

پروژه شماره ۲ : تحلیل مولفه اصلی با داده واقعی

مقدمه

سرمایه‌گذاری و انباشت سرمایه در تحول اقتصادی کشور نقش بسزایی داشته است. اهمیت این عامل و نقش مؤثر آن را میتوان به وضوح در سیستم کشورهای با نظام سرمایه‌داری مشاهده کرد. بدون شک بورس یکی از مناسب ترین جایگاه‌ها جهت جذب سرمایه‌های کوچک و استفاده از آنها در جهت رشد یک شرکت، در سطح کلان و نیز رشد شخصی فرد سرمایه‌گذار است. از آنجایی که هدف و تعریف سرمایه‌گذاری، به تعویق انداختن مصرف جهت مصرف بیشتر و بهتر در آینده است؛ افراد با سرمایه گذاری انتظار دستیابی به سود مورد انتظار خود را دارند.

ما در این پروژه سعی داریم تحلیل مولفه اصلی را برای قیمت جفت ارز یورو به دلار انجام دهیم.

داده‌های ما به کمک نرم‌افزار متاتریدر ۴ در بازه زمانی ۶ ساله در تایم فریم روزانه استخراج شده اند.

ابتدا داده‌های خود را فراخوانی میکنیم :

```
> EURUSD<-read.csv("C:/Users/12345/Desktop/Processed_EURUSD.csv",header=T)
> dim(EURUSD)
[1] 1313    7
```

داده‌های ما شامل ۷ متغیر(متغیر اول تاریخ است که جلوتر از محاسبات حذف میشود) برای ۱۳۱۳ روز کاری نماد معاملاتی EURUSD میباشد.

در زیر لیست تمام متغیرها را مشاهده میکنیم :

```
> names(EURUSD)
[1] "Date"    "mom1"    "mom3"    "Momentum" "RSI"      "StdDev"   "StdDevSma"
```

متغیرهای ما اندیکاتور هستند که در زیر آنها را تعریف میکنیم

اندیکاتور مومنتوم (Momentum)

اندیکاتور مومنتوم برای شناسایی سرعت حرکت قیمت طراحی شده است. این اندیکاتور ، جزء اندیکاتورهای پیشرو می باشد .نحوه محاسبه مومنتوم به این صورت است که قیمت فعلی پایانی [سهم](#) را با [قیمت پایانی](#) n دوره قبل سهم مقایسه می کند. منظور از n دوره بازه زمانی است که توسط تحلیلگر انتخاب می شود.در اینجا از سه اندیکاتور مومنتوم استفاده کردیم.

اندیکاتور RSI

اندیکاتور RSI که مخفف Relative Strength Index می باشد با روابط پیچیده ریاضی محاسبه می گردد که ذکر این روابط از حوصله این بحث خارج می باشد. در تنظیمات RSI نیز عموماً از تنظیم ۱۴ روزه استفاده می شود. این اندیکاتور همانند مکدی (MACD) در زمره اسیلاتورها طبقه بندی می شود چراکه همواره بین دو سطح ۰ و ۱۰۰ در حال نوسان می باشد.

اندیکاتور StdDev

اندیکاتور StdDev اختصاری از Standard Deviation است که یک اندیکاتور انحراف معیار است و انحراف معیار واژه آماری است. انحراف معیار یعنی چگونه ارقام وسیع از میانگین تفریق می شوند. این اندیکاتور مقیاس تغییرات قیمت مربوط به میانگین متحرک را مشخص می کند. بنابراین اگر حجم اندیکاتور بالا باشد، بازار بی ثبات است و میله های نشان دهنده قیمت نسبت به میانگین متحرک پراکنده هستند. اگر حجم اندیکاتور بالا نباشد، به این معنی است که بازار بی ثبات نیست و قیمت ها به میانگین متحرک نزدیک هستند. در این پروژه از دو اندیکاتور StdDev استفاده شده.

اولین قدم تحلیل مولفه اصلی بدست آوردن بردار میانگین است :

```
> EURUSD<-EURUSD[, -1]
> colMeans(EURUSD)
      mom3      Momentum      RSI      StdDev      StdDevSma
1.000148e+02 1.000422e+02 4.993945e+01 5.777226e-03 5.496093e-03
```

اکنون به سراغ بدست آوردن ماتریس واریانس-کواریانس میرویم :

```
> (S<-cov(EURUSD))
      mom3      Momentum      RSI      StdDev      StdDevSma
mom3      7.180242e-01 6.359884e-01 6.142336e+00 1.793856e-07 1.960776e-05
Momentum  6.359884e-01 1.938842e+00 1.341774e+01 1.137701e-04 1.457327e-04
RSI        6.142336e+00 1.341774e+01 1.279793e+02 2.086328e-03 2.226654e-03
StdDev     1.793856e-07 1.137701e-04 2.086328e-03 8.451479e-06 8.033296e-06
StdDevSma  1.960776e-05 1.457327e-04 2.226654e-03 8.033296e-06 7.727530e-06
```

محاسبه مقدار ویژه و بردار ویژه :

```
> eigen(S)
eigen() decomposition
$values
[1] 1.296875e+02 5.273421e-01 4.212860e-01 1.598059e-05 4.759178e-08

$vectors
      [,1]      [,2]      [,3]      [,4]      [,5]
[1,] -4.782602e-02 -0.1001622476 0.9938209571 -3.181338e-04 -1.449469e-05
```

```
[2,] -1.045737e-01  0.9899938601  0.0947440099 -2.613022e-04 -1.737674e-05
[3,] -9.933665e-01 -0.0993963099 -0.0578217838  6.647095e-05  1.206342e-06
[4,] -1.607243e-05 -0.0001796979 -0.0002603502 -7.228335e-01 -6.910222e-01
[5,] -1.718022e-05 -0.0001498329 -0.0002265896 -6.910222e-01  7.228335e-01
```

کاهش بعد به کمک مقادیر ویژه:

```
> landa<-eigen(S)$values
> which(landa>mean(landa))
[1] 1
```

براساس مقادیر ویژه مولفه‌های با مقادیر ویژه بزرگتر از میانگین مقادیر ویژه را نگه می‌داریم که در اینجا مولفه‌های اصلی اول را نگه می‌داریم.

تحلیل مولفه اصلی به کمک ماتریس واریانس-کوواریانس :

```
> EURUSD.pca<-princomp(EURUSD,cor=F,scores = T)
> summary(EURUSD.pca)
```

Importance of components:

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5
Standard deviation	11.3837060	0.725906624	0.64881824	3.99605e-03	2.180723e-04
Proportion of Variance	0.9927383	0.004036723	0.00322488	1.22329e-07	3.643078e-10
Cumulative Proportion	0.9927383	0.996774998	0.99999988	1.00000e+00	1.000000e+00

در اینجا نیز میتوان نتیجه گرفت که میتوان کاهش بعد انجام داد چون که مولفه اصلی اول ۹۹ درصد کل واریانس را نمایش میدهد .

همچنین میتوان انحراف استاندارد هر متغیر را در بخش (standard deviation) مشاهده کرد.

در اینجا نیز میتوانیم بردارهای ویژه را مشاهده کنیم:

```
> loadings(EURUSD.pca)
```

Loadings:

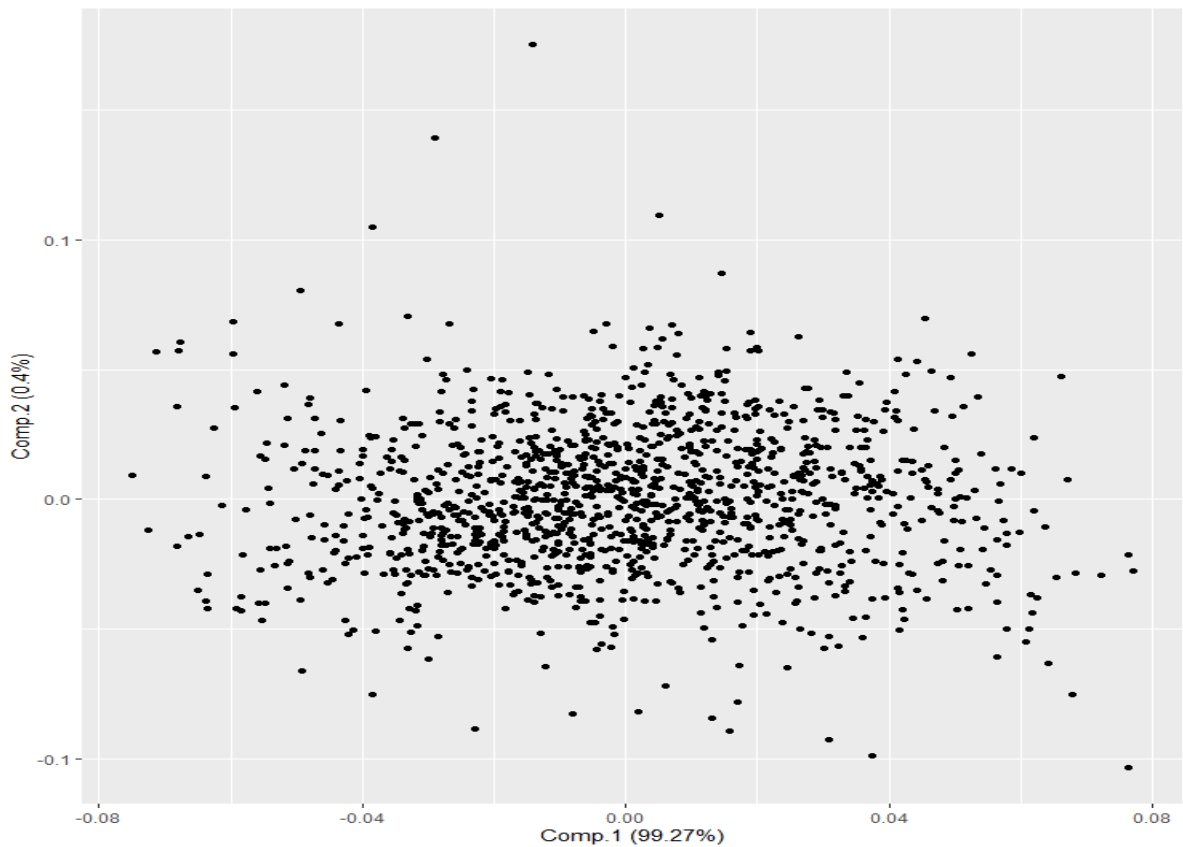
	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5
mom3		0.100	0.994		
Momentum	0.105	-0.990			
RSI	0.993				
StdDev			0.723	0.691	
StdDevsma			0.691	-0.723	

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5
SS loadings	1.0	1.0	1.0	1.0	1.0
Proportion Var	0.2	0.2	0.2	0.2	0.2
Cumulative Var	0.2	0.4	0.6	0.8	1.0

رسم نمودار :

```
> require(ggfortify)
```

```
> autoplot(EURUSD.pca)
```



نمودار بالا اطلاعات خوبی در زمینه نرمال بودن داده‌ها، نقاط دور افتاده و همچنین میزان پراکنش داده‌ها نسبت به PCA نمایش می‌دهد.

طبق نمودار بالا داده‌ها دارای توزیع نرمال هستند.

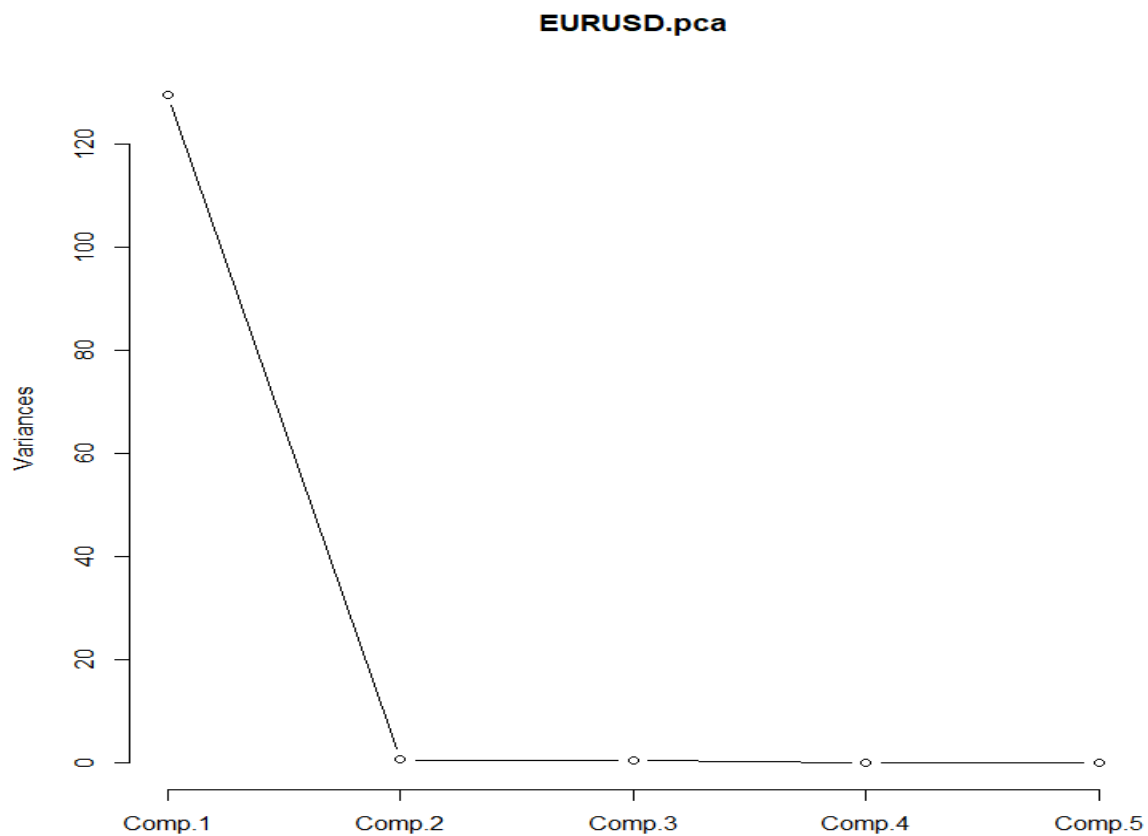
همچنین به نظر می‌رسد که دارای نقطه پرت هستیم.

رسم نمودار سنگ ریزه :

```
> plot(EURUSD.pca,type ="lines")
```

طبق نمودار پایین ما باید در کاهش بعد ۱ مولفه اصلی را نگه داریم.

در نمودار پایین باید نگاه به شیب خط کنیم یعنی تا جایی که شیب خط زیاده مولفه‌ها را نگه می‌داریم و از جایی که شیب خط کم شد مولفه‌ها را حذف می‌کنیم.



تحلیل مولفه اصلی با استفاده از ماتریس همبستگی (R)

ماتریس R را تعریف میکنیم :

```
> (R<-cor(EURUSD))
```

	mom3	Momentum	RSI	StdDev	StdDevsma
mom3	1.000000e+00	0.53902443	0.64075814	7.282018e-05	0.008324113
Momentum	5.390244e-01	1.00000000	0.85180158	2.810546e-02	0.037650048
RSI	6.407581e-01	0.85180158	1.00000000	6.343754e-02	0.070804763
StdDev	7.282018e-05	0.02810546	0.06343754	1.000000e+00	0.994047343
StdDevsma	8.324113e-03	0.03765005	0.07080476	9.940473e-01	1.000000000

مقادیر ویژه را بدست می آوریم:

```
> eigen(R)
eigen() decomposition
$values
```

```
[1] 2.383626898 1.975662847 0.498800144 0.136015835 0.005894276
```

```
$vectors
```

	[,1]	[,2]	[,3]	[,4]	[,5]
[1,]	-0.5046476	0.1427260	0.83693252	0.156475370	-0.004416471
[2,]	-0.5755957	0.1326337	-0.48963489	0.641309408	-0.008819194
[3,]	-0.6033991	0.1147648	-0.24301098	-0.750771313	0.005166033
[4,]	-0.1550094	-0.6896256	0.01887635	0.008188685	-0.707082189
[5,]	-0.1609409	-0.6879558	0.01977337	0.022651335	0.707043709

کاهش بعد به کمک مقادیر ویژه :

```
> landa<-eigen(R)$values
```

```
> which(landa>1)
```

```
[1] 1 2
```

در تحلیل مولفه اصلی با استفاده از ماتریس همبستگی مولفه‌هایی که مقادیر ویژه بزرگتر از ۱ دارند را نگه می‌داریم و بقیه حذف می‌کنیم پس در اینجا بر خلاف نتیجه قبل کاهش بعد به ۲ بعد است.

تحلیل مولفه اصلی :

```
> EURUSD.pca.R<-princomp(covmat=R)
```

```
> summary(EURUSD.pca.R)
```

Importance of components:

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5
Standard deviation	1.5438999	1.4055827	0.70625785	0.36880325	0.076774190
Proportion of Variance	0.4767254	0.3951326	0.09976003	0.02720317	0.001178855
Cumulative Proportion	0.4767254	0.8718579	0.97161798	0.99882114	1.000000000

Loadings:

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6
WDIM	0.511		0.446		0.621	0.393
CIRCUM	0.561		-0.320		0.228	-0.723
FBEYE	0.462	0.147	-0.475	-0.473	-0.311	0.467
EYEHD	0.144	-0.664	-0.314	0.592		0.283
EARHD	0.110	-0.644	0.470	-0.488	-0.311	-0.129
JAW	0.421	0.339	0.392	0.431	-0.600	

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6
SS loadings	1.000	1.000	1.000	1.000	1.000	1.000
Proportion Var	0.167	0.167	0.167	0.167	0.167	0.167
Cumulative Var	0.167	0.333	0.500	0.667	0.833	1.000

در اینجا می‌بینیم که تحلیل مولفه اصلی به ما می‌گوید در کاهش بعد باید ۲ مولفه اصلی اول را نگه داریم و ۲ مولفه اصلی اول ۸۷ درصد کل واریانس را نمایش می‌دهند.