به نام آن‌که جان را فکرت آموخت

دانشگاه صنعت آب و برق (شهید عباسپور)

مباحث ویژه در

(ویراست نوشتار ۱٫۰٫۰)

سیدسعید موسوی‌ندوشنی

s_mousavi@pwut.ac.ir

پاییز ۱۳۹۱

فهرست مطالب

پیش‌گفتار

نوشتار عمومی برای نرم‌افزار R به منظور آشنایی کسانی که در این زمینه مبتدی و در ابتدای راه هستند ارائه شد و در سایت R به آدرس زیر قرار گرفت که به‌صورت رایگان قابل پیاده‌سازی است.

http://cran.r-project.org/doc/contrib/Mousavi-R-lang_in_Farsi.pdf

به نظر می‌رسد که نوشته گفته شده مورد استفاده علاقه‌مندان قرار گرفته است و عده‌ای هم حقیر را از طریق پست الکترونیک مورد لطف و محبت خویش قرار دادند که از همین جا از آنها سپاسگزاری می‌شود.
بنده در ادامه راه به فکر تهیه پاره‌ای از مباحث خاص به مدد R افتادم. بنابراین نوشتار حاضر زیر عنوان «مباحث ویژه در R» (مانند سری‌های زمانی، مقادیر حدی، زمین‌آمار و . . .) در حد بضاعت اندک بنده مطرح می‌گردد.
اولین مبحث آن اختصاص به سری‌های زمانی دارد که شرح آن در چند فصل تقدیم می‌شود و این مباحث به مرور ادامه می‌یابد و ویرایش می‌گردد تا نوشته منقح‌تری حاصل شود. در اینجا خاطر نشان می‌شود که فرض بر این است که خواننده گرامی برای تعقیب مباحث ویژه، آشنایی اولیه با نرم‌افزار R را کسب نموده است.
مسلماً وجیزه حاضر خالی از خلل نیست، نویسنده از هر گونه اظهار نظر و پیشنهادی برای اصلاح و ارتقاء آن استقبال و استفاده خواهد نمود.

سیدسعید موسوی‌ندوشنی
تهران - پاییز ۱۳۹۱

فصل اول

سری‌های زمانی

۱-۱ سری زمانی

سری‌های زمانی عبارتند از ثبت فرآیندهایی است که در زمان تغییر می‌کنند. عمل ثبت مشاهدات می‌تواند به صورت گسسته و یا پیوسته انجام پذیرد. در اینجا روی ثبت گسسته با فواصل زمانی مساوی تاکید می‌گردد. با انتخاب مبداء زمانی و مقیاس زمانی مناسب، مجموعه زمانی به صورت $1,2,\cdots,n$ و مجموعه مشاهدات به شکل $\{x_t : t = 1, 2, \cdots , n\}$ و یا $x_1, x_2, \cdots , x_n$ نشان داده می‌شود.

۱-۱-۱ ترسیم سری زمانی، روند و تغییرات فصلی

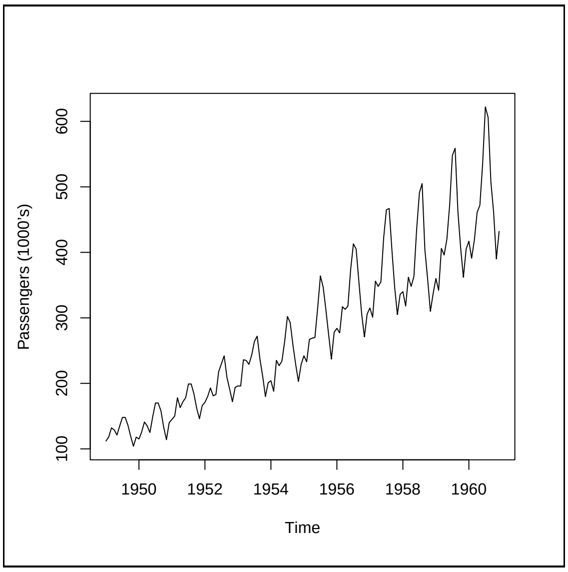
۱-۱-۱-۱ ترسیم سری‌ها

برای ترسیم و . . . از فایل داده‌هایی در خود R زیر عنوان AirPassengers استفاده می‌شود که تعداد مسافرین بین‌المللی ماهانه (بر حسب هزار نفر) را نشان می‌دهد. این آمار در دوره 1960-1949 ثبت شده است.
مثال:

```
> data(AirPassengers)
```

```
> AP <- AirPassengers
> plot(AP, ylab = "Passengers (1000's)")
```

شکل ۱-۱ حاصل اجرای کُدهای بالا است. تابع دیگری در زبان R به نام window() وجود دارد که می‌تواند



شکل ۱-۱: نمایش سری زمانی ماهانه مسافرین

یک زیر مجموعه‌ای از سری زمانی را به‌دست دهد. به مثال زیر توجه کنید.

```
> x <- AirPassengers
> y <- window(x, 1950, 1951)
```

که نتیجه آن به صورت زیر است.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1950	115	126	141	135	125	149	170	170	158	133	114	140
1951	145											

اگر بخواهید روی فصلی از سال کنترل داشته باشید، مثلاً ماه به مثال زیر توجه کنید.

```
> x <- AirPassengers
> y <- window(x, 1950, c(1951,0))
```

که نتیجه آن به صورت زیر است.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1950	115	126	141	135	125	149	170	170	158	133	114	140

یا به مثال زیر توجه کنید.

```
> x <- AirPassengers
> y <- window(x, c(1950,2), c(1951,3))
```

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1950		126	141	135	125	149	170	170	158	133	114	140
1951	145	150	178									

اکنون به سری زمانی ماهانه‌ی زیر توجه کنید که تاریخ ندارد.

	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12
1	3.59	2.20	4.53	7.01	10.45	43.86	59.97	92.29	47.00	19.16	12.45	8.84
2	9.60	19.12	21.93	24.20	25.22	42.28	37.32	14.11	6.75	4.40	2.62	2.82
3	3.13	3.97	17.52	11.77	9.65	18.26	72.26	32.45	12.09	5.12	3.03	2.46
4	2.76	4.17	6.71	10.57	9.87	7.58	15.45	29.85	4.25	1.74	1.12	1.04
5	1.22	2.81	6.25	9.07	28.12	20.00	36.48	45.32	12.19	2.98	2.85	2.19
		⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

برای ترسیم داده‌های ماهانه ابتدا باید سال‌های آماری را پشت سر هم قرار داده و سپس آن‌ها را رسم نمود. به دو صورت می‌توان این کار را انجام داد. کُدهای صورت اول به شرح زیر است.

```
> x <- matrix(scan("F:/R_files/data/ghar.txt"), ncol=12, byrow=T)
> y <- t(x)
> y <- matrix(y,ncol=1, byrow=T)
> ts.plot(y)
```

کُدهای صورت دوم به شرح زیر است.

```
> x <- read.table("F:/R_files/data/ghar.txt")
> y <- as.matrix(y)
> y <- t(x)
> y <- matrix(y, ncol=1, byrow=T)
> ts.plot(y)
```

۱-۱-۱-۲ چند سری زمانی

می‌توان چند سری زمانی را به‌صورت توأم نشان داد. مثلاً در استرالیا از سال 1958 تا سال 1990 فایل تولید برق (بر حسب کیلووات ساعت)، ماءالشعیر (بر حسب میلیون لیتر) و شکلات (بر حسب تُن) وجود دارد. این سه سری به‌صورت online توسط کُد R خوانده می‌شود.

```
> "http://www.massey.ac.nz/~pscowper/ts/cbe.dat"
> CBE <- read.table(www, header = T)
CBE[1:4, ]
```

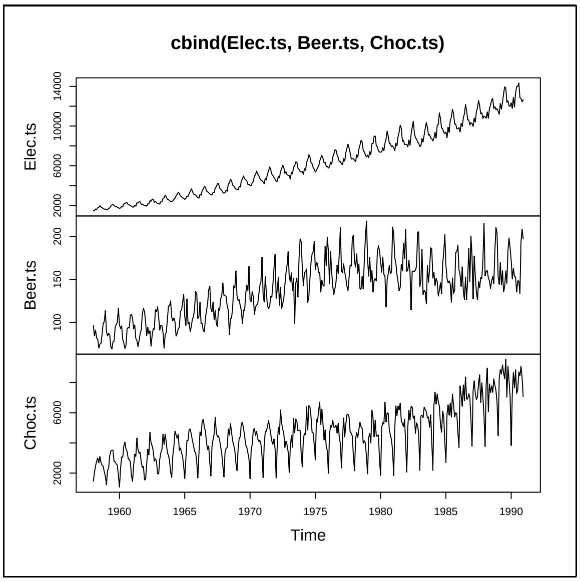
خروجی کُدهای بالا به‌صورت زیر است.

```
 choc beer elec
1 1451 96.3 1497
2 2037 84.4 1463
3 2477 91.2 1648
4 2785 81.9 1595
```

همان‌طور که ملاحظه می‌شود داده‌های بالا فاقد سال و ماه هستند. به عبارت دیگر تاریخ ندارند و دارای ساختار سری زمانی نیستند. بنابراین باید این اشیاء را با استفاده از تابع ts() به سری زمانی تبدیل نمود. برای این کار به کُدهای زیر توجه کنید.

```
> "http://www.massey.ac.nz/~pscowper/ts/cbe.dat"
> CBE <- read.table(www, header = T)
> Elec.ts <- ts(CBE[, 3], start = 1958, freq = 12)
> Beer.ts <- ts(CBE[, 2], start = 1958, freq = 12)
> Choc.ts <- ts(CBE[, 1], start = 1958, freq = 12)
> plot(cbind(Elec.ts, Beer.ts, Choc.ts))
```

نمودار ۱-۲ حاصل اجرای کُدهای بالا است.



شکل ۱-۲: نمایش چند سری زمانی

1. trend 2. seasonal variation

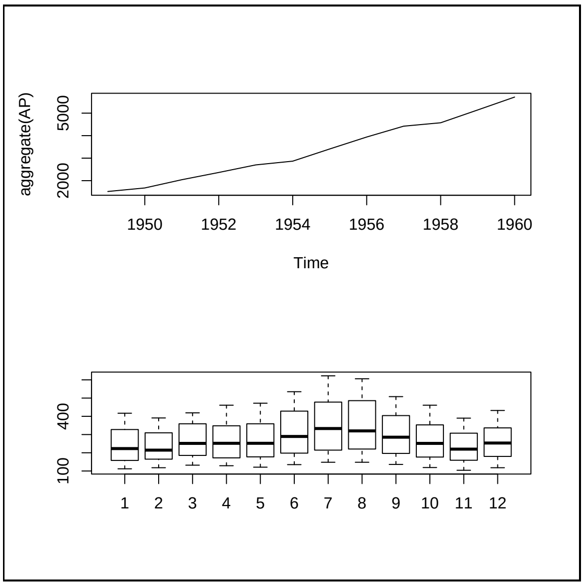
۱-۱-۱-۳ روند و تغییرات فصلی

رسم سری زمانی نه تنها الگوی سری را به‌دست می‌دهد بلکه روند[۱]، تغییرات فصلی[۲]، مقادیر پرت[۳] و احیاناً مقادیر خطا را نیز ظاهر می‌سازد، در مثال بالا تکرار الگو در هر سال تغییرات فصلی است و در حالت کلی، تغییرات سیستماتیک در سری زمانی که به‌صورت تناوبی ظاهر نمی‌شوند، روند گویند. ساده‌ترین مدل برای تبین روند، افزایش و کاهش خطی است. برای به‌دست آوردن واضح روند، می‌توان تأثیر تغییرات فصلی با تجمیع آمار ماهانه به آمار سالانه از بین برد. این کار در R با استفاده از تابع aggregate() انجام می‌شود. خلاصه مقادیر هر فصل توسط تابع boxplot() مشخص می‌شود و تابع cycle() فصل را برای هر یک از داده‌ها معین می‌سازد.

```
> data(AirPassengers)
> AP <- AirPassengers
> layout(1:2)
> plot(aggregate(AP))
> boxplot(AP ~ cycle(AP))
```

نمودار **۱-۳** حاصل اجرای کُدهای بالا است.

مثال: داده‌های بیکاری ماهانه به‌صورت online توسط کُد R خوانده می‌شود. همان‌طور که ملاحظه می‌گردد



شکل ۱-۳: نمایش سری سالانه تجمیع شده و boxplot آنها

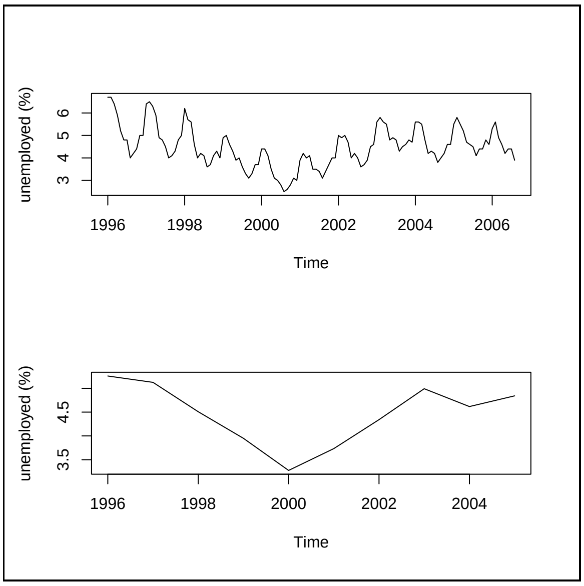
نام متغیر unemploy است. این داده‌ها سری زمانی نیستند و برای تبدیل آنها به سری زمانی از تابع ts() استفاده می‌شود، که پارامتر شروع و دوره تناوب freq در آن ملاحظه می‌گردد. متوسط سالانه توسط تابع تجمیع و عملگر تقسیم معین می‌شود.

```
> www <- "http://www.massey.ac.nz/~pscowper/ts/Maine.dat"
> Maine.month <- read.table(www, header = TRUE)
```

3. outliers

```
> Maine.month.ts <- ts(unemploy, start = c(1996, 1), freq = 12)
> Maine.annual.ts <- aggregate(Maine.month.ts)/12
> layout(1:2)
> plot(Maine.month.ts, ylab = "unemployed (%)")
> plot(Maine.annual.ts, ylab = "unemployed (%)")
```

نمودار ۱-۴ حاصل اجرای کُدهای بالا است.



شکل ۱-۴: نمایش سری بیکاری و متوسط آنها

۱-۲ تجزیه سری‌های زمانی

۱-۲-۱ مدل‌ها

همان‌طورکه قبلاً ملاحظه شد، در خیلی از سری‌ها روند و یا تغییرات فصلی مشاهده می‌شود. بنابراین مدل‌ها دارای مؤلفه‌هایی هستند. یک تجزیه ساده سری زمانی، مدل جمعی[۴] است، که به‌صورت زیر می‌باشد.

$$x_t = m_t + s_t + z_t$$

که در آن t زمان و x_t سری مشاهده شده است. مؤلفه m_t نشانه روند و s_t نماینده تغییرات فصلی است. مؤلفه z_t نیز خطا را نشان می‌دهد.

اگر تغییرات فصلی به‌صورت یک روند ظاهر شود، مدل مناسب یک مدل ضربی[۵] است، که به‌صورت زیر نشان داده می‌شود.

$$x_t = m'_t . s'_t . z'_t$$

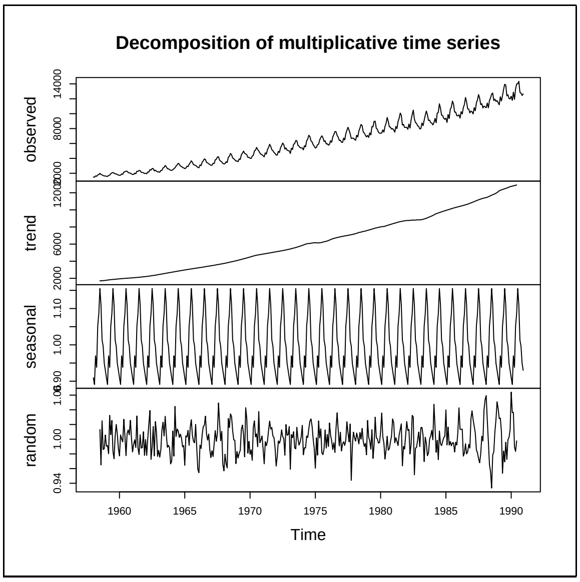
4. additive 5. multiplicative

همان‌طور که ملاحظه می‌شود با گرفتن لگاریتم از رابطه اخیر، مدل ضربی به مدل جمعی تبدیل می‌گردد.

$$y_t = \ln(x_t) = \ln(m'_t) + \ln(s'_t) + \ln(z'_t)$$
$$= m_t + s_t + z_t$$

۱-۲-۲ تجزیه مؤلفه‌ها در R

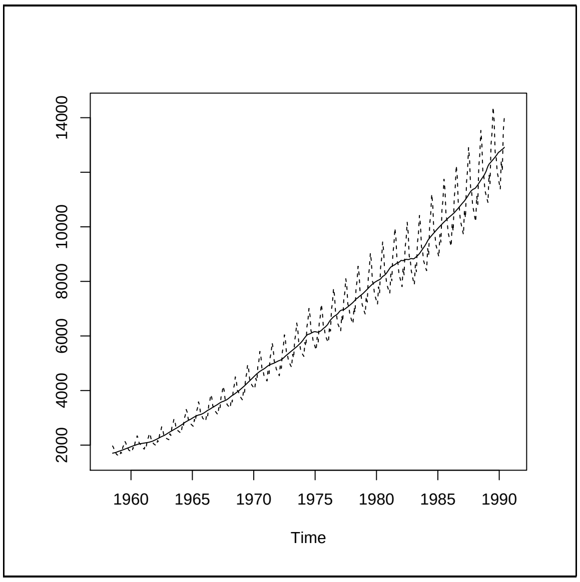
در R، تابع decompose() روند و تغییرات فصلی را با استفاده از روش میانگین متحرک برآورد می‌کند. اگر تابع فوق داخل تابع plot() قرار گیرد، آنگاه یک شکل با سه نمودار ایجاد می‌گردد. برای مثال داده مربوط به تولید برق (مثال بالا)، درکُدهای زیر به‌صورت مدل جمعی و ضربی (شکل ۱-۵) ترسیم می‌شود و در انتهای کُدها مدل ضربی دنبال می‌شود و بالاخره روی یک نمودار روند و تغییرات فصلی یکجا ملاحظه می‌گردد (شکل ۱-۶).



شکل ۱-۵: نمایش مؤلفه‌های سری

۱-۳ همبستگی

در خیلی از حالات، متغیرهای متوالی همبسته خواهند بود. برای تجزیه و تحلیل آنها باید چنین همبستگی معین گردد. ساختار همبستگی در سری‌های زمانی توسط تابع همبستگی تعریف می‌شود و از طریق داده‌های مشاهده شده برآورد می‌گردد.



شکل ۱-۶: نمایش روند و تغییرات فصلی

۱-۳-۱ تعریف میانگین و واریانس

امید ریاضی که با حرف E نشان داده می‌شود، در واقع میانگین جامعه است. بنابراین $E(x)$ میانگین متغیر x است که با μ نیز نمایش داده می‌شود. عبارت $E[(x-\mu)^2]$ میانگین مربع انحرافات حول μ است که به واریانس متغیر x شناخته می‌شود که با σ^2 نیز نمایش داده می‌شود.

۱-۳-۲ تعریف کوواریانس و همبستگی

اگر دو متغیر (x, y) داشته باشید، مفهومی به نام کوواریانس به‌صورت زیر قابل تعریف است.

$$\gamma(x, y) = E[(x-\mu_x)(y-\mu_y)]$$

کوواریانس در واقع میزان ارتباط خطی بین دو متغیر (x, y) را اندازه می‌گیرد. اگر یک نمونه‌ای به اندازه n از (x_i, y_i) داشته باشید می‌توان کوواریانس نمونه را برآورد نمود که رابطه آن به‌صورت زیر است.

$$\mathrm{Cov}(x, y) = \sum (x_i - \bar{x})(y_i - \bar{y})/n$$

ضریب همبستگی میزان ارتباط بین دو متغیر (x, y) را به‌صورت بدون بُعد معین می‌کند. این ضریب بین 1- و 1+ قرار دارد. اگر این ضریب صفر باشد بدین معنی است که ارتباط خطی بین متغیرهای (x, y) وجود ندارد. همبستگی جامعه یعنی ρ بین متغیرهای (x, y) به‌صورت زیر تعریف می‌شود.

$$\rho(x, y) = \frac{E[(x-\mu_x)(y-\mu_y)]}{\sigma_x \sigma_y} = \frac{\gamma(x, y)}{\sigma_x \sigma_y}$$

ضریب همبستگی نمونه به‌صورت زیر محاسبه می‌گردد.

$$\mathrm{Cor}(x,y)=\frac{\mathrm{Cov}(x,y)}{\mathrm{sd}(x)\,\mathrm{sd}(y)}$$

۱-۳-۳ تعریف توابع اتوکواریانس و خودهمبستگی

برای تعریف همبستگی سری‌های زمانی ابتدا به میانگین پرداخته می‌شود. تابع میانگین یک سری زمانی به‌صورت زیر است و تابعی از t می‌باشد.

$$\mu(t)=E(x_t)$$

اگر تابع میانگین ثابت باشد، یعنی تابعی از زمان نباشد، آنگاه سری زمانی در میانگین ایستا[6] است و برآورد نمونه‌ای آن به‌صورت زیر است.

$$\bar{x}=\sum_{i=1}^{n} x_t/n$$

تابع واریانس یک سری زمانی که در میانگین ایستا است، به‌صورت زیر می‌باشد.

$$\sigma^2(t)=E[(x_t-\mu)^2]$$

اگر سری زمانی در واریانس نیز ایستا باشد، آنگاه $\sigma^2(t)$ تبدیل به ثابت σ^2 می‌گردد. برآورد نمونه‌ای آن به‌صورت زیر است.

$$\mathrm{Var}(x)=\frac{\sum(x_t-\bar{x})^2}{n-1}$$

تاکنون سری زمانی را ملاحظه نمودید که در میانگین و واریانس ایستا بودند. اما متغیرها ممکن است همبسته نیز باشند. سری زمانی ایستای مرتبه دوم است که همبستگی بین متغیرها فقط به تفاوت گام‌های آنها بستگی داشته باشد. تعداد گام‌های بین متغیرها را تأخیر[7] گویند.

همبستگی یک متغیر با خودش در زمان‌های مختلف را خودهمبستگی[8] و یا همبستگی پیاپی[9] نامند. اگر سری زمانی ایستای مرتبه دوم باشد، آنگاه می‌توان تابع اتوکوواریانس (acvf) یعنی γ_k که تابعی از k است را به‌صورت زیر بیان نمود.

$$\gamma_k=E[(x_t-\mu)(x_{t+k}-\mu)]$$

تابع γ_k بستگی به زمان ندارد زیرا میانگین وابسته به زمان نیست، یعنی $E(x_t)=E(x_{t+k})=\mu$ است. تابع خودهمبستگی (acf) با تأخیر k به‌صورت زیر تعریف می‌شود.

$$\rho_k=\frac{\gamma_k}{\sigma^2}$$

با استفاده از تعریف بالا می‌توان نتیجه گرفت که $\rho_0=1$ است، زیرا $\gamma_0=\sigma^2$ می‌باشد.

6. stationary 7. lag 8. Autocorrelation 9. Serial correlation

۱-۳-۳-۱ محاسبه خودهمبستگی

اکنون فرض کنید n مشاهده $x_1, x_2, \cdots, x_n$ یک سری زمانی گسسته باشد. می‌توان $n-1$ جفت مشاهده یعنی

$$(x_1, x_2), (x_2, x_3), \cdots, (x_{n-1}, x_n)$$

را ساخت. اگر اولین مشاهده را در هر زوج به عنوان متغیر اول و دومین مشاهده را به عنوان متغیر دوم فرض کنید، ضریب همبستگی به صورت زیر است.

$$r_1 = \frac{\sum_{t=1}^{n-1} \left(x_t - \bar{x}_{(1)}\right)\left(x_{t+1} - \bar{x}_{(2)}\right)}{\sqrt{\sum_{t=1}^{n-1} \left(x_t - \bar{x}_{(1)}\right)^2 \sum_{t=1}^{n-1} \left(x_{t+1} - \bar{x}_{(2)}\right)^2}}$$

که در آن $\bar{x}_1 = \frac{1}{n-1}\sum_{t=1}^{n-1} x_t$ و $\bar{x}_2 = \frac{1}{n-1}\sum_{t=2}^{n} x_t$ است و آن را ضریب خودهمبستگی بین مشاهدات نامند. که در واقع همبستگی بین متغیر x در زمان‌های مختلف است.
اگر n به اندازه کافی بزرگ باشد r_1 از فرمول زیر استفاده می‌شود.

$$r_1 = \frac{\sum_{t=1}^{n-1} (x_t - \bar{x})(x_{t+1} - \bar{x})}{\sum_{t=1}^{n} (x_t - \bar{x})^2}$$

که در آن $\bar{x} = \frac{1}{n}\sum_{t=1}^{n} x_t$ است. به همین طریق می‌توان ضریب خودهمبستگی را با تأخیر k محاسبه نمود.

$$r_k = \frac{\sum_{t=1}^{n-k} (x_t - \bar{x})(x_{t+k} - \bar{x})}{\sum_{t=1}^{n} (x_t - \bar{x})^2} \qquad k = 0, 1, 2, \cdots$$

معمولاً برای مقادیر $k > \frac{n}{4}$ نباید r_kها را محاسبه کرد. (زیرا در این صورت خیلی از اطلاعات سری زمانی را از دست می‌دهید.)
برای برآورد فاصله‌ای r_k با سطح معنی‌دار بودن 5% مقدار $\pm\frac{2}{\sqrt{n}}$ در نظر گرفته می‌شود.

۲-۳-۳-۱ همبستگی‌نگار[۱۰]

برای تعبیر و تفسیر سری‌های زمانی، از نمودار r_k (محور عرض‌ها) برحسب k (محور طول‌ها) استفاده می‌شود. این نمودار را همبستگی‌نگار نامند. محور عرض‌ها که مقادیر خودهمبستگی روی قرار دارد بدون بُعد است. محور

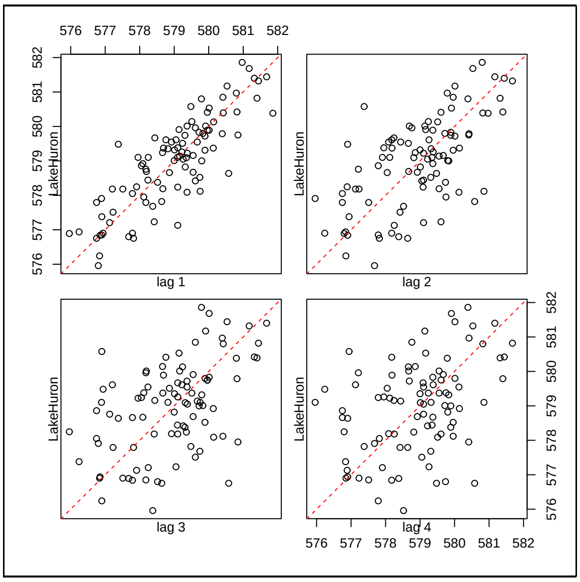
10. correlogram

طول‌ها که تأخیرها را نشان می‌دهد بستگی به فاصله زمانی نمونه‌گیری دارد.

برای ترسیم تأخیرهای مختلف نسبت به هم از تابع lag.plot() در R استفاده می‌شود. به‌عنوان مثال

```
> lag.plot(LakeHuron, lag=4, labels=F, do.lines=F, diag.col = "red")
```

در مثال بالا، چون lag=4 است، داده‌های LakeHuron تا چهار تأخیر ترسیم می‌گردد (شکل ۱-۷). برای



شکل ۱-۷: نمودارهای تأخیر

محاسبه r_k و ترسیم همبستگی‌نگار در R از تابع acf() استفاده می‌شود. شکل کلی آن به‌صورت زیر است.

```
acf(x, lag.max=NULL, type=c("correlation","covariance","partial"), plot=TRUE)
```

که در آن

x: بردار داده‌ها است.

lag.max: تعداد تأخیر مورد نظر است. اگر این آرگومان حذف گردد، R دارای پیش‌فرض است که از رابطه‌ی 10*log(N) محاسبه می‌گردد، که N اندازه نمونه‌ها و لگاریتم آن در مبنای 10 می‌باشد.

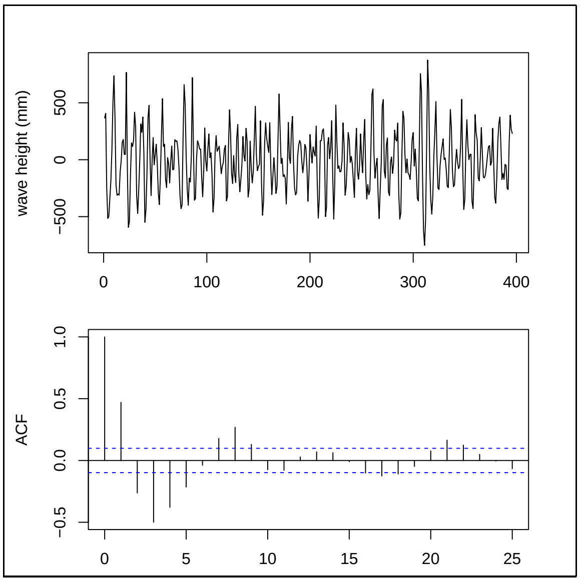
type: رشته‌ای از کاراکترها است و نوع تابع acf را معین می‌کند. اگر این آرگومان حذف گردد، "correlation" پیش‌فرض این تابع است.

plot: این آرگومان منطقی دارای پیش‌فرض TRUE است و در این صورت نمودار همبستگی‌نگار رسم می‌گردد. اگر این آرگومان FALSE باشد، نمودار گفته شده رسم نمی‌شود و مقادیر acf به ازای تأخیرها چاپ می‌شود. توجه داشته باشید که این مقادیر در برداری به‌صورت acf(x)$acf ذخیره می‌گردد و برای دسترسی به آنها از زمان تأخیر به عنوان اندیس بردار استفاده می‌شود. بنابراین بردار بالا به‌صورت acf$acf[k+1] است. اگر k=0 باشد، مقدار acf$acf[1] برابر یک است.

اکنون به‌عنوان مثال تابع acf کُدهای زیر را ملاحظه نمایید.

```
> www <- "http://www.massey.ac.nz/~pscowper/ts/wave.dat"
> wave.dat <- read.table (www, header=T)
> par(mfrow=c(2,1), mar=c(2,4,2,2))
> plot(ts(waveht), ylab = 'wave height (mm)')
> acf(waveht, main="")
```

نمودار ۱-۸ حاصل اجرای کُدهای بالا است.



شکل ۱-۸: نمایش سری زمانی و همبستگی‌نگار آن

در ادامه مثال قبل مقدار acf در k=1 به‌صورت زیر محاسبه می‌شود.

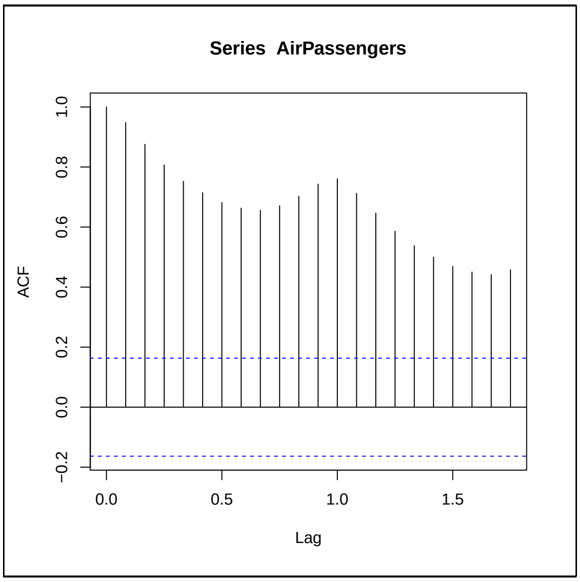
```
> acf(waveht)$acf[2]
[1] 0.4702564
```

مجدداً اگر منظور محاسبه اتوکوواریانس در k=1 باشد، به‌صورت زیر محاسبه می‌شود.

```
> acf(waveht, type = c("covariance"))$acf[2]
[1] 33328.39
```

همان‌طور که ملاحظه می‌کنید دو خط به‌صورت خطچین در بالا و پایین r_kها وجود دارند که نشانه فاصله اطمینان هستند و با سطح معنی‌دار بودن 5% برابر $\pm\frac{2}{\sqrt{N}}$ است، که در آن N اندازه نمونه‌ها می‌باشد.
مقادیر r_k ساختاری است که از آن برای تعیین همبستگی سری مورد استفاده قرار می‌گیرد. اگر مقادیر r_k از بین خطوط اخیر بیرون آمد، آنگاه می‌توان فرض صفر یعنی $\rho_k = 0$ در سطح 5% را رد کرد. البته در اینجا باید به سه نکته توجه داشت. اول این که $r_0 = 1$ از این قاعده مستثنی است. دوم آن که مقادیر درون فاصله دو خط ممکن است دقیقاً صفر نباشند اما با صفر تفاوت معنی‌داری ندارند. سوم آن که در سطح 5% مثلاً برای 40 تأخیر (متناسب با تعداد داده‌ها)، دو تا بیرون افتادن (بجزء r_0) را مشمول خطا ندانید و بر اساس آن داوری نکنید.
معمولاً روند داده‌ها خود را در همبستگی‌نگار به‌صورت یک کاهش آهسته نشان می‌دهد. در همبستگی‌نگار

داده‌های acf(AirPassengers) شکل ۱-۹ مقادیر بزرگ و مثبت (که ناشی از مقادیر مشابه در سری است و در زمان‌های نزدیک به‌هم هستند.) را ملاحظه می‌کنید.

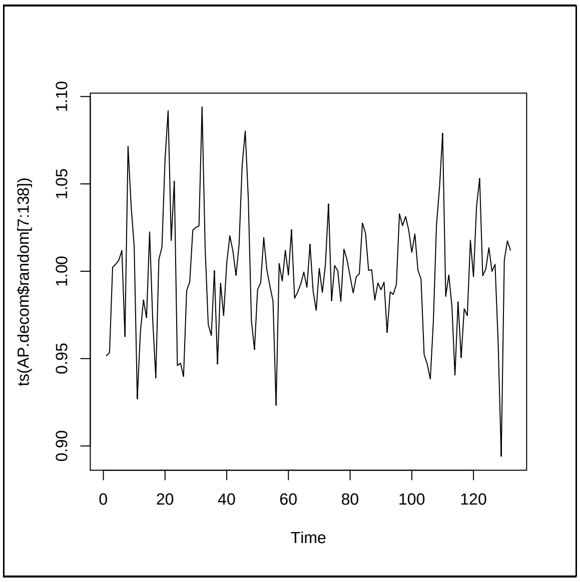


شکل ۱-۹: همبستگی‌نگار داده‌های AirPassengers

اگر تغییرات فصلی نیز وجود داشته باشد، روی همبستگی‌نگار ظاهر می‌شود. تناوب سالانه سری داده‌ها، در همبستگی‌نگار آن نیز با همین تناوب ملاحظه می‌شود. حداکثر مقدار همبستگی در تأخیر یکسال است که رابطه مثبت خطی بین زوج (x_t, x_{t+12}) که توسط 12 ماه از هم جدا شده‌اند را نشان می‌دهد. بالعکس چون رفتار تغییرات فصلی تقریباً سینوسی است، مقادیر به‌صورت 6 ماه (0.5 سال) از هم جدا شده‌اند که رابطه منفی با هم دارند. برای مثال مقادیر بالا که در تابستان رخ می‌دهد به دنبال مقادیر کوچک زمستان است.
هدف اصلی همبستگی‌نگار تشخیص خودهمبستگی در سری زمانی (پس از حذف روند و تغییرات فصلی که برآورد شده‌اند.) است. در کُد زیر با استفاده از تابع decompose() تغییرات فصلی تصحیح و مؤلفه روند حذف می‌گردد. سپس سری مؤلفه تصادفی در شکل ۱-۱۰ و همبستگی‌نگار در شکل ۱-۱۱ ترسیم می‌گردد.

```
> data(AirPassengers)
> AP <- AirPassengers
> AP.decom <- decompose(AP, "multiplicative")
> plot(ts(AP.decom$random[7:138]))
> acf(AP.decom$random[7:138])
```

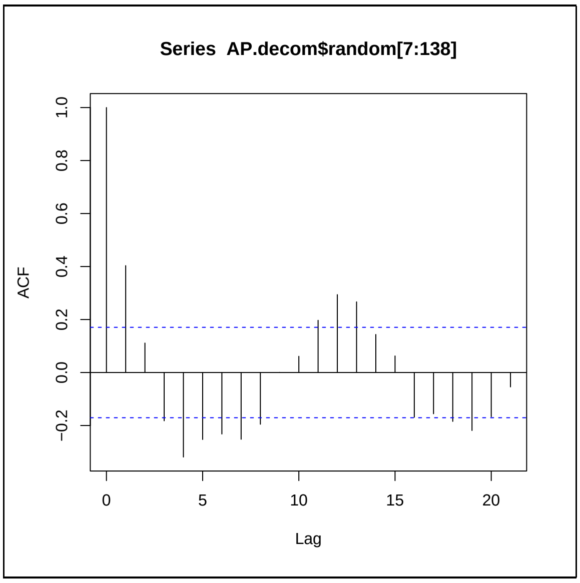
همبستگی‌نگار در شکل ۱-۱۱ نشان می‌دهد که میرایی r_kها کسینوسی شکل است و تصحیح تغییرات فصلی کاملاً مؤثر نبوده است. این تعبیر به نظر درست نمی‌رسد زیرا تجزیه به مؤلفه‌ها، دوازده ماه را به‌صورت مستقل برآورد می‌کند. اگر بررسی بیشتری انجام شود، ملاحظه می‌کنید که انحراف معیار سری داده‌های اصلی از July تا June برابر 109 است. انحراف معیار، پس از حذف روند 41 است و انحراف معیار پس از تصحیح تغییرات



شکل ۱-۱۰: مؤلفه تصادفی AirPassengers

فصلی برابر 0.03 است. برای محاسبه انحراف معیارها به کُدهای زیر توجه کنید.

```
> AP <- AirPassengers
> AP.decom <- decompose(AP, "multiplicative")
> sd(AP[7:138])
[1] 109.4187
> sd(AP[7:138] - AP.decom$trend[7:138])
[1] 41.11491
> sd(AP.decom$random[7:138])
[1] 0.0333884
```



شکل ۱-۱۱: همبستگی‌نگار مؤلفه تصادفی AirPassengers

فصل دوم

مدل‌های تصادفی پایه

در این فصل پاره‌ای از مدل‌های تصادفی که کاربرد زیادی در فصول بعدی دارد، بحث می‌شود. ضمناً عملگرهایی معرفی می‌گردد که سری‌های زمانی از آنها خیلی استفاده می‌کنند و ابزار جدیدی برای تشخیص مناسب‌تر مدل مورد بررسی قرار می‌گیرد. همچنین پاره‌ای از بحث‌های گذشته نیز بسط بیشتری می‌یابد و واضح‌تر می‌گردد.

۲-۱ نوفه سفید[1]

۲-۱-۱ مقدمه

باقیمانده در یک مدل، خطای ناشی از تفاوت بین سری داده‌های مشاهده و مدل را در زمان t بیان می‌کند. سری y_t و مقادیر برآورد شده آن یعنی $\hat{y}_t$ توسط مدل مفروض است. در این صورت سری باقیمانده‌ها یعنی x_t به‌صورت زیر است.

$$x_t = y_t - \hat{y}_t$$

1. White noise

در فصل قبل ملاحظه شد که مؤلفه‌های سری زمانی مانند روند و یا تغییرات فصلی در همبستگی‌نگار منعکس می‌شود. همچنین دیده شد که نوعاً سری‌های زمانی همبسته هستند، اما سری باقیمانده‌ها نباید همبسته باشند و لذا همبستگی‌نگار آنها نباید هیچ نوع الگوی مشخصی را نشان دهد. این ایده موجب تعریف زیر می‌گردد.

۲-۱-۲ تعریف

سری زمانی $\{w_t : t = 1, 2, \cdots, n\}$ را نوفه سفید گسسته (DWN)[۲] گویند، اگر متغیرهای $w_1, w_2, \cdots, w_n$ از هم مستقل و دارای توزیع یکسان با میانگین صفر و واریانس σ^2 باشند. از خاصیت استقلال می‌توان نتیجه گرفت که $i \neq j$ $\mathrm{Cor}(w_i, w_j) = 0$ است. اگر متغیرها از تابع چگالی احتمال نرمال پیروی کند، آنگاه می‌توان سری نوفه سفید گاوسی داشت که $w_t \sim N(0, \sigma^2)$ است.

۳-۱-۲ شبیه‌سازی در R

مدل برازش یافته برای شبیه‌سازی[۳] داده‌ها استفاده می‌شود. سری شبیه‌سازی شده توسط مدل را بعضی اوقات داده‌های ساختگی[۴] نیز گویند تا از داده‌های مشاهده شده متمایز گردد.
شبیه‌سازی فرآیند مفیدی است. برای مثال، می‌تواند سناریوهایی را برای آینده ایجاد نماید و می‌توان از آن برای بازه اطمینان پارامترهای مدل استفاده نمود. در R این کار بخوبی سامان می‌یابد. برای مثال rnorm(100) می‌تواند 100 متغیر مستقل که دارای توزیع یکسان نرمال استاندارد هستند را ایجاد کند.
اکنون به کُدهای زیر توجه کنید.

```
> set.seed(1)
> w <- rnorm(100)
> plot.ts(w, type = "l")
```

تابع set.seed() برای آن است که نقطه شروع هر بار اجرای شبیه‌سازی یکسان باشد، بدین معنی که در هر بار اجرای برنامه 100 عدد یکسان ایجاد می‌گردد. اما اگر این تابع حذف شود با اجراهای مختلف 100تایی‌های متفاوت حاصل می‌گردد. نتیجه اجرای کُدهای بالا شکل ۲-۱ است.

۴-۱-۲ خواص مرتبه دوم و همبستگی‌نگار

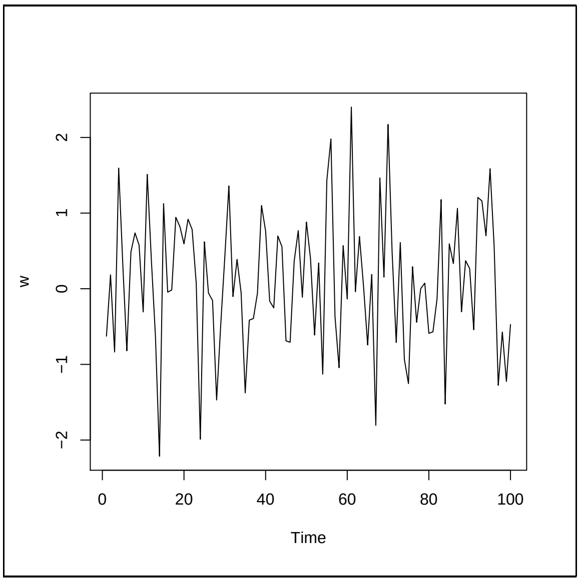
اگر متغیر تصادفی به صورت $\{w_t\}$ و ناهمبسته و دارای میانگین صفر و واریانس σ^2 باشد، دارای تابع اتوکوواریانس زیر است.

$$\gamma(k) = \begin{cases} \sigma^2 & k = 0 \\ 0 & k \neq 0 \end{cases}$$

زیرا

$$\mathrm{Cov}(w_t, w_{t+k}) = E[(w_t - \mu)(w_{t+k} - \mu)]$$

2. Discrete white noise 3. Simulation 4. Synthetic



شکل ۲-۱: نمایش سری زمانی نوفه سفید

چون $\mu = E(w_t) = E(w_{t+k}) = 0$ است، آنگاه نتیجه می‌شود.

$$E(w_t, w_{t+k}) = E(w_t)E(w_{t+k}) = \begin{cases} E(w_t^2) = \sigma_a^2 & k = 0 \\ 0 \times 0 & k \neq 0 \end{cases}$$

و تابع خودهمبستگی عبارتست از:

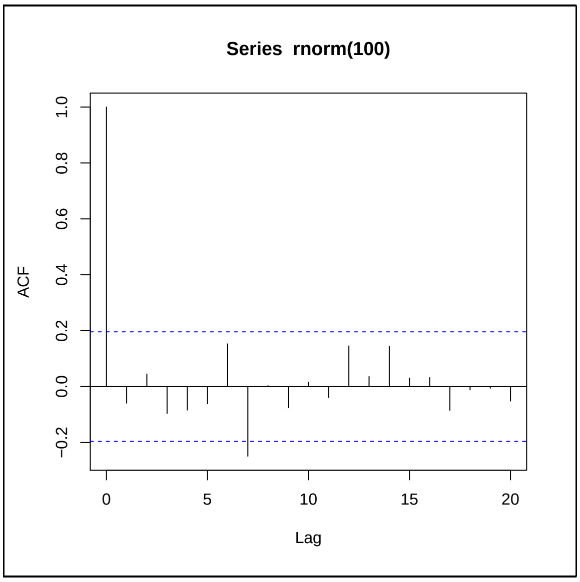
$$\rho_k = \frac{c_k}{c_0} = \begin{cases} 1 & k = 0 \\ 0 & k \neq 0 \end{cases}$$

تابع خودهمبستگی داده‌های نوفه سفید شبیه‌سازی شده به ازای $k \neq 0$ باید برابر صفر باشند اما در عمل به علت تغییرات نمونه‌گیری مقادیر r_kها دقیقاً صفر نیست، اما داخل بازه اطمینان واقع می‌شوند و این بدین معنی است که آنها تفاوت معنی‌داری با صفر ندارند. حتی با توجه به سطح معنی‌دار بودن 5% و 20 تأخیر، بیرون آمدن یک مورد از r_kها نیز بلا اشکال است. اکنون به مثال زیر توجه کنید.

```
> set.seed(2)
> acf(rnorm(100))
```

نتیجه اجرای کُدهای بالا شکل ۲-۲ است، که در تأخیر 7 یک بیرون‌زدگی دارد. البته در R تابعی وجود دارد که تابع خودهمبستگی r_k یک نوفه سفید را به‌صورت نظری محاسبه می‌کند. این تابع ()ARMAacf است. حاصل این تابع را می‌توان با تابع ()plot ترسیم نمود. به مثال زیر توجه کنید.

```
> wn <- ARMAacf(ar=0, ma=0, lag.max=20)
```



شکل ۲-۲: نمایش تابع خودهمبستگی نوفه سفید

```
> plot(wn, type='h')
```

نتیجه اجرای کُدهای بالا شکل ۲-۳ است، که در $k = 0$ برابر یک است و برای سایر تأخیرها مقادیر برابر صفر است.

۲-۲ گام‌زدن تصادفی[5]

۲-۲-۱ مقدمه

گام‌زدن تصادفی غالباً برازش مناسبی برای روندهای تصادفی است. معمولاً حتی برازش آن از مدل فراگیری چون ARIMA مناسب‌تر است.

۲-۲-۲ تعریف

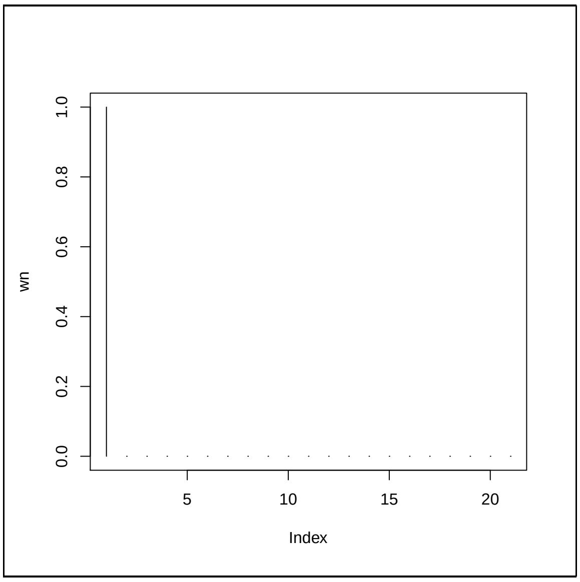
فرض کنید که $\{x_t\}$ یک سری زمانی است. آنگاه $\{x_t\}$ گام‌زدن تصادفی است اگر:

$$x_t = x_{t-1} + w_t$$

که در آن $\{w_t\}$ سری نوفه سفید است. با جایگزینی $x_{t-1} = x_{t-2} + w_{t-1}$ در معادله بالا و همچنین با جایگزینی x_{t-3} در قبلی و ادامه این روش (جایگزینی پسرو[6]) نتیجه می‌شود که:

$$x_t = w_t + w_{t-1} + w_{t-2} + \cdots$$

5. Random walk 6. Back substitution



شکل ۲-۳: نمایش تابع خودهمبستگی نظری نوفه سفید

در عمل سری فوق بی‌نهایت نیست و از زمان $t = 1$ شروع می‌شود. لذا:

$$x_t = w_1 + w_2 + \cdots + w_t$$

جایگزینی پسرو برای مدل‌های سری‌های زمانی پیچیده نیز اعمال می‌گردد. بنابراین می‌توان یک عملگر به‌صورت زیر تعریف نمود.

۲-۲-۳ عملگر انتقال پسرو[7]

عملگر B که به‌صورت زیر است.

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

عملگر انتقال پسرو نامیده می‌شود، که در پاره‌ای از موارد به آن عملگر تأخیر نیز گویند. اگر عملگر گفته شده مرتباً تکرار گردد، آنگاه:

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

اکنون گام‌زدن تصادفی با این عملگر نوشته می‌شود.

$$x_t = B(x_t) + w_t \;\Rightarrow\; (1-B)x_t = w_t \;\Rightarrow\; x_t = (1-B)^{-1}w_t$$
$$= (1 + B + B^2 + \cdots)w_t \;\Rightarrow\; x_t = w_t + w_{t-1} + w_{t-2} + \cdots$$

7. Backward shift operator

۲-۲-۴ خواص مرتبه دوم گام‌زدن تصادفی

خواص مرتبه دوم فرآیند گام‌زدن تصادفی با توجه به $\mu_x = 0$ به‌صورت زیر است.

$$\gamma_k(t) = \mathrm{Cov}(x_t, x_{t+k}) = \mathrm{Cov}\left(\sum_{i=1}^{t} w_i, \sum_{j=1}^{t+k} w_j\right) = \sum_{i=j} \mathrm{Cov}(w_i, w_j)$$

کوواریانس تابعی از زمان است، بنابراین فرآیند ناایستا است. اکنون ضریب خودهمبستگی به‌صورت زیر است.

$$\rho_k(t) = \frac{\mathrm{Cov}(x_t, x_{t+k})}{\sqrt{\mathrm{Var}(x_t)\,\mathrm{Var}(x_{t+k})}} = \frac{t\sigma^2}{\sqrt{t\sigma^2(t+k)\sigma^2}} = \frac{1}{\sqrt{1+k/t}}$$

برای tهای بزرگ بهرحال مقدار k نسبت به t به‌صورت قابل ملاحظه‌ای کوچکتر است. بنابراین مقدار ρ_k نزدیک به یک است و همبستگی‌نگار برای گام‌زدن تصادفی دارای مقادیر مثبت خواهد بود و به‌صورت خیلی کُند از مقدار یک کاهش می‌یابد.

۲-۲-۵ عملگر تفاضل[۸]

این عملگر می‌تواند سری ناایستا را به سری ایستا تبدیل کند. برای مثال اگر سری $\{x_t\}$ گام‌زدن تصادفی باشد که ناایستا است، تفاضل مرتبه اول $\{x_t\}$ سری نوفه سفید را ایجاد می‌کند که رابطه آن به‌صورت $x_t - x_{t-1} = w_t$ است.

عملگر تفاضل یعنی ∇ به‌صورت زیر تعریف می‌شود.

$$\nabla x_t = x_t - x_{t-1}$$

توجه داشته باشید که $\nabla x_t = (1-B)x_t$ است. بنابراین ∇ می‌تواند بر حسب عملگر انتقال پسرو بیان گردد. مرتبه‌های بالاتر عملگر تفاضل به‌صورت زیر است.

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

۲-۲-۶ شبیه‌سازی

غالباً شبیه‌سازی برای مطالعه یک سری زمانی مفید است و خصوصیات اصلی مدل در نمودارها ظاهر می‌شود. بنابراین وقتی که داده تاریخی خصوصیات مشابه را از خود نشان داد، مدل می‌تواند به‌عنوان یک امکان بالقوه برای آن معرفی گردد.

کُدهای زیر می‌تواند یک گام‌زدن تصادفی را شبیه‌سازی کند.

```
> set.seed(2)
> x <- w <- rnorm(1000)
```

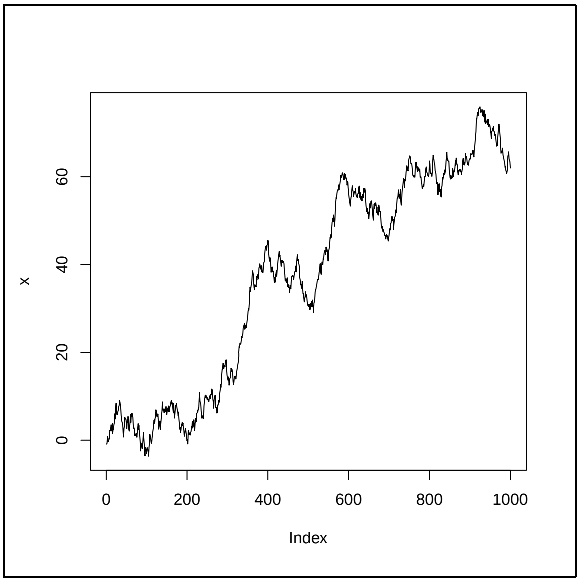
8. Difference operator

```
> for (t in 2:1000) x[t] <- x[t - 1] + w[t]
> plot(x, type = "l")
```

اولین دستور سری نوفه سفید را به w نسبت می‌دهد و ضمناً به x مقداردهی می‌گردد. حلقه for گام‌زدن تصادفی را ایجاد می‌کند. البته می‌توان شبیه‌سازی گام‌زدن تصادفی را با کُدهایی ایجاد نمود که در آن از حلقه استفاده نشده باشد.

```
> set.seed(2) # so you can reproduce the results
> v <- rnorm(1000) # v contains 1000 iid N(0,1) variates
> x <- cumsum(v) # x is a random walk
> plot(x, type = "l")
```

بهر روی سری ایجاد شده در شکل ۲-۴ نشان داده شده است. برای به‌دست آوردن همبستگی‌نگار سری



شکل ۲-۴: نمایش گام‌زدن تصادفی

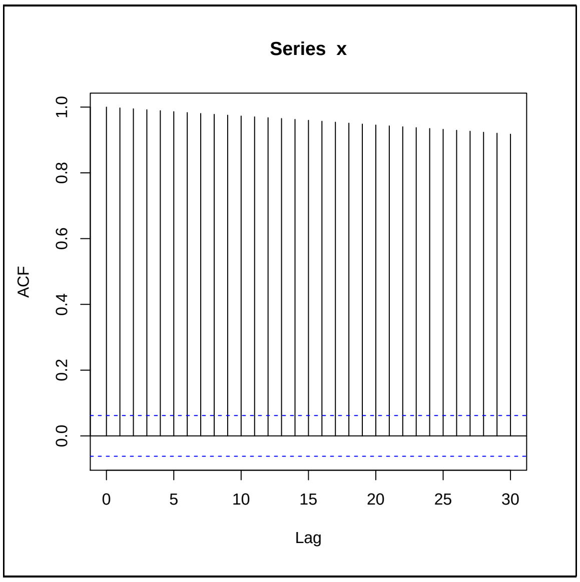
شبیه‌سازی شده، در ادامه برنامه‌های قبل، دستور زیر اجرا می‌گردد. سپس نتیجه در شکل ۲-۵ مشاهده می‌شود.

```
> acf(x)
```

همان‌طور که قبلاً ملاحظه شد، گام‌زدن تصادفی در یک تفاضل تبدیل به سری داده‌های نوفه سفید می‌گردد. اکنون باید دید که مطلب گفته شده در مورد سری شبیه‌سازی شده نیز صادق است. برای اعمال عملگر تفاضل، تابع ()diff در R موجود است. در ادامه برنامه‌های اخیر، دستور زیر اجرا می‌شود. سپس نتیجه در شکل ۲-۶ مشاهده می‌شود.

```
> acf(diff(x))
```

در شکل اخیر، الگوی مشخصی برای همبستگی‌نگار ملاحظه نمی‌شود. دو مورد بیرون‌زدگی از بازه اطمینان وجود دارد که در سطح معنی‌دار بودن 5% قابل اعتنا نیست. بنابراین همن‌طور که انتظار می‌رفت سری شبیه‌سازی



شکل ۲-۵: نمایش خودهمبستگی‌نگار گام‌زدن تصادفی

شده x از یک فرآیند گام‌زدن تصادفی پیروی می‌کند.

۲-۳ مدل‌های اتورگرسیو

۲-۳-۱ تعریف

سری زمانی $\{x_t\}$ را فرآیند اتورگرسیو مرتبه p به اختصار $\mathrm{AR(p)}$ گویند، اگر

$$x_t = \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + \cdots + \alpha_p x_{t-p} + w_t$$

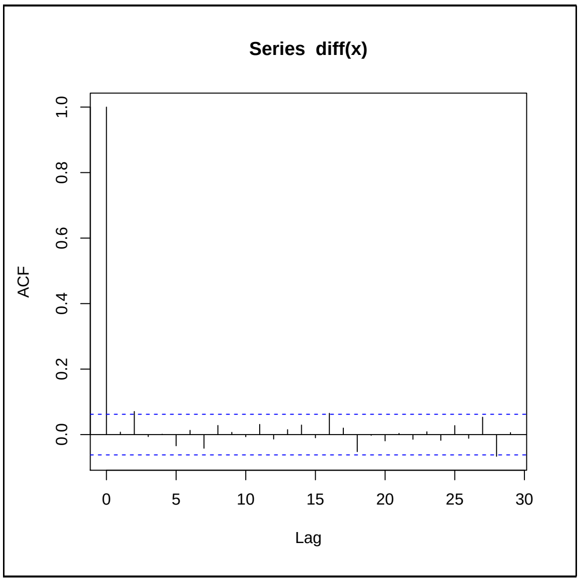
که در آن $\{w_t\}$ نوفه سفید و α_i پارامترهای مدل با $\alpha_p \neq 0$ هستند. اگر معادله اخیر بر حسب عملگر انتقال پسرو بیان گردد، به‌صورت زیر خواهد بود.

$$\theta_p(B)x_t = (1 - \alpha_1 B - \alpha_2 B^2 - \cdots - \alpha_p B^p)x_t = w_t$$

اکنون به نکات زیر توجه کنید.

- گام‌زدن تصادفی حالت خاص $\mathrm{AR}(1)$ است که در آن $\alpha_1 = 1$ می‌باشد.
- این مدل در واقع رگرسیون x_t نسبت به xها در زمان‌های گذشته است و به همین سبب به آن اتورگرسیو گویند.
- برآورد در زمان t برابر است با:

$$\hat{x}_t = \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + \cdots + \alpha_{t-p} x_t$$



شکل ۲-۶: نمایش خودهمبستگی‌نگار گام‌زدن تصادفی با اعمال تفاضل

- پارامترهای مدل به روش حداقل مربعات خطا برآورد می‌گردد.

۲-۳-۲ ایستایی و ناایستایی در اتورگرسیو

معادله $\theta_p(B) = 0$ را معادله مشخصه نامند که می‌تواند دارای ریشه‌های حقیقی و یا مختلط باشد. اگر قدر مطلق تمام ریشه‌های چندجمله‌ای $\theta_p(B)$ بیشتر از واحد گردد، آنگاه فرآیند ایستا است. در فرآیند گام‌زدن تصادفی $\theta = 1 - B$ است که $B = 1$ می‌شود و فرآیند ناایستا است. اکنون به چهار مثال زیر برای تعیین ایستایی و ناایستایی فرآیند AR توجه کنید.

۱) مدل $\mathrm{AR}(1)$ به‌صورت $x_t = \frac{1}{2}x_{t-1} + w_t$ ایستا است، زیرا ریشه معادله $1 - \frac{1}{2}B = 0$ برابر $B = 2$ است که از واحد بزرگتر است.

۲) مدل $\mathrm{AR}(2)$ به‌صورت $x_t = x_{t-1} - \frac{1}{4}x_{t-2} + w_t$ ایستا است. زیرا با استفاده از عملگر انتقال پسرو معادله $\frac{1}{4}(B^2 - 4B + 4)x_t = w_t$ یعنی $\frac{1}{4}(B-2)^2 x_t = w_t$ نتیجه می‌شود. ریشه این چندجمله‌ای با حل معادله $\theta(B) = \frac{1}{4}(B-2)^2$ به‌دست می‌آید که $B = 2$ است و بزرگتر از واحد می‌باشد.

۳) مدل $\mathrm{AR}(2)$ به‌صورت $x_t = \frac{1}{2}x_{t-1} + \frac{1}{2}x_{t-2} + w_t$ ناایستا است، زیرا یکی از ریشه‌ها واحد است. با استفاده از عملگر انتقال پسرو $-\frac{1}{2}(B^2 + B - 2)x_t = w_t$ یعنی $-\frac{1}{2}(B-1)(B+2)x_t = w_t$ نتیجه می‌شود. چندجمله‌ای $\theta(B) = -\frac{1}{2}(B-1)(B+2)$ دارای ریشه‌های $B = 1, -2$ است. فقط مقدار $B = 1$ سبب ناایستایی می‌گردد. زیرا قدر مطلق $B = -2$ بزرگتر از واحد است.

۴) مدل $\mathrm{AR}(2)$ به‌صورت $x_t = -\frac{1}{4}x_{t-2} + w_t$ ایستا است، زیرا ریشه‌های $1 + \frac{1}{4}B^2 = 0$ یعنی

$B = \pm 2i$ هستند که اعداد مختلط است و $i = \sqrt{-1}$ می‌باشد. قدر مطلق هر یک از ریشه‌ها برابر 2 می‌باشد که بزرگتر از واحد است.

برای محاسبه ریشه‌های چندجمله‌ای در R تابعی به نام polyroot() وجود دارد که می‌توان ریشه‌های معادله مشخصه را محاسبه نمود و ایستایی سری‌ها را بررسی کرد. به‌عنوان مثال، موارد سوم و چهارم با تابع گفته شده محاسبه می‌شود.

مثال سوم:

```
> polyroot(c(-2,1,1))
[1] 1-0i -2+0i
```

قدر مطلق ریشه‌ها عبارتند از:

```
> Mod(polyroot(c(-2,1,1)))
[1] 1 2
```

مثال چهارم:

```
> polyroot(c(1,0,1/4))
[1] 0+2i 0-2i
```

قدر مطلق ریشه‌ها عبارتند از:

```
> Mod(polyroot(c(1,0,1/4)))
[1] 2 2
```

۲-۳-۳ خواص مرتبه دوم مدل AR(1)

همان‌طورکه قبلاً مشاهده شد، مدل AR(1) به‌صورت زیر است.

$$x_t = \alpha x_{t-1} + w_t$$

که در آن $\{w_t\}$ نوفه سفید با میانگین صفر و واریانس σ^2 می‌باشد. می‌توان نشان داد که

$$\mu_x = 0$$
$$\gamma_k = \alpha^k \sigma^2/(1-\alpha^2)$$

با استفاده از عملگر B و شرط ایستایی فرآیند $|\alpha| < 1$ می‌توان نوشت که

$$(1-\alpha B)x_t = w_t$$

اکنون می‌توان x_t را محاسبه نمود.

$$\begin{aligned} x_t &= (1-\alpha B)^{-1} w_t \\ &= w_t + \alpha w_{t-1} + \alpha^2 w_{t-2} + \cdots = \sum_{i=0}^{\infty} \alpha^i w_{t-i} \end{aligned}$$

اکنون به محاسبه میانگین توجه کنید.

$$E(x_t) = E\left(\sum_{i=0}^{\infty} \alpha^i w_{t-i}\right) = \sum_{i=0}^{\infty} \alpha^i E(w_{t-i}) = 0$$

اتوکوواریانس به شرح زیر است.

$$\begin{aligned}\gamma_k = \mathrm{Cov}(x_t, x_{t+k}) &= \mathrm{Cov}\left(\sum_{i=0}^{\infty} \alpha^i w_{t-i}, \sum_{j=0}^{\infty} \alpha^j w_{t+k-j}\right) \\ &= \sum\nolimits_{j=k+i} \alpha^i \alpha^j \,\mathrm{Cov}(w_{t-i}, w_{t+k-j}) \\ &= \alpha^k \sigma^2 \sum\nolimits_{i=0}^{\infty} \alpha^{2i} = \alpha^k \sigma^2/(1-\alpha^2)\end{aligned}$$

۲-۳-۴ همبستگی‌نگار فرآیند AR(1)

برای محاسبه تابع خودهمبستگی کافی است که اتوکوواریانس را بر واریانس تقسیم کنید.

$$\rho_k = \alpha^k \qquad (k \geq 0)$$

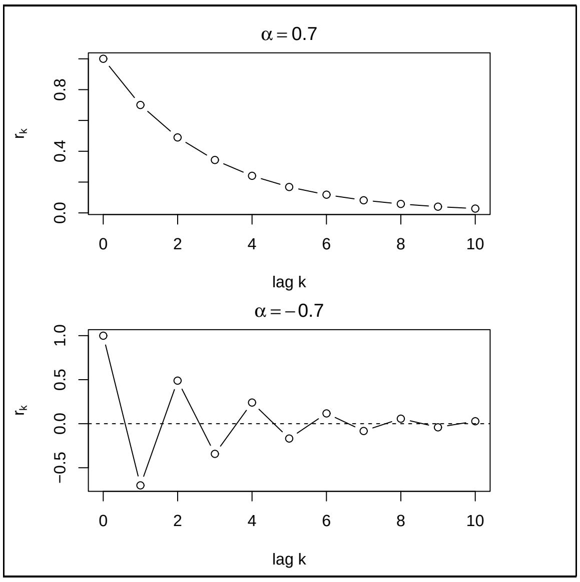
که در آن $|\alpha| < 1$ است. بنابراین همبستگی‌نگار به ازای αهای کوچکتر با سرعت بیشتری به سمت صفر میل می‌کند. در ادامه مثال دو همبستگی‌نگار برای مقادیر مثبت و منفی α در شکل ۲-۷ می‌آید.

```
> plot(0:10, rho(0:10, -0.7), type = "b", xlab="lag k")
> rho <- function(k, alpha) alpha^k
> par(mfrow=c(2,1), mar=c(4,4,2,4))
> plot(0:10, rho(0:10, 0.7), type = "b", xlab="lag k",
+ ylab=expression(r[k]), main=expression(alpha==0.7))
> plot(0:10, rho(0:10, -0.7), type = "b", xlab="lag k",
+ ylab=expression(r[k]), main=expression(alpha==-0.7))
> abline(h=0, lty=2)
```

۲-۳-۵ تابع خودهمبستگی جزیی [۹]

مطابق معادله اخیر، ρ_k، مقادیر خودهمبستگی‌ها برای تمام تأخیرها مخالف صفر هستند، حتی اگر x_t فقط به x_{t-1} بستگی داشته باشد. تابع خودهمبستگی جزیی در تأخیر k، نوعی از همبستگی است که اثر همبستگی‌ها با تأخیر کوچکتر حذف گردد. برای مثال تابع خودهمبستگی جزیی AR(1) برای تمام $k > 1$ صفر خواهد بود. در حالت کلی، تابع خودهمبستگی جزیی در تأخیر k برابر با kامین ضریب برازش یافته مدل AR(k) است، یعنی در فرآیند AR(p) ضرایب α_k به ازای $k > p$ برابر صفر است. بنابراین همبستگی‌نگار تابع خودهمبستگی جزیی برای تعیین مرتبه فرآیند AR می‌تواند مورد استفاده قرارگیرد. در R تابعی به نام pacf() برای محاسبه و ترسیم تابع خودهمبستگی جزیی وجود دارد.

9. Partial autocorrelation function



شکل ۲-۷: نمایش همبستگی‌نگار فرآیند $\mathrm{AR}(1)$ برای $\alpha = 0.7, -0.7$

۲-۳-۶ شبیه‌سازی

مدل $\mathrm{AR}(1)$ در R به‌صورت زیر شبیه‌سازی شده و فرآیند و توابع خودهمبستگی و خودهمبستگی جزیی در شکل ۲-۸ ترسیم می‌گردد. همان‌طورکه ملاحظه می‌کنید برای مقادیر $k > 1$ در همبستگی‌نگار خودهمبستگی جزیی، همبستگی معنی‌داری وجود ندارد.

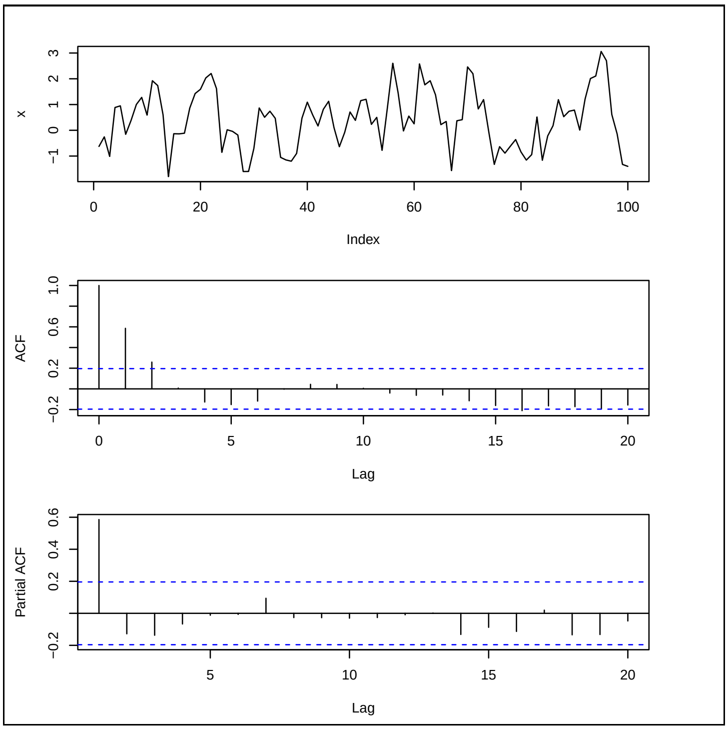
۲-۳-۷ مدل‌های برازش یافته

۲-۳-۸ مدل‌های برازش یافته بر داده‌های شبیه‌سازی شده

مدل $\mathrm{AR(p)}$ در R توسط تابع ar() بر داده‌ها برازش می‌یابد. درکُدهای زیر، مدل اتورگرسیون x.ar برای سری شبیه‌سازی شده با بازه اطمینان 95% و پارامتر داده شده برازش می‌یابد، که واریانس مجانبی[10] آن با استفاده از x.ar$asy.var استخراج می‌شود.

```
> set.seed(1)
> x <- w <- rnorm(100)
> for (t in 2:100) x[t] <- 0.7 * x[t - 1] + w[t]
> x.ar <- ar(x, method = "mle")
> x.ar$order
رتبه مدل
[1] 1
> x.ar$ar
پارامتر مدل
```

10. asymtotic



شکل ۲-۸: نمایش فرآیند $\mathrm{AR}(1)$ و خودهمبستگی‌هایش

```
[1] 0.6009459
> x.ar$ar + c(-2, 2) * sqrt(x.ar$asy.var)
```
بازه اطمینان پارامتر مدل
```
[1] 0.4404031 0.7614886
```

روش mle[11] که در فرآیند برازش در کُدها استفاده شد، بر اساس تابع حداکثر درستنمایی استوار است. انتخاب مقدار p مربوط به فرآیند، از ضابطه AIC[12] انجام می‌شود که بیانگر حداقل پارامتر مدل است. محاسبه ضابطه گفته شده به‌صورت زیر است.

$$\mathrm{AIC} = -2 \times \text{log-likelihood} + 2 \times \text{number of parameters}$$

در تابع ()ar، مدل با کوچکترین AIC بهترین برازش را اختیار می‌کند. توجه کنید که کُدهای بالا، مرتبه درست مدل یعنی $p = 1$ را بازیابی نمود. برآورد پارامتر مدل $\mathrm{AR}(1)$ برابر $\hat{\alpha} = 0.6$ به‌دست آمد، که کوچکتر از مقدار $\alpha = 0.7$ است. اما بازه اطمینان 95% شامل مقدار پارامتر مدل می‌گردد و ایرادی از این حیث وجود ندارد.

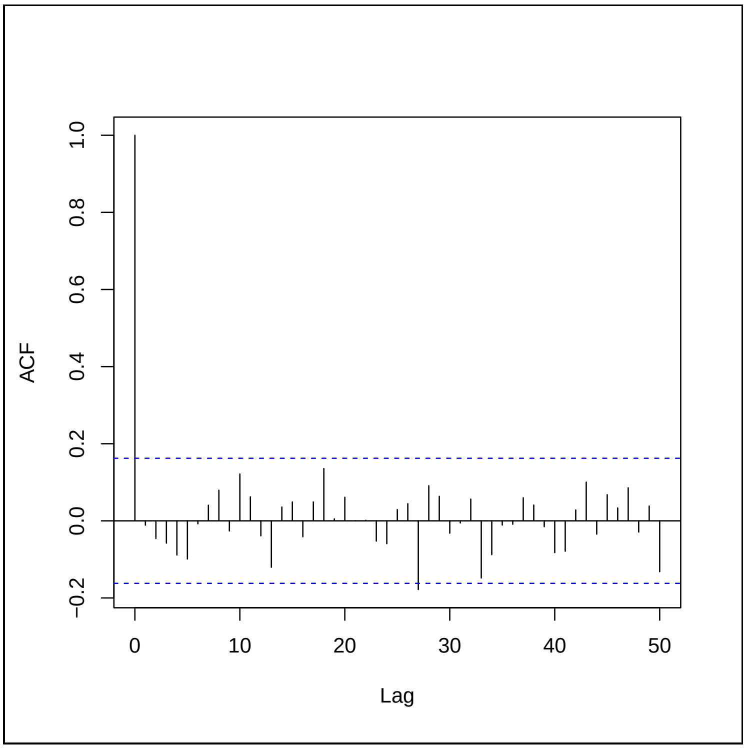
۲-۳-۹ سری دمای جهانی: مدل AR برازش یافته

این سری از سال 1970 افزایش روند دما را نشان می‌دهد که ناشی از اثر گلخانه‌ای است. با تردید می‌توان ادعا نمود که روند فزاینده‌ای ناشی از انتقال یک پدیده تصادفی حذف شده است. برای سازگاری این ادعا با سری

11. Maximum likelihood estimation 12. Akaike Information Criterion

داده‌ها، ممکن است مؤلفه روند را مدل‌سازی نموده بدون آن که از توابع معین استفاده نمود.
به مدل AR برازش یافته به میانگین سری دمای سالانه توجه کنید.

```
> www <- "http://www.massey.ac.nz/~pscowper/ts/global.dat"
> Global <- scan(www)
> Global.ts <- ts(Global, st = c(1856, 1), end = c(2005, 12), fr = 12)
> Global.ar <- ar(aggregate(Global.ts, FUN = mean), method = "mle")
> mean(aggregate(Global.ts, FUN = mean))
[1] -0.1382628
> Global.ar$order
[1] 4
> Global.ar$ar
[1] 0.58762026 0.01260253 0.11116731 0.26763656
> acf(Global.ar$res[-(1:Global.ar$order)], lag = 50, main="")
```



شکل ۲-۹: نمایش همبستگی‌نگار باقیمانده‌ها

همبستگی‌نگار (شکل ۲-۹) باقیمانده‌ها، یک سری نوفه سفید را نشان می‌دهد. زیرا به جز مقدار r_k در تأخیر 27 بقیه آنها در بازه اطمینان قرار دارند. لذا برازش مدل AR(4) بر داده‌ها، مناسب به نظر می‌رسد.
اکنون با توجه به نتایج کُدهای اخیر که مقدار میانگین دمای متوسط سالانه و پارامترهای مدل را به‌دست داده است، می‌توان نوشت.

$$\hat{x}_t = -0.14+0.59(x_{t-1}+0.14)+0.013(x_{t-2}+0.14)+0.11(x_{t-3}+0.15)+0.27(x_{t-4}+0.15)$$

فصل سوم

مدل‌های ایستا

قبلاً راجع به ایستایی مطالبی مطرح شد. اکنون ادامه آن بحث پیگیری می‌شود. سری زمانی $\{x_t\}$ را اکیداً ایستا[1] گویند اگر تابع توزیع توام سری در اثر انتقال در زمان تغییر نکند، یعنی

$$f\left(x_{t_1}, x_{t_2}, \cdots, x_{t_k}\right) = f\left(x_{t_1+h}, x_{t_2+h}, \cdots, x_{t_k+h}\right)$$

توجه کنید که ایستایی مؤکد، میانگین و واریانس ثابت در زمان را نیز شامل می‌شود و کوواریانس $\mathrm{Cov}(x_t, x_s)$ فقط به تأخیر $k = |t-s|$ بستگی داشته باشد.

۳-۱ مدل‌های میانگین متحرک

۳-۱-۱ فرآیند MA(q): تعریف و خواص

فرآیند میانگین متحرک (MA) از مرتبه q ترکیب خطی از جمله جاری نوفه سفید و همچنین آخرین q جمله قبلی نوفه سفید می‌باشد. رابطه آن به‌صورت زیر است.

$$x_t = w_t + \beta_1 w_{t-1} + \cdots + \beta_q w_{t-q}$$

1. Strictly stationary

که در آن $\{w_t\}$ نوفه سفید با میانگین صفر و واریانس σ_w^2 است. معادله بالا را می‌توان بر حسب عملگر انتقال پسرو B، بازنویسی نمود.

$$x_t = (1 + \beta_1 B + \beta_2 B^2 + \cdots + \beta_q B^q) w_t = \phi_q(B) w_t$$

که در آن ϕ_q چندجمله‌ای از درجه q است. نظر به این که فرآیندهای MA شامل جمع متناهی از جملات ایستای نوفه سفید است، لذا ایستایی برقرار است و ثبات در میانگین و اتوکوواریانس نسبت به زمان وجود دارد.
میانگین و واریانس $\{x_t\}$ به سادگی قابل محاسبه است.

$$\begin{aligned} E(x_t) &= E(w_t + \beta_1 w_{t-1} + \cdots + \beta_q w_{t-q}) \\ &= E(w_t) + \beta_1 E(w_{t-1}) + \cdots + \beta_q E(w_{t-q}) \\ &= 0 \end{aligned}$$

زیرا هر یک از میانگین‌ها برابر صفر هستند.

$$\begin{aligned} \mathrm{Var}(x_t) &= \mathrm{Var}(w_t + \beta_1 w_{t-1} + \cdots + \beta_q w_{t-q}) \\ &= \mathrm{Var}(w_t) + \beta_1^2 \mathrm{Var}(w_{t-1}) + \cdots + \beta_q^2 \mathrm{Var}(w_{t-q}) \\ &= \sigma_w^2 (1 + \beta_1^2 + \cdots + \beta_q^2) \end{aligned}$$

زیرا تمام جملات دارای واریانس یکسان σ_w^2 و از یکدیگر مستقل هستند.
تابع خودهمبستگی برای $k \geq 0$ به‌صور زیر است.

$$\rho(k) = \begin{cases} 1 & k = 0 \\ \dfrac{\sum_{i=0}^{q-k} \beta_i \beta_{i+k}}{\sum_{i=0}^{q} \beta_i^2} & k = 1, \ldots, q \\ 0 & k > q \end{cases}$$

که در آن $\beta_0 = 1$ است. برای $k > q$ تابع برابر صفر است، زیرا x_t و x_{t-1} شامل جملات مستقل نوفه سفید هستند و همچنین کوواریانس آنها صفر است.
فرآیند MA را وارون‌پذیر[2] گویند اگر بتوان آن را بر حسب فرآیند اتورگرسیو ایستا با مرتبه نامتناهی بدون جمله خطا نوشت. برای مثال، فرآیند $\mathrm{MA}(1)$ با رابطه $x_t = (1 - \beta B) w_t$ می‌تواند به‌صورت زیر بیان شود.

$$w_t = (1 - \beta B)^{-1} x_t = x_t + \beta x_{t-1} + \beta^2 x_{t-2} + \cdots$$

که در آن $|\beta| < 1$ است تا شرط همگرایی صدق کند.
به‌طور کلی، فرآیند MA(q) یک فرآیند وارون‌پذیر است اگر قدر مطلق ریشه‌های $\phi(B)$ همگی بزرگتر از واحد باشند.

2. invertible

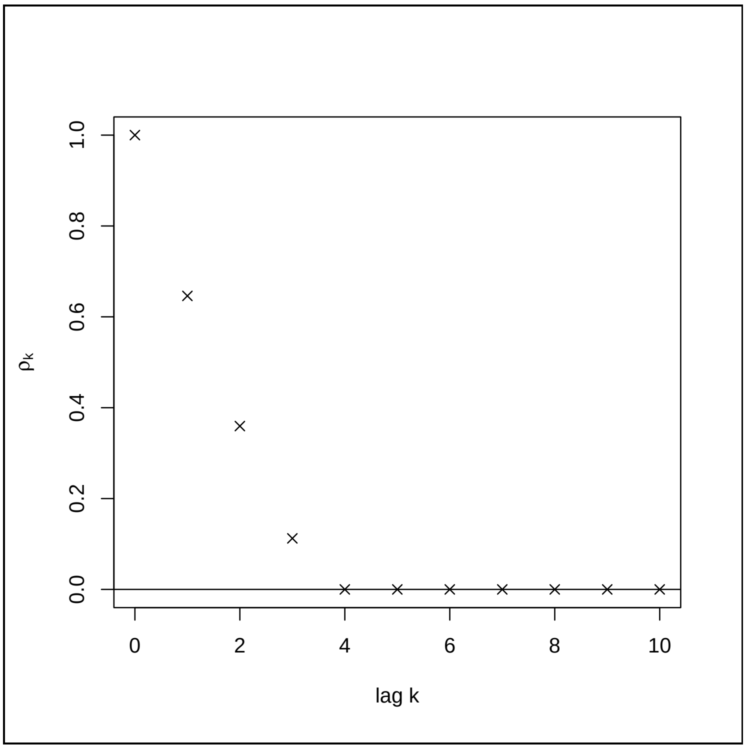
۳-۱-۲ مثال‌های R: همبستگی‌نگار و شبیه‌سازی

تابع خودهمبستگی برای فرآیند MA(q) را می‌توان در R پیاده نمود. برای این کار تابعی با استفاده از حلقه و شرط تعریف می‌شود.

```
> rho <- function(k, beta) {
+ q <- length(beta) - 1
+ if (k > q) ACF <- 0 else {
+ s1 <- 0; s2 <- 0
+ for (i in 1:(q-k+1)) s1 <- s1 + beta[i] * beta[i+k]
+ for (i in 1:(q+1)) s2 <- s2 + beta[i]^2
+ ACF <- s1 / s2}
+ ACF}
```

با استفاده از کُدهای بالا، تابع خودهمبستگی برای فرآیند MA(q) قابل محاسبه است. کُدهای زیر برای فرآیند MA(3) با پارامترهای $\beta_1 = 0.7$، $\beta_=0.5$ و $\beta_3 = 0.2$ به صورت شکل ۳-۱ است.

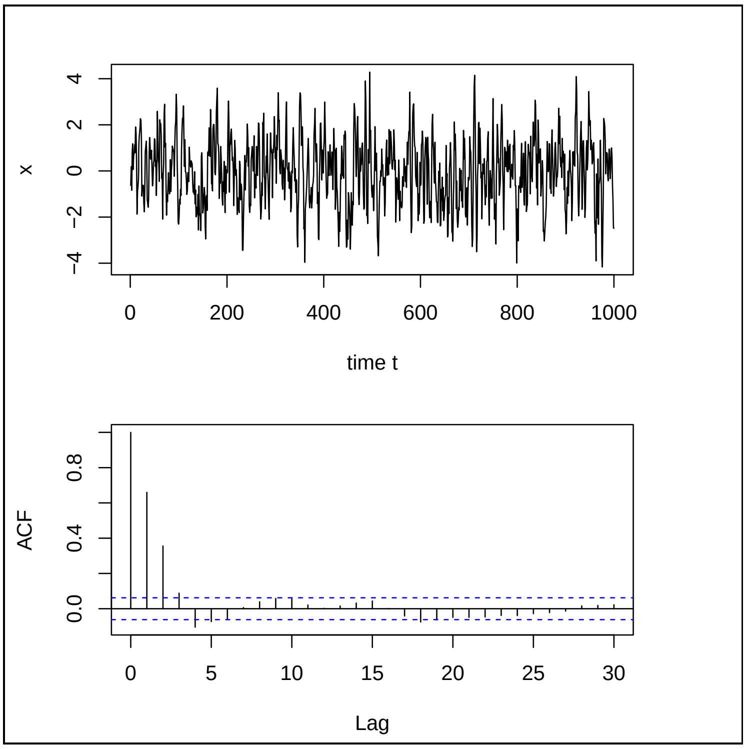
```
> beta <- c(1, 0.7, 0.5, 0.2)
> rho.k <- rep(1, 10)
> for (k in 1:10) rho.k[k] <- rho(k, beta)
> plot(0:10, c(1, rho.k), pch = 4, ylab = expression(rho[k]), xlab="lag k")
> abline(0, 0)
```



شکل ۳-۱: نمایش همبستگی‌نگار مدل MA(3)

کُدهای زیر برای شبیه‌سازی فرآیند $\mathrm{MA}(3)$ و ترسیم همبستگی‌نگار مورد استفاده قرار می‌گیرد. سری زمانی و همبستگی‌نگار در شکل **۳-۲** ترسیم شده است. همان‌طورکه انتظار می‌رفت، اولین سه مقدار تابع خودهمبستگی دارای تفاوت معنی‌داری از صفر هستند و برای تأخیرهای بیشتر از 3، مقادیر r_kها در بازه اطمینان قرار دارند. البته 5% هم ممکن است بیرون‌زدگی وجود داشته باشد که ناشی از تغییرات نمونه‌گیری است.

```
> set.seed(1)
> b <- c(0.8, 0.6, 0.4)
> x <- w <- rnorm(1000)
> for (t in 4:1000) {
+ for (j in 1:3) x[t] <- x[t] + b[j] * w[t - j]
+ }
> par(mfrow=c(2,1), mar=c(4,4,2,4))
> plot(x, type = "l", xlab="time t")
> acf(x, main="")
```



شکل **۳-۲**: نمایش فرآیند شبیه‌سازی $\mathrm{MA}(3)$ و همبستگی‌نگار آن

۳-۲ مدل‌های MA برازش یافته

۳-۲-۱ مدل برازش یافته به سری شبیه‌سازی شده

مدل MA(q) می‌تواند بر داده‌ها برازش یابد. در R تابعی به نام arima() وجود دارد که مرتبه مدل با آرگومان order=c(0,0,q) تنظیم می‌شود. بر خلاف تابع ar()، تابع گفته شده به‌صورت پیش‌فرض میانگین را کسر نمی‌کند و در جمله عرض از مبداء[۳] برآورد می‌گردد.
پارامترهای مدل توسط یک الگوریتم عددی برآورد می‌شوند. تابع arima() برای برآورد پارامترها، مجموع مربعات شرطی را حداقل می‌سازد البته در صورتی که آرگومان method=c("CSS") باشد.
توضیح الگوریتم مجموع مربعات شرطی برای برازش فرآیند MA(q) به این صورت است که به ازای پارامترها، مجموع مربعات باقیمانده‌ها به‌طور مکرر محاسبه می‌گردد. باقیمانده‌ها w_t و برآوردش یعنی $\hat{w}_t$ به‌صورت زیر است.

$$S(\hat{\beta}_1,\ldots,\hat{\beta}_q)=\sum_{t=1}^{n}\hat{w}_t^2=\sum_{t=1}^{n}\left\{x_t-(\hat{\beta}_1\hat{w}_{t-1}+\cdots+\hat{\beta}_q\hat{w}_{t-q})\right\}^2$$

مشروط بر آن که مقادیر $\hat{w}_0,\ldots,\hat{w}_{t-q}$ در شروع تکرار آنها، برابر با صفر باشند. جستجوی عددی، مقادیری از پارامترها را معین می‌کند که مجموع مربعات بالا حداقل گردد.
در کُدهای زیر، مدل میانگین متحرک، x.ma به داده‌های شبیه‌سازی شده اخیر برازش می‌یابد. پارامترهای برآورد شده در خروجی برنامه در بازه اطمینان 95% قرار می‌گیرد (تقریباً $coeff. \pm 2\,\text{s.e. of}\, coeff.$) شامل مقادیر پارامترهای (0.8,0.6,0.4) می‌شود که در شبیه‌سازی از آنها استفاده شده است. ضمناً همچنان که انتظار می‌رفت عرض از مبداء تفاوت معنی‌داری با صفر ندارد.
می‌توان مقدار میانگین را برابر صفر تنظیم نمود. برای این کار از آرگومان include.mean=FALSE در تابع arima() استفاده می‌شود. بنابراین پیش‌فرض include.mean=TRUE است.

```
> set.seed(1)
> b <- c(0.8, 0.6, 0.4)
> x <- w <- rnorm(1000)
> for (t in 4:1000) {
+ for (j in 1:3) x[t] <- x[t] + b[j] * w[t - j]
+ }
> x.ma <- arima(x, order = c(0, 0, 3))
> x.ma


Call:
arima(x = x, order = c(0, 0, 3))
```

3. Intercept

```
Coefficients:
 ma1     ma2     ma3  intercept
 0.7898  0.5665  0.3959    -0.0322
s.e.  0.0307  0.0351  0.0320     0.0898

sigma^2 estimated as 1.068:  log likelihood = -1452.41,  aic = 2914.83
```

۳-۳ مدل‌های ترکیبی: فرآیند ARMA

۳-۳-۱ تعریف

همان‌طور که قبلاً مشاهده شد، سری $\{x_t\}$ فرآیند اتورگرسیو از مرتبه p به‌صورت زیر است.

$$x_t = \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + \cdots + \alpha_p x_{t-p} + w_t$$

که در آن $\{w_t\}$ نوفه سفید و α_iها پارامترهای مدل هستند. دسته مفیدی از مدل‌ها وقتی که فرآیند AR و MA به هم افزوده می‌شوند، حاصل می‌گردد. سری زمانی $\{x_t\}$ دارای فرآیند اتورگرسیو میانگین متحرک (ARMA) از مرتبه (p,q) است و با ARMA(p,q) نشان داده می‌شود. رابطه آن به‌صورت زیر است.

$$x_t = \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + \cdots + \alpha_p x_{t-p} + w_t + \beta_1 w_{t-1} + \beta_2 w_{t-2} + \cdots + \beta_q w_{t-q}$$

که در آن $\{w_t\}$ نوفه سفید است. رابطه فوق را می‌توان به‌صورت عملگر انتقال پسرو نوشت.

$$\theta_p(B) x_t = \phi_q(B) w_t$$

موارد زیر در مورد ARMA(p,q) درخور توجه است.

۱) فرآیند ایستا است اگر قدر مطلق تمام ریشه‌های θ بزرگتر از واحد باشند.

۲) فرآیند وارون‌پذیر است اگر قدر مطلق تمام ریشه‌های ϕ بزرگتر از واحد باشند.

۳) مدل AR(p) حالت خاص مدل ARMA(p,0) است.

۴) مدل MA(q) حالت خاص مدل ARMA(0,q) است.

۵) حداقل پارامتر: وقتی برازش انجام می‌شود، مدل ARMA غالباً کارآیی بیشتری در خصوص پارامترها نسبت به مدل‌های MA و یا AR به تنهایی دارد.

۲-۳-۳ خواص مرتبه دوم مدل

برای تعیین خواص مرتبه دوم ARMA(p,q) سری زمانی $\{x_t\}$، بهتر است که در ابتدا آن را بر حسب جملات نوفه سفید بنویسید. زیرا آنها از هم مستقل هستند. برای مثال مدل ARMA(1,1) را در نظر بگیرید.

$$x_t = \alpha x_{t-1} + w_t + \beta w_{t-1}$$

که در آن w_t نوفه سفید است. میانگین و واریانس آن $E(w_t)$ و $\mathrm{Var}(w_t) = \sigma_w^2$ است. معادله اخیر را بر اساس x_t طوری مرتب نموده که بر حسب مؤلفه‌های نوفه سفید در آید.

$$x_t = (1-\alpha B)^{-1}(1+\beta B)w_t$$

بسط طرف راست رابطه بالا به‌صورت زیر است.

$$\begin{aligned} x_t &= (1+\alpha B+\alpha^2B^2+\cdots)(1+\beta B)w_t \\ &= \left(\sum_{i=0}^{\infty}\alpha^i B^i\right)(1+\beta B)w_t \\ &= \left(1+\sum_{i=0}^{\infty}\alpha^{i+1}B^{i+1}+\sum_{i=0}^{\infty}\alpha^i\beta B^{i+1}\right)w_t \\ &= w_t+(\alpha+\beta)\sum_{i=0}^{\infty}\alpha^{i-1}w_{t-i} \end{aligned}$$

با توجه به رابطه بالا، میانگین $E(x_t)$ برابر صفر است، زیرا به ازای تمام iها $E(w_{t-i})=0$ است و واریانس

$$\begin{aligned} \mathrm{Var}(x_t) &= \mathrm{Var}\left[w_t+(\alpha+\beta)\sum_{i=1}^{\infty}\alpha^{i-1}w_{t-i}\right] \\ &= \sigma_w^2+\sigma_w^2(\alpha+\beta)(1+\alpha^2)^{-1} \end{aligned}$$

برای اتوکوواریانس γ_k برای $k>0$ نتیجه می‌شود.

$$\begin{aligned} \mathrm{Cov}(x_t,x_{t+k}) &= (\alpha+\beta)\alpha^{k-1}\sigma_w^2+(\alpha+\beta)^2\sigma_w^2\alpha^k\sum_{i=0}^{\infty}\alpha^{2i-2} \\ &= (\alpha+\beta)\alpha^{k-1}\sigma_w^2+(\alpha+\beta)^2\sigma_w^2\alpha^k(1-\alpha^2)^{-1} \end{aligned}$$

برای تابع خودهمبستگی ρ_k نتیجه می‌شود.

$$\begin{aligned} \rho_k &= \gamma_k/\gamma_0 = \mathrm{Cov}(x_t,x_{t+k})/\,\mathrm{Var}(x_t) \\ &= \frac{\alpha^{k-1}(\alpha+\beta)(1+\alpha\beta)}{1+\alpha\beta+\beta^2} \end{aligned}$$

از رابطه اخیر می‌توان نتیجه گرفت $\rho_k = \alpha\rho_{k-1}$

۳-۴ مدل‌های ARMA: تحلیل‌تجربی

۳-۴-۱ شبیه‌سازی و برازش

فرآیند ARMA و عمومی‌تر از آن فرآیندهای ARIMA که در فصل بعدی بحث می‌شود را می‌توان توسط تابع arima.sim() در R شبیه‌سازی نمود. مدل ARMA(p,q) می‌تواند توسط تابع arima() با مرتبه c(p,0,q) برازش داد.
در کُدهای زیر داده‌ها برای فرآیند ARMA(1,1) با $\alpha = -0.6$ و $\beta = 0.5$ شبیه‌سازی می‌شود، سپس مدل ARMA(1,1) بر آنها برازش می‌یابد. همان‌طور که انتظار می‌رفت برآورد پارامترهای α و β از نمونه نزدیک مقادیر آنها در بالا است.

```
> set.seed(1)
> x <- arima.sim(n = 10000, list(ar = -0.6, ma = 0.5))
> coef(arima(x, order = c(1, 0, 1)))


 ar1          ma1    intercept
-0.596966371  0.502703368 -0.006571345
```

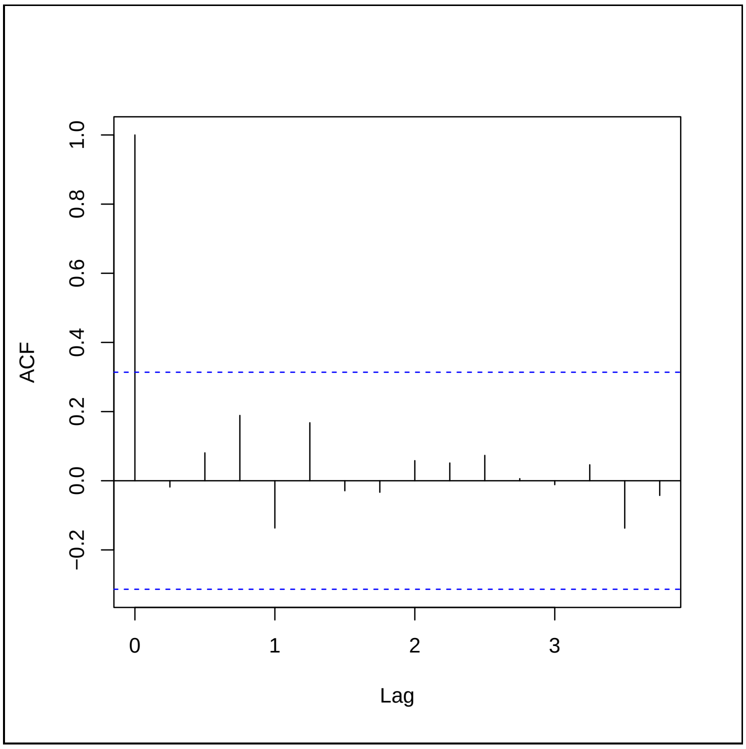
۳-۴-۲ سری نرخ تبدیل

در کُدهای زیر، مدل‌های MA(1)، AR(1) و ARMA(1,1) بر سری نرخ تبدیل برازش می‌یابد و ضابطه AIC آنها با هم مقایسه می‌گردند. فرآیند ARMA(1,1) مدل مناسب‌تری برای داده‌ها می‌باشد. شکل ۳-۳ همبستگی‌نگار باقیمانده مدل ARMA(1,1) را نشان می‌دهد که دارای خودهمبستگی‌های کوچک است، و سازگار با نوفه سفید هستند و استفاده از این مدل را تایید می‌کند.

```
> www <- "http://www.massey.ac.nz/~pscowper/ts/pounds_nz.dat"
> x <- read.table(www, header = T)
> x.ts <- ts(x, st = 1991, fr = 4)
> x.ma <- arima(x.ts, order = c(0, 0, 1))
> x.ar <- arima(x.ts, order = c(1, 0, 0))
> x.arma <- arima(x.ts, order = c(1, 0, 1))
> AIC(x.ma)
[1] -3.526895
> AIC(x.ar)
[1] -37.40417
> AIC(x.arma)
[1] -42.27357
> x.arma
```

```
Call:
arima(x = x.ts, order = c(1, 0, 1))
Coefficients:
ar1 ma1 intercept
0.892 0.532 2.960
s.e. 0.076 0.202 0.244
sigma^2 estimated as 0.0151: log likelihood = 25.1, aic = -42.3
```

> acf(resid(x.arma))



شکل ۳-۳: همبستگی‌نگار باقیمانده سری ARMA(1,1) برای سری نرخ تبدیل

فصل چهارم

مدل‌های ناایستا

همان‌طورکه در فصل‌های گذشته آمد بسیاری از سری‌ها ناایستا هستند، زیرا متأثر از تغیرات فصلی و روند می‌باشند. به‌عنوان مثال گام‌زدن تصادفی یکی از آنها بود. که با یک بار تفاضل به سری ایستا تبدیل شد. در این فصل مدل گام‌زدن تصادفی بسط می‌یابد که شامل جملات اتورگرسیو و میانگین متحرک است.
سری تفاضلی محتاج تجمیع[1] و یا یکپارچگی[2] است، که به اختصار ARIMA[3] نامیده می‌شود.
فرآیند ARIMA شامل جملات فصلی است که آن را ناایستا می‌سازد و می‌توان آن را به‌صورت SARIMA نشان داد. مدل‌های ARIMA فصلی ابزار توانمندی در تحلیل بسیاری از سری‌های زمانی می‌باشد.
در بعضی از سری‌های ناایستا واریانس به‌طور وضوح تغیر می‌کند. این مطلب در سری‌های زمانی مالی شایع است. در رکوردهای اقلیمی نیز امکان بروز دارد. یکی از رویکردها برای مدل‌سازی این نوع سری‌ها، مدل اتورگرسیو برای واریانس است و به اختصار مدل ARCH[4] نامیده می‌شود. نوع تعمیم یافته آن به اختصار GARCH[5] نامیده می‌شود.

1. aggregated 2. integrated 3. **A**uto**R**egressive **I**ntegreted **M**oving **A**vrage 4. **A**uto**R**egressive **C**onditional **H**eteroskedastic 5. **G**eneralised **A**uto**R**egressive **C**onditional **H**eteroskedastic

۴-۱ مدل ARIMA غیر فصلی

۴-۱-۱ تفاضل و سری تولید برق

اعمال تفاضل بر سری $\{x_t\}$ می‌تواند روند را حذف کند، خواه تصادفی باشد مثل گام‌زدن تصادفی و خواه معین باشد مثل موردی که روند خطی است. در گام‌زدن تصادفی یعنی $x_t = X_{t-1} + w_t$، اولین تفاضل $\nabla x_t = x_t - x_{t-1} = w_t$ است که ایستا می‌باشد.
در مقابل فرآیند $x_t = a+bt+w_t$ که در یک روند خطی با خطاهای نوفه سفید است، آنگاه $\nabla x_t = x_t - x_{t-1}$ یک فرآیند میانگین متحرک ایستا است که می‌توان آن را با MA(1) نشان داد. تابع arima() در R برازش مدل‌ها را با مقدار ثابت مجاز نمی‌شمارد. اگر شما بخواهید مدل تفاضلی بر روند معین توسط R برازش دهید، آنگاه سری تفاضلی با میانگین صفر تنظیم می‌شود. البته می‌توان مدل ARMA را بر سری تفاضلی با استفاده از تابع arima() و include.mean=FALSE و d=0 برازش داد.
در ارتباط با مدل ARIMA فرض کنید که $x_t = a + bt + w_t$ باشد، بنابراین $\nabla x_t = y_t = b + w_t + w_{t-1}$ است. با استفاده از تعریف معکوس $y_t = \nabla x_t$ رابطه $x_t = x_0 + \sum_{i=1}^{t} y_i$ به‌دست می‌آید. اگر MA(1) بر سری زمانی تفاضلی $\{y_t\}$ برازش یابد، ضریب w_{t-1} برابر 1- است. بنابراین واریانس $\{x_t\}$ شبیه‌سازی شده حول خط مستقیم افزایش خواهد یافت.
جزییات پاراگراف اخیر را می‌توانید در زیر مشاهده کنید.

$$
\begin{aligned}
y_t &= \nabla x_t \\
&= x_t - x_{t-1} \\
&= a + bt + w_t - \{a + b(t-1) + w_{t-1}\} \\
&= b + w_t + w_{t-1} \\
\Rightarrow x_0 + \sum_{i=1}^{t} y_i &= x_0 + \sum_{i=1}^{t} (b + w_i - w_{t-1}) \\
&= x_0 + bt + w_t = x_t
\end{aligned}
$$

توجه داشته باشید که $w_0 = 0$ است و $x_0 = a$ می‌گردد.
با جایگزینی پسرو برای x_{t-1} نتیجه می‌شود.

$$\begin{aligned}
x_t &= x_{t-1} + b + w_t + \beta w_{t-1} \\
&= x_{t-2} + b + w_{t-1} + \beta w_{t-2} + b + w_t + \beta w_{t-1} \\
&= x_{t-2} + 2b + w_t + (1+\beta) w_{t-1} + \beta w_{t-2} \\
&= x_{t-3} + 3b + w_t + (1+\beta) w_{t-1} + (1+\beta) w_{t-2} + \beta w_{t-3} \\
&\vdots \\
&= x_0 + bt + w_t + (1+\beta) \sum_{i=1}^{t} w_t
\end{aligned}$$

واریانس عبارتست از:

$$\mathrm{Var}(x_t) = \sigma_w^2 \{1 + (1+\beta)^2 (t-1)\}$$

که با افزایش مقدار t واریانش نیز افزایش می‌یابد، مگر این که $\beta = -1$ باشد که در این صورت واریانس برابر σ_w^2 است.

همان‌طور که قبلاً ملاحظه شد، اولین تفاضل در R توسط diff صورت می‌گیرد. برای مثال داده‌های لگاریتم طبیعی مربوط به تولید الکتریسته را در نظر بگیرید. کُدهای زیر سه نمودار (شکل ۴-۱) سری زمانی، سری تفاضلی و سری تفاضلی لگاریتم را نشان می‌دهد. همان‌طور که مشاهده می‌شود در دو سری تفاضلی مثال گفته شده، مؤلفه روند دیگر ظاهر نمی‌شود.

```
> www <- "http://www.massey.ac.nz/~pscowper/ts/cbe.dat"
> CBE <- read.table(www, he = T)
> Elec.ts <- ts(CBE[, 3], start = 1958, freq = 12)
> par(mfrow=c(3,1), mar=c(4,4,2,4))
> plot(Elec.ts)
> plot(diff(Elec.ts))
> plot(diff(log(Elec.ts)))
```

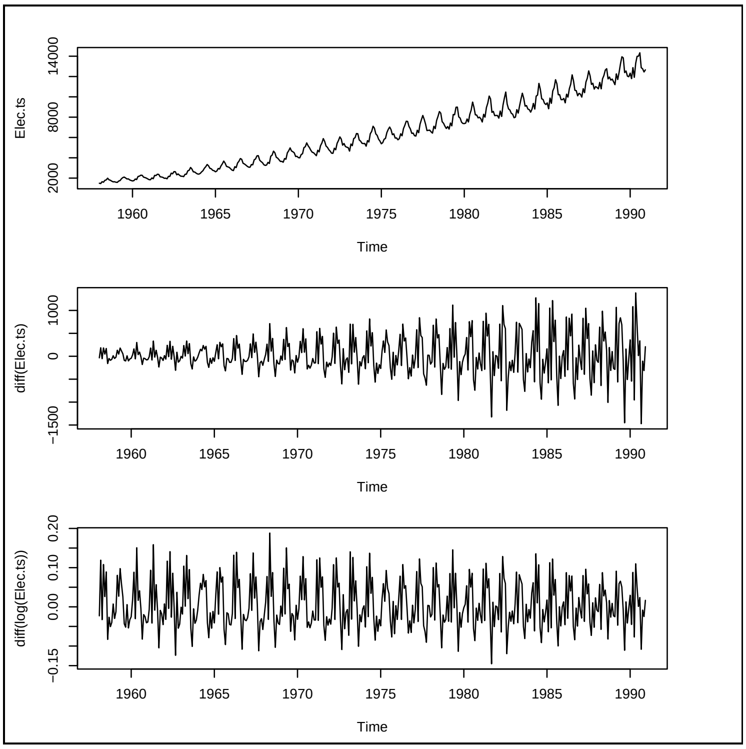
۴-۱-۲ مدل تلفیقی[۶]

اگر سری $\{x_t\}$ پس از dامین تفاضل نوفه سفید $\{w_t\}$ گردد، آنگاه آن را مدل تلفیقی از مرتبه d گویند و با $I(d)$ نشان داده می‌شود. دیدیم که $\nabla^d \equiv (1-B)^d$ است که در آن B عملگر انتقال پسرو می‌باشد. سری $\{x_t\}$ تلفیقی از مرتبه d است، اگر

$$(1-B)^d = w_t$$

باشد. گام‌زدن تصادفی حالت خاص $I(1)$ است. می‌توان تابع diff() را با مرتبه بالا نیز داشت. به‌عنوان مثال برای دوبار تفاضل diff(diff(x)) و یا diff(x, d=2) قابل ذکر است. آرگومان دیگری در این تابع زیر عنوان

6. Integrated model



شکل ۴-۱: نمایش سری زمانی، سری تفاضلی و سری تفاضلی لگاریتمی

lag موجود است که می‌تواند تأخیر تفاضل را مشخص کند. مقدار آن به‌صورت پیش‌فرض برابر واحد است. این آرگومان می‌تواند سبب حذف مؤلفه تغییرات فصلی در مدل جمعی گردد. برای مثال diff(x, lag=12) مؤلفه روند خطی و اثر تغییرات فصلی مدل جمعی را در سری‌های ماهانه حذف می‌نماید. در اینجا ممکن است تفاوت بین دو آرگومان d و lag ابهام داشته باشد. برای رفع ابهام به مثال $w_t = v_t - v_{t-3}$ توجه کنید، سپس کُد آن در تابع diff() به‌صورت زیر است.

```
> w <- diff(v, lag=2, d=1)
```

۴-۱-۳ تعریف و مثال‌ها

سری زمانی $\{x_t\}$ دارای فرآیند ARIMA(p,d,q) است، اگر dامین تفاضل سری $\{x_t\}$ تبدیل به فرآیند ARMA(p,q) گردد. اگر $y_t = (1-B)^d x_t$ باشد، آنگاه $\theta_p(B) y_t = \phi_q(B) w_t$ است. اکنون اگر به‌جای y_t متغیر x_t قرار گیرد، نتیجه می‌شود که

$$\theta_p(B)(1-B)^d x_t = \phi_q(B) w_t$$

که در آن θ_p و ϕ_q چند جمله‌ای‌هایی به ترتیب از مرتبه p و q هستند. اکنون به چند مثال از مدل‌های ARIMA توجه کنید.

- فرآیند $x_t = x_{t-1} + w_t + \beta w_{t-1}$ که در آن β پارامتر مدل است. اکنون به‌صورت عملگر انتقال پسرو نوشته می‌شود.

$$x_t - x_{t-1} = w_t + \beta w_{t-1} \Rightarrow (1-B)x_t = (1+\beta B)w_t$$

بنابراین با مقایسه، سری $\{x_t\}$ به‌صورت ARIMA(0,1,1) است، و به آن مدل تلفیقی میانگین متحرک گویند که به‌صورت IMA(1,1) نشان داده می‌شود. در حالت کلی ARIMA(0,d,q)≡IMA(d,q) است.

- فرآیند $x_t = \alpha x_{t-1} + x_{t-1} - \alpha x_{t-2} + w_t$ که در آن α پارامتر مدل است. اکنون به‌صورت عملگر انتقال پسرو نوشته می‌شود.

$$x_t - x_{t-1} - \alpha x_{t-1} + \alpha x_{t-2} = w_t \Rightarrow (1-B)(1-\alpha B)x_t = w_t$$

و به آن مدل اتورگرسیو تلفیقی گویند که به‌صورت ARI(1,1) نشان داده می‌شود. در حالت کلی ARI(p,d)≡ARIMA(p,d,0) است.

۴-۱-۴ شبیه‌سازی و برازش

فرآیند ARIMA(p,d,q) بر داده‌ها با تابع ()arima در R برازش می‌یابد که مرتبه مدل با c(p,d,q) تعیین می‌گردد. برای مثال در کُدهای زیر داده‌های مربوط به $x_t = 0.5x_{t-1} + x_{t-1} + w_t + 0.3w_{t-1}$ شبیه‌سازی شده و سپس مدل ARIMA(1,1,1) بر آنها برازش داده می‌شود.

```
> set.seed(1)
> x <- w <- rnorm(1000)
> for (i in 3:1000) x[i] <- 0.5 * x[i - 1] + x[i - 1] - 0.5 *
+ x[i - 2] + w[i] + 0.3 * w[i - 1]
> arima(x, order = c(1, 1, 1))

Call:
arima(x = x, order = c(1, 1, 1))

Coefficients:
     ar1    ma1
   0.4235 0.3308
s.e. 0.0433 0.0450

sigma^2 estimated as 1.067: log likelihood = -1450.13, aic = 2906.26
```

البته در R تابع کتابخانه‌ای وجود دارد که کار همان کُد بالا را انجام می‌دهد. این تابع ()arima.sim نام دارد. اکنون سیستم شبیه‌سازی شده به‌صورت زیر است.

```
> x <- arima.sim(model = list(order = c(1, 1, 1), ar = 0.5,
```

```
+ ma = 0.3), n = 1000)
> arima(x, order = c(1, 1, 1))

Call:
arima(x = x, order = c(1, 1, 1))

Coefficients:
        ar1     ma1
     0.5567  0.2502
s.e. 0.0372  0.0437

sigma^2 estimated as 1.079: log likelihood = -1457.45, aic = 2920.91
```

۴-۲ مدل‌های ARIMA فصلی

۴-۲-۱ تعریف

مدل فصلی ARIMA از تفاضل با تأخیری به اندازه فصل (s) استفاده می‌کند تا اثر تغییرات فصلی مدل جمعی حذف گردد. با تأخیر واحد عملگر تفاضل مؤلفه روند را حذف می‌کند و تأخیر s تفاضل بیانگر جمله میانگین متحرک است. مدل فصلی ARIMA شامل جملات اتورگرسیو و میانگین متحرک با تأخیر s می‌باشد. مدل فصلی $\mathrm{ARIMA}(p,d,q)(P,D,Q)_s$ را می‌توان بر حسب عملگر انتقال پسرو به‌صورت زیر نوشت

$$\Theta_P(B^s)\theta_p(B)(1-B^s)^D(1-B)^d x_t = \Phi_Q(B^s)\phi_q(B)w_t$$

که در آن Θ_P، θ_p، Φ_Q و ϕ_q به ترتیب چند جمله‌های از مرتبه P، p، Q و q می‌باشند. در حالت عمومی این مدل ناایستا است. اما اگر $D=d=0$ و قدر مطلق ریشه‌های معادله مشخصه (جملات چند جمله‌ای طرف چپ معادله اخیر) همگی بزرگتر از واحد باشند، آنگاه ایستا خواهد بود. اکنون به چند مثال مدل فصلی ARIMA توجه کنید.

- مدل ساده AR با تناوب فصلی 12، $\mathrm{ARIMA}(0,0,0)(1,0,0)_{12}$ که به‌صورت $x_t = \alpha x_{12} + w_t$ است. چنین مدلی برای داده‌های ماهانه مناسب است مشروط بر آن که ماه سال گذشته روی ماه سال جاری اثر گذاشته باشد. این مدل ایستا است اگر $|\alpha^{-1/12}| > 1$ باشد.
- مدل بند بالا را می‌توان به‌صورت $x_t = x_{t-1} + \alpha x_{t-12} - \alpha x_{t-13} + w_t$ بسط داد. با جابجایی و فاکتورگیری رابطه $(1-\alpha B^{12})(1-B)x_t = w_t$ و یا $\Theta_1(B^{12})(1-B)x_t = w_t$ به‌دست می‌آید، که با مقایسه با رابطه کلی ARIMA فصلی، فرآیند $\mathrm{ARIMA}(0,1,0)(1,0,0)_{12}$ حاصل می‌شود. توجه کنید که این مدل را به‌صورت $\nabla x_t = \alpha \nabla x_{t-12} + w_t$ نیز می‌توان نوشت. با توجه به این مطلب که تغییر در زمان t بستگی به تغییر در همان زمان (یعنی ماه) در سال گذشته دارد. این مدل ناایستا است، زیرا چند جمله‌ای طرف چپ شامل جمله $(1-B)$ است که دارای ریشه $B=1$ می‌باشد.

- یک مدل میانگین متحرک چهار فصل به‌صورت $x_t = (1-\beta B^4)w_t = w_t - \beta w_{t-4}$ است. این مدل ایستا است و فقط برای داده‌های بدون روند مناسب است. اگر داده‌ها دارای روند تصادفی باشند، مدل به‌صورت تفاضل مرتبه اول بسط می‌یابد که $x_t = x_{t-1} + w_t - \beta w_{t-4}$ است، و می‌توان به‌صورت $\mathrm{ARIMA}(0,1,0)(0,0,1)_4$ نیز نوشت. اگر جملات فصلی دارای روند تصادفی باشد، تفاضل در تناوب فصلی نیز اعمال می‌گردد و رابطه $x_t = x_{t-4} + w_t - \beta w_{t-4}$ را به‌دست می‌دهد، و آن را به‌صورت $\mathrm{ARIMA}(0,0,0)(0,1,1)_4$ نیز می‌توان نوشت.

باید توجه داشت که تفاضل با تأخیر s، روند خطی را حذف می‌کند، بنابراین می‌توان تفاضل با تأخیر یک داشت یا خیر؟ اگر تأخیر یک در تفاضل اعمال گردد در حالی که روند خطی وجود دارد، آنگاه جملات میانگین متحرک به‌صورت نوفه سفید معرفی می‌گردد.
به‌عنوان مثال، به سری زمانی با تناوب 4 که در آن روند خطی و تغییرات فصلی و نوفه سفید به‌صورت یک مدل جمعی وجود دارد توجه کنید.

$$x_t = a + bt + s_{[t]} + w_t$$

که $[t]$ باقیمانده t بر 4 را نشان می‌دهد. بنابراین $s_{[4]} = s_{[t-4]}$ اکنون به اولین تفاضل با تأخیر 4 توجه کنید.

$$\begin{aligned}(1-B^4)x_t &= x_t - x_{t-4}\\ &= a + bt - (at + b(t-4)) + s_{[t]} - s_{[t-4]} + w_t - w_{t-4}\\ &= 4b + w_t - w_{t-4}\end{aligned}$$

که می‌توان آن را به‌صورت ARIMA(0,0,0)(0,1,1) با جمله ثابت 4b بیان نمود. اکنون فرض کنید که اولین مرتبه تفاضل با تأخیر یک قبل از تفاضل با تأخیر 4 اعمال گردد، آنگاه

$$\begin{aligned}(1-B^4)(1-B)x_t &= (1-B^4)(b + s_{[t]} - s_{[t-1]} + w_t - w_{t-1})\\ &= w_t - w_{t-1} - w_{t-4} + w_{t-5}\end{aligned}$$

که می توان آن را به‌صورت $\mathrm{ARIMA}(0,1,1)(0,1,1)_4$ نوشت که فاقد مقدار ثابت است.

۴-۲-۲ رویه برازش

مدل‌های ARIMA فصلی، به‌صورت بالقوه می‌تواند شامل پارامترهای متعدد و ترکیبی از جملات باشد. بنابراین شایسته است که محدوده وسیعی از مدل برای برازش بر داده‌ها امتحان شود و برای انتخاب مدل مناسب‌تر از ضابطه AIC استفاده گردد. مدل مناسب برازش یافته شده، منطقاً دارای همبستگی‌نگار باقیمانده به‌صورت نوفه سفید است.
در R، از تابع ()arima برای برازش استفاده می‌شود. در این تابع آرگومانی زیر عنوان seasonal وجود دارد که مؤلفه‌های فصلی در آن درج می‌گردد. در مثال زیر سری لگاریتم داده‌های تولید الکتریسته در نظر گرفته شده

است. پارامتر تفاضل در هر دو مورد $d = 1$ است که روند خطی را حذف می‌کند. مدل ARI برازش بهتری را نشان می‌دهد، زیرا AIC کوچکتری دارد.

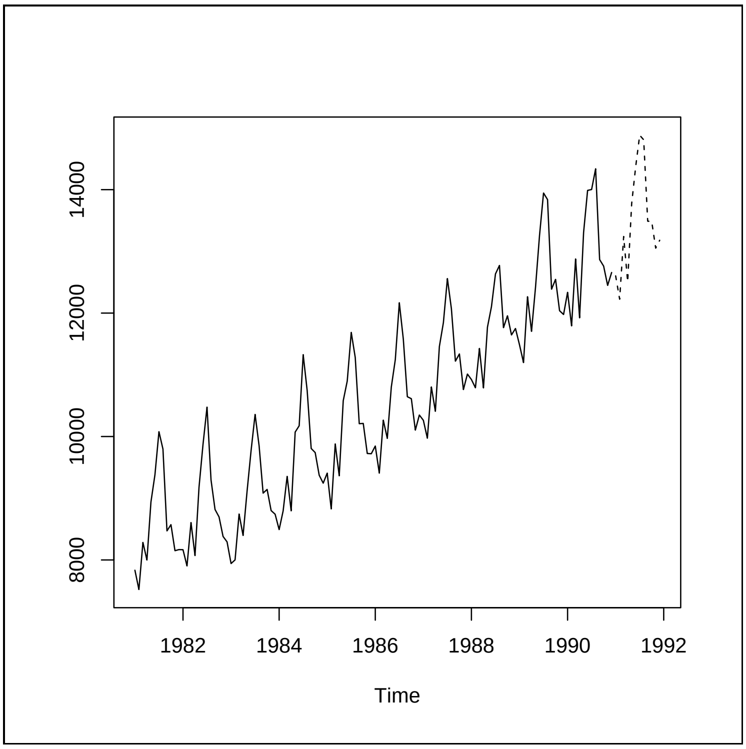
```
> www <- "http://www.massey.ac.nz/~pscowper/ts/cbe.dat"
> CBE <- read.table(www, he = T)
> Elec.ts <- ts(CBE[, 3], start = 1958, freq = 12)
> AIC (arima(log(Elec.ts), order = c(1,1,0),
+ seas = list(order = c(1,0,0), 12)))
[1] -1764.741
> AIC (arima(log(Elec.ts), order = c(0,1,1),
+ seas = list(order = c(0,0,1), 12)))
[1] -1361.586
```

برای کنترل دامنه مدل‌ها و یافتن مدل مناسب‌تر از سعی و خطا استفاده می‌شود. برای سهولت در این امر تابع کوچکی را می‌توان نوشت تا بر اساس ضابطه AIC مدل مناسب‌تر را به‌دست دهد. این رویکرد وقتی روش CSS باشد بهتر جواب می‌دهد و الگوریتم استوارتری است. اکنون به کُدهای زیر توجه کنید.

```
> www <- "http://www.massey.ac.nz/~pscowper/ts/cbe.dat"
> CBE <- read.table(www, he = T)
> Elec.ts <- ts(CBE[, 3], start = 1958, freq = 12)
> get.best.arima <- function(x.ts, maxord = c(1,1,1,1,1,1))
+ {
+ best.aic <- 1e8
+ n <- length(x.ts)
+ for (p in 0:maxord[1]) for(d in 0:maxord[2]) for(q in 0:maxord[3])
+ for (P in 0:maxord[4]) for(D in 0:maxord[5]) for(Q in 0:maxord[6])
+ {
+ fit <- arima(x.ts, order = c(p,d,q),
+ seas = list(order = c(P,D,Q),
+ frequency(x.ts)), method = "CSS")
+ fit.aic <- -2 * fit$loglik + (log(n) + 1) * length(fit$coef)
+ if (fit.aic < best.aic)
+ {
+ best.aic <- fit.aic
+ best.fit <- fit
+ best.model <- c(p,d,q,P,D,Q)
+ }
+ }
+ list(best.aic, best.fit, best.model)
+ }
> best.arima.elec <- get.best.arima( log(Elec.ts),
+ maxord = c(2,2,2,2,2,2))
> best.fit.elec <- best.arima.elec[[2]]
> acf( resid(best.fit.elec) )
> best.arima.elec [[3]]
[1] 0 1 1 2 0 2
```

```
> ts.plot( cbind( window(Elec.ts,start = 1981),
+ exp(predict(best.fit.elec,12)$pred) ), lty = 1:2)
```

با اجرای کُدهای بالا، مدل مناسب‌تر از مرتبه دوم $\text{ARIMA}(0,1,1)(2,0,2)_{12}$ است. آزمون مراتب بالاتر مدل ضروری به نظر نمی‌رسد، زیرا باقیمانده‌های مدل گفته شده تقریباً نوفه سفید (شکل ۴-۳) هستند. شکل ۴-۲ مقادیر پیش‌بینی شده را نشان می‌دهد.



شکل ۴-۲: نمایش پیش‌بینی سری ماهانه تولید برق

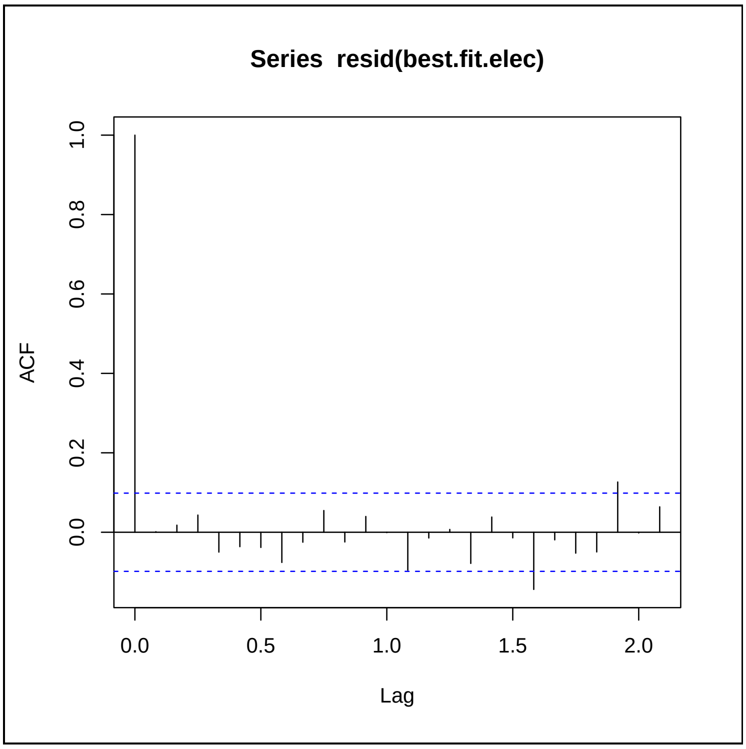
۴-۳ مدل‌های ARCH

۴-۳-۱ سری‌های SP500

داده‌های SP500 مربوط به کمپانی McGraw-Hill درکتابخانه MASS در R وجود دارد.

```
> library(MASS)
> data(SP500)
> par(mfrow=c(2,1), mar=c(4,4,2,4)) > plot(SP500, type = 'l')
> acf(SP500, main="")
```

شکل ۴-۴ سری زمانی اطلاعات گفته شده را نشان می‌دهد. در نگاه اول فرآیند ایستا را نشان می‌دهد. با دقت بیشتر معلوم می‌شود که واریانس در ثلث وسط سری کوچکترین و مقدار آن در ثلث آخر، بزرگترین مقدار

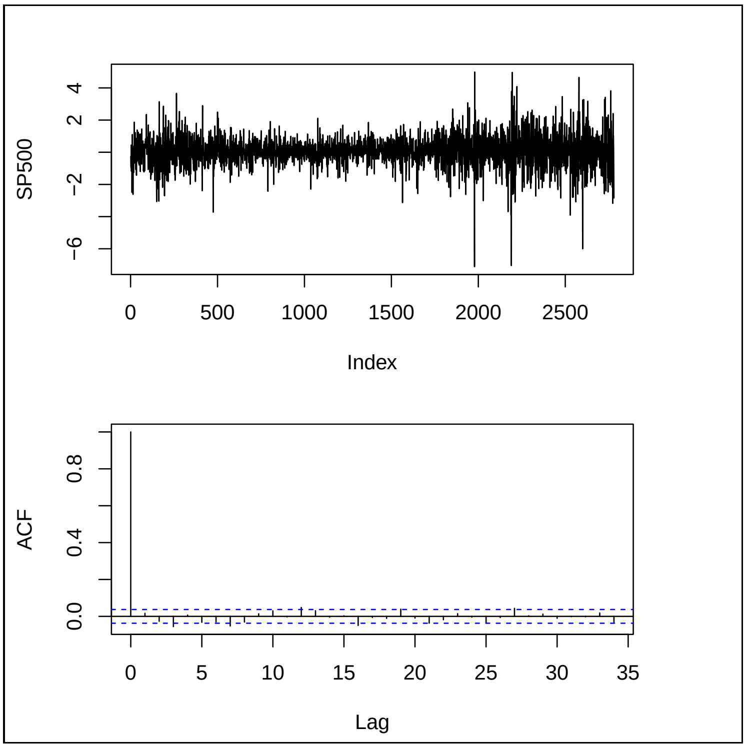


شکل ۴-۳: نمایش باقیمانده‌های سری ماهانه تولید برق

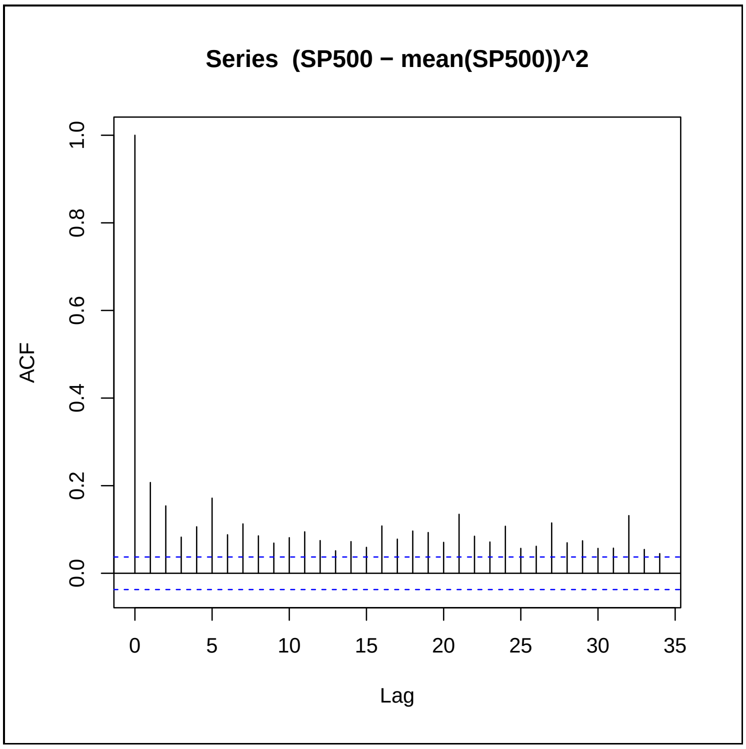
را دارد. به عبارت دیگر افزایش در مسیر منظمی انجام نمی‌شود. اگر واریانس نسبت به زمان ثابت نباشد و تغییرات منظم انجام گیرد، مانند داده‌های هواپیمایی و تولید برق (که واریانس با روند افزایش می‌یابد) آنگاه به آن heteroskedastic گویند. اگر واریانس سری دارای پریود افزایشی باشد و وابسته به زمان گردد (مثل داده SP500) آنگاه به سری heteroskedastic شرطی گویند. توجه داشته باشید که همبستگی‌نگار سری متغیر قیمت‌ها (SP500) تفاوت معنی‌داری با نوفه سفید ندارد، اما سری ناایستا است به دلیل این که در زمان متفاوت واریانس متغیر است. اگر مطابق شکل ?? همبستگی‌نگار نوفه سفید را نشان می‌دهد، آنگاه همبستگی‌نگار مربع مقادیر تغییرات نسبت به زمان را نشان می‌دهد. این مربعات معادل واریانس هستند (به شرط آن که میانگین آنها به‌صورت صفر تنظیم گردد). میانگین داده SP500 از 2 ژانویه 1900 تا 31 دسامبر 1999 برابر 0.048 است، که در مقایسه با واریانس مقدار کوچکی است. بنابراین اندیس از 360 تا 1469 در نظر گرفته می‌شود. همبستگی‌نگار مقادیر مربع با میانگین تصحیح شده SP500 به‌صورت زیر است.

```
library(MASS)
data(SP500)
acf((SP500 - mean(SP500))^2)
```

شکل ۴-۵ به وضوح heteroskedastic شرطی و ناپایداری را نشان می‌دهد.



شکل ۴-۴: نمایش سری داده‌های SP500



شکل ۴-۵: نمایش مربع مقادیر با میانگین تصحیح شده

۴-۳-۲ مدل کردن ناپایداری: تعریف مدل ARCH

برای مدل کردن این رفتار، باید مدلی به‌دست داد که تغییرات شرطی در واریانس را تبین کند. برای رسیدن به این مطلب، سری $\{\epsilon\}$ اولین مرتبه اتورگرسیو شرطی heteroskedastic است که با ARCH(1) نشان داده می‌شود، اگر

$$\epsilon_t = w_t\sqrt{\alpha_0 + \alpha_1\epsilon_{t-1}^2}$$

باشد. در رابطه بالا $\{w_t\}$ نوفه سفید با میانگین صفر و واریانس واحد است و α_0، α_1 پارامترهای مدل هستند. برای بیان volality از رابطه اخیر واریانس بگیرید.

$$\begin{aligned}
\mathrm{Var}(\epsilon_t) &= E(\epsilon_t^2)\\
&= E(\epsilon_t^2)E(\alpha_0 + \alpha_1\epsilon_{t-1}^2)\\
&= E(\alpha_0 + \alpha_1\epsilon_{t-1}^2)\\
&= \alpha_0 + \alpha_1\,\mathrm{Var}(\epsilon_{t-1})
\end{aligned}$$

نظر به این که $\{w_t\}$ دارای واریانس واحد و $\{\epsilon\}$ دارای میانگین صفر است، اگر معادله اخیر را با فرآیند AR(1) مقایسه کنید

$$x_t = \alpha_0 + \alpha_1 x_{t-1} + w_t$$

حاصل می‌شود. ملاحظه می‌کنید که واریانس فرآیند ARCH(1) درست مثل AR(1) رفتار می‌کند. خودهمبستگی‌های مربع خطاها مشخص می‌کند که برازش مدل ARCH مناسب است؟

۴-۳-۳ بسط‌های مربوط به مدل‌های GARCH

اولین مرتبه مدل ARCH می‌تواند به یک فرآیند مرتبه p با تأخیرهای بالا بسط یابد. فرآیند ARCH(p) به‌صورت زیر است.

$$\epsilon_t = w_t\sqrt{\alpha_0 + \sum_{i=1}^{p}\alpha_p\epsilon_{t-i}^2}$$

که در آن $\{w_t\}$ نوفه سفید با میانگین صفر و واریانس واحد است.
بسط‌های وسیع‌تر با کاربرد داده‌های مالی از مدل ARCH به‌صورت GARCH(q,p) است، که ARCH(p) حالت خاص GARCH(0,p) است. سری $\{\epsilon_t\}$ دارای مدل GARCH(q,p) می‌باشد، اگر

$$\epsilon_t = w_t\sqrt{h_t}$$

که در آن

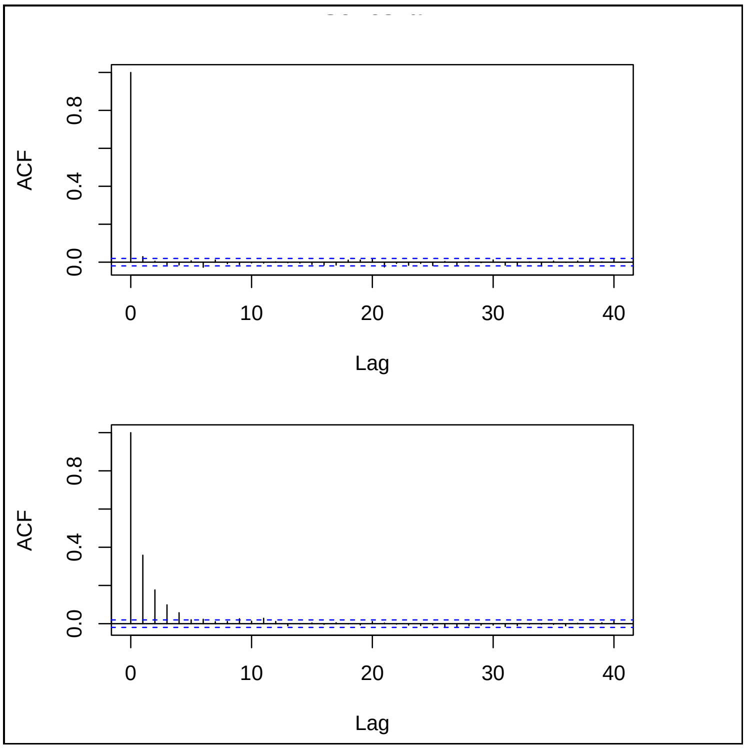
$$h_t = \alpha_0 + \sum_{i=1}^{p}\alpha_i\epsilon_{t-i}^2 + \sum_{j=1}^{q}\beta_j h_{t-j}$$

است و $\alpha_i\ (i=0,1,\ldots,p)$ و $\beta_j\ (j=0,1,\ldots,q)$ پارامترهای مدل هستند.

۴-۳-۴ شبیه‌سازی و مدل GARCH برازش یافته

در کُدهای زیر داده‌هایی با مدل GARCH(1,1) به‌صورت $a_t = w_t\sqrt{h_t}$ که $h_t = \alpha_0 + \alpha_{t-1} + \beta_{t-1}$ است شبیه‌سازی می‌شود. برای پایداری باید $\alpha_1 + \beta_1 < 1$ باشد و همبستگی‌نگار به‌صورت شکل ۴-۶ است.

```
> set.seed(1)
> alpha0 <- 0.1
> alpha1 <- 0.4
> beta1 <- 0.2
> w <- rnorm(10000)
> a <- rep(0, 10000)
> h <- rep(0, 10000)
> for (i in 2:10000) {
+ h[i] <- alpha0 + alpha1 * (a[i - 1]^2) + beta1 * h[i -1]
+ a[i] <- w[i] * sqrt(h[i])
+ }
> par(mfrow=c(2,1), mar=c(4,4,2,4))
> acf(a)
> acf(a^2, main="")
```



شکل ۴-۶: نمایش همبستگی‌نگار مقادیر و مربع آنها

سری a خواص مدل GARCH را نشان می‌دهد که مقادیر ناهمبسته دارد که در مربع مقادیر همبستگی ظاهر می‌گردد.

در مثال زیر، مدل GARCH بر روی داده‌های شبیه‌سازی شده برازش می‌یابد که با استفاده از تابع garch() انجام می‌شود که در بسته نرم‌افزاری tseries یافت می‌شود. ملاحظه می‌گردد که پارامترهای مدل دوباره بازیابی می‌شود و در محدوده بازه اطمینان 95% قرار می‌گیرد. پیش‌فرض تابع اخیر برابر GARCH(1,1) است که نوعاً پاسخگو است، اما برای رتبه‌های بالاتر از آرگومان order=c(p,q) استفاده می‌شود.

```
> set.seed(1)
> alpha0 <- 0.1
> alpha1 <- 0.4
> beta1 <- 0.2
> w <- rnorm(10000)
> a <- rep(0, 10000)
> h <- rep(0, 10000)
> for (i in 2:10000) {
+ h[i] <- alpha0 + alpha1 * (a[i - 1]^2) + beta1 * h[i -1]
+ a[i] <- w[i] * sqrt(h[i])
+ }
> library(tseries)
> library(quadprog)
> library(zoo)
> a.garch <- garch(a, grad = "numerical", trace = FALSE)
> confint(a.garch)
```

با اجرای کُدهای بالا، نتایج زیر حاصل می‌گردد.

```
       2.5 %      97.5 %
a0 0.0882393 0.1092903
a1 0.3307897 0.4023932
b1 0.1928344 0.2954660
```

در مثال آرگومان trace=F خلاصه محاسبات را به‌دست می‌دهد و مراحل محاسبات عددی را به‌صورت تفضیلی نشان نمی‌دهد. آرگومان grad="numerical" روش عددی استوارتری را به‌دست می‌دهد.

۵-۳-۴ برازش بر سری SP500

مدل GARCH بر سری SP500 برازش می یابد. سری باقیمانده مدل GARCH یعنی $\{w_t\}$ از رابطه زیر محاسبه می‌شود.

$$\hat{w}_t = \frac{\epsilon_t}{\sqrt{\hat{h}_t}}$$

اگر مدل GARCH برای سری باقیمانده‌ها مناسب باشد، آنگاه نتیجه باید به‌صورت نوفه سفید با میانگین صفر و واریانس واحد ظاهر می‌شود. در حالت GARCH(1,1)

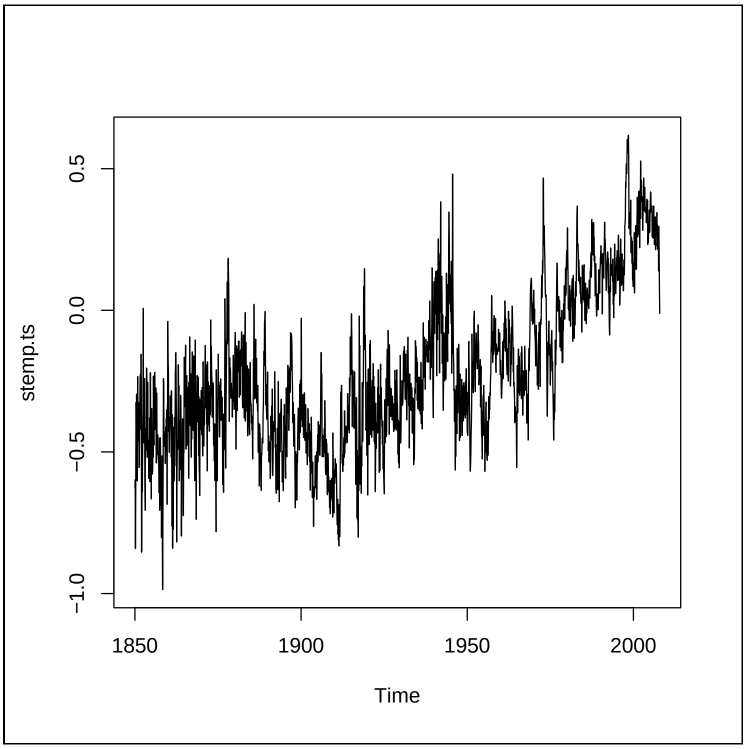
$$\hat{h}_t = \hat{\alpha}_0 + \hat{\alpha}_1\epsilon_{t-1}^2 + \hat{\beta}\hat{h}_{t-1}$$

که در آن $\hat{h}_1 = 0$ به ازای $t = 2, \ldots, n$ است.
محاسبات توسط تابع garch انجام می‌گیرد. اولین مقدار سری باقیمانده موجود نیست. بنابراین آن مقدار با استفاده از [-1] حذف می‌شود و همبستگی‌نگار سری باقیمانده و سری مربعات باقیمانده‌ها به‌دست می‌آید.

۴-۳-۶ ناپایداری در سری اقلیم

اخیراً راجع به ناپایداری در اقلیم مطالعاتی انجام شده است. داده‌های (2007-1850) مربوط به واحد تحقیقات دانشگاه East Anglia ملاک عمل قرارگرفته است. سری مورد نظر خوانده شده و در شکل **۴-۷** ترسیم شده است. مدل ARIMA مناسب توسط تابع get.best.arima که در کُدهای قسمت **۴-۲-۲** آمده است، به‌دست می‌آید. بازه اطمینان برای پارامترهای مدل با استفاده از تابع t() از حالت سطری به ستونی تبدیل شده است.

```
> stemp <- scan("http://www.massey.ac.nz/~pscowper/ts/stemp.dat")
> stemp.ts <- ts(stemp, start = 1850, freq = 12)
> plot(stemp.ts)
> stemp.best <- get.best.arima(stemp.ts, maxord = rep(2,6))
> stemp.best[[3]]
[1] 1 1 2 2 0 1
```



شکل ۴-۷: نمایش سری ناپایداری اقلیم

در ادامه کُدهای فوق پارامترهای مدل برآورد می‌گردد.

```
> stemp.arima <– arima(stemp.ts, order = c(1,1,2),
+ seas = list(order = c(2,0,1), 12))
> t( confint(stemp.arima) )
```

	ar1	ma1	ma2	sar1	sar2	sma1
2.5 %	0.8317389	-1.447400	0.3256699	0.8576802	-0.02501886	-0.9690530
97.5 %	0.9127947	-1.312553	0.4530474	1.0041394	0.07413434	-0.8507036

دومین مؤلفه فصلی AR یعنی sar2 تفاوت معنی‌داری با صفر ندارد، لذا مدل با کُدهای زیر مجدداً برازش می‌یابد.

```
> stemp.arima <– arima(stemp.ts, order = c(1,1,2),
+ seas = list(order = c(1,0,1), 12))
> t( confint(stemp.arima) )
```

	ar1	ma1	ma2	sar1	sma1
2.5 %	0.8304007	-1.445057	0.3242728	0.9243495	-0.9694897
97.5 %	0.9108115	-1.311246	0.4508999	0.9956639	-0.8679223

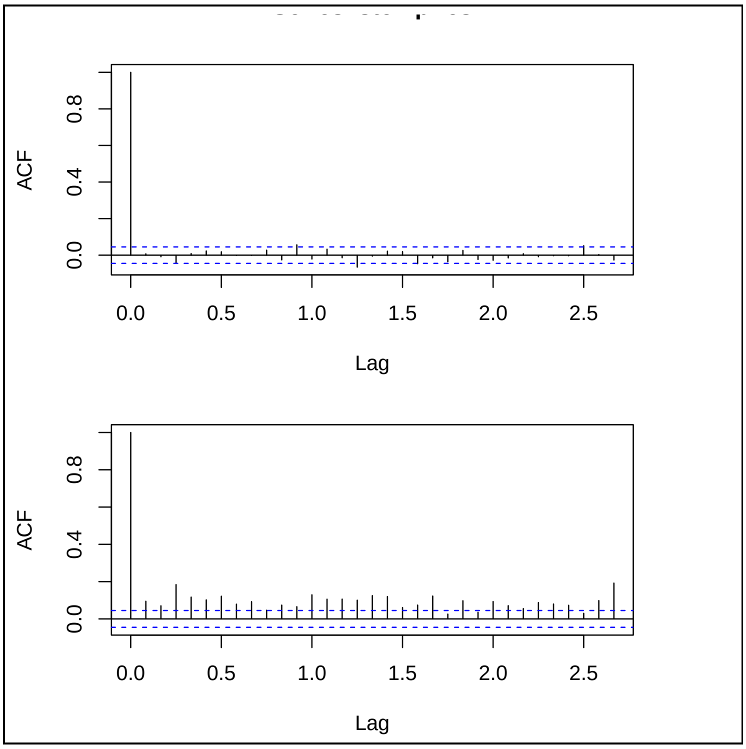
برای کنترل نکویی برازش، همبستگی‌نگار باقیمانده‌های مدل ARIMA در شکل ۴-۸ ترسیم شده است. ضمناً برای بررسی همبستگی‌نگار مربع باقیمانده‌ها نیز در شکل اخیر ملاحظه می‌شود.

```
> stemp.res <- resid(stemp.arima)
> par(mfrow=c(2,1), mar=c(4,4,2,4))
> acf(stemp.res)
> acf(stemp.res^2, main="")
```

کاملاً روشن است که ناپایداری وجود دارد، بنابراین مدل GARCH بر سری باقیمانده‌ها برازش می‌یابد.

۴-۳-۷ GARCH در پیش‌بینی و شبیه‌سازی

همان‌طور که قبلاً ملاحظه شد مدل GARCH اثر خود را در تحلیل سری زمانی در باقیمانده‌ها نشان می‌دهد. لذا تأثیری در پیش‌بینی نقاط ندارد. بنابراین کاربرد اصلی این مدل در مطالعات سری‌های زمانی مالی، بیمه و اقلیم برای فرآیند شبیه‌سازی است.



شکل ۴-۸: نمایش همبستگی‌نگار باقیمانده‌ها و مربعات آن

مراجع

[1] **Cowpertwait, Paul S.P., Metcalf Andrew V. (2009)** *Introductory Time Series with R,* Springer, 254p.

[2] **Ihaka Ross, (2005)** *Time Series Analysis,* University of Auckland, 105p.

[3] **Shumway Robert H., Stoffer David S., (2011)** *Time Series Analysis and Its Applications (With R Examples),* Third edition, Spriger, 596p.