

1 Анализ конкретного временного ряда

2 Модельные ряды

SSA

Romanova

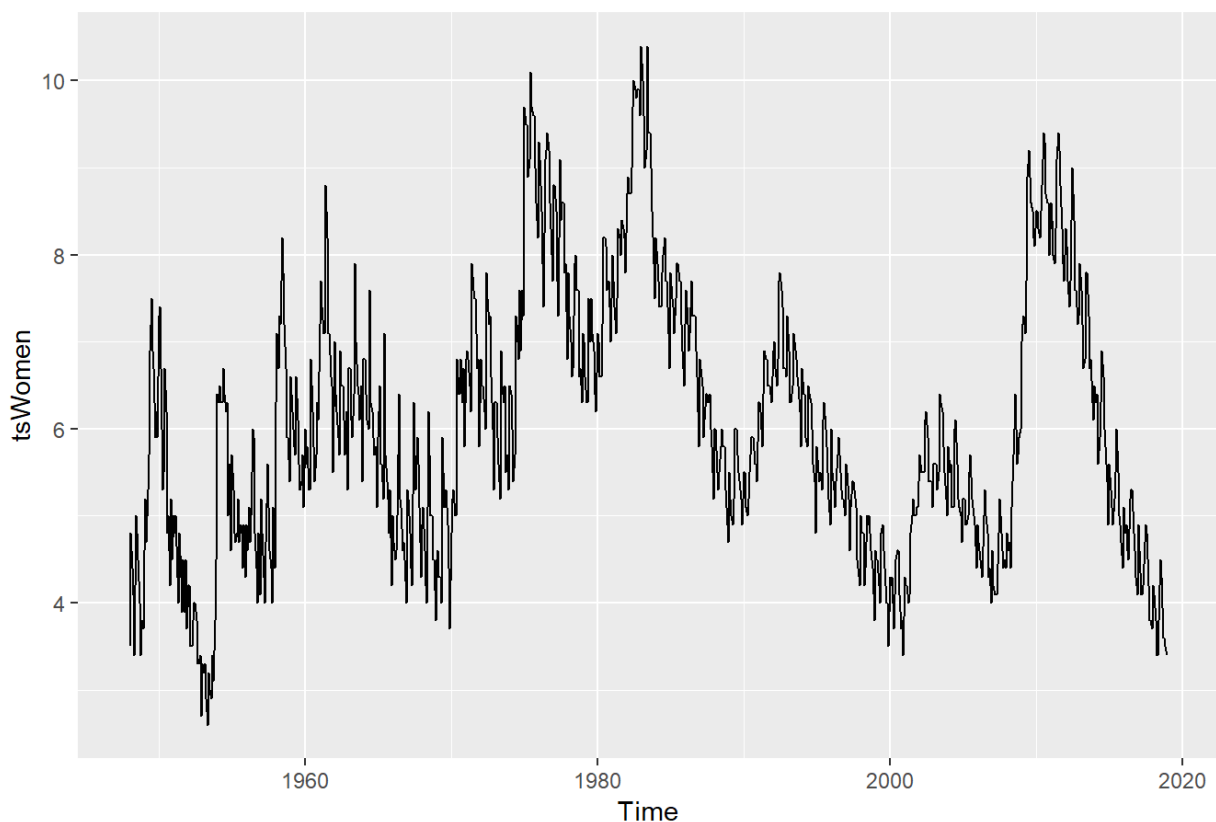
25 апреля 2019 г

1 Анализ конкретного временного ряда

1.1 Общие вещи

Рассматриваемый ряд – уровень женской безработицы с 1948 по 2018 год, измеренный в процентах неработающих женщин относительно всех женщин, которые имеют возможность работать.

```
tsWomen<-ts(TS.data$Value, frequency = 12, start = 1948)
autoplot(tsWomen)
```

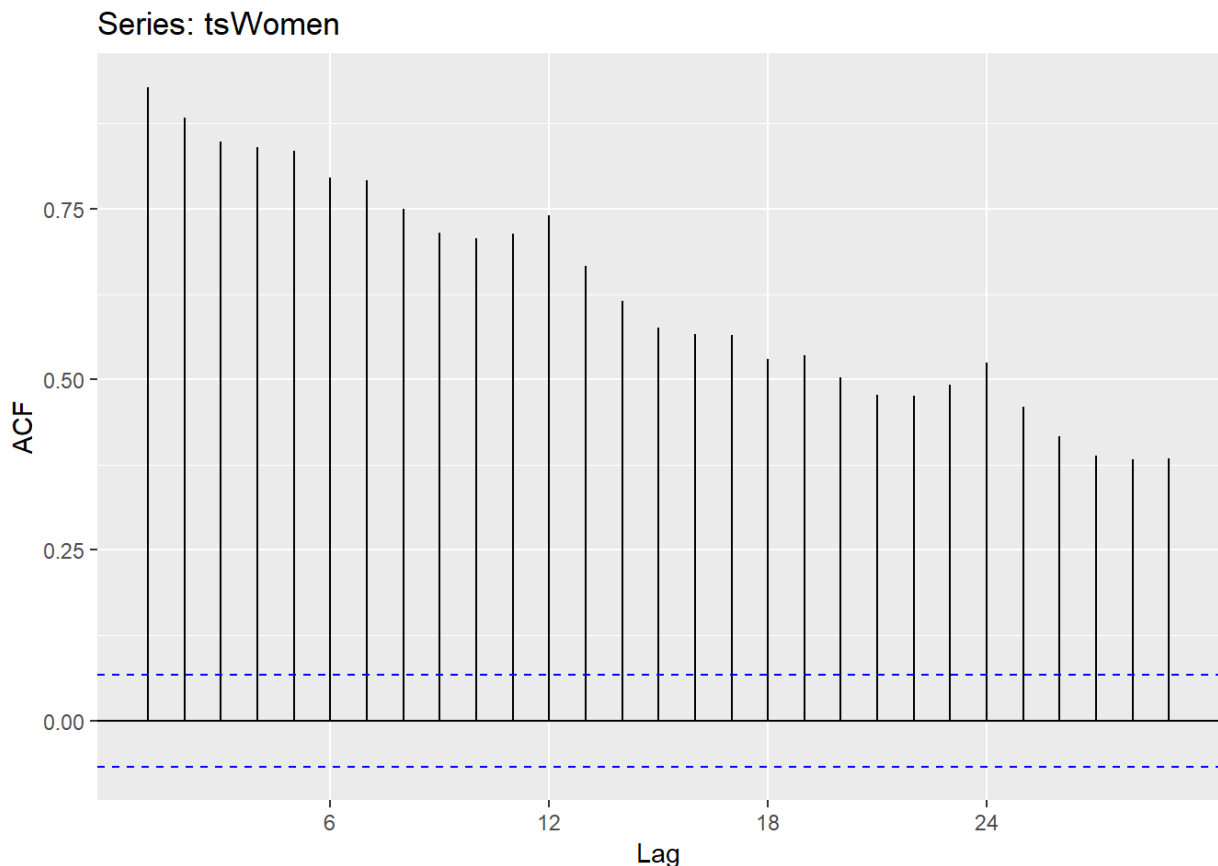


Модель ряда скорее всего нельзя назвать ни аддитивной, ни мультипликативной. На графике можно заметить, что колебания сильнее в левой части, в то время как значение тренда там ниже, чем в других местах (так что модель не мультипликативная), при этом колебания нельзя назвать постоянными, так что и не вполне аддитивная. (Это все будет более заметно на графике остатков после удаления тренда.)

Автокорреляция ряда – это его корреляция со сдвинутой версией самого себя. Это важный показатель (для недетерминированных рядов), потому что, он показывает, влияют ли предыдущие состояния (запаздывающие наблюдения) временного ряда на текущее состояние (линейно). Таким образом, по диаграмме автокорреляции можно увидеть, есть ли тренд (опять же линейный, так как при сильно нелинейном тренде автокорреляции могут быть и очень малы).

На диаграмме автокорреляции, если автокорреляция пересекает пунктирную синюю линию, это означает, что конкретное запаздывание значимо коррелирует с текущим рядом (пунктирная синяя линия – граница доверительной области равенства нулю автокорреляции).

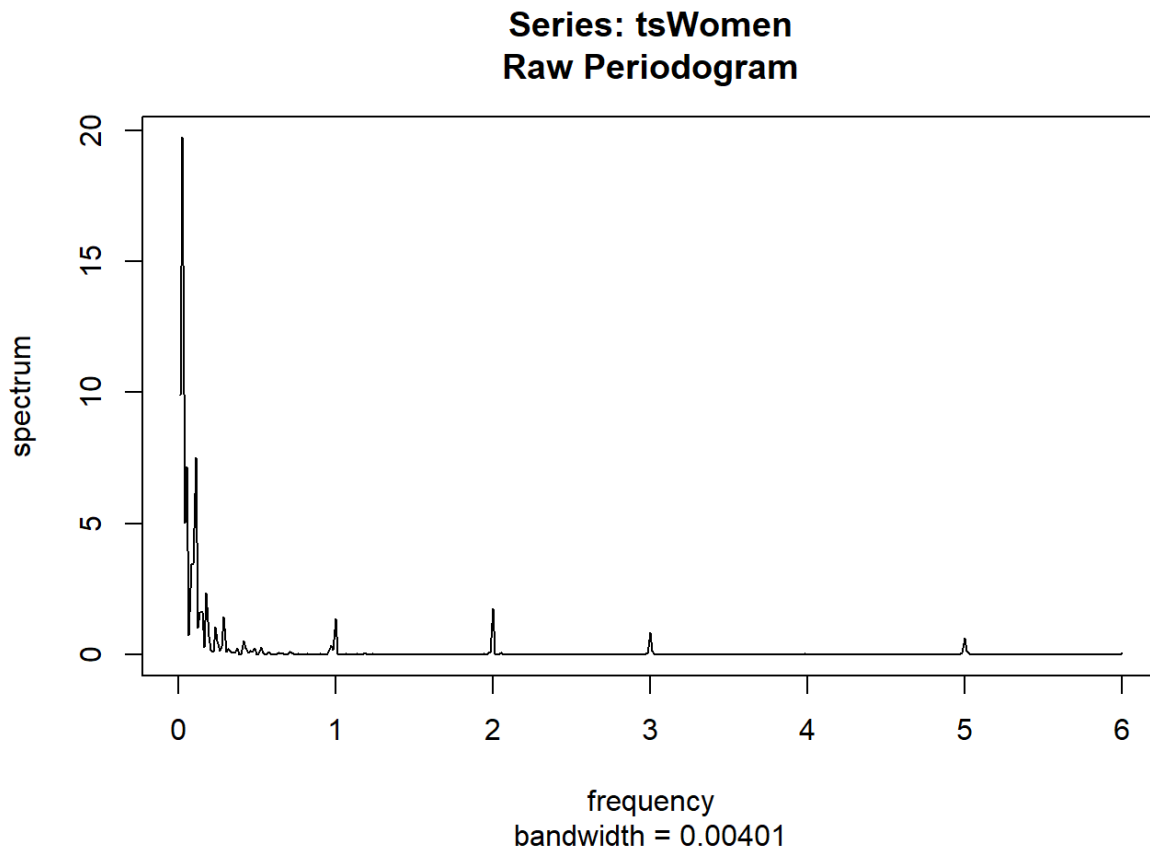
```
ggAcf (tsWomen)
```



Тут видно, что есть тренд и периодичность (по годам).

Периодограмма ряда в обычной шкале:

```
spec.pgram (tsWomen, log="no")
```



Присутствует периодичность по годам, полугодиям и кварталам. Также есть небольшой пик для периода 2.4.

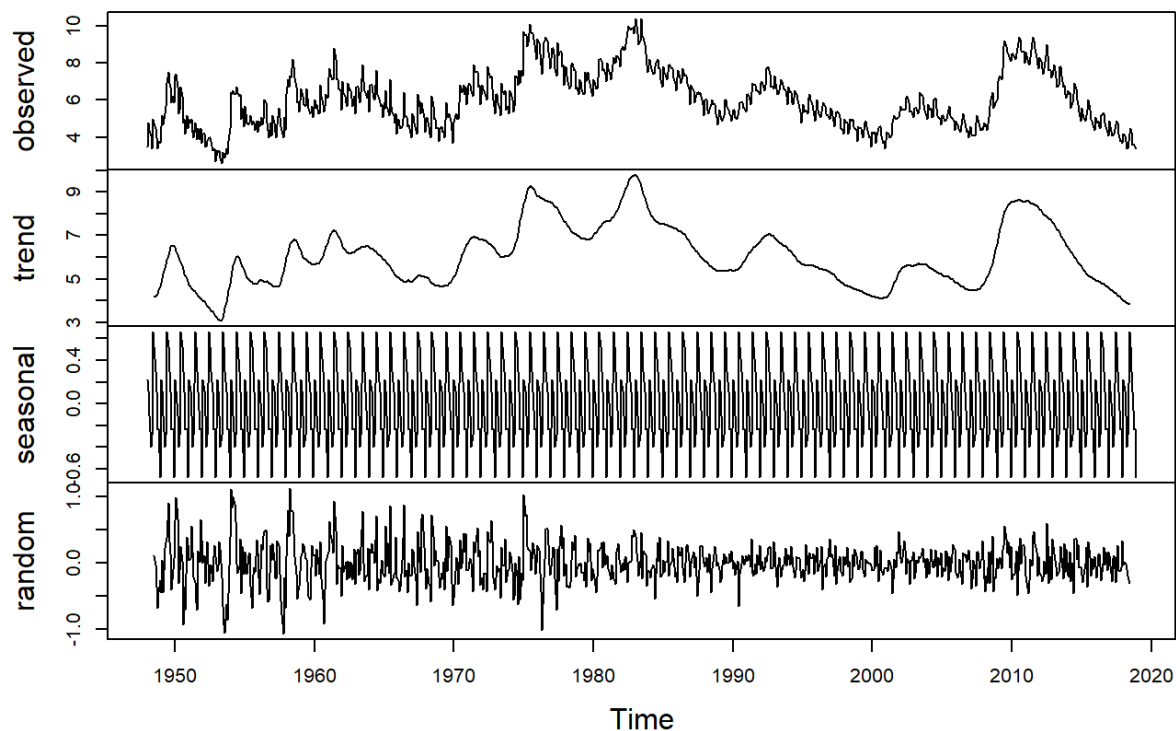
1.2 Разложение ряда

1.2.1 Классическое сезонное разложение

Классическое сезонное разложение реализуется функцией `decompose()`. В этой функции можно выбрать, аддитивная модель или мультипликативная. У нас, как мы обсуждали выше, ни то, ни другое. Но в силу того, что увеличение шума в левой части ряда не слишком численно большое, будем считать, что это допустимо – применять метод, ориентированный на аддитивную модель.

```
Women.decompose<-stats::decompose(tsWomen, type="additive", filter = c(1/24, rep(1/12, 11), 1/24))
plot(Women.decompose)
```

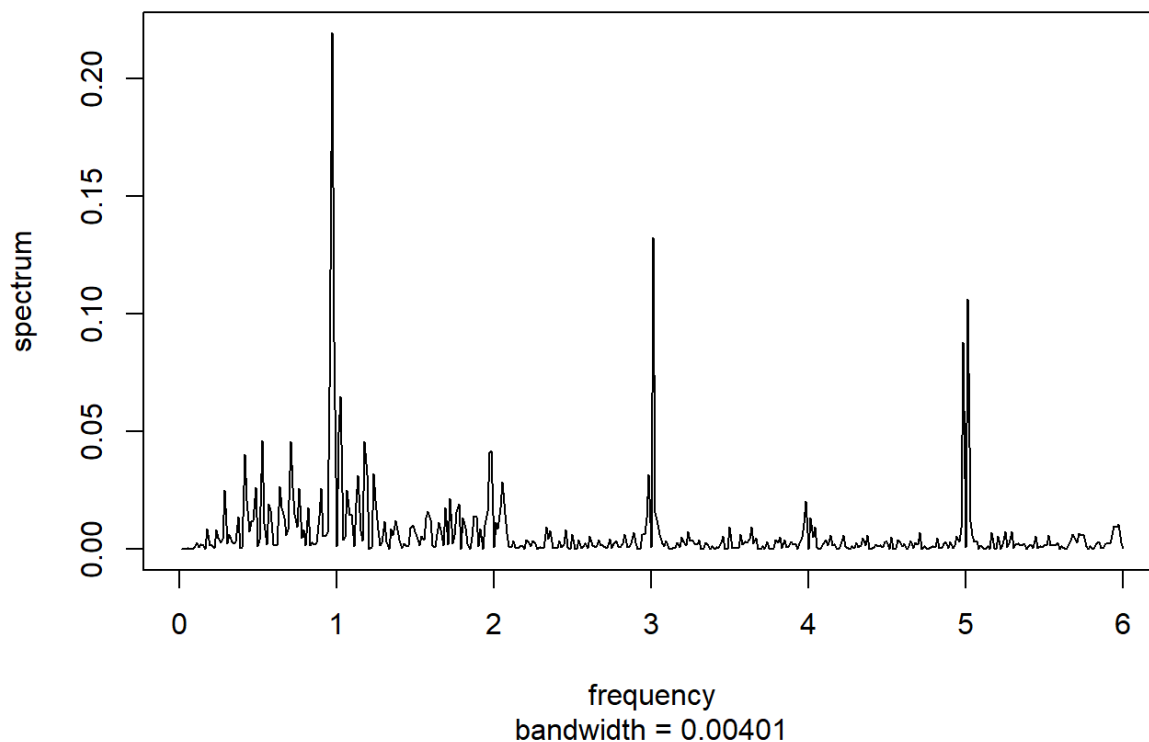
Decomposition of additive time series



Периодограмма шума:

```
Women.noise<-ts(tsWomen[-c(1:6,847:852)]-Women.decompose$seasonal[-c(1:6,847:852)]-Women.decompose$trend[-c(1:6,847:852)],frequency = 12)
spec.pgram(Women.noise, log="no")
```

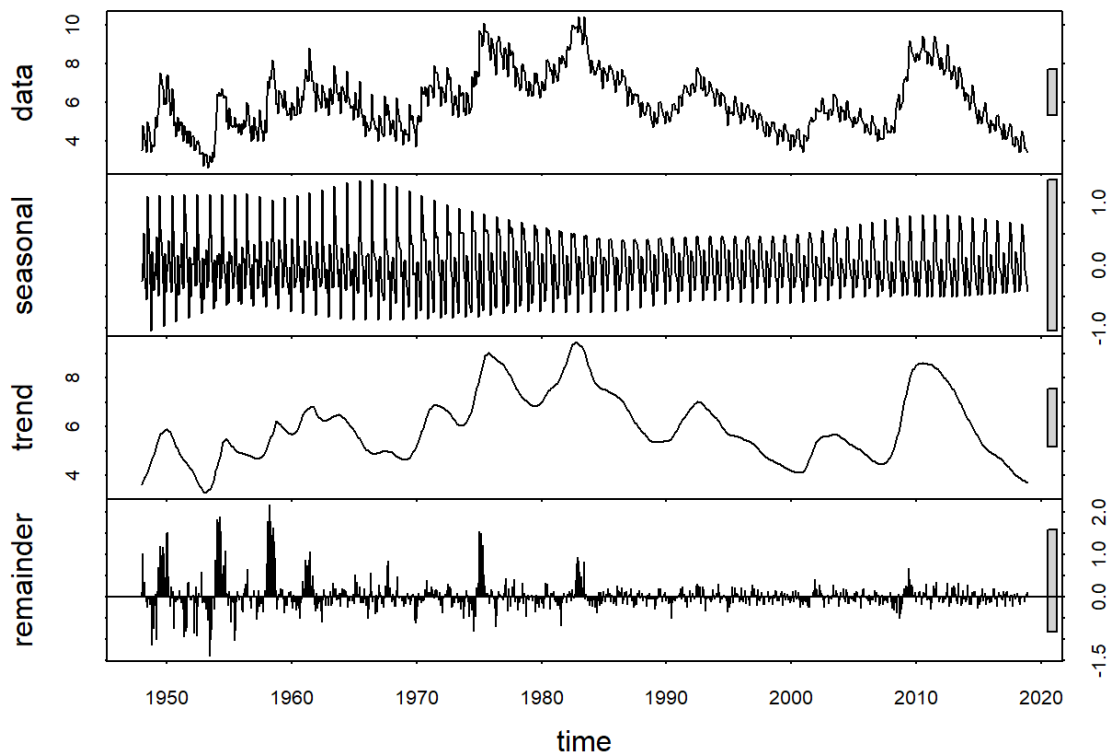
Series: Women.noise Raw Periodogram



Вот так вот. Периодичность в шуме есть и довольно приличная.

1.2.2 Усовершенствованное классическое сезонное разложение

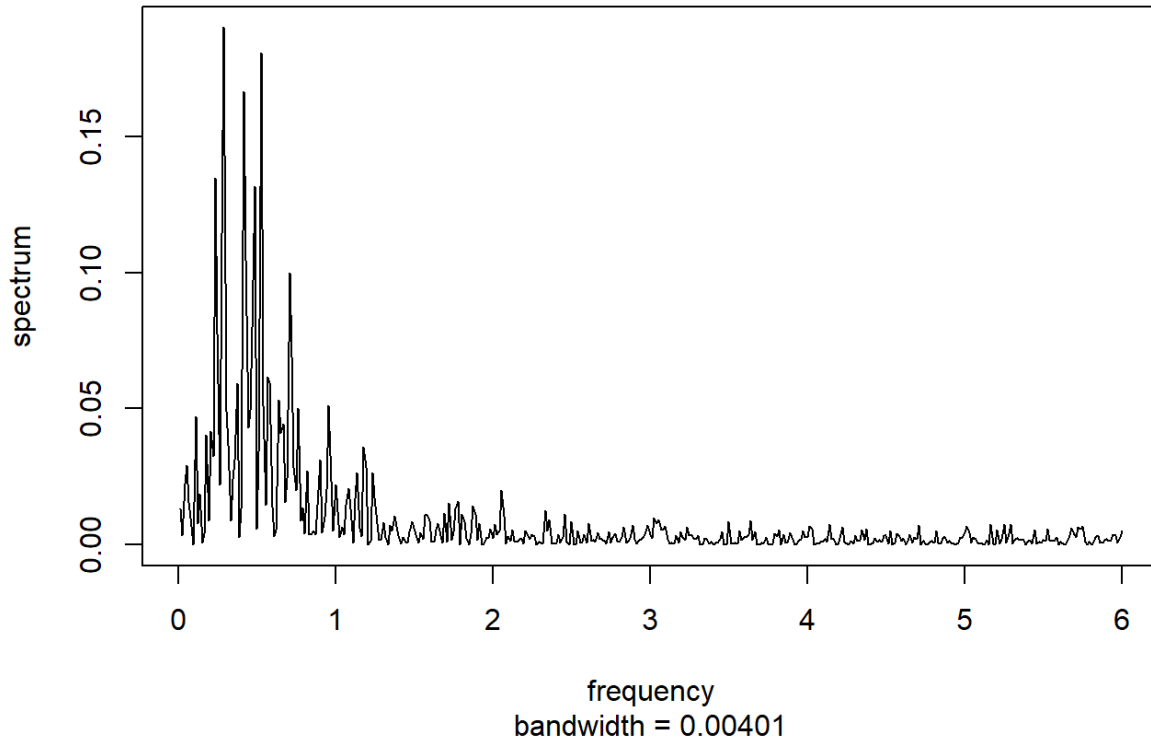
```
Women.stl<-stl(tsWomen,s.window = 15,s.degree = 1, l.window = 36,robust = TR
  UE, outer=10)
plot(Women.stl)
```



Периодограмма шума:

```
Women.noise.stl<-ts((Women.stl$time.series)[,"remainder"],frequency = 12)
spec.pgram(Women.noise.stl, log="no")
```

Series: Women.noise.stl Raw Periodogram



От периодичности избавиться получилось. Однако в периодограмме шума выделяются низкие частоты. Либо мы оставили часть тренда, либо, что вполне вероятно, мы имеем дело с красным шумом. Скорее всего второй вариант...

Таким образом, надежд на то, что остаток – белый шум, особо нет.

```
autocor.noise.stl<- autocorrelations(Women.noise.stl, maxlag = 30)
#plot(autocor.noise.stl)
whiteNoiseTest(autocor.noise.stl, h0 = "iid",method = "LjungBox",interval=FALSE) # "LiMcLeod" (default), "LjungBox", "BoxPierce".
```

```
## $test
##          ChiSq DF      pvalue
## [1,] 389.5655  1 1.029106e-86
## attr(,"method")
## [1] "LjungBox"
##
## $ci
##      int
## [1,]  0 0
## attr(,"level")
## [1] FALSE
```

Не белый шум.

1.2.3 SSA

Посмотрим на график собственных чисел всего ряда.

```
Women.ssa<-ssa(tsWomen,L=420)
plot(Women.ssa)
```

Component norms

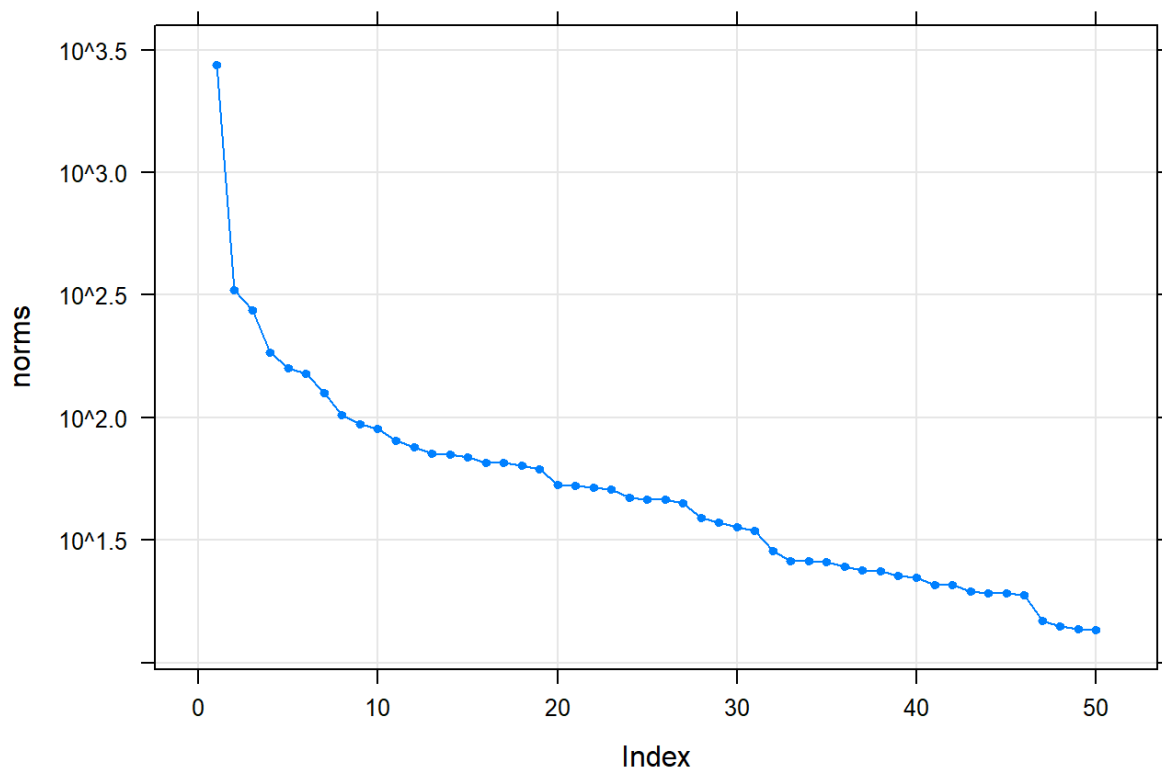


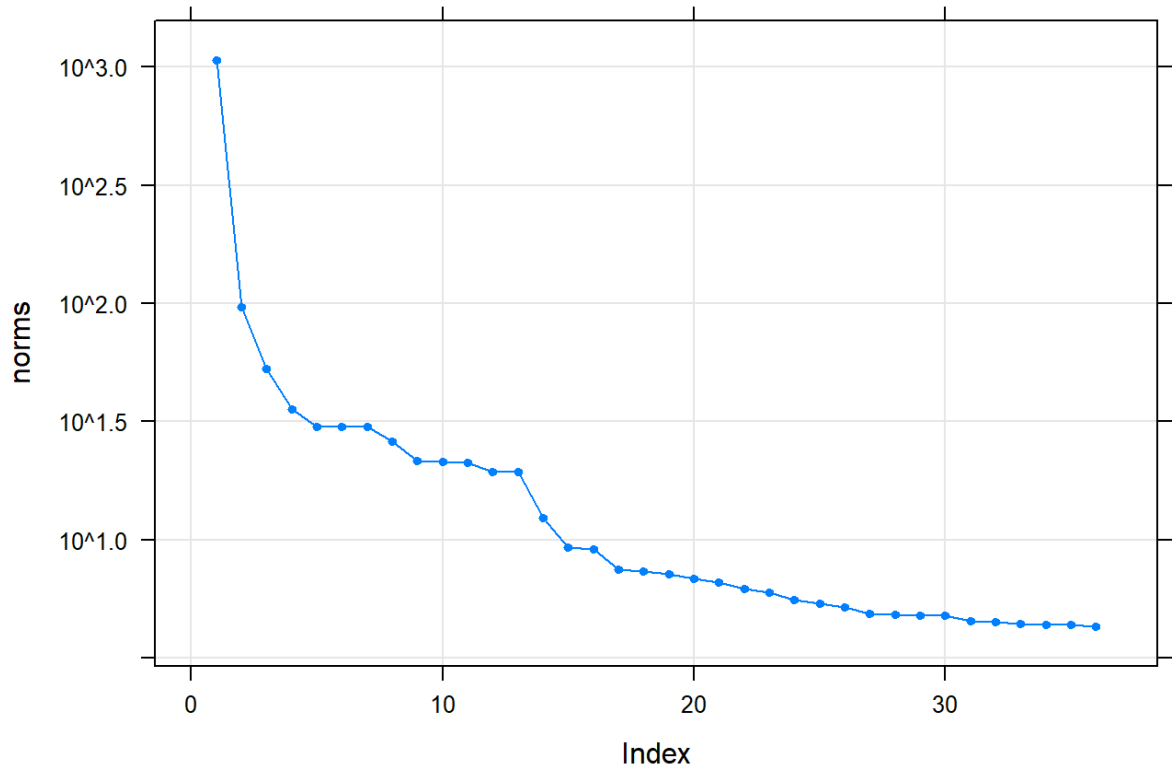
График не очень приятный. Будет сложно отделить сигнал от шума (собственные числа близки).

Как показали эксперименты, сразу делать `ssa` для нашего ряда – это ужас. Тренд вылезает всюду и вне очереди (как мы и ожидали, посмотрев на предыдущий график, все перемешалось. Этого можно было ожидать и на основе периодограммы, там все пики примерно одинаковой величины). И рассматривать нам придется очень много собственных векторов. Поэтому лучше сначала выделим тренд, а потом остаток разделим на периодичность и шум.

Будем применять последовательный SSA. Для выделения тренда возьмем небольшую длину окна.

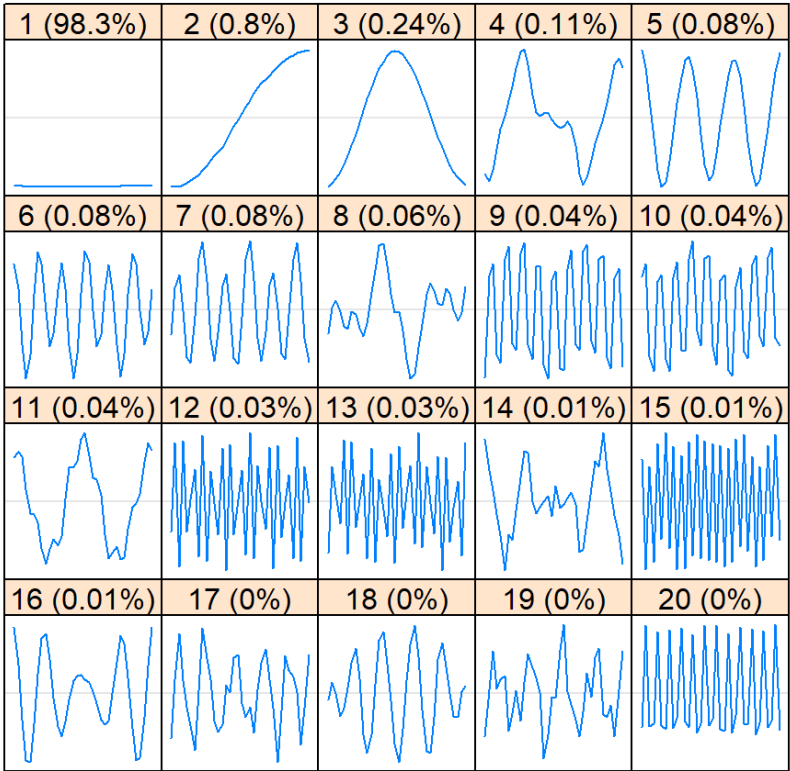
```
trend.SSA<-ssa(tsWomen,L=36)
plot(trend.SSA)
```

Component norms



```
plot(trend.SSA,type="vectors",idx=1:20)
```

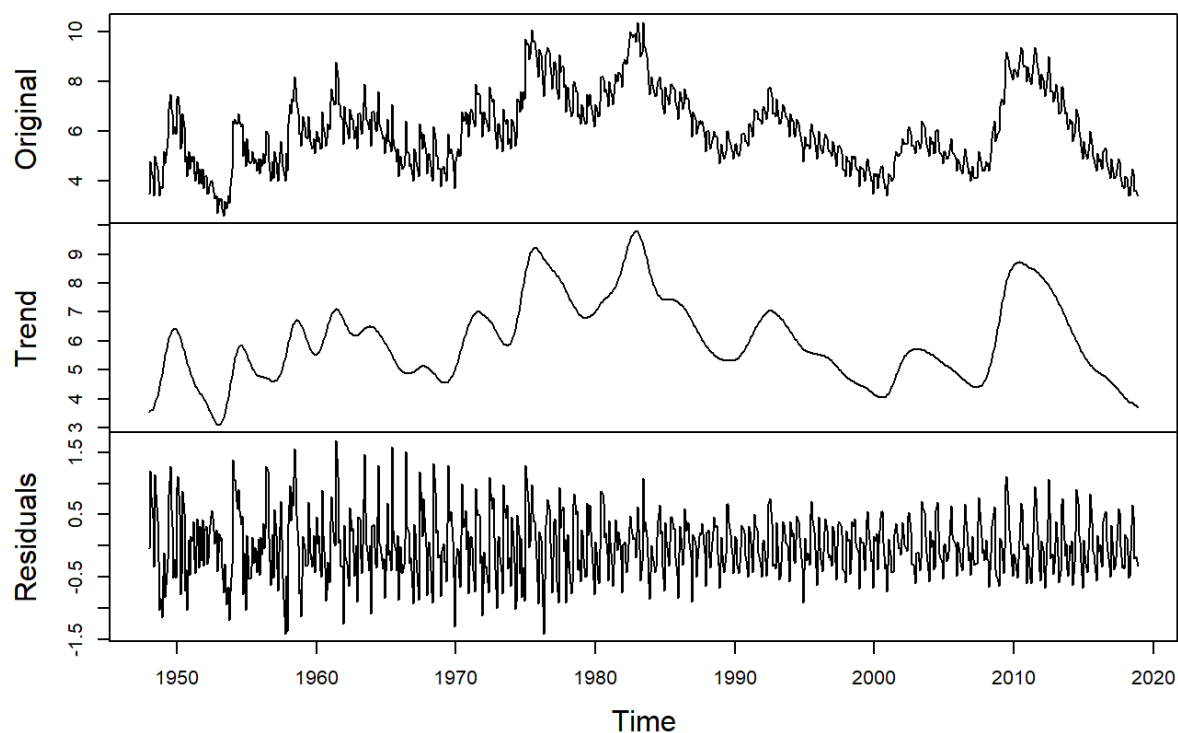
Eigenvectors



Первые три компоненты есть тренд.


```
plot(reconstruct(trend.SSA, groups = list(Trend=c(1:3))),
     plot.method = "xyplot", layout = c(1,7),
     add.residuals = FALSE, add.original = FALSE))
```

Reconstructed Series

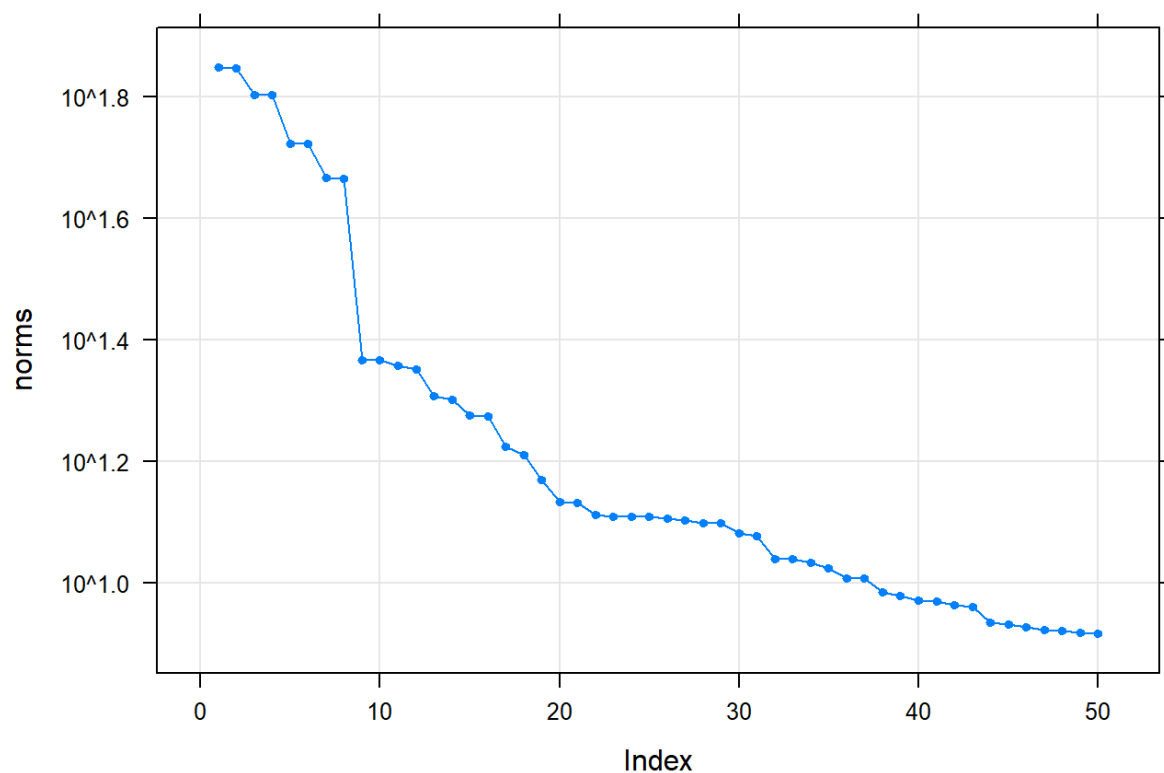


```
trend.ssa.reconstruct<-reconstruct(trend.SSA, groups = list(Trend = 1:3))
Women.trend<-trend.ssa.reconstruct$Trend
```

Переходим к анализу ряда без тренда.

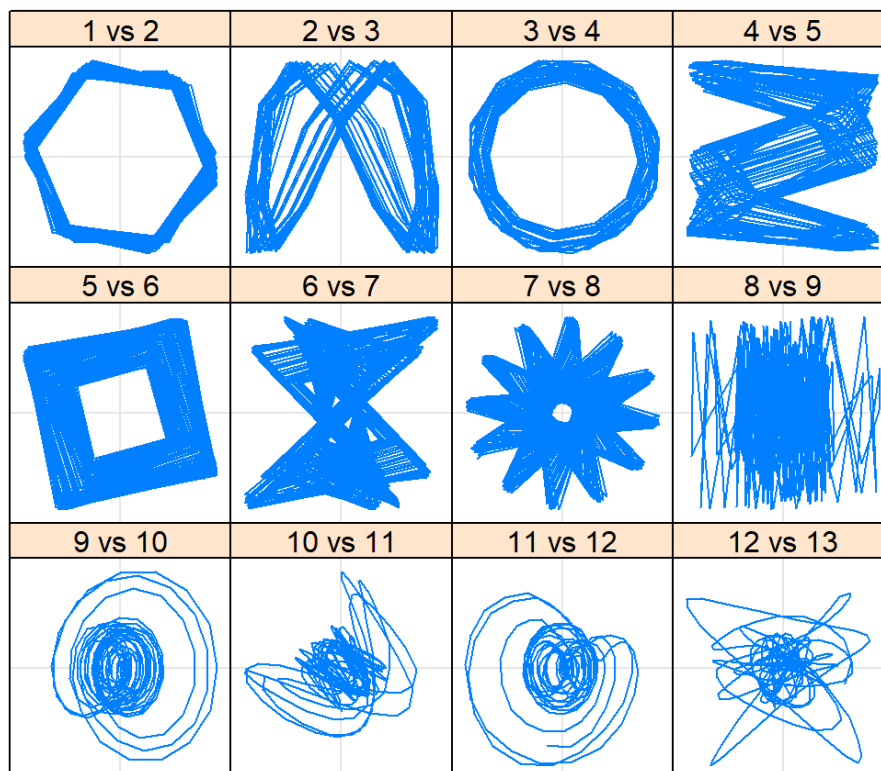
```
Women.detrend<-ts(tsWomen-Women.trend,frequency = 12)
Women.detrend.ssa<-ssa(Women.detrend,L=420)
plot(Women.detrend.ssa)
```

Component norms



```
plot(Women.detrend.ssa, type = "paired", idx = 1:12, plot.contrib = FALSE)
```

Pairs of eigenvectors



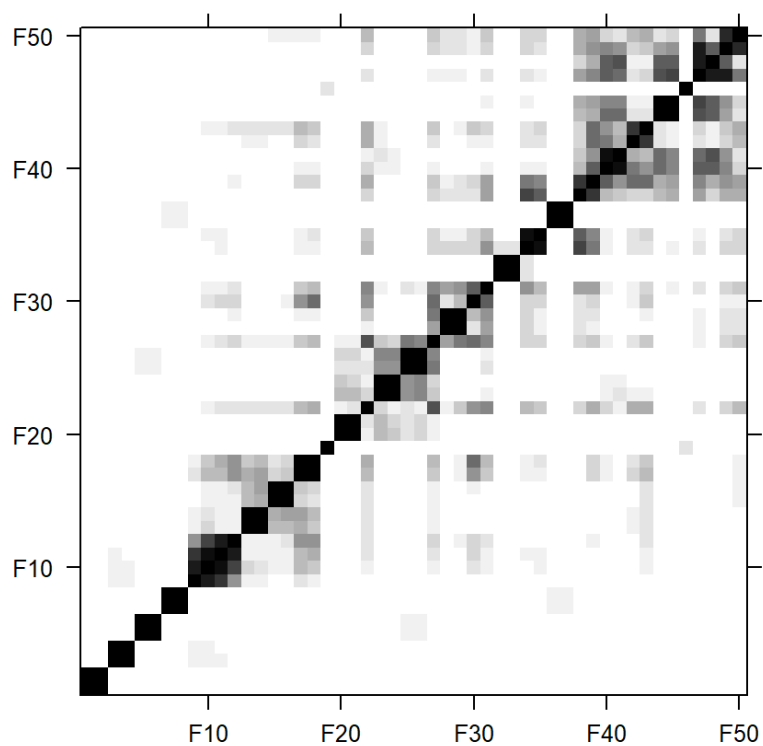
```
parestimate(Women.detrend.ssa, groups = list(1:2, 3:4, 5:6, 7:8, 9:10, 11:12), method = "esprit")
```

```
## $F1
##   period    rate |   Mod   Arg |    Re    Im
##     6.003 -0.000503 | 0.99950  1.05 | 0.50015  0.86536
##    -6.003 -0.000503 | 0.99950 -1.05 | 0.50015 -0.86536
##
## $F2
##   period    rate |   Mod   Arg |    Re    Im
##    12.033 -0.000421 | 0.99958  0.52 | 0.86638  0.49854
##   -12.033 -0.000421 | 0.99958 -0.52 | 0.86638 -0.49854
##
## $F3
##   period    rate |   Mod   Arg |    Re    Im
##     3.999 -0.001438 | 0.99856  1.57 | -0.00045  0.99856
##    -3.999 -0.001438 | 0.99856 -1.57 | -0.00045 -0.99856
##
## $F4
##   period    rate |   Mod   Arg |    Re    Im
##     2.399 -0.001808 | 0.99819  2.62 | -0.86496  0.49823
##    -2.399 -0.001808 | 0.99819 -2.62 | -0.86496 -0.49823
##
## $F5
##   period    rate |   Mod   Arg |    Re    Im
##    15.100 -0.012844 | 0.98724  0.42 | 0.90300  0.39903
##   -15.100 -0.012844 | 0.98724 -0.42 | 0.90300 -0.39903
##
## $F6
##   period    rate |   Mod   Arg |    Re    Im
##    19.595 -0.019228 | 0.98096  0.32 | 0.93095  0.30919
##   -19.595 -0.019228 | 0.98096 -0.32 | 0.93095 -0.30919
```

Первые 4 периодики не вызывают сомнений. Две последние и выглядят странно, и периоды у них странные.

```
plot(wcor(Women.detrend.ssa, groups = 1:50),
     scales = list(at = c(10, 20, 30, 40, 50)))
```

W-correlation matrix

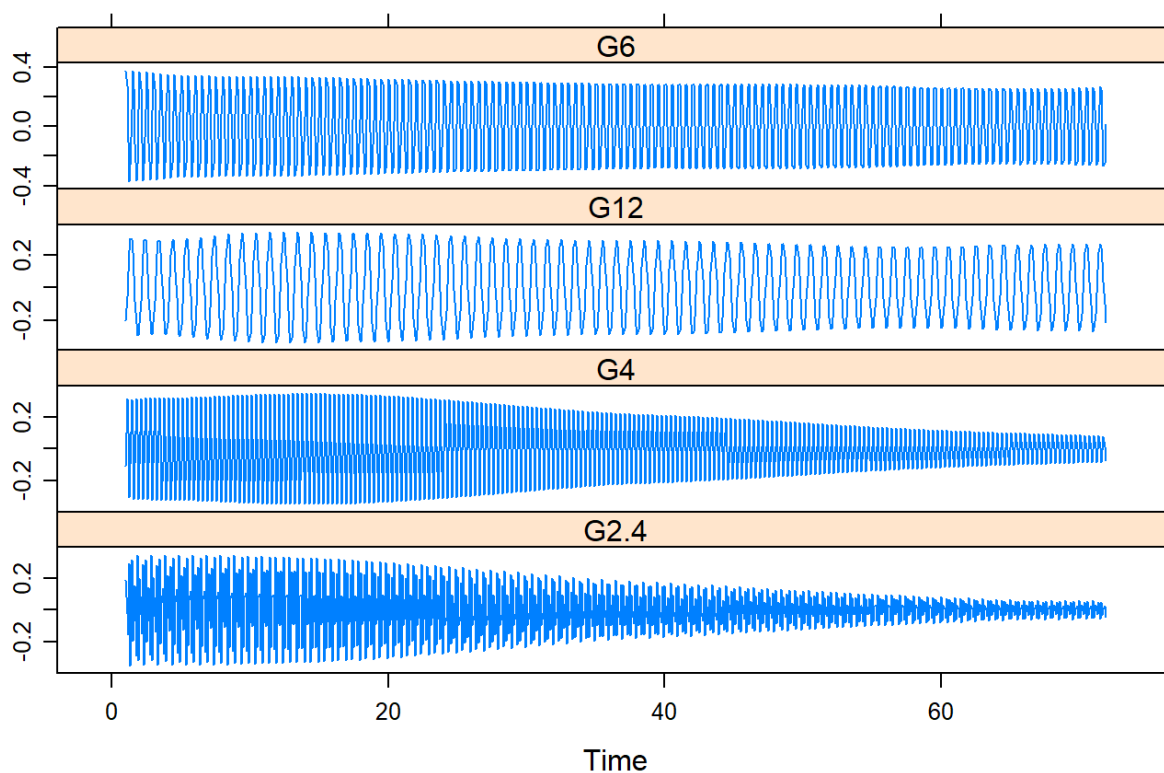


Также выделяются 4 периодики.

Reconstructed series:

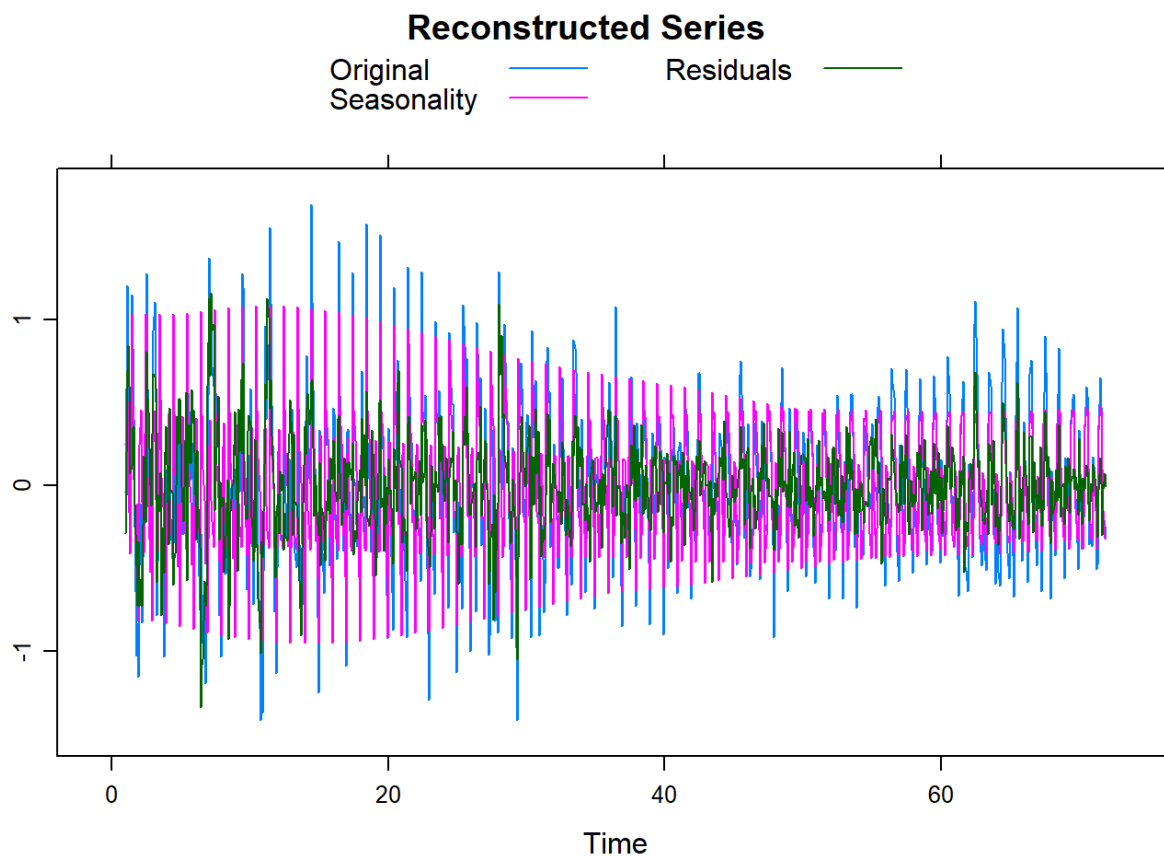
```
plot(reconstruct(Women.detrend.ssa, groups = list(G6 = 1:2, G12 = 3:4, G4=5:
6,G2.4=7:8)),
  plot.method = "xyplot", layout = c(1,4),
  add.residuals = FALSE, add.original = FALSE)
```

Reconstructed Series



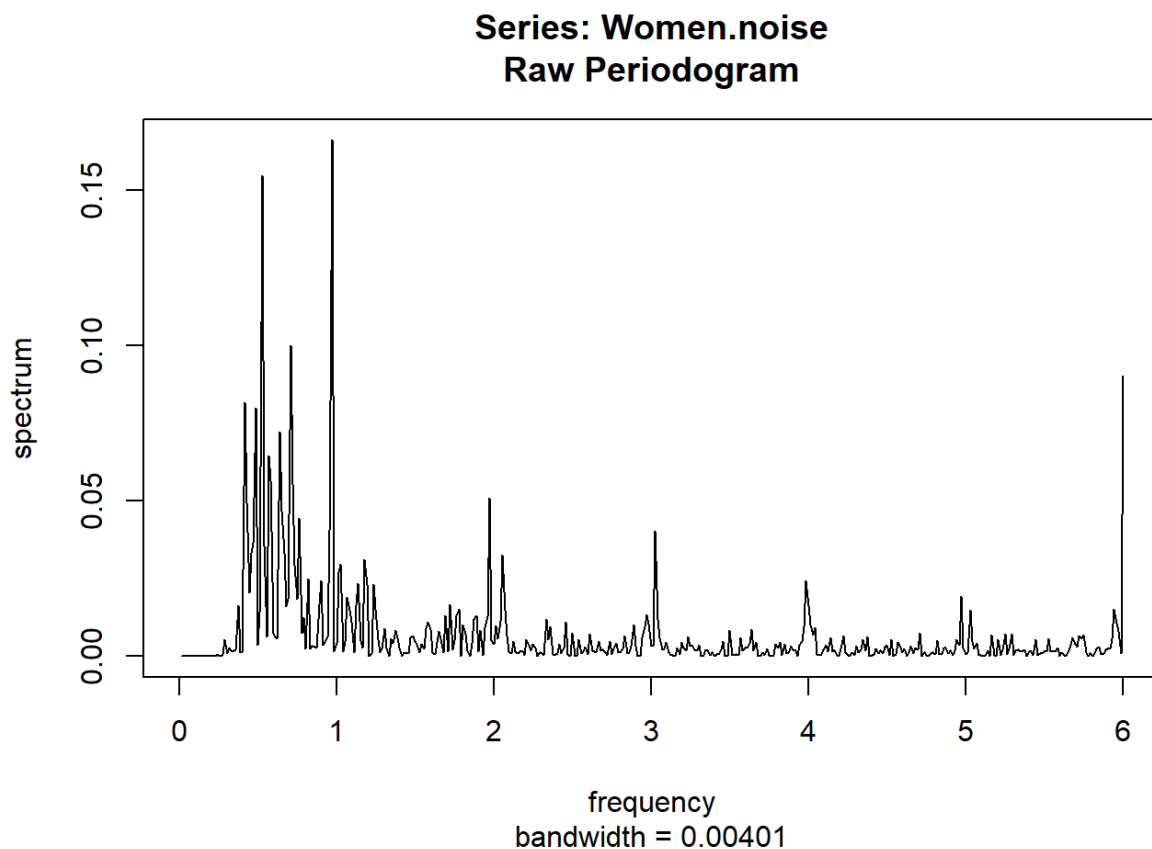
Берем 4 периодики в сезонность. Периоды: 6, 12, 4, 2.4.

```
Women.detrend.reconstruct <- reconstruct(Women.detrend.ssa,
                                          groups = list(Seasonality = 1:8))
#plot(Women.detrend.reconstruct , add.residuals = FALSE, add.original = TRUE,
#     plot.method = "xyplot",
#     superpose = TRUE, auto.key = list(columns = 2))
plot(Women.detrend.reconstruct , add.residuals = TRUE, add.original = TRUE,
      plot.method = "xyplot",
      superpose = TRUE, auto.key = list(columns = 2))
```



Периодограмма шума:

```
Women.noise<-ts(Women.detrend-Women.detrend.reconstruct$Seasonality,frequency = 12)
spec.pgram(Women.noise, log="no")
```



Небольшая часть периодичности осталась. И так как есть подозрения насчет того, что мы имеем дело с красным шумом, проверять гипотезу о том, что остаток – белый шум, – бесполезно.

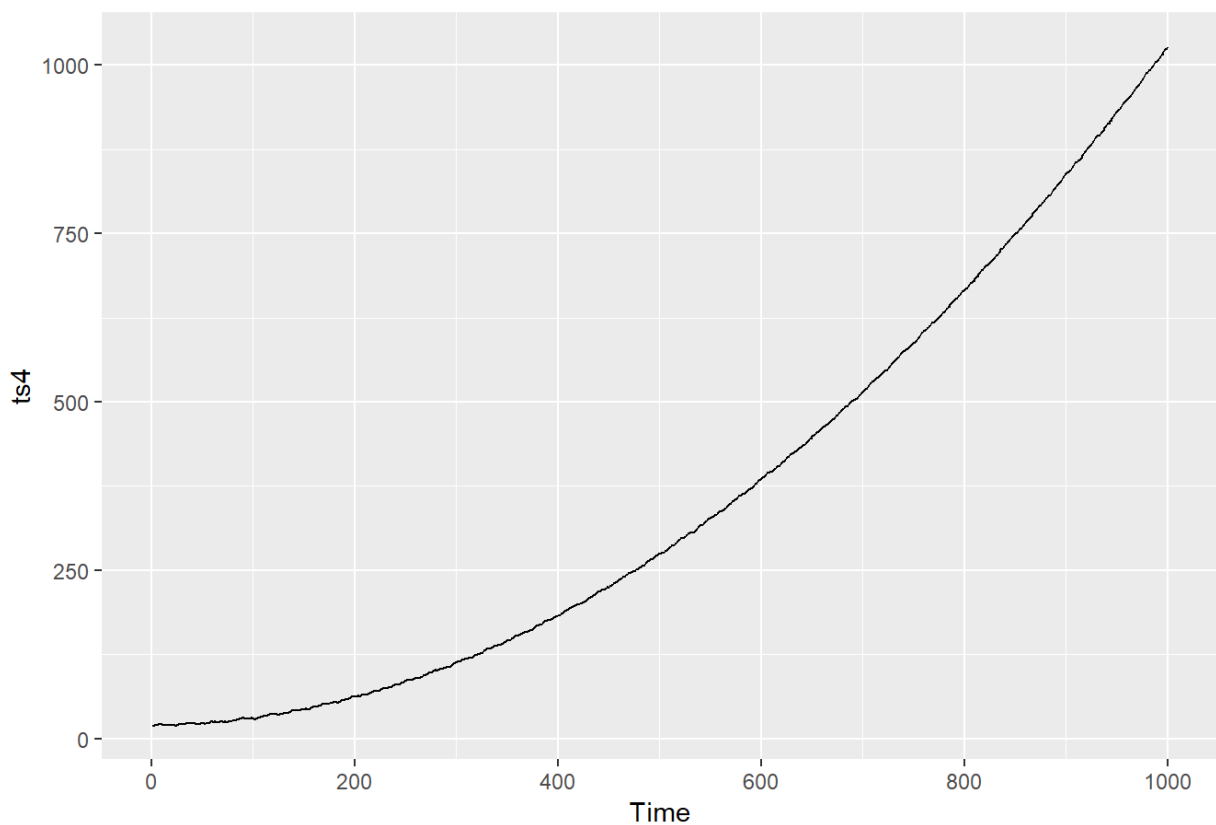
2 Модельные ряды

Модельные данные (COL4, COL10).

```
ts.data<-read.table("stud2005.csv",header = TRUE,dec=".",sep = ";")  
ts4<-ts(ts.data$COL4)  
ts10<-ts(na.omit(ts.data$COL10))
```

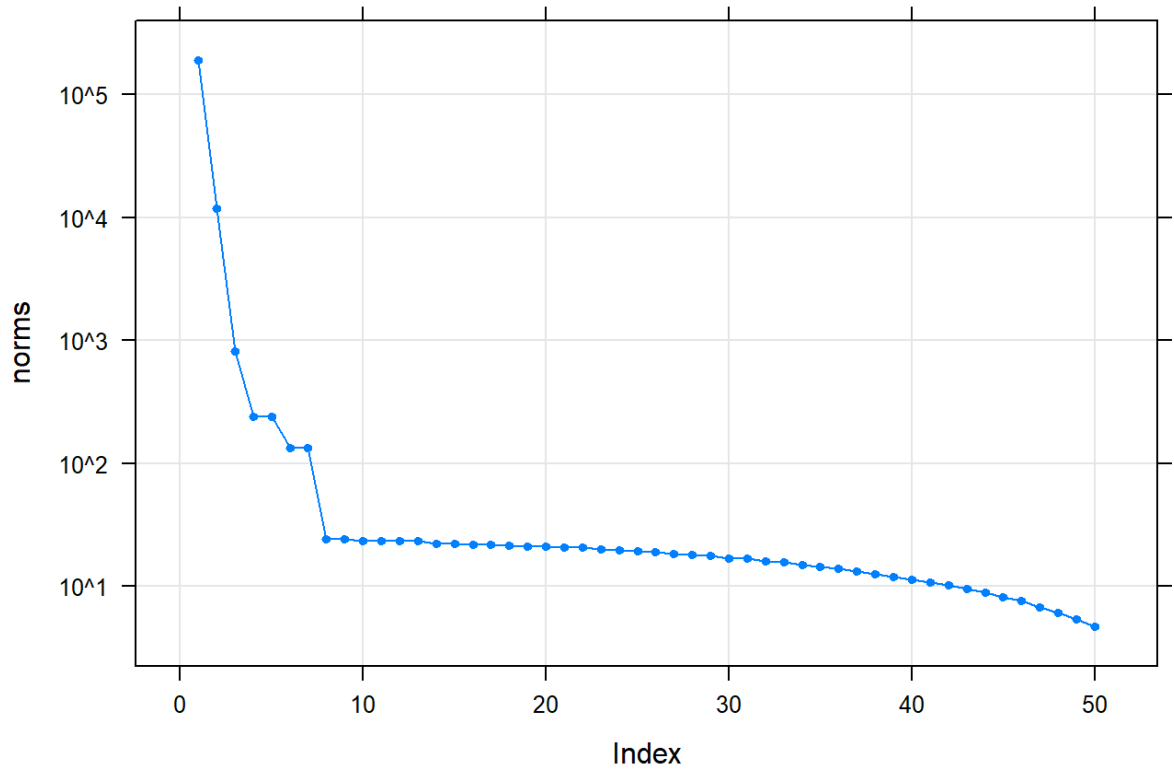
Разберемся сначала с рядом COL4.

```
autoplot(ts4)
```



```
ssa4<-ssa(ts4, L=500)  
plot(ssa4)
```

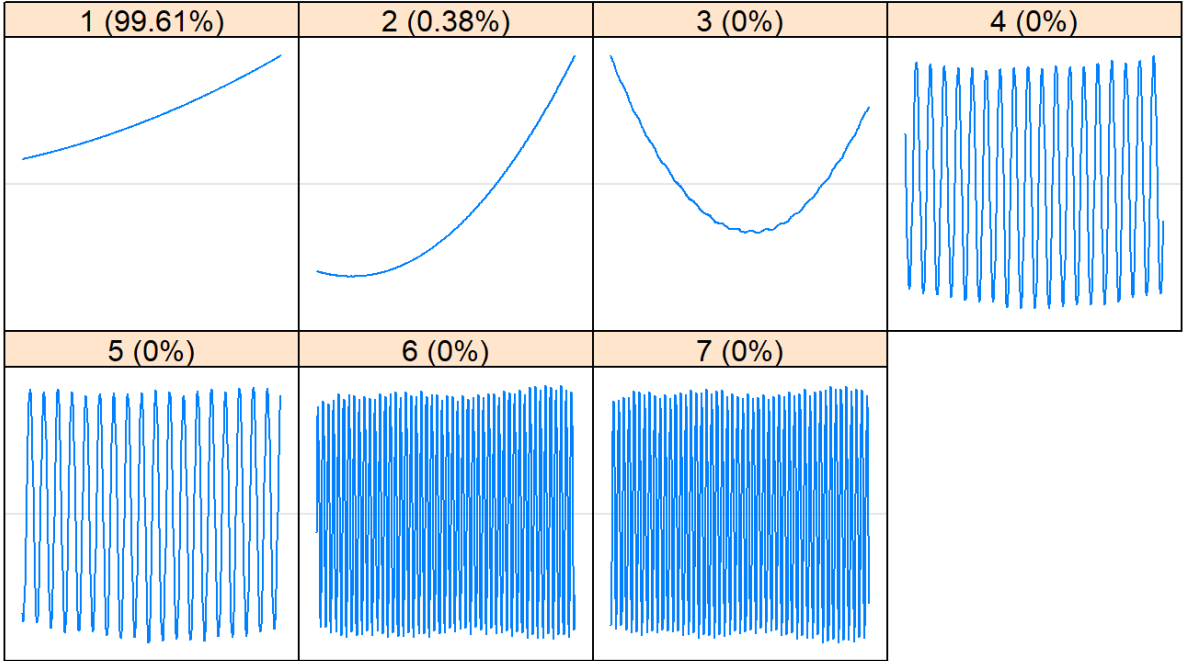
Component norms



Уже по этому графику можно увидеть пары одинаковых собственных чисел (4-5 и 6-7), а также то, что, начиная с 8й компоненты, собственные числа мало отличаются. Возможно, стоит рассматривать как раз 7 первых компонент.

```
plot(ssa4, type = "vectors", idx = 1:7)
```

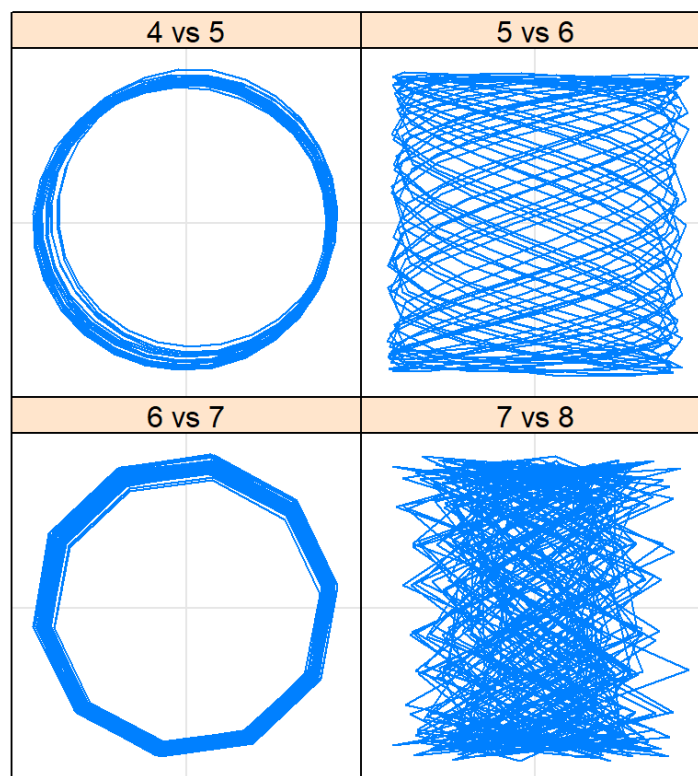
Eigenvectors



Первые три собственных вектора соответствуют тренду. Рассмотрим попарно остальные.


```
plot(ssa4, type = "paired", idx = 4:7, plot.contrib = FALSE)
```

Pairs of eigenvectors



Видим, что векторы 4-5 и 6-7 образуют пары. 6-7 с периодом 10, а 4-5 не понятно, какой-то большой период, уточним:

```
print(parestimate(ssa4, groups = list(4:5),
                  method = "esprit"))
```

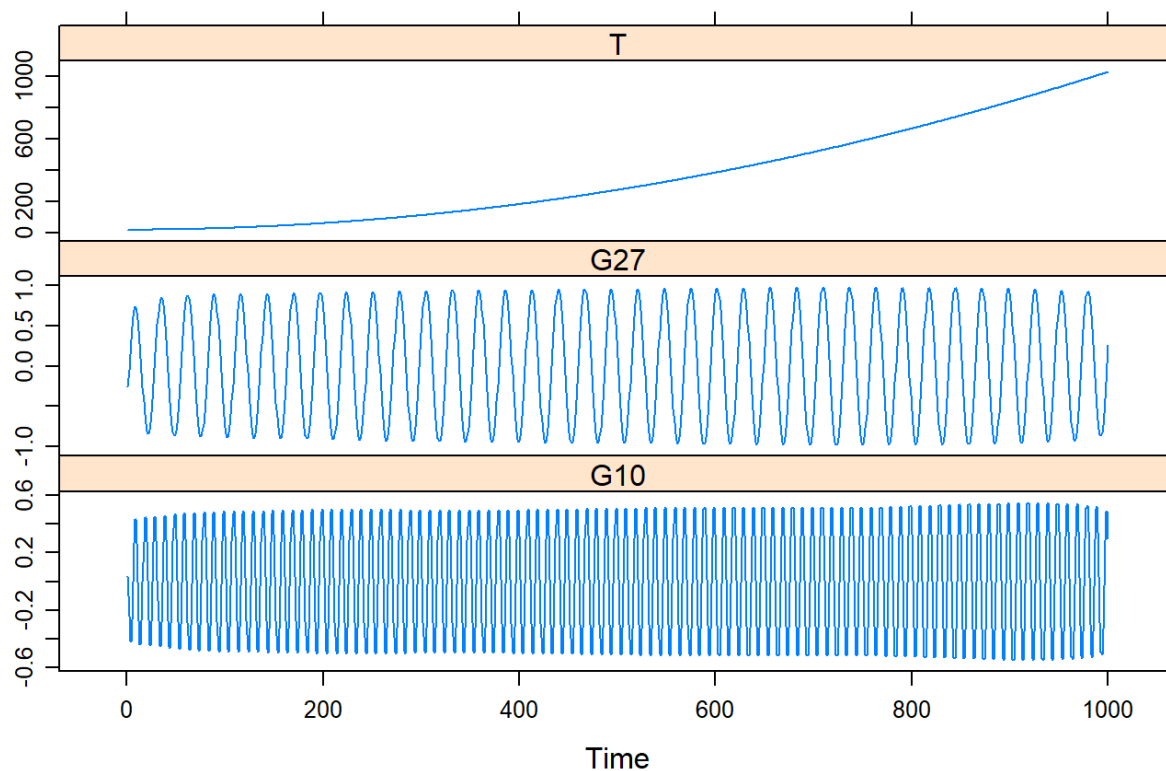
```
##    period    rate |   Mod   Arg |    Re    Im
##    26.987  0.000111 | 1.00011  0.23 | 0.97313  0.23075
##   -26.987  0.000111 | 1.00011 -0.23 | 0.97313 -0.23075
```

Период пары 4-5 равен 27.

Reconstructed series:

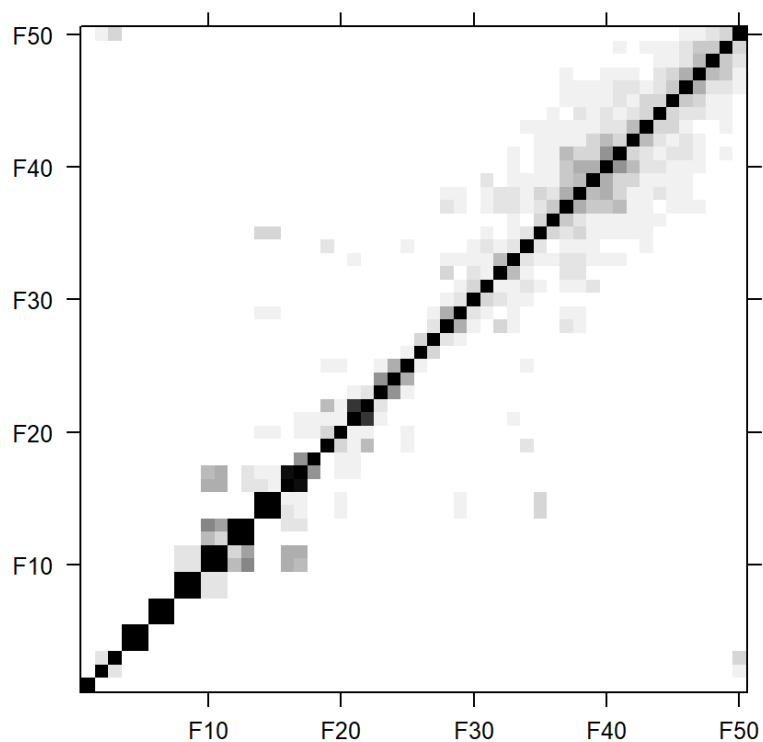
```
plot(reconstruct(ssa4, groups = list(T=1:3, G27 = 4:5, G10 = 6:7)),
     plot.method = "xyplot", layout = c(1,3),
     add.residuals = FALSE, add.original = FALSE)
```

Reconstructed Series



```
plot(wcor(ssa4, groups = 1:50),
     scales = list(at = c(10, 20, 30, 40, 50)))
```

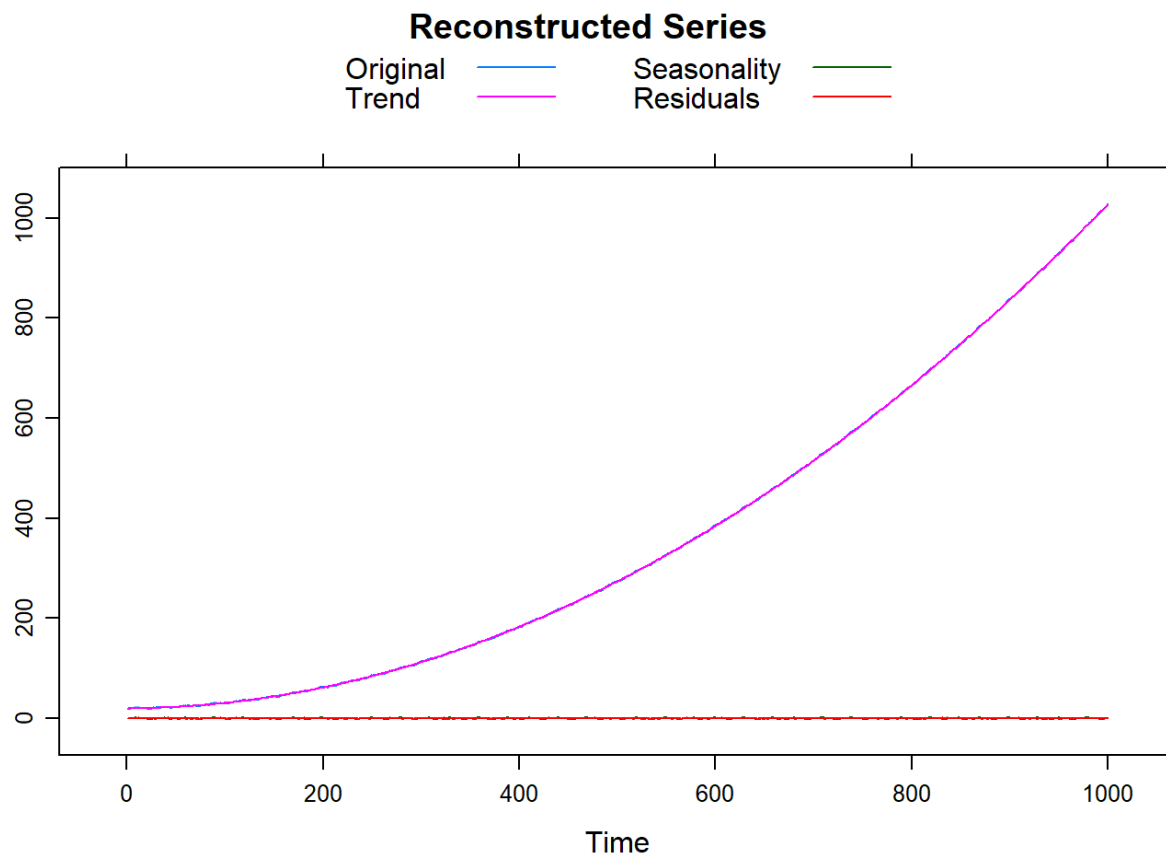
W-correlation matrix



Первые три квадрата – тренд (не коррелируют с остальными...почти (2 и 3 коррелируют друг с другом немного)). Следующие два квадрата – 2 гармоник (синусы сильно коррелируют с соответствующими косинусами в паре и не коррелируют ни с чем остальным). Остальное – шум.

Положим первые 3 компоненты в тренд, 4:7 – в периодичность.

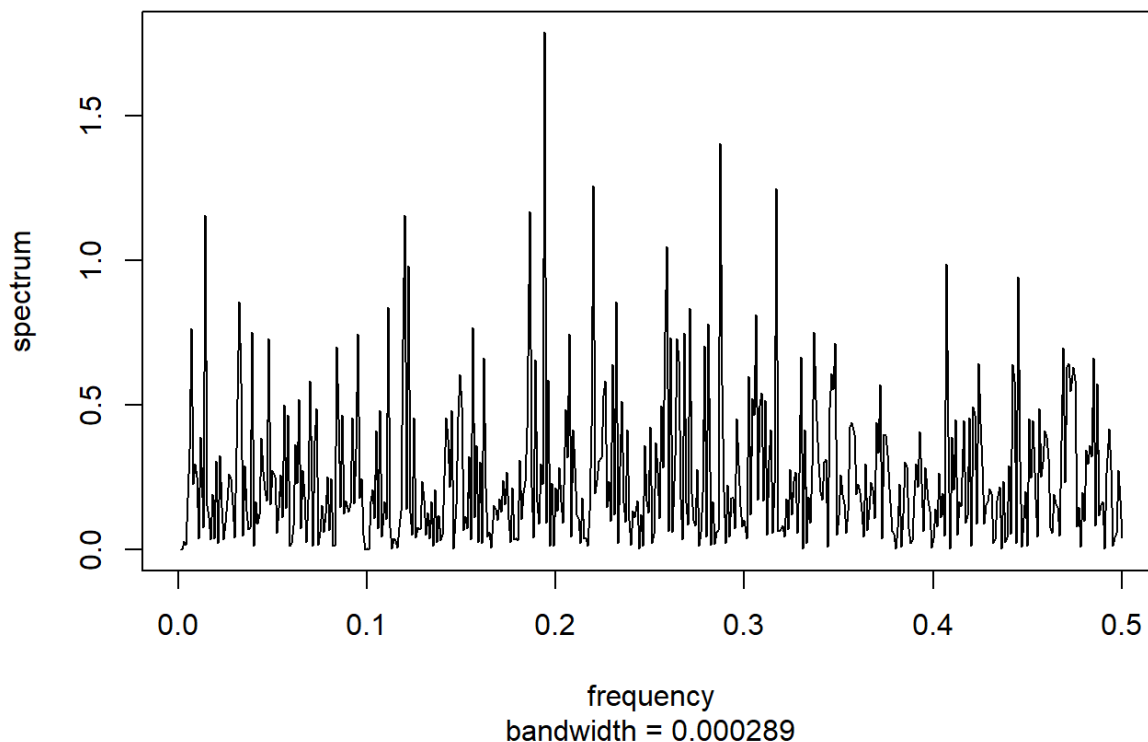
```
reconstruct4<- reconstruct(ssa4,
                           groups = list(Trend = 1:3,
                                           Seasonality = 4:7))
plot(reconstruct4, add.residuals = TRUE, add.original = TRUE,
     plot.method = "xyplot",
     superpose = TRUE, auto.key = list(columns = 2))
```



Периодограмма остатков:

```
noise4<-ts4-reconstruct4$Trend-reconstruct4$Seasonality
spec.pgram(noise4, log="no")
```

Series: noise4 Raw Periodogram



Внешне похоже на белый шум. Проверим.

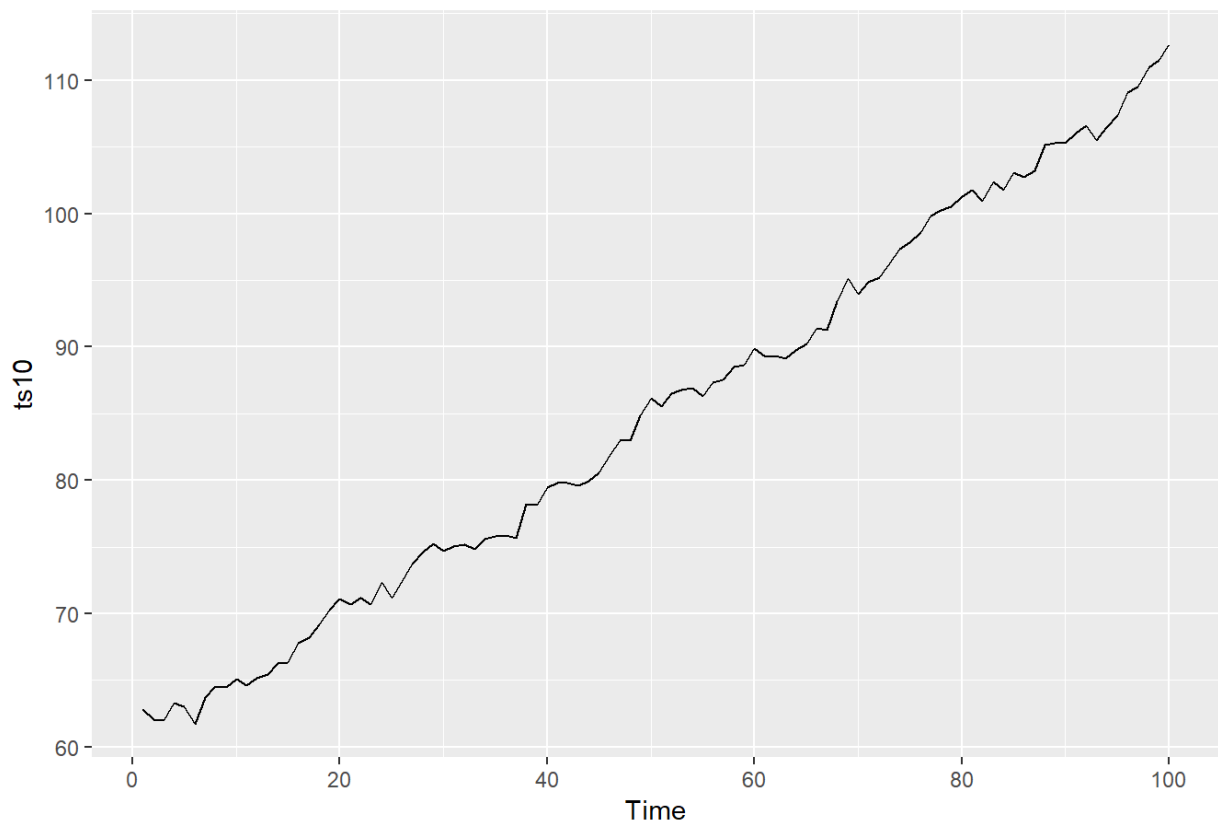
```
autocor.noise<- autocorrelations(noise4, maxlag = 30)
#plot(autocor.noise)
whiteNoiseTest(autocor.noise, h0 = "iid",method = "LjungBox",interval=FALSE)
#"LiMcLeod" (default), "LjungBox", "BoxPierce".
```

```
## $test
##           ChiSq DF    pvalue
## [1,] 0.8473242  1 0.3573104
## attr("method")
## [1] "LjungBox"
##
## $ci
##      int
## [1,]  0 0
## attr("level")
## [1] FALSE
```

Гипотеза о том, что остался белый шум не отверглась.

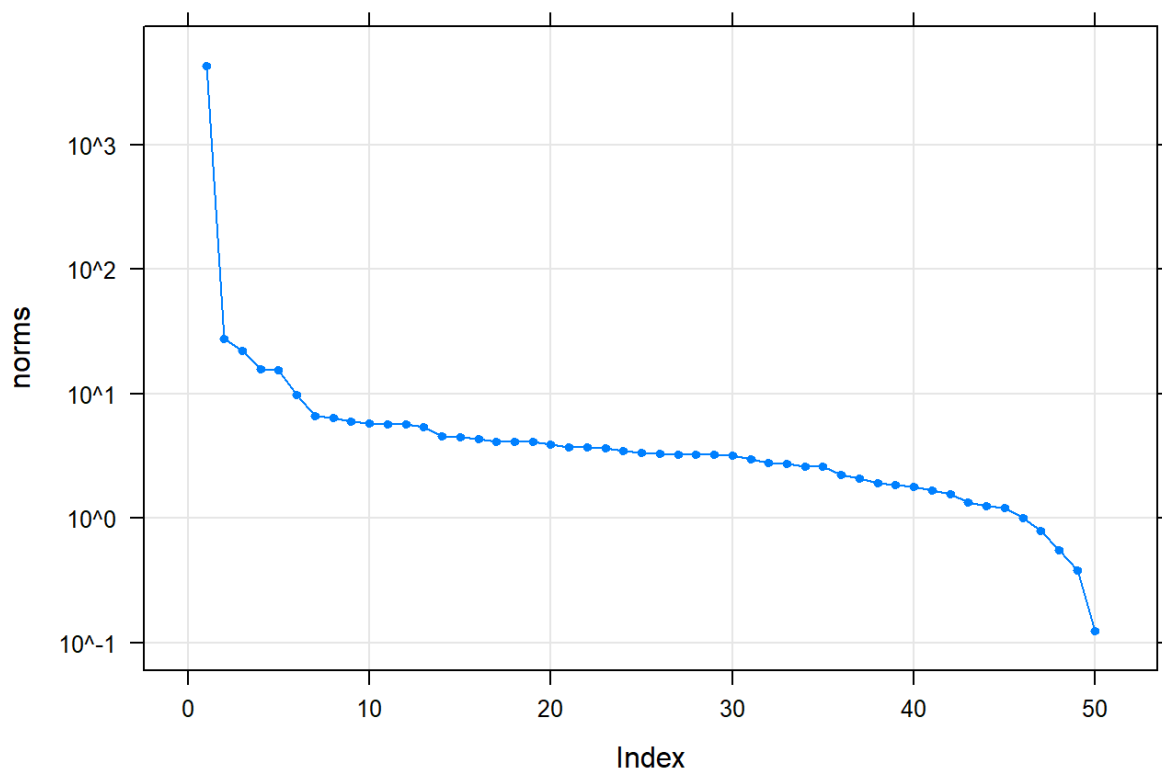
Разбираемся со вторым модельным рядом.

```
autoplot(ts10)
```



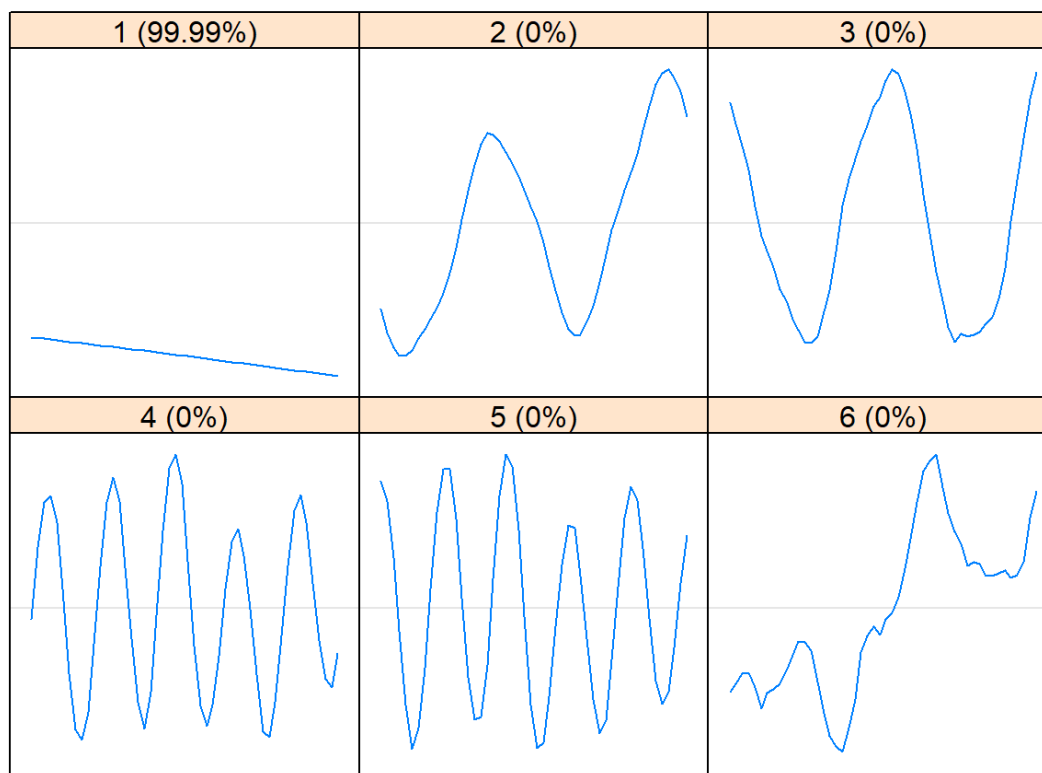
```
ssa10<-ssa(ts10, L=50)  
plot(ssa10)
```

Component norms



```
plot(ssa10, type = "vectors", idx = 1:6)
```

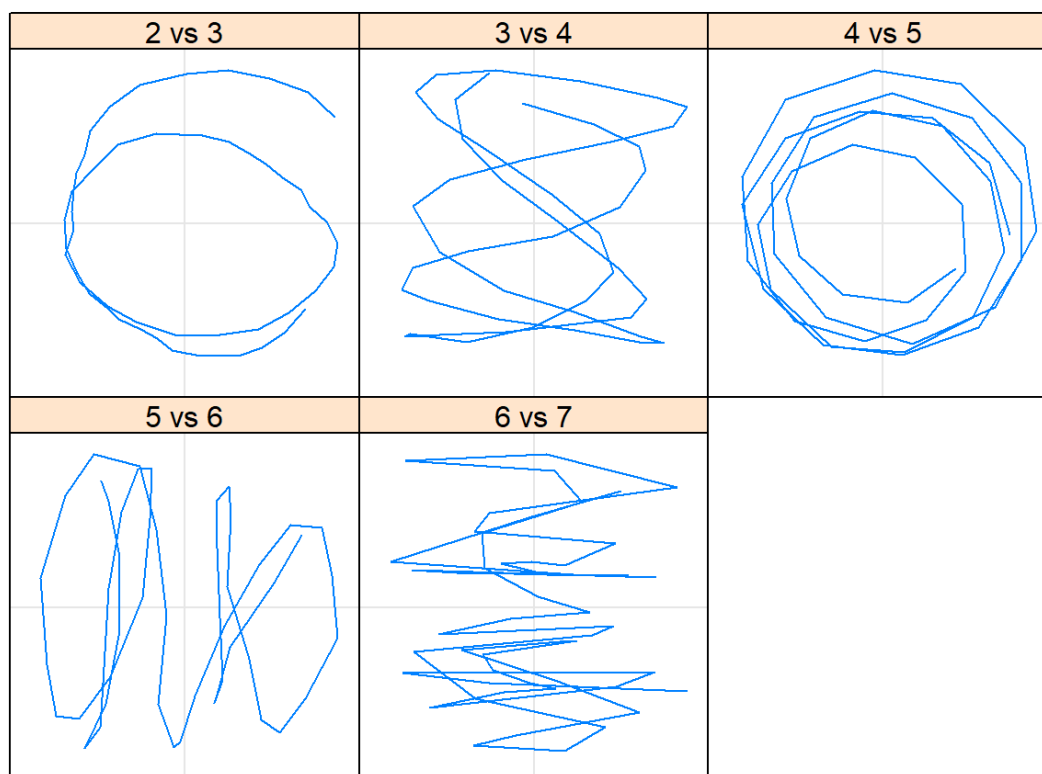
Eigenvectors



Первая компонента – тренд. Рассмотрим попарно остальные.

```
plot(ssa10, type = "paired", idx = 2:6, plot.contrib = FALSE)
```

Pairs of eigenvectors



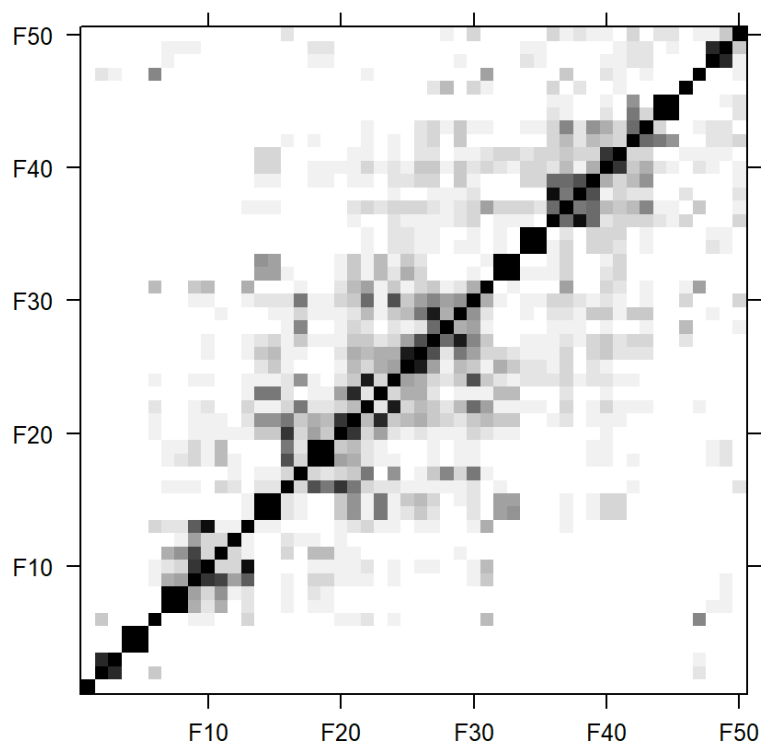
```
print(parestimate(ssa10, groups = list(2:3,4:5), method = "esprit"))
```

```
## $F1
##   period    rate |   Mod   Arg |    Re    Im
##   28.276  0.004682 | 1.00469  0.22 | 0.97999  0.22142
##   -28.276  0.004682 | 1.00469 -0.22 | 0.97999 -0.22142
##
## $F2
##   period    rate |   Mod   Arg |    Re    Im
##   10.221 -0.009592 | 0.99045  0.61 | 0.80914  0.57121
##   -10.221 -0.009592 | 0.99045 -0.61 | 0.80914 -0.57121
```

Получили, что периоды равны 28 и 10 (примерно).

```
plot(wcor(ssa10, groups = 1:50),
     scales = list(at = c(10, 20, 30, 40, 50)))
```

W-correlation matrix



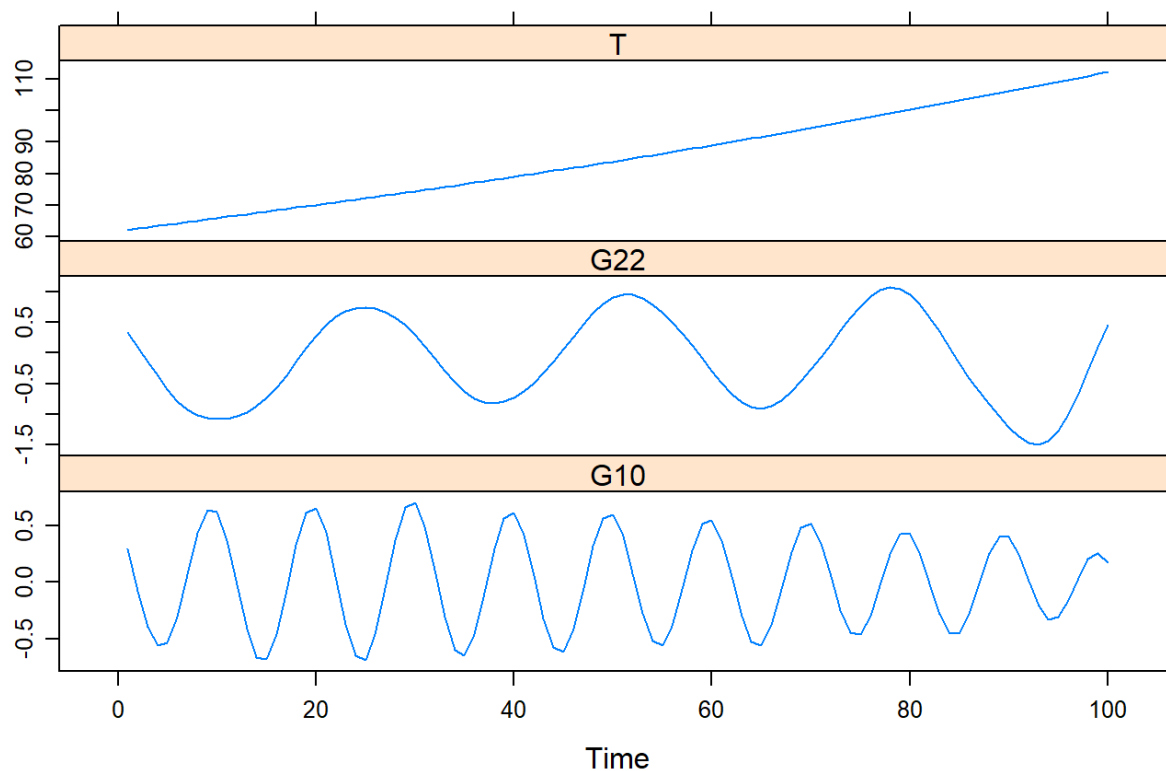
То же самое. 1 тренд, 2 гармоники. То, что одна из гармоник коррелирует с чем-то еще может значить, что между ней и каким-то шумом нет сильной разделимости. Тогда в этот кусок могла попасть часть шума. Но что делать...

Положим первую компоненту в тренд, 2:5 – в периодичность.

Reconstructed series:

```
plot(reconstruct(ssa10, groups = list(T=1, G22 = 2:3, G10 = 4:5)),
     plot.method = "xyplot", layout = c(1,3),
     add.residuals = FALSE, add.original = FALSE)
```

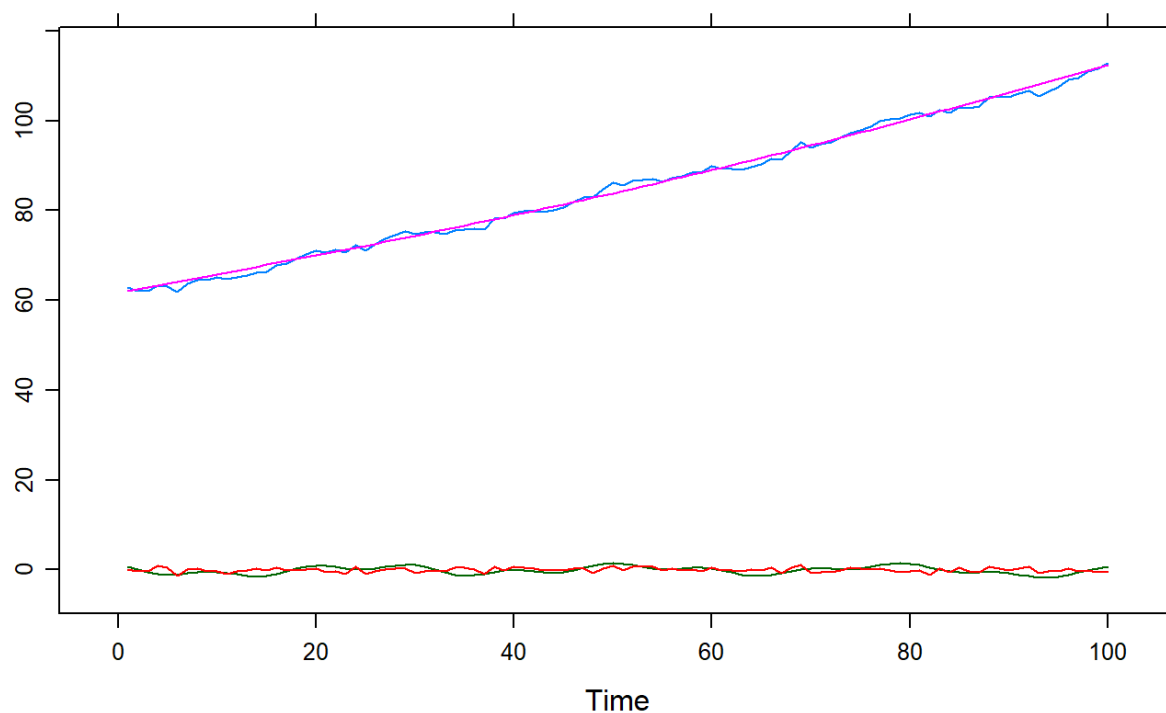
Reconstructed Series



```
reconstruct10<- reconstruct(ssa10,
                           groups = list(Trend = 1,
                                           Seasonality = 2:5))
plot(reconstruct10, add.residuals = TRUE, add.original = TRUE,
     plot.method = "xyplot",
     superpose = TRUE, auto.key = list(columns = 2))
```

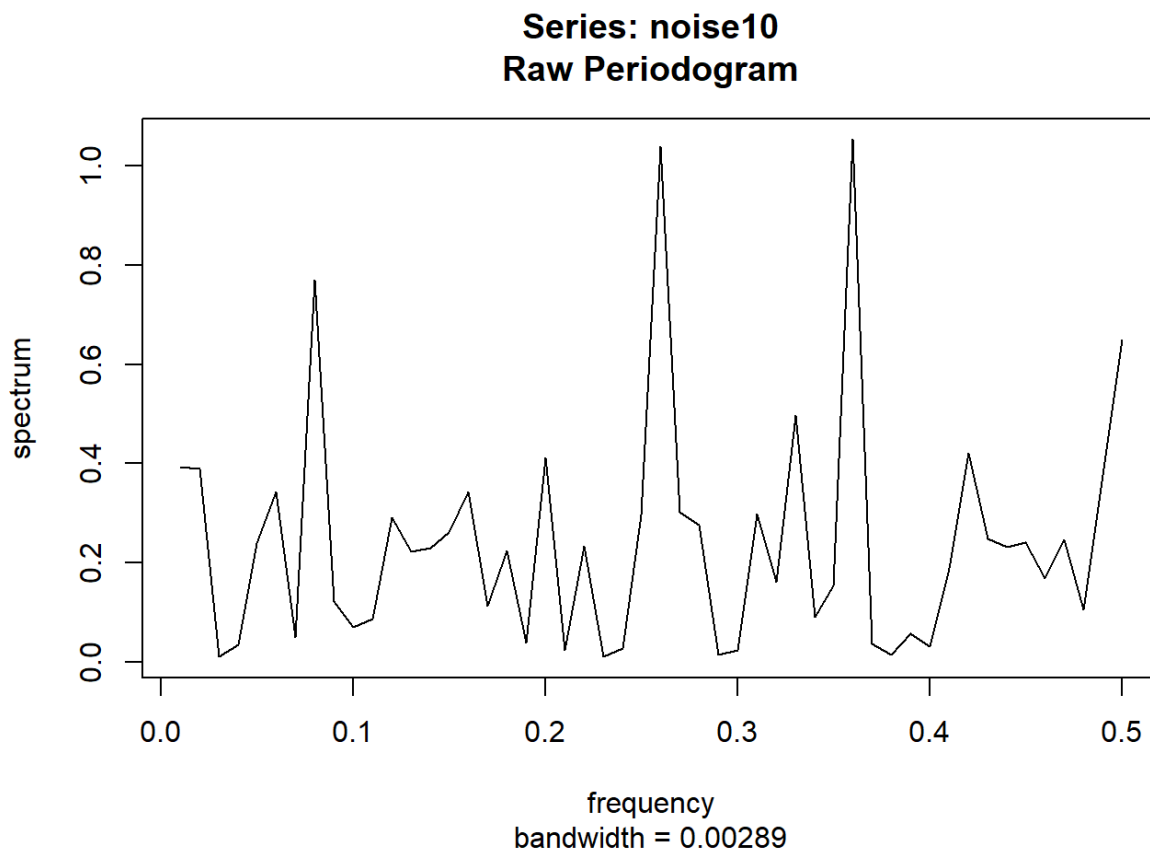
Reconstructed Series

Original ——— Seasonality ———
Trend ——— Residuals ———



Периодограмма остатков:

```
noise10<-ts10-reconstruct10$Trend-reconstruct10$Seasonality
spec.pgram(noise10,log="no")
```



С одной стороны, вроде как похоже на оставшуюся периодичность (пики). Но наблюдений так мало, что это вполне могут быть просто случайные пики.

Проверка гипотезы о том, что остаток – белый шум.

```
autocor.noise10<- autocorrelations(noise10, maxlag = 10)
whiteNoiseTest(autocor.noise10, h0 = "iid",method = "LjungBox",interval=FALS
E) # "LiMcLeod" (default), "LjungBox", "BoxPierce".
```

```
## $test
##      ChiSq DF    pvalue
## [1,] 0.160193 1 0.6889789
## attr(,"method")
## [1] "LjungBox"
##
## $ci
##      int
## [1,] 0 0
## attr(,"level")
## [1] FALSE
```

Гипотеза не отвергается.

