Лабораторная работа №1
«Временные ряды и случайные процессы»

Задание. Даны значения временного ряда. Найти выборочное среднее значение. Найти выборочную дисперсию значений.

В качестве элементов временного ряда будут использованы возвраты значений цен следующих фирм:

«Алроса» (ALRS),
«Русолово» (ROLO),
«Институт стволовых клеток человека» (ISKJ),
«Газпром» (GAZP),
«ЛСР» (LSRG),
«ПИК» (PIKK),
«СберБанк» (SBER),
«Московская Биржа» (MOEX),
«Магнит» (MGNT),
«Черкизово» (GCHE).

Рассматриваемый период: с 27.06.2014 по 19.03.2021. Приведём краткое описание компаний. Аббревиатура, указанная в скобках, – листинг на бирже.

«Алроса» – российская группа алмазодобывающих компаний, занимающая лидирующую позицию в мире по объёму добычи алмазов. Корпорация занимается разведкой месторождений, добычей, обработкой и продажей алмазного сырья. Основная деятельность сосредоточена в Якутии, а также Архангельской области и Африке.

«Русолово» – публичное акционерное общество, являющееся российским предприятием горнодобывающей и металлургической промышленности. Предприятие специализируется на добыче и переработке олова, цинка и медной руды.

«Институт стволовых клеток человека» – российский биотехнологический холдинг, основанный в 2003 году. Ведёт разработки и предоставляет услуги, связанные с клеточными, генными и постгеномными технологиями, развивая сферу персонализированной и профилактической медицины.

«Газпром» – публичное акционерное общество, являющееся российской транснациональной энергетической компанией, более половины акций которой принадлежит государству. Является холдинговой компанией Группы «Газпром». Непосредственно ПАО «Газпром» осуществляет только продажу природного газа и сдаёт в аренду свою газотранспортную систему.

«ЛСР» – российская компания, работающая в сфере производства стройматериалов, развития и строительства недвижимости. Головной офис – в Санкт-Петербурге. По состоянию на 1 августа 2021 года является вторым по объёму текущего строительства застройщиком в России.

«ПИК» – российская строительная компания со штаб-квартирой в Москве.

«СберБанк» – российский финансовый конгломерат, крупнейший универсальный банк России и Восточной Европы. По итогам 2019 года у Сбербанка 96,2 млн активных частных клиентов и 2,6 млн активных

корпоративных клиентов. Среди крупнейших банков мира по размеру активов находится в восьмом десятке.

«Московская Биржа» – крупнейший российский биржевой холдинг, созданный в 2011 году в результате слияния ММВБ, основанной в 1992 году, и биржи РТС, открытой в 1995 году.

«Магнит» – сеть розничных магазинов, третья по выручке частная компания России. Обладает также тепличным комплексом, собственным автопарком на 5,7 тыс. автомобилей.

«Черкизово» – публичное акционерное общество, являющееся российской группой компаний, занимающихся производством мясной продукции. Крупнейший производитель и переработчик мяса птицы, свинины и комбикормов в России.

Для расчёта выборочного среднего значения используем формулу $\overline{R}_c = \frac{1}{N}\sum_{i=1}^{N} R_i$ (смещённая оценка) или $\overline{R}_н = \frac{1}{N-1}\sum_{i=1}^{N} R_i$ (несмещённая оценка). Для расчёта выборочной дисперсии используем формулу $\overline{V}_c = \frac{1}{N}\sum_{i=1}^{N}\left(R_i - \overline{R}\right)^2$ (смещённая оценка) или $\overline{V}_н = \frac{1}{N-1}\sum_{i=1}^{N}\left(R_i - \overline{R}\right)^2$ (несмещённая оценка). Несмещенная оценка в математической статистике – это точечная оценка, математическое ожидание которой равно оцениваемому параметру. В противном случае оценка называется смещённой. Рассмотрим пример.

Пусть $N = 3, R_1 = 0.1, R_2 = 0.2, R_3 = 0.3$. Тогда

$$\overline{R}_c = \frac{1}{3}(0.1 + 0.2 + 0.3) = 0.2, \overline{R}_н = \frac{1}{2}(0.1 + 0.2 + 0.3) = 0.3,$$

$$\overline{V}_c = \frac{1}{3}\left((0.1 - 0.2)^2 + (0.2 - 0.2)^2 + (0.3 - 0.2)^2\right) = 0.0066,$$

$$\overline{V}_н = \frac{1}{2}\left((0.1 - 0.3)^2 + (0.2 - 0.3)^2 + (0.3 - 0.3)^2\right) = 0.025.$$

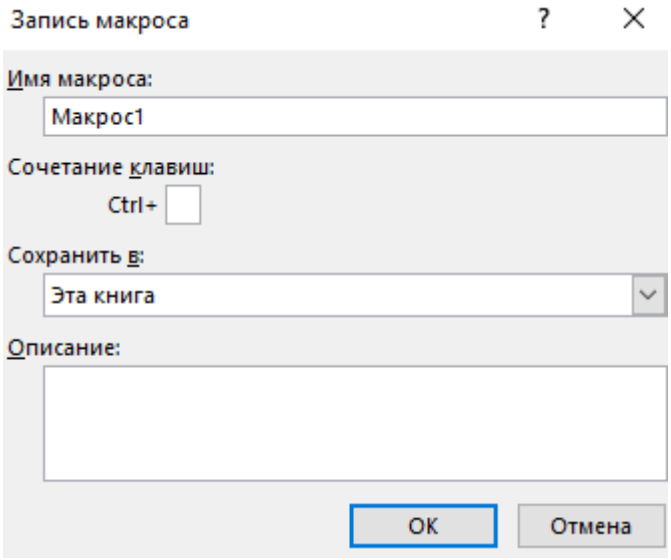
Для расчёта смещённой и несмещённой оценки элементов временного ряда можно использовать методы языка VBA Excel.

Язык VBA (расшифровывается как Visual Basic for Application) разработан компанией Microsoft и предназначен для автоматизации процессов в MS Office. VBA широко используется в Excel, Word, Access и других программах пакета. Суть языка заключается в оперировании объектами, что относит его к объектно-ориентированному программированию. Под объектом понимается элемент, структурная частица Excel: книга, лист, диапазон, ячейка. Данные объекты имеют следующую иерархию: Application→Workbooks→Worksheets. Например, обратиться к ячейке «A1» на листе можно одним из способов:

Application.Workbooks(«Имя книги»).Worksheets(«Имя листа»).Range(«A1»)
Application.Workbooks(«Имя книги»).Worksheets(«Имя листа»).Cells(1,1)
Application.Workbooks(«Имя книги»).Worksheets(«Имя листа»).Cells(1, «A»)

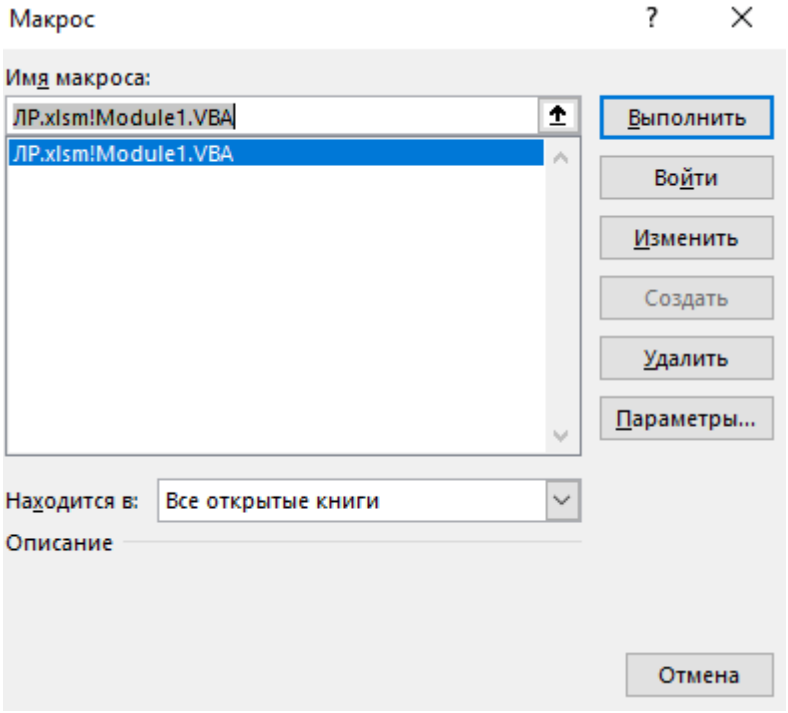
Для вставки макроса в VBA Excel воспользуемся командой Вид→Вставка

макроса. В диалоговом окне

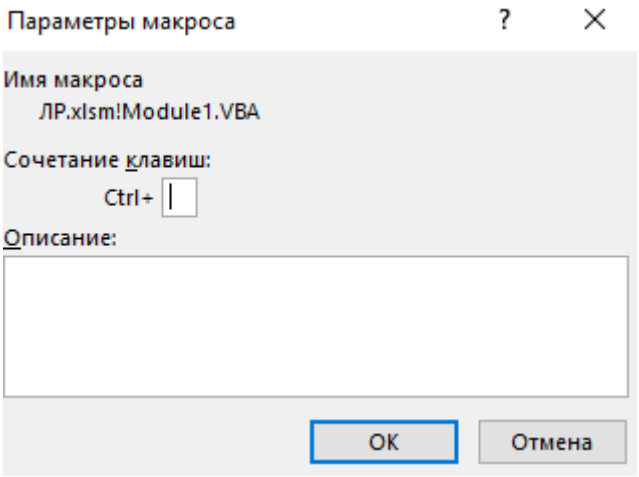


необходимо задать имя макроса, а также сочетание клавиш, начиная с клавиши Ctrl, необходимое для обеспечения быстрого доступа к макросу. Кроме этого, необходимо указать путь для сохранения готового макроса. Можно привести краткое описание макроса.

Открыть уже готовый макрос можно с помощью команды Вид→Макросы. В диалоговом окне



необходимо выбрать имя макроса, и нажать кнопку «Войти». Можно также запустить выбранный макрос, нажав кнопку «Выполнить». С помощью кнопки «Изменить» можно изменить имя макроса. С помощью кнопки «Удалить» можно удалить макрос. С помощью кнопки «Параметры» можно задать сочетание клавиш, начиная с Ctrl, для быстрого доступа к макросу, а также привести его краткое описание.



Приведём краткое описанные основных методов языка VBA Excel. Переменные, являющиеся константами, объявляются как Const. Процедура начинается со слов Sub «Имя процедуры» и заканчивается словами End Sub. Например, зададим константу N=298 (количество элементов в заданной выборке) в процедуре

```
Sub VBA
Const N = 298
End Sub
```

Существуют разные типы данных:

-числовой: Byte , Integer, Long, Currency, Single, Double

-текстовый: String

-дата: Date

-логический: Boolean

-объект: Object

-другой: Variant.

Описание переменных происходит с помощью Dim. Например, статический массив размера N можно задать как Dim A(1 to N) as Integer (Double). Динамический массив размера можно задать как Dim A() as Integer (Double). Определить его размер можно с помощью ReDim A(N). По умолчанию все массивы начинаются с 0. Условный оператор записывается как If Then ElseIf Then Else End If. Циклы оформляются как For To Next и Do While Loop.

Пусть в диапазоне A1:A298 заданы значения возвратов компании «Алроса». Найдем минимальное и максимальное значения возврата, а также моменты времени их достижения. Введём соответствующие переменные и присвоим им начальные значения.

```
value_max = Worksheets("1").Cells(1, 2).Value: ind_max = 1
value_min = Worksheets("1").Cells(1, 2).Value: ind_min = 1
For i = 2 To N
If Worksheets("1").Cells(i, 2).Value > value_max Then value_max = Worksheets("1").Cells(i, 2).Value: ind_max = i
If Worksheets("1").Cells(i, 2).Value < value_min Then value_min = Worksheets("1").Cells(i, 2).Value: ind_min = i
Worksheets("1").Cells(3, 4).Value = value_max
```

```
Worksheets("1").Cells(3, 5).Value = ind_max
Worksheets("1").Cells(3, 3).Value = "maximal value"
Worksheets("1").Cells(4, 4).Value = value_min
Worksheets("1").Cells(4, 5).Value = ind_min
Worksheets("1").Cells(4, 3).Value = "minimal value"
```

Итак, максимальное значение элементов выборки равно 0,174137375 и достигается в момент времени 219; минимальное значение элементов выборки равно -0,12244898 и достигается в момент времени 231.

Найдём $\overline{R}_c$ для элементов диапазона A1:A298:

```
s = 0
For i = 1 To N
s = s + Worksheets("1").Cells(i, 2).Value
Next
s = s / N
Worksheets("1").Cells(1, 3).Value = "sample average"
Worksheets("1").Cells(1, 4).Value = s
```

Найдём $\overline{V}_c$ для элементов диапазона A1:A298:

```
s1 = 0
For i = 1 To N
s1 = s1 + (Worksheets("1").Cells(i, 2).Value * Worksheets("1").Cells(i, 2).Value-
s)^2
Next
s1 = s1 / N
Worksheets("1").Cells(2, 3).Value = "sample variance"
Worksheets("1").Cells(2, 4).Value = s1
```

Итак, $\overline{R}_c = -0.001840035$, $\overline{V}_c = 0.001943609$.

Найдём $\overline{R}_н$ для элементов диапазона A1:A298:

```
s = 0
For i = 1 To N
s = s + Worksheets("1").Cells(i, 2).Value
Next
s = s / (N-1)
Worksheets("1").Cells(5, 3).Value = "sample average"
Worksheets("1").Cells(5, 4).Value = s
```

Найдём $\overline{V}_н$ для элементов диапазона A1:A298:

```
s1 = 0
For i = 1 To N
s1 = s1 + (Worksheets("1").Cells(i, 2).Value-s)^2
Next
s1 = s1 / (N-1)
Worksheets("1").Cells(6, 3).Value = "sample variance"
Worksheets("1").Cells(6, 4).Value = s1
```

Итак, $\overline{R}_н = -0.001840623$, $\overline{V}_н = 0.001950153$.

Лабораторная работа №2

«Вероятностное описание временного ряда»

Задание. Даны значения временных рядов, отражающих доходности 10 фирм. Найти выборочную ковариационную матрицу.

Выборочная ковариационная матрица находится из следующей формулы: $C = \frac{1}{N}\sum_{i=1}^{N} R_i R_i^T - \overline{R}\,\overline{R}^T$ и является квадратной матрицей размера M. Рассмотрим примеры.

Пусть $N = 2, M = 2, R_1 = \begin{pmatrix} 0.1 \\ 0.2 \end{pmatrix}, R_2 = \begin{pmatrix} 0.3 \\ 0.4 \end{pmatrix}$.

Тогда

$$\overline{R} = \frac{1}{2}\left(\begin{pmatrix} 0.1 \\ 0.2 \end{pmatrix} + \begin{pmatrix} 0.3 \\ 0.4 \end{pmatrix}\right) = \begin{pmatrix} 0.2 \\ 0.3 \end{pmatrix},$$

$$C = \frac{1}{2}\left(\begin{pmatrix} 0.1 \\ 0.2 \end{pmatrix}\begin{pmatrix} 0.1 & 0.2 \end{pmatrix} + \begin{pmatrix} 0.3 \\ 0.4 \end{pmatrix}\begin{pmatrix} 0.3 & 0.4 \end{pmatrix}\right) - \begin{pmatrix} 0.2 \\ 0.3 \end{pmatrix}\begin{pmatrix} 0.2 & 0.3 \end{pmatrix} =$$

$$= \frac{1}{2}\left(\begin{pmatrix} 0.01 & 0.02 \\ 0.02 & 0.04 \end{pmatrix} + \begin{pmatrix} 0.09 & 0.12 \\ 0.12 & 0.16 \end{pmatrix}\right) - \begin{pmatrix} 0.04 & 0.06 \\ 0.06 & 0.09 \end{pmatrix} =$$

$$= \frac{1}{2}\begin{pmatrix} 0.1 & 0.14 \\ 0.14 & 0.2 \end{pmatrix} - \begin{pmatrix} 0.04 & 0.06 \\ 0.06 & 0.09 \end{pmatrix} = \begin{pmatrix} 0.01 & 0.01 \\ 0.01 & 0.01 \end{pmatrix}$$

Заметим, что ковариационная матрица получилась вырожденной. Её определитель равен 0.

Пусть $N = 3, M = 2, R_1 = \begin{pmatrix} 0.1 \\ 0.2 \end{pmatrix}, R_2 = \begin{pmatrix} 0.3 \\ 0.4 \end{pmatrix}, R_3 = \begin{pmatrix} 0.8 \\ 0.9 \end{pmatrix}$.

Тогда

$$\overline{R} = \frac{1}{3}\left(\begin{pmatrix} 0.1 \\ 0.2 \end{pmatrix} + \begin{pmatrix} 0.3 \\ 0.4 \end{pmatrix} + \begin{pmatrix} 0.8 \\ 0.9 \end{pmatrix}\right) = \begin{pmatrix} 0.4 \\ 0.5 \end{pmatrix},$$

$$C = \frac{1}{-}\left(\begin{pmatrix} 0.1 \\ 0.2 \end{pmatrix}\begin{pmatrix} 0.1 & 0.2 \end{pmatrix} + \begin{pmatrix} 0.3 \\ 0.4 \end{pmatrix}\begin{pmatrix} 0.3 & 0.4 \end{pmatrix} + \begin{pmatrix} 0.8 \\ 0.9 \end{pmatrix}\begin{pmatrix} 0.8 & 0.9 \end{pmatrix}\right) - \begin{pmatrix} 0.4 \\ 0.5 \end{pmatrix}\begin{pmatrix} 0.4 & 0.5 \end{pmatrix} =$$

$$= \frac{1}{2}\left(\begin{pmatrix} 0.01 & 0.02 \\ 0.02 & 0.04 \end{pmatrix} + \begin{pmatrix} 0.09 & 0.12 \\ 0.12 & 0.16 \end{pmatrix} + \begin{pmatrix} 0.64 & 0.72 \\ 0.72 & 0.81 \end{pmatrix}\right) - \begin{pmatrix} 0.16 & 0.2 \\ 0.2 & 0.25 \end{pmatrix} =$$

$$= \frac{1}{2}\begin{pmatrix} 0.74 & 0.84 \\ 0.84 & 1.01 \end{pmatrix} - \begin{pmatrix} 0.16 & 0.2 \\ 0.2 & 0.25 \end{pmatrix} = \frac{1}{2}\begin{pmatrix} 0.42 & 0.44 \\ 0.44 & 0.51 \end{pmatrix}$$

Заметим, что ковариационная матрица получилась невырожденной. Её определитель равен 0.0103.

Для частного случая, когда $M = 1, R = (\rho_i)_{i=1}^{N}$ и необходимо найти ковариацию элементов ρ_i и ρ_j, используют формулу:

$$\operatorname{cov}(\rho_i, \rho_j) = E((\rho_i - E\rho_i)(\rho_j - E\rho_j)) = E(\rho_i \rho_j) - E\rho_i E\rho_j,$$

$i = 1,\ldots,N;\ j = 1,\ldots,N.$

Перечислим основные свойства ковариации.

1. Если случайные величины ρ_i и ρ_j независимы, то $\operatorname{cov}(\rho_i, \rho_j) = E\rho_i E\rho_j - E\rho_i E\rho_j = 0$.
2. Если $i = j$, то $\operatorname{cov}(\rho_i, \rho_i) = E(\rho_i \rho_i) - E\rho_i E\rho_i = D\rho_i$.
3. $\operatorname{cov}(\rho_j, \rho_i) = E(\rho_j \rho_i) - E\rho_j E\rho_i = \operatorname{cov}(\rho_i, \rho_j)$.
4. $\operatorname{cov}(c\rho_i + x, d\rho_j + y) = cd \operatorname{cov}(\rho_i, \rho_j)$.
5. $\left|\operatorname{cov}(\rho_i, \rho_j)\right| \le \sqrt{D\rho_i D\rho_j}$
6. $\left|\operatorname{cov}(\rho_i, \rho_j)\right| = \sqrt{D\rho_i D\rho_j} \Leftrightarrow \rho_i = a\rho_j + b$
7. $D(\rho_i \pm \rho_j) = D\rho_i + D\rho_j \pm 2\operatorname{cov}(\rho_i, \rho_j)$

Для нахождения ковариационной матрицы использовались векторы значений возврата следующих фирм: ФосАгро (PHOR), Пятёрочка (FIVE), М.Видео (MVID), Рост (ROST), Fix Price (FIXP), Акрон (AKRN), Объединённая вагонная компания (UWGN), Детский мир (DSKY), Роскосмос (RKKE), ВТБ (VTBE). Приведём краткое описание компаний.

ФосАгро – публичное акционерное общество, российский химический холдинг.

Пятёрочка – ведущая компания современной розничной торговли.

М.Видео – публичное акционерное общество, российская торговая сеть по продаже бытовой техники и электроники. После слияния с сетью «Эльдорадо» в 2018 году оба бренда были сформированы в группу «М.Видео-Эльдорадо».

Рост – группа компаний, лидер рынка овощей защищённого грунта РФ.

Fix Price – российская сеть магазинов в формате «магазин фиксированных цен», управляющая компания ООО «Бэст Прайс». Сеть включает в себя более 4800 магазинов, работающих в более чем 1300 населённых пунктах в 79 регионах России, а также в Грузии, Казахстане, Латвии, Белоруссии, Узбекистане и Киргизии.

Акрон – группа компаний-производителей минеральных удобрений. Главный офис в Москве, основные предприятия в Великом Новгороде, Доргобуже, а также Северо-Западная Фосфорная компания в Мурманской области.

Объединённая вагонная компания – публичное акционерное общество, научно-производственная корпорация, российский производитель грузовых железнодорожных вагонов. Включена Министерством экономического развития РФ в перечень системообразующих организаций.

Детский мир – советская и российская сеть магазинов товаров для детей, созданная в 1947 году и ставшая крупнейшей.

Роскосмос – ракетно-космическая корпорация «Энергия» имени С.П.Королёва. Одно из ведущих предприятий космической промышленности

СССР и России. Главная организация корпорации находится в городе Королёве, филиал – на космодроме Байконур.

ВТБ – российский универсальный коммерческий банк с государственным участием. Банк ВТБ является головной структурой группы ВТБ.

Зададим константы

```
Const M = 10
Const N=298
```

Определим выборочную ковариационную матрицу как

```
Dim C(1 To M, 1 To M) As Double.
```

Приравняем её элементы к 0.

```
For i = 1 To M
For j = 1 To M
C(i, j) = 0
Next
Next
```

Определим векторы

```
Dim V(1 To M, 1) As Double
Dim VT(1, 1 To M) As Double
```

Найдем значение $\sum_{i=1}^{N} R_i R_i^T$ следующим образом:

```
For k = 1 To N
For j = 1 To M
V(j, 1) = Worksheets("2").Cells(k, j).Value
VT(1, j) = Worksheets("2").Cells(k, j).Value
Next
For i = 1 To M
For j = 1 To M
C(i, j) = C(i, j) + V(i, 1) * VT(1, j)
Next
Next
Next
```

Найдем значение $\frac{1}{N}\sum_{i=1}^{N} R_i R_i^T$ следующим образом:

```
For i = 1 To M
For j = 1 To M
C(i, j) = C(i, j) / N
Next
Next
```

Определим векторы:

```
Dim AV(1 To M, 1) As Double
Dim AVT(1, 1 To M) As Double
```

Найдём значение $\overline{R}$ следующим образом:

```
s = 0: j = 1
Do While j < M
```

```
For i = 1 To N
s = s + Worksheets("2").Cells(i, j).Value
Next
AV(j, 1) = s / N
AVT(1, j) = s / N
j = j + 1
Loop
```

Найдем значения матрицы $C = \frac{1}{N}\sum_{i=1}^{N} R_i R_i^T - \overline{R}\,\overline{R}^T$.

```
For i = 1 To M
For j = 1 To M
C(i, j) = C(i, j) - AV(i, 1) * AVT(1, j)
Next
Next
```

Запишем элементы матрицы C в ячейки К1:Т10.

```
For i = 1 To M
For j = 1 To M
Worksheets("2").Cells(i, j + M).Value = C(i, j)
Next
Next
```

Проанализируем полученный результат. Найдем максимальный и минимальный элементы ковариационной матрицы. Зададим переменные

value_max – максимальный элемент

ind_max_y – номер строки с максимальным элементом

ind_max_y – номер столбца с минимальным элементом

value_min – минимальный элемент

ind_min_x – номер строки с минимальным элементом

ind_min_y – номер столбца с минимальным элементом

```
value_max = Worksheets("2").Cells(1, M + 1).Value: ind_max_x = 1: ind_max_y = 1
value_min = Worksheets("2").Cells(1, M + 1).Value: ind_min_x = 1: ind_min_y = 1
For i = 1 To N
For j = M + 1 To 2 * M
If Worksheets("2").Cells(i, j).Value > value_max Then value_max = Worksheets("2").Cells(i, j).Value: ind_max_x = i: ind_max_y = j - M:
If Worksheets("2").Cells(i, j).Value < value_min Then value_min = Worksheets("2").Cells(i, j).Value: ind_min_x = i: ind_min_y = j - M:
Next
Next
Worksheets("2").Cells(M + 1, M + 4).Value = ind_max_y
Worksheets("2").Cells(M + 1, M + 3).Value = ind_max_x
Worksheets("2").Cells(M + 1, M + 2).Value = value_max
Worksheets("2").Cells(M + 1, M + 1).Value = "maximal value"
Worksheets("2").Cells(M + 2, M + 4).Value = ind_min_y
Worksheets("2").Cells(M + 2, M + 3).Value = ind_min_x
Worksheets("2").Cells(M + 2, M + 2).Value = value_min
```

Worksheets("2").Cells(M + 2, M + 1).Value = "minimal value"

Получили, что максимальный элемент ковариационной матрицы $c_{22} = 0.010464$, а минимальный элемент ковариационной матрицы $c_{26} = -0.00078$.

Если ковариация положительна, то с ростом значений одной случайной величины значения второй имеют тенденцию возрастать, а если знак отрицательный, то убывать. Однако только по абсолютному значению ковариации нельзя судить о том, насколько сильно величины взаимосвязаны, так как масштаб ковариации зависит от их дисперсий. Значение ковариации можно нормировать, поделив её на произведение среднеквадратических отклонений (квадратных корней из дисперсий) случайных величин. Полученная величина называется коэффициентом корреляции Пирсона, который всегда находится в интервале от -1 до 1: $r(X,Y) = \frac{\operatorname{cov}(X,Y)}{\sigma_X \sigma_Y}$. Случайные величины, имеющие нулевую ковариацию, называются некоррелированными.

Перечислим основные свойства корреляции.

1. $r(X,X) = 1$
2. $r(X,Y) = r(Y,X)$
3. $r(X,Y) = 0$ для независимых случайных величин X и Y
4. $r(aX + b, cY + d) = \operatorname{sgn}(ac) r(X,Y)$
5. $|r(X,Y)| \le 1$
6. $|r(X,Y)| = 1 \Leftrightarrow x = aY + b$

Коэффициент корреляции $r(X,Y)$ отражает степень линейной зависимости между двумя случайными величинами X и Y.

При $r(X,Y) > 0$ можно сделать вывод о том, что с ростом одной случайной величины вторая случайная величина имеет тенденцию к увеличению. Например, рост и вес человека связаны положительной корреляционной зависимостью.

При $r(X,Y) < 0$ можно сделать вывод о том, что с ростом одной случайной величины вторая случайная величина имеет тенденцию к уменьшению. Например, температура и время сохранности продуктов питания связаны отрицательной корреляционной зависимостью.

При $r(X,Y) = 0$ случайные величины X и Y называются некоррелированными. Отметим, что некоррелированность случайных величин не означает их статистическую независимость, это говорит лишь о том, что между ними нет линейной зависимости.

Из всего сказанного можно сделать вывод о том, что при описании двумерных случайных величин бывает недостаточно таких хорошо известных характеристик, как математическое ожидание, дисперсия и среднее квадратическое отклонение. Поэтому часто для их описания используются ещё две характеристики: ковариация и корреляция. А понятие «корреляция» вообще стало широко использоваться не только в науке, но и в повседневной жизни.

Лабораторная работа №3
«Автоковариационная и автокорреляционная функции.
Декомпозиция временного ряда, тренд,
сезонная и циклическая компоненты»

Задание. Даны значения временного ряда. Разделить их на два кластера, используя метод максимального правдоподобия.

Метод максимального правдоподобия в математической статистике – это метод оценивания неизвестного параметра путём максимизации функции правдоподобия. Основан на предположении о том, что вся информация о статистической выборки содержится в функции правдоподобия. Метод максимального правдоподобия был проанализирован Р. Фишером между 1912 и 1922 годами.

Предположим, что выборка значений возврата $R = \{\rho_i\}_{i=1}^{N}$ состоит из независимых и одинаково распределённых случайных величин с плотностью общего закона распределения $p(x) = p_1 p(x/q_1) + p_2 p(x/q_2)$. Здесь q_1 и q_2 – наборы параметров распределений, подлежащих оценке. Оценки по методу максимального правдоподобия являются решением оптимизационной задачи: $\max\limits_{p_1,p_2,q_1,q_2}\left[\sum\limits_{i=1}^{N}\ln\left(p_1 p(\rho_i/q_1) + p_2 p(\rho_i/q_2)\right)\right]$. Максимум ищется по параметрам законов распределений и вероятностям p_1 и p_2. Следующий алгоритм является монотонным и сходящимся к множеству стационарных точек, среди которых есть решение. Для нормальных законов алгоритм сходится к максимально правдоподобным оценкам.

Алгоритм.

1. Определяются вероятности p_1 и p_2, а также параметры распределений q_1 и q_2.
2. Вычисляются

$$\alpha_{i,1} = \frac{p(r_i/q_1)p_1}{p(r_i/q_1)p_1 + p(r_i/q_2)p_2}, \alpha_{i,2} = \frac{p(r_i/q_2)p_2}{p(r_i/q_1)p_1 + p(r_i/q_2)p_2},$$

$$p_1 = \frac{1}{N}\sum_{i=1}^{N}\alpha_{i,1}, p_2 = \frac{1}{N}\sum_{i=1}^{N}\alpha_{i,2},$$

$$q_1 = \arg\max\sum_{i=1}^{N}\alpha_{i,1}\ln p_1(\rho_i/q), q_2 = \arg\max\sum_{i=1}^{N}\alpha_{i,2}\ln p_2(\rho_i/q)$$

3. Если «критерий остановки», то стоп. Иначе 2.

В качестве критерия остановки можно взять малую различимость между предыдущим и последующим центром для каждого кластера.

Для нормальных законов параметры: $q_1 = \langle m_1, \sigma_1^2 \rangle, q_2 = \langle m_2, \sigma_2^2 \rangle$. Очередное приближение к максимально правдоподобным оценкам вычисляется следующим образом:

$$\alpha_{i,1}=\frac{\frac{1}{\sqrt{2\pi\sigma_1^2}}\exp\left(-\frac{(\rho_i-m_1)^2}{2\sigma_1^2}\right)p_1}{\frac{1}{\sqrt{2\pi\sigma_1^2}}\exp\left(-\frac{(\rho_i-m_1)^2}{2\sigma_1^2}\right)p_1+\frac{1}{\sqrt{2\pi\sigma_2^2}}\exp\left(-\frac{(\rho_i-m_2)^2}{2\sigma_2^2}\right)p_2},$$

$$\alpha_{i,2}=\frac{\frac{1}{\sqrt{2\pi\sigma_2^2}}\exp\left(-\frac{(\rho_i-m_2)^2}{2\sigma_2^2}\right)p_2}{\frac{1}{\sqrt{2\pi\sigma_1^2}}\exp\left(-\frac{(\rho_i-m_1)^2}{2\sigma_1^2}\right)p_1+\frac{1}{\sqrt{2\pi\sigma_2^2}}\exp\left(-\frac{(\rho_i-m_2)^2}{2\sigma_2^2}\right)p_2}$$

$$p_1=\frac{1}{N}\sum_{i=1}^{N}\alpha_{i,1},\ p_2=\frac{1}{N}\sum_{i=1}^{N}\alpha_{i,2},$$

$$m_1=\frac{1}{\sum_{i=1}^{N}\alpha_{i,1}}\sum_{i=1}^{N}\alpha_{i,1}\rho_i,\ m_2=\frac{1}{\sum_{i=1}^{N}\alpha_{i,2}}\sum_{i=1}^{N}\alpha_{i,2}\rho_i,$$

$$\sigma_1^2=\frac{1}{\sum_{i=1}^{N}\alpha_{i,1}}\sum_{i=1}^{N}\alpha_{i,1}\rho_i^2-m_1^2,\sigma_2^2=\frac{1}{\sum_{i=1}^{N}\alpha_{i,2}}\sum_{i=1}^{N}\alpha_{i,2}\rho_i^2-m_2^2$$

Для выбора начальных значений p_1, p_2, q_1, q_2 выборка разбивается на два подмножества медианой выборки: $R_1=\{\rho\in R,\rho\le \mathrm{med}(R)\}, R_2=R\setminus R_1$, после этого вычисляются $p_1=\frac{|R_1|}{|R|}, p_2=\frac{|R_2|}{|R|}$, $m_1=\frac{1}{|R_1|}\sum_{\rho\in R_1}\rho, m_2=\frac{1}{|R_2|}\sum_{\rho\in R_2}\rho$,

$$\sigma_1^2=\frac{1}{|R_1|}\sum_{\rho\in R_1}\rho^2-m_1^2,\sigma_2^2=\frac{1}{|R_2|}\sum_{\rho\in R_2}\rho^2-m_2^2.$$

Отметим, что под медианой набора чисел понимается число, которое находится в середине этого набора, если его упорядочить по возрастанию. То есть такое число, что половина из элементов набора не меньше его, а другая половина не больше. Например, медианой набора {11,9,3,5,5} является число 5, так как оно стоит в середине этого набора после его упорядочивания: {3,5,5,9,11}. Если в выборке чётное число элементов, медиана может быть не определена однозначно: тогда для числовых данных чаще всего используют полусумму двух соседних значений. То есть, медиана набора {1,3,5,7} равна 4. Также определяется медиана случайной величины: в этом случае она определяется как число, которое делит пополам распределение. Если распределение непрерывно, то медиана является одним из решений уравнения: $F(x)=0.5$, где F – функция распределения случайной величины x, связанная с плотностью распределения f как $F(x)=\int_{-\infty}^{y}f(y)dy$.

После остановки разбиение выборки R на два класса R_1 и R_2 осуществляется следующим образом. Элемент выборки ρ_i относится к R_1, если $p(\rho_i/q_1)p_1\ge p(\rho_i/q_2)p_2$. Иначе ρ_i относится к R_2.

Определим массивы

```
Dim R(1 To N) As Double
Dim R1(1 To N) As Double
Dim R2(1 To N) As Double
Dim I1(1 To N) As Integer
Dim I2(1 To N) As Integer
```

В массиве R содержится вся выборка, в массивах R1 и R2 записаны элементы первого и второго кластеров соответственно. Массивы I1 и I2 содержат индексы и являются вспомогательными переменными.

```
For i = 1 To N
R(i) = Worksheets("4").Cells(i, 2).Value
Next
R1(1) = Worksheets("4").Cells(1, 2).Value
R1(2) = Worksheets("4").Cells(2, 2).Value
For i = 3 To N
R2(i - 2) = Worksheets("3").Cells(i, 2).Value
Next
```

Сначала первый кластер содержит первый и второй элементы выборки, второй кластер – все остальные элементы выборки. Обозначим через p1, p2 вероятности принадлежности к каждому кластеру. Обозначим через n1, n2 выборочное среднее элементов в первом и втором кластере соответственно. Обозначим через sigma1, sigma2 выборочную дисперсию элементов в первом и втором кластере соответственно.

```
p1 = 2 / N: p2 = (N - 2) / N
n1 = (R1(1) + R1(2)) / 2
sigma1 = (R1(1) * R1(1) + R1(2) * R1(2)) / 2 - n1 * n1
s2 = 0: s3 = 0
For i = 3 To N
s2 = s2 + R2(i - 2): s3 = s3 + R2(i - 2) * R2(i - 2)
Next
n2 = s2 / (N - 2)
sigma2 = s3 / (N - 2) - n2 * n2
```

Введём переменные eps = 0.001: dif1 = 0.1: dif2 = 0.1.

Приведём реализацию метода максимального правдоподобия.

```
Do While dif1 > eps And dif2 > eps
k1 = n1: k2 = n2: s1 = 0: s2 = 0: s3 = 0: s4 = 0: s5 = 0: s6 = 0
For i = 1 To N
u = (1 / Sqr(2 * 3.14 * sigma1)) * Exp(-(R (i) - n1) * (R (i) - n1) / (2 * sigma1)) * p1
+ (1 / Sqr(2 * 3.14 * sigma2)) * Exp(-(R (i) - n2) * (R (i) - n2) / (2 * sigma2)) * p2
u1 = (1 / Sqr(2 * 3.14 * sigma1)) * Exp(-(R (i) - n1) * (R (i) - n1) / (2 * sigma1)) * p1
u2 = (1 / Sqr(2 * 3.14 * sigma2)) * Exp(-(R (i) - n2) * (R (i) - n2) / (2 * sigma2)) * p2
alpha1 = u1 / u: alpha2 = u2 / u
s1 = s1 + alpha1: s2 = s2 + alpha2
s3 = s3 + R (i) * alpha1: s4 = s4 + R (i) * alpha2
s5 = s5 + R (i) * R (i) * alpha1: s6 = s6 + R (i) * R (i) * alpha2
```

```
Next
p1 = s1 / N: p2 = s2 / N: n1 = s3 / s1: n2 = s4 / s2: sigma1 = s5 / s1 - n1 * n1: sigma2 = s6 / s2 - n2 * n2
dif1 = Abs(n1 - k1): dif2 = Abs(n2 - k2)
Loop
```

Отметим, что в вышеприведённой процедуре используются следующие функции языка VBA Excel: Abs, Sqr, Exp (абсолютное значение, квадратный корень, экспонента)

Разделим выборку на кластеры.

```
m1 = 0: m2 = 0:
For i = 1 To N
u1 = (1 / Sqr(2 * 3.14 * sigma1)) * Exp(-(R(i) - n1) * (R(i) - n1) / (2 * sigma1)) * p1
u2 = (1 / Sqr(2 * 3.14 * sigma2)) * Exp(-(R(i) - n2) * (R(i) - n2) / (2 * sigma2)) * p2
If (u1 >= u2) Then m1 = m1 + 1: R1(m1) = R(i): I1(m1) = i:
If (u1 < u2) Then m2 = m2 + 1: R2(m2) = R(i): I2(m2) = i:
Next
```

Получили, что m1=147, m2=151. Выведем значения первого кластера в ячейки B1:B147, второго кластера в ячейки B148:B298. Выведем значения номеров элементов в ячейки A1:A147 и A148:A298 соответственно. Выведем в ячейки C1:C147 число 0 (принадлежность первому кластеру), в ячейки C148:C298 число 1 (принадлежность второму кластеру).

Для графического отображения информации введём динамические массивы

```
Dim Rml1() As Double
Dim Rml2() As Double
Dim Iml1() As Integer
Dim Iml2() As Integer
```

Определим их размеры

```
ReDim Iml1(m1 - 1): ReDim Iml2(m2 - 1): ReDim Rml1(m1 - 1): ReDim Rml2(m2 - 1)
```

Определим их значения

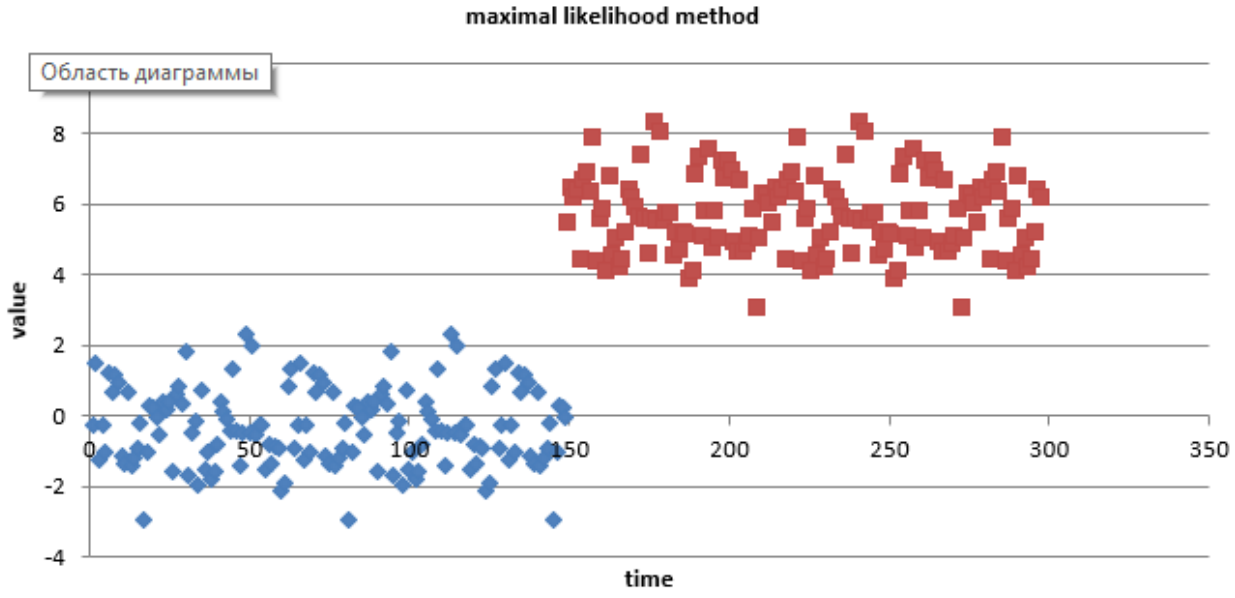
```
For i = 1 To m1
Worksheets("4").Cells(i, 1).Value = I1(i)
Worksheets("4").Cells(i, 2).Value = R1(i)
Worksheets("4").Cells(i, 3).Value = 0
Iml1(i - 1) = I1(i): Rml1(i - 1) = R1(i):
Next
For i = 1 To m2
Worksheets("4").Cells(i + m1, 1).Value = I2(i)
Worksheets("4").Cells(i + m1, 2).Value = R2(i)
Worksheets("4").Cells(i + m1, 3).Value = 1
Iml2(i - 1) = I2(i): Rml2(i - 1) = R2(i):
Next
```

Построим точечные диаграммы для каждого кластера.

```
If Worksheets("4").ChartObjects.Count Then Worksheets("4").ChartObjects.Delete
```

```
With Worksheets("4").ChartObjects.Add(500, 250, 500, 250)
With .Chart
.ChartType = xlXYScatter
.HasLegend = False
.SeriesCollection.Add Source:=Range("A1:A2")
.SeriesCollection(1).XValues = Iml1
.SeriesCollection(1).Values = Rml1
.SeriesCollection.Add Source:=Range("A1:A2")
.SeriesCollection(2).XValues = Iml2
.SeriesCollection(2).Values = Rml2
.HasTitle = True
.ChartTitle.Text = "maximal likelihood method"
.ChartTitle.Font.Size = 10
.Axes(xlCategory, xlPrimary).HasTitle = True
.Axes(xlCategory, xlPrimary).AxisTitle.Text = "time"
.Axes(xlCategory, xlPrimary).AxisTitle.Font.Size = 10
.Axes(xlCategory, xlPrimary).TickLabels.Font.Size = 10
.Axes(xlValue, xlPrimary).HasTitle = True
.Axes(xlValue, xlPrimary).AxisTitle.Text = "value"
.Axes(xlValue, xlPrimary).AxisTitle.Font.Size = 10
.Axes(xlValue, xlPrimary).TickLabels.Font.Size = 10
End With
End With
```

Заметим, что элементы первого кластера покрашены в синий цвет, а элементы второго кластера – в красный цвет. Отметим, что центр одного из кластеров находится в точке 0, а центр другого из кластеров находится в точке 6. Также стоит отметить, что данный алгоритм интересно было бы применить сначала к тестовой выборке, состоящей из сгенерированных нормальных случайных величин с заданными выборочными средними значениями и дисперсиями; а затем применить к выборке, состоящей из реальных данных.



Лабораторная работа №4

«Элементы кластерного анализа. Метод k-средних»

Задание. Даны значения временного ряда. Разделить их на два кластера, используя метод k-средних.

Кластерный анализ – многомерная статистическая процедура, выполняющая сбор данных, содержащих информацию о выборке объектов, и затем упорядочивающая объекты в сравнительно однородные группы. Задача кластеризации относится к статистической обработке, а также к широкому классу задач обучения без учителя.

Слово «кластер» произошло от английского слова «cluster», что означает гроздь, сгусток, пучок. Большинство исследователей склоняются к тому, что впервые термин «кластерный анализ» был предложен психологом Р. Трионом. Впоследствии возник ряд терминов, которые в настоящее время принято считать синонимами термина «кластерный анализ». К примеру, термин «автоматическая классификация».

Спектр применений кластерного анализа очень широк. Его используют в геологии, социологии, маркетинге, антропологии, филологии, государственном управлении, биологии, химии, психологии, медицине, археологии и других дисциплинах. Однако универсальность применения привела к появлению большого количества несовместимых терминов, методов и подходов, затрудняющих однозначное использование и непротиворечивую интерпретацию кластерного анализа.

Кластерный анализ выполняет следующие основные задачи:

1. Разработка классификации или типологии.
2. Исследование полезных концептуальных схем группирования объектов.
3. Порождение гипотез на основе исследования данных.
4. Проверка гипотез.

Кластерный анализ предполагает следующие основные этапы:

1. Отбор выборки для кластеризации. Предполагается, что подвергаться кластеризации будут лишь количественные данные.
2. Определение признакового пространства, то есть множества переменных, по которым будут оцениваться объекты в выборке.
3. Вычисление значений меры сходства (различия) между объектами.
4. Применение метода кластерного анализа для создания групп сходных объектов.
5. Проверка достоверности результатов кластерного решения.

Заметим, что данные, используемые в кластерном анализе, должны быть однородными. Однородность требует, чтобы все данные были одной природы, описывались сходным набором характеристик.

Цели кластеризации:

1. Понимание данных путём выявления кластерной структуры. Разбиение выборки на группы схожих объектов позволяет упростить дальнейшую обработку данных и принятия решений, применяя к каждому кластеру свой метод анализа.

2. Сжатие данных. Если исходная выборка достаточно большая, то можно сократить её, остаив по одному наиболее типичному представителю от каждого кластера.
3. Обнаружение новизны. Выделяются нетипичные объекты, которые не удаётся присоединить ни к одному из кластеров.

Выделяют следующие подходы к кластеризации данных:

1. Вероятностный подход. К нему относится метод k-средних.
2. Подход на основе искусственного интеллекта.
3. Логический подход.
4. Теоретико-графовый подход.
5. Иерархический подход.

Метод k-средних является наиболее популярным методом кластеризации. Был изобретён в 1950-х годах математиком Гуго Штейнгаузом и почти одновременно Стюартом Ллойдом. Особую популярность приобрёл после работы Маккуина.

Основным преимуществом метода k-средних являются его скорость, простота реализации, универсальность и сходимость за конечное число шагов. К числу недостатков можно отнести то, что число кластеров нужно знать заранее.

Метод k-средних, наряду с методом нечётких k-средних и методом иерархической кластеризации, относится к непараметрическим методам. В основе метода k-средних лежит оптимизационная задача:

$$\min_{m,x} \sum_{i=1}^{N} (\rho_i - m_1)^2 x_i + \sum_{i=1}^{N} (\rho_i - m_2)^2 (1 - x_i),\ x_i \in \{0,1\}.$$

Применение координатного спуска к этой задаче – основной способ кластеризации методом k-средних. Хорошо известно, что, являясь монотонным, метод не гарантирует сходимости к решению оптимизационной задачи. Существенно упрощающей особенностью рассматриваемой задачи является её одномерность. Одномерность данных позволяет перед кластеризацией упорядочить данные в порядке возрастания. При условии, что число кластеров равно двум, можно предложить следующий алгоритм.

Суть метода k-средних заключается в следующем. Сначала выборка $R = \{\rho_1,...,\rho_N\}$ разбивается на 2 кластера R_1 и R_2 произвольным образом. Например, $R_1 = \{\rho_1\}, R_2 = \{\rho_2,...,\rho_N\}$. Затем рассчитываются центры кластеров: $m_1 = \frac{1}{|R_1|}\sum_{\rho_i \in R_1} \rho_i, m_2 = \frac{1}{|R_2|}\sum_{\rho_i \in R_2} \rho_i$, и полагаем $R_1 = \{\ \}, R_2 = \{\ \}$. После этого в цикле по количеству элементов выборки для каждого элемента выборки подсчитывается расстояние от него до центра каждого кластера: $d_1 = |\rho_i - m_1|, d_2 = |\rho_i - m_2|$. Заполнение кластеров происходит по следующей схеме: $\rho_i \in R_1$, если $d_1 < d_2$ и $\rho_i \in R_2$, если $d_1 \geq d_2$. Процесс необходимо повторять до тех пор, пока центры m_1 и m_2 начинают мало отличаться друг от друга.

Определим массивы

```
Dim R(1 To N) As Double
Dim R1(1 To N) As Double
Dim R2(1 To N) As Double
Dim I1(1 To N) As Integer
Dim I2(1 To N) As Integer
```

В массиве R содержится вся выборка, в массивах R1 и R2 записаны элементы первого и второго кластеров соответственно. Массивы I1 и I2 содержат индексы и являются вспомогательными переменными.

```
For i = 1 To N
R(i) = Worksheets("3").Cells(i, 2).Value
Next
R1(1) = Worksheets("3").Cells(1, 2).Value
For i = 2 To N
R2(i - 1) = Worksheets("3").Cells(i, 2).Value
Next
```

Сначала первый кластер содержит первый элемент выборки, второй кластер – все остальные элементы выборки. Обозначим через m1, m2 количество элементов в первом и втором кластере соответственно. Обозначим через s1, s2 сумму элементов в первом и втором кластере соответственно. Обозначим через n1, n2 среднее значение элементов в первом и втором кластере соответственно.

```
m1 = 1: s1 = R1(1)
n1 = s1 / m1
m2 = N – 1: s2 = 0
For i = 2 To N
s2 = s2 + R2(i-1)
Next
n2 = s2 / m2
```

Введём переменные eps = 0.001: dif1 = 0.1: dif2 = 0.1.

Приведём реализацию метода k-средних.

```
Do While dif1 > eps And dif2 > eps
k1 = n1: k2 = n2: m1 = 0: m2 = 0: s1 = 0: s2 = 0
For i = 1 To N
d1 = Abs(R(i) - n1): d2 = Abs(R(i) - n2)
If d1 < d2 Then m1 = m1 + 1: R1(m1) = R(i): I1(m1) = i: s1 = s1 + R(i)
If d1 >= d2 Then m2 = m2 + 1: R2(m2) = R(i): I2(m2) = i: s2 = s2 + R(i)
Next
n1 = s1 / m1: n2 = s2 / m2: dif1 = Abs(n1 - k1): dif2 = Abs(n2 - k2)
Loop
```

Получили, что m1=114, m2=184. Выведем значения первого кластера в ячейки B1:B114, второго кластера в ячейки B115:B298. Выведем значения номеров элементов в ячейки A1:A114 и A115:A298 соответственно. Выведем в ячейки C1:C114 число 0 (принадлежность первому кластеру), в ячейки C115:C298 число 1 (принадлежность второму кластеру).

Для графического отображения информации введём динамические

```
массивы
Dim Rkm1() As Double
Dim Rkm2() As Double
Dim Ikm1() As Integer
Dim Ikm2() As Integer
        Определим их размеры
ReDim Ikm1(m1 - 1): ReDim Ikm2(m2 - 1): ReDim Rkm1(m1 - 1): ReDim Rkm2(m2 - 1)
        Определим их значения
For i = 1 To m1
Worksheets("3").Cells(i, 1).Value = I1(i)
Worksheets("3").Cells(i, 2).Value = R1(i)
Worksheets("3").Cells(i, 3).Value = 0
Ikm1(i - 1) = I1(i): Rkm1(i - 1) = R1(i):
Next
For i = 1 To m2
Worksheets("3").Cells(i + m1, 1).Value = I2(i)
Worksheets("3").Cells(i + m1, 2).Value = R2(i)
Worksheets("3").Cells(i + m1, 3).Value = 1
Ikm2(i - 1) = I2(i): Rkm2(i - 1) = R2(i):
Next
        Построим точечные диаграммы для каждого кластера.
If Worksheets("3").ChartObjects.Count Then Worksheets("3").ChartObjects.Delete
With Worksheets("3").ChartObjects.Add(500, 250, 500, 250)
With .Chart
.ChartType = xlXYScatter
.HasLegend = False
.SeriesCollection.Add Source:=Range("A1:A2")
.SeriesCollection(1).XValues = Ikm1
.SeriesCollection(1).Values = Rkm1
.SeriesCollection.Add Source:=Range("A1:A2")
.SeriesCollection(2).XValues = Ikm2
.SeriesCollection(2).Values = Rkm2
.HasTitle = True
.ChartTitle.Text = "k-means method"
.ChartTitle.Font.Size = 10
.Axes(xlCategory, xlPrimary).HasTitle = True
.Axes(xlCategory, xlPrimary).AxisTitle.Text = "time"
.Axes(xlCategory, xlPrimary).AxisTitle.Font.Size = 10
.Axes(xlCategory, xlPrimary).TickLabels.Font.Size = 10
.Axes(xlValue, xlPrimary).HasTitle = True
.Axes(xlValue, xlPrimary).AxisTitle.Text = "value"
.Axes(xlValue, xlPrimary).AxisTitle.Font.Size = 10
.Axes(xlValue, xlPrimary).TickLabels.Font.Size = 10
End With
```

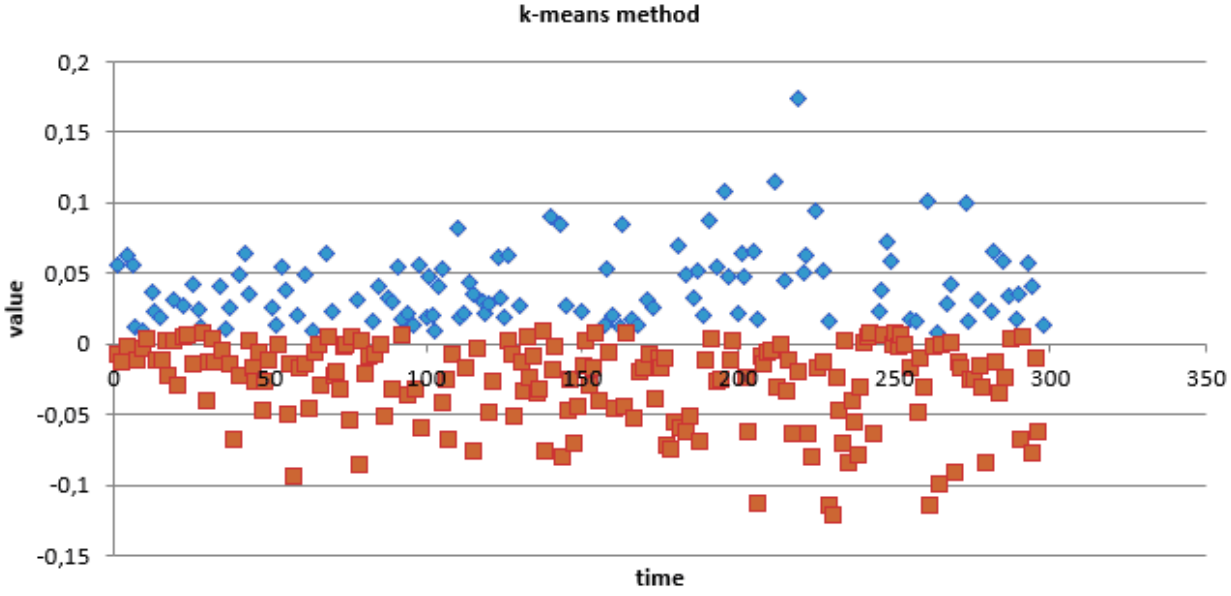
End With

Заметим, что элементы первого кластера покрашены в синий цвет, а элементы второго кластера – в красный цвет. Также отметим, что перед добавлением нового графика происходит удаление старого графика. При этом перед удалением графика происходит проверка того, что график есть на листе. В процедуре построения диаграммы задаётся её тип xlXYScatter, что означает точечная диаграмма. Для осей абсцисс xlCategory и ординат xlValue можно задать подписи осей и размер шрифта. Также можно задать название диаграммы. На представленном рисунке на оси абсцисс отмечено время, на оси ординат – значение элемента выборки в данный момент времени. Видно, что в одном кластере присутствуют в основном положительные элементы, а в другом кластере – отрицательные. Количество элементов в обоих кластеров получилось почти одинаковым.

Интересной является задача нахождения эмпирической функции распределения в каждом кластере и построения её графика. Функция распределения возрастает и принимает значения на отрезке $[0,1]$. Для построения функции распределения в каждом кластере нужно сначала упорядочить элементы кластера по неубыванию. По оси абсцисс находятся значения элементов кластера. По оси ординат находятся относительные частоты попадания случайной величины в заданный интервал.

В качестве продолжения исследования можно рассмотреть задачу выделения интервала минимальной длины в каждом кластере с заданной доверительной вероятностью попадания в этот интервал. Очевидно, что с увеличением доверительной вероятности длина интервала будет уменьшаться. Данная задача является очень актуальной в стохастической финансовой математике, так как позволяет обобщить известные модели с помощью замены точечных параметров на интервалы. Поскольку при такой замене возникает неопределённость, будем использовать методы робастной оптимизации.

Одна из задач робастной оптимизации –это задача расчёта интервала справедливых цен для обобщённой модели Кокса-Росса-Рубинштейна.



Лабораторная работа №5

«Линейные стохастические модели. Модель скользящего среднего MA, авторегрессионная модель AR»

Задание. Построить график компьютерной реализации последовательности $h=(h_n)$, подчиняющейся $MA(1)$-модели с $h_n=\mu+b_1\varepsilon_{n-1}+b_0\varepsilon_n$ с параметрами $\mu=1, b_1=1, b_0=0.1$ и $\varepsilon_n \sim N(0,1)$.

Модель скользящего среднего является распространённым подходом для моделирования одномерных временных рядов. Согласно модели, оценка прогнозируемых членов ряда линейно зависит от текущего и прошлых значений, а также некоторого стохастического члена, который отражает вероятностный характер модели.

Модель скользящего среднего может применяться в разных областях, в том числе, в стохастической финансовой математике, когда необходимо описать поведение процесса с помощью некоторой модели. С этой целью на рынке ценных бумаг инвесторы проявляют большую заинтересованность в том, чтобы ознакомиться с разными классами стохастических процессов, которые могут быть применены при построении моделей динамики финансовых показателей (цен, индексов, обменных курсов…) и при проведении различных расчётов (рисков, хеджирующих стратегий, рациональных цен опционов).

Приведём определение и свойства модели $MA(q)$, хотя в работе необходимо построить график компьютерной реализации модели $MA(1)$. Заметим, что процесс белого шума можно формально считать процессом скользящего среднего нулевого порядка $MA(0)$.

В модели скользящего среднего $MA(q)$, описывающей эволюции последовательности $h=(h_n)$, предполагается следующий способ формирования значений h_n по белому шуму в широком смысле: $\varepsilon=(\varepsilon_n)$:

$h_n=(\mu+b_1\varepsilon_{n-1}+\ldots+b_q\varepsilon_{n-q})+b_0\varepsilon_n$, где параметр q определяет порядок зависимости от прошлого, а $F_n=\sigma(\varepsilon_n,\varepsilon_{n-1},\ldots)$.

Введём оператор сдвига назад L, действующий на числовых последовательностях $x=(x_n)$ по формуле: $Lx_n=x_{n-1}$. Так как $L(Lx_n)=Lx_{n-1}=x_{n-2}$, то $L^2x_n=x_{n-2}$ и $L^kx_n=x_{n-k}, k\geq 0$. В литературе оператор L иногда называют лаговым оператором.

Пусть c, c_1, c_2 константы. Свойства оператора L:

1. $L(cx_n)=cLx_n$
2. $L(x_n+y_n)=Lx_n+Ly_n$
3. $(c_1L+c_2L^2)x_n=c_1Lx_n+c_2L^2x_n=c_1x_{n-1}+c_2x_{n-2}$
4. $(1-\lambda_1L)(1-\lambda_2L)x_n=x_n-(\lambda_1+\lambda_2)x_{n-1}+\lambda_1\lambda_2x_{n-2}$

Из 1 и 2 следует, что оператор L обладает свойством линейности.

Перепишем выражение для h_n с помощью оператора L:

$h_n=\mu+\beta(L)\varepsilon_n$, где $\beta(L)=b_0+b_1L+\ldots+b_qL^q$.

Рассмотрим вероятностные характеристики последовательности $h=(h_n)$.

Пусть $q=1$. Тогда $h_n=\mu+b_0\varepsilon_n+b_1\varepsilon_{n-1}$. Имеем:

$$Eh_n=\mu, Dh_n=b_0^2+b_1^2, \operatorname{cov}(h_n,h_{n+1})=b_0b_1, \operatorname{cov}(h_n,h_{n+k})=0, k>1$$

Последние два свойства означают, что $h=(h_n)$ является последовательностью с коррелированными соседними значениями h_n и h_{n+1}, в то время как корреляция значений h_n и h_{n+k} при $k\geq 2$ равна 0. Следовательно, ковариационная матрица будет трёхдиагональной. Интерес представляет изучение её свойств, кроме свойства положительной определённости и симметричности. К примеру, вычислить определитель или найти обратную матрицу.

Заметим, что если $b_0b_1>0$ (то есть b_0 и b_1 одного знака), то величины h_n и h_{n+1} положительно коррелированны. То есть с ростом h_n растёт и h_{n+1}, и наоборот. Если $b_0b_1<0$ (то есть b_0 и b_1 имеют разные знаки), то величины h_n и h_{n+1} отрицательно коррелированны. То есть с ростом h_n убывает h_{n+1}, и наоборот. Также отметим, что у элементов последовательности $h=(h_n)$ среднее значение, дисперсия и ковариация не зависят от n. Это в предположении о том, что $\varepsilon=(\varepsilon_n)$ является белым шумом в широком смысле с $E\varepsilon_n=0, D\varepsilon_n^2=1$ и коэффициенты в выражении для h_n не зависят от n. То есть, последовательность $h=(h_n)$ является стационарной в широком смысле. Если к тому же предположить, что последовательность $\varepsilon=(\varepsilon_n)$ является гауссовской, что последовательность $h=(h_n)$ тоже будет гауссовской. Это означает, что все её вероятностные свойства выражаются лишь в терминах среднего, дисперсии и ковариации. В этом случае последовательность $h=(h_n)$ является стационарной в узком смысле, то есть $\operatorname{Law}(h_{i_1},...,h_{i_n})=\operatorname{Law}(h_{i_1+k},...,h_{i_n+k}), n\geq 1$ и произвольных k.

Пусть $(h_1,...,h_n)$ – некоторая реализация, полученная в результате наблюдений величин h_k в моменты $k=1,...,n$ и $\bar{h}_n=\frac{1}{n}\sum_{k=1}^{n}h_k$ – временное среднее, позволяющее оценить среднее значение μ. Будем измерять качество этой оценки величиной среднеквадратического отклонения $\Delta_n^2=E\left|\bar{h}_n-\mu\right|^2$. Справедлива следующая

Теорема. $\Delta_n^2\to 0\Leftrightarrow\frac{1}{n}\sum_{k=1}^{n}R(k)\to 0$ при $n\to\infty$, где $R(k)=\operatorname{cov}(h_n,h_{n+k})$ и $h=(h_n)$ – произвольная стационарная в широком смысле последовательность.

Доказательство. Пусть $\mu=Eh_n=0$. Тогда

$\left|\frac{1}{n}\sum_{k=1}^{n}R(k)\right|^2=\left|E\left(\frac{1}{n}\sum_{k=1}^{n}h_k\right)h_0\right|^2\leq Eh_0^2E\left|\frac{1}{n}\sum_{k=1}^{n}h_k\right|^2$. Необходимость доказана.

С другой стороны,

$$E\left|\frac{1}{n}\sum_{k=1}^{n}h_k\right|^2=\frac{1}{n^2}E\left(\sum_{k=1}^{n}h_k^2+2\sum_{k=1,k\neq l}^{n}h_kh_l\right)=\frac{2}{n^2}\sum_{l=1}^{n}\sum_{k=0}^{l-1}R(k)-\frac{1}{n}R(0).$$

Выберем $\delta > 0$ и пусть $n(\delta)$ таково, что для всех $l \ge n(\delta)$ $\left|\frac{1}{l}\sum_{k=0}^{l-1} R(k)\right| \le \delta$.

Тогда для $n \ge n(\delta)$

$$\left|\frac{1}{n^2}\sum_{l=1}^{n}\sum_{k=0}^{l-1} R(k)\right| = \left|\frac{1}{n^2}\sum_{l=1}^{n(\delta)}\sum_{k=0}^{l-1} R(k) + \frac{1}{n^2}\sum_{l=n(\delta)+1}^{n}\sum_{k=0}^{l-1} R(k)\right| \le \frac{1}{n^2}\left|\sum_{l=1}^{n(\delta)}\sum_{k=0}^{l-1} R(k)\right| + \frac{1}{n^2}\left|\sum_{l=n(\delta)+1}^{n} l\cdot\frac{1}{l}\sum_{k=0}^{l-1} R(k)\right| \le$$

$$\le \frac{1}{n^2}\left|\sum_{l=1}^{n(\delta)}\sum_{k=0}^{l-1} R(k)\right| + \delta. \text{ Так как } n(\delta) < \infty, |R(0)| \le \text{const}, \text{ то } \overline{\lim_n} E\left|\frac{1}{n}\sum_{k=1}^{n} h_k\right|^2 \le \delta.$$

Учитывая то, что δ произвольное, достаточность доказана.

Таким образом, рассматриваемая модель $MA(1)$ является эргодической в том смысле, что $\bar{h}_n$ сходятся к μ в L^2.

Напомним, что по ковариации $\operatorname{cov}(h_n, h_m)$ определяется корреляция $\operatorname{corr}(h_n, h_m)$ формулой: $\operatorname{corr}(h_n, h_m) = \frac{\operatorname{cov}(h_n, h_m)}{\sqrt{Dh_n Dh_m}}$. Из неравенства Коши-Буняковского следует, что $|\operatorname{corr}(h_n, h_m)| \le 1$. В стационарном случае $\operatorname{cov}(h_n, h_{n+k})$ не зависит от n. Обозначим $R(k) = \operatorname{cov}(h_n, h_{n+k})$, $\rho(k) = \operatorname{corr}(h_n, h_{n+k})$, тогда $\rho(k) = \frac{R(k)}{R(0)}$. В случае модели $MA(1)$ имеем: $\rho(k) = \begin{cases} 1, k = 0 \\ 0, k > 1 \\ \frac{b_0 b_1}{b_0^2 + b_1^2}, k = 1 \end{cases}$. Если обозначить $\theta_1 = \frac{b_1}{b_0}$, то получим, что $\rho(1) = \frac{\theta_1}{1+\theta_1^2} = \frac{(1/\theta_1)}{1+(1/\theta_1)^2}$. Следовательно, разные значения θ_1 и $1/\theta_1$ приводят к одному и тому же результату $\rho(1)$.

Заметим, что если процесс стоимости рискового актива записать в виде $S_n = \exp(h_n)$ и $F_n = \sigma(\varepsilon_1, ..., \varepsilon_n)$, то некоторые условные математические ожидания можно вычислить следующим образом:

$$E(S_n / F_{n-1}) = \exp(\mu + b_1\varepsilon_{n-1})\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}\exp\left(b_0 x - \frac{x^2}{2}\right)dx = \exp\left(\mu + b_1\varepsilon_{n-1} + \frac{b_0^2}{2}\right),$$

$$E(S_n / F_n) = \exp(\mu + b_1\varepsilon_{n-1} + b_0\varepsilon_n) = S_n,$$

$$E(S_n / F_{n-2}) = \exp(\mu)\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}\exp\left(b_1 x - \frac{x^2}{2}\right)dx\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}\exp\left(b_1 x - \frac{x^2}{2}\right)dx =$$

$$= \exp\left(\mu + \frac{b_1^2 + b_0^2}{2}\right).$$

В то же время

$E(h_n / F_{n-1}) = \mu + b_1\varepsilon_{n-1}$, $E(h_n / F_n) = \mu + b_1\varepsilon_{n-1} + b_0\varepsilon_n = h_n$, $E(h_n / F_{n-2}) = \mu$.

Определим константу

```
Const L = 1000
Определим массивы
Dim MA(1 To L) As Double
Dim hMA() As Double
Dim iMA() As Integer
```

Положим i=2:mu = 1: b0 = 0.1: b1 = 1.

Применим метод Бокса-Мюллера для генерации стандартной нормальной случайной величины. В некоторых языках программирования есть встроенный генератор нормальных случайных величин, но в VBA Excel такой возможности нет.

```
Randomize
x1 = Rnd: x2 = Rnd
epsilon = Sqr(-2 * Log(x1)) * Cos(2 * myPi * x2)
MA(1) = mu + b0 * epsilon:
Do While i < L
Randomize
x1 = Rnd: x2 = Rnd
k=epsilon
epsilon = Sqr(-2 * Log(x1)) * Cos(2 * myPi * x2)
MA(i) = mu + b1 * k + b0 * epsilon:
i = i + 1
Loop
```

Выведем значения массива MA в столбец 1 листа.

Определим размер динамических массивов iMA и hMA равным L/5-1. Чтобы не все точки выводить на графике. Заполним динамические массивы iMA и hMA.

```
k = 0
ReDim iMA(L / 5-1): ReDim hMA(L / 5-1)
For i = 1 To L
Worksheets("6").Cells(i, 2).Value = MA(i)
If i Mod 5 = 0 Then iMA(k) = i: hMA(k) = MA(i): k = k + 1
Next
```

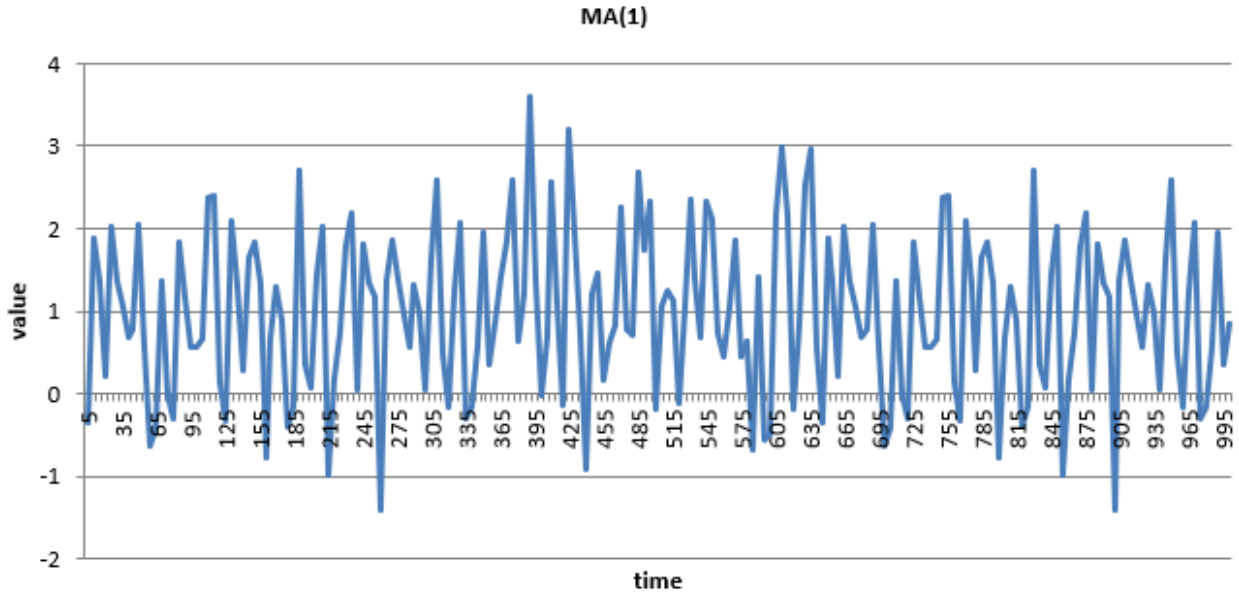
Построим график зависимости h от Ind.

```
If Worksheets("6").ChartObjects.Count Then Worksheets("6").ChartObjects.Delete
With Worksheets("6").ChartObjects.Add(500, 250, 500, 250)
With .Chart
.ChartType = xlLine
.HasLegend = False
.SeriesCollection.Add Source:=Range("A1:A2")
.SeriesCollection(1).XValues = iMA
.SeriesCollection(1).Values = hMA
.HasTitle = True
.ChartTitle.Text = "MA(1)"
.ChartTitle.Font.Size = 10
.Axes(xlCategory, xlPrimary).HasTitle = True
```

```
.Axes(xlCategory, xlPrimary).AxisTitle.Text = "time"
.Axes(xlCategory, xlPrimary).AxisTitle.Font.Size = 10
.Axes(xlCategory, xlPrimary).TickLabels.Font.Size = 10
.Axes(xlValue, xlPrimary).HasTitle = True
.Axes(xlValue, xlPrimary).AxisTitle.Text = "value"
.Axes(xlValue, xlPrimary).AxisTitle.Font.Size = 10
.Axes(xlValue, xlPrimary).TickLabels.Font.Size = 10
End With
End With
```

По графику видно, что значения случайного процесса колеблются около единицы, то есть это среднее значение. Дисперсия случайного процесса равна 1.01. То есть, значения процесса должны попадать а полосу $[1-3\cdot 1.005,1+3\cdot 1.005]$, то есть в интервал $[-2.015,4.015]$, что видно на графике. Данную модель необходимо применять в тех случаях, когда значения случайного процесса «скользят» возле среднего значения. Интерес представляет смена параметров модели μ, b_0, b_1 и анализ поведения графика процесса скользящего среднего первого порядка $MA(1)$. Также было бы интересно построить графики компьютерной реализации процессов скользящего среднего более высоких порядков. В общем случае порядок модели должен быть входным параметром и определяться пользователем. В теоретической части будет рассмотрен вопрос оценки параметров модели.

На графике присутствуют не все $L=1000$ значений последовательности h_n, а только 200 точек, что придаёт рисунку большую наглядность. Особый интерес представляет изучение свойств процесса $MA(\infty)$. В качестве дополнительного задания можно предложить на одном рисунке изобразить разными цветами графики модели скользящего среднего с разными параметрами и разных порядков. В заключение отметим, что большой интерес представляет модель $ARMA$, являющаяся смесью модели скользящего среднего MA и авторегрессионной модели AR. Также будет рассмотрена интегральная модель $ARMA$, а именно, модель $ARIMA$.



Лабораторная работа №6

«Модель авторегрессии и скользящего среднего ARMA и интегральная модель ARIMA»

Задание. Построить график компьютерной реализации последовательности $h=(h_n)$, подчиняющейся $ARMA(1,1)$-модели с $h_n = a_0 + a_1 h_{n-1} + b_1 \varepsilon_{n-1} + \sigma\varepsilon_n$ с параметрами $a_0 = -1, a_1 = 0.5, b_1 = 0.1, \sigma = 0.1, h_0 = 0$ и $0 \le n \le 1000$.

Построение вероятностно-статистических моделей по прошлым данным является необходимым для того, чтобы дать прогноз будущего движения цен. Безошибочный прогноз можно дать в очень редких случаях. К примеру, если мы имеем дело с сингулярными стационарными последовательностями. В общем случае при прогнозировании совершается ошибка, величина которой определяет степень риска решений, основанных на полученном прогнозе. В линейных стационарных моделях имеется теория построения оптимальных в среднеквадратическом смысле линейных оценок. В основном эта теория развита в работах А.Н.Колмогорова и Н.Винера. Предположим, что задано фильтрованное вероятностное пространство (Ω, F, F_n, P), $F_n = \sigma(...,\varepsilon_{-1},\varepsilon_0,\varepsilon_1,...,\varepsilon_n)$ с белым шумом в широком смысле $\varepsilon = (\varepsilon_n)$. Напомним, что под сигма-алгеброй F_n понимается информация, доступная к моменту времени n. По определению, последовательность $h=(h_n)$ является $ARMA$-моделью, если $h_n = \mu_n + \sigma\varepsilon_n$, где $\mu_n = (a_0 + a_1 h_{n-1} + ... + a_p h_{n-p}) + (b_1\varepsilon_{n-1} + ... + b_q\varepsilon_{n-q})$. Без ограничения общности можно считать $\sigma = 1$. Тогда $h_n - (a_1 h_{n-1} + ... + a_p h_{n-p}) = a_0 + (\varepsilon_n + b_1\varepsilon_{n-1} + ... + b_q\varepsilon_{n-q})$, или $\alpha(L)h_n = a_0 + \beta(L)\varepsilon_n$, где $\alpha(L) = 1 - a_1 L - ... - a_p L^p, \beta(L) = 1 + b_1 L + ... + b_q L^q$. Если $q = 0$, то $\alpha(L)h_n = w_n, w_n = a_0 + \varepsilon_n$, то есть получаем $AR(p)$-модель. Если $p = 0$, то $h_n = a_0 + \beta(L)\varepsilon_n$, то есть получаем $MA(q)$-модель.

Модель $ARMA$ может интерпретироваться как линейная модель множественной регрессии, в которой в качестве объясняющих переменных выступают прошлые значения самой зависимой переменной, а в качестве регрессионного остатка – скользящие средние из элементов белого шума. $ARMA$-процессы имеют более сложную структуру по сравнению со схожими по поведению AR или MA процессами, но при этом $ARMA$-процессы характеризуются меньшим количеством параметров, что является одним из их преимуществ.

Пусть $a_1 + ... + a_p \ne 1$. Тогда $h_n = \mu + \frac{\beta(L)}{\alpha(L)}\varepsilon_n$, где $\mu = \frac{a_0}{1 - (a_1 + ... + a_p)}$. Для стационарного решения $Eh_n = \frac{a_0}{1 - (a_1 + ... + a_p)}$.

Ковариация $R(k) = \operatorname{cov}(h_n, h_{n+k})$ для $k > q$ удовлетворяет соотношениям: $R(k) = a_1 R(k-1) + ... + a_p R(k-p)$.

В случае $k < q$ выражение для $R(k)$ выглядит более сложно, так как надо

учитывать также корреляционную зависимость между h_{n-k} и ε_{n-k}.

Рассмотрим модель $ARMA(1,1)$, являющуюся комбинацией моделей $AR(1)$ и $MA(1)$: $h_n - a_1 h_{n-1} = a_0 + \varepsilon_n + b_1\varepsilon_{n-1}$. Предположим, что $|a_1| < 1$ (стационарность). В этом случае $\alpha(L) = 1 - a_1 L, \beta(L) = 1 + b_1 L$ и

$$h_n = \frac{a_0}{1-a_1} + \frac{1+b_1 L}{1-a_1 L}\varepsilon_n = \frac{a_0}{1-a_1} + \left(\sum_{k=0}^{\infty} a_1^k L^k\right)(1+b_1 L)\varepsilon_n = \frac{a_0}{1-a_1} + (a_1+b_1)\sum_{k=1}^{\infty} a_1^{k-1}\varepsilon_{n-k} + \varepsilon_n .$$

Отсюда находим, что

$R(k) = a_1 R(k-1), k \geq 2; R(1) = a_1 R(0) + b_1, R(0) = a_1 R(1) + (1 + a_1 b_1 + b_1^2)$, то есть

$$R(0) = Dh_n = \frac{1+2a_1 b_1 + b_1^2}{1-a_1^2}, \rho(k) = \frac{R(k)}{R(0)} = \frac{(1+a_1 b_1)(a_1+b_1)}{1+2a_1 b_1 + b_1^2} a_1^{k-1}.$$

Отметим, что при $|a_1| < 1$ корреляция убывает геометрическим образом при $k \to \infty$.

Вычислим числовые характеристики временного ряда при заданных значениях параметров модели. Имеем: $Eh_n = \frac{a_0}{1-a_1} = \frac{-1}{1-0.5} = -2$. Заметим, что по графику видно, что значения временного ряда колеблются возле отметки -2. Также имеем, что $R(0) = Dh_n = \frac{1+2a_1 b_1 + b_1^2}{1-a_1^2} = \frac{1+2\cdot 0.5\cdot 0.1 + 0.01}{1-0.25} = 1.48$. Вычислим $R(1) = a_1 R(0) + b_1 = 0.84, R(2) = a_1 R(1) = 0.42$. Видно, что значения временного ряда лежат в полосе $[-2.5, -1.5]$.

Поскольку временные ряды в общем и модель $ARMA$ в частности в основном используются для прогнозирования будущего движения цен, опишем кратко финансовые структуры и инструменты, их ключевые объекты.

Теория финансов и финансовая инженерия призваны исследовать свойства финансовых структур и исследовать, как наиболее рациональным образом распоряжаться финансовыми ресурсами с учётом факторов окружающей среды (времени, риска и т.д.), используя разнообразные финансовые инструменты и операции.

Выделяют следующие ключевые объекты:

1. индивидуумы
2. фирмы
3. посреднические структуры
4. финансовый рынок

На финансовом рынке принято различать:

1. основные (первичные) инструменты
(банковский счёт, облигации, акции)
2. производные (вторичные) инструменты
(опционы, фьючерсные и форвардные контракты, варранты, свопы, комбинации, спрэды, сочетания)

Банковский счёт может формироваться по формуле простых, либо сложных процентов. В первом случае проценты начисляются определённое

количество раз в год. Во втором случае проценты начисляются непрерывно. Суть банковского счёта состоит в том, что банк обязуется выплачивать определённый процент от суммы счёта. Облигации (боны) бывают государственными, муниципальными, либо могут принадлежать какой-либо корпорации. Это долговые обязательства, выпускаемые различными финансовыми институтами (государством, банками, корпорациями, акционерными компаниями) с целью аккумулирования капитала. Каждую облигацию характеризует ряд числовых параметров: номинальная стоимость, момент погашения, купонная процентная ставка, начальная цена. Акция – это долевая ценная бумага, выпуская различными финансовыми институтами для увеличения капитала. Опцион – это ценная бумага (контракт), выпускаемая различными финансовыми институтами и дающая покупателю право купить (продать) определённую ценность в установленный период (момент времени) на заранее оговариваемых условиях. В отличие от опциона, фьючерс (форвард) – это соглашение (обязательство) купить (продать) определённую ценность в установленный период (момент времени) на заранее оговариваемых условиях. Опционы бывают двух типов: опционы покупателя (опционы колл) и опционы продавца (опционы пут). По времени исполнения опционы классифицируются на два типа: Европейские (предъявляются к исполнению только в финальный момент времени) и Американские (предъявляются к исполнению в любой момент времени). В зависимости от соотношений между начальной и контрактной ценой опционы бывают трёх типов: с выигрышем, с нулевым выигрышем и с проигрышем. Важной является задача расчёта интервала справедливых цен опциона, нижней границей которого является цена продавца, а верхней границей – цена покупателя. В некоторых случаях цены покупателя и продавца совпадают. В зависимости от контрактной цены различают стандартные опционы (когда контрактная цена это заранее известное число), арифметические азиатские опционы (когда контрактная цена это среднее арифметическое всех рыночных значений цен от начального до финального моментов времени), опционы с последействием (когда контрактная цена это минимальное значение из всех рыночных значений цен от начального до финального моментов времени). Существует выражение для паритета колл-пут, связывающее справедливые цены Европейских опционов колл и пут. Стоит также отметить, что в поведении покупателя и продавца наблюдается большая разница. Покупатель, купив опцион, просто наблюдает за ценами и дожидается финального момента времени. Продавец же, купив опцион, не просто наблюдает за ценами и дожидается финального момента времени, а предлагает стратегию, позволяющую в финальный момент времени выполнить финансовой обязательство. Существует специальная терминология для тех, кто играет на повышение и на понижение. «Бык» – дилер на фондовой бирже, валютном или товарном рынке, ожидающий, что цены поднимутся. На рынке «быков» дилер с большей вероятностью будет покупать, а не продавать. «Медведь» – дилер на фондовой бирже, валютном или товарном рынке, ожидающий, что цены упадут. На рынке «медведей» дилер с большей вероятностью будет продавать, а не покупать. В реальном мире с целью снижения (редуцирования) риска

инвесторы широко прибегают к методам диверсификации, хеджирования, инвестируя средства в самые разнообразные ценные бумаги. При этом различают риск несистематический (то есть, тот, на который инвестор может повлиять своими действиями) и систематический. Дальнейшее развитие идёт в двух направлениях: в предположении определённости и неопределённости.

При описании динамики цен и расчётах мы будем считать, что на рынке отсутствуют арбитражные возможности. Это означает, что никто не может получить прибыль без риска. С математической точки зрения это означает, что существует так называемая мартингальная (риск-нейтральная) вероятностная мера, относительно которой нормированные (дисконтированные) цены оказываются мартингалами. Под дисконтированием в финансовой математике понимается отношение цен рисковых активов к безрисковым. Если такая мера единственная, то рынок полный и справедливая цена опциона, устраивающая и покупателя, и продавца, единственная. Опишем основные вопросы, которые естественно возникают в теории и практике финансового рынка:

1. Как функционирует финансовый рынок в условиях неопределённости.
2. Как описываются цены и какова их динамика во времени.
3. На какие концепции и теории следует опираться при расчётах.
4. Предсказуемо ли будущее движение цен.
5. Каков риск тех или иных финансовых инструментов.

Определим константу

```
Const L = 1000
```

Определим массивы

```
Dim ARMA(1 To L) As Double
Dim hARMA() As Double
Dim iARMA() As Integer
```

Положим

```
i = 2: a0 = -1: a1 = 0.5: b1 = 0.1: sigma = 0.1
myPi = WorksheetFunction.Pi
```

Применим метод Бокса-Мюллера для генерации стандартной нормальной случайной величины.

```
Randomize
x1 = Rnd: x2 = Rnd
epsilon = Sqr(-2 * Log(x1)) * Cos(2 * myPi * x2)
ARMA(1) = a0 + sigma * epsilon:
Do While i < L
Randomize
x1 = Rnd: x2 = Rnd
k = epsilon
epsilon = Sqr(-2 * Log(x1)) * Cos(2 * myPi * x2)
ARMA(i) = a0 + a1 * ARMA(i - 1) + b1 * k + sigma * epsilon:
i = i + 1
Loop
```

Выведем значения массива ARMA в столбец 1 листа. Определим размер динамических массивов iARMA и hARMA равным L/5-1. Чтобы не все точки

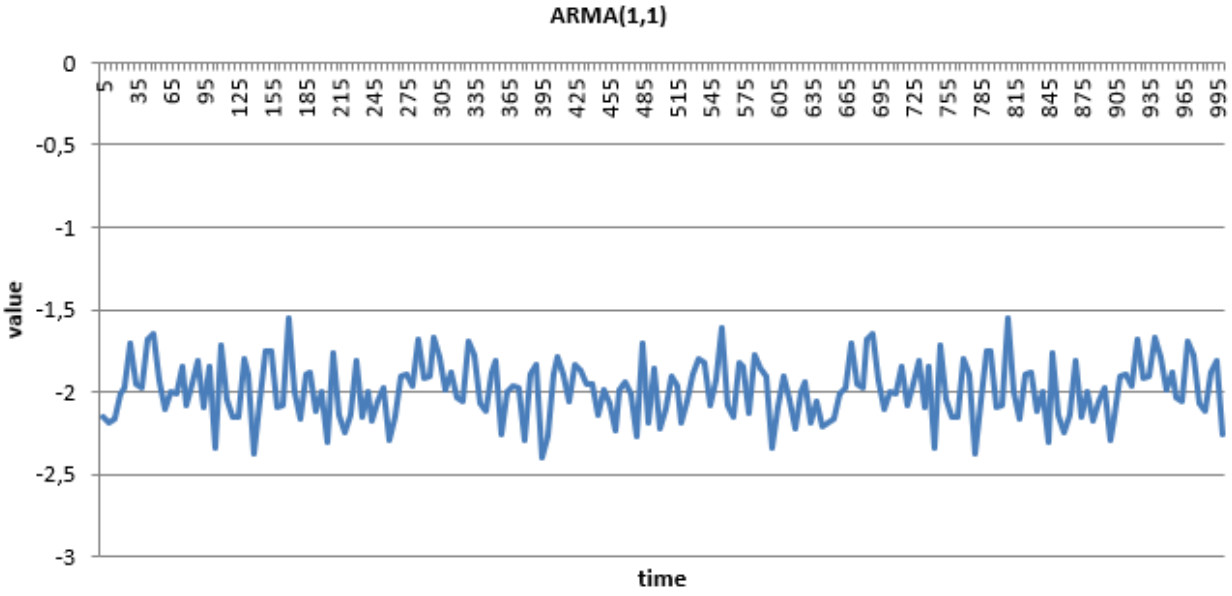
выводить на графике. Заполним динамические массивы iARMA и hARMA.

```
k = 0
ReDim iARMA(L / 5-1): ReDim hARMA(L / 5-1)
For i = 1 To L
Worksheets("8").Cells(i, 1).Value = ARMA(i)
If i Mod 5 = 0 Then iARMA(k) = i: hARMA(k) = ARMA(i): k = k + 1
Next
```

Построим график зависимости hARMA от iARMA.

```
If Worksheets("8").ChartObjects.Count Then Worksheets("8").ChartObjects.Delete
With Worksheets("8").ChartObjects.Add(500, 250, 500, 250)
With .Chart
.ChartType = xlLine
.HasLegend = False
.SeriesCollection.Add Source:=Range("A1:A2")
.SeriesCollection(1).XValues = iARMA
.SeriesCollection(1).Values = hARMA
.HasTitle = True
.ChartTitle.Text = "ARMA(1,1)"
.ChartTitle.Font.Size = 10
.Axes(xlCategory, xlPrimary).HasTitle = True
.Axes(xlCategory, xlPrimary).AxisTitle.Text = "time"
.Axes(xlCategory, xlPrimary).AxisTitle.Font.Size = 10
.Axes(xlCategory, xlPrimary).TickLabels.Font.Size = 10
.Axes(xlValue, xlPrimary).HasTitle = True
.Axes(xlValue, xlPrimary).AxisTitle.Text = "value"
.Axes(xlValue, xlPrimary).AxisTitle.Font.Size = 10
.Axes(xlValue, xlPrimary).TickLabels.Font.Size = 10
End With
End With
```

Отметим, что по аналогии можно построить графики компьютерной реализации моделей $ARMA(1,2)$, $ARMA(2,1)$, $ARMA(2,2)$ и так далее.



Лабораторная работа №7

«Оценивание параметров на основе метода максимального правдоподобия»

Задание. Построить график компьютерной реализации последовательности $h=(h_n)$, подчиняющейся $AR(2)$-модели с $h_n=a_0+a_1h_{n-1}+a_2h_{n-2}+\sigma\varepsilon_n$ с параметрами $a_0=0, a_1=-0.5, a_2=0.01, \sigma=0.1, h_0=h_1=0$ и $\varepsilon_n \sim N(0,1)$, $2\le n\le 1000$.

Говорят, что последовательность $h=(h_n)_{n\ge 1}$ подчиняется авторегрессионной модели (схеме) $AR(p)$ порядка p, если $h_n=\mu_n+\sigma\varepsilon_n$, где $\mu_n=a_0+a_1h_{n-1}+\ldots+a_ph_{n-p}$. Иначе можно сказать, что последовательность $h=(h_n)_{n\ge 1}$ подчиняется разностному уравнению порядка p: $h_n=a_0+a_1h_{n-1}+\ldots+a_ph_{n-p}+\sigma\varepsilon_n$, которое с помощью оператора сдвига L может быть переписано в следующем виде: $(1-a_1L-\ldots-a_pL^p)h_n=a_0+\sigma\varepsilon_n$, или $\alpha(L)h_n=w_n$, где $\alpha(L)=1-a_1L-\ldots-a_pL^p, w_n=a_0+\sigma\varepsilon_n$. В случае $n\ge 1$ необходимо задавать начальные условия $(h_{1-p},h_{2-p},\ldots,h_0)$. Часто полагают $h_{1-p}=h_{2-p}=\ldots=h_0=0$. Их можно также считать случайными, не зависящими от последовательности значений $\varepsilon_1,\varepsilon_2,\ldots$ В эргодических случаях асимптотическое поведение h_n при $n\to\infty$ не зависит от начальных условий, поэтому их конкретизация не является существенной.

Рассмотрим случай $p=1$, тогда $h_n=a_0+a_1h_{n-1}+\sigma\varepsilon_n$. В этом случае из прошлых величин $h_{n-1},\ldots,h_{n-p}$ вклад в h_n вносит только ближайшее по времени значение h_{n-1}. Если $\varepsilon=(\varepsilon_n)_{n\ge 1}$ – последовательность независимых случайных величин, h_0 не зависит от $\varepsilon=(\varepsilon_n)_{n\ge 1}$, то последовательность $h=(h_n)_{n\ge 1}$ будет классическим примером конструктивно заданной марковской цепи. Рекуррентным образом находим $h_n=a_0(1+a_1+\ldots+a_1^{n-1})+a_1^nh_0+\sigma(\varepsilon_n+a_1\varepsilon_{n-1}+\ldots+a_1^{n-1}\varepsilon_1)$. Отсюда видно, что свойства последовательности $h=(h_n)_{n\ge 1}$ существенно зависят от значений параметра a_1; при этом следует различать три случая: $|a_1|<1, |a_1|>1, |a_1|=1$.

Имеем:

$$Eh_n=a_1^nEh_0+a_0(1+a_1+\ldots+a_1^{n-1}),\ Dh_n=a_1^{2n}Dh_0+\sigma^2(1+a_1^2+\ldots+a_1^{2(n-1)}),$$

$$\operatorname{cov}(h_n,h_{n-k})=a_1^{2n-k}Dh_0+\sigma^2a_1^k(1+a_1^2+\ldots+a_1^{2(n-k-1)}), n-k\ge 1.$$

Если $|a_1|<1, E|h_0|<\infty, Dh_0<\infty$, то при $n\to\infty$

$$Eh_n=a_1^nEh_0+\frac{a_0(1-a_1^n)}{1-a_1}\to\frac{a_0}{1-a_1}, Dh_n=a_1^{2n}Dh_0+\frac{\sigma^2(1-a_1^{2n})}{1-a_1^2}\to\frac{\sigma^2}{1-a_1^2},$$

$$\operatorname{cov}(h_n,h_{n-k})\to\frac{\sigma^2a_1^k}{1-a_1^2}.$$

В этом случае последовательность $h=(h_n)_{n\ge 0}$ при $n\to\infty$ «стационаризируется». Более того, если начальное распределение для h_0

является гауссовским, то есть $h_0 \sim N\left(\frac{a_0}{1-a_1}, \frac{\sigma^2}{1-a_1^2}\right)$, то $h=(h_n)_{n\geq 0}$ образует гауссовскую стационарную последовательность с $Eh_n = \frac{a_0}{1-a_1}, Dh_n = \frac{\sigma^2}{1-a_1^2}$ и $\operatorname{cov}(h_n, h_{n+k}) = \frac{\sigma^2 a_1^k}{1-a_1^2}$.

Напомним, что стационарность в узком смысле понимается как выполнение для всех допустимых m и k свойства $\operatorname{Law}(h_0, h_1, ..., h_m) = \operatorname{Law}(h_k, h_{1+k}, ..., h_{m+k})$. Стационарность в широком смысле означает, что $\operatorname{Law}(h_i, h_j) = \operatorname{Law}(h_{i+k}, h_{j+k})$. Если $\rho(k) = \operatorname{corr}(h_n, h_{n+k}) = \frac{\operatorname{cov}(h_n, h_{n+k})}{\sqrt{Dh_n Dh_{n+k}}}$, то $\rho(k) = a_1^k$. То есть, корреляция между значениями h_n и h_{n+k} убывает к нулю геометрическим образом.

Заметим, что для каждого фиксированного n значение h_n в модели $AR(1)$ можно интерпретировать как соответствующее значение h_n в модели $MA(q)$ с $q = n-1$.

В модели $AR(1)$ случай $|a_1| = 1$ соответствует классическому случайному блужданию. Если $a_1 = 1$, то $h_n = a_0 n + h_0 + \sigma(\varepsilon_1 + ... + \varepsilon_n)$. Следовательно, $Eh_n = a_0 n + Eh_0$ и $Dh_n = \sigma^2 n \to \infty$ при $n \to \infty$.

Случай $|a_1| > 1$ является «взрывающимся» в том смысле, что среднее значение Eh_n и дисперсия Dh_n экспоненциально быстро растут с ростом n.

Рассмотрим теперь случай $p=2$: $h_n = a_0 + a_1 h_{n-1} + a_2 h_{n-2} + \sigma\varepsilon_n$, или $(1 - a_1 L - a_2 L^2)h_n = a_0 + \sigma\varepsilon_n$. Если $a_2 = 0$, то $(1-a_1 L)h_n = a_0 + \sigma\varepsilon_n$, и мы получаем модель $AR(1)$. Положим $w_n = a_0 + \sigma\varepsilon_n$. Тогда $(1-a_1 L)h_n = w_n$. Имеем:
$(1 + a_1 L + a_1^2 L^2 + ... + a_1^k L^k)(1 - a_1 L) = (1 - a_1^{k+1} L^{k+1})$. Тогда
$h_n = (1 + a_1 L + a_1^2 L^2 + ... + a_1^k L^k) w_n + a_1^{k+1} L^{k+1} h_n$. Положим $k = n-1$. Тогда
$h_n = (a_0 + \sigma\varepsilon_n) + a_1(a_0 + \sigma\varepsilon_{n-1}) + ... + a_1^{n-1}(a_0 + \sigma\varepsilon_1) + a_1^n h_0$, или
$h_n = (1 + a_1 L + a_1^2 L^2 + ... + a_1^{n-1} L^{n-1})(1 - a_1 L)h_n + a_1^n h_0$.
Если $|a_1| < 1$ и n достаточно велико, то приближённо
$h_n \approx (1 + a_1 L + a_1^2 L^2 + ... + a_1^{n-1} L^{n-1})(1 - a_1 L)h_n$

Рассмотрим стационарную последовательность $\tilde{h} = (\tilde{h}_n)$: $\tilde{h}_n = \sum_{j=0}^{\infty} a_1^j w_{n-j}$. Отметим, что $\tilde{h}_n$ является решением уравнения $(1-a_1 L)h_n = a_0 + \sigma\varepsilon_n$. Покажем, что в классе стационарных решений с конечным вторым моментом это решение является единственным. Пусть $\tilde{h} = (\tilde{h}_n)$ – какое-то другое стационарное решение. Тогда $h_n = \sum_{j=0}^{k} a_1^j w_{n-j} + a_1^{k+1} h_{n-(k+1)}$. Следовательно,

$E\left|h_n - \sum_{j=0}^{k} a_1^j w_{n-j}\right|^2 = a_1^{2(k+1)} E h_{n-(k+1)}^2 = a_1^{2(k+1)} E h_0^2 \to 0$ при $k \to \infty$.

Выразим h_n через w_n. Поскольку для любых λ_1, λ_2 справедливо равенство: $(1-\lambda_1 L)(1-\lambda_2 L) = 1-(\lambda_1+\lambda_2)L + \lambda_1\lambda_2 L^2$. Определим λ_1 и λ_2 из системы $\begin{cases}\lambda_1+\lambda_2 = a_1\\ \lambda_1\lambda_2 = -a_2\end{cases}$. Получим, что $(1-\lambda_1 L)(1-\lambda_2 L) = 1-a_1 L - a_2 L^2$. Отметим, что λ_1 и λ_2 являются корнями квадратного уравнения: $\lambda^2 - a_1\lambda - a_2 = 0$, то есть $\lambda_1 = \frac{a_1+\sqrt{a_1^2+4a_2}}{2}, \lambda_2 = \frac{a_1-\sqrt{a_1^2+4a_2}}{2}$. По-другому можно сказать, что $\lambda_1 = z_1^{-1}, \lambda_2 = z_2^{-1}$, где z_1, z_2 – корни уравнения $1-a_1 z - a_2 z^2 = 0$, а алгебраическое выражение $1-a_1 z - a_2 z^2$ получается из операторного выражения $1-a_1 L - a_2 L^2$ заменой $L \to z$.

Перепишем уравнение $(1-a_1 L - a_2 L^2)h_n = a_0 + \sigma\varepsilon_n$ в виде: $(1-\lambda_1 L)(1-\lambda_2 L)h_n = w_n$. Отсюда $h_n = (1-\lambda_2 L)^{-1}(1-\lambda_1 L)^{-1} w_n$. Если $\lambda_1 \neq \lambda_2$, то $\frac{1}{(1-\lambda_1 L)(1-\lambda_2 L)} = (\lambda_1-\lambda_2)^{-1}\left(\frac{\lambda_1}{1-\lambda_1 L} - \frac{\lambda_2}{1-\lambda_2 L}\right)$, поэтому

$h_n = \frac{\lambda_1}{\lambda_1-\lambda_2}(1-\lambda_1 L)^{-1} w_n - \frac{\lambda_2}{\lambda_1-\lambda_2}(1-\lambda_2 L)^{-1} w_n$.

Предположим, что $|\lambda_i| < 1, i = 1,2$. То есть, корни характеристического уравнения $1-a_1 z - a_2 z^2 = 0$ лежат вне единичного круга. Тогда $(1-\lambda_i L)^{-1} = 1 + \lambda_i L + \lambda_i^2 L^2 + \ldots, i = 1,2$. Следовательно, стационарное решение уравнения $(1-a_1 L - a_2 L^2)h_n = a_0 + \sigma\varepsilon_n$ имеет вид: $h_n = \sum_{j=0}^{\infty}(c_1\lambda_1^j + c_2\lambda_2^j)w_{n-j}$ с коэффициентами $c_1 = \frac{\lambda_1}{\lambda_1-\lambda_2}, c_2 = \frac{\lambda_2}{\lambda_2-\lambda_1}$.

Определим константу

```
Const L = 1000
```

Определим массивы

```
Dim AR(0 To L) As Double
Dim hAR() As Double
Dim iAR() As Integer
```

Положим

```
i = 3: a0 = 0: a1 = -0.5: a2 = 0.01: sigma = 0.1: AR(0) = 0: AR(1) = 0:
myPi = WorksheetFunction.Pi
```

Применим метод Бокса-Мюллера для генерации стандартной нормальной случайной величины.

```
Randomize
x1 = Rnd: x2 = Rnd
epsilon = Sqr(-2 * Log(x1)) * Cos(2 * myPi * x2)
```

```
AR(2) = a0 + sigma * epsilon:
Do While i < L
Randomize
x1 = Rnd: x2 = Rnd
epsilon = Sqr(-2 * Log(x1)) * Cos(2 * myPi * x2)
AR(i) = a0 + a1 * AR(i - 1) + a2 * AR(i - 2) + sigma * epsilon:
i = i + 1
Loop
```

Выведем значения массива AR в столбец 1 листа. Определим размер динамических массивов iAR и hAR равным L/5-1. Чтобы не все точки выводить на графике. Заполним динамические массивы iAR и hAR.

```
k = 0
ReDim iAR(L / 5-1): ReDim hAR(L / 5-1)
For i = 1 To L
Worksheets("7").Cells(i, 1).Value = h(i)
If i Mod 5 = 0 Then iAR(k) = i: hAR(k) = AR(i): k = k + 1
Next
```

Построим график зависимости h от Ind.

```
If Worksheets("7").ChartObjects.Count Then Worksheets("7").ChartObjects.Delete
With Worksheets("7").ChartObjects.Add(500, 250, 500, 250)
With .Chart
.ChartType = xlLine
.HasLegend = False
.SeriesCollection.Add Source:=Range("A1:A2")
.SeriesCollection(1).XValues =iAR
.SeriesCollection(1).Values = hAR
.HasTitle = True
.ChartTitle.Text = "AR(2)"
.ChartTitle.Font.Size = 10
.Axes(xlCategory, xlPrimary).HasTitle = True
.Axes(xlCategory, xlPrimary).AxisTitle.Text = "time"
.Axes(xlCategory, xlPrimary).AxisTitle.Font.Size = 10
.Axes(xlCategory, xlPrimary).TickLabels.Font.Size = 10
.Axes(xlValue, xlPrimary).HasTitle = True
.Axes(xlValue, xlPrimary).AxisTitle.Text = "value"
.Axes(xlValue, xlPrimary).AxisTitle.Font.Size = 10
.Axes(xlValue, xlPrimary).TickLabels.Font.Size = 10
End With
End With
```

В общем, модель линейной авторегрессии можно применять везде, где значения временных рядов линейно зависят от предыдущих значений этого же ряда. Также её можно сочетать с моделью скользящего среднего. Отметим, что для конкретных начальных данных все числовые характеристики случайного процесса, полученные практическим путём из графика, совпадают с теоретическими значениями.

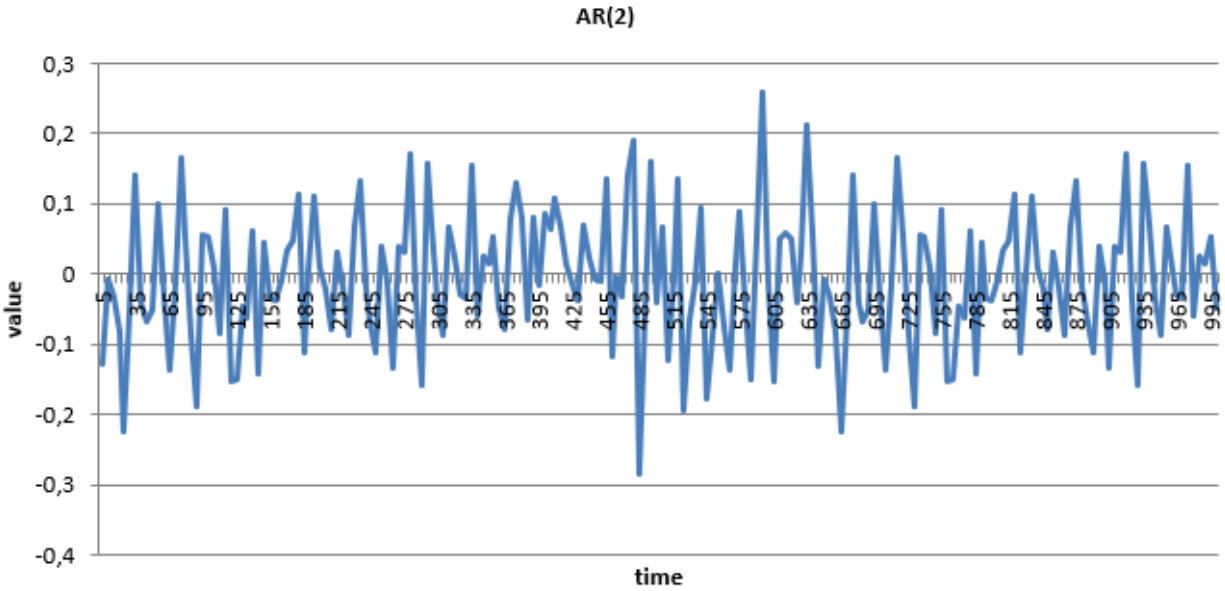
Немаловажными являются наличие стационарного решения и оценка параметров авторегрессионной модели. Как было показано выше, стационарность решения получается, если все корни заданного уравнения лежат вне единичного круга. Параметры достаточно легко оцениваются с помощью метода максимального правдоподобия.

Предположим, что временной ряд представляет собой ежедневные продажи (сегодня, вчера, позавчера и т.д.), а случайная величина отражает влияние на продажи факторов, которые невозможно учесть в модели (например, погода, колебания курса доллара). Тогда, зная параметры модели и прошлые значения временного ряда, можно предсказать его будущие значения. Поэтому основное назначение авторегрессионной модели – прогнозирование. Кроме этого, с её помощью можно производить анализ временных рядов – выявлять тенденцию, сезонность и другие особенности. Заметим, что наблюдения можно производить с разными промежутками времени: ежедневно, еженедельно, поквартально, ежегодно и т.д.

Временной ряд может отражать ежедневное число заболевших в период пандемии, стоимость недвижимости, ценных бумаг, товаров, услуг и многое другое.

Также стоит отметить, что для решения многих задач может применяться метод Монте-Карло. Это численный метод, основанный на моделировании случайных величин с целью вычисления различных характеристик. Интересным был бы анализ поведения авторегрессионной модели с параметрами, порядком и начальными данными, определяемыми пользователем.

Как видно из графика, среднее значение временного ряда равно нулю, что соответствует теоретическому значению, так как начальные значения временного ряда и коэффициент a_0 равны нулю. Значения временного ряда находятся в полосе $[-0.3,0.3]$. При этом наибольшее значение временной ряд достигает в момент времени 590, а наименьшее значение – в момент времени 485. Начальное и конечное значения процесса отрицательны.



Лабораторная работа №8

«Прогнозирование в линейных моделях»

Задание. Построить график компьютерной реализации белого шума $h_n = \sigma\varepsilon_n$ с $\sigma = 0.1$ и $\varepsilon_n \sim N(0,1)$.

Эмпирический анализ эволюции финансовых индексов начинается с построения подходящей вероятностно-статистической модели. В теории временных рядов имеются стандартные модели, например, модель скользящего среднего, модель авторегрессии и многие другие. Уже с небольшим числом параметров ими можно аппроксимировать широкий класс стационарных последовательностей. Однако далеко не все временные ряды являются стационарными, более того, они могут иметь сложную структуру. Анализ показывает, что в статистических данных часто присутствуют три составляющие:

1. Тренд. Он может меняться быстро или медленно (например, в случае инфляции).
2. Периодические или непериодические циклы.
3. Нерегулярная флуктуирующая компонента (стохастическая или хаотическая)

При этом в наблюдаемые данные они могут входить разными способами.

Некоторые модели обладают такими феноменами, которые обнаруживаются только при эмпирическом анализе. К примеру, отклонение от гауссовости, эффекты кластерности и долгая память в ценах.

Поэтому имеет смысл рассмотреть нелинейные модели, например, модель стохастической волатильности. Основная цель эконометрического анализа временных рядов заключается в предсказании будущего поведения цен. Качество предсказания зависит от того, насколько удачно выбрана модель и насколько точно произведена оценка параметров. Обычно при прогнозировании находится доверительная область, расположенная между некоторыми кривыми. Это область, в которой с той или иной степенью надёжности будет происходить предполагаемое движение цен в будущем. Важным является вопрос, что же принять в качестве базисной последовательности. В теории временных рядов в качестве базисной последовательности рассматривается последовательность $\varepsilon = (\varepsilon_n)$, называемая белым шумом.

Белый шум идентифицируется с источником случайности, определяющим стохастический характер исследуемых вероятностно-статистических объектов. При этом говорят, что последовательность $\varepsilon = (\varepsilon_n)$ является белым шумом в широком смысле, если $E\varepsilon_n = 0, E\varepsilon_n^2 < \infty, E\varepsilon_n\varepsilon_m = 0$ для всех $n \neq m$. Иначе говоря, белый шум в широком смысле – это квадратично интегрируемая последовательность некоррелированных случайных величин с нулевыми средними. Если в этом определении добавить ещё требование гауссовости (нормальности), то получаемую последовательность $\varepsilon = (\varepsilon_n)$ называют белым шумом в узком смысле или просто белым шумом. Это равносильно тому, что $\varepsilon = (\varepsilon_n)$ есть последовательность независимых

нормально распределённых случайных величин ($\varepsilon_n \sim N(0,\sigma_n^2)$). В дальнейшем мы будем считать $\sigma_n^2 \equiv 1$. В этом случае часто говорят, что $\varepsilon = (\varepsilon_n)$ – стандартная гауссовская последовательность.

В стохастической финансовой математике очень часто цену рискового актива представляют как $S_n = S_0 \exp(h_n)$. Вычислим среднее значение цены ES_n. Имеем:

$$ES_n = \frac{S_0}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \exp(\sigma_n x) \exp\left(-\frac{x^2}{2}\right) dx = \frac{S_0}{\sqrt{2\pi}} \exp\left(\frac{\sigma_n^2}{2}\right) \int_{-\infty}^{+\infty} \exp\left(-\frac{(x-\sigma_n)}{2}\right)^2 dx = S_0 \exp\left(\frac{\sigma_n^2}{2}\right)$$

Часто требуется посчитать среднее значение функции, зависящей от цены, – $f(S_n)$. Необходимо использовать следующую формулу:

$$Ef(S_n) = Ef(S_0 \exp(\sigma_n \varepsilon_n)) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} f(S_0 \exp(\sigma_n x)) \exp\left(-\frac{x^2}{2}\right) dx.$$

К примеру, в качестве функции f можно рассмотреть Европейский опцион колл (или пут): $f_N = \max(S_N - K, 0)$ ($f_N = \max(K - S_N, 0)$). Параметр K – контрактная цена, то есть цена, по которой в финальный момент времени N можно купить (или продать) опцион.

В отличие от Европейских опционов, на рынке ценных бумаг есть ещё и Американские опционы, которые могут быть предъявлены к исполнению (погашены) в любой момент времени $n = 0,...,N$. Платёжная функция Американского опциона колл (пут) имеет вид: $f_n = \max(S_n - K, 0)$ ($f_n = \max(K - S_n, 0)$). Если Европейский опцион может быть погашен только в финальный момент времени $n = N$, то Американский опцион может быть погашен даже в начальный момент времени $n = 0$.

В общем, существует много процессов на рынке ценных бумаг, основанных на белом шуме. Это различные процессы, описывающие стоимость рисковых активов.

Определим константу

```
Const L = 1000
```

Определим массивы

```
Dim WN(1 To L) As Double
Dim hWN() As Double
Dim iWN() As Integer
```

Положим

```
i = 1: sigma = 0.1
myPi = WorksheetFunction.Pi
```

Отметим, что для числа π нужно применить функцию рабочего листа WorksheetFunction.Pi.

Приведём известные способы генерации нормально распределённых случайных величин.

1. $V_1 = \sum_{i=1}^{12} x_i - 6, x_i$ – независимые, равномерно распределённые на $[0,1]$

случайные величины. Тогда $V_1 \sim N(0,1)$.

2. $V_2 = \sqrt{-2\ln x_1}\cos(2\pi x_2), V_3 = \sqrt{-2\ln x_1}\sin(2\pi x_2), x_1, x_2$ – независимые, равномерно распределённые на $(0,1)$ случайные величины. Тогда $V_2 \sim N(0,1), V_3 \sim N(0,1)$. Случайные величины V_2 и V_3 независимы.

3. $V_4 = x_1\sqrt{\frac{-2\ln s}{s}}, V_5 = x_2\sqrt{\frac{-2\ln s}{s}}, x_1, x_2$ – независимые, равномерно распределённые на $[-1,1]$ случайные величины, $s = \sqrt{x_1^2 + x_2^2}$. Тогда $V_4 \sim N(0,1), V_5 \sim N(0,1)$. Случайные величины V_4 и V_5 независимы. При этом если $s > 1$ или $s = 0$, то такие значения s отбрасываются.

Способы (2) и (3) носят название методов Бокса-Мюллера, так как были получены в 1958 году Джорджем Боксом и Мервином Мюллером. Выразим из (2) переменные x_1, x_2 через переменные V_2, V_3: $x_1 = \exp\left(-\frac{1}{2}(V_2^2 + V_3^2)\right)$, $x_2 = \frac{1}{2\pi}\operatorname{arctg}\left(\frac{V_3}{V_2}\right)$. Тогда $\left|\frac{\partial(x_1, x_2)}{\partial(V_2, V_3)}\right| = \left(\frac{\exp\left(-\frac{V_2^2}{2}\right)}{\sqrt{2\pi}}\right)\left(\frac{\exp\left(-\frac{V_3^2}{2}\right)}{\sqrt{2\pi}}\right)$, то и объясняет возможность применения формул (2) и (3) для генерации нормальных случайных величин.

Способ (1) основан на центральной предельной теореме. В нём для генерации одной нормальной случайной величины необходимо сгенерировать 12 равномерно распределённых случайных величин. В способе (2) для генерации двух нормальных случайных величин необходимо сгенерировать 2 равномерно распределённые случайные величины. В способе (3) для генерации двух нормальных случайных величин необходимо сгенерировать 2 равномерно распределённые случайные величины.

Приведём адаптивный алгоритм генерации нормально распределённой случайной величины. Адаптивным он называется потому, что адаптирован к конкретному компьютеру. Пусть $V = \alpha V_2 + \beta V_3, \alpha^2 + \beta^2 = 1$. Тогда $V = \sqrt{-2\ln x_1}\left(\alpha\cos(2\pi x_2) + \beta\sin(2\pi x_2)\right) = \sqrt{-2\ln x_1}\left(\cos\varphi\cos(2\pi x_2) + \sin\varphi\sin(2\pi x_2)\right) =$ $= \sqrt{-2\ln x_1}\cos(2\pi x_2 - \varphi), 0 \le \varphi \le \frac{\pi}{2}$, то есть если $\varphi = 0$, то $V = V_2 = \sqrt{-2\ln x_1}\cos(2\pi x_2)$; а если $\varphi = \frac{\pi}{2}$, то $V = V_3 = \sqrt{-2\ln x_1}\sin(2\pi x_2)$. Решим следующую задачу: $\chi^2(\varphi) \to \min\limits_{\varphi \in \left[0, \frac{\pi}{2}\right]}$, χ^2 – критерий согласия Пирсона. Результатом решения этой задачи является оптимальный угол φ^*, который мы будем использовать для генерации нормальной случайной величины по формуле: $V = \sqrt{-2\ln x_1}\cos(2\pi x_2 - \varphi^*)$.

Критерий согласия Пирсона для фиксированного угла φ рассчитывается следующим образом. Разобьём отрезок $[-3,3]$ на N частей с шагом $h=\frac{6}{N}$ точками $-3=y_1<y_2<\ldots<y_{N+1}=3, y_0=-\infty, y_{N+2}=+\infty$. Вместе с интервалами $(-\infty,-3)$ и $(3,\infty)$ имеем $N+2$ интервала. Для каждого интервала (y_{i-1},y_i) теоретическая вероятность попадания в интервал равна $p_m(i)=\frac{1}{\sqrt{2\pi}}\int\limits_{y_{i-1}}^{y_i}\exp\left(-\frac{x^2}{2}\right)dx$. Сгенерируем n нормальных случайных величин по адаптивному алгоритму с заданным углом φ. Для каждого интервала (y_{i-1},y_i) практическая вероятность попадания в интервал $p_n(i)$ равна отношению $\frac{n_i}{n}$, где n_i – число случайных величин, попавших в интервал (y_{i-1},y_i). Тогда $\chi^2=n\sum\limits_{i=1}^{N+2}\frac{(p_n(i)-p_m(i))^2}{p_m(i)}$. Чтобы найти минимальное значение критерия, необходимо произвести разбиение отрезка $\left[0,\frac{\pi}{2}\right]$ точками $0=\varphi_0<\varphi_1<\ldots<\varphi_K=\frac{\pi}{2}$. После этого найти $\min\chi^2(\varphi_i), i=0,\ldots,K$.

Применим метод Бокса-Мюллера (2) для генерации стандартной нормальной случайной величины. Обратим внимание, что для генерации равномерно распределённой на $(0,1)$ случайной величины в VBA Excel используется сочетание команд Randomize и Rnd.

```
Do While i < L
Randomize
x1 = Rnd: x2 = Rnd
epsilon1 = Sqr(-2 * Log(x1)) * Cos(2 * myPi * x2)
epsilon2 = Sqr(-2 * Log(x1)) * Sin(2 * myPi * x2)
WN(i) = sigma * epsilon1
WN(i + 1) = sigma * epsilon2
i = i + 2
Loop
```

Выведем значения массива WN в столбец 1 листа.

Определим размер динамических массивов iWN и hWN равным L/5-1. Чтобы не все точки выводить на графике. Заполним динамические массивы iWN и hWN.

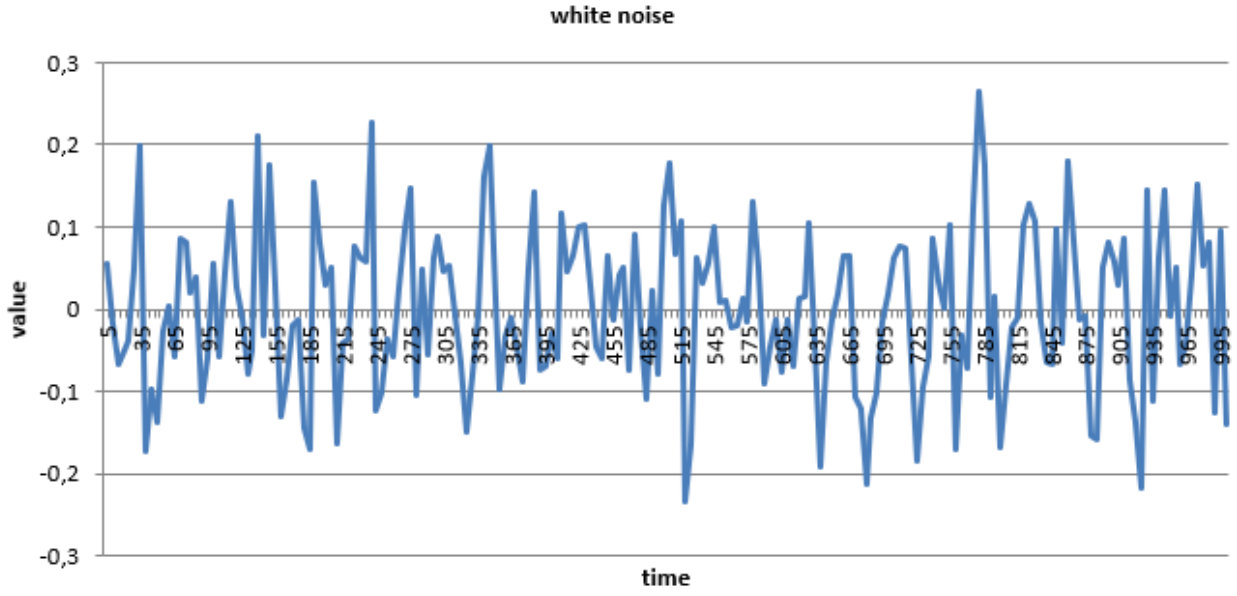
```
k = 0
ReDim iWN(L / 5-1): ReDim hWN(L / 5-1)
For i = 1 To L
Worksheets("5").Cells(i, 2).Value = WN(i)
If i Mod 5 = 0 Then iWN(k) = i: hWN(k) = WN(i): k = k + 1
Next
```

Построим график зависимости hWN от iWN.

```
If Worksheets("5").ChartObjects.Count Then Worksheets("5").ChartObjects.Delete
With Worksheets("5").ChartObjects.Add(500, 250, 500, 250)
With .Chart
.ChartType = xlLine
.HasLegend = False
.SeriesCollection.Add Source:=Range("A1:A2")
.SeriesCollection(1).XValues = iWN
.SeriesCollection(1).Values = hWN
.HasTitle = True
.ChartTitle.Text = "white noise"
.ChartTitle.Font.Size = 10
.Axes(xlCategory, xlPrimary).HasTitle = True
.Axes(xlCategory, xlPrimary).AxisTitle.Text = "time"
.Axes(xlCategory, xlPrimary).AxisTitle.Font.Size = 10
.Axes(xlCategory, xlPrimary).TickLabels.Font.Size = 10
.Axes(xlValue, xlPrimary).HasTitle = True
.Axes(xlValue, xlPrimary).AxisTitle.Text = "value"
.Axes(xlValue, xlPrimary).AxisTitle.Font.Size = 10
.Axes(xlValue, xlPrimary).TickLabels.Font.Size = 10
End With
End With
```

Заметим, что волатильность σ_n характеризует степень разброса значений случайного процесса относительно её математического ожидания. Чем больше σ_n, тем больше разброс, и наоборот. Также отметим, что теперь тип графика xlLine, что означает линия. Цвет графика по умолчанию синий, но может быть изменён в программе.

Кроме этого, стоит отметить, что по правилу «трёх сигм» значения случайной величины h_n должны с большой вероятностью лежать в интервале $[-3\sigma_n, 3\sigma_n]$, то есть в при $\sigma_n = 0.1$ в интервале $[-0.3, 0.3]$.



Лабораторная работа №9

«Нелинейные стохастические условно-гауссовские модели ARCH и GARCH»

Задание. Построить график компьютерной реализации последовательности $h=(h_n)$, подчиняющейся $ARCH(1)$-модели с $h_n=\sqrt{\alpha_0+\alpha_1 h_{n-1}^2}\varepsilon_n$ с параметрами $\alpha_0=0.9, \alpha_1=0.2, h_0=3$ и $0\le n\le 100$.

Пусть (Ω,F,P) – исходное вероятностное пространство, $\varepsilon=(\varepsilon_n)_{n\ge 1}$ – последовательность независимых нормально распределённых случайных величин, $\varepsilon_n \sim N(0,1)$, моделирующих случайность в рассматриваемых далее моделях.

Через F_n обозначаем $\sigma-$алгебры $\sigma(\varepsilon_1,...,\varepsilon_n)$, $F_0=\{\varnothing,\Omega\}$.

Будем интерпретировать $S_n=S_n(\omega)$ как значение цены (акции, обменного курса и т.п.) в момент времени $n=0,1,...$Время может измеряться в годах, месяцах и т.д.

Величины $h=(h_n)_{n\ge 1}$ имеют следующий вид: $h_n=\ln\frac{S_n}{S_{n-1}}=\sigma_n\varepsilon_n$, где волатильности σ_n определяются следующим образом: $\sigma_n^2=\alpha_0+\sum_{i=1}^{p}\alpha_i h_{n-i}^2$ с $\alpha_0>0, \alpha_i\ge 0, h_0=h_0(\omega)$ – случайная величина, не зависящая от $\varepsilon=(\varepsilon_n)_{n\ge 1}$. Часто h_0 считается константой или выбирается из соображений стационарности значений $Eh_n^2, n\ge 0$.

Волатильности σ_n являются предсказуемыми функциями от $h_{n-1}^2,...,h_{n-p}^2$. При этом большие (малые) значения h_{n-i}^2 приводят к большим (малым) значениям σ_n^2. Возникновение же больших h_n^2 в предположении, что предшествующие $h_{n-1}^2,...,h_{n-p}^2$ были малыми, происходит за счёт появления больших значений ε_n. Таким образом, становится понятным, почему рассматриваемые нелинейные модели могут объяснять эффекты типа кластерности, т.е. группирования значений (h_n) в классы больших и малых значений.

Эти рассмотрения оправдывают данное Р. Энглем название для этой модели $ARCH(p)$. Расшифровывается как Autoregressive conditional heteroskedastic model. Переводится как Авторегрессионная модель условной неоднородности. В этой модели условная дисперсия (волатильность) ведёт себя неоднородным образом, поскольку зависит от прошлых значений $h_{n-1}^2, h_{n-2}^2,...$.

Обратимся к рассмотрению ряда свойств последовательности $h=(h_n)_{n\ge 1}$, описываемой $ARCH(1)$-моделью: $\sigma_n^2=\alpha_0+\alpha_1 h_{n-1}^2$. То есть, для $h_n=\sigma_n\varepsilon_n$ имеем следующие свойства: $Eh_n=0, Eh_n^2=\alpha_0+\alpha_1 Eh_{n-1}^2, E(h_n^2/F_{n-1})=\sigma_n^2=\alpha_0+\alpha_1 h_{n-1}^2$. В предположении, что $0<\alpha_1<1$ рекуррентное соотношение $Eh_n^2=\alpha_0+\alpha_1 Eh_{n-1}^2$ имеет единственное стационарное решение $Eh_n^2=\frac{\alpha_0}{1-\alpha_1}, n\ge 0$. То есть, в данном

случае следует взять $h_0^2 = \dfrac{\alpha_0}{1-\alpha_1}$.

Далее имеем: $Eh_n^4 = E\sigma_n^4 E\varepsilon_n^4 = 3E\sigma_n^4 = 3E\left(\alpha_0 + \alpha_1 h_{n-1}^2\right)^2 =$

$= 3\left(\alpha_0^2 + 2\alpha_0\alpha_1 Eh_{n-1}^2 + \alpha_1^2 Eh_{n-1}^4\right) = \dfrac{3\alpha_0^2(1+\alpha_1)}{1-\alpha_1} + 3\alpha_1^2 Eh_{n-1}^4$. Отсюда в предположении, что $0 < \alpha_1 < 1$ и $3\alpha_1^2 < 1$ находим стационарное решение ($Eh_n^4 \equiv \text{const}$): $Eh_n^4 = \dfrac{3\alpha_0^2(1+\alpha_1)}{(1-\alpha_1)(1-3\alpha_1^2)}$. Следовательно, стационарное значение коэффициента эксцесса $K \equiv \dfrac{Eh_n^4}{\left(Eh_n^2\right)^2} - 3 = \dfrac{6\alpha_1^2}{1-3\alpha_1^2}$. Отметим, что $K > 0$. Это говорит о том, что плотность распределения величин (h_n) в окрестности среднего значения вытянута вверх (тем сильнее, чем больше α_1^2). Для нормального распределения эксцесс $K = 0$.

Эмпирическое значение $\overline{K}_N$ коэффициента эксцесса, подсчитываемое по значениям $h_1,...,h_N$, находится по формуле: $\overline{K}_N = \dfrac{\dfrac{1}{N}\sum_{k=1}^{N}\left(h_k - \overline{h}_N\right)^4}{\left(\dfrac{1}{N}\sum_{k=1}^{N}\left(h_k - \overline{h}_N\right)^2\right)^2} - 3$, где $\overline{h}_N = \dfrac{1}{N}\left(h_1 + ... + h_N\right)$. Положительность коэффициента эксцесса для финансовых индексов является скорее правилом, чем исключением. Случаи отрицательных значений эксцесса на практике очень редки.

Последовательность $h = (h_n)$ с $h_n = \sigma_n\varepsilon_n$ является при $0 < \alpha_1 < 1$ квадратично интегрируемой мартингал-разностью и, тем самым, является последовательностью с ортогональными значениями: $\text{cov}(h_n, h_m) = 0, n \neq m$. Это свойство не означает независимости величин h_n и h_m, поскольку их совместное распределение $\text{Law}(h_n, h_m)$не является гауссовским при $\alpha_1 > 0$.

О характере зависимости величин h_n и h_m можно получить представление, рассматривая корреляционную зависимость их квадратов h_n^2 и h_m^2 или модулей $|h_n|$ и $|h_m|$.

Имеем:

$$Dh_n^2 = \frac{2}{1-3\alpha_1^2}\left(\frac{\alpha_0}{1-\alpha_1}\right)^2, Eh_n^2 h_{n-1}^2 = \frac{1+3\alpha_1}{1-3\alpha_1^2}\cdot\frac{\alpha_0^2}{1-\alpha_1},$$

$$\rho(1) = \text{corr}\left(h_n^2, h_{n-1}^2\right) = \frac{\text{cov}\left(h_n^2, h_{n-1}^2\right)}{\sqrt{Dh_n^2 Dh_{n-1}^2}} = \alpha_1.$$

Далее для $k < n$ имеем:

$$Eh_n^2 h_{n-k}^2 = E\left[h_{n-k}^2 E\left(h_n^2 / F_{n-1}\right)\right] = E\left[h_{n-k}^2 E\left(\sigma_n^2\varepsilon_n^2 / F_{n-1}\right)\right] = E\left[h_{n-k}^2\left(\alpha_0 + \alpha_1 h_{n-1}^2\right)\right] = \alpha_0 + \alpha_1 Eh_{n-1}^2 h_{n-k}^2,$$

что даёт в стационарном случае простое рекуррентное соотношение для

$\rho(1)=\dfrac{\operatorname{cov}(h_n^2,h_{n-1}^2)}{\sqrt{Dh_n^2 Dh_{n-1}^2}}$: $\rho(k)=\alpha_1\rho(k-1)$, откуда $\rho(k)=\alpha_1^k$.

Отметим, что $ARCH(p)$-модели тесно связаны с общими авторегрессионными схемами $AR(p)$.

Пусть имеется $ARCH(p)$-модель и $\nu_n=h_n^2-\sigma_n^2$. Тогда если $Eh_n^2<\infty$, то последовательность $\nu=(\nu_n)$ образует относительно потока (F_n) мартингал-разность, и величины $x_n=h_n^2$ удовлетворяют авторегрессионной модели $AR(p)$: $x_n=\alpha_0+\alpha_1 x_{n-1}+\ldots+\alpha_p x_{n-p}+\nu_n$ с шумом $\nu=(\nu_n)$, являющимся мартингал-разностью. Если $p=1$, то $x_n=\alpha_0+\alpha_1 x_{n-1}+\nu_n$.

$ARCH(p)$-модели также тесно связаны с авторегрессионными моделями со случайными коэффициентами, которые используются при описании случайных блужданий в случайных средах.

Пусть $p=1$. Тогда $h_n=\sigma_n\varepsilon_n$, $\sigma_n^2=\alpha_0+\alpha_1 h_{n-1}^2$ и $h_n=\sqrt{\alpha_0+\alpha_1 h_{n-1}^2}\,\varepsilon_n$.

Рассмотрим авторегрессионную модель первого порядка со случайными коэффициентами: $x_n=B_1\eta_n x_{n-1}+B_0\delta_n$, где (η_n) и (δ_n) – две независимые стандартные гауссовские последовательности.

С точки зрения конечномерных распределений последовательность $x=(x_n)$ с $x_0=0$ устроена так же, как и последовательность $\tilde{x}=(\tilde{x}_n)$ с $\tilde{x}_n=\sqrt{\tilde{\varepsilon}_n},\tilde{x}_0=0$, где $(\tilde{\varepsilon}_n)$ – стандартная гауссовская последовательность.

Следовательно, структура образования последовательностей $h=(h_n)$ и $\tilde{x}=(\tilde{x}_n)$ одна и та же. Значит, при $B_0^2=\alpha_0, B_1^2=\alpha_1$ вероятностные законы последовательностей $h=(h_n)$ и $\tilde{x}=(\tilde{x}_n)$ с $h_0=\tilde{x}_0=0$ одни и те же.

Предположим, что величины $h=(h_n)$ подчиняются $AR(1)/ARCH(1)$-модели, то есть $h=(h_n)$ удовлетворяет авторегрессионной схеме $AR(1)$ с $ARCH(1)$-шумом $\left(\sqrt{\alpha_0+\alpha_1 h_{n-1}^2}\,\varepsilon_n\right)_{n\ge1}$.

Условно-гауссовский характер этой модели даёт возможность представить плотность $p_\theta(h_1,\ldots,h_n)$ совместного распределения P_θ величин $h_1,\ldots,h_n$ для заданного значения параметра $\theta=(\alpha_0,\alpha_1,\beta_0,\beta_1)$ в следующем виде ($h_0=0$): $p_\theta(h_1,\ldots,h_n)=(2\pi)^{-n/2}\prod_{k=1}^{n}(\alpha_0+\alpha_1 h_{k-1}^2)^{-1/2}\exp\left(-\dfrac{1}{2}\sum_{k=1}^{n}\dfrac{(h_k-\beta_0-\beta_1 h_{k-1})^2}{\alpha_0+\alpha_1 h_{k-1}^2}\right)$.

В качестве примера использования этого представления рассмотрим задачу оценивания методом максимального правдоподобия неизвестного значения параметра β_1, считая остальные параметры $\alpha_0,\alpha_1,\beta_0$ известными.

Оценка $\hat{\beta}_1$ максимального правдоподобия для параметра β_1 определяется как корень уравнения: $\dfrac{dP(\alpha_0,\alpha_1,\beta_0,\beta_1)}{d\beta_1}(h_1,\ldots,h_n)=0$. Следовательно,

$$\hat{\beta}_1 = \frac{\sum_{k=1}^{n} \frac{(h_k - \beta_0)h_{k-1}}{\alpha_0 + \alpha_1 h_{k-1}^2}}{\sum_{k=1}^{n} \frac{h_{k-1}^2}{\alpha_0 + \alpha_1 h_{k-1}^2}}, \hat{\beta}_1 = \beta_1 + \frac{M_n}{\langle M_n \rangle}, M_n = \sum_{k=1}^{n} \frac{h_{k-1}\varepsilon_k}{\sqrt{\alpha_0 + \alpha_1 h_{k-1}^2}}, \langle M_n \rangle = \sum_{k=1}^{n} \frac{h_{k-1}^2}{\alpha_0 + \alpha_1 h_{k-1}^2},$$ где

M_n – мартингал, $\langle M_n \rangle$ – его квадратическая характеристика.

Здесь $\langle M_n \rangle \to \infty$ (P-п.н.) и, согласно усиленному закону больших чисел для квадратично интегрируемых мартингалов $\frac{M_n}{\langle M_n \rangle} \to 0$ (P-п.н.). Следовательно, построенные оценки $\hat{\beta}_1$ являются сильно состоятельными в том смысле, что $P_\theta\left(\hat{\beta}_1 \to \beta_1\right) = 1$ для значений $\theta = (\alpha_0, \alpha_1, \beta_0, \beta_1)$, где $\beta \in R$.

Определим константу

```
Const L = 1000
```

Определим массивы

```
Dim ARCH(1 To L) As Double
Dim hARCH() As Double
Dim iARCH() As Integer
```

Положим

```
i = 1: alpha0 = 0.9: alpha1 = 0.2: ARCH(0) = 3
myPi = WorksheetFunction.Pi
```

Применим метод Бокса-Мюллера для генерации стандартной нормальной случайной величины.

```
Do While i <= L
Randomize
x1 = Rnd: x2 = Rnd
epsilon = Sqr(-2 * Log(x1)) * Cos(2 * myPi * x2)
ARCH(i) = Sqr(alpha0 + alpha1 * ARCH(i - 1) * ARCH(i - 1)) * epsilon:
i = i + 1
Loop
```

Выведем значения массива ARCH в столбец 1 листа. Определим размер динамических массивов iARCH и hARCH равным L/5-1. Чтобы не все точки выводить на графике. Заполним динамические массивы iARCH и hARCH.

```
k = 0
ReDim iARCH(L / 5 - 1): ReDim hARCH(L / 5 - 1)
For i = 1 To L
Worksheets("10").Cells(i, 1).Value = ARCH(i)
If i Mod 5 = 0 Then iARCH(k) = i: hARCH(k) = ARCH(i): k = k + 1
Next
```

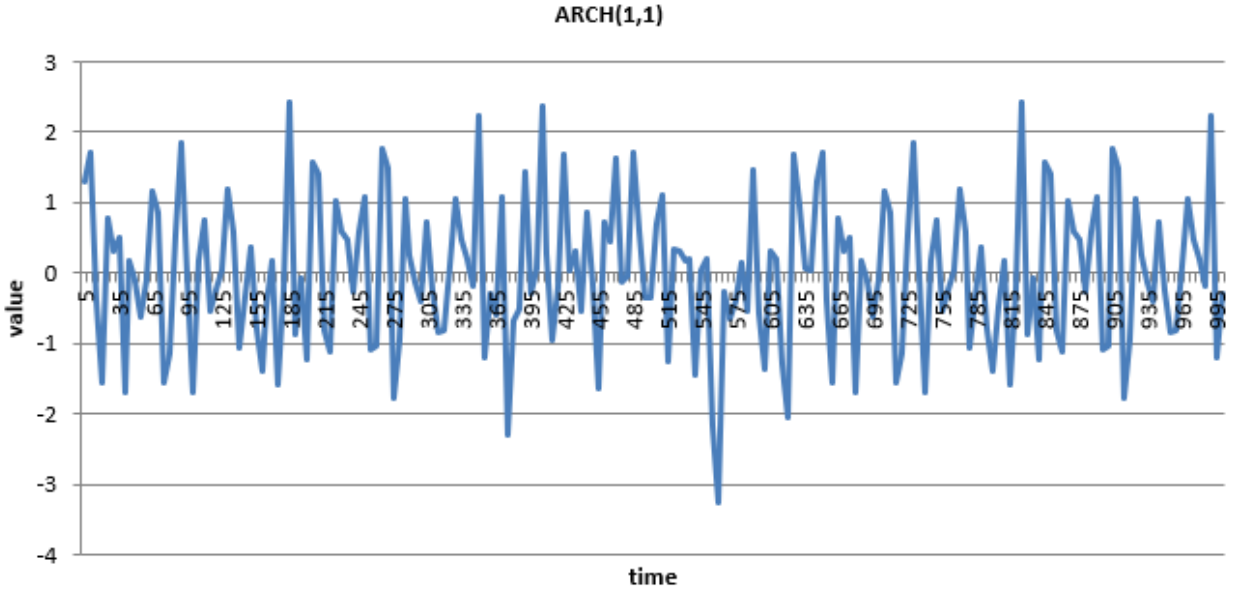
Построим график зависимости hARCH от iARCH.

```
If Worksheets("10").ChartObjects.Count Then
Worksheets("10").ChartObjects.Delete
With Worksheets("10").ChartObjects.Add(500, 250, 500, 250)
With .Chart
```

```
.ChartType = xlLine
.HasLegend = False
.SeriesCollection.Add Source:=Range("A1:A2")
.SeriesCollection(1).XValues = iARCH
.SeriesCollection(1).Values = hARCH
.HasTitle = True
.ChartTitle.Text = "ARCH(1,1)"
.ChartTitle.Font.Size = 10
.Axes(xlCategory, xlPrimary).HasTitle = True
.Axes(xlCategory, xlPrimary).AxisTitle.Text = "time"
.Axes(xlCategory, xlPrimary).AxisTitle.Font.Size = 10
.Axes(xlCategory, xlPrimary).TickLabels.Font.Size = 10
.Axes(xlValue, xlPrimary).HasTitle = True
.Axes(xlValue, xlPrimary).AxisTitle.Text = "value"
.Axes(xlValue, xlPrimary).AxisTitle.Font.Size = 10
.Axes(xlValue, xlPrimary).TickLabels.Font.Size = 10
End With
End With
```

Таким образом, обращение к нелинейным моделям вызвано необходимостью найти объяснение явлениям, которые нельзя объяснить в рамках линейных моделей. Это такие явления, как кластерность цен, катастрофические изменения цен, наличие «тяжёлых хвостов» в распределениях величин $h_n = \ln(S_n / S_{n-1})$, наличие «долгой памяти в ценах» и многие другие.

На рынке флуктуируют макроэкономические и микроэкономические индексы, однако какими моделями следует пользоваться – стохастическими или хаотическими, единодушного мнения нет. Поэтому вопрос выбора правильной модели является самым важным. Необходимо при выборе модели также учитывать, что многие экономические показатели имеют трендовый характер, однако рост может идти как бы циклами, как периодическими, так и непериодическими. То есть, движение может то убыстряться, то замедляться.



Лабораторная работа №10

«Модели стохастической волатильности, интегральные модели»

Задание. Построить график компьютерной реализации последовательности $h=(h_n)$, подчиняющейся $ARIMA(0,1,1)$-модели с $h_n=\mu_n+h_{n-1}+b_1\varepsilon_{n-1}+b_0\varepsilon_n$ с параметрами $\mu=1, b_1=1, b_0=0.1, h_0=0$.

$ARIMA$ – интегрированная модель авторегрессии и скользящего среднего. Это модель временного ряда. Является расширением модели $ARMA$ для нестационарных временных рядов, которые можно сделать стационарными взятием разностей некоторого порядка от исходного временного ряда.

Рассмотренные модели $ARMA(p,q)$ хорошо изучены и успешно применяются при описании стационарных временных рядов. Когда во временном ряду $x=(x_n)$ имеет место нестационарность, иногда простым взятием разностей $\Delta x_n=x_n-x_{n-1}$ или разностей $\Delta^d x_n$ порядка d удаётся получить более стационарную последовательность $\Delta^d x=(\Delta^d x_n)$. Говорят, что последовательность $x=(x_n)$ является $ARIMA(p,d,q)$-моделью, если $\Delta^d x=(\Delta^d x_n)$ образует $ARMA(p,q)$-модель, то есть $\Delta^d ARIMA(p,d,q)=ARMA(p,q)$.

Рассмотрим частный случай – модель $ARIMA(0,1,1)$. Это означает, что $\Delta x_n=h_n$, где (h_n) подчиняется модели $MA(1)$, то есть $\Delta x_n=\mu+(b_0+b_1L)\varepsilon_n$.

Если ввести оператор суммирования («интегрирования») S, определяя его формулой $S=\Delta^{-1}$, или $S=1+L+L^2+\ldots=(1-L)^{-1}$, то $x_n=(Sh)_n$, где $h_n=\mu+(b_0+b_1L)\varepsilon_n=\mu+b_0\varepsilon_n+b_1\varepsilon_{n-1}$.

Тем самым, $x=(x_n)$ можно рассматривать как результат «интегрирования» последовательности $h=(h_n)$, подчиняющейся модели $MA(1)$, что и объясняет происхождение названия $ARIMA=AR+I+MA$, в котором I – Integrated, AR – Autoregression, MA – Moving Average.

Модели $ARIMA$ широко используются в теории Джорджа Бокса и Гвилема Дженкинса. В основу их книги «Анализ временных рядов. Прогноз и управление», первоначально опубликованной в 1970 году, положено использование данных о корреляционных функциях одномерного и многомерного временных рядов. Особое внимание уделено нестационарным временным рядам, содержащим либо стационарные приращения, либо периодические нестационарности. В первый выпуск вошли главы, содержащие основные сведения из корреляционной теории случайных процессов, выбор модели, оценивание её параметров и проверку модели, а также модели для сезонных временных рядов. Во второй выпуск вошли главы, содержащие вопросы оценивания передаточных функций линейных фильтров, задачи автоматического управления в цепях с прямой и обратной связями, а также некоторые другие задачи теории регулирования и управления. Книга очень полезна всем специалистам, встречающимся на практике с анализом и прогнозированием эмпирических величин, меняющихся со временем.

Джордж Бокс – британский статистик, внёсший заметный вклад в такие области, как контроль качества, планирование эксперимента, анализ временных

рядов и Байесовский вывод. Член Лондонского королевского общества. Бокс писал: «В сущности, все модели неправильны, но некоторые полезны».

Гвилем Мейрион Дженкинс – британский статистик и системный инженер, родился в Гауэртоне, Суонси, Уэльс. Он наиболее известен своей новаторской работой с Джорджем Боксом над моделями авторегрессионного скользящего среднего, также называемыми моделями Бокса-Дженкинса, в анализе временных рядов.

Методология Бокса-Дженкинса состоит из пяти этапов:

1. Проверка на стационарность или нестационарность и проверка данных, если необходимо

2. Определение подходящей модели *ARMA*

3. Оценка параметров выбранной модели

4. Диагностическая проверка адекватности модели

5. Прогнозирование или повторение некоторых шагов (при необходимости)

Данный процесс является итеративным, и в нём присутствует субъективная роль разработчика модели при интерпретации двух инструментов: оценочной автокорреляционной функции и частичной автокорреляционной функции.

Во времена Бокса и Дженкинса возможности компьютеров были значительно ограничены, поэтому для оценивания коэффициентов разрабатывались отдельные методы для каждой модели. Сейчас учёные разработали общий метод максимального правдоподобия.

Главная идея применения метода состоит в предположении, что данные имеют некоторое вероятностное распределение, и исчисляется вероятность нужного события. Это в общем случае зависит от нескольких неизвестных параметров. Используя данные, можно максимизировать вероятность этого события. Коэффициенты, при которых достигается максимум вероятности соответствующего события, являются необходимыми оценками параметров. Иногда очень тяжело найти эти оценки в аналитическом виде. В таком случае используют числовые методы оптимизации функции правдоподобия.

В *ARIMA*-моделях во время прогнозирования будущих значений значения объясняющих переменных (регрессоров) можно рассматривать или фиксированными на выборочных значениях, или случайными. Первая возможность приводит к условному прогнозу (наподобие множественной регрессии), вторая – к безусловному прогнозу. Известно, что условная дисперсия случайной величины не превышает её безусловную дисперсию, поэтому точность условного прогноза всегда высшая.

В общем случае модель Бокса-Дженкинса – это математическая модель, предназначенная для прогнозирования диапазонов данных на основе входных данных из указанного временного ряда. Методология позволяет определить тенденции с помощью авторегрессии, скользящих средних и сезонной разницы для создания прогноза. Модель авторегрессионного интегрированного скользящего среднего является разновидностью модели Бокса-Дженкинса. Модель Бокса-Дженкинса – это модель прогнозирования на основе

регрессионных исследований временных рядов. Подходит лучше всего на срок до 18 месяцев. Модель прогнозирует данные на основе трёх принципов: авторегрессии (p), скользящего среднего (q) и интегрирования (d). Процесс авторегрессии (p) проверяет данные на стационарность. Если используемые данные являются стационарными, это может упростить процесс прогнозирования. Если используемые данные нестационарны, их необходимо отличить (d). Данные также проверяются на соответствие скользящему среднему, что используется в части (q) процесса анализа. В целом, первоначальный анализ данных подготавливает их к прогнозированию путём определения параметров p,q,d, которые применяются при разработке прогноза. Одно из применений модели Бокса-Дженкинса – прогноз цен на акции. Этот анализ обычно строится и кодируется с помощью программного обеспечения R. Известно, что модель Бокса-Дженкинса лучше всего подходит для наборов данных, которые в основном стабильны с низкой волатильностью.

Язык программирования R используется для статистической обработки данных и работы с графикой. Широко используется как статистическое программное обеспечение для анализа данных. Кроме Microsoft Excel и R, для анализа временных рядов также хорошо подходят языки Python и MathCAD.

В литературе имеется большой материал по применениям моделей $ARIMA$ в статистике финансовых данных. Подход $ARIMA$ к временным рядам заключается в том, что в первую очередь оценивается стационарность ряда. Различными тестами выявляются наличие единичных корней и порядок интегрированности временного ряда (обычно ограничиваются первым или вторым порядком). Далее при необходимости (если порядок интегрированности больше нуля) ряд преобразуется взятием разности соответствующего порядка, и уже для преобразованной модели строится некоторая $ARMA$-модель, поскольку предполагается, что полученный процесс является стационарным, в отличие от исходного нестационарного процесса (разностно-стационарного или интегрированного процесса порядка d). Теоретически порядок интегрированности d временного ряда может быть не целой величиной, а дробной. В этом случае говорят о дробно-интегрированных моделях авторегрессии-скользящего среднего ($ARFIMA = AR + F + I + MA$, где F – Fractional). Для понимания сущности дробного интегрирования необходимо рассмотреть разложение оператора взятия d-й разности в степенной ряд по степеням оператора L для дробных d (разложение в ряд Тейлора). Отметим также, что часто используется модель $SARIMA$, полученная из модели $ARIMA$ путём включения в неё дополнительных сезонных компонент.

В литературе модель $ARIMA$ и её разновидности представлены очень хорошо в монографии А.Н.Ширяева «Основы стохастической финансовой математики». Том 1. Факты, модели. В монографии, написанной в 1998 году, а затем переизданной в 2004 году, представлены основные линейные и нелинейные модели финансовых индексов, а также их приложение к задачам стохастической финансовой математики. Это задачи прогнозирования, хеджирования, построения оптимального портфеля и многие другие.

Приведены способы оценки параметров, а также ошибки прогноза. Также присутствует анализ хаотических моделей и способы их отличия от стохастических моделей. Приведены компьютерные реализации известных моделей различных порядков и при разных начальных данных и значениях параметров, в том числе модели динамического хаоса.

Приведём код программы для построения компьютерной реализации модели $ARIMA(0,1,1)$ на языке MS Excel. Данный язык программирования удобен тем, что позволяет сочетать работу с таблицами и работу с программным кодом. Причём начальные данные можно задавать в таблице, результирующие данные также выводит в таблицу. Например, в таблицу можно выводить значения полученного временного ряда.

Определим константу

```
Const L = 1000
```

Определим массивы

```
Dim ARIMA(1 To L) As Double
Dim hARIMA() As Double
Dim iARIMA() As Integer
```

Положим

```
i = 1: mu = 1: b1 = 1: b0 = 0.1
myPi = WorksheetFunction.Pi
```

Применим метод Бокса-Мюллера для генерации стандартной нормальной случайной величины.

```
Randomize
x1 = Rnd: x2 = Rnd
epsilon = Sqr(-2 * Log(x1)) * Cos(2 * myPi * x2)
ARIMA(0) = 0:
Do While i <= L
Randomize
x1 = Rnd: x2 = Rnd
k = epsilon
epsilon = Sqr(-2 * Log(x1)) * Cos(2 * myPi * x2)
ARIMA(i) = ARIMA(i - 1) + mu + b1 * k + b0 * epsilon:
i = i + 1
Loop
```

Выведем значения массива ARIMA в столбец 1 листа. Определим размер динамических массивов iARIMA и hARIMA равным L/5-1. Чтобы не все точки выводить на графике. Заполним динамические массивы iARIMA и hARIMA.

```
k = 0
ReDim iARIMA(L / 5 - 1): ReDim hARIMA(L / 5 - 1)
For i = 1 To L
Worksheets("9").Cells(i, 1).Value = ARIMA(i)
If i Mod 5 = 0 Then iARIMA(k) = i: hARIMA(k) = ARIMA(i): k = k + 1
Next
```

Построим график зависимости hARIMA от iARIMA.

```
If Worksheets("9").ChartObjects.Count Then Worksheets("9").ChartObjects.Delete
```

```
With Worksheets("9").ChartObjects.Add(500, 250, 500, 250)
With .Chart
.ChartType = xlLine
.HasLegend = False
.SeriesCollection.Add Source:=Range("A1:A2")
.SeriesCollection(1).XValues = iARIMA
.SeriesCollection(1).Values = hARIMA
.HasTitle = True
.ChartTitle.Text = "ARIMA(1,1)"
.ChartTitle.Font.Size = 10
.Axes(xlCategory, xlPrimary).HasTitle = True
.Axes(xlCategory, xlPrimary).AxisTitle.Text = "time"
.Axes(xlCategory, xlPrimary).AxisTitle.Font.Size = 10
.Axes(xlCategory, xlPrimary).TickLabels.Font.Size = 10
.Axes(xlValue, xlPrimary).HasTitle = True
.Axes(xlValue, xlPrimary).AxisTitle.Text = "value"
.Axes(xlValue, xlPrimary).AxisTitle.Font.Size = 10
.Axes(xlValue, xlPrimary).TickLabels.Font.Size = 10
End With
End With
```

Обратим внимание, что значения временного ряда возрастают, начиная с 0 и заканчивая значением, равным примерно 1000. По оси абсцисс отмечены значения времени, начиная с 0 и заканчивая значением 1000. Видно, что поведение графика временного ряда практически линейное, что подтверждает тот факт, что модели *ARIMA* хорошо подходят для анализа временных рядов с низкой волатильностью. Интересным был бы анализ моделей *ARIMA* с другими параметрами p,q,d с последующим построением графиков моделей. Также можно предложить построить компьютерную реализацию процесса *ARIMA* с другим параметром μ, другими значениями коэффициентов b_i и другими предыдущими значениями h_i.