Roots

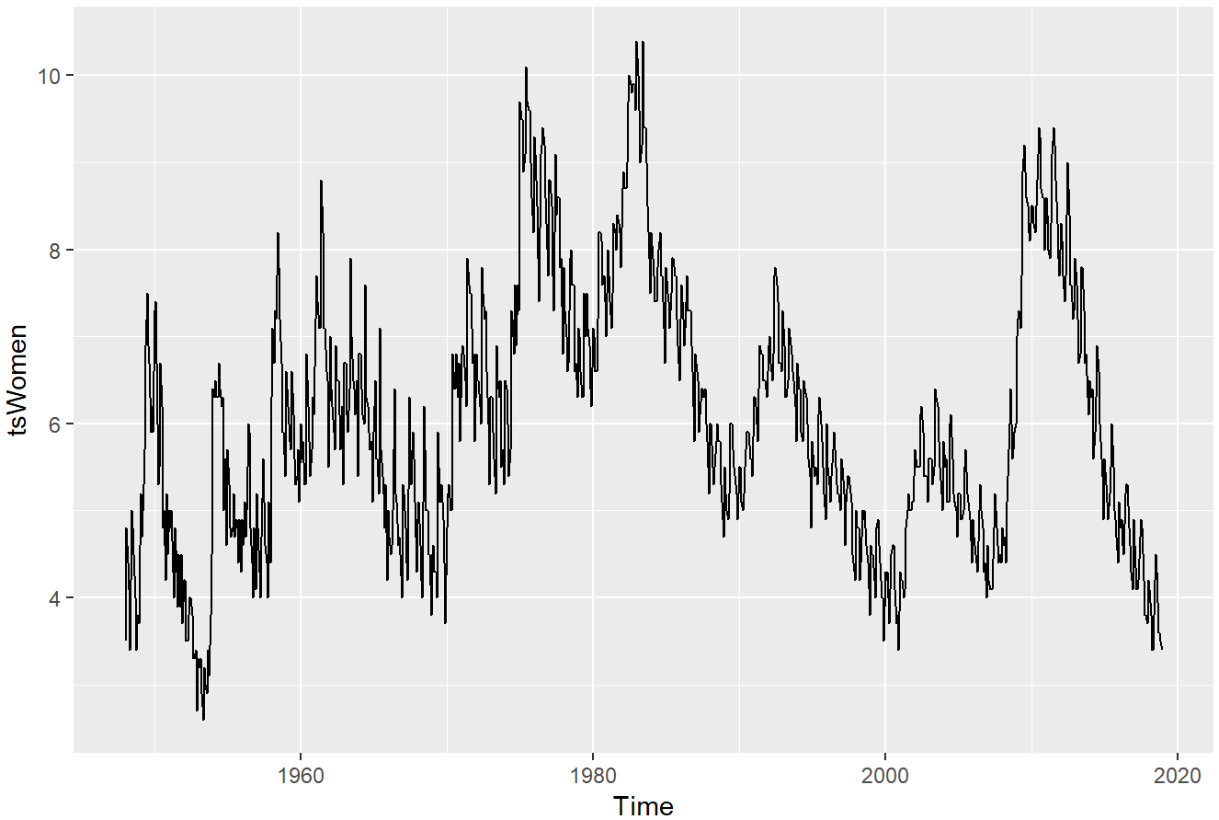
Romanova

25 апреля 2019 г

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)
```



2 SSA

Посмотрим на график собственных чисел всего ряда.

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

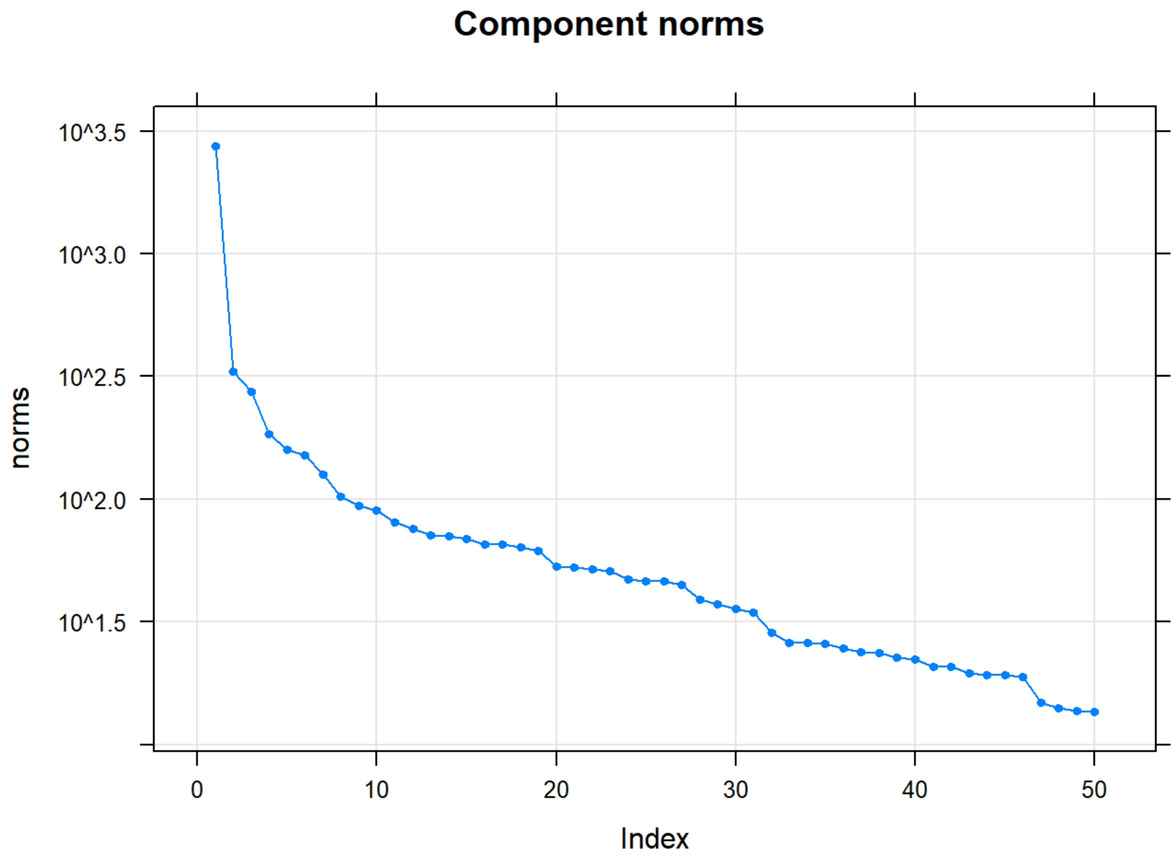
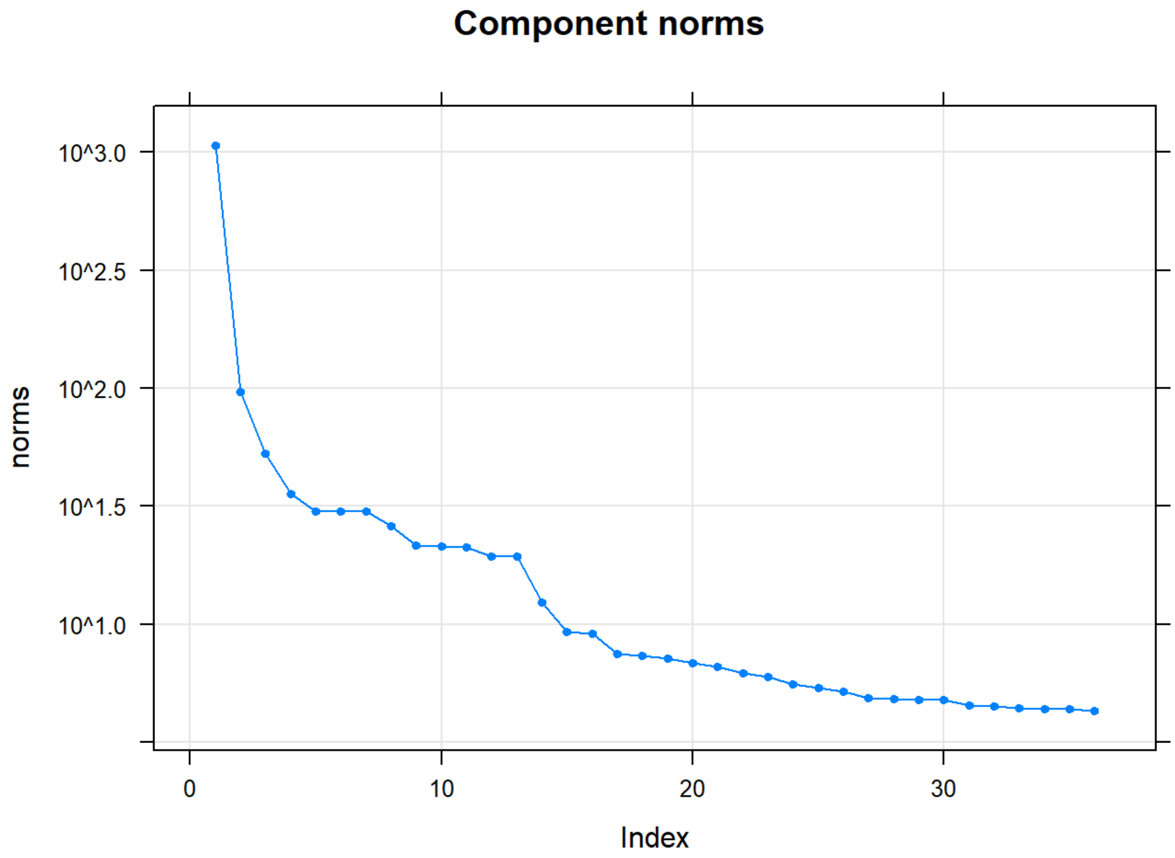


График не очень приятный. Будет сложно отделить сигнал от шума (собственные числа близки).

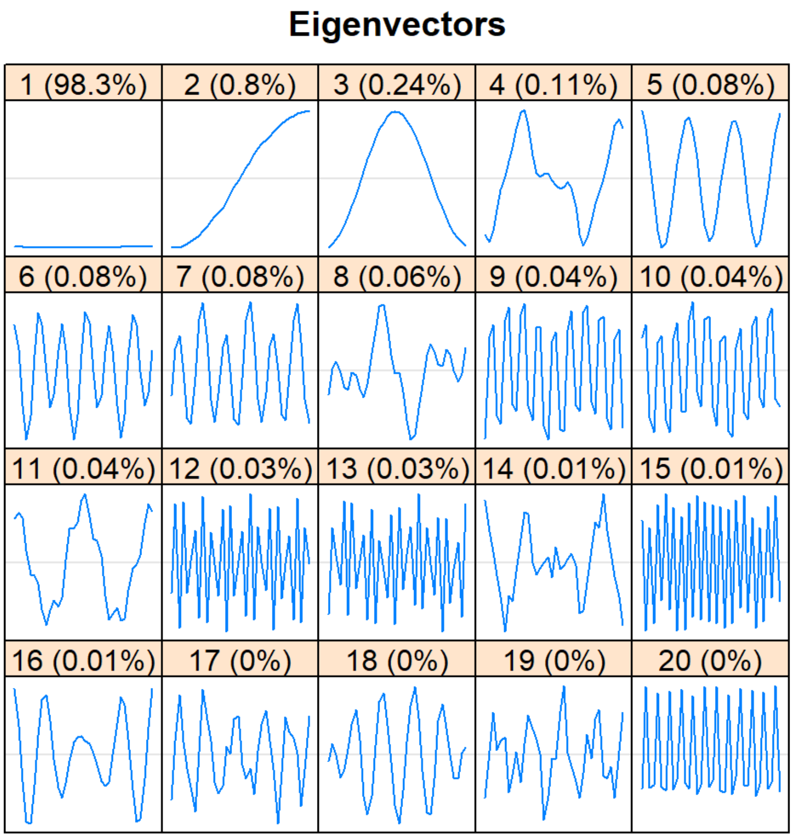
Как показали эксперименты, сразу делать ssa для нашего ряда – это ужас. Тренд вылезает всюду и вне очереди (как мы и ожидали, посмотрев на предыдущий график, все перемешалось. Этого можно было ожидать и на основе периодограммы, там все пики примерно одинаковой величины). И рассматривать нам придется 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)
```

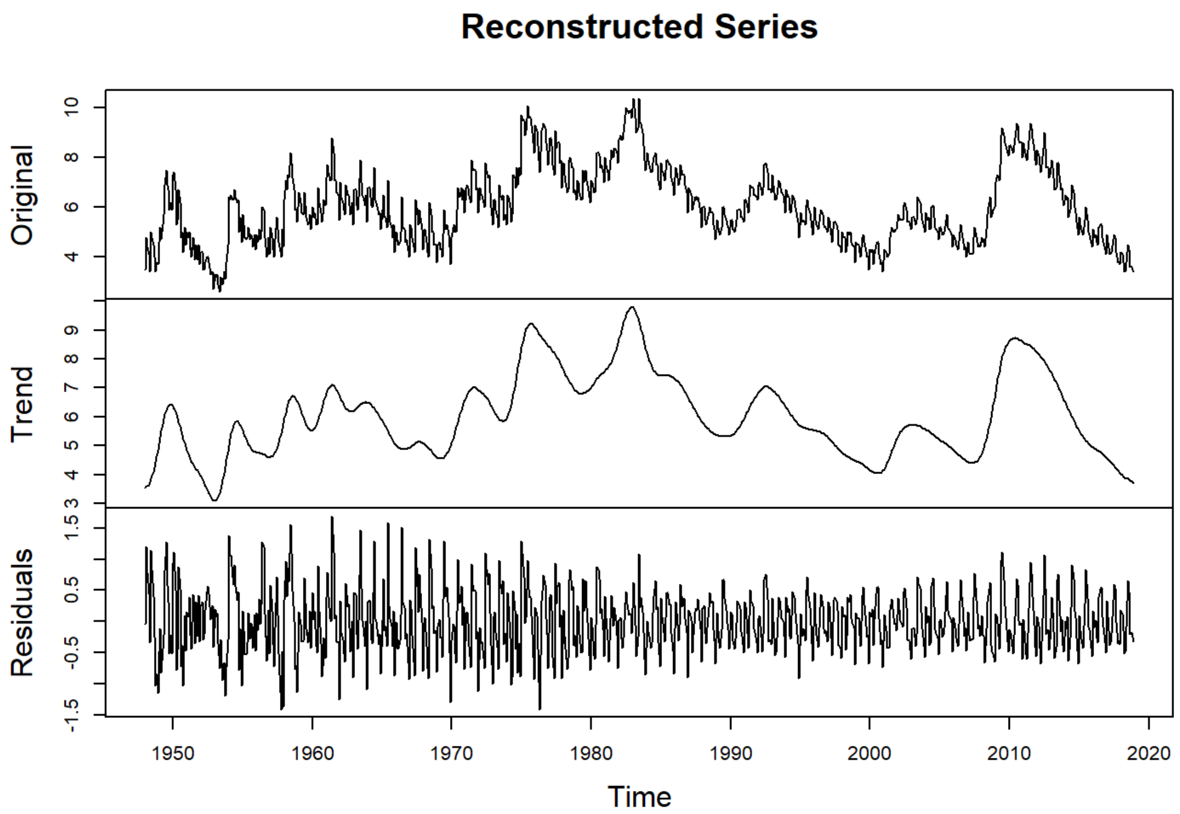



```
plot(trend.SSA,type="vectors",idx=1:20)
```



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))
```



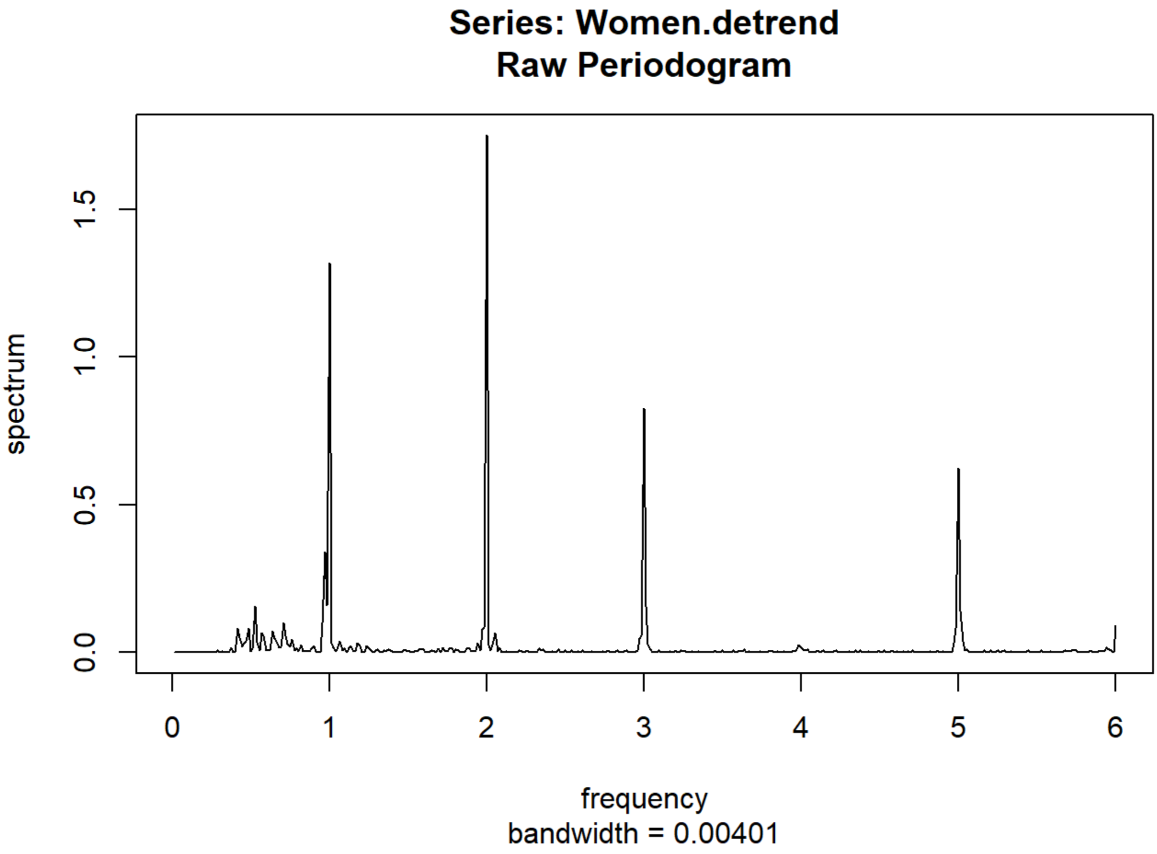

```
trend.ssa.reconstruct<-reconstruct(trend.SSA, groups = list(Trend = 1:3))
Women.trend<-trend.ssa.reconstruct$Trend
```

Переходим к анализу ряда без тренда.

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

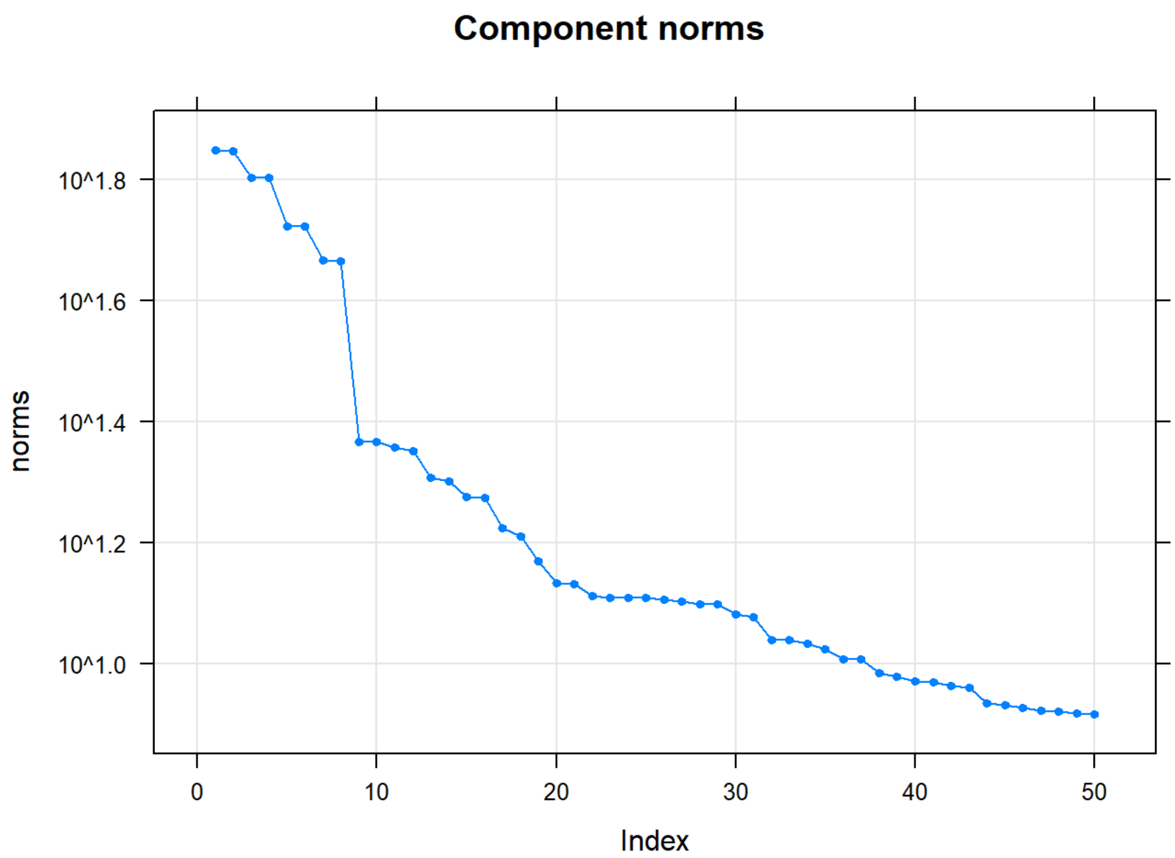
Построим периодограмму для полученного ряда без тренда:

```
spec.pgram(Women.detrend,log='no')
```



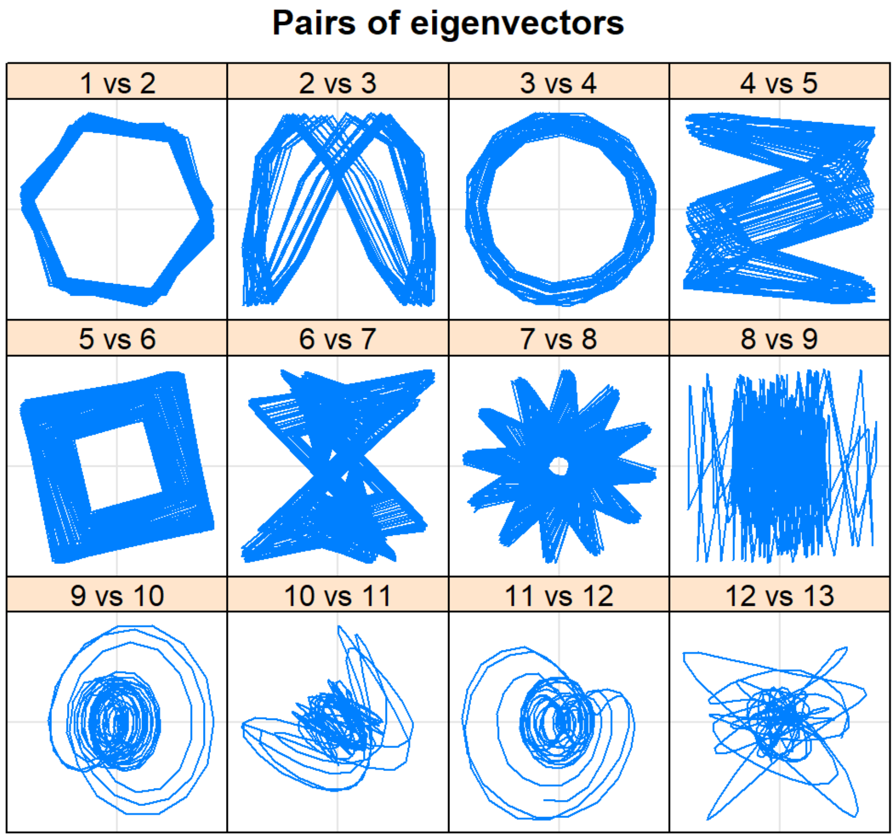
Как мы видели и раньше, присутствуют периоды 12, 6, 4, 2.4. И теперь еще немного выделились периоды 2 и 3, которых на периодограмме исходного ряда было не видно (из-за масштаба). Так что далее будем искать в том числе и пилу. И заметны повышенные низкие частоты.

```
plot(Women.detrend.ssa)
```



Четыре пары выделяются сильно. Но нет уверенности в том, что они возьмут всю периодичность (точнее, точно не возьмут, хотя бы потому, что мы теперь обнаружили больше периодов в ряде). Но сначала разберемся с очевидными.

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

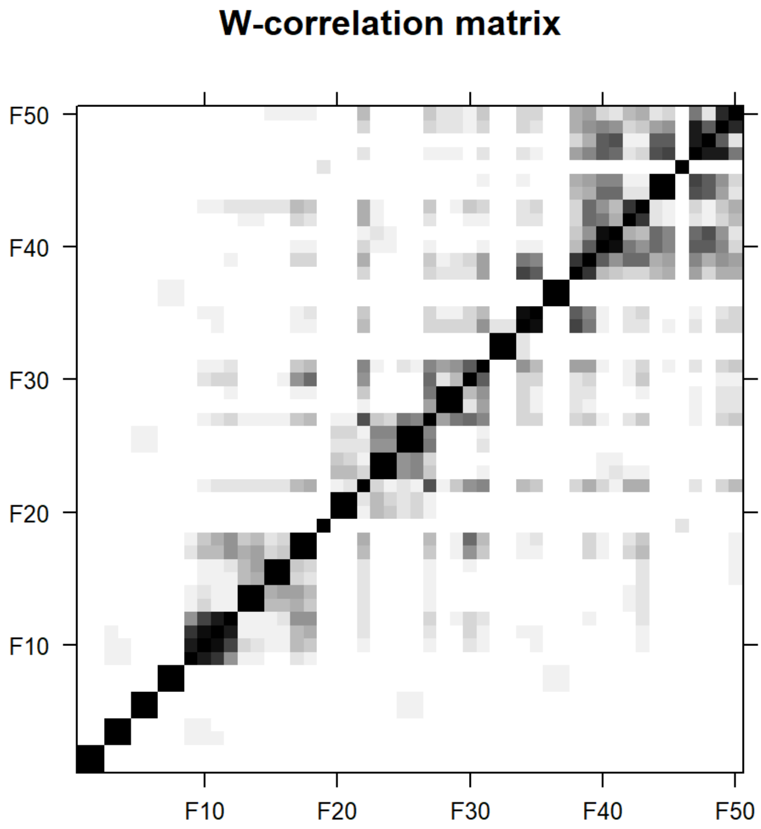



```
parestimate(Women.detrend.ssa, groups = list(1:2,3:4,5:6,7:8,9:10,11:12), me
thod = "esprit")
```

```
## $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.

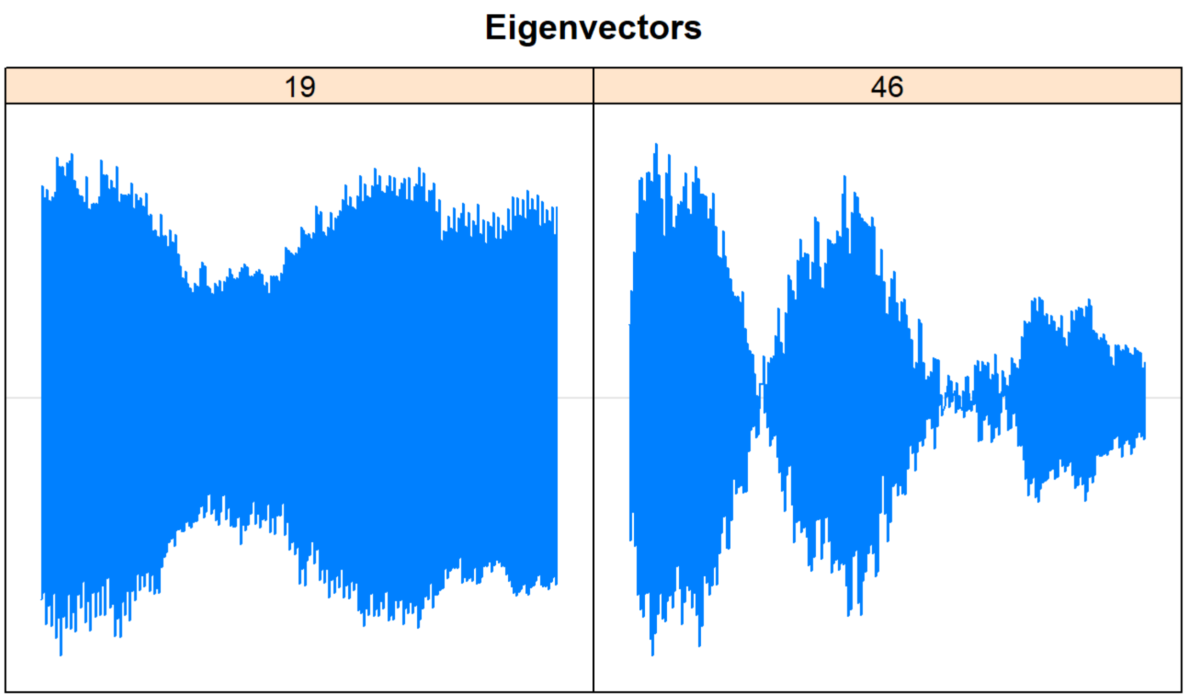
```
plot(wcor(Women.detrend.ssa, groups = 1:50),
          scales = list(at = c(10, 20, 30,40,50)))
```



Также отчетливо выделяются 4 периодики. Но нам этого не достаточно. Сначала попробуем выведать все, что можно, из матрицы взвешенных корреляций. На ней можно заметить квадратики, особо не коррелирующие с остальными, но имеющие уже не очень большой вклад. Это пары 32-33 и 36-37. Также выделяется 19 и 46 компонента, они не коррелируют с остальными (только немного между собой).

Возможно, компоненты 19 и 46 – это две пилы. Посмотрим, как они выглядят:

```
plot(Women.detrend.ssa, type = "vectors", idx = c(19,46), plot.contrib = FAL
SE)
```



Удостоверимся:

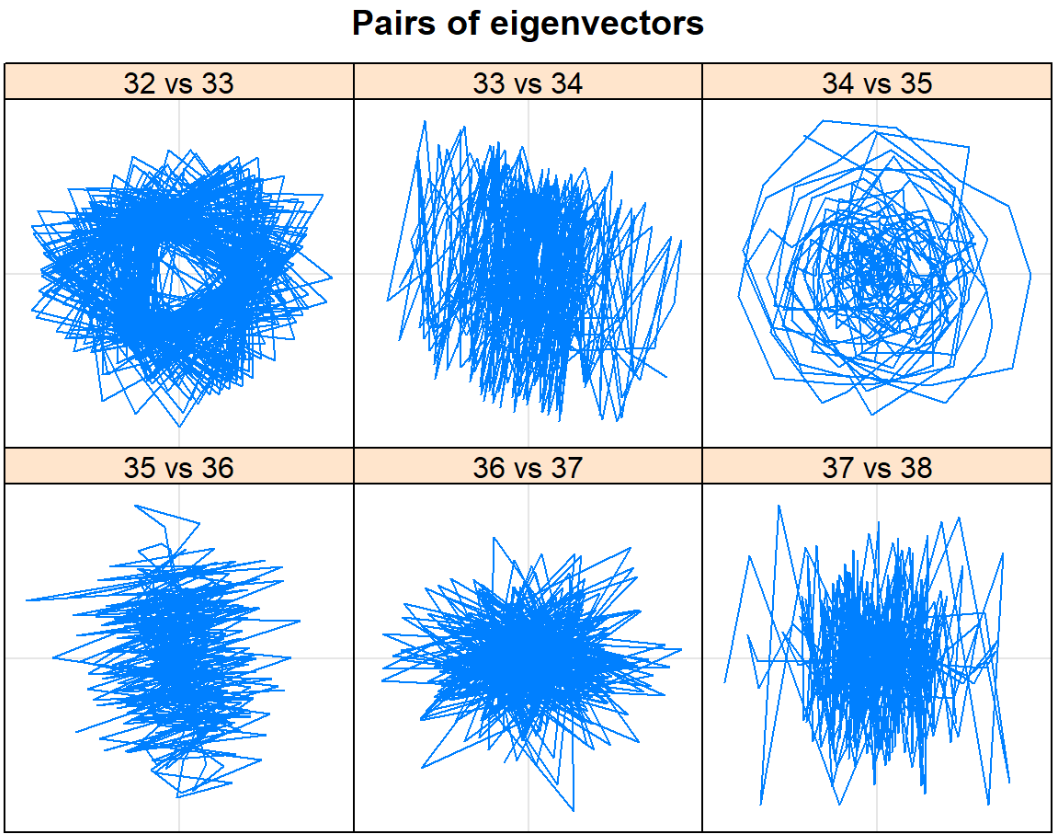
```
parestimate(Women.detrend.ssa, groups = list(19,46), method = "esprit")
```

```
## $F1
##    period      rate   |    Mod     Arg  |     Re        Im
##     2.000  -0.003630  |  0.99638   3.14 | -0.99638   0.00000
##
## $F2
##    period      rate   |    Mod     Arg  |     Re        Im
##     2.000  -0.018527  |  0.98164   3.14 | -0.98164   0.00000
```

19 похожа на пилу, а 46 смесь чего-то (имеет сильную модуляцию).

Проверим теперь пары 32-33 и 36-37.

```
plot(Women.detrend.ssa, type = "paired", idx = 32:37, plot.contrib = FALSE)
```

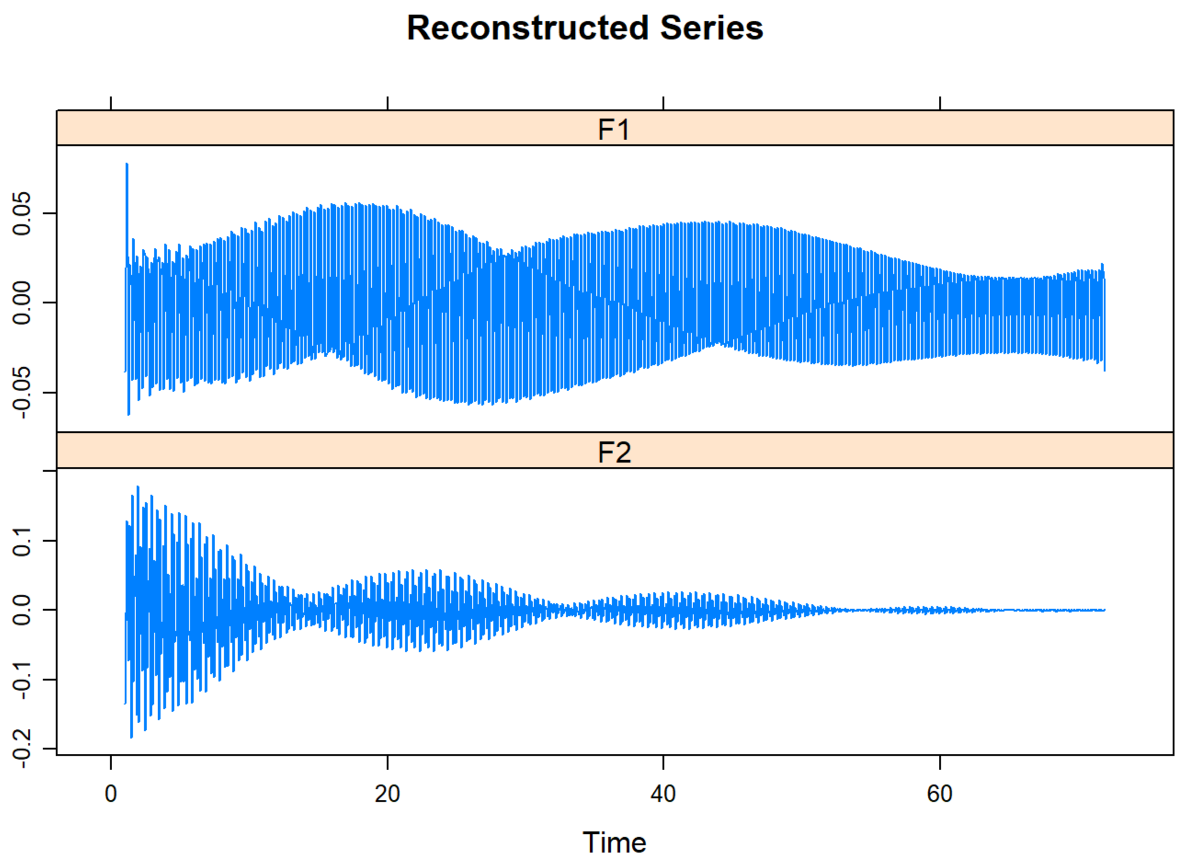



```
parestimate(Women.detrend.ssa, groups = list(32:33,36:37), method = "esprit"
)
```

```
## $F1
##    period      rate   |    Mod     Arg  |      Re        Im
##     3.066  -0.044602  |  0.95638   2.05 | -0.44015   0.84907
##    -3.066  -0.044602  |  0.95638  -2.05 | -0.44015  -0.84907
##
## $F2
##    period      rate   |    Mod     Arg  |      Re        Im
##     2.415  -0.041079  |  0.95975   2.60 | -0.82323   0.49338
##    -2.415  -0.041079  |  0.95975  -2.60 | -0.82323  -0.49338
```

Периоды этих парочек выглядят правдоподобно: 3 и 2.4. Такие периоды для нас актуальны.

```
plot(reconstruct(Women.detrend.ssa, groups = list(32:33,36:37)),
     plot.method = "xyplot", layout = c(1,2),
     add.residuals = FALSE, add.original = FALSE)
```

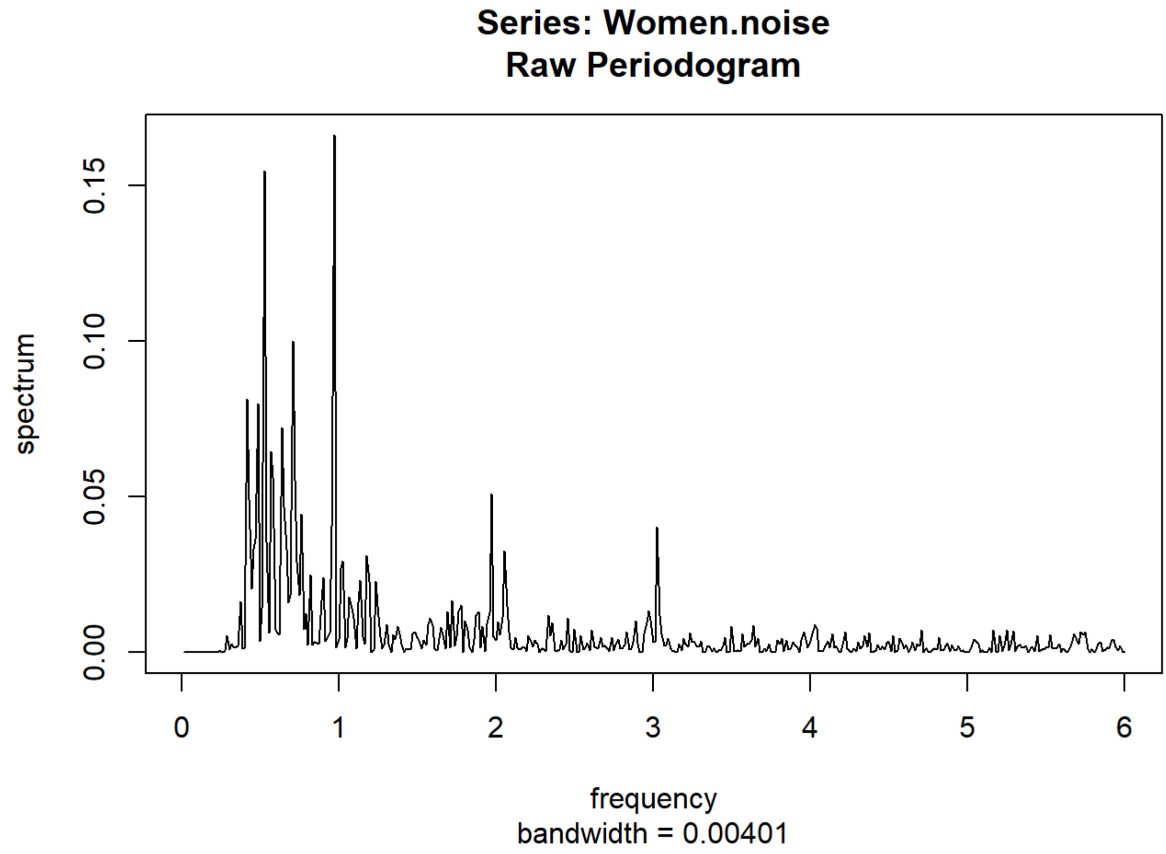


Берем все рассмотренные компоненты в сезонность (они описывают периоды 6,12, 4, 2.4, 3, 2):

```
Women.detrend.reconstruct <- reconstruct(Women.detrend.ssa,
                              groups = list(Seasonality = c(1:8,19,46,32:33,36:37)))
```

Периодограмма остатка:

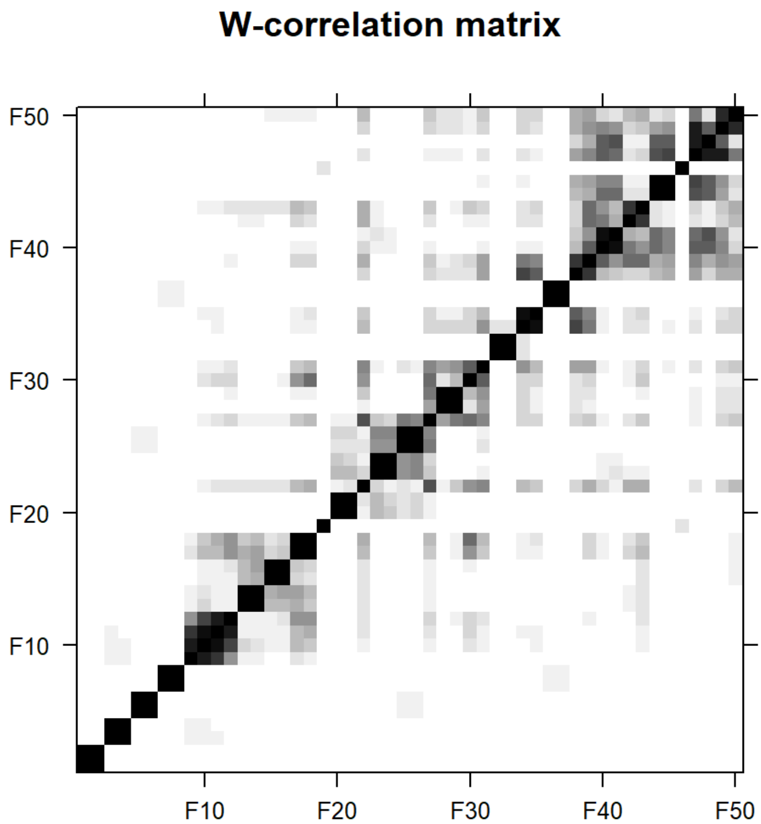
```
Women.noise<-ts(Women.detrend-Women.detrend.reconstruct$Seasonality,frequenc
y = 12)
spec.pgram(Women.noise, log="no")
```



Все еще остались периоды 12, 6 и 4.

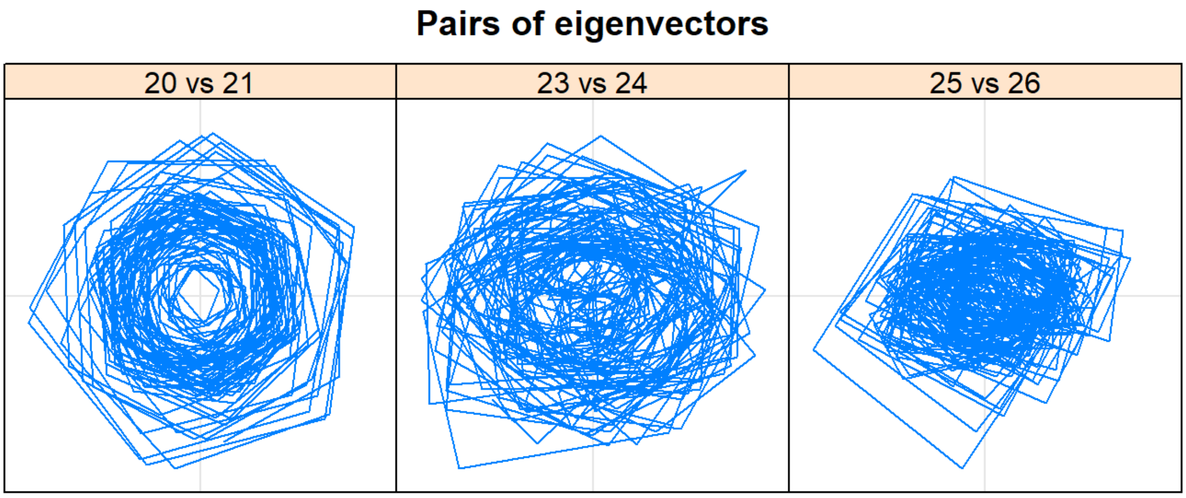
Снова посмотрим на матрицу взвешенных корреляций:

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



Заметим, что компоненты 20-21,23-24,25-26 образуют блок компонент, которые коррелируют между собой, но не коррелируют с другими.

```
plot(Women.detrend.ssa, type = "paired", idx = c(20,23,25), idy = c(21,24,26
), plot.contrib = FALSE)
```




```
parestimate(Women.detrend.ssa, groups = list(20:21,23:24,25:26), method = "e
sprit")
```

```
## $F1
##    period      rate   |    Mod     Arg  |     Re        Im
##     6.033  -0.015663 |  0.98446   1.04 |  0.49708   0.84975
##    -6.033  -0.015663 |  0.98446  -1.04 |  0.49708  -0.84975
##
## $F2
##    period      rate   |    Mod     Arg  |     Re        Im
##     6.002  -0.140996 |  0.86849   1.05 |  0.43449   0.75200
##    -6.002  -0.140996 |  0.86849  -1.05 |  0.43449  -0.75200
##
## $F3
##    period      rate   |    Mod     Arg  |     Re        Im
##     4.733  -0.236549 |  0.78935   1.33 |  0.19004   0.76613
##    -4.733  -0.236549 |  0.78935  -1.33 |  0.19004  -0.76613
```

Похоже, что пары 20-21 и 23-24 – это перемешавшиеся периодики с периодом 6. Пара 25-26 имеет неправдоподобный период 4.7, но при этом очень сильно коррелирует с парой 23-24. Это странно. В итоге эта пара перебивает период 4, так что ее тоже нужно брать. (возможно, при помощи iossa это тоже получится подкрутить).

Другой блок (9:18) не похож на перемешавшиеся периодики, тк коррелирует и с остальными. Однако мы до сих пор не нашли недостающую гармонику с периодом 12. Проверим периоды на всякий случай:

```
parestimate(Women.detrend.ssa, groups = list(9:10,11:12,13:14,15:16,17:18),
method = "esprit")
```

```
## $F1
##    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
##
## $F2
##    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
##
## $F3
##    period     rate    |    Mod     Arg  |     Re         Im
##    17.935  -0.005452 |  0.99456   0.35 |  0.93415   0.34134
##   -17.935  -0.005452 |  0.99456  -0.35 |  0.93415  -0.34134
##
## $F4
##    period     rate    |    Mod     Arg  |     Re         Im
##    19.511  -0.001420 |  0.99858   0.32 |  0.94725   0.31604
##   -19.511  -0.001420 |  0.99858  -0.32 |  0.94725  -0.31604
##
## $F5
##    period     rate    |    Mod     Arg  |     Re         Im
##    21.931  -0.011408 |  0.98866   0.29 |  0.94836   0.27939
##   -21.931  -0.011408 |  0.98866  -0.29 |  0.94836  -0.27939
```

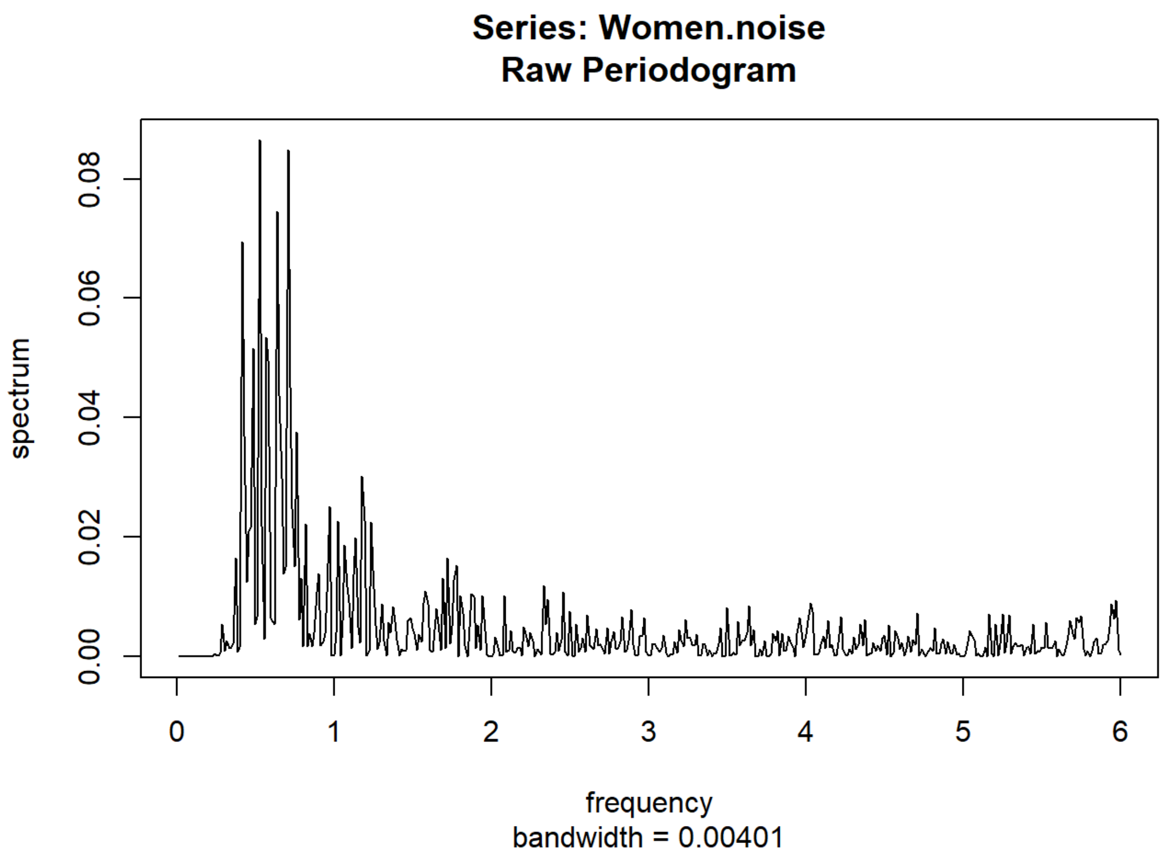
Пара 9:10 имеет период примерно 15. Однако, так как ближе к 12ти ничего другого нет, пробуем ее добавить и, о чудо, эта пара справляется с периодом 12. (так, видимо, тоже получилось из-за того, что гармоники перемешались)

Обновим результат.

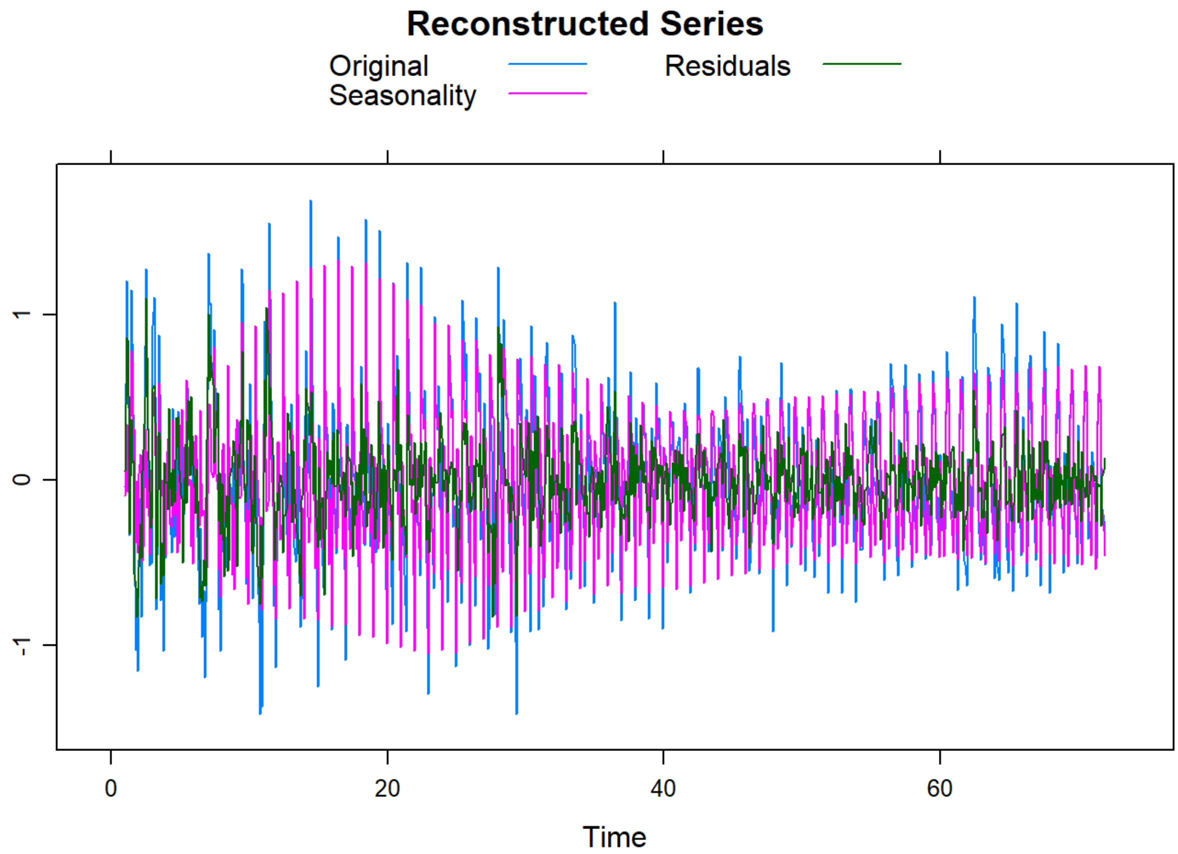
```
Women.detrend.reconstruct <- reconstruct(Women.detrend.ssa,
                      groups = list(Seasonality = c(1:8,9:10,19,20:21,23:24,
25:26,32:33,36:37)))
```

Периодограмма шума:

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




```
plot(Women.detrend.reconstruct , add.residuals = TRUE, add.original = TRUE,
     plot.method = "xyplot",
     superpose = TRUE, auto.key = list(columns = 2))
```

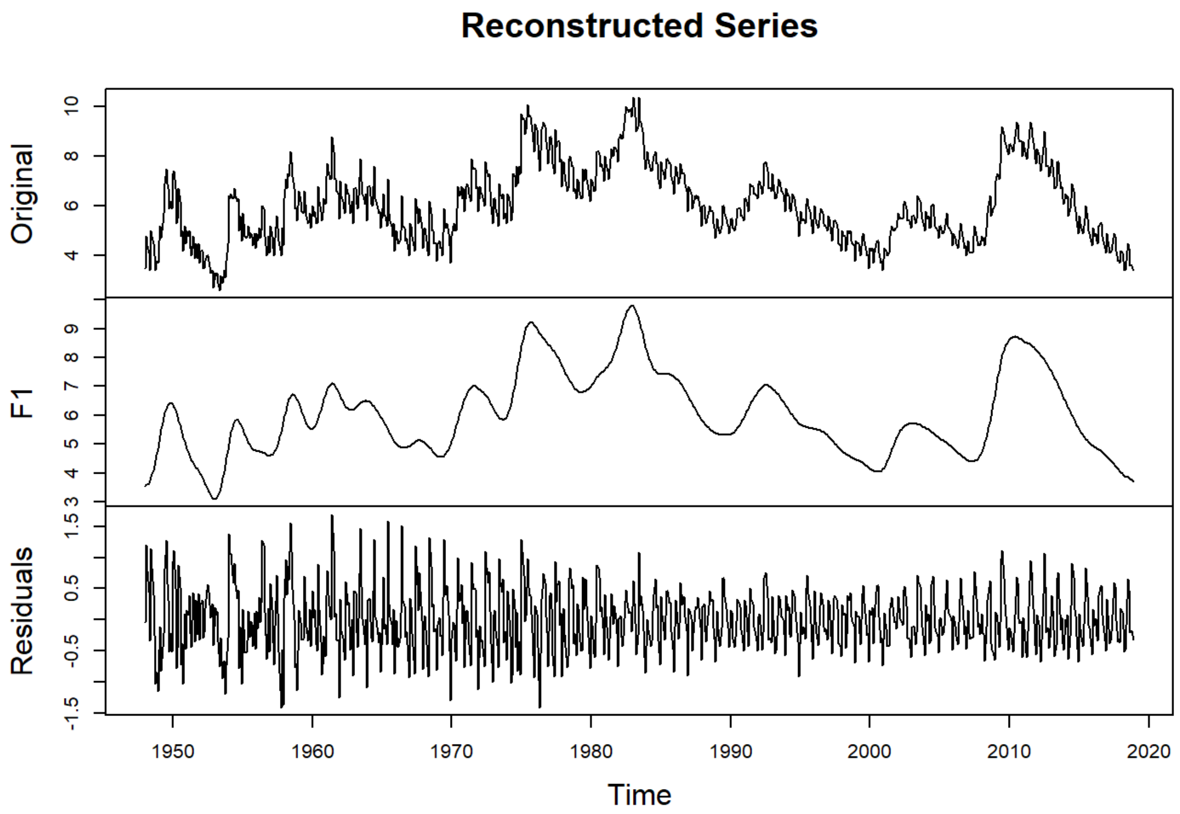


3 Автоматическая идентификация

Посмотрим, какие результаты дает автоматическая группировка.

Для тренда. Частотная группировка:

```
gr1 <- grouping.auto(trend.SSA, grouping.method = "pgram", groups = 1:20, ba
se = "series",freq.bins = list(1/24), threshold = 0.95)
plot(reconstruct(trend.SSA, groups = gr1))
```




```
head(gr1)
```

```
## $F1
## [1] 1 2 3
```

Автоматическая группировка взяла первые 3 компоненты разложения в тренд. Мы поступили так же.

Кластерная группировка для ряда без тренда. Возьмем сначала в качестве компонент, из которых выбирать, с 1 по 37 компоненту. (последняя взятая нами вручную компонента – 37, но внутри брали не все)

```
gr <- grouping.auto(Women.detrend.ssa, grouping.method = "wcor", groups=1:37
)
head(gr,50)
```

```
## $`1`
## [1] 1 2
##
## $`2`
## [1] 3 4
##
## $`3`
## [1] 5 6
##
## $`4`
## [1] 7 8
##
## $`5`
## [1]  9 10 11
##
## $`6`
## [1] 12 17 18
##
## $`7`
## [1] 13 14
##
## $`8`
## [1] 15 16
##
## $`9`
## [1] 19
##
## $`10`
## [1] 20 21
##
## $`11`
## [1] 22 27
##
## $`12`
## [1] 23 24
##
## $`13`
## [1] 25 26
##
## $`14`
## [1] 28 29
##
## $`15`
## [1] 30 31
##
## $`16`
## [1] 32 33
##
## $`17`
## [1] 34 35
##
## $`18`
## [1] 36 37
```

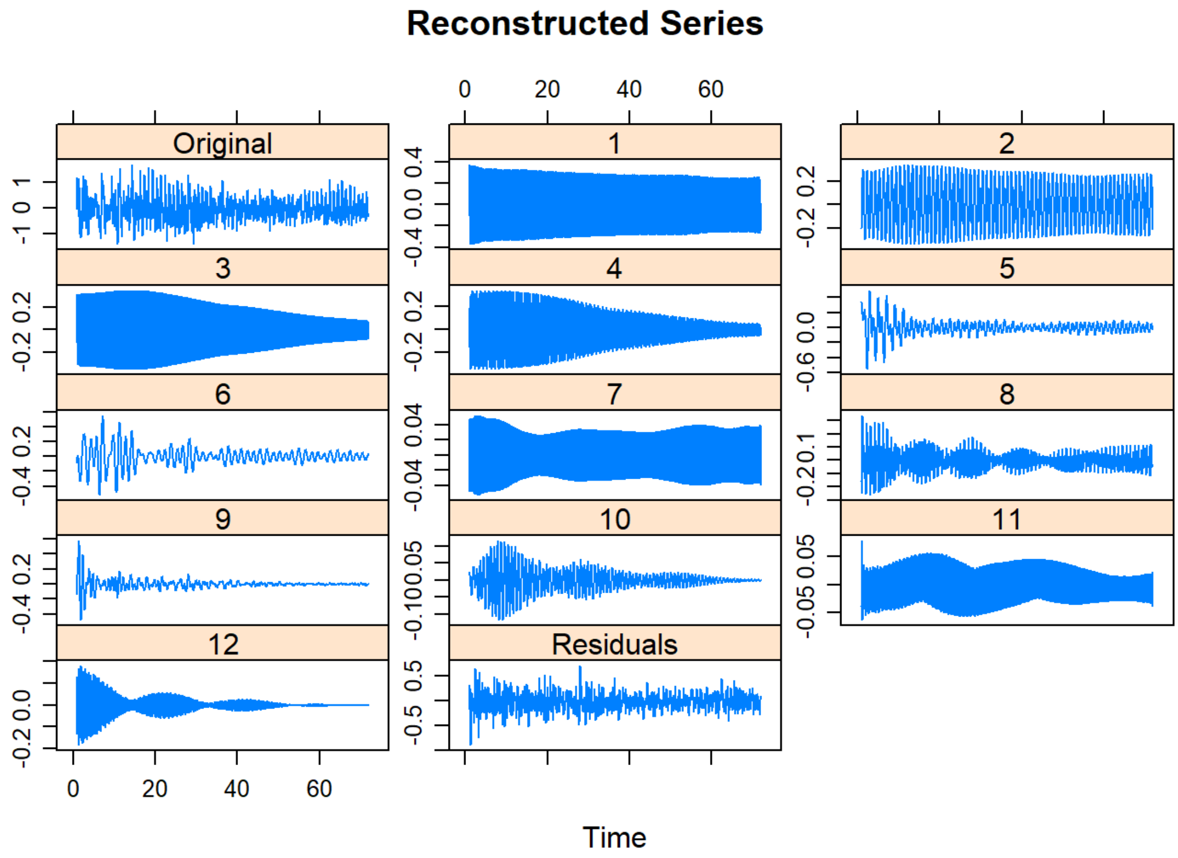
Метод выделил 18 кластеров (мы – 11). Также выделил одинокую пилу и большинство групп (только их здесь больше). Ну и в целом на данный момент это просто распаривание всего подряд.

Если теперь укажем желаемое количество кластеров (+1 для шума, чтобы чисто теоретически метод мог разделить так же, как мы):

```
gr <- grouping.auto(Women.detrend.ssa, grouping.method = "wcor", groups=1:37
,nclust=12)
head(gr,50)
```

```
## $`1`
## [1] 1 2
##
## $`2`
## [1] 3 4
##
## $`3`
## [1] 5 6
##
## $`4`
## [1] 7 8
##
## $`5`
## [1]  9 10 11
##
## $`6`
## [1] 12 13 14 15 16 17 18
##
## $`7`
## [1] 19
##
## $`8`
## [1] 20 21 23 24 25 26
##
## $`9`
## [1] 22 27 30 31 34 35
##
## $`10`
## [1] 28 29
##
## $`11`
## [1] 32 33
##
## $`12`
## [1] 36 37
```

```
plot(reconstruct(Women.detrend.ssa, groups = gr),plot.method = "xyplot")
```



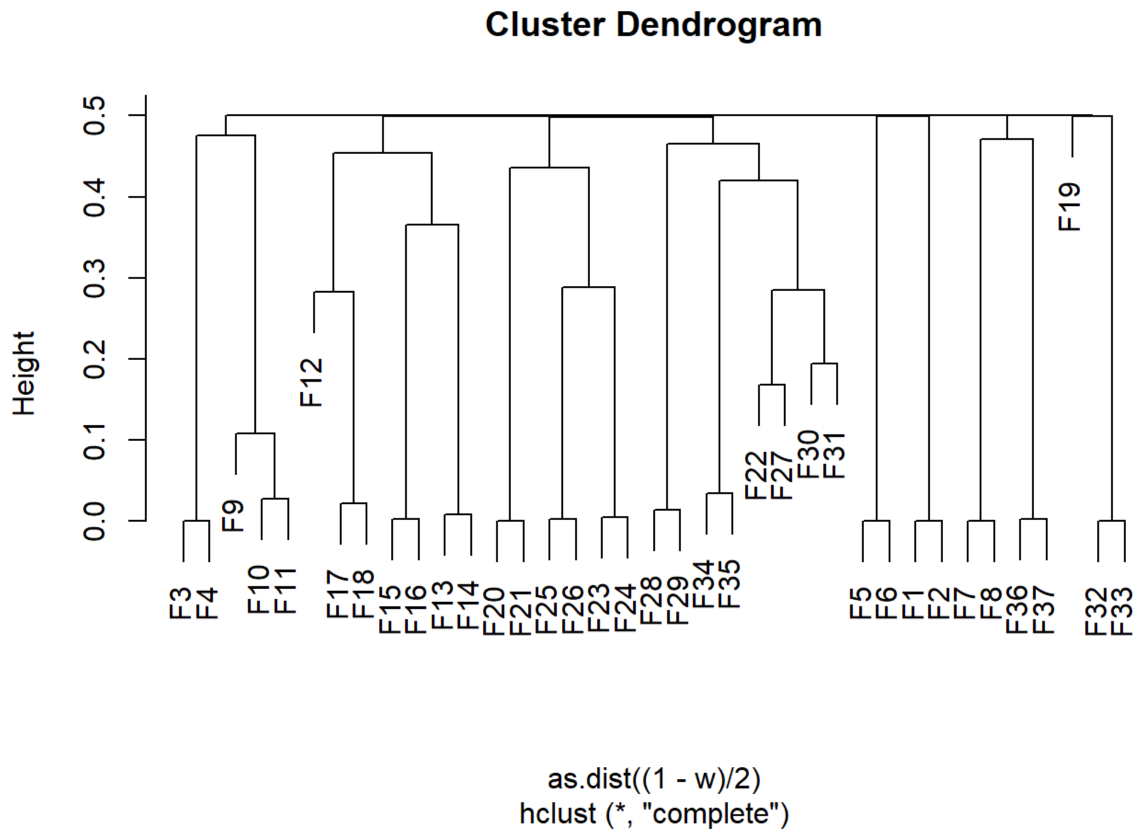
Наиболее очевидные вещи метод заметил. Он попарно выделил первые 4 периодики (которые мы выделили первым делом), 19 (пила) – выделил отдельно, компоненты 20,21,23,24,25,26, которые есть смешавшиеся периодики, – выделил в одну группу, но не разделил попарно (это и не удивительно). Выделил периодики 32:33, 36:37.

Если считать, что если в кластере больше 2х компонент, то кластер соответствует шуму, то можно сказать, что в шум отправились компоненты 12:18, 22, 27, 30, 31, 34, 35, блок периодик 20,21,23:26 и 9,10,11 (мы брали пару 9:10, но она тоже совсем не очевидная).

Понятно, что такой подход не способен выделить весь шум в одну компоненту в данном случае, так как сигнал в матрице взвешенных корреляций у нас разбросан (то есть не так, что сигнал в начале, а шум в конце, все вперемешку).

Попробуем увеличть количество кластеров, вдруг еще что хорошее разделится.

```
gr <- grouping.auto(Women.detrend.ssa, grouping.method = "wcor", groups=1:37
,nclust=15)
plot(gr)
```


Cluster Dendrogram
Height
0.0
0.1
0.2
0.3
0.4
0.5
F3
F4
F9
F10
F11
F12
F17
F18
F15
F16
F13
F14
F20
F21
F25
F26
F23
F24
F28
F29
F34
F35
F22
F27
F30
F31
F5
F6
F1
F2
F7
F8
F36
F37
F19
F32
F33
as.dist((1 - w)/2)
hclust (*, "complete")

```
head(gr,50)
```

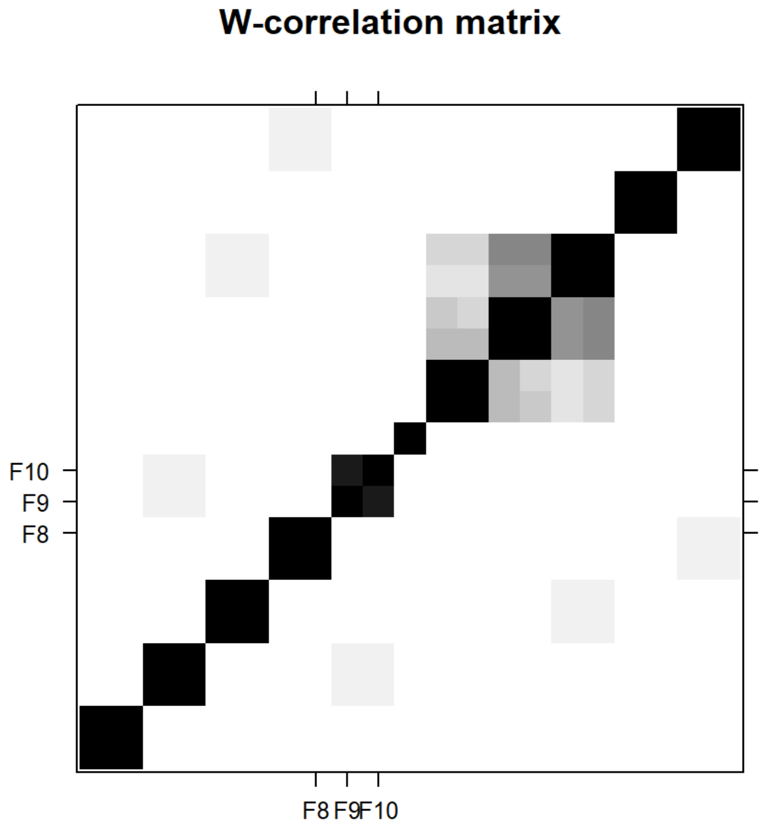
```
## $`1`
## [1] 1 2
##
## $`2`
## [1] 3 4
##
## $`3`
## [1] 5 6
##
## $`4`
## [1] 7 8
##
## $`5`
## [1]  9 10 11
##
## $`6`
## [1] 12 17 18
##
## $`7`
## [1] 13 14 15 16
##
## $`8`
## [1] 19
##
## $`9`
## [1] 20 21
##
## $`10`
## [1] 22 27 30 31
##
## $`11`
## [1] 23 24 25 26
##
## $`12`
## [1] 28 29
##
## $`13`
## [1] 32 33
##
## $`14`
## [1] 34 35
##
## $`15`
## [1] 36 37
```

Разделился скорее шум на несколько кластеров, чем периодики друг от друга. И в нашем случае результат этой автоматической группировки почти невозможно адекватно интерпретировать. Разбито на пары, но совершенно не ясно, какие пары шум, а какие – нет (раз в нашем случае нельзя сказать, что все, что по номеру больше – шум).

4 Улучшение разделимости

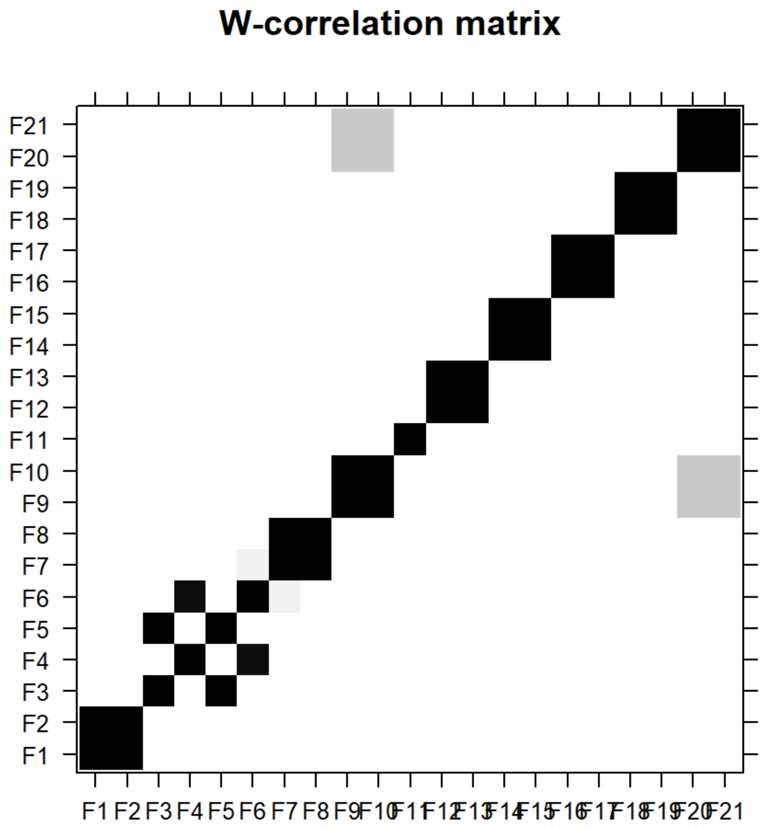
Итак, выделили сезонность. Попробуем подкрутить, чтобы гармоники лучше разделились. Удобнее снова смотреть на матрицу взвешенных корреляций (сигнала).

```
plot(wcor(Women.detrend.ssa, groups = c(1:8,9:10,19,20:21,23:24,25:26,32:33,
36:37)),
     scales = list(at = c(8,9,10)))
```



Подозреваем, что 20-21,23-24,25-26 – три смешавшиеся гармоники. Проблема с сильной разделимостью, собственные числа у них почти равны.

```
Wiossa<-iossa(Women.detrend.ssa, nested.groups = list(3:4,5:6,7:8,9:10,20:21
,23:24,25:26,36:37))
plot(wcor(Wiossa, groups = c(1:8,9:10,19,20:21,23:24,25:26,32:33,36:37)))
```



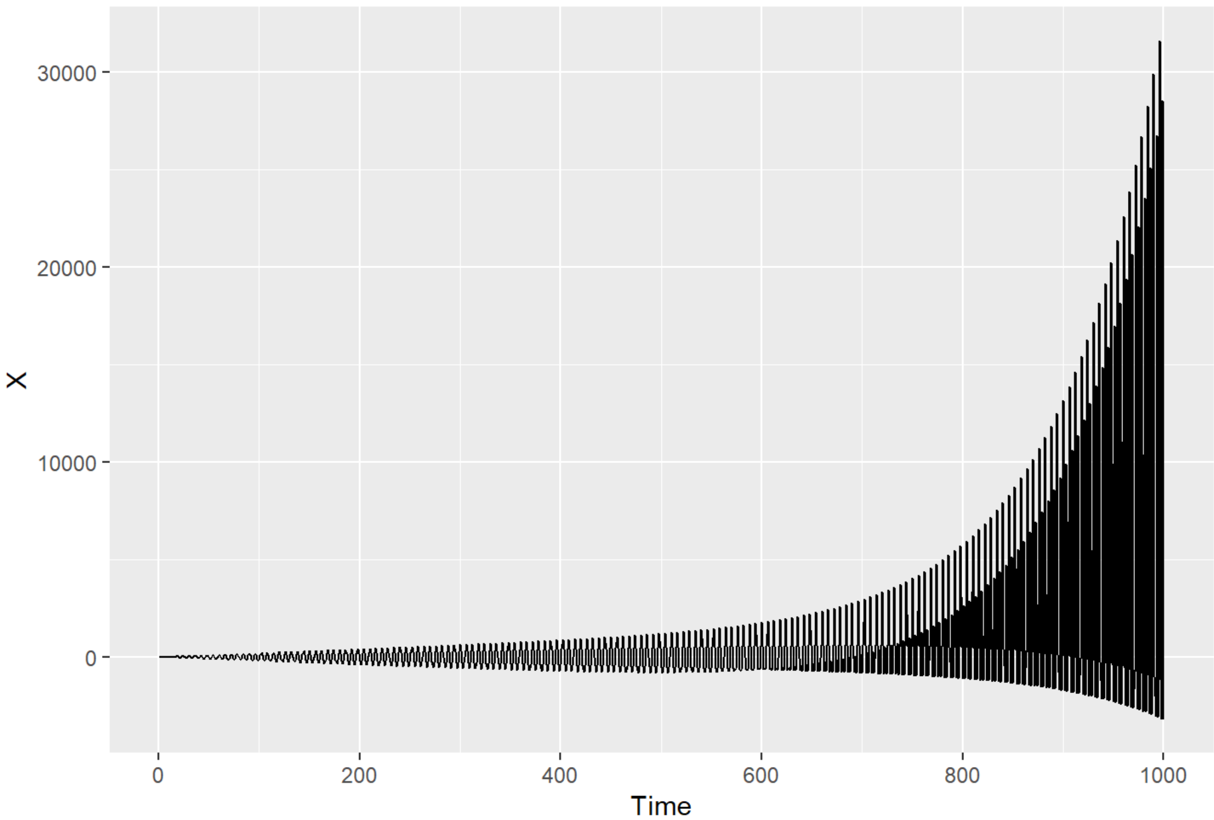
Некоторые компоненты поменялись местами. Матрица взвешенных корреляций почистилась, но закоррелировались компоненты 9:10 и 36:37. (fossa тоже не помогла)

5 Корни характеристического многочлена

5.1 Моделированный ряд

Рассмотрим ряд
$x_n = 0.4e^{n/100} + 3cos(2\pi * 1/2 * n) + e^{n/100}cos(2\pi * 4/12 * n) + 2ncos(2 * \pi * 1/6 * n)$.

```
N<-1:1000
X<-ts(0.4*exp(N/100)+3*cos(2*pi*1/2*N)+exp(N/100)*cos(2*pi*4/12*N)+2*N*cos(2
*pi*1/6*N))
autoplot(X)
```



Ранг ряда равен 8 (1+1+2+4). По порядку. Корень характеристического многочлена первого слагаемого равен $e^{1/100} \approx 1.01005$, т.е. почти единица, будет на вещественной оси чуть правее единичной окружности. Второе слагаемое ранга 1 (частота косинуса = 1/2), это будет вещественный корень (-1). Третье слагаемое – экспоненциально модулированный косинус, имеет ранг 2 и два сопряженных корня, чуть вылезающие за единичную окружность (угол $2\pi/3$). Ранг четвертого слагаемого равен 4. Тут уже сложнее сказать, какие будут корни. Предполагаем, что это будет пара сопряженных комплексных корней кратности 2 (линейная функция дает кратность).

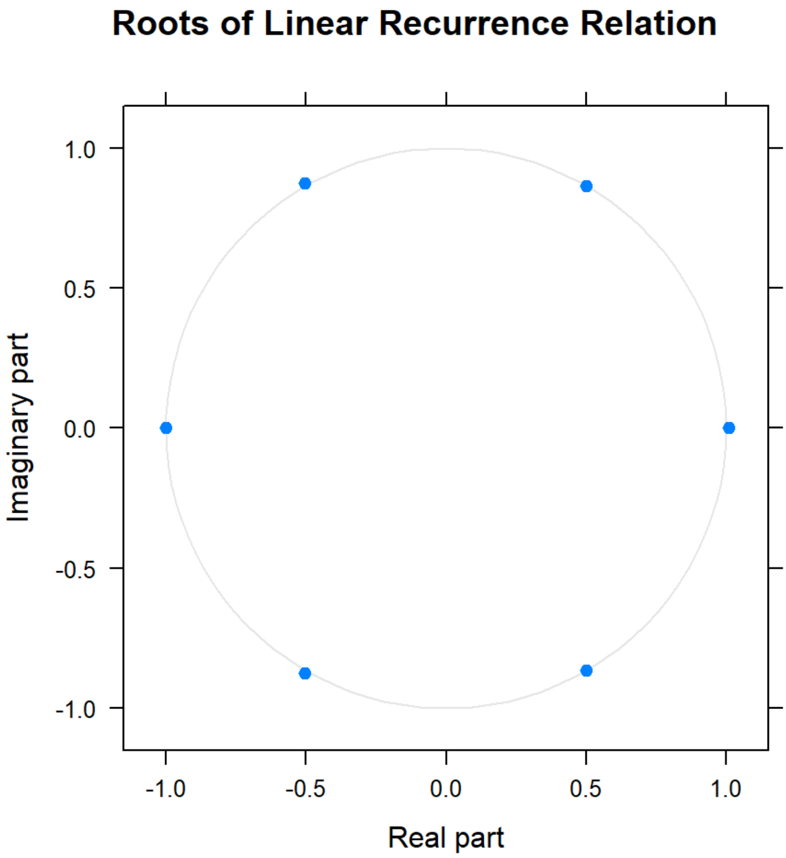
```
d<-8
X.ssa<-ssa(X,L=(d+1))
l <- lrr(X.ssa,groups = list(1:d))
r <- roots(l)
Mod(r)
```

```
## [1] 1.0100502 1.0100502 1.0100502 1.0000007 1.0000007 1.0000000 0.9999993
## [8] 0.9999993
```

```
Arg(r)
```

```
## [1]  2.094395 -2.094395  0.000000  1.047198 -1.047198  3.141593  1.047197
## [8] -1.047197
```

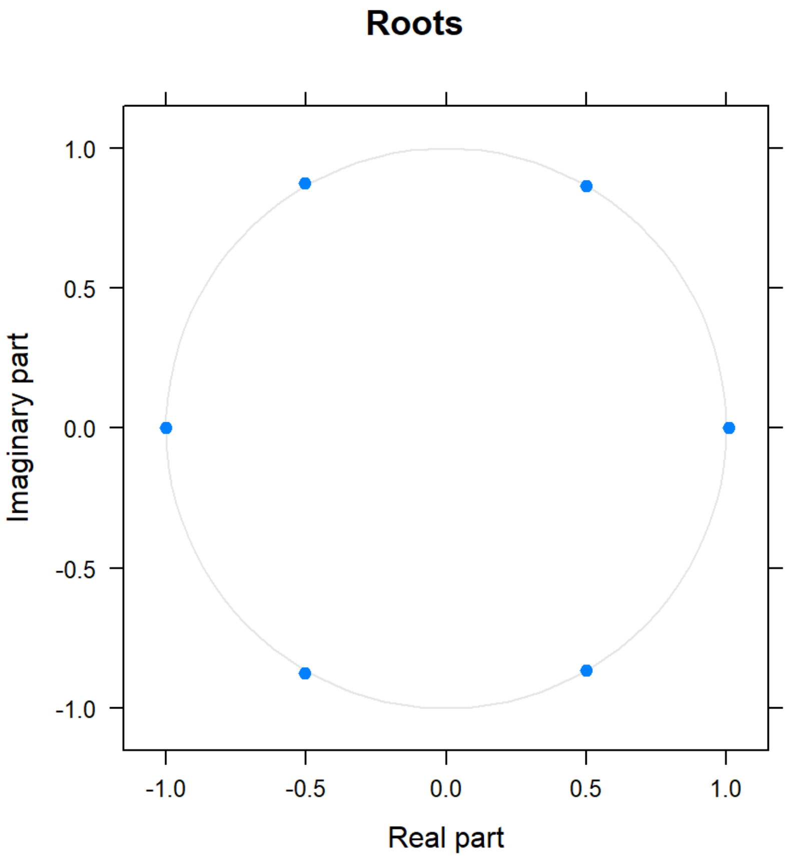
```
plot(l)
```



Корни 4го слагаемого очень близки друг к другу. Остальное все тоже совпадает с предположениями.

Сравним с тем, что выдает esprit. Возьмем $L = 500$.

```
X.ssa<-ssa(X,L=500)
par<-parestimate(X.ssa, groups=list(1:d), method = 'esprit')
plot(par)
```




```
print(par)
```

```
##    period     rate   |    Mod      Arg   |     Re         Im
##     3.000   0.010000 |  1.01005   2.09 | -0.50503   0.87473
##    -3.000   0.010000 |  1.01005  -2.09 | -0.50503  -0.87473
##       Inf   0.010000 |  1.01005  -0.00 |  1.01005  -0.00000
##     6.000   0.000000 |  1.00000   1.05 |  0.50000   0.86603
##    -6.000   0.000000 |  1.00000  -1.05 |  0.50000  -0.86603
##    -2.000   0.000000 |  1.00000  -3.14 | -1.00000  -0.00000
##     6.000  -0.000000 |  1.00000   1.05 |  0.50000   0.86603
##    -6.000  -0.000000 |  1.00000  -1.05 |  0.50000  -0.86603
```

Вроде все так.

Тут (для удобства выпишем)“”

1,2 – $e^{n/100} cos(2\pi * 4/12 * n)$,

3 – $e^{n/100}$,

4,5,7,8 – $2ncos(2 * \pi * 1/6 * n)$,

6 – $cos(2\pi * 1/2 * n) = (-1)^n$.

Решаем систему линейных уравнений.

```
mu<-par$roots
W<-t(matrix(mu,nrow = d,ncol = d))^(1:d)
C<-solve(W,X[1:d])
print(C)
```

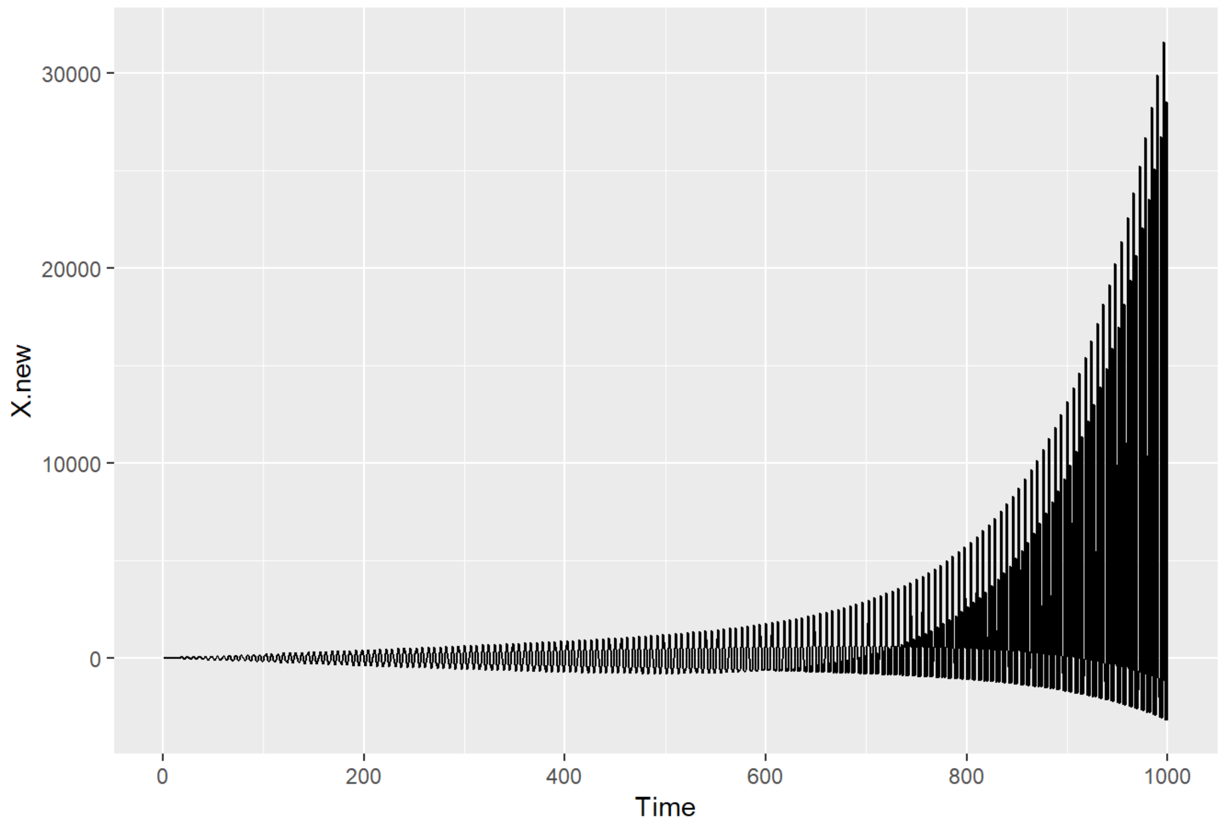
```
## [1]           0.5-         0i           0.5+         0i           0.4+
0i
## [4]  117177355.4+152164610i  116515238.1-150861735i           3.0-
0i
## [7] -117177355.4-152164610i -116515238.1+150861735i
```

Это коэффициенты для формулы в комплексной форме. Мы генерировали через вещественную, так что и сравнивать удобнее в ней.

Со слагаемыми первого ранга все просто, полученные коэффиценты стоят и в нужной (вещественной) формуле (0.4 перед $e^{1/100}$ и 3 перед $cos(\pi n)$). Второму слагаемому соответствуют вещественные коэффициенты, поэтому $C_1\mu_1^n + C_2\mu_2^n = 2Re(C_1)\rho^n \cos(2\pi n\omega) = e^{n/100}\cos(2\pi n * 1/3)$. (смотрим тут на модули и частоты соответствующих корней). С последним пока не особо получается разобраться. Понятно, что там надо учитывать кратность и все такое, но пока что получилось только запутаться.

Если просто сгенерировать полученный ряд и сравнить с исходным внешне:

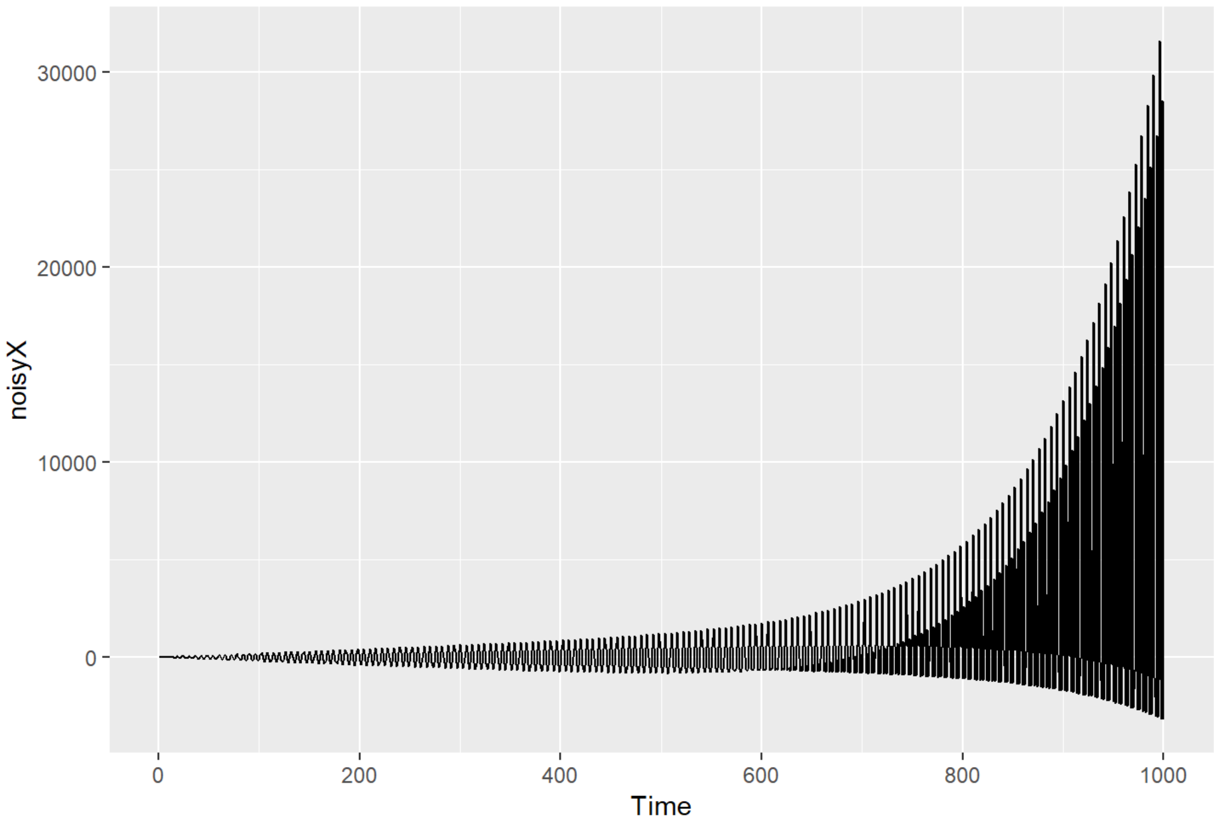
```
W.full<-t(matrix(mu,nrow = d,ncol = length(N)))^N
X.new<-ts(W.full%*%C)
autoplot(X.new)
```



Да, очень похоже.

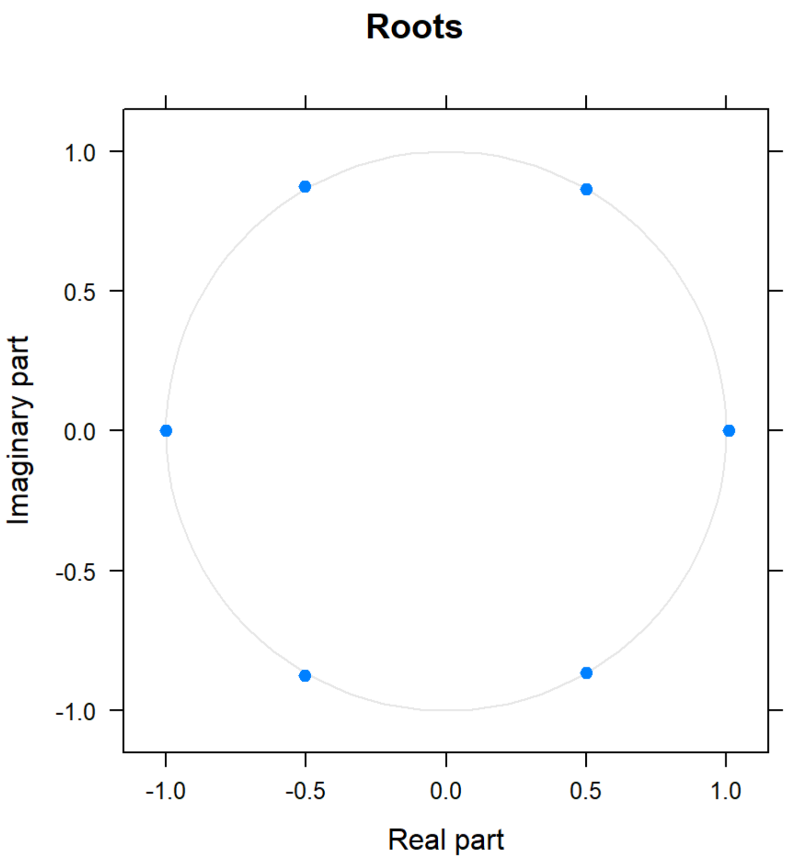
Добавим к сигналу шум:

```
noisyX<-X+rnorm(length(X),0,10)
autoplot(noisyX)
```



Находим корни:

```
d<-8
noisyX.ssa<-ssa(noisyX,L=500)
npar<-parestimate(noisyX.ssa, groups=list(1:d), method = 'esprit')
plot(npar)
```



```
print(npar)
```

```
##    period     rate   |    Mod     Arg  |     Re        Im
##     3.000   0.010001 |  1.01005   2.09 | -0.50502   0.87473
##    -3.000   0.010001 |  1.01005  -2.09 | -0.50502  -0.87473
##       Inf   0.009996 |  1.01005   0.00 |  1.01005   0.00000
##     5.999   0.000154 |  1.00015   1.05 |  0.49996   0.86623
##    -5.999   0.000154 |  1.00015  -1.05 |  0.49996  -0.86623
##     2.000  -0.000017 |  0.99998   3.14 | -0.99998   0.00000
##     6.001  -0.000164 |  0.99984   1.05 |  0.50006   0.86580
##    -6.001  -0.000164 |  0.99984  -1.05 |  0.50006  -0.86580
```

Получили примерно то же, что и раньше.

Находим коэффициенты.

```
mu<-npar$roots
W.full<-t(matrix(mu,nrow = d,ncol = length(N)))^N

noisyC<-ginv(W.full)%*%noisyX
print(noisyC)
```

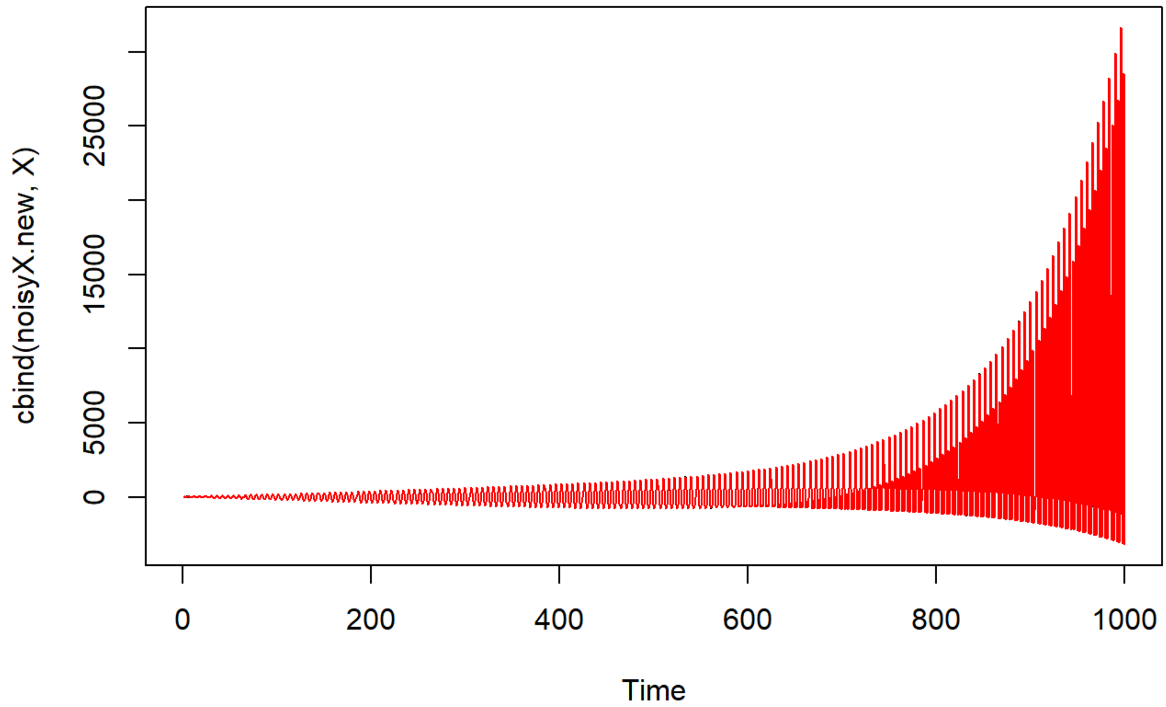
```
##                              [,1]
## [1,]  4.994697e-01+6.308636e-04i
## [2,]  4.994697e-01-6.308636e-04i
## [3,]  4.016777e-01-0.000000e+00i
## [4,]  1.672271e+03-1.573314e+03i
## [5,]  1.672271e+03+1.573314e+03i
## [6,]  3.623154e+00+0.000000e+00i
## [7,] -1.673959e+03+1.573065e+03i
## [8,] -1.673959e+03-1.573065e+03i
```

```
print(C)
```

```
## [1]          0.5-        0i          0.5+        0i          0.4+
0i
## [4]  117177355.4+152164610i  116515238.1-150861735i          3.0-
0i
## [7] -117177355.4-152164610i -116515238.1+150861735i
```

Все, кроме коэффициентов для последнего слагаемого, примерно совпадает. С последним не совсем понятно, являеется это ошибкой или нет. Может, у них внутри могут быть разные числа, важно только соотношения между ними. Возможно, это прояснится, если все-таки понять, как связаны коэффициенты комплексной формы этого слагаемого с коэффициентом перед $n\cos(2*\pi*1/6*n)$.

```
noisyX.new<-ts(W.full%*%noisyC)
plot.ts(cbind(noisyX.new,X), plot.type = 'single', col = c('black', 'red'))
```



Ну, вроде все хорошо, но у нас и шум был очень простой.

5.2 Реальный ряд

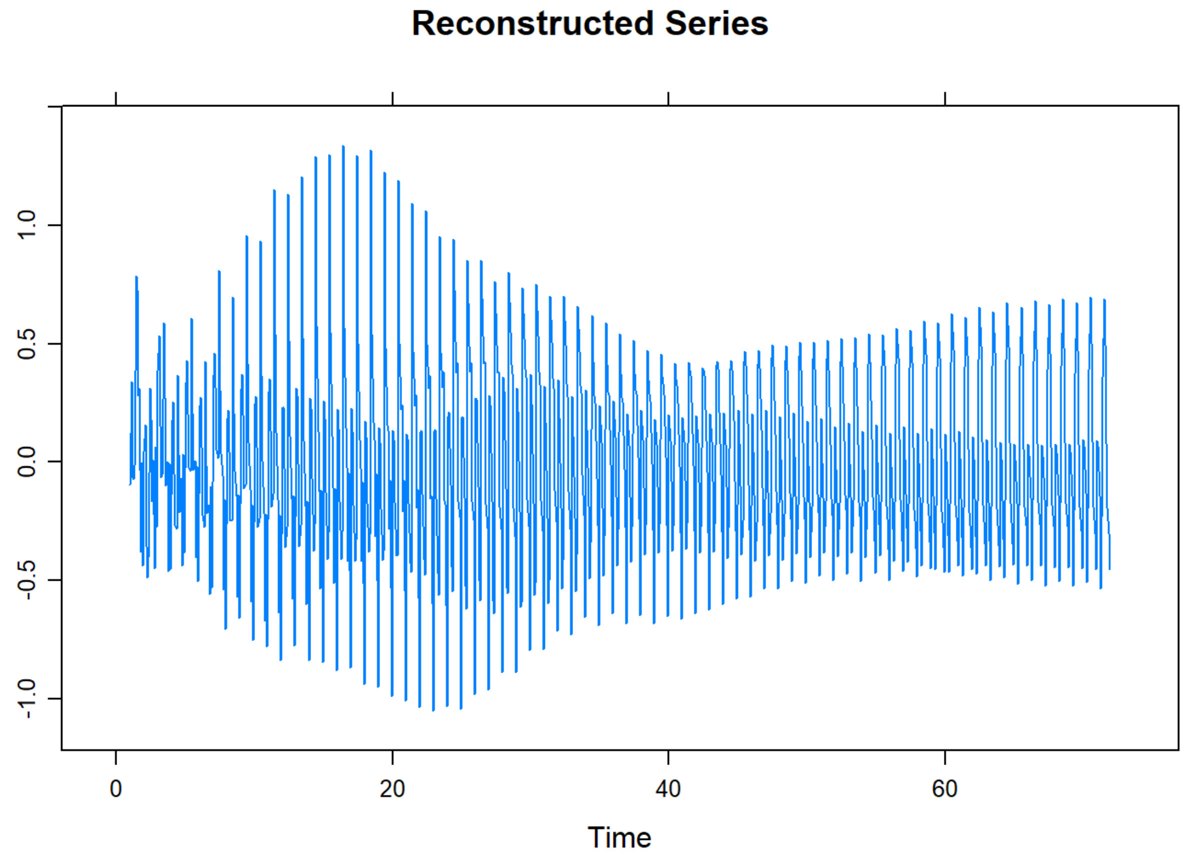
Пробуем с нашим рядом.

Будем проделывать это для сезонности. (Пробовала сделать для всего ряда, но там выходит очень уж громоздко: чтобы как-то побороть периодичность в шуме, пришлось взять очень много компонент (50) и такое ощущение, что в итоге очен много шума прихватили в сигнал).

```
Women.detrend.ssa<-Wiossa
Women.detrend.reconstruct <- reconstruct(Women.detrend.ssa,
                        groups = list(Seasonality = c(1:8,9:10, 19,20:21,23:24
,25:26,32:33,36:37)))
```

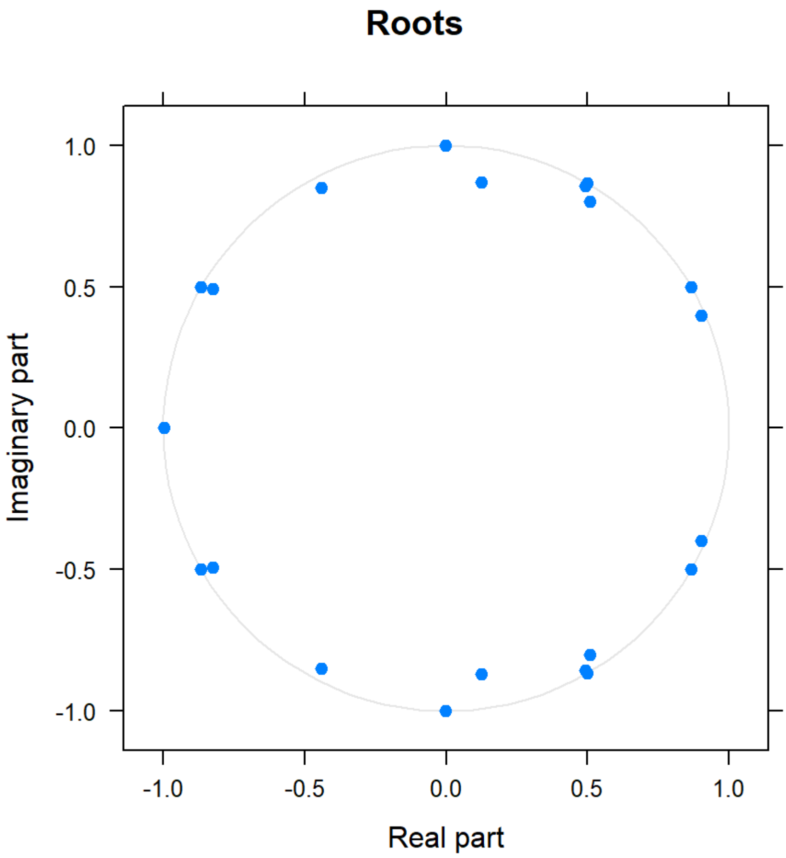
Выделенная сезонность:

```
plot(Women.detrend.reconstruct , add.residuals = FALSE, add.original = FALSE
,
     plot.method = "xyplot",
     superpose = TRUE)
```



По числу взятых в сезонность компонент считаем, что ранг сигнала равен 21. (Хотя тут сложно использовать термин “ранг”“, учитывая, что компоненты мы выбирали не по порядку, а сейчас”полагаем" нулями собственные числа компонент, которые больше тех ненулевых, которые мы выбрали в сигнал. Ну ладно.)

```
d<-21
Wpar<-parestimate(Women.detrend.ssa, groups=list(c(1:8,9:10, 19,20:21,23:24,
25:26,32:33,36:37)), method = 'esprit')
plot(Wpar)
```



На рисунке видим, что тренда нет, видим пилу и модулированные периодики соответствующих порядков. Заметим, что многие из них есть пара или даже тройка близлежащих сопряженных корней. Понятно, что в реальном случае мы не могли получить кратные корни, как в моделированном примере, но такие близкие корни могут означать, что истинная модуляция не экспоненциальная, а имеет более сложную форму. Получается, что такая модуляция будет как бы аппроксимироваться суммой близких экспонентально моделированных гармоник.

```
print(Wpar)
```

```
##    period     rate    |    Mod     Arg  |      Re        Im
##    12.034  -0.000182 |  0.99982   0.52 |  0.86661   0.49862
##   -12.034  -0.000182 |  0.99982  -0.52 |  0.86661  -0.49862
##     6.001  -0.000582 |  0.99942   1.05 |  0.49981   0.86546
##    -6.001  -0.000582 |  0.99942  -1.05 |  0.49981  -0.86546
##     3.999  -0.001384 |  0.99862   1.57 | -0.00057   0.99862
##    -3.999  -0.001384 |  0.99862  -1.57 | -0.00057  -0.99862
##     2.399  -0.001946 |  0.99806   2.62 | -0.86489   0.49808
##    -2.399  -0.001946 |  0.99806  -2.62 | -0.86489  -0.49808
##     2.000  -0.002106 |  0.99790   3.14 | -0.99790   0.00000
##    15.101  -0.012831 |  0.98725   0.42 |  0.90302   0.39902
##   -15.101  -0.012831 |  0.98725  -0.42 |  0.90302  -0.39902
##     5.988  -0.013255 |  0.98683   1.05 |  0.49162   0.85565
##    -5.988  -0.013255 |  0.98683  -1.05 |  0.49162  -0.85565
##     2.414  -0.039195 |  0.96156   2.60 | -0.82514   0.49371
##    -2.414  -0.039195 |  0.96156  -2.60 | -0.82514  -0.49371
##     3.069  -0.043370 |  0.95756   2.05 | -0.43894   0.85103
##    -3.069  -0.043370 |  0.95756  -2.05 | -0.43894  -0.85103
##     6.248  -0.052756 |  0.94861   1.01 |  0.50801   0.80112
##    -6.248  -0.052756 |  0.94861  -1.01 |  0.50801  -0.80112
##     4.396  -0.130980 |  0.87724   1.43 |  0.12385   0.86845
##    -4.396  -0.130980 |  0.87724  -1.43 |  0.12385  -0.86845
```

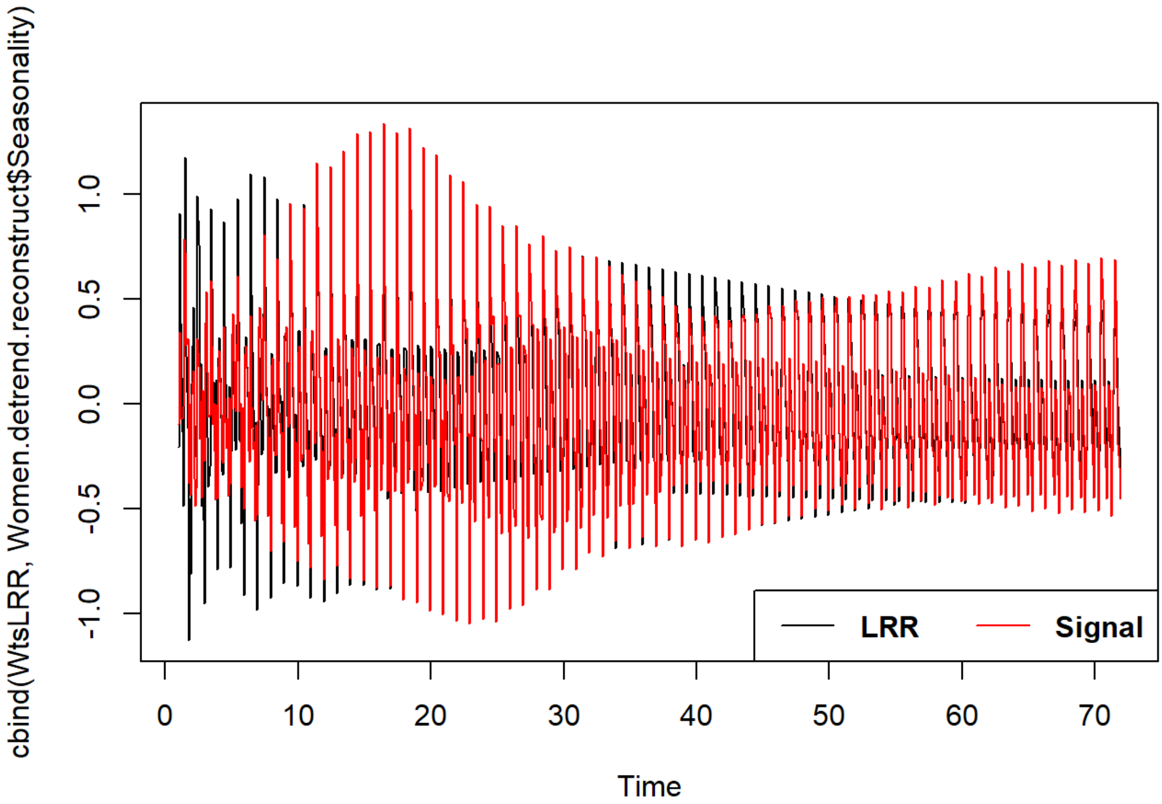
Находим коэффициенты:

```
mu<-Wpar$roots
N<-1:length(Women.detrend)
W.full<-t(matrix(mu,nrow = d,ncol = length(N)))^N
C<-ginv(W.full)%*%Women.detrend
print(C)
```

```
##                            [,1]
##   [1,] -0.15398222-0.03873846i
##   [2,] -0.15398222+0.03873846i
##   [3,]  0.02666346-0.21901948i
##   [4,]  0.02666346+0.21901948i
##   [5,] -0.17871227+0.05164097i
##   [6,] -0.17871227-0.05164097i
##   [7,] -0.18881805+0.08702148i
##   [8,] -0.18881805-0.08702148i
##   [9,] -0.07113974+0.00000000i
##  [10,] -0.00774397-0.10809445i
##  [11,] -0.00774397+0.10809445i
##  [12,]  0.12253905+0.13196577i
##  [13,]  0.12253905-0.13196577i
##  [14,]  0.15784156+0.07658942i
##  [15,]  0.15784156-0.07658942i
##  [16,] -0.01008753+0.02910608i
##  [17,] -0.01008753-0.02910608i
##  [18,] -0.00685809-0.30980695i
##  [19,] -0.00685809+0.30980695i
##  [20,]  0.04361390+0.39270277i
##  [21,]  0.04361390-0.39270277i
```

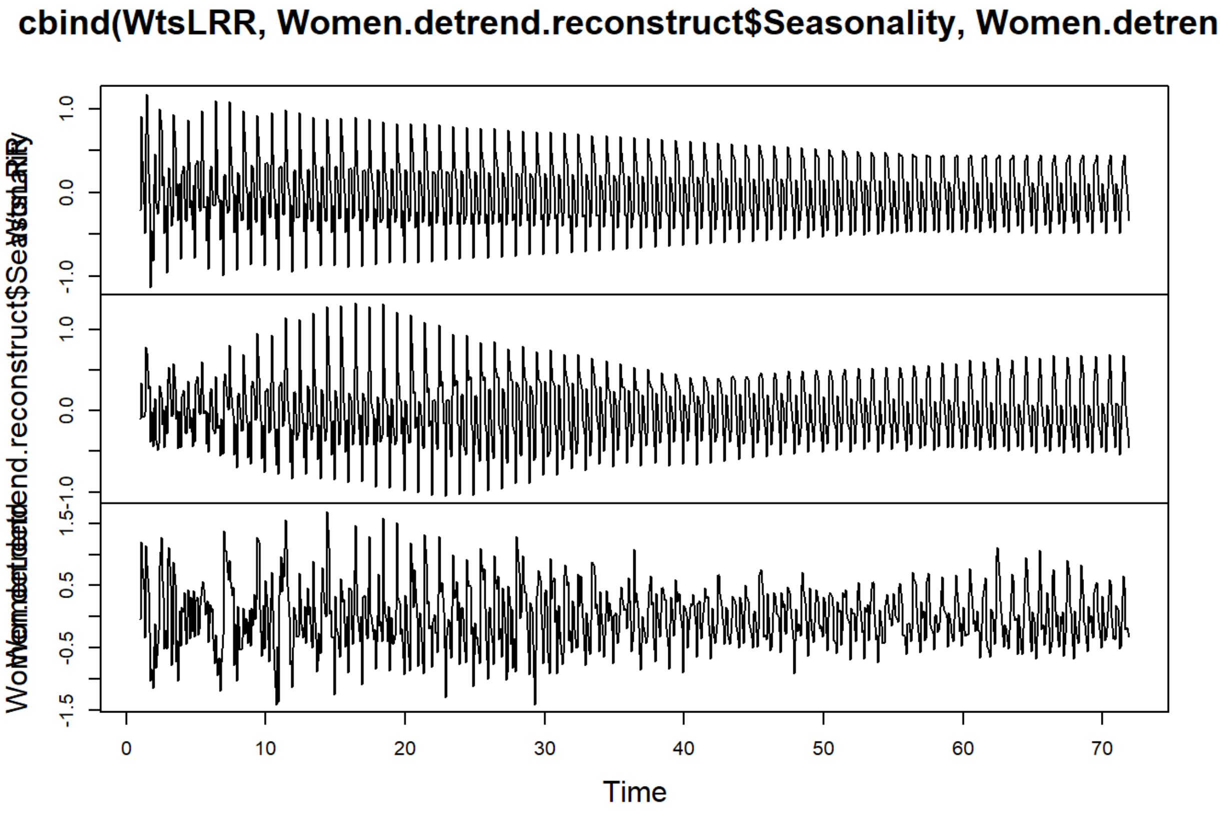
По полученной комплексной формуле генерируем ряд и сравниваем его с выделенным сигналом:

```
WtsLRR<-ts(as.numeric(W.full%*%C),frequency = 12)
plot.ts(cbind(WtsLRR,Women.detrend.reconstruct$Seasonality), plot.type = 'si
ngle',col = c('black', 'red'))
legend("bottomright",paste(c("LRR","Signal")), col = c('black', 'red'),lty=1
,lw=c(1,1),text.font = 2,horiz=TRUE)
```



Кажется, что сгенерированный по формуле ряд слишком грубо оценивает сезонность. Сезонность, которую мы выделили, имеет более сложную форму и это соответствует виду исходного ряда:

```
plot.ts(cbind(WtsLRR,Women.detrend.reconstruct$Seasonality,Women.detrend))
```


cbind(WtsLRR, Women.detrend.reconstruct$Seasonality, Women.detren
1.0
0.0
-1.0
1.0
0.0
-1.0
1.5
0.5
-0.5
-1.5
0
10
20
30
40
50
60
70
Time