$(1-B)(1-B^4)x_t = (1+0.8B)(1+0.6B^4)z_t$$

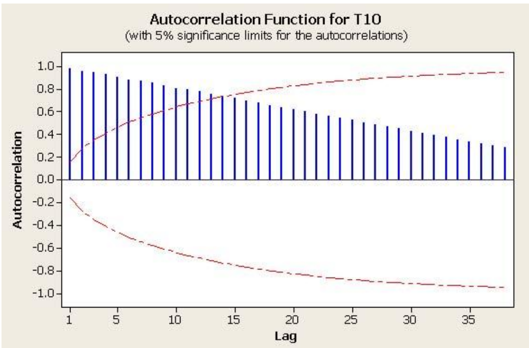
در واقع ما در این مثال می خواهیم بدانیم که آیا با بررسی داده ها ما به همان مدلی که خودمان داده ها را از آن تولید کرده ایم می رسیم یا نه. یعنی مدل احتمالی مولد داده ها برای ما از قبل شناخته شده است. بنابراین با بررسی داده ها ما باید به همان مدل شناخته شده برسیم.

نمودار این سری نشان دهنده یک رفتار فصلی و یک روند روبه بالا می باشد. همانطور که ملاحظه می شود این سری ناایستا است. برای بررسی بیشتر acf و pacf این سری را رسم می کنیم.

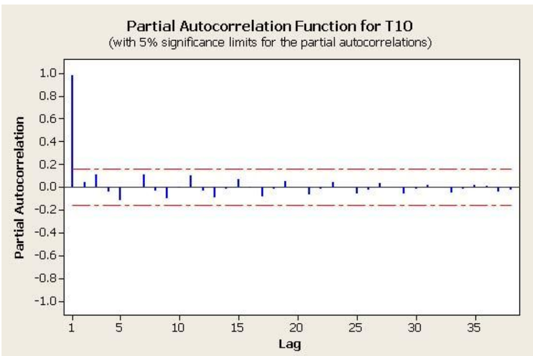


شکل ٥ – ٩٨ : نمودار سری زمانی شبیه سازی شده از مدل ARIMA(٠,١,١)*(٠,١,١)٤

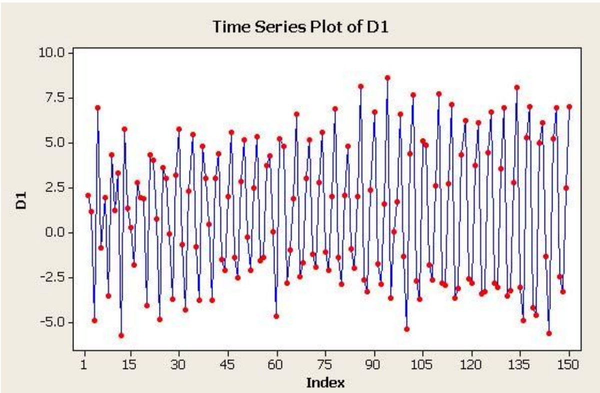
.



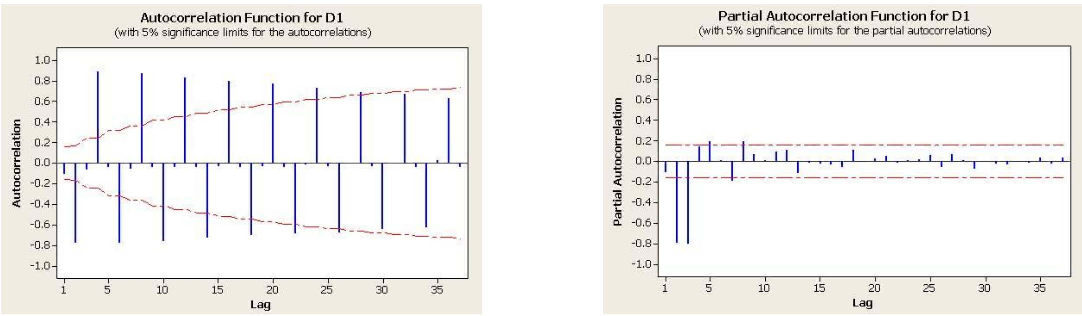
شکل ٥ – ٩٩ : acf و pacf سری زمانی T١٠



همانطور که ملاحظه می شود acf بزرگ بوده و بسیار کند تنزل می کند. درصورتی که pacf فقط یک مقدار بزرگ در تأخیر ۱ دارد. از این رو متوجه می شویم که سری ناایستا است و نیاز به تفاضلی کردن دارد. برای حذف ناایستایی سری را تفاضلی نموده، acf و pacf سری تفاضلی شده را رسم می کنیم.



شکل ۵ – ۱۰۰ : نمودار سری تفاضلی شده

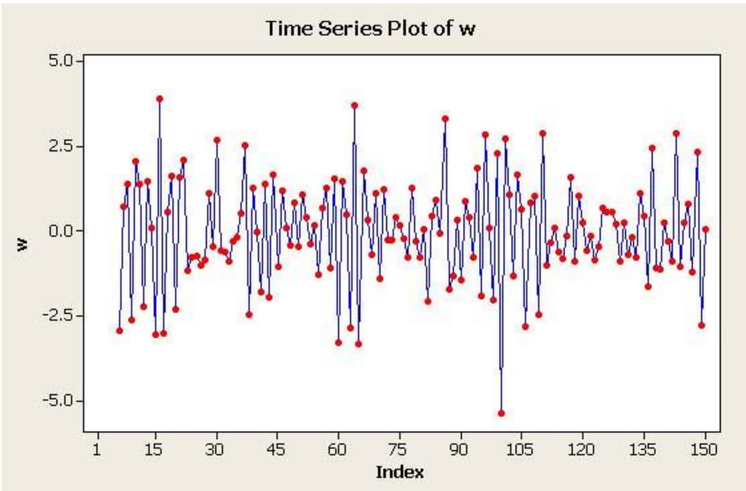


شکل ۵– ۱۰۱ :acf و pacf سری تفاضلی شده

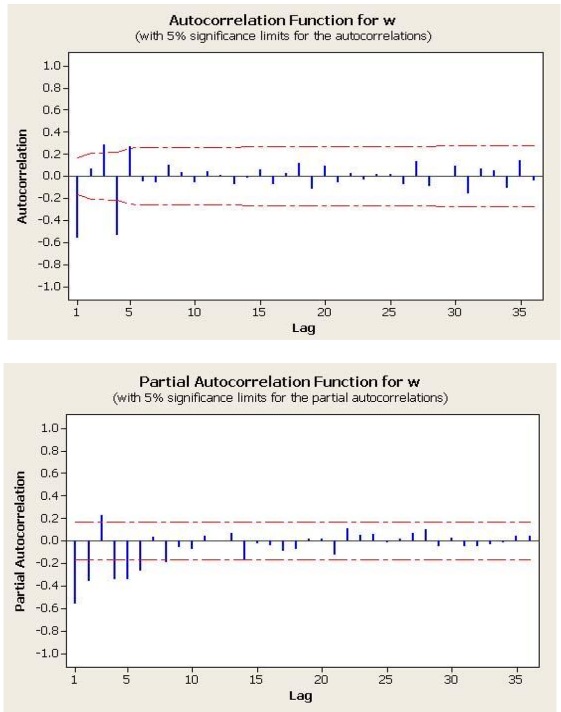
از اینکه acf در مضربهای دوره فصلی ٤ بسیار کند تنزل می کند، نتیجه می گیریم که یک پدیده فصلی با دوره ٤ در داده ها وجود دارد. بنابراین برای حذف این اثر فصلی، تفاضل فصلی $(1-B^4)$ به داده ها اعمال می کنیم.

بنابراین سری $(1-B)x_t$ را یک بار دیگر و از مرتبه ٤ تفاضلی می کنیم. اکنون نمودارهای مربوط به سری جدید $w_t = (1-B)(1-B^4)x_t$ را رسم می کنیم.

همانطور که می دانید برای انجام این کار در مینی تب، با استفاده از پنجره Differences ابتدا سری T۱۰ را از مرتبه یک تفاضلی نموده و سپس سری تفاضلی شده را یک بار دیگر از مرتبه ٤ تفاضلی می کنیم.



شکل ۵ – ۱۰۲ : نمودار سری زمانی w_t



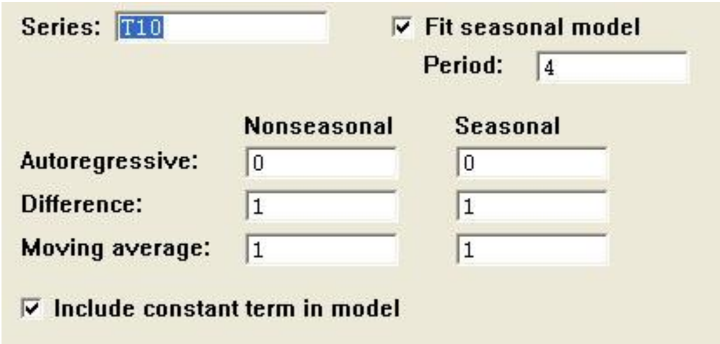
شکل ۵ – ۱۰۳ : acf و pacf سری w_t

می دانیم که acf مدل ARIMA(۰,۰,۱)*(۰,۰,۱)$_4$ بجز در تأخیرهای ۱,۳,٤,۵ برابر صفر است. بنابراین acf نمونه ای w_t یک مدل ARIMA(۰,۱,۱)*(۰,۱,۱)$_4$ را برای سری اولیه x_t می دهد. یعنی

$$(1-B)(1-B^4)x_t = (1+\beta B)(1+\gamma B^4)z_t$$

برازش مدل ARIMA(۰،۱،۱)*(۰،۱،۱)$_{۴}$

برای برازش چنین مدلی در مینی تب، پنجرهARIMA را به شکل زیر تکمیل می کنیم:



شکل ۵ – ۱۰٤ : تکمیل پنجره آریما برای برازش مدل ARIMA(۰,۱,۱)*(۰,۱,۱)$_{۴}$

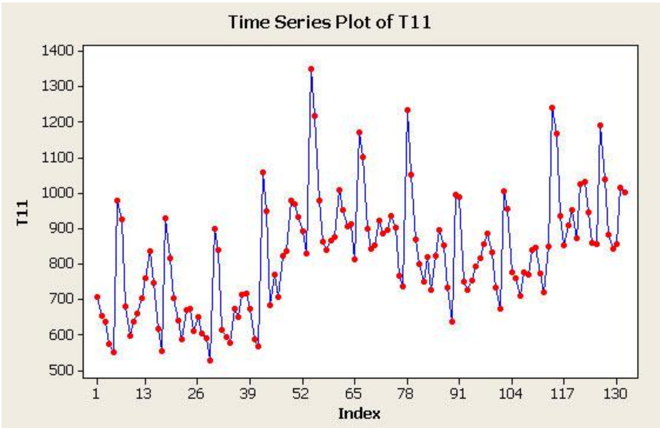
بررسی ها نشان می دهد که جمله ثابت معنی دار نیست. همچنین نمودارهای مربوط به باقیمانده ها نشان می دهد که این مدل برازش بسیار خوبی به داده ها دارد.

برآورد پارامترهای میانگین متحرک غیر فصلی و فصلی حاصل از برازش مدل فوق(بدون جمله ثابت) عبارت است از $\beta = 0.758$ و $\gamma = 0.643$، که بسیار نزدیک به پارامترهای مدل مولد داده ها است.

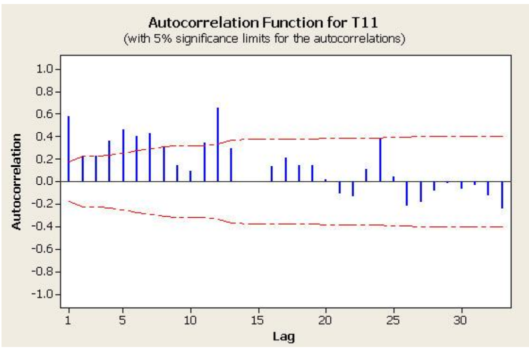
مثال ۱۱

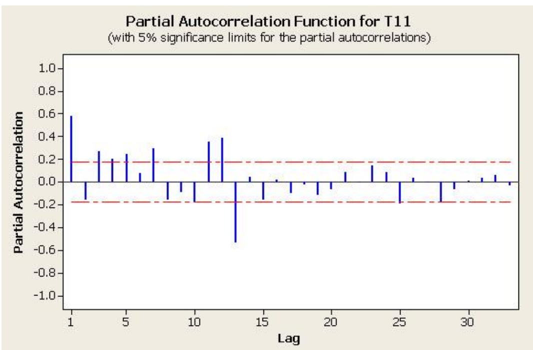
ارقام ماهانه استخدام مردان ١٦ تا ١٩ سال(به هزار نفر) از سال ۱۹۷۱ تا ۱۹۸۱ که در پیوست با نام T۱۱ لیست شده اند را در نظر می گیریم . می خواهیم رفتار سری را با یک مدل مناسب ARIMA مدل سازی کنیم.

تشخیص مدل آزمایشی



شکل ۵ – ۱۰۵ : نمودار سری زمانی ارقام ماهانه استخدام مردان ١٦ تا ١٩ سال از ۱۹۷۱ تا ۱۹۸۱





شکل ۵ – ١٠٦ : acf و pacf سری زمانی T۱۱

با توجه به نمودار سری یک روند تصادفی در داده ها مشاهده می شود. بنابراین قبل از تشخیص مدل باید به وسیله تفاضلی کردن این روند حذف شود. acf این سری نیز به کندی به سمت صفر میل می کند که دال بر ناایستایی سری مربوطه می باشد. طبیعت فصلی سری نیز آشکار است. ارقام استخدام در ماه های تابستان افزایش می یابد و در ماه ژوئن وقتی مدارس دایر نیست به اوج خود می رسد. سپس در پاییز وقتی مدارس دوباره باز می شود، کاهش پیدا می کند. این پدیده هر ۱۲ ماه یک بار خودش را تکرار می کند. بنابراین طول دوره فصلی ۱۲ می باشد.

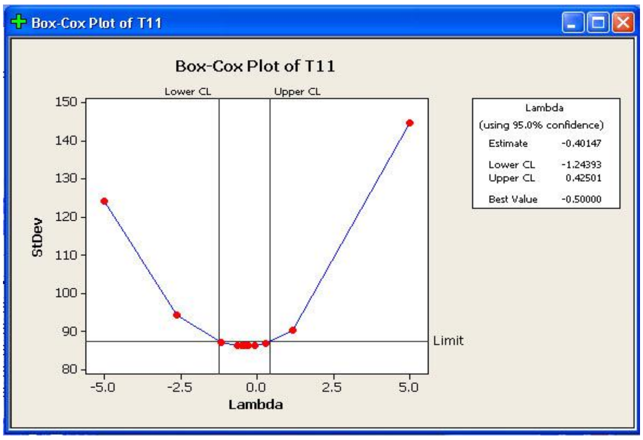
همچنین می توان با نگه داشتن نشانگر موس بر روی نقاط اوج در نمودار سری, علاوه بر مشاهده مقدار سری در این نقاط ملاحظه نمود که این نقطه اوج چندمین مشاهده سری زمانی می باشد. با تکرار این کار در چند نقطه اوج دیگر، متوجه می شویم که این نقاط به فاصله ۱۲ ماه از یکدیگر قرار گرفته اند. به عبارتی هر ۱۲ ماه شاهد یک نقطه اوج هستیم. بنابراین می توان گفت که این داده ها تحت تأثیر یک پدیده فصلی با دوره ۱۲ می باشند.

قبل از هر گونه اقدامی برای ایستایی میانگین سری، ابتدا باید از ایستا بودن واریانس سری اطمینان حاصل کرد.

بررسی پایایی واریانس

برای این منظور در پنجره session بعد از فعال کردن خط فرمان عبارت زیر را تایپ کرده و با فشردن اینتر آن را تأیید می کنیم.

```
MTB > boxcox T۱۱ ۱.
```



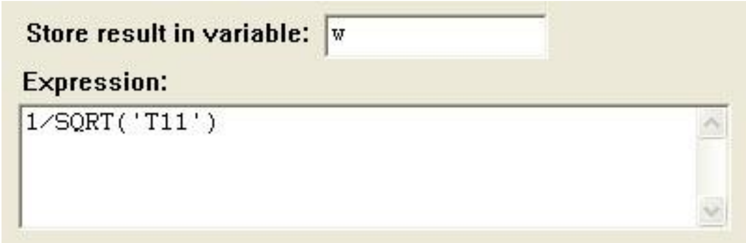
شکل ۵ – ۱۰۷ : نمودار باکس-کاکس برای سری زمانی T۱۱

همانطور که ملاحظه می شود بهترین مقدار پیشنهادی برای پارامتر تبدیل ۰.۵- می باشد. با توجه به جدول(۲-۱)تبدیل مناسب به ازای ۰.۵- = λ عبارت است از $\frac{1}{\sqrt{x_t}}$. بنابراین واریانس داده ها پایدار نیست و نیاز به تبدیل دارد.

اعمال تبدیل $\frac{1}{\sqrt{x_t}}$

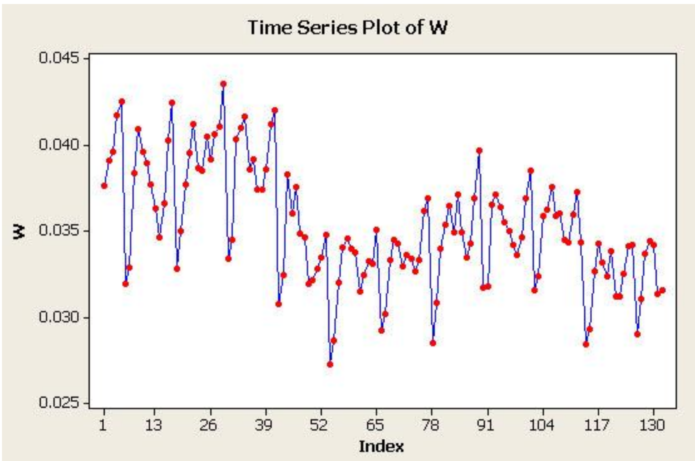
برای انجام این کار از منوی Calc گزینه Calculator را انتخاب کرده و سپس در کادر Expression عبارت ۱/sqrt(T۱۱) را تایپ می کنیم. داده های تبدیل شده را با نام w ذخیره می کنیم.

شکل زیر نیز چگونگی تکمیل کردن این پنجره را نشان می دهد.



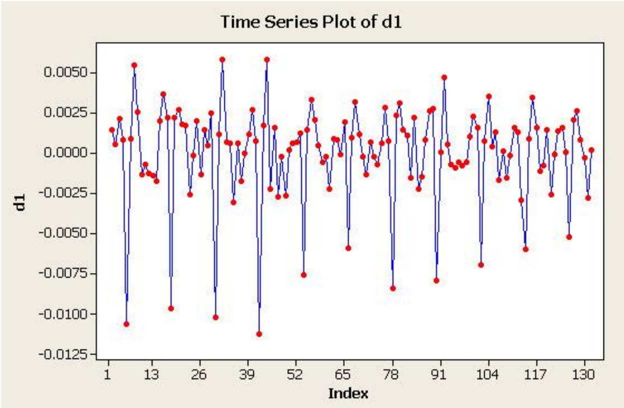
شکل ۵ – ۱۰۸ : تبدیل عکس ریشه دوم

برای اطمینان از تأثیر تبدیل اعمال شده در پایا سازی واریانس، رویه باکس–کاکس را یک بار دیگر و برای سری جدید w_t انجام داده ایم. در نتیجه اجرای این رویه بهترین مقدار پیشنهادی برای پارامتر تبدیل یک بدست آمده است که دلالت بر پایا بودن واریانس داده های تبدیل شده دارد. نمودار سری زمانی برای داده های تبدیل شده بصورت زیر است:

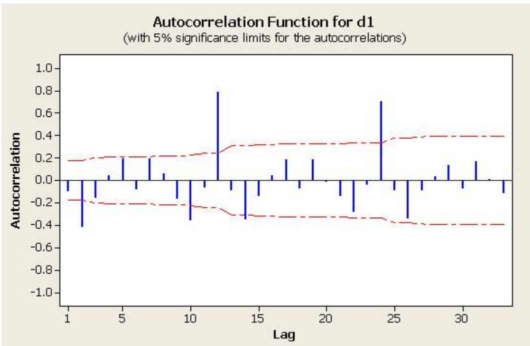


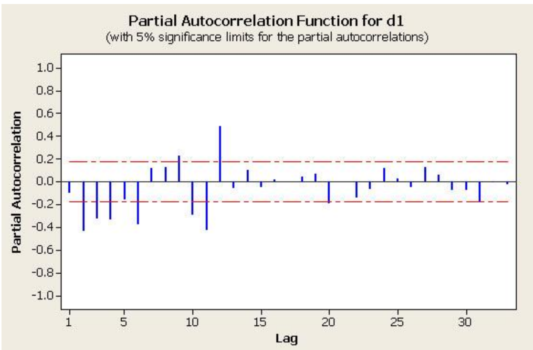
شکل ۵ – ۱۰۹ : نمودار سری زمانی تبدیل شده

اکنون سری تبدیل شده را که دارای واریانس ثابت می باشد، برای حصول ایستایی در میانگین یک بار تفاضلی می کنیم. نمودارهای مربوط به سری تفاضلی شده بصورت زیر است:



شکل ۵ – ۱۱۰ : نمودار سری زمانی تفاضلی شده

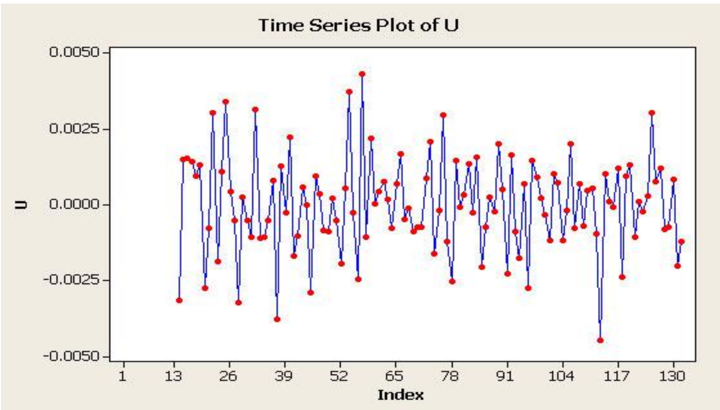




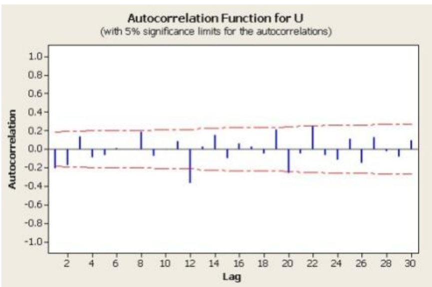
شکل ۵ – ۱۱۱ : acf و pacf سری تفاضلی شده

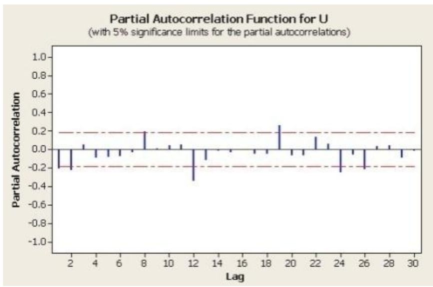
acf سری تفاضلی شده همبستگی های بزرگی را در تأخیرهای ۱۲و۲٤ نشان می دهد که بر لزوم تفاضل فصلی $(1-B^{12})$ تأکید می کند.

بنابراین برای حذف اثر فصلی، سری تفاضلی شده $(1-B)w_t$ را یک بار دیگر و از مرتبه ۱۲ تفاضلی می کنیم. اکنون نمودارهای مربوط به سری جدید $u_t = (1-B)(1-B^{12})w_t$ را رسم می کنیم.



شکل ٥- ۱۱۲ : نمودار سری زمانی u_t





شکل ٥ – ۱۱۳ : acf و pacf سری زمانی u_t

acf سری جدید یک مقدار معنی دار در تأخیر ۱۲ دارد. مقدار آماره tبرای acf در تأخیر ۱۲ برابر ۳.۵۳- است. البته مقدار acf در تأخیر یک و بیست نیز معنی دار می باشد. آماره tبرای acf در تأخیر یک برابر ۲.۲۷- و در تأخیر بیست برابر ۲.۱۷- است، که کمی بزرگتر از ۲ می باشد.

فعلا برای امساک در پارامتر از مقدار acf در تأخیر یک صرفنظر کرده و یک مدل آزمایشی بصورت $ARIMA(0,1,0)*(0,1,1)_{12}$را به داده ها برازش می دهیم. در مراحل بعد با برازش این مدل و بررسی نمودارهای مربوط به باقیمانده ها چنانچه لازم باشد آن را اصلاح می کنیم. شکل این مدل بصورت زیر می باشد:

$$(1-B)(1-B^{12})w_t = (1+\gamma B^{12})z_t$$

بطور کلی می توان گفت در سطح اطمینان ۰.۰۵ تحت فرض صفر که باقیمانده ها تصادفی هستند، از هر بیست ضریب یکی می تواند معنی دار باشد. بنابراین به راحتی می توان همبستگی در تأخیر ۲۰ را نادیده گرفت.

برازش مدل$ARIMA(0,1,0)*(0,1,1)_{12}$

نتیجه برازش این مدل در قسمت مربوط به برآورد پارامترها بصورت زیر می باشد:

ARIMA Model: W

```
Final Estimates of Parameters

Type                Coef     SE Coef       T       P
```

```
SMA   ١٢           ٠.٦٢٩٩         ٠.٠٧٧٤    ٨.١٤   ٠.٠٠٠

Constant  -٠.٠٠٠٠٠٨٥٩  ٠.٠٠٠٠٥٤٢٣  -٠.١٦  ٠.٨٧٤
```

همانطور که ملاحظه می شود مقدار آماره t برای جمله ثابت معنی دار نیست. بنابراین نیازی به حضور جمله ثابت در مدل نمی باشد.

برازش مدل $ARIMA(٠،١،١)(٠،١،١)_{١٢}$ بدون جمله ثابت

ARIMA Model: W

```
Final Estimates of Parameters

Type        Coef  SE Coef      T       P

SMA  ١٢  ٠.٦٣٠١    ٠.٠٧٧٠   ٨.١٨   ٠.٠٠٠


Differencing: ١ regular, ١ seasonal of order ١٢

Number of observations:  Original series ١٣٢, after
differencing ١١٩

Residuals:    SS =  ٠.٠٠٠٢١٩٠٢٤ (backforecasts excluded)

MS =  ٠.٠٠٠٠٠١٨٥٦  DF = ١١٨


Modified Box-Pierce (Ljung-Box) Chi-Square statistic
```

Lag	١٢	٢٤	٣٦	٤٨
Chi-Square	١٤.٩	٣٤.٦	٤٩.٦	٦٢.٧
DF	١١	٢٣	٣٥	٤٧
P-Value	٠.١٨٨	٠.٠٥٦	٠.٠٥٢	٠.٠٦٢

با توجه به مقدار برآورد شده برای تنها پارامتر مدل، این مدل بصورت زیر خواهد بود:

$$(1-B)(1-B^{12})w_t = (1-0.63B^{12})z_t$$

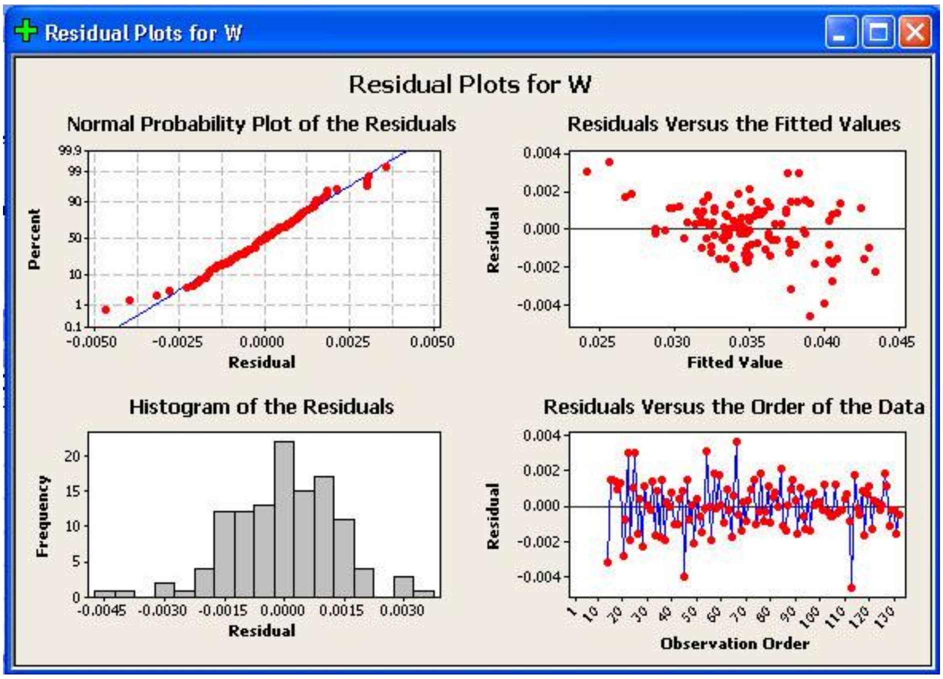
بررسی مناسبت مدل

همانطور که ملاحظه می شود فرضیه ناهمبسته بودن باقیمانده های حاصل از برازش این مدل بوسیله آزمون پرت–مانتو در سطح ٠.٠٥ تأیید می شود. خروجی این آزمون بصورت زیر است:

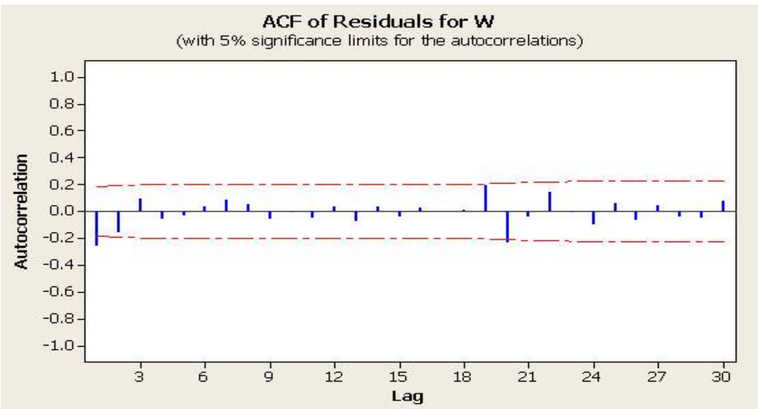
Modified Box-Pierce (Ljung-Box) Chi-Square statistic

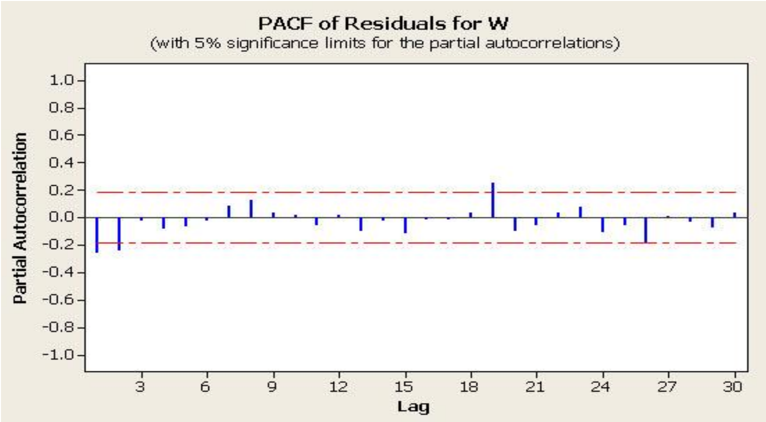
Lag	١٢	٢٤	٣٦	٤٨
Chi-Square	١٤.٩	٣٤.٦	٤٩.٦	٦٢.٧
DF	١١	٢٣	٣٥	٤٧
P-Value	٠.١٨٨	٠.٠٥٦	٠.٠٥٢	٠.٠٦٢

نمودار های زیر نیز فرضیات مربوط به نرمال بودن باقیمانده ها و ثابت بودن واریانس باقیمانده ها را تأیید می کنند.



شکل ۵ – ۱۱۴ : نمودار باقیمانده های مدل ARIMA(۰,۱,۰)*(۰,۱,۱)$_{۱۲}$





شکل ۵– ۱۱۵ : acf و pacf باقیمانده های مدل $ARIMA(0,1,0)*(0,1,1)_{12}$

در نمودار acf یک مقدار معنی دار در تأخیر یک به چشم می خورد. بنابراین تصمیم می گیریم مدل را بصورت $ARIMA(0,1,1)*(0,1,1)_{12}$ اصلاح کنیم. از همان اول هم با توجه به معنی دار بودن acf در تأخیر ۱ و ۱۲ می توانستیم مدل کاملتر فوق را به داده ها برازش دهیم. اما ما عمدا یک مدل با پارامتر کمتر را انتخاب کردیم تا در مرحله بعد اگر تجزیه و تحلیل باقیمانده ها نشان داد که مدل نیاز به اصلاح دارد، آن را اصلاح کنیم.

توجه: در این مثال با بررسی acf و pacf باقیمانده های حاصل از برازش مدل $ARIMA(0,1,0)*(0,1,1)_{12}$ ملاحظه کردیم که رفتار سری باقیمانده ها بجای آنکه شبیه یک فرآیند تصادفی محض باشد مشابه رفتار یک فرآیند MA(۱) می باشد. بنابراین مدل برازش شده را بصورت $ARIMA(0,1,1)*(0,1,1)_{12}$ اصلاح کردیم.

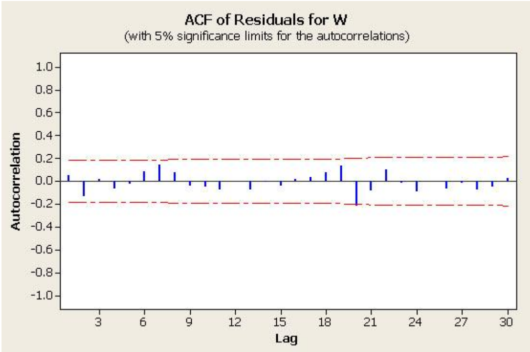
برازش مدل $ARIMA(0,1,1)*(0,1,1)_{12}$

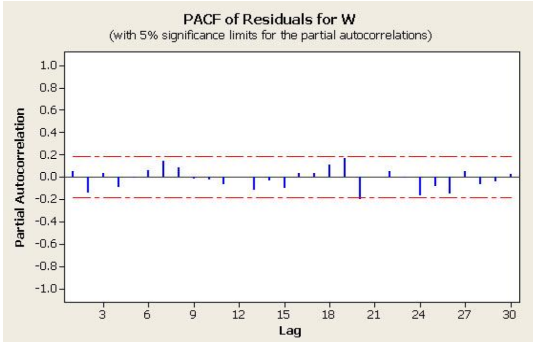
نتیجه برازش این مدل به صورت زیر می باشد.

```
Final Estimates of Parameters
```

```
Type        Coef  SE Coef      T      P
MA   ١   ٠.٤٠١٠   ٠.٠٨٥٩  ٤.٦٧  ٠.٠٠٠
SMA  ١٢  ٠.٦٥٢٩   ٠.٠٧٦٠  ٨.٥٩  ٠.٠٠٠

Differencing: ١ regular, ١ seasonal of order ١٢
Number of observations:  Original series ١٣٢, after
differencing ١١٩
Residuals:    SS =  ٠.٠٠٠١٩٧٠٠٣ (backforecasts excluded)
MS =  ٠.٠٠٠٠٠١٦٨٤  DF = ١١٧

Modified Box-Pierce (Ljung-Box) Chi-Square statistic

Lag          ١٢     ٢٤     ٣٦     ٤٨
Chi-Square   ٨.٩    ٢٤.٣   ٣٤.٥   ٤٧.٠
DF           ١٠     ٢٢     ٣٤     ٤٦
P-Value      ٠.٥٤٥  ٠.٣٣٠  ٠.٤٤٣  ٠.٤٣٠
```

acf و pacf باقیمانده های حاصل از برازش این مدلنیز در شکل (٥-١١٦) رسم شده اند. همانطور که ملاحظه می شود، رفتار باقیمانده های حاصل از برازش این مدل مشابه رفتار یک فرآیند تصادفی محض می باشد. بررسی سایر معیارها نیز نشان می دهد که این مدل برازش بهتری به داده ها دارد.





شکل ۵ – ۱۱۶ : acf وpacf باقیمانده های مدل

ARIMA(۰,۱,۱)*(۰,۱,۱)۱۲

بنابراین ما مدل فوق را به عنوان یک مدل مناسب که برازنده داده ها می باشد در نظر می گیریم. با توجه به برآورد پارامترها شکل مدل بصورت زیر خواهد بود:

$$(1-B)(1-B^{12})w_t = (1-0.4B)(1-0.65B^{12})z_t$$

با ساده کردن مدل خواهیم داشت:

$$w_t = w_{t-1} + w_{t-12} - w_{t-13} + z_t - 0.4z_{t-1} - 0.65z_{t-12} + 0.26z_{t-13}$$

پیش بینی

در این مثال مسأله پیش بینی را بیشتر مورد کنکاش قرار می دهیم. برای پیش بینی با مبداء معلوم مثلا t=۱۳۲، می توان مستقیما از معادله تفاضلی استفاده کرد. برای پیش بینی l مرحله بعد داریم:

$$w_{t+l} = w_{t+l-1} + w_{t+l-12} - w_{t+l-13} + z_{t+l} - 0.4z_{t+l-1} - 0.65z_{t+l-12} + 0.26z_{t+l-13}$$

بنابراین پیش بینی l مرحله بعد از مبداء زمانی t=۱۳۲به صورت زیر است:

$$\hat{w}_{132}(l) = \hat{w}_{132}(l-1) + \hat{w}_{132}(l-12) - \hat{w}_{132}(l-13) + E(z_{132+l} \mid z_{132}, z_{131}, \ldots)$$
$$-0.4E(z_{132+l-1} \mid z_{132}, z_{131}, \ldots) - 0.65E(z_{132+l-12} \mid z_{132}, z_{131}, \ldots) + 0.26E(z_{132+l-13} \mid z_{132}, z_{131}, \ldots)$$

که در آن داریم:

$$\hat{w}_{132}(j) = z_{132+j} \qquad j \le 0$$

و

$$E(z_{132+j} \mid z_{132}, z_{131}, \ldots) = \begin{cases} \hat{z}_{132+j}, \; j \le 0 \\ 0, \; j > 0 \end{cases}$$

مقادیر پیش بینی شده برای شش ماه آینده که با استفاده از نرم افزار محاسبه شده اند عبارتند از:

Forecasts from period ١٣٢

٩٥ Percent Limits

Period	Forecast	Lower	Upper	Actual
١٣٣	٠.٠٣٠٢٤٤٩	٠.٠٢٧٧٠١١	٠.٠٣٢٧٨٨٧	
١٣٤	٠.٠٢٩٩٥٢٢	٠.٠٢٦٩٨٦٩	٠.٠٣٢٩١٧٥	
١٣٥	٠.٠٣١١٦٧٤	٠.٠٢٧٨٣٣٥	٠.٠٣٤٥٠١٤	
١٣٦	٠.٠٣٢٩١٦١	٠.٠٢٩٢٥٠٤	٠.٠٣٦٥٨١٧	
١٣٧	٠.٠٣٣٠٠٥٨	٠.٠٢٩٠٣٦٠	٠.٠٣٦٩٧٥٦	
١٣٨	٠.٠٢٦٥٥٣٨	٠.٠٢٢٣٠١٦	٠.٠٣٠٨٠٦٠	

توجه : این پیش بینی ها مربوط به سری تبدیل شده می باشد. برای به دست آوردن پیش بینی های سری واقعی، عکس تبدیل اعمال شده را به مقادیر پیش بینی شده اعمال

می کنیم. چون ما داده ها رابه صورت $\frac{1}{\sqrt{x_t}}$ تبدیل کرده ایم، بنابراین برای معکوس کردن این تبدیل داریم:

$$\left(\frac{1}{\frac{1}{\sqrt{x_t}}}\right)^2 = \left(\sqrt{x_t}\right)^2 = x_t$$

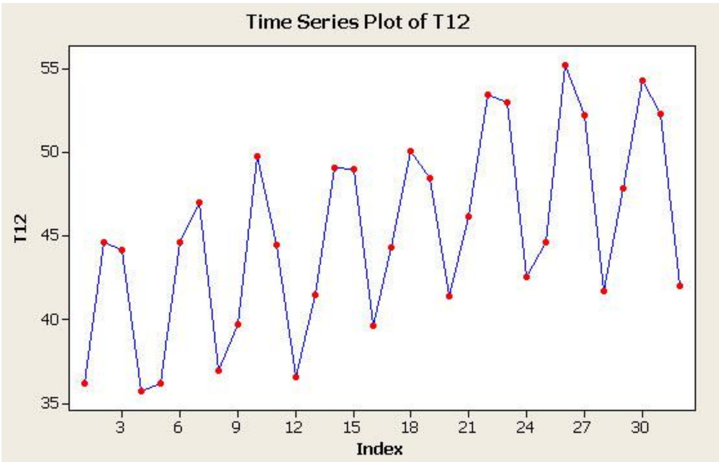
بنابراین کافی است که مقادیر پیش بینی شده فوق را معکوس نمود و سپس به توان ۲ برسانیم. به عنوان مثال مقدار پیش بینی یک ماه بعد سری تبدیل شده عبارت است از ۰.۰۳۰۲ . در نتیجه برای محاسبه پیش بینی یک ماه بعد برای سری اصلی داریم:

$$\left(\frac{1}{0.0302}\right)^2 = 1093.19$$

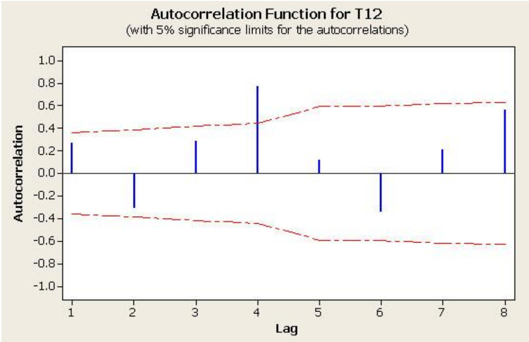
مثال ۱۲

داده های مربوط به مقادیر سه ماهه فروش یک شرکت(برحسب میلیون ریال) برای ۸ سال متوالی در پیوست با نامT۱۲ لیست شده است. با بررسی این داده ها می خواهیم یک مدل آزمایشی مناسب را شناسایی کنیم. بررسی مناسبت مدل شناسایی شده و در صورت لزوم اصلاح آن به خواننده واگذار می شود.

تشخیص مدل آزمایشی

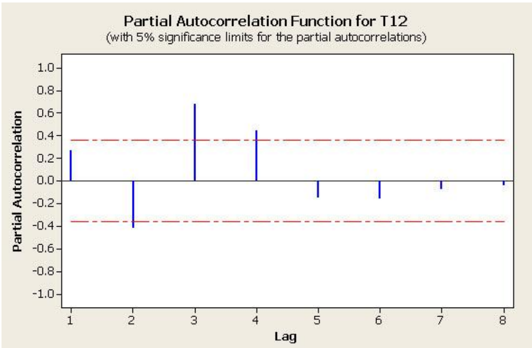


شکل ۵ – ۱۱۷ : نمودار سری زمانی مقادیر سه ماهه فروش



شکل ٥ – ١١٨ : acf و pacf سری زمانی

مقادیر سه ماهه فروش



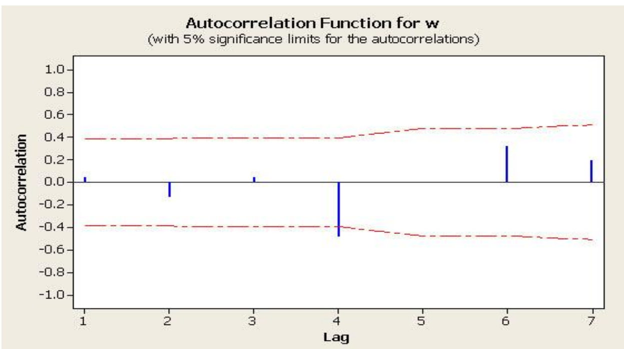
اجرای رویه باکس- باکس پایایی واریانس داده ها را تأیید می کند. بنابراین نیازی به تبدیلات پایداری واریانس نداریم. نمودار سری زمانی و acf آن نشان دهنده یک طرح فصلی قوی با دوره ٤ می باشد. سری تفاضلی شده فصلی $w_t = (1-B^4)x_t$ را در نظر می گیریم. نمودارهای مربوطه به صورت زیر می باشد:

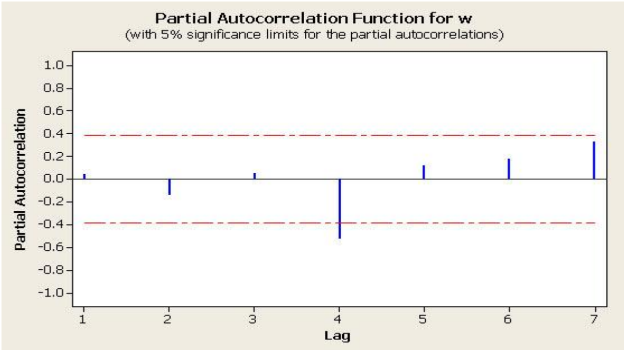
شکل ٥- ١١٩ : نمودار سری زمانی تفاضلی شده از مرتبه ٤

با توجه به اینکه acf این سری فقط یک مقدار معنی دار در تأخیر ٤ دارد، مدل آزمایشی $(1-B^4)x_t = \theta_0 + (1+\gamma B^4)z_t$ پیشنهاد می شود.

acf باقیمانده ها نامناسب بودن این مدل را نشان نداده است. با وجود این همانطور که ملاحظه می شود، pacf این سری نیز فقط یک مقدار معنی دار در تأخیر ٤ دارد که اندازه آن تقریبا مساوی acf در همان تأخیر می باشد. این مطلب نشان می دهد که فرآیند AR فصلی به شکل $(1-\alpha B^4)(1-B^4)x_t = \theta_0 + z_t$ نیز می تواند یک مدلآزمایشی خوب باشد.در این مورد نیز acf باقیمانده ها نامناسب بودن مدل را نشان نمی دهد. بررسی بیشتر و انتخاب مدل نهایی را به خواننده واگذار می کنیم.

شکل ٥ – ١٢٠ : acf و pacf سری زمانی تفاضلی شده از مرتبه ٤





مثال۱۳ (سری مگس)

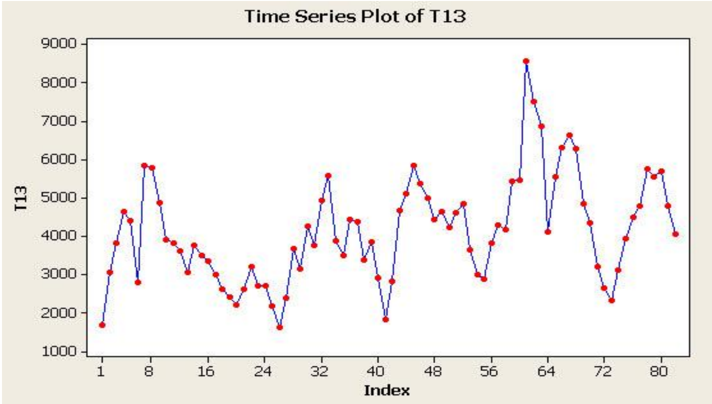
یک تعداد ثابت از مگسهای گوشتی بالغ با نسبت مساوی از نر و ماده در داخل قفسی نگهداری می شوند و روزانه مقدار ثابتی غذا به آنها داده می شود. جمعیت مگسها هر روز(به مدت تقریبا دوسال) شمرده شده و n=۳٦٤ مشاهده حاصل گردیده است. این داده ها توسط نیکلسن در سال۱۹۵۰ تهیه شده اند.

بریلنجر، گوکنهیمر، گوترب و اوستر(۱۹۸۰) از تحلیل سری های زمانی برای این مجموعه داده ها استفاده کرده اند. تانگ این سری را به صورت دو سری مگس A و سری مگس B در نظر می گیرد و معتقد است که سری مگس A غیر خطی است.

سری مگس A ؛ $20 \le t \le 145$ ، z_t

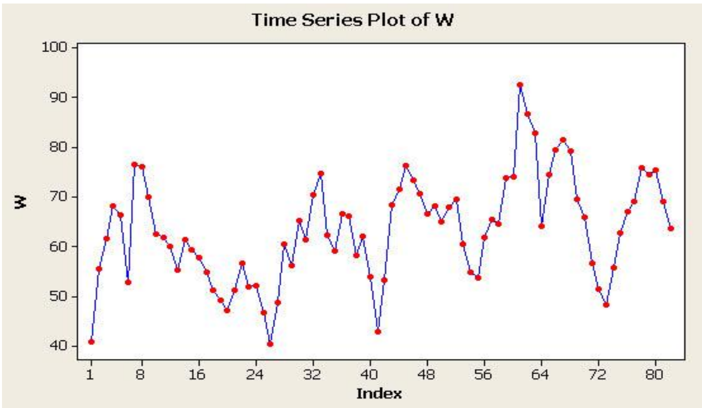
سری مگس B ؛ $218 \le t \le 299$ ، z_t

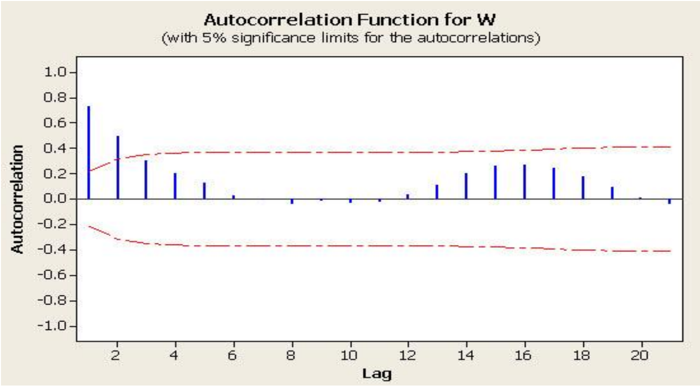
سری ای که ما در اینجا بررسی می کنیم سری مگس B می باشد که خطی بوده و شامل ۸۲ مشاهده است. این سری را در پیوست با نام T۱۳ آورده ایم. نمودار این سری بصورت زیر می باشد:

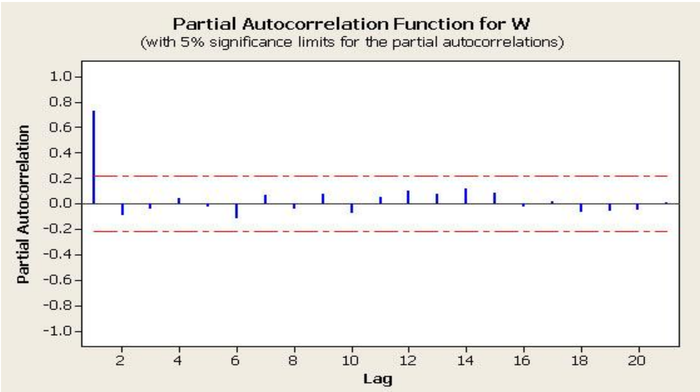


شکل ۵ – ۱۲۱ : نمودار سری زمانی مگس B

اجرای رویه باکس–کاکس نشان می دهد که این داده ها نیاز به یک تبدیل ریشه دوم دارند. زیرا با اجرای این رویه بهترین مقدار پیشنهادی برای پارامتر تبدیل ۰.۵ می باشد که دلالت بر یک تبدیل ریشه دوم دارد. بنابراین سری تبدیل شده به صورت $w_t = \sqrt{x_t}$ می باشد. acf وpacfسری تبدیل شده به شکل زیر است:







شکل ۵– ۱۲۲ : نمودار سری زمانی تبدیل شده و acfو pacfآن

نمودار سری نشان می دهد که این سری در میانگین تقریبا ایستا است. همانطور که ملاحظه می شود acf این سری بطور نمائی تنزل می کند در حالی که pacfآن بعد از تأخیر یک قطع می شود و این دلالت بر یک مدل(۱)ARدارد. بنابراین مدل آزمایشی بصورت زیر خواهد بود:

$$(1-\alpha B)\sqrt{x_t} = z_t$$

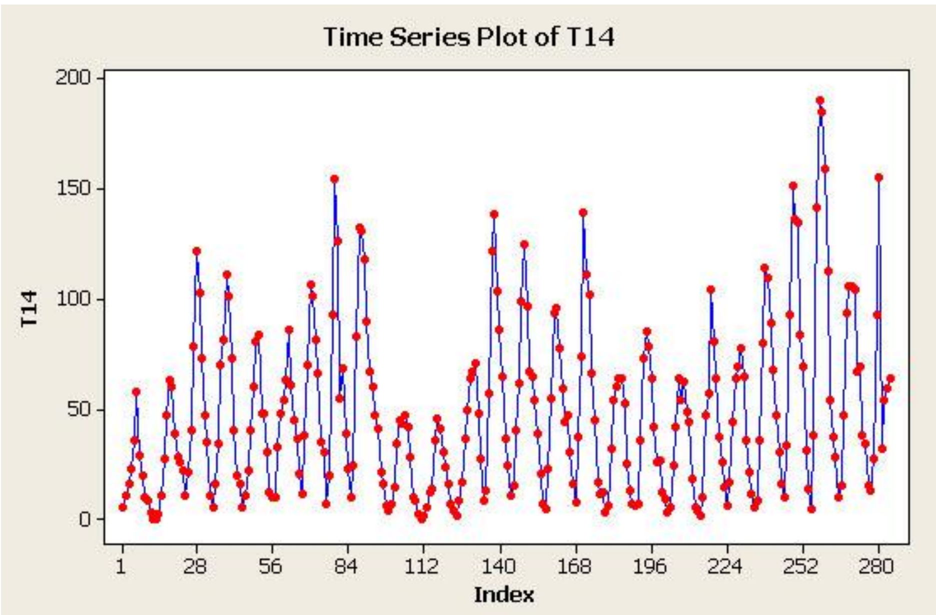
برازش مدل و بررسی مناسبت آن را به خواننده واگذار می کنیم.

مثال ۱٤ (سری لکه های خورشیدی والفر)

سری مربوط به تعداد لکه های خورشیدی سالانه والف که در پیوست این کتاب آمده است را در نظر می گیریم. این سری شامل ۲۸۵ مشاهده برای سالهای ۱۷۰۰ تا ۱۹۸٤ می باشد. دانشمندان براین عقیده اند که تعداد لکه های خورشیدی, روی هوای زمین و در نتیجه فعالیتهای بشر نظیر کشاورزی، ارتباطات از راه دور و غیره اثر می گذارد.

سری تعداد لکه های خورشیدی والف در تاریخچه سری های زمانی بطور وسیعی مورد بحث واقع شده است. این سری به سری تعداد لکه های خورشیدی والفر نیز معروف است. زیرا او دانشجوی والف بوده است. این سری در پیوست با نام T۱٤آمده است.

برای برازش یک مدلARIMA ی مناسب، ابتدا نمودار سری زمانی را رسم می کنیم.



شکل ٥ – ۱۲۳ : نمودار سری زمانی تعداد لکه های خورشیدی والفر

این نمودار نشان می دهد که این سری در میانگین ایستا است. برای بررسی ایستایی در واریانس با فعال کردن خط فرمان در پنجره session ،دستور زیر را تایپ می کنیم.

```
MTB > BOXCOX T١٤ ١.
```

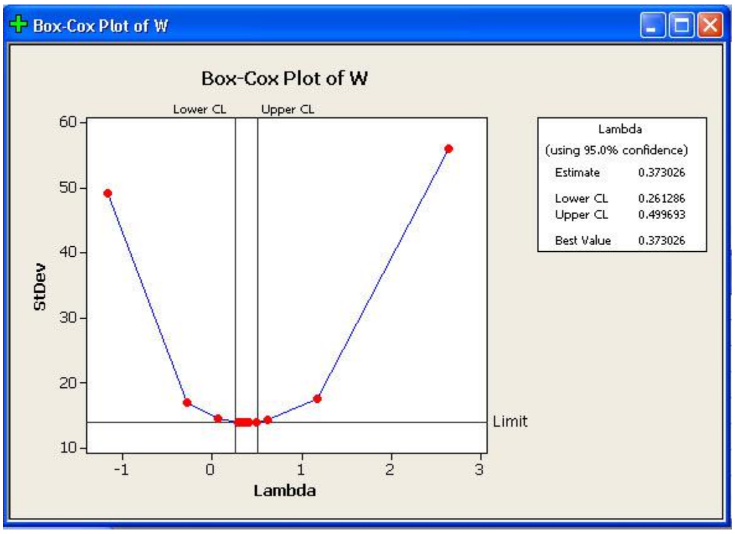
اما با تایپ این فرمان و فشردن کلید اینتر پیغام زیر ظاهر می شود:

```
* ERROR * All data must be positive when using the Box-Cox
transformation.
```

چون در بین داده ها چند صفر وجود دارد لذا رویه باکس–کاکس را نمی توان اجرا کرد. همانطور که می دانید برای اجرای این رویه همه داده ها باید بزرگتر از صفر باشند. برای رفع این مشکل می توان یک مقدار ثابت را به همه داده ها اضافه کرد. ما در اینجا عدد یک را به همه داده ها اضافه می کنیم. برای انجام این کار در پنجره session دستورات زیر را تایپ می کنیم.

```
MTB > name C٢ 'w'
MTB > let 'w'='T١٤'+١
MTB > boxcox w ١.
```

نتیجه اجرای مراحل فوق شکل زیر می باشد.

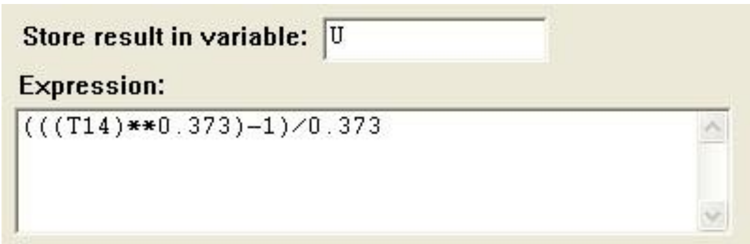


شکل ٥ – ١٢٤ : نمودار باکس– کاکس برای سری والفر

بهترین مقدار پیشنهادی برای پارامتر تبدیل λ، ٠.٣٧٣ به دست آمده است. همانطور که می دانید تبدیل توانی باکس–کاکس برای پایدار کردن واریانس بصورت زیر است:

$$T(x_t) = \frac{x_t^{\lambda} - 1}{\lambda}$$

برای اعمال این تبدیل از منوی Calc گزینهCalculatorرا انتخاب کرده و سپس در پنجره Expression فرمول فوق را تایپ می کنیم.سری تبدیل شده را U می نامیم.



شکل ٥ – ١٢٥ : تبدیل توانی سری والفر

همچنین می توان از پنجره session بصورت زیر استفاده کرد:

MTB > Name C٤'U'

MTB > Let 'U' = (((T١٤)**٠.٣٧٣)-١)/٠.٣٧٣

اکنون برای آنکه از پایدار شدن واریانس داده ها و مؤثر واقع شدن این تبدیل اطمینان حاصل کنیم، رویه باکس–کاکس را برای سری تبدیل شده u_t اجرا می کنیم.

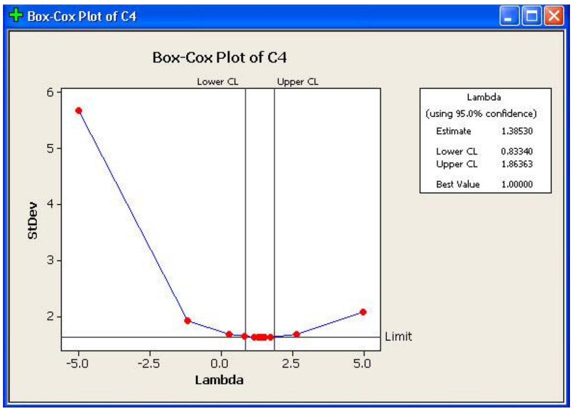
البته با توجه به آنکه برخی از داده ها منفی می باشد، ابتدا به همه داده ها یک مقدار ثابت مثلا عدد ۱۰ را اضافه می کنیم.

MTB > Let C٤ = 'U'+١٠

اکنون رویه باکس–کاکس را برای داده های ستون C٤ اجرا می کنیم.

MTB > boxcox C٤ ١.

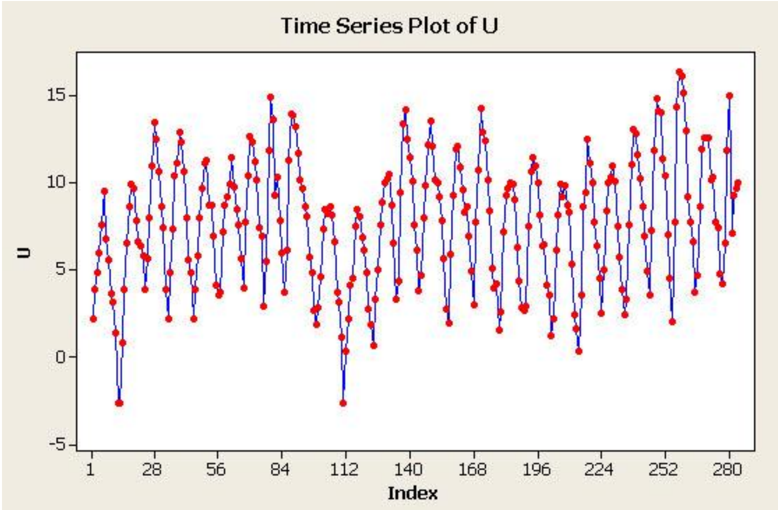
نتیجه اجرای این رویه بصورت زیر است:



شکل ٥ – ١٢٦ : نمودار باکس- کاکس برای سری تبدیل شده

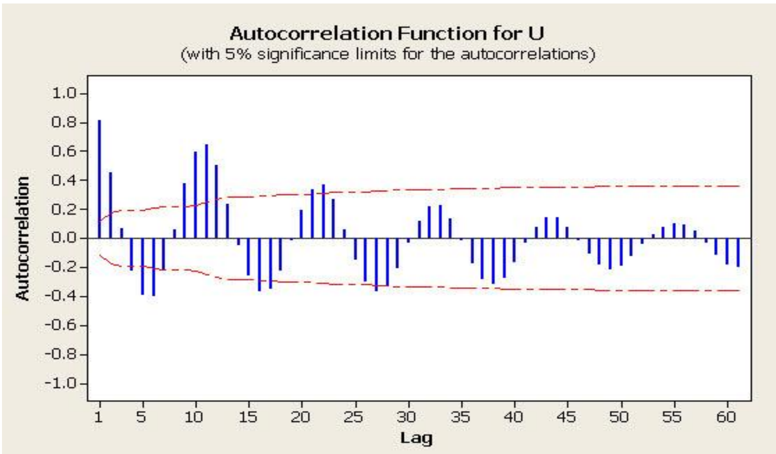
همانطور که ملاحظه می شود بهترین مقدار پیشنهادی برای λ، یک می باشد که به ثابت شدن واریانس داده های تبدیل اشاره شده دارد. لازم به ذکر است که استفاده از یک تبدیل ریشه دوم نیز باعث ثابت شدن واریانس داده ها می شود.

بنابراین سری تبدیل شده U دارای میانگین و واریانس ثابت می باشد. نمودار این سری به شکل زیر است:

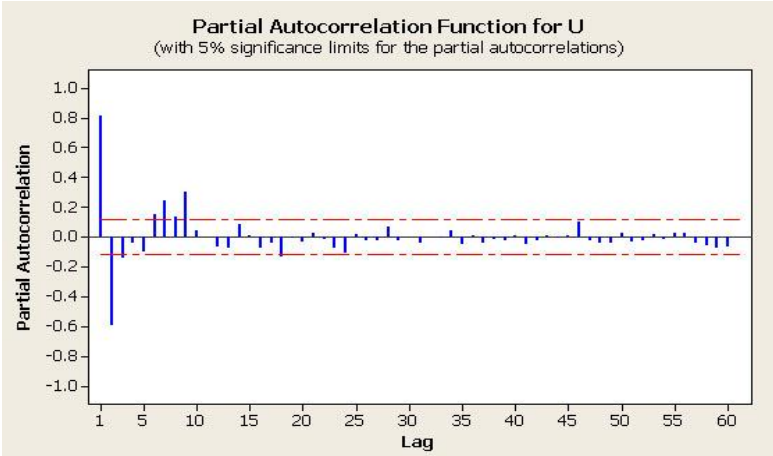


شکل ۵ – ۱۲۷ : نمودار سری u_t (سری والفر پس از اعمال تبدیل مورد نظر)

اکنون acf و pacf سری تبدیل شده u_t را رسم می کنیم.



شکل ۵ – ۱۲۸ : acf سری u_t



شکل ۵ – ۱۲۹ : pacf سری u_t

acf نمونه یک موج سینوس–کسینوس میرا را نشان می دهد. pacf نیز مقادیر بزرگی را در تأخیرهای ۱ ، ۲ و ۹ نشان می دهد. این مسأله امکان یک مدل آزمایشی AR(۲)

$$(1-\alpha_1 B-\alpha_2 B^2-\alpha_3 B^3)u_t=z_t$$

و یا یک مدل آزمایشی AR(۹)

$$(1-\alpha_1 B-\ldots-\alpha_9 B^9)u_t=z_t$$

را تأیید می کند.

اگر از مقادیر تابع خود همبستگی بعد از تأخیر ۳ صرف نظر کنیم، باکس و جنکینز امکان یک مدلآزمایشی AR(۳)را نیز پیشنهاد می کنند.اگر چه تحلیل آنها مبتنی بر داده های تبدیل نشده بین ۱۷۷۰ تا ۱۸۶۹ می باشد.

به واسطه خود همبستگی بزرگ ۰.٦٤ در تأخیر ۱۱ بسیاری از دانشمندان معتقدند که این سری شامل یک دوره یازده ساله است.

برازش مدل(۲)AR

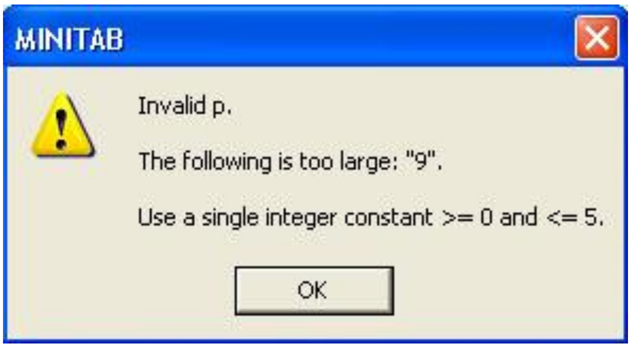
نتیجه برازش این مدل بصورت زیر می باشد:

```
Final Estimates of Parameters

Type          Coef  SE Coef       T      P
        AR   ١      ١.٣١٠٧    ٠.٠٤٧٣    ٢٧.٧٢   ٠.٠٠٠
        AR   ٢     -٠.٦١٠٣    ٠.٠٤٧٢   -١٢.٩٢   ٠.٠٠٠
        Constant   ٢.٢٨٩٩    ٠.١٠١١    ٢٢.٦٥   ٠.٠٠٠
                   Mean       ٧.٦٤٤٣    ٠.٣٣٧٥

Number of observations:  ٢٨٤
Residuals:    SS =  ٨١٥.٥٧٨ (backforecasts excluded)
              MS =   ٢.٩٠٢   DF = ٢٨١

Modified Box-Pierce (Ljung-Box) Chi-Square statistic

Lag              ١٢      ٢٤      ٣٦      ٤٨
Chi-Square      ٥٠.٢    ٦٣.٧    ٧٨.٦    ٨٦.٩
DF               ٩       ٢١      ٣٣      ٤٥
P-Value         ٠.٠٠٠   ٠.٠٠٠   ٠.٠٠٠   ٠.٠٠٠
```

همانطور که ملاحظه می شود آزمون پرت–مانتو فرضیه ناهمبسته بودن باقیمانده ها را به شدت رد می کند. acf و pacf نیز نامناسب بودن مدل را نشان می دهد.

برازش مدل(۹)AR

در صورت برازش این مدل پیغام زیر ظاهر می شود:



شکل ۵ – ۱۳۰ : پیغام ظاهر شده در صورت برازش مدل(۹)AR

در مینی تب نمی توان یک مدل(۹)ARرا برازش داد. در برازش مدلهای آریما در مینی تب هیچکدام از پارامترهای p ، P ، qو Q نمی توانند بیشتر از ۵ باشند و بیشترین تعداد پارامتر که در مجموع می توان برآورد کرد ۱۰ تا است.

ما در اینجا برای برازش یک مدل(۹)AR به داده ها از نرم افزار SAS استفاده می کنیم.

برازش مدل(۹)AR با استفاده از SAS

برای برازش این مدل با استفاده از نرم افزار SAS ، کافی است در پنجره program editorبرنامه زیر را تایپ کنیم.

```
data wolfer;

input U @@;

cards;
```

داده های ستون U را در این قسمت کپی کنید.

```
run;

proc arima;

identify var=U;

estimate p=۹;


run;
```

برخی از نتایج اجرای این رویه در SAS بصورت زیر است.

Conditional Least Squares Estimation

Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MU	2.68054	1.21612	2.20	0.0283	0
AR1,1	1.16068	0.05667	20.48	<.0001	1
AR1,2	-0.42488	0.08865	-4.79	<.0001	2
AR1,3	-0.08494	0.09147	-0.93	0.3539	3
AR1,4	0.24634	0.09679	2.55	0.0115	4
AR1,5	-0.19356	0.10155	-1.91	0.0577	5
AR1,6	-0.09884	0.10107	-0.98	0.3290	6
AR1,7	0.31385	0.10154	3.09	0.0022	7
AR1,8	-0.31286	0.09868	-3.17	0.0017	8
AR1,9	0.38427	0.06119	6.28	<.0001	9

Constant Estimate	0.026663
Variance Estimate	2.386944
Std Error Estimate	1.544974
AIC	1062.861

برآورد پارامترها:

The ARIMA Procedure

Model for variable U

Estimated Mean ٢.٦٨٠٥٤٤

Autoregressive Factors

Factor ١: ١ - ١.١٦٠٦٨ B**(١) + ٠.٤٢٤٨٨ B**(٢) + ٠.٠٨٤٩٤ B**(٣) - ٠.٢٤٦٣٤ B**(٤) + ٠.١٩٣٥٦ B**(٥) + ٠.٠٩٨٨٤ B**(٦)
- ٠.٣١٣٨٥ B**(٧) + ٠.٣١٢٨٦ B**(٨) - ٠.٣٨٤٢٧ B**(٩)

با توجه به پارامترهای برآورد شده، شکل این مدل بصورت زیر می باشد:

$$(1+1.16B-0.42B^2-0.08B^3+0.24B^4-0.19B^5$$
$$-0.09B^6+0.31B^7-0.31B^8+0.38B^9)(u_t-2.68)$$
$$=z_t$$

در مرحله بعد می توان پارامترهایی را که اختلاف معنی داری با صفر ندارند از مدل حذف کرد.

بررسی همبستگی باقیمانده ها

در قسمت دیگری از خروجی SAS برای بررسی همبستگی باقیمانده ها داریم:

Autocorrelation Check of Residuals

To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	.	0	.	-0.012	-0.012	-0.005	0.007	-0.037	-0.022
12	3.14	3	0.3710	-0.019	-0.019	0.038	-0.011	0.015	0.076
18	11.19	9	0.2632	-0.066	-0.059	0.083	0.031	0.103	-0.017
24	14.69	15	0.4741	-0.031	-0.057	-0.032	-0.001	0.058	-0.053
30	18.89	21	0.5922	-0.081	0.055	-0.044	0.002	-0.042	0.005
36	22.57	27	0.7077	-0.047	0.036	-0.039	0.053	-0.020	-0.055
42	26.35	33	0.7873	-0.074	0.007	-0.065	0.022	0.021	-0.027
48	27.20	39	0.9227	0.016	-0.004	-0.036	-0.016	0.013	0.022

همانطور که ملاحظه می شود در ستون چهارم با توجه به مقادیر $p-value$ فرضیه ناهمبسته بودن باقیمانده ها در تمامی تأخیرها پذیرفته می شود. بنابراین مدل(۹)ARیک مدل مناسب می باشد. بررسی سایر فرضیات مربوط به باقیمانده ها با استفاده از نرم افزار SAS به خواننده واگذار می شود.

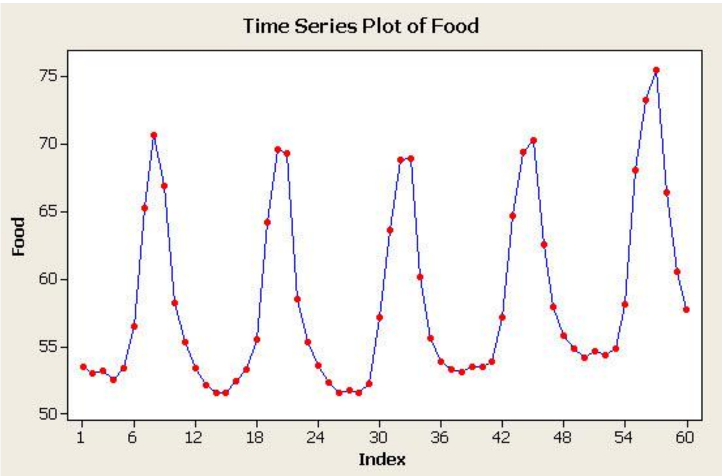
مثال ۱۵

در نرم افزار مینی تب از منوی File گزینه Open Worksheetرا انتخاب کرده و از پنجره باز شده از پوشه DATA فایل .EMPLOY.MTWرا باز می کنیم. سری مشاهدات متغییر Food را که مربوط به ارقام استخدام ماهانه در صنایع غذایی می باشد، در نظر

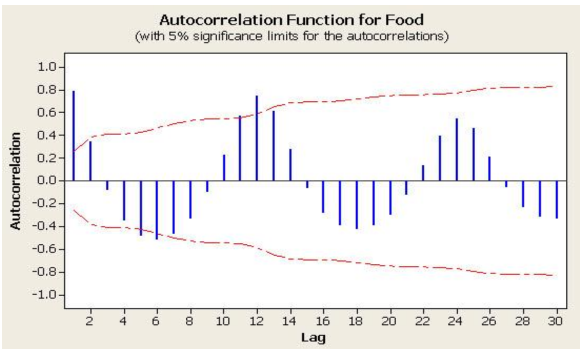
می گیریم. می خواهیم یک مدل مناسب از خانواده مدلهای ARIMA برای این داده ها شناسایی کنیم. بررسی مناسبت مدل شناسایی شده را به خواننده محترم واگذار می کنیم.

تشخیص مدل آزمایشی

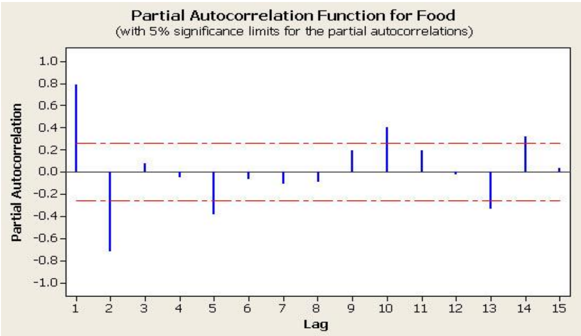
برای تشخیص مدل مناسب همانطور که در مثالهای قبل دیدید، باید نمودار سری زمانی و نمودارهای توابع خود همبستگی و خود همبستگی جزئی را مورد تجزیه و تحلیل قرار داد.



شکل ۵ – ۱۳۱ : نمودار سری زمانی Food



شکل ۵ – ۱۳۲ : acf و pacf سری زمانی Food



همچنین می توان برای درک بهتر رفتارتابع خود همبستگی نمودار قدر مطلق مقادیر خود همبستگی را رسم کرد تا اثر همبستگی منفی موجب سردرگمی نشود.

با توجه به نمودار سری زمانی و نمودار تابع خود همبستگی، وجود یک مؤلفه فصلی در داده ها به وضوح مشخص است. با توجه به اینکه این داده ها ماهانه می باشند، طول دوره فصلی ۱۲ می باشد. بنابراین برای تبدیل این سری به یک سری ایستا به نظر می رسد یک دیفرانس گیری فصلی از مرتبه ۱۲ مفید باشد. البته ممکن است یک روند بلند مدت نیز در داده ها وجود داشته باشد. اما مقدار آن در مقایسه با عامل فصلی ناچیز است. اگر عامل روند قوی تر می بود، برای ایستا کردن سری ممکن بود به یک بار تفاضلی کردن دیگر از مرتبه اول نیاز می داشتیم.

دیفرانس گیری فصلی

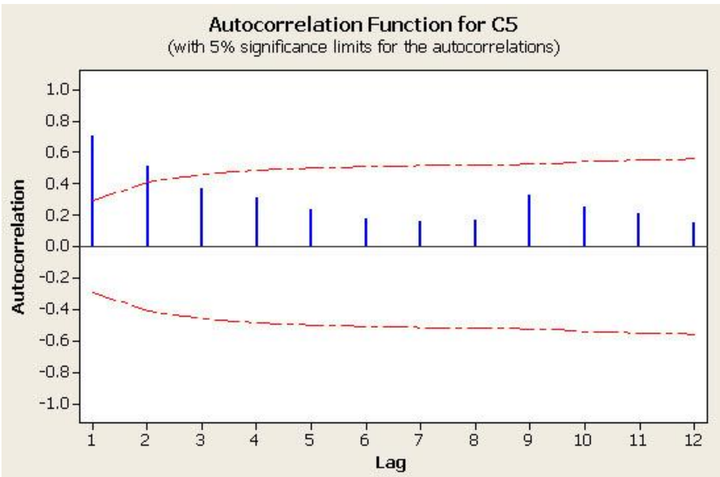
همانطورکه قبلا نیز ملاحظه کردید، دیفرانس گیری فصلی(از مرتبه ۱۲) بصورت زیر تعریف می شود:

$$\nabla_{12}x_t = (1-B^{12})x_t = x_t - x_{t-12}$$

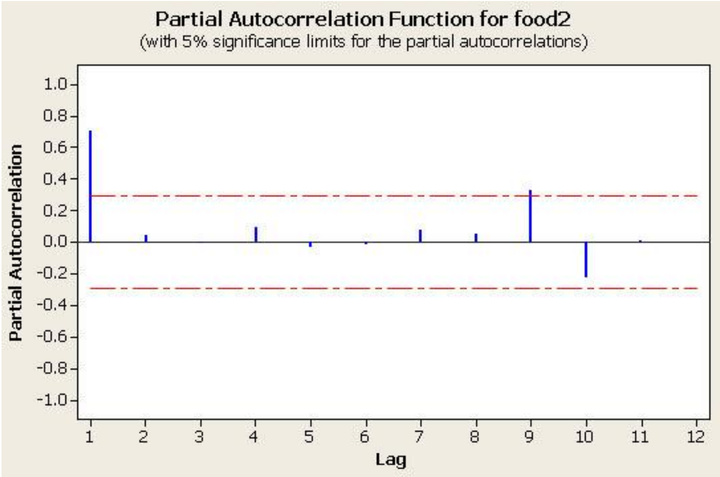
حال برای بررسی بیشتر سری تفاضلی شده تابع خود همبستگی و تابع خود همبستگی جزئی آن را رسم می کنیم.

مقادیر تابع خود همبستگی برای تأخیرهای ۱ و ۲ مثبت و معنی دار می باشند و به دنبال آن همبستگی های مثبتی که به سرعت به صفر میل نمی کنند(میرا نمی شوند) مشاهده می شود. این مدل به طور کلی به یک فرآیند اتورگرسیو اشاره می کند.

در تابع خود همبستگی جزئی یک نقطه اوج به مقدار ۰.۷ در تأخیر اول مشاهده می شود که دلالت بر یک مدل(۱)ARدارد. البته مقدار این تابع در تأخیر نهم نیز معنی دار است اما هیچ دلیلی بر غیر تصادفی بودن آن نداریم. بنابراین ما مدل آزمایشی را به صورت یک مدل(۱)ARتشخیص داده ایم.



شکل ۵- ۱۳۳ : acf و pacf سری تفاضلی شده



نکته

۱. در تابع خود همبستگی اگر r_1 بطور معنی داری با صفر اختلاف داشته باشد ولی مقادیر بعدی r_k تماما به صفر نزدیک باشند در این صورت یک مدلMA مشخص می شود. زیرا acfنظری آن به این شکل است.

۲. اگر به نظر برسد که $r_1, r_2, r_3, \ldots$ بطور نمائی کاهش پیدا می کنند، در این صورت یک مدلAR مناسب است. زیرا همانطور که می دانیدacfیک فرآیند AR که ترکیبی از نمائی و سینوسی است بطور آهسته میرا(کم) می شود.

مراحل برازش مدل و بررسی مناسبت مدل را به خواننده واگذار می کنیم.

۵-۶ تمرین

در این قسمت برای تمرین بیشتر چهار سری زمانی از کتاب باکس–جنکینز را در نظر می گیریم. مشاهدات مربوط به این سریها در پیوست آمده است. با رسم نمودار سری زمانی ,acf و pacf آنها و با توجه به متدولوژی باکس–جنکینز در مدل سازی سری های زمانی با روش تکراری، مدل آزمایشی را تشخیص داده و سپس کفایت آن را مورد بررسی قرار دهید.

برای پرهیز از گمراهی و راهنمائی بیشتر خوانندگان محترم ما در جدول (۵ – ۲) مدل آزمایشی مناسب را برای هر سری معرفی کرده ایم. انتظار می رود با توجه به مثالهای متعدد مطرح شده در این فصل، خواننده مشکل خاصی در مدل سازی سری های زمانی خطی یک متغیره گسسته نداشته باشد.

معرفی سری ها

سری A: قیمتهای بورس عمومی IBM از هفدهم مه ۱۹۶۱ تا دوم نوامبر ۱۹۴۲(شامل ۳۶۹ مشاهده).

سری B : داده های مربوط به اندازه گیری درجه حرارت یک فرآیند شیمیایی هر دقیقه یک بار(شامل ۲۲۶ مشاهده).

سری C : داده های مربوط به اندازه گیری غلظت یک فرآیند شیمیایی هر ساعت یک بار(شامل ۳۱۰ مشاهده).

سری D : داده های مربوط به عملکرد یک فرآیند شیمیایی پخت(شامل ۷۰ مشاهده).

جدول (۵ – ۲) تشخیص مدل آزمایشی سری ها

تشخیص برای سری اصلی	ماهیت ظاهری سری تفاضل گیری شده	درجه تفاضل گیری	سری
ARIMA(۰,۱,۱)	MA مرتبه اول	۱	A
ARIMA(۱,۱,۰) ARIMA(۰,۲,۲)*	AR مرتبه اول فرآیند تصادفی محض	۱ یا ۲	B
ARIMA(۱,۰,۰) ARIMA(۰,۱,۱)*	AR مرتبه اول فرآیند تصادفی محض	۰ ۱	C
ARIMA(۲,۰,۰)	AR مرتبه دوم	۰	D

*بنظر می رسد مرتبه اپراتور میانگین متحرک صفر باشد اما برای بررسیهای بعدی فرم کلی تر حفظ شده است.

۵ – ۷ مباحث دیگر

در ادامه این فصل توجه خواننده محترم را به چند مبحث جالب توجه دیگر جلب می کنیم. این مباحث به جهت تکمیل مطلب و وسعت دید خواننده مفید خواهد بود. البته همانطور که گفته شد مبحث سری های زمانی بسیار گسترده می باشد و مدل سازی سری های زمانی نیاز به اطلاعات و تبحر زیادی دارد و آنچه در این کتاب مورد توجه قرار گرفته تنها محدود به مدل سازی سری های زمانی خطی یک متغیره گسسته در قلمرو زمان می باشد. ما مدل سازی این رده از سری ها را به روش باکس–جنکینز مورد مطالعه قرار دادیم و مثالهای متنوعی را در این رابطه بررسی کردیم.اکنون در این قسمت چند مبحث دیگر را مورد توجه قرار می دهیم.

۵–۷–۱ انتقاد به مدلهای ARIMA

مدلهای ARIMA برای حل بسیاری از مسائل پیش بینی روش نیرومندی به شمار می آیند. این روش می تواند پیش بینی های خیلی دقیق از سری های زمانی تهیه کند و یک روش رسمی و سازمان یافته برای مدل سازی و تجزیه و تحلیل ارائه می دهد. به هر حال این مدلها خالی از چندین محدودیت مهم نیستند.

به طور کلی برای توسعه یک مدل ARIMAی قابل قبول ما نیاز به حداقل ۵۰ مشاهده و ترجیحا ۱۰۰ مشاهده داریم. این تعداد مشاهده نسبتا حجم بزرگی از داده است. انواع بسیاری از مسائل پیش بینی وجود داردکه در آنها این مقدار داده تاریخی در دسترس نیستند.

به عنوان مثال در ساخت کالاهای مد فصلی، تاریخ فصلهای قبلی اغلب به علت تغییرات در سبکها، استراتژیهای بازاریابی وغیره از ارزش کمی در پیش بینی فصل جاری برخوردارند. حتی برای پیش بینی تقاضا در مبنای ماهانه برای محصول نسبتا پایدار، ممکن است ما نتوانیم داده تاریخی از ۵۰ تا ۱۰۰ ماه که کاملا قابل اعتماد است جمع آوری کنیم.

تا آنجایی که به تشخیص مدل وتخمین پارامترها مربوط است، هر فصل کامل از تاریخ درست شبیه یک مشاهده منفرد است. به عبارت دیگر، مثلا برای ۵ سال داده ماهانه ماتنها ۵ مشاهده فروردین، ۵ مشاهده اردیبهشت و ... داریم. بنابراین فصلهای بسیاری از داده تاریخی ممکن است لازم باشد تا یک مدل فصلی مناسبی را بنا کرد. به هر حال باید خاطر نشان کنیم که در غیبت داده تاریخی کافی، می توان داده در دسترس را تجزیه و تحلیل کرد تا یک مدل اولیه به دست آید. به محض اینکه داده تازه در دسترس قرار می گیرد می توان این مدل را پریودوار تجدید نظر کرد.

عدم امتیاز دیگر مدلهای ARIMA این است که در حال حاضر روش مناسبی وجود ندارد که به محض اینکه مشاهده جدیدی در دسترس قرار می گیرد، تخمین های پارامترهای مدل را اصلاح یا به روز در آورد. تحلیلگر ناگریز است که پریودوار به طور کامل مدل را از نو برازش دهد. شاید حتی مطلب جدی تر این باشد که ما باید فرض کنیم که تکامل آینده سری زمانی با گذشته یکسان خواهد بود. یعنی "شکل" مدل با زمان تغییر نخواهد کرد. هیچگونه روشهای سیستماتیک برای هدایت کارآیی مدل یا کنترل تطبیقی اتوماتیک وار مقادیر پارامترها در مقابل تغییرات در فرآیند مربوطه وجود ندارد.

عدم امتیاز نهایی مدلهای ARIMA سرمایه گذاری مورد نیاز در زمان وسایر منابع برای ساختن یک مدل رضایت بخش است. این موضوع ممکن است مهم نباشد اگر تنها تعداد محدودی سری زمانی بررسی شوند. به هر حال نمونه ای از کاربرد پیش بینی در زمینه سیستمهای تولید-موجودی ممکن است از چند صد تا چندین هزار سری زمانی مختلف را دربرداشته باشد.

با وجود این نقاط ضعف، مدلهای ARIMA احتمالا دقیقترین طبقه از مدلهای پیش بینی اند که امروزه در دسترسند. این مدلها به خصوص برای سری های زمانی ای که در آنها فاصله نمونه گیری خیلی کوچکند، به دلیل اینکه تاریخ نسبتا طولانی ای را به آسانی می توان تهیه کرد، به نحو شایسته ای مناسبند. به این جهت آنها به طور گسترده برای سری ها یی به کار برده شده اند که مشاهدات ساعتی، روزانه و یا هفتگی مورد توجه اند. به عنوان مثال، سری های زمانی ویژگی های خروجی فرآیندهای شیمیایی، نظیر مقدار وخلوص و ... اغلب از این نوع کلی اند.

مدلهای ARIMA بیشترین سود آوری را موقعی خواهند داشت که تنها تعداد اندکی سری زمانی مورد توجه اند و مدیریت مایل است که منابع لازم را صرف کند تا درجه بالایی از دقت پیش بینی را به دست آورد.

۵-۷-۲ مدلهای توابع انتقال

درمباحث گذشته کاربرد مدلهای ARIMA را برای مدل بندی و پیش بینی سری های زمانی توضیح دادیم. این مدلها را می توان به صورت مدلهای سری های زمانی تک متغیره مطالعه کرد. زیرا آنها تنها از تاریخ گذشته متغیرمورد بررسی استفاده می کنند. اگر سری زمانی مورد بررسی را x_t بنامیم که با یک یا تعداد بیشتری از سری های زمانی دیگر در رابطه باشد، می توان مدلی بنا کرد که از محتوای اطلاعاتی سری های زمانی اخیر استفاده کند تا به پیش بینی x_t کمک نماید. این مدلها، مدلهای توابع انتقال نامیده می شوند و بسط منطقی مدلهای ARIMA ی تک متغیره هستند.

به عنوان مثال فرض کنید که بخواهیم سری زمانی گسسته x_t را مدل سازی و پیش بینی کنیم. یک مدل تک متغیره سعی می کند که دریافت جاری x_t را با دریافتهای قبلی $x_{t-1}, x_{t-2}, \ldots$ مرتبط کند. حال فرض کنید که بدانیم x_t با یک سری زمانی دومی مانند y_t مرتبط است. اگر این سری های زمانی ایستا باشند، مناسب است فکر کنیم که x_t و y_t انحرافات از مقادیر تعادل یا میانگین هستند. یا اگر سری ها غیرایستا

باشند، به صورت انحرافاتی از مبداهای مناسب خواهند بود. یک مدل ساده که دو سری فوق را به یکدیگر مرتبط می کند به صورت زیر است:

$$x_t = \upsilon_0 y_t + \upsilon_1 y_{t-1} + \ldots + \upsilon_k y_{t-k} + n_t \qquad (I)$$

که در آن $\upsilon_0, \upsilon_1, \ldots \upsilon_k$ وزنهای مجهولند و n_t مؤلفه خطا بوده که ممکن است(ولی نه لزوما) فرض شود به صورت نرمال و به طور مستقل با میانگین صفر و واریانس ثابت توزیع شده است. ما معمولا y_t را **سری زمانی ورودی**و x_t را **سری زمانی خروجی** و معادله (I)را یک **مدل تابع انتقال** می نامیم.

مدلهای تابع انتقال کاربرد عملی وسیعی پیدا کرده اند. این مدلها به طور طبیعی از زمینه هایی ناشی می شوند که یک ساختار همبسته کننده یا ساختار علی بین متغیرهایی که به طور موقت یا به طور مداوم با هم مرتبط بوده اند، وجود دارد. به عنوان مثال فروش در پریود t(سری خروجی) با هزینه های آگهی در پریودهای قبلی(سری ورودی) مرتبط است. حداکثر تولید الکتریسیته روزانه(خروجی) با متغیر های خاص هوا نظیر حداکثر درجه حرارت روزانه، رطوبت نسبی و پوشش ابری(ورودی)ارتباط دارد. ماکزیمم تخلیه سیل از یک رودخانه(خروجی) با شدت بارندگی در نقاط مختلف بالای رودخانه(ورودی) ارتباط دارد.

این مدلها در بسیاری از انواع فرآیندها و مسائل کنترل کیفیت نیز مفیدند. در این مسائل مقدار یک ویژگی کیفی در زمان tبا تنظیم متغیرهای قابل کنترل فرآیند در پریودهای زمانی قبل مرتبط است.

مدلهای تابع انتقال فرض می کنند که ورودی y_t متغیر خروجی x_t را تحت تأثیر قرار می دهد. اما این رابطه تک سویی است. یعنی x_t روی y_t تأثیری ندارد. این موضوع در عمل می تواند غیر واقعی باشد. مثلا اگر ما دو متغیر حجم فروش و مخارج آگهیرا درنظر بگریم، درحالت کلی مخارج آگهی بالا وقتی روی می دهد که فروش بالا

باشد. تا حدودی به این دلیل که مخارج آگهی زیاد، فروش را بالا می برد. به هر حال در بسیاری از شرکتها به میزانی که درآمد حاصل از فروش افزایش می یابد، آگهی نیز افزایش می یابد. بنابراین در این مثال بین دو متغیر بازخور وجود دارد. یک مدل تابع انتقال به طور شایسته این رابطه را بیان نخواهد کرد. برای بیان رابطه بین فروش و آگهی یک **مدل سری زمانی چند گانه**می تواند مورد استفاده قرار گیرد.

۵-۷-۳ آشنایی با سری های زمانی غیر خطی

فرض خطی بودن اغلب خیلی مشکوک می باشد. فرآیندهای زیادی در طبیعت زمینه های مهندسی و فیزیک اتفاق می افتد که یک رفتار غیر خطی را نشان می دهند. معروفترین این پدیده ها "سیکل حدی" می باشد که در زمینه های مختلف مانند زمین شناسی و تئوری مدارهای الکتریکی بسیار مفید شناخته شده است. بسیاری از سری های زمانی بصورت مکفی توسط یک مدل خطی نشان داده نمی شوند. این حقیقت در سالهای اخیر باعث بوجود آمدن مدلهای جالبی مانند مدلهای دوخطی، مدلهای اتورگرسیو نمائی, مدلهای اتورگرسیو آستانه ای و ... شده است.

پریستلی در سال ۱۹۸۰ یک کلاس عمومی از مدلهای غیرخطی به نام مدل وابسته به حالت (state dependent model) را ارائه دادکه شامل همه مدلهای خطی و مدلهای غیرخطی استاندارد می باشد. این مدلها را به اختصار SDM می نامند. اکنون به طور مختصر به معرفی برخی از مدلهای غیر خطی می پردازیم.

۱- مدلهای دوخطی (bilinear model)

یک کلاس خاص از مدلهای غیر خطی که بطور گسترده ای در مقالات تئوری کنترل بحث شده است، مدلهای دوخطی می باشد. خصیصه مورد توجه فرآیند دوخطی

این است که با وجود غیر خطی بودن ساختمان تئوری آن شبیه به یک فرآیند خطی است.

مدل دوخطی ARMA

مدل دوخطی ARMA از مرتبه (p,q,P,Q) را با BARMA(p,q,P,Q) نشان می دهیم و بصورت زیر تعریف می کنیم:

$$x_t = \sum_{j=1}^{p} \alpha_j x_{t-j} + \sum_{i=0}^{q} b_i z_{t-i} + \sum_{k=0}^{Q}\sum_{l=1}^{P} \beta_{kl} z_{t-k} x_{t-l}$$

که در آن z_t همان فرآیند تصادفی محض است. مدل فوق نسبت به ها x و z ها بطور مجزا خطی است ولی نسبت به هردوی آنها خطی نیست. واضح است که همه مدلهای ARMA(p,q) حالت خاصی از مدل اصلی فوق می باشند. زیر کلاسهای خاصی از مدل اصلی فوق به اختصار معرفی می شود.

مدلهای بالا قطری(super diagonal

شکل کلی این مدلها بصورت زیر است:

$$x_t = \sum_{k=1}^{Q}\sum_{l=2}^{P} \beta_{kl} z_{t-k} x_{t-l} + z_t$$
$$k < l$$

به عنوان مثال یک حالت خاص آن بصورت $x_t = \beta x_{t-2} z_{t-1} + z_t$ می باشد.

مدل زیر قطری(subdiagonal

یک حالت خاص این مدل بصورت $x_t = \beta x_{t-k} z_{t-k-k} + z_t$ می باشد. مانند $x_t = \beta x_{t-2} z_{t-3} + z_t$.

مدل قطری(diagonal)

یک خالت خاص این مدل بصورت $x_t = \beta x_{t-k} z_{t-k} + z_t$ می باشد.مانند $x_t = \beta x_{t-1} z_{t-1} + z_t$.

۲- مدل آتورگرسیو نمائی (exponential autoregresive)

رفتار غیر خطی ارتعاشات توسط مؤلفین زیادی مطالعه شده است و بطور خاص شامل خصیصه های پدیده فرکانس وابسته به دامنه و سیکل حدی می باشد. یک مدل سری زمانی گسسته که این پدیده ها را بیان می کند، دارای خواصی مشابه با این ارتعاشات تصادفی غیر خطی است. این مدل به شکل اتو رگرسیو با ضرایب غیر ثابت می باشد که ضرایب بطور نمائی به مشاهدات وابسته هستند و از این رو اتو رگرسیو نمائی نامیده می شود.

مدل اتورگرسیو نمائی برای بیان شکلهای معینی از نوسانات تصادفی غیر خطی معرفی شده اند. این نوسانات بطور خاص توسط معادلات دیفرانسیل مرتبه دوم زیر توصیف می شوند:

$$\ddot{x}_{(t)} + f\{\dot{x}_{(t)}\} + g\{x_{(t)}\} = y_{(t)}$$

که در آن $f\{.\}$ نیروی کاهش دهنده و $g\{.\}$ نیروی تقویت کننده نوسانات می باشد و $y(t)$ یک ورودی تصادفی است. مثلا اگر $y_{(t)} = 0$ و $f(\dot{x}) = 0$ و $g(x) = x$ آنگاه یک معادله ساده حرکت یکنواخت حاصل می گردد.

در ارتعاشات غیر خطی $f(\dot{x})$ و $g(x)$ شکل غیر خطی خواهند داشت و یک نوع خاص از رفتار غیر خطی را نشان می دهند که بصورت سیکل حدی و پدیده فرکانس وابسته به دامنه اتفاق می افتند. برای ساختن یک مدل سری زمانی که معرف اثر فرکانسهای وابسته به دامنه است، اوزاکی پیشنهاد می کند که با یک مدل AR(۲)(با پارامتر گسسته) شروع کنیم.

$$x_t - \alpha_1 x_{t-1} - \alpha_2 x_{t-2} = z_t \qquad (I)$$

و سپس ضرایب α_2, α_1 را به x_{t-1} وابسته نمائیم. مخصوصا او پیشنهاد می کند که ضرایب را بصورت توابع نمائی از x_{t-1} به شکل زیر بسازیم:

$$\alpha_1 = \phi_1 + \pi_1 \exp(-\gamma x_{t-1}^2)$$
$$\alpha_2 = \phi_2 + \pi_2 \exp(-\gamma x_{t-1}^2)$$

بنابراین مدل (I) به مدل زیر تبدیل می شود:

$$x_t = (\phi_1 + \pi_1 e^{-\gamma x_{t-1}^2}) x_{t-1} + (\phi_2 + \pi_2 e^{-\gamma x_{t-1}^2}) x_{t-2} + z_t$$

که $\gamma, \pi_2, \pi_1, \phi_2, \phi_1$ مقادیر حقیقی هستند و $\gamma > 0$ می باشد.

برای چنین مدلی x_t یک فرآیند متقارن با میانگین صفر است و یک مدل اتورگرسیو نمائی مرتبه دوم نامیده می شود.

۳- مدل اتورگرسیو آستانه ای (threshold autoregressive)

یک کلاس خاص از مدلهای غیر خطی، مدل آتورگرسیو آستانه ای می باشد که توسط تانگ(۱۹۷۸) معرفی شده است. ایده اصلی این است که فرآیند x_t را با یک مدل خطی اتورگرسیو شروع کنیم و سپس پارامترها را طوری درنظر بگیریم که مطابق با تعداد متناهی از مقادیر گذشته x_t و یا مطابق با تعداد متناهی از مقادیر گذشته یک فرآیند متحد با آن مثل y_t تغییر کنند.

تعریف)فرض کنید $\{r_0, r_1, ..., r_p\}$ یک زیر مجموعه مرتب شده از اعداد حقیقـی باشـند بطوری که $-\infty < r_1 < r_2 < ... < r_{p-1} < \infty$ ، که r_0 و r_p به ترتیب $-\infty$ و ∞ مـی باشند. این مجموعه یک افراز برای اعداد حقیقی R تعریف می کند بطوری که :

$$R = R_1 \cup R_2 \cup ... \cup R_P \qquad ; \quad R_i = (r_{i-1}, r_i]\ i = 1, ..., p$$

r_i را آستانه i ام و R_i را ناحیه آستانه i ام می نامیم.

تعریف)یک مدل اتو رگرسیو آستانه ای مرتبه اول TAR(۱)بصورت زیر توصیف می شود:

$$x_t = \begin{cases} a^{(1)} x_{t-1} + z_t^{(1)} & x_{t-1} \le r_1 \\ a^{(2)} x_{t-1} + z_t^{(2)} & x_{t-1} > r_1 \end{cases}$$

این مدل اولین بار توسط تانگ در ارتباط با آنالیز داده های مربوط به جریان رودخانه ها مطالعه شده است. این مدل را می توان به شکل k-آستانه ای توسعه داد:

$$x_t = a^{(i)} x_{t-1} + z_t^{(i)} \qquad ; \quad x_{t-1} \in R_i$$

که $R_1, R_2, ..., R_k$ ناحیه های آستانه ای زیر مجموعه های خط حقیقی R^1 هستند. مدل k-آستانه ای را می توان تقریب خوبی برای یک مدل غیر خطی مرتبه اول در حالت کلی در نظر گرفت. مانند $x_t = \lambda(x_{t-1}) + z_t$ که λ یک تابع کلی از x است.

تعریف)یک مدل اتورگرسیو آستانه ای مرتبه k ، TAR(k)بصورت زیر توصیف می شود:

$$x_t + a_1^{(i)} + ... + a_k^{(i)} x_{t-k} = z_t^{(i)} \qquad ; (x_{t-1}, ..., x_{t-k}) \in R^{(i)} ; i = 1, ..., k$$

که $R^{(i)}$ ناحیه ای از فضای اقلیدسی k بعدی(R^k) است. این مدل تقریبی برای مدل کلی $x_t = f(x_{t-1}, ..., x_{t-k}) + z_t$ در نظر گرفته می شود.

امام علی (ع) :

العا لم من علم قدره.

عالم کسی است که قدر خودش را بداند.

پیوست

داده ها ی سری های T۱ تا T۱۵ را بصورت افقی بخوانید.

سری T۱: داده های مربوط به مصالح خط راه آهن از ژانویه ۱۹٦۸ تا ژوئن ۱۹۷٦ (n=۱۰۲)

639	643	640	653	667	667	663	654	649	651	659	672	670
675	692	702	706	710	722	729	740	755	763	788	818	826
821	819	827	848	881	879	878	878	868	856	844	824	820
819	813	815	822	818	815	792	769	775	771	773	780	779
774	772	775	770	766	771	773	772	767	775	777	777	776
779	787	790	791	792	802	799	792	780	790	799	810	814
828	862	874	892	872	869	870	859	857	870	867	856	854
862	861	855	846	847	845	838	828	823	814	812		

سری T۲ : متوسط تعداد نقایص مشاهده شده در هر کامیون در بازرسی نهایی درانتهای خط تولید کارخانه تولید کامیون(n=٤٥)

1.20	1.50	1.54	2.70	1.95	2.40	3.44	2.83	1.76	2.00	2.09
1.89	1.80	1.25	1.58	2.25	2.50	2.05	1.46	1.54	1.42	1.57
1.40	1.51	1.08	1.27	1.18	1.39	1.42	2.08	1.85	1.82	2.07
2.32	1.23	2.91	1.77	1.61	1.25	1.15	1.37	1.79	1.68	1.78
1.84										

سری T۳ : فروش تعداد تخته پوستهای خز در کانادا توسط شرکت HUDNON BAY بین سالهای ۱۸۵۷ تا ۱۹۱۱ (n=۵۵)

23362	31642	33757	23226	15178	7272	4448	4926	5437
16498	35971	76556	68392	37447	45686	7942	5123	7106
11250	18774	30508	42834	27345	17834	15386	9443	7599
8061	27187	51511	74050	78773	33899	18886	11520	8352
8660	12902	20331	36853	56407	39437	26761	15185	4473
5781	9117	19267	36116	58850	61478	36300	9704	3410
3774								

سری T٤ : تعداد زنان بیکار ۱٦ تا ۱۹ ساله (به هزار نفر) از ژانویه ۱۹٦۱ تا دسامبر ۱۹۸۵ (n=۳۰۰)

375	384	383	326	344	375	419	424	429	399	376	288	360
376	360	381	354	301	333	339	316	352	378	360	388	398
377	383	449	415	429	369	414	462	447	403	409	390	380
438	431	426	348	394	396	451	384	491	466	454	442	475
401	406	385	380	422	397	430	433	421	374	401	451	465
456	469	466	412	427	414	384	328	395	381	360	383	383
403	425	422	414	382	390	320	412	437	421	450	442	450
412	422	372	375	392	356	392	426	442	426	406	392	426
445	464	379	409	497	459	513	549	447	445	432	514	565
557	601	582	587	560	590	556	582	527	585	556	574	556
582	583	644	620	618	623	546	568	595	605	598	592	558
595	549	637	568	605	594	567	545	545	592	576	593	603
631	614	617	546	632	673	732	593	693	730	731	733	802
755	805	751	855	769	800	825	799	802	765	827	760	781
769	766	752	751	761	873	750	758	772	791	813	781	797
802	782	838	756	764	796	781	780	679	748	759	749	756
802	754	792	772	769	731	746	741	712	723	698	746	754
735	722	737	728	773	723	741	738	765	748	707	808	746
773	751	721	731	735	701	762	783	796	803	806	765	781
768	812	854	858	818	856	897	817	872	895	825	922	915
902	908	911	919	861	827	855	867	836	916	828	835	792
771	757	756	712	733	746	728	707	666	636	676	696	654
613	677	705	680	699	650	687	638	670	555	631	676	659
689												

سری T۵: میزان مرگ و میر سالانه ناشی از تصادفات (در صد هزار نفر) در ایالت پنسیلوانیا بین سالهای ۱۹۵۰ تا ۱۹۸٤(n=۳۵)

54.2	55.0	53.9	53.1	50.2	50.8	48.8	48.1	45.5	44.5	43.0
41.5	44.2	45.1	45.1	46.4	44.9	46.9	47.5	45.4	45.2	45.3
44.1	43.2	40.6	39.8	38.2	39.5	38.2	40.7	38.7	36.9	34.6
34.9	35.1									

سری T٦ : تولید سالانه تنباکو در آمریکا از ۱۸۷۱ تا ۱۹۸٤ (n=۱۱٤)

327	385	382	217	609	466	621	455	472	469	426
579	509	580	611	609	469	661	525	648	747	757
767	767	745	760	703	909	870	852	886	960	976
857	939	973	886	836	1054	1142	941	1117	992	1037
1157	1207	1326	1445	1444	1509	1005	1254	1518	1245	1376
1289	1211	1373	1533	1648	1565	1018	1372	1085	1302	1163
1569	1386	1881	1460	1262	1408	1406	1951	1991	1315	2107
1980	1969	2030	2332	2256	2059	2244	2193	2176	1668	1736
1796	1944	2061	2315	2344	2228	1855	1887	1968	1710	1804
1906	1705	1749	1742	1990	2182	2137	1914	2025	1527	1786
2064	1994	1429	1728							

سری T۷ : غلظت یک فرآیند شیمیایی (n=۱۹۷)

17.0	16.6	16.0	16.1	17.1	16.9	16.8	17.4	17.1	17.0	16.7
17.4	17.2	17.4	17.4	17.0	17.0	17.2	17.4	16.8	17.1	17.4
17.4	17.5	17.4	17.6	17.4	17.3	17.0	17.8	17.5	18.1	17.5
17.4	17.4	17.1	17.6	17.7	17.4	17.8	17.6	17.5	16.5	17.8
17.3	17.3	17.1	17.4	16.9	17.3	17.6	16.9	16.7	16.8	16.8
17.2	16.8	17.6	17.2	16.6	17.1	16.9	16.6	18.0	17.2	17.3
17.0	16.9	17.3	16.8	17.3	17.4	17.7	16.8	16.9	17.0	16.9
17.0	16.6	16.7	16.8	16.7	16.4	16.5	16.4	16.6	16.5	16.7
16.4	16.4	16.2	16.4	16.3	16.4	17.0	16.9	17.1	17.1	16.7
16.9	16.5	17.2	16.4	17.0	17.0	16.7	16.2	16.6	16.9	16.5
16.6	16.6	17.0	17.1	17.1	16.7	16.8	16.3	16.6	16.8	16.9
17.1	16.8	17.3	17.2	17.3	17.2	17.3	17.2	17.2	17.5	16.9
16.9	16.9	17.0	16.5	16.7	16.8	16.7	16.7	16.6	16.5	17.0
16.7	16.7	16.9	17.4	17.1	17.0	18.8	17.2	17.2	17.4	17.2
16.9	16.8	17.0	17.4	17.2	17.2	17.1	17.1	17.1	17.4	17.2
16.9	16.9	17.0	16.7	16.9	17.3	17.8	17.8	17.6	17.5	17.0
16.9	17.1	17.2	17.4	17.5	17.9	17.0	17.0	17.0	17.2	17.3
17.4	17.4	17.0	18.0	18.2	17.6	17.8	17.7	17.2	17.4	

سری T۸ : داده های مربوط به چسبندگی یک فرآیند شیمیایی (n=۱۰۰)

29.33	19.98	25.76	29.00	31.03	32.68	33.56	27.50	26.75
30.55	28.94	28.50	28.19	26.13	27.79	27.63	29.89	28.18
26.65	30.01	30.80	30.45	36.61	31.40	30.83	33.22	30.15
27.08	33.66	36.58	29.04	28.08	30.28	29.35	33.60	30.29
20.11	17.51	23.71	24.22	32.43	32.44	29.39	23.45	23.62
28.12	29.94	30.56	32.30	31.58	27.99	24.13	29.20	34.30
26.41	28.78	21.28	21.71	21.47	24.71	33.61	36.54	35.70
33.68	29.29	25.12	27.23	30.61	29.06	28.48	32.01	31.89
31.72	29.02	31.92	24.28	22.69	26.60	28.86	28.27	28.17
28.58	30.76	30.62	20.84	16.57	25.23	31.79	32.52	30.28
26.14	19.03	24.34	31.53	31.95	31.68	29.10	23.15	26.74
32.44								

سریT۹ : مسافت پیموده شده توسط مسافرین خطوط هوایی ایالات متحده از ژانویه ۱۹۶۰ تا دسامبر ۱۹۷۷ (n=۲۱٦)

2.42	2.14	2.28	2.50	2.44	2.72	2.71	2.74	2.55
2.49	2.13	2.28	2.35	1.82	2.40	2.46	2.38	2.83
2.68	2.81	2.54	2.54	2.37	2.54	2.62	2.34	2.68
2.75	2.66	2.96	2.66	2.93	2.70	2.65	2.46	2.59
2.75	2.45	2.85	2.99	2.89	3.43	3.25	3.59	3.12
3.16	2.86	3.22	3.24	2.95	3.32	3.29	3.32	3.91
3.80	4.02	3.53	3.61	3.22	3.67	3.75	3.25	3.70
3.98	3.88	4.47	4.60	4.90	4.20	4.20	3.80	4.50
4.40	4.00	4.70	5.10	4.90	5.70	3.90	4.20	5.10
5.00	4.70	5.50	5.30	4.60	5.90	5.50	5.40	6.70
6.80	7.40	6.00	5.80	5.50	6.40	6.20	5.70	6.40
6.70	6.30	7.80	7.60	8.60	6.60	6.50	6.00	7.60
7.00	6.00	7.10	7.40	7.20	8.40	8.50	9.40	7.10
7.00	6.60	8.00	10.45	8.81	10.61	9.97	10.69	12.40
13.38	14.31	10.90	9.98	9.20	10.94	10.53	9.06	10.17
11.17	10.84	12.09	13.66	14.06	11.14	11.10	10.00	11.98
11.74	10.27	12.05	12.27	12.03	13.95	15.10	15.65	12.47
12.29	11.52	13.08	12.50	11.05	12.94	13.24	13.16	14.95
16.00	16.98	13.15	12.88	11.99	13.13	12.99	11.69	13.78
13.70	13.57	15.12	15.55	16.73	12.68	12.65	11.18	13.27
12.64	11.01	13.30	12.19	12.91	14.90	16.10	17.30	12.90
13.36	12.26	13.93	13.94	12.75	14.19	14.67	14.66	16.21
17.72	18.15	14.19	14.33	12.99	15.19	15.09	12.94	15.46
15.39	15.34	17.02	18.85	19.49	15.61	16.16	14.84	17.04

سری T۱۰ : ۱۵۰ داده شبیه سازی شده از الگوی $_{4}$(۰،۱،۱)*(۰،۱،۱)ARIMA با $\beta = 0.8$ و $\gamma = 0.6$

97.323	100.169	110.752	119.851	121.799	125.721	136.152
145.765	152.012	157.871	169.188	181.318	183.128	189.861
203.344	213.626	216.468	222.034	233.512	245.057	245.504
251.052	99.386	104.475	108.949	120.605	127.540	130.557
134.049	145.491	156.293	159.769	167.274	179.960	191.285
191.439	197.990	211.808	223.600	225.794	230.428	241.983
251.616	253.507	100.581	105.712	111.713	115.752	126.888
133.554	136.034	143.401	156.306	166.352	170.048	177.068
188.676	200.030	202.407	209.169	219.971	231.908	233.996
237.094	250.315	260.501	95.671	109.055	113.635	119.380
122.581	134.013	141.636	145.873	151.636	163.872	175.631
179.141	185.391	196.391	210.073	211.793	216.864	228.497
240.955	242.367	244.670	102.601	103.344	115.535	112.423
124.873	130.221	140.264	151.227	156.866	162.214	174.562
183.945	187.751	196.406	207.403	219.515	221.199	225.202
237.421	249.358	249.887	101.751	109.132	111.439	122.320
130.316	133.251	137.761	149.660	161.653	165.239	172.450
183.056	194.459	198.099	203.713	216.688	227.439	229.665
234.174	245.153	256.824	103.672	110.472	115.798	118.610
129.496	137.640	140.608	148.268	158.831	170.421	174.456
181.098	192.726	204.684	208.789	213.734	224.872	236.350
236.984	240.533	254.345				

سری T۱۱ : ارقام مربوط به استخدام ماهانه مردان ۱٦ تا ۱۹ ساله (به هزار نفر) از ۱۹۷۱ تا ۱۹۸۱ (n=۱۳۲)

707	655	638	574	552	980	926	680	597	637	660
704	758	835	747	617	554	929	815	702	640	588
669	675	610	651	605	592	527	898	839	614	594
576	672	651	714	715	672	588	567	1057	949	683
771	708	824	835	980	969	931	892	828	1350	1218
977	863	838	866	877	1007	951	906	911	812	1172
1101	900	841	853	922	886	896	936	902	765	735
1234	1052	868	798	751	820	725	821	895	851	734
636	994	990	750	727	754	792	817	856	886	833
733	675	1004	956	777	761	709	777	771	840	847
774	720	848	1240	1168	936	853	910	953	874	1026
1030	946	860	856	1190	1038	883	843	857	1016	1003

سری T۱۲ : مقادیر فروش یک شرکت هر سه ماه یک بار برای ۸ سال متوالی (بر حسب میلیون ریال)
(n=۳۲)

36.14	44.60	44.15	35.72	36.19	44.63	46.95	36.90	39.66
49.72	44.49	36.54	41.44	49.07	48.98	39.59	44.29	50.09
48.42	41.39	46.11	53.44	53.00	42.52	44.61	55.18	52.24
41.66	47.84	54.27	52.31	42.03				

سری T۱۳ : سری مگس (n=۸۲)

1676	3075	3815	4639	4424	2784	5860	5781	4897	3920	3835
3618	3050	3772	3517	3350	3018	2625	2412	2221	2619	3203
2706	2717	2175	1628	2388	3677	3156	4272	3771	4955	5584
3891	3501	4436	4369	3394	3869	2922	1843	2837	4690	5119
5838	5389	4993	4446	4651	4243	4620	4849	3664	3016	2881
3821	4300	4168	5448	5477	8579	7533	6884	4127	5546	6316
6650	6304	4842	4352	3215	2652	2330	3123	3955	4494	4780
5753	5555	5712	4786	4066						

سری T۱٤ : سری لکه های خورشیدی والفر(n=۲۸٤)

5.00	11.00	16.00	23.00	36.00	58.00	29.00	20.00
10.00	8.00	3.00	0.00	0.00	2.00	11.00	27.00
47.00	63.00	60.00	39.00	28.00	26.00	22.00	11.00
21.00	40.00	78.00	122.00	103.00	73.00	47.00	35.00
11.00	5.00	16.00	34.00	70.00	81.00	111.00	101.00
73.00	40.00	20.00	16.00	5.00	11.00	22.00	40.00
60.00	80.90	83.40	47.70	47.80	30.70	12.20	9.60
10.20	32.40	47.60	54.00	62.90	85.90	61.20	45.10
36.40	20.90	11.40	37.80	69.80	106.10	100.80	81.60
66.50	34.80	30.60	7.00	19.80	92.50	154.40	125.90
54.80	68.10	38.50	22.80	10.20	24.10	82.90	132.00
130.90	118.10	89.90	66.60	60.00	46.90	41.00	21.30
16.00	6.40	4.10	6.80	14.50	34.00	45.00	43.10
47.50	42.20	28.10	10.10	8.10	2.50	0.00	1.40
5.00	12.20	13.90	35.40	45.80	41.10	30.10	23.90
15.60	6.60	4.00	1.80	8.50	16.60	36.30	49.60
64.20	67.00	70.90	47.80	27.50	8.50	13.20	56.90
121.50	138.30	103.20	85.70	64.60	36.70	24.20	10.70
15.00	40.10	61.50	98.50	124.70	96.30	66.60	64.50
54.10	39.00	20.60	6.70	4.30	22.70	54.80	93.80
95.80	77.20	59.10	44.00	47.00	30.50	16.30	7.30
37.60	74.00	139.00	111.20	101.60	66.20	44.70	17.00
11.30	12.40	3.40	6.00	32.30	54.30	59.70	63.70
63.50	52.20	25.40	13.10	6.80	6.30	7.10	35.60
73.00	85.10	78.00	64.00	41.80	26.20	26.70	12.10
9.50	2.70	5.00	24.40	42.00	63.50	53.80	62.00
48.50	43.90	18.60	5.70	3.60	1.40	9.60	47.40
57.10	103.90	80.60	63.60	37.60	26.10	14.20	5.80
16.70	44.30	63.90	69.00	77.80	64.90	35.70	21.20
11.10	5.70	8.70	36.10	79.70	114.40	109.60	88.80
67.80	47.50	30.60	16.30	9.60	33.20	92.60	151.60
136.30	134.70	83.90	69.40	31.50	13.90	4.40	38.00
141.70	190.20	184.80	159.00	112.30	53.90	37.50	27.90
10.20	15.10	47.00	93.80	105.90	105.50	104.50	66.60
68.90	38.00	34.50	15.50	12.60	27.50	92.50	155.40
32.27	54.25	59.65	63.62				

تمرینات

داده های این قسمت را بصورت ستونی از بالا به پائین بخوانید.

سری A : قیمتهای بورس عمومی IBM ؛ روزانه از ۱۷ مه ۱۹۶۱ تا ۲ نوامبر۱۹۶۲ (۳۶۹ مشاهده)

٣٦٦	٣٨٧	٣٨٧	٣٢٠	٤٥٧	٥٢٣	٥٤٧	٥٣٢	٥٧٥	٥٩٨	٥٥٧	٥٣١	٤٩٠	٤٧٧	٤٦٠
٣٥٩	٣٨٣	٣٨٧	٣٣٢	٤٤٩	٥١٦	٥٤٨	٥٢٥	٥٧٣	٥٩٥	٥٦٠	٥٤٧	٤٨٩	٤٧٩	٤٥٧
٣٥٦	٣٨٨	٣٧٦	٣٢٠	٤٥٠	٥١١	٥٤٩	٥٤٢	٥٧٧	٥٩٥	٥٧١	٥٥١	٤٨٩	٤٧٥	٤٥٢
٣٥٥	٣٨٢	٣٨٥	٣٣٣	٤٣٥	٥١٨	٥٥٣	٥٥٥	٥٨٢	٥٩٢	٥٧١	٥٤٧	٤٨٥	٤٧٩	٤٥٩
٣٦٧	٣٨٤	٣٨٥	٣٤٤	٤١٥	٥١٧	٥٥٣	٥٥٨	٥٨٤	٥٨٨	٥٦٩	٥٤١	٤٩١	٤٧٦	٤٦٢
٣٥٧	٣٨٢	٣٨٠	٣٣٩	٣٩٨	٥٢٠	٥٥٢	٥٥١	٥٧٩	٥٨٢	٥٧٥	٥٤٥	٤٩٢	٤٧٦	٤٥٩
٣٦١	٣٨٣	٣٧٣	٣٥٠	٣٩٩	٥١٩	٥٥١	٥٥١	٥٧٢	٥٧٦	٥٨٠	٥٤٩	٤٩٤	٤٧٨	٤٦٣
٣٥٥	٣٨٣	٣٨٢	٣٥١	٣٦١	٥١٩	٥٥٠	٥٥٢	٥٧٧	٥٧٨	٥٨٤	٥٤٥	٤٩٩	٤٧٩	٤٧٩
٣٤٨	٣٨٨	٣٧٧	٣٥٠	٣٨٣	٥١٩	٥٥٣	٥٥٣	٥٧١	٥٨٩	٥٨٥	٥٤٩	٤٩٨	٤٧٧	٤٩٣
٣٤٣	٣٩٥	٣٧٦	٣٥٤	٣٩٣	٥١٨	٥٥٤	٥٥٧	٥٦٠	٥٨٥	٥٩٠	٥٤٧	٥٠٠	٤٧٦	٤٩٠
٣٣٠	٣٩٢	٣٧٩	٣٥٠	٣٨٥	٥١٣	٥٥١	٥٥٧	٥٤٩	٥٨٠	٥٩٩	٥٤٣	٤٩٧	٤٧٥	٤٩٢
٣٤٠	٣٨٦	٣٨٦	٣٥٩	٣٦٠	٤٩٩	٥٥١	٥٤٨	٥٥٦	٥٧٩	٦٠٣	٥٤٠	٤٩٤	٤٧٥	٤٩٨
٣٣٩	٣٨٣	٣٨٧	٣٧٥	٣٦٤	٤٨٥	٥٤٥	٥٤٧	٥٥٧	٥٨٤	٥٩٩	٥٣٩	٤٩٥	٤٧٣	٤٩٩
٣٣١	٣٧٧	٣٨٦	٣٧٩	٣٦٥	٤٥٤	٥٤٧	٥٤٥	٥٦٣	٥٨١	٥٩٦	٥٣٢	٥٠٠	٤٧٤	٤٩٧
٣٤٥	٣٦٤	٣٨٩	٣٧٦	٣٧٠	٤٦٢	٥٤٧	٥٤٥	٥٦٤	٥٨١	٥٨٥	٥١٧	٥٠٤	٤٧٤	٤٩٦
٣٥٢	٣٦٩	٣٤٩	٣٨٢	٣٧٤	٤٧٣	٥٣٧	٥٣٩	٥٦٧	٥٧٧	٥٨٧	٥٢٧	٥١٣	٤٧٤	٤٩٠
٣٤٦	٣٥٥	٣٩٣	٣٧٠	٣٥٩	٤٨٢	٥٣٩	٥٣٩	٥٦١	٥٧٧	٥٨٥	٥٤٠	٥١١	٤٦٥	٤٨٩
٣٥٢	٣٥٠	٤٠٩	٣٦٥	٣٣٥	٤٨٦	٥٣٨	٥٣٥	٥٥٩	٥٧٨	٥٨١	٥٤٢	٥١٤	٤٦٦	٤٧٨
٣٥٧	٣٥٣	٤١١	٣٦٧	٣٢٣	٤٧٥	٥٣٣	٥٣٧	٥٥٣	٥٨٠	٥٨٣	٥٣٨	٥١٠	٤٦٧	٤٨٧
	٣٤٠	٤٠٩	٣٧٢	٣٠٦	٤٥٩	٥٢٥	٥٣٥	٥٥٣	٥٨٦	٥٩٢	٥٤١	٥٠٩	٤٧١	٤٩١
	٣٥٠	٤٠٨	٣٧٦	٣٣٣	٤٥١	٥١٣	٥٣٦	٥٥٣	٥٨٣	٥٩٢	٥٤١	٥١٥	٤٧١	٤٨٧
	٣٤٩	٣٩٣	٣٦٣	٣٣٠	٤٥٣	٥١٠	٥٣٧	٥٤٧	٥٨١	٥٩٦	٥٤٧	٥١٩	٤٦٧	٤٨٢
	٣٥٨	٣٩١	٣٧١	٣٣٦	٤٤٦	٥٢١	٥٤٣	٥٥٠	٥٧٦	٥٩٦	٥٥٣	٥٢٣	٤٧٣	٤٧٩
	٣٦٠	٣٨٨	٣٦٩	٣٢٨	٤٥٥	٥٢١	٥٤٨	٥٤٤	٥٧١	٥٩٥	٥٥٩	٥١٩	٤٨١	٤٧٨
	٣٦٠	٣٩٦	٣٧٦	٣١٦	٤٥٢	٥٢١	٥٤٦	٥٤١	٥٧٥	٥٩٨	٥٥٧	٥٢٣	٤٨٨	٤٧٩

سری B : اندازه گیری درجه حرارت یک فرآیند شیمیایی هر دقیقه یک بار (۲۲٦ مشاهده)

۱۸.۸	۲۳.٤	۲٤.٥	۲۱.۷	۲۱.۳	۲٤	۲٦	۱۹.۷	۲۲.٤	۲٦.٦
	۲۳.۳	۲٤.٥	۲۱.۸	۲۱.۲	۲۳.۹	۲٥.۸	۱۹.۹	۲۲.۲	۲۷
	۲۳.۳	۲٤.٥	۲۱.۹	۲۱.۲	۲۳.۷	۲٥.٦	۲۰	۲۲	۲۷.۱
	۲۳.۳	۲٤.٥	۲۲.۲	۲۱.۱	۲۳.٦	۲٥.٤	۲۰.۱	۲۱.۸	۲۷.۱
	۲۳.٤	۲٤.٤	۲۲.٥	۲۱	۲۳.٥	۲٥.۲	۲۰.۲	۲۱.٤	۲۷.۱
	۲۳.٤	۲٤.٤	۲۲.۸	۲۰.۹	۲۳.٥	۲٤.۹	۲۰.۳	۲۰.۹	۲۷.۱
	۲۳.۳	۲٤.۲	۲۳.۱	۲۱	۲۳.٥	۲٤.۷	۲۰.۳	۲۰.۳	۲٦.۹
	۲۳.۲	۲٤.۲	۲۳.٤	۲۱	۲۳.٥	۲٤.٥	۲۱.٦	۱۹.۷	۲٦.۸
	۲۳.۳	۲٤.۱	۲۳.۸	۲۱	۲۳.٥	۲٤.٤	۲۱.۹	۱۹.٤	۲٦.۷
	۲۳.۳	۲٤.۱	۲٤.۱	۲۱.۲	۲۳.۷	۲٤.٤	۲۱.۷	۱۹.۳	۲٦.٤
	۲۳.۲	۲٤	۲٤.٦	۲۱.۱	۲۳.۸	۲٤.٤	۲۱.۳	۱۹.۲	۲٦
	۲۳.۱	۲٤	۲٤.۹	۲۰.۹	۲۳.۸	۲٤.٤	۲۱.۲	۱۹.۱	۲٥.۸
	۲۲.۹	۲٤	۲٤.۹	۲۰.۸	۲۳.۹	۲٤.٤	۲۱.٤	۱۹	۲٥.٦
	۲۲.۸	۲۳.۹	۲٥.۱	۲۰.۸	۲۳.۹	۲٤.۳	۲۱.۷	۱۸.۹	۲٥.۲
	۲۲.٦	۲۳.۸	۲٥	۲۰.۸	۲۳.۸	۲٤.٤	۲۲.۲	۱۸.۹	۲٥
	۲۲.٤	۲۳.۸	۲٥	۲۰.۸	۲۳.۷	۲٤.٤	۲۳	۱۹.۲	۲٤.٦
	۲۲.۲	۲۳.۷	۲٥	۲۰.۹	۲۳.٦	۲٤.٤	۲۳.۸	۱۹.۳	۲٤.۲
	۲۱.۸	۲۳.۷	۲٥	۲۰.۸	۲۳.٤	۲٤.٤	۲٤.٦	۱۹.۳	۲٤
	۲۱.۳	۲۳.٦	۲٥	۲۰.۸	۲۳.۲	۲٤.٤	۲٥.۱	۱۹.٤	۲۳.۷
	۲۰.۸	۲۳.۷	۲٤.۸	۲۰.۷	۲۳	۲٤.٥	۲٥.٦	۱۹.٥	۲۳.٤
	۲۰.۲	۲۳.٦	۲٤.۷	۲۰.۷	۲۲.۸	۲٤.٥	۲٥.۸	۱۹.٦	۲۳.۱
	۱۹.۷	۲۳.٦	۲٤.٦	۲۰.۸	۲۲.٦	۲٤.٤	۲٦.۱	۱۹.٦	۲۲.۹
	۱۹.۳	۲۳.٦	۲٤.٥	۲۰.۹	۲۲.٤	۲٤.۳	۲٦.۳	۱۹.٦	۲۲.۸
	۱۹.۱	۲۳.٥	۲٤.٥	۲۱.۲	۲۲	۲٤.۲	۲٦.۳	۱۹.٦	۲۲.۷
	۱۹	۲۳.٥	۲٤.٥	۲۱.٤	۲۱.٦	۲٤.۲	۲٦.۲	۱۹.٦	۲۲.٦

سری C : اندازه گیری غلظت یک فرآیند شیمیایی هر ساعت یک بار (۳۱۰ مشاهده)

٨.٨	٨	٩.٨	١٠	٩.٦	٩.٦	٩.٤	٩.٥	٩	٩.٥	٨.٤	٨.٩	٨
٨.٧	٨	٩.٨	١٠	٩.٨	٩.٦	٩.٤	٩.٣	٨.٨	٩.٥	٨.٣	٩.١	٨
٨.٦	٨.٤	٩.٦	٩.٨	١٠.٢	٩.٢	٩	٩.٥	٨.٦	٩.٩	٨.٣	٩.٥	٧.٤
٨.٣	٨.٨	٩.٢	٩.٨	١٠	٩.٢	٩.٤	٩.٥	٨.٦	٩.٥	٨.١	٨.٥	٨
٧.٩	٨.٤	٩.٦	٩.٧	١٠	٩	٩.٤	٩.١	٨	٩.٧	٨.٢	٨.٤	٨
٨.٥	٨.٤	٩.٢	٩.٦	١٠	٩	٩.٦	٩.٣	٨	٩.١	٨.٣	٨.٣	٨
٨.٧	٩	٩.٢	٩.٤	٩.٤	٩	٩.٤	٩.٥	٨	٩.١	٨.٥	٨.٢	٨
٨.٩	٩	٩.٦	٩.٢	٩.٢	٩.٤	٩.٢	٩.٣	٨	٨.٩	٨.١	٨.١	٨.٨
٩.١	٩.٤	٩.٦	٩	٩.٦	٩	٨.٨	٩.١	٨.٦	٩.٣	٨.١	٨.٣	٨.٤
٩.١	١٠	٩.٦	٩.٤	٩.٧	٩	٨.٨	٩.٣	٨	٩.١	٧.٩	٨.٤	٨.٤
	١٠	٩.٦	٩.٦	٩.٧	٩.٤	٩.٢	٩.١	٨	٩.١	٨.٣	٨.٧	٨
	١٠	٩.٦	٩.٦	٩.٨	٩.٤	٩.٢	٩.٥	٨	٩.٣	٨.١	٨.٨	٨.٢
	١٠.٢	٩.٦	٩.٦	٩.٨	٩.٦	٩.٦	٩.٤	٧.٦	٩.٥	٨.١	٨.٨	٨.٢
	١٠	١٠	٩.٦	٩.٨	٩.٤	٩.٦	٩.٥	٨.٦	٩.٣	٨.١	٩.٢	٨.٢
	١٠	١٠	٩.٦	١٠	٩.٦	٩.٨	٩.٦	٩.٦	٩.٣	٨.٤	٩.٦	٨.٤
	٩.٦	١٠.٤	٩.٦	١٠	٩.٦	٩.٨	١٠.٢	٩.٦	٩.٣	٨.٧	٩	٨.٤
	٩	١٠.٤	٩	٨.٦	٩.٦	١٠	٩.٨	١٠	٩.٩	٩	٨.٨	٨.٤
	٩	٩.٨	٩.٤	٨.٦	١٠	١٠	٩.٦	٩.٤	٩.٧	٩.٣	٨.٦	٨.٦
	٨.٦	٩	٩.٤	٩	١٠	٩.٤	٩.٦	٩.٣	٩.١	٩.٣	٨.٦	٨.٨
	٩	٩.٦	٩.٤	٩.٤	٩.٦	٩.٨	٩.٤	٩.٢	٩.٣	٩.٥	٨.٨	٨.٦
	٩.٦	٩.٨	٩.٦	٩.٤	٩.٢	٨.٨	٩.٤	٩.٥	٩.٥	٩.٣	٨.٨	٨.٦
	٩.٦	٩.٦	٩.٤	٩.٤	٩.٢	٨.٨	٩.٤	٩.٥	٩.٤	٩.٥	٨.٦	٨.٦
	٩	٩.٦	٩.٦	٩.٤	٩.٢	٨.٨	٩.٤	٩.٥	٩	٩.٥	٨.٦	٨.٦
	٩	٨	٩.٦	٩.٤	٩	٨.٨	٩.٦	٩.٩	٩	٩.٥	٨.٤	٨.٦
	٨..٩	٨	٩.٨	٩.٦	٩	٩.٦	٩.٦	٩.٩	٨.٨	٩.٥	٨.٣	٨.٨

سری D : عملکرد یک فرآیند شیمیایی پخت (۷۰ مشاهده)

٦٠	٣٤	٥٢	٤٥	٥٨	٢٥	٥٧	٥٨	٤١	٤٧
٣٩	٣٥	٣٨	٥٧	٤٥	٥٩	٥٠	٤٤	٥٩	٦٤
٥٩	٥٤	٥٩	٥٠	٥٤	٥٠	٦٠	٨٠	٤٨	٢٣
٤٠	٤٥	٥٥	٦٢	٣٦	٧١	٤٥	٥٥	٧١	٧١
٥٧	٦٨	٤١	٤٤	٥٤	٥٦	٥٧	٣٧	٣٥	٣٨
٥٤	٣٨	٥٣	٦٤	٤٨	٧٤	٥٠	٧٤	٥٧	٦٤
٢٣	٥٠	٤٩	٤٣	٥٥	٥٠	٤٥	٥١	٤٠	٥٥

منابع و مواخذ

۱. **تحلیل سری های زمانی روشهای یک متغیری و چند متغیری**

تالیف : ویلیام دبلیو.اس. وی

ترجمه : دکتر حسینعلی نیرومند

۲. **تجزیه و تحلیل سری های زمانی و پیش بینی**

تالیف : جی . ای . پی . باکس و جی . ام . جنکینس

ترجمه : دکتر محمد رضا مشکانی

۳. **پیش بینی و تجزیه و تحلیل سری های زمانی**

تالیف : مونتگمری، جانسون و گاردینر

ترجمه: دکتر محمد تقی فاطمی قمی

٤. **تجزیه و تحلیل سری های زمانی**

تالیف : جاناتان کرایر

ترجمه: دکتر حسینعلی نیرومند

۵. **پیش بینی سری های زمانی**

تالیف : باورمن اکانل

ترجمه: دکتر رضا شیوا

٦. **مقدمه ای بر تحلیل سری های زمانی**

تالیف : سی- چتفیلد

ترجمه: دکتر حسینعلی نیرومند - دکتر ابوالقاسم بزرگ نیا

۷. **آمار کاربردی**
تالیف : ژیل گرینون و سوزان ویو

ترجمه: دکتر حمزه گنجی و مهدی گنجی

۸. **سری های زمانی**
تالیف : دکتر حسینعلی نیرومند و دکتر ابوالقاسم بزرگ نیا - انتشارات دانشگاه پیام نور

۹. **مقدمه ای بر سری های زمانی و پیش بینی**
تالیف : پتر.جی.براکول - ریچارد.آ. دیویس

ترجمه : دکتر محمد امینی _ دکتر ابوالقاسم بزرگ نیا _ دکتر محمد حسین دهقان

۱۰. پایان نامه کارشناسی ارشد با عنوان" **آزمونهای ریشه واحد در مدلهای ARIMA از مرتبه نامعلوم**"
مولف : مهرناز نصر اصفهانی _ دانشگاه فردوسی مشهد

۱۱. پایان نامه کارشناسی ارشد با عنوان" **برآورد داده های گمشده در سری های زمانی و مقایسه روشهای مختلف** "
مولف : امید خزین قناد _ دانشگاه فردوسی مشهد

۱۲. پایان نامه کارشناسی ارشد با عنوان" **سری های زمانی غیر خطی و الگوی SDM** "
مولف : محمد حسین کربلایی _ دانشگاه فردوسی مشهد

۱۳. پایان نامه کارشناسی ارشد با عنوان" **سری های زمانی دو خطی**"
مولف : سید محمد امیر جهانشاهی _ دانشگاه فردوسی مشهد

ناشر الکترونیکی این کتاب:

شرکت داده پردازی آماری اطمینان شرق- ارائه دهنده خدمات تخصصی تحلیل آماری در کشور

آدرس وب سایت : www.spss-iran.ir

تلفن دریافت سفارشات تحلیل و خدمات آماری: ثابت ۰۵۱۳۷٤۱۰۷۳۹ - همراه ۰۹۱۹۸۱۸۰۹۹۱

ایمیل برای دریافت سفارش خدمات تجزیه و تحلیل آماری با نرم افزار:

mojtaba.farshchi@gmail.com