Министерство образования и науки Российской Федерации

Федеральное государственное автономное образовательное учреждение высшего образования

«Национальный исследовательский Томский политехнический Университет»



Инженерная школа ядерных технологий

01.03.02 «Прикладная математика и информатика»

**Лабораторная работа №2**

по дисциплине:

**Статистическое моделирование и прогнозирование**

Вариант 2

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Выполнил:** |  | | | | |
| Студент группы | 0В01 |  | Белясов А.А |
|  |  |  |
| **Проверил:** | Крицкий О.Л. | | | | |
| преподаватель |  |  |  |  |  | |
|  |  |  |  |  |  | |

г. Томск 2023 г.

**Задание №1**

В соответствии с вариантом задания построить модель ***AR*(*p*) с оптимальным значением лага**, для чего

1. Вычислить выборочные автокорреляции до порядка *n*=10 включительно;
2. проверить исторические данные на случайность, вычислив статистику Льюнга– Бокса;
3. Оценить коэффициенты модели *AR*(*p*) с помощью формул Юла – Уокера;
4. Подобрать оптимальный параметр *p*, проверяя при различных лагах критерий единичного корня Дики-Фулера;
5. Подобрать параметр *p* с помощью информационного критерия Акаике. Сравнить его с найденным в п.4 лагом *p*;
6. Построить алгоритмом *AR*(*p*) прогноз на 25 день торгов.

**Задание №2**

В соответствии с вариантом задания построить модель ***ARMA*(*p*,*q*) с оптимальным значением лага *p***, для чего

1. Вычислить дополнительно заданию №1 выборочные автоковариации до порядка *n*=10 включительно (они нужны для оценки параметров модели ***ARMA*(*p*,1)**). Построить по найденным ранее автокорреляциям коррелограмму и оценить по ней значение лага *p* модели *ARMA*(*p*,1), если это возможно;
2. Центрировав исходные данные и перейдя к вспомогательному ряду , оценить коэффициенты модели *ARMA*(*p*,1), задействовав при решении системы нелинейных уравнений метод Ньютона;
3. Проверяя значимость статистики Бокса-Пирса, найти максимальное значение *m* равных нулю первых автокорреляций ошибок центрированной модели *ARMA*(*p*,1). Будет ли такой *ARMA*(*p*,1)–процесс стационарным?
4. Подобрать параметр *p* с помощью информационного критерия Акаике. Сравнить его с найденным в п.1 лагом *p*;
5. Возвращаясь к исходным данным , построить методом *ARMA*(*p*,1**)** прогноз на 25 день торгов;
6. С помощью программного пакета Statistica проверить качество сделанного прогноза на 25 день торгов, построив *ARMA*(*p*,1) для исходных данных и перейдя по необходимости к дифференцированному с первым порядком ряду (если построение напрямую невозможно).

**Задание №3**

В соответствии с вариантом задания построить модель ***GARCH*(1,1)**, для чего:

1. Перейти к **относительным приращениям** исходных данных;
2. Вычислить выборочное математическое ожидание и смещенную оценку дисперсии для последних *k* = 5, 10, 15 значений временного ряда. Правда ли, что выборочное среднее очень близко к нулю?
3. Вычислить отдельно две последовательности автокорреляций до порядка 5 включительно для рядов и при лаге *k* = 15;
4. Проверить значимость статистики Льюнга–Бокса для вычисленных ранее двух последовательностей автокорреляций. Показать, что для ряда статистика Льюнга-Бокса может быть значима;
5. Возвращаясь к исходным данным, построить методом ***GARCH*(1,1)** прогноз на 25 день торгов, если это возможно;
6. C помощью программного пакета Matlab проверить качество сделанного прогноза на 25 день торгов, построив ***GARCH*(1,1)** для исходных данных (использовать model = garch('garchlags',1,'archlags',1); [estM1,H,logL] = estimate(model,a), где a – ряд ценовых приращений; функцию forecast).

Таблица 1

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| День | Номер варианта | | | | | | | | | |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 1 | 43.88 | 18.58 | 49.76 | 8.50 | 59.90 | 63.11 | 45.27 | 8.64 | 7.01 | 41.59 |
| 2 | 44.15 | 18.75 | 50.06 | 8.55 | 59.95 | 63.54 | 45.34 | 8.66 | 7.02 | 41.76 |
| 3 | 44.47 | 18.83 | 50.37 | 8.54 | 60.16 | 63.52 | 45.51 | 8.69 | 6.99 | 41.89 |
| 4 | 44.61 | 18.61 | 50.49 | 8.52 | 59.50 | 63.28 | 45.55 | 8.62 | 7.00 | 41.75 |
| 5 | 44.52 | 18.62 | 50.59 | 8.49 | 59.37 | 63.12 | 45.17 | 8.61 | 6.97 | 41.58 |
| 6 | 44.51 | 18.50 | 50.44 | 8.50 | 59.61 | 63.23 | 45.35 | 8.64 | 6.99 | 41.70 |
| 7 | 44.78 | 18.32 | 50.84 | 8.52 | 59.40 | 63.29 | 45.43 | 8.64 | 7.00 | 41.77 |
| 8 | 44.73 | 18.43 | 50.82 | 8.51 | 59.18 | 63.23 | 45.32 | 8.66 | 6.99 | 41.68 |
| 9 | 44.73 | 18.37 | 50.39 | 8.50 | 59.35 | 63.18 | 44.72 | 8.63 | 7.01 | 41.56 |
| 10 | 45.06 | 18.67 | 50.92 | 8.56 | 59.67 | 63.73 | 44.83 | 8.68 | 7.06 | 41.83 |
| 11 | 44.95 | 18.78 | 50.96 | 8.59 | 59.50 | 63.94 | 44.78 | 8.68 | 7.09 | 41.90 |
| 12 | 44.78 | 18.72 | 50.74 | 8.56 | 59.22 | 63.62 | 44.62 | 8.63 | 7.09 | 41.73 |
| 13 | 44.54 | 18.66 | 50.75 | 8.54 | 59.15 | 63.43 | 44.98 | 8.60 | 7.09 | 41.65 |
| 14 | 45.10 | 18.81 | 51.39 | 8.61 | 59.64 | 64.02 | 45.60 | 8.67 | 7.17 | 42.07 |
| 15 | 45.42 | 18.98 | 51.40 | 8.66 | 60.32 | 64.34 | 46.01 | 8.77 | 7.24 | 42.24 |
| 16 | 45.38 | 19.15 | 51.06 | 8.65 | 60.16 | 64.43 | 45.72 | 8.74 | 7.22 | 42.16 |
| 17 | 45.39 | 19.21 | 51.65 | 8.66 | 60.09 | 64.29 | 45.84 | 8.73 | 7.24 | 42.36 |
| 18 | 45.71 | 19.14 | 52.23 | 8.75 | 60.31 | 65.03 | 46.14 | 8.76 | 7.31 | 42.68 |
| 19 | 45.94 | 19.17 | 52.31 | 8.71 | 59.99 | 64.73 | 46.16 | 8.72 | 7.32 | 42.54 |
| 20 | 45.38 | 18.99 | 51.70 | 8.58 | 59.31 | 63.82 | 45.52 | 8.63 | 7.19 | 41.96 |
| 21 | 44.97 | 18.83 | 51.61 | 8.48 | 58.71 | 63.16 | 45.05 | 8.55 | 7.13 | 41.65 |
| 22 | 45.13 | 19.00 | 51.62 | 8.49 | 59.19 | 63.24 | 44.91 | 8.60 | 7.11 | 41.76 |
| 23 | 45.41 | 19.07 | 51.88 | 8.53 | 59.51 | 63.45 | 45.17 | 8.65 | 7.14 | 41.95 |
| 24 | 45.03 | 18.93 | 51.51 | 8.49 | 59.02 | 63.03 | 44.91 | 8.59 | 7.09 | 41.68 |
| валют | Австр. долл | Бр. реал | Вон, Корея | Дат. крона | доллар | евро | Кан. доллар | юань | Норв. крона | Син. Д. |

**Теоретическая часть**

*Авторегрессионная модель AR(p)*

Рассмотрим математическую модель, которая позволяет обработать эмпирические данные (например, котировки акций и т.п.) и оценить их будущее значение. Предлагаемая авторегрессионная модель позволяет получить оценку временного ряда о его предыдущих состояниях.

Рассмотрим авторегрессионную модель *AR(p)* порядка *p*:

|  |  |
| --- | --- |
|  | (1) |

где *a*0, *a*1, ..., *ap*, – некоторые коэффициенты, – нормальное распределенные случайные величины с нулевым средним и дисперсией единица.

В общем случае оценку коэффициентов модели *AR(p)* ленче всего проводить с помощью рекуррентных формул Юла-Уокера:

где – выборочные автокорреляции процесса - неизвестные коэффициенты модели *AR(p)*, *j=1, 2, …, p*. При этом кулевой коэффициент выражается из формулы для математического ожидания:

*Статистическая проверка нестационарности модели AR(p)*

Для проверки статистической гипотезы о нестационарности ряда значений модели (1) используется критерий Дики-Фулера (Dickey-Fuller test, ADF). Нулевая гипотеза *H0* состоит в том, что ряд нестационарен и = 1 при альтернативной гипотезе, что ряд стационарен и :

При построении статистического критерия дополнительно предполагается, что шумы (1) некоррелированы и в силу нормальности независимы: без выполнения этого условия критерий работать не будет. Для проверки некоррелированности шумов нужно использовать критерий Дарбина-Уотсона (DW-test).

Критическая статистика: , где – диагональный элемент обратной матрицы модели (1), *Х –* центрированные столбцы данных правой части в (1).

Нулевая гипотеза *H*0 принимается, если *t*кр<γ. В противном случае принимается альтернативная гипотеза (при γ< *t*кр).

**Замечание**: очень часто до проверки *H*0 порядок авторегрессии *p* неизвестен. Для его нахождения прибегают к одной из следующих процедур:

* 1. Выбрать относительно большое значение *p* и, проверяя значимость гипотезы *H*0, последовательно уменьшать *p* на единицу. Остановиться при первой статистически неподтвержденной *H*0 (т.е. при переходе модели от нестационарной к стационарной).
  2. При различных значениях *p* использовать информационный критерий Акаике.

,

где – погрешности модели (остатки), выбирая ту модель, у которой AIC меньше.

При каждом выбранном *p* проверять некоррелированность шумов, для чего использовать критерий Дарбина-Уотсона. Если гипотеза *H0* статистически значима, нужно увеличить *p* и повторить процедуру.

*Aвторегрессионная модель со скользящим средним ARMA(p,q)*

Обобщим модель (1), добавив к ней нормально распределенные ошибки наблюдений , с нулевым средним и дисперсиями , которые вносят так называемую скользящую среднюю ошибку, и убрав константу:

|  |  |
| --- | --- |
|  | (2) |

По аналогии с *AR(p)* – процессом, *ARMA(p,q)* будет стационарным, если первые m его автокорреляций остатков модели равны нулю. Для проверки гипотезы используют статистику Бокса – Пирса:

где n – число данных, p, q – параметры модели, – выборочные автокорреляции между эмпирическими и теоретическими значениями модели:

Для этой статистики нулевая гипотеза о равенстве нулю первых m автокорреляций остатков подтверждается, если при заданном уровне значимости .

Наряду со статистикой Бокса-Пирса можно использовать и упоминавшуюся ранее статистику Льюнга-Бокса. Для *ARMA(p,q)* – процесса она имеет вид:

Нулевая гипотеза о равенстве нулю первых m автокорреляций ошибок подтверждается (а значит, *ARMA(p,q)* – процесс будет стационарным), если при заданном уровне значимости .

*Оценивание коэффициентов модели ARMA(p,q)*

Заметим, что один и тот же временной ряд может быть описан моделью *ARMA(p,q)* с различными значениями параметров *p*, *q*. Поэтому при практических вычислениях требуется выбрать порядки модели, оценить ее параметры, провести статистическое оценивание правильности модели в целом.

Если , то математическое ожидание нужно вводить в модель в качестве параметра, потому что если процесс стационарен, то из (2)

и эту конструкцию можно использовать в авторегрессионной модели.

В целом, вместо этого удобнее рассмотреть вспомогательный временной ряд с нулевым математическим ожиданием, который будет удовлетворять (2), а затем перейти к исходному ряду с помощью преобразования: . В качестве так же можно использовать центрированные значения :.

Перейдем к центрированным значениям исходного временного ряда . Далее, нам нужно оценить первые *p* параметров (шумовые слагаемые пока не разделяем на составляющие, считая их одним шумом). Как и при вычислении автоковариаций, будем умножать (2) последовательно на и вычислять математические ожидания от обоих частей равенства. Имеем:

…

Для нахождения решения системы относительно неизвестных вместо теоретических значений автоковариации будем использовать их выборочные аналоги:

Найдя оценки , перейдем к процессу *,* для которого рассчитаем первые q выборочные автокорреляции. Их используем для вычисления оценок параметров . Действительно, в соответствие (2) фактически

– процесс скользящего среднего. Его автокорреляции можно сравнить с выборочными аналогами:

где – выборочные автокорреляции. Эта система нелинейная относительно неизвестных .

*Определение порядка модели ARMA(p,q)*

Для определения порядков модели *p* и *q* необходимо вычислить достаточное количество выборочных автокорреляций и построить коррелограмму. Первым признаком стационарности общего *ARMA(p,q)*–процесса является быстрое убывание с ростом *k*. Выбирая поэтому в качестве *p* порядок последней достаточно большой по модулю выборочной частной автокорреляции, мы с большой точностью находим требуемую величину лага.

Аналогично, для определения порядка *q* (если *p*=0) для скользящего среднего можно использовать выборочные автокорреляции, потому что его теоретические автокорреляции становятся равными нулю, начиная с лага *q*.

В случае общего *ARMA(p,q)* – процесса, когда оба коэффициента не равны нулю, такой подход не приносит успеха. В этом случае можно отслеживать только скорость убывания с ростом *k*. Кроме того, хороший результат дает информационный критерий Акаике:

,

где – погрешности модели (остатки), – исторические значения временного ряда.

Перебор значений *p* и *q* останавливается для той модели, у которой *AIC* меньше.

Вместе с использованием численных процедур оправдана дополнительная проверка статистических гипотез относительно некоррелированности и гомоскедастичности (стационарности) погрешностей , рассмотренных ранее.

*Обобщенная авторегрессионная модель условной неоднородности GARCH(p,q)*

Метод *GARCH(p,q)* позволяет прогнозировать волатильность, проводить анализ коррелированных и высокочастотных данных. Данный метод основан на предположении авторегрессионной зависимости вида:

|  |  |
| --- | --- |
| , | (3) |

где – коэффициенты модели, подлежащие оценке, , – относительные приращения значений временного ряда или логарифмические приращения – волатильность, .

Для простоты изложения рассмотрим (3) c *p*=1 и *q*=1. Тогда модель будет иметь вид:

|  |  |
| --- | --- |
|  | (4) |

где – коэффициенты модели, подлежащие оценке – долговременное среднее отклонение в структуре данных, , – относительные приращения значений временного ряда – волатильность, .

Предположим, произвольность распределения дневных приращений , – цена некоторого актива а *n*-ый день торгов. Так как , то (4) можно представить в виде:

|  |  |
| --- | --- |
|  | (5) |

Записывая (5) в будущий момент времени (*n*+*k*), получаем, что:

Оценим дневную волатильность по последним *k* наблюдениям:

где – выборочное среднее, *k* – лаг (задержка) временного ряда.

Так как , то по определению дневной волатильности следует, что . Учитывая это, имеем:

|  |  |
| --- | --- |
|  | (6) |

Равенство (6) определяет условие устойчивости *GARCH(1,1)*. Действительно, если , то последнее слагаемое вносит все меньший вклад в математическое ожидание и с ростом лага k стремится к V:

|  |  |
| --- | --- |
|  | (7) |

Если же *α+β ≥ 1*, то и случайный процесс будет в среднем стремительно возрастать, что свидетельствует о его неустойчивости.

Исходя из (7) можно сделать вывод, что *V* характеризует уровень возврата временного ряда к прежнему состоянию с коэффициентом возврата γ.

Перейдем к определению параметров модели (3). Наиболее общим способом их оценки является нахождение максимума функции правдоподобия:

или логарифмической функции правдоподобия

|  |  |
| --- | --- |
|  | (8) |

где – функции плотности распределения наблюдений, *m –* число наблюдений. Например, если предположить, что имеет место нормальное распределение приращений с математическим ожиданием *a* и дисперсией , то

В случае относительных приращений цен акций математическое ожидание . Легко показать, что задача определения максимума выражения (8) совпадает с нахождением максимума функции (это метод квазимаксимального правдоподобия, если закон распределения неизвестен или отличен от нормального)

|  |  |
| --- | --- |
|  | (9) |

Поиск максимума (9) осуществляется в соответствие с выполнением необходимого условия экстремума функции трех переменных:

где (нужно посчитать все частные производные по параметрам от (9) и получить реккурентные соотношения).

*Случайность данных*

Используем несколько выборочных автокорреляций , для доказательства гипотезы о случайности значений исходного временного ряда . Рассмотрим Q-статистику Льюнга-Бокса в предположении о нормальности распределения значений :

Было показано, что статистический критерий работает даже в условиях отсутствия нормального закона распределения для исходного временного ряда (выполнена ЦПТ, т.е. дисперсия D( ) < ∞). Нулевая гипотеза состоит в том, что ряд является винеровским процессом, т.е. процессом с независимыми приращениями и нулевым средним. Если значение статистики Q(r) больше критического (табличного) значения функции распределения при заданном уровне значимости и числе степеней свободы, то признается наличие ненулевых автокорреляций до порядка m включительно.

*Некоторые статистические критерии, используемые при анализе временных рядов*

Выборочной автоковариацией *k*-го порядка временного ряда называется число:

Выборочной автокорреляцией *k*-го порядка временного ряда называется число:

Известно, что выборочные автокорреляции имеют нормальное асимптотическое распределение.

*Стационарность временного ряда*

Автокорреляции удобны и для проверки временного ряда на стационарность. В целом можно заметить, что для выявления стационарности нужно вычислять автокорреляции до некоторого порядка и заметить, что коррелограмма быстро убывает после нескольких первых значений (первое значение может быть любым). Если же автокорреляция первого порядка близка к единице, а коррелограмма медленно убывает по экспоненте, то это свидетельствует о нестационарности .

*Статистическая значимость временного ряда*

Пусть для временного ряда вычислены автокорреляции до k-го порядка включительно. Проверим статистическую гипотезу о величине первой автокорреляции ρ1, вычисляя статистику Дарбина-Уотсона (DW-test):

где – погрешность модели (разность между наблюдаемым и модельным значением).

Значения статистики γ лежат в интервале [0, 4]. Распределение статистики известно и имеет два критических значения: dl и du, их значения приведены в таблицах. Выдвигаем нулевую гипотезу H0: ρ1=0, где ρ1 - первая автокорреляция (остальные автокорреляции статистически не проверяются). Нулевая гипотеза о нулевой автокорреляции подтверждается, если du < γ < 4-du. H0 отклоняется в пользу альтернативы о наличии положительной автокорреляции, если γ < dl. H0 отклоняется в пользу альтернативы о наличии отрицательной автокорреляции, если 4-dl<γ. Зона неопределенности критерия, когда нельзя ни принять основную гипотезу, ни принять альтернативную, состоит из двух интервалов, описываемых неравенствами: dl < γ < du и 4-du < γ < 4-dl .

**Практическая часть**

***Построение AR(p) модели с оптимальным значением лага***

Вычислим автокорреляции до порядка *n*=10 включительно. Результаты вычисления представлены в табл. 2.

Таблица 2

Значение автокорреляций до порядка *n*=10

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Порядок | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| Значение | 1 | 0.831 | 0.644 | 0.515 | 0.372 | 0.280 | 0.160 | 0.003 | -0.148 | -0.339 | -0.449 |

Проверим исторические данные на случайность, вычислив статистику Льюнга– Бокса. Значение статистики Люнга-Бокса составила , а значение квантиля распределения хи-квадрат на уровне значимости 0.05 – . Следовательно, гипотезу о случайности исходных данных отвергаем.

Первоначально построим модель *AR(p)* для *p*=3. Произведем оценку параметров, выбранной модели при помощи формул Юла-Уокера. В результате получим следующие значения параметров модели:

Подберем оптимальный параметр *p*, проверяя при различных лагах критерий единичного корня Дики-Фулера. Проверим данный критерий при *p*=3. Получим, что значение критерия единичного корня Дики-Фулера составляет , а критическая точка . В итоге, получаем, что при *p*=1 принимаем нулевую гипотезу о нестационарности данных.

Подберем параметр *p* с помощью информационного критерия Акаике. Результаты расчета данного критерия при различных *p* :

AIC(p=3)=2.09

AIC(p=2)=3.99

Отсюда делаем вывод, что модель с лагом 3 оптимальна.

Исходя из полученных данных в табл. 4, выберем значение *p*=3. Проверим для данного *p* некоррелированность шумов, для чего используем критерий Дарбина-Уотсона. При данном *p* значение , а значение статистики Дарбина-Уотсона равно .

Таким образом, , следовательно, гипотеза *H0* принимается.

Построим прогноз на 25 день торгов, а так же найдем прогнозы для уже имеющихся исторический данных (проведем бэк-тестинг)

Предсказанное значение равно 18.879

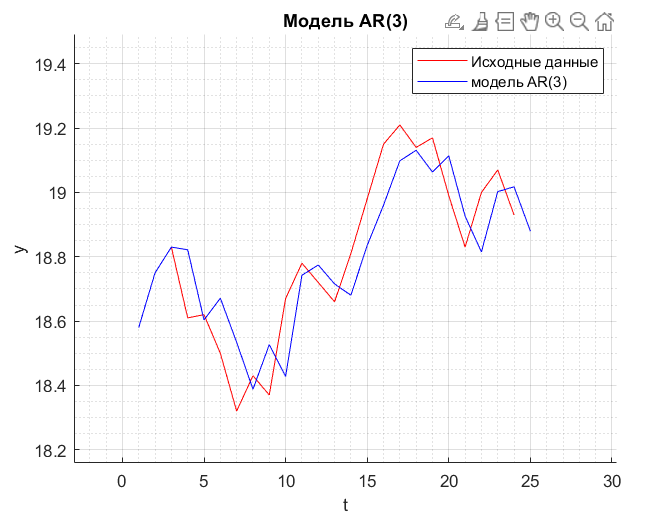


Рисунок 1. Модель AR(3)

***Построение ARMA(p,1) модели с оптимальным значением лага***

Первоначально вычислим дополнительно заданию №1 выборочные автоковариации до порядка *n*=10 включительно, которые необходимы для оценки параметров модели *ARMA(p,1)*. По найденным автокорреляциям построим коррелограмму и оценим по ней значение лага *p* модели *ARMA*(*p*,1). Результат вычисления представлен в табл. 3.

Таблица 3. Значение автокорреляций до порядка *n*=10

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Порядок | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| Значение | 1 | 0.831 | 0.644 | 0.515 | 0.372 | 0.280 | 0.160 | 0.003 | -0.148 | -0.339 | -0.449 |

Выберем *p*=3 и оценим коэффициенты модели *ARMA*(*p*,1), центрировав исходные данные и перейдя к вспомогательному ряду . Результат оценки коэффициентов модели представлен ниже:

Проверим значимость статистики Бокса-Пирса при данном *р*=3. Получим значение , что меньше

В итоге, модель *ARMA*(*p*,1) при *р*=3 построить можно.

Далее подберем параметр *p* с помощью информационного критерия Акаике. Результаты расчета данного критерия при различных *p* представлен в табл. 4.

Таблица 4

Значение информационного критерия Акаике

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| p | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| AIC | 3.622 | 6.766 | 7.358 | 6.898 | 6.892 | 6.691 | 7.563 | 7.317 | 7.125 |

Исходя из полученных результатов, представленных в табл. 7, наилучшее значение *p*=1. Коэффициенты модели имеют следующий вид:

Проверим значимость статистики Бокса-Пирса при данном *р*=1. Получим значение , что меньше В итоге, модель *ARMA*(*p*,1) при *р*=1 построить можно.

Так же проверим статистику Льюнга-Бокса. Получим

Таким образом ряд стационарен, а модель значима. Поэтому мы можем сделать прогноз на 25 день торгов. А так же провести бэк-тестинг.

Предсказанное значение равно 18.9064

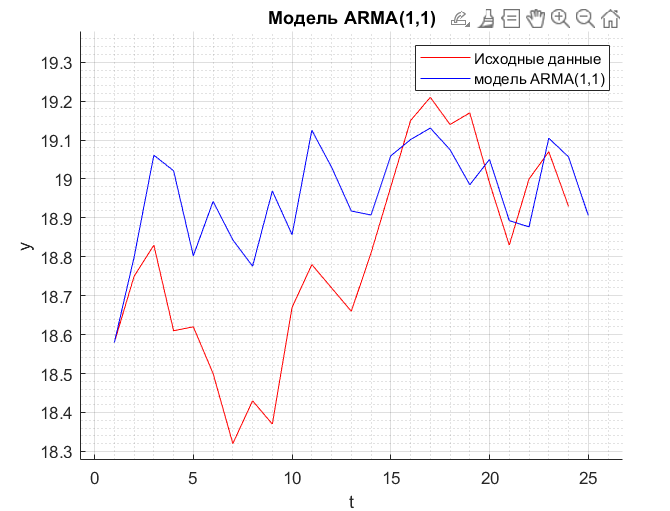


Рисунок 2. Модель ARMA(1,1)

1. *Построение GARCH(1,1) модели с оптимальным значением лага*

Перейдем к относительным приращениям исходных данных по следующей формуле: .

Теперь вычислим выборочное математическое ожидание и смещенную оценку дисперсии для последних *k* = 5, 10, 15 значений временного ряда. Результаты вычислений представлены в табл. 5.

Таблица 5

Значения выборочного мат. ожидания и смещенной дисперсии

|  |  |  |
| --- | --- | --- |
| *k* | *E(x)* | *D(x)* |
| 5 | -0.0025 |  |
| 10 |  |  |
| 15 | 0.002 |  |

Анализируя полученные данные в табл. 9, можно сказать, что выборочные средние очень близки к нулю.

Отдельно вычислим две последовательности автокорреляций до порядка 5 включительно для рядов и при лаге *k* = 15.

В табл. 6 представлен результат вычисления автокорреляций до порядка 5 включительно для ряда .

Таблица 6

Автокорреляций до порядка 5 включительно для ряда

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Порядок | 0 | 1 | 2 | 3 | 4 | 5 |
| Значение | 1 | 0.1417 | 0.0328 | -0.542 | -0.0443 | -0.1064 |

Вычислим значения при лаге *k* = 15. Полученные значения представлены в табл. 7.

Таблица 7

Значения при лаге *k* = 15

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| № | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| Значение |  |  |  |  |  |  |  |  |

Таким образом, построив ряд при лаге *k* = 15, вычислим его автокорреляции до 5 порядка включительно. Результат вычисления представлен в табл. 8.

Таблица 8

Автокорреляций до порядка 5 включительно для ряда

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Порядок | 0 | 1 | 2 | 3 | 4 | 5 |
| Значение | 1 | 0.0995 | 0.0232 | -0.5456 | 0.0035 | -0.0869 |

Следующим шагом проверим значимость статистики Льюнга–Бокса для вычисленных ранее двух последовательностей автокорреляций. Для ряда значение статистики Люнга-Бокса составило , а значение квантиля распределения хи-квадрат на уровне значимости 0.05 – . Следовательно, гипотезу о случайности исходных данных принимаем. Для ряда значение статистики Люнга-Бокса составило , а значение квантиля распределения хи-квадрат на уровне значимости 0.05 – . Следовательно, гипотезу о случайности исходных данных принимаем. Следовательно, ряды и являются стационарными.

Вернемся к исходным данным и построим методом *GARCH(1,1)*прогноз на 25 день торгов. Результат прогноза: . При этом значения коэффициентов, следующие (они были вычислены при помощи встроенного метода в пакете Matlab):

При этом НЕ выполняется условие .

Кроме этого с помощью программного пакета Matlab проверим качество сделанного прогноза на 25 день торгов, построив *GARCH(1,1)*для исходных данных. Для этого использовался следующий программный код:

*model = garch('garchlags', 1, 'archlags', 1);*

*[estM1, H, logL, info] = estimate(model, transpose(data), 'Display', {'params'});*

*frcst = forecast(estM1, 1, transpose(data));*

Результат работы программного кода представлен на рис. 3.

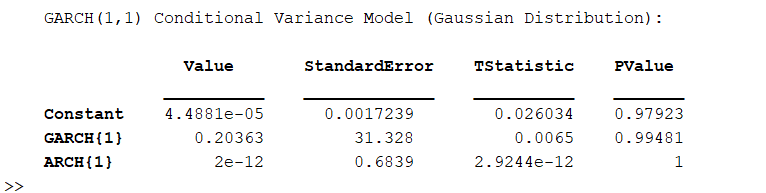


Рисунок 3. Результат работы программного кода

Представим результат нашего предсказания на графике.



Рисунок 4 Модель GARCH(1,1)

**Вывод**

В ходе выполнения лабораторной работы были изучены три математических модели, позволяющие обработать эмпирические данные (например, котировки акций и т.п.) и оценить их будущее значение, такие как *AR(p)*, *ARMA(p,q)*, *GARCH(p,q)*.

Так же были изучены способы оценивания коэффициентов этих моделей.

**Приложение А(задание 1)**

main.m

clc, clearvars, close all, format compact

y= [18.58,18.75,18.83,18.61,18.62,18.50,18.32,18.43,18.37,18.67,18.78,18.72,18.66,18.81,18.98,19.15,19.21,19.14,19.17,18.99,18.83,19.00,19.07,18.93];

n=24;

t=1:n;

[cov,cor,a,Q,a0,h1,gamma,gammaDF,AIK,h2,h\_pred]=find\_characteristics(n,y);

create\_AR\_plot(y,h1)

find\_characteristics

function [c , r,a,Q,a0,h1,gamma,gammaDF,AIK,h2,h\_pred] = find\_characteristics(n, y);

m = 10;

c = zeros(11);

r = zeros(11);

y\_mean=0;

for i=1:n

y\_mean=y\_mean+y(i);

end

y\_mean=y\_mean/n;

%Нахождение автокорелляций и автоковариаций

for i = 1:m+1

sum = 0;

for j = 1:n-(i-1)

sum = sum + (y(j) - y\_mean) \* (y(j+(i-1)) - y\_mean);

end

c(i) = sum / n;

end

for i = 2:m+1

r(i) = c(i) / c(1);

end

r(1)=1;

%Нахождение коэффициентов с помощью критерия Юла-Уокера

A=[r(2),r(3),r(4)];

B=[1,r(2),r(3);r(2),1,r(2);r(3),r(2),1];

a=A/B;

%Статистика Льюнга-Бокса

Q=0;

for i=1:3

Q=Q+r(i)\*r(i)\*n\*(n+2)/(n-i+1);

end

%коэффициент модели а0

a0=(1-a(1)-a(2)-a(3))\*y\_mean;

%создание модели

% h1(1)=a0+a(1)\*a(1)\*a0;

% h1(2)=a0+a(1)\*y(1)+a(2)\*a0;

% h1(3)=a0+a(1)\*y(2)+a(2)\*y(1)+a(3)\*a0;

h1(1)=y(1);

h1(2)=y(2);

h1(3)=y(3);

for i=4:n

h1(i)=a0+a(1)\*y(i-1)+a(2)\*y(i-2)+a(3)\*y(i-3);

% % % h1(i)=a0+a(1)\*h1(i-1)+a(2)\*h1(i-2)+a(3)\*h1(i-3);

end

%тест Дарбина-Уотсонa

for i = 1:n

e(i) = h1(i) - y(i);

end

gamma = 0;

sum\_gamma = 0;

for i = 2:n

gamma = gamma + (e(i) - e(i-1))^2;

sum\_gamma = sum\_gamma + e(i)^2;

end

gamma = gamma / sum\_gamma;

%Тест Дики-Фуллера

h\_mean=0;

for i=1:n

h\_mean=h\_mean+h1(i);

end

h\_mean=h\_mean/n;

for i=1:n

h\_cn(i)=h1(i)-h\_mean;

end

c1=inv(transpose(h\_cn)\*h\_cn);

c1(1,1)

s=sqrt(abs(c1(1,1))/(n-1));

gammaDF=(a(1)-1)/s;

%for i=1:n

% y\_cn(i)=y(i)-y\_mean;

%end

%c=inv(transpose(y\_cn)\*y\_cn);

%s=sqrt(c(1,1)/(n-1));

%gammaDF=(a(1)-1)/s;

%модель для р=2

h2(1)=a0+a(1)\*a0;

h2(2)=a0+a(1)\*h2(1)+a(2)\*a0;

for i=3:n

h2(i)=a0+a(1)\*h2(i-1)+a(2)\*h2(i-2);

end

%Критерий Акаике для р=3

p=3;

sum=0;

for i=1:n

sum=sum+(y(i)-h1(i))^2;

end

AIK(1)=2\*p/n+log(sum/n);

%Критерий Акаике для р=2

p=2;

sum=0;

for i=1:n

sum=sum+(y(i)-h2(i))^2;

end

AIK(2)=2\*p/n+log(sum/n);

%прогноз на 25 день торгов

h\_pred=0;

%for i=1:n

h\_pred=a0+a(1)\*y(24)+a(2)\*y(23)+a(3)\*y(22);

h1(25)=h\_pred;

create\_AR\_plot

function create\_AR\_plot(y, h1)

t = 1:length(y);

figure('Color', 'w')

hold on

plot(t, y, Color='red')

t(25)=25;

plot(t, h1, Color='blue')

hold off

grid on

grid minor

title('Модель AR(3)');

legend('Исходные данные', 'модель AR(3)');

xlabel('t')

ylabel('y')

end

**Приложение В (задание 2)**

main

clc, clearvars, close all, format compact

%y= [49.76, 50.06, 50.37, 50.49, 50.59, 50.44, 50.84, 50.82, 50.39, 50.92, 50.96, 50.74, 50.75, 51.39, 51.40, 51.06, 51.65, 52.23, 52.31, 51.70, 51.61, 51.62, 51.88, 51.51]

y= [18.58,18.75,18.83,18.61,18.62,18.50,18.32,18.43,18.37,18.67,18.78,18.72,18.66,18.81,18.98,19.15,19.21,19.14,19.17,18.99,18.83,19.00,19.07,18.93];

n=24;

t=1:n;

[cov,omega,cor,a,r\_cd,gamma\_bp]=find\_characteristics(n,y);

[a\_AIC,AIC\_c,A1,B1]=AIC(n,y);

[r\_cd1,gamma\_bp1,LB,y\_pred]=Lag\_1(n,omega,a\_AIC,y);

create\_ARMA\_plot(y,y\_pred)

find\_characteristics

function [c ,omega, r,a,r\_cd,gamma\_bp] = find\_characteristics(n, y);

m = 11;

c = zeros(12, 1);

r = zeros(12, 1);

y\_mean=0;

for i=1:n

y\_mean=y\_mean+y(i);

end

y\_mean=y\_mean/n;

%Нахождение автокорелляций и автоковариаций

for i = 1:m+1

sum = 0;

for j = 1:n-(i-1)

sum = sum + (y(j) - y\_mean) \* (y(j+(i-1)) - y\_mean);

end

c(i) = sum / n;

end

for i = 2:m+1

r(i) = c(i) / c(1);

end

r(1)=1;

%Центрируем данные

omega=0;

for i=1:n

omega(i)=y(i)-y\_mean;

end

%Найдем коэффициенты a

B=[c(3),c(4),c(5)];

A=[c(2),c(1),c(2);c(3),c(2),c(1);c(4),c(3),c(2)];

a=B/A;

%модель для центрированных данных

x=0;

x(1)=omega(1);

x(2)=omega(2)-a(1)\*omega(1);

x(3)=omega(3)-a(1)\*omega(2)-a(2)\*omega(1);

for i=4:n

x(i)=omega(i)-a(1)\*omega(i-1)-a(2)\*omega(i-2)-a(3)\*omega(i-3);

end

% автокорреляции для центрированных данных

x\_mean=0;

for i=1:n

x\_mean=x\_mean+x(i);

end

x\_mean=x\_mean/n;

%Нахождение автокорелляций и автоковариаций

for i = 1:m+1

sum = 0;

for j = 1:n-(i-1)

sum = sum + (x(j) - x\_mean) \* (x(j+(i-1)) - x\_mean);

end

c\_cd(i) = sum / n;

end

for i = 2:m+1

r\_cd(i) = c\_cd(i) / c\_cd(1);

end

r\_cd(1)=1;

%проверим статистику Бокса-Пирса

gamma\_bp=0;

for k=1:m-1

r\_ch=0;

r\_zn=0;

for j=1:n

r\_zn=r\_zn+(x(j))^2;

end

for i=k+1:n

r\_ch=r\_ch+(x(i))\*(x(i-k));

end

gamma\_bp=gamma\_bp+(r\_ch/r\_zn)^2;

end

gamma\_bp=gamma\_bp\*n;

chi2inv(0.95,8)

AIC

function [a\_AIC,AIC\_c,A1,B1] = AIC(n, y);

%найдем все коэффициенты а

[cov,omega]=find\_characteristics(n,y);

B1=0;

for i=1:10

for j=1:i

for k=1:i

if k-j>0

A1(j,k,i)=cov(k-j);

else

if k-j<0

A1(j,k,i)=cov(j-k+2);

else

A1(j,k,i)=cov(2);

end

end

end

end

B1(i)=cov(i+2);

%A1=A1(:,:,i);

%a1(i)=squeeze(A1([1:i],[1:i],i))/B1([1:i]);

end

a1= B1([1:1])/squeeze(A1([1:1],[1:1],1));

a2= B1([1:2])/squeeze(A1([1:2],[1:2],2));

a3= B1([1:3])/squeeze(A1([1:3],[1:3],3));

a4= B1([1:4])/squeeze(A1([1:4],[1:4],4));

a5= B1([1:5])/squeeze(A1([1:5],[1:5],5));

a6= B1([1:6])/squeeze(A1([1:6],[1:6],6));

a7= B1([1:7])/squeeze(A1([1:7],[1:7],7));

a8= B1([1:8])/squeeze(A1([1:8],[1:8],8));

a9= B1([1:9])/squeeze(A1([1:9],[1:9],9));

a10= B1([1:10])/squeeze(A1([1:10],[1:10],10));

a=0;

a1(2:10)=0;

a2(3:10)=0;

a3(4:10)=0;

a4(5:10)=0;

a5(6:10)=0;

a6(7:10)=0;

a7(8:10)=0;

a8(9:10)=0;

a9(10:10)=0;

a1=transpose(a1);

a2=transpose(a2);

a3=transpose(a3);

a4=transpose(a4);

a5=transpose(a5);

a6=transpose(a6);

a7=transpose(a7);

a8=transpose(a8);

a9=transpose(a9);

a10=transpose(a10);

a\_AIC=horzcat(a1,a2,a3,a4,a5,a6,a7,a8,a9,a10);

a\_AIC=transpose(a\_AIC);

%Нахождение самих коэффициентов AIK

AIC\_c=0;

sum=0;

for p=1:10

SUM=0;

for i=1:n

sum=y(i);

for j=1:p

if i-j>0

sum=sum-a\_AIC(p,j)\*y(i-j);

end

end

SUM=SUM+sum^2;

sum=0;

end

AIC\_c(p)=2\*(p+1)/n+log(SUM/n);

end

Lag\_1

function[r\_cd1,gamma\_bp1,LB,y\_pred]=Lag\_1(n,omega,a\_AIC,y)

m=11;

%модель для центрированных данных

x=0;

x(1)=omega(1);

for i=2:n

x(i)=omega(i)-a\_AIC(1,1)\*omega(i-1);

end

%%%%%%%%%%%% автокорреляции для центрированных данных

x\_mean=0;

for i=1:n

x\_mean=x\_mean+x(i);

end

x\_mean=x\_mean/n;

%Нахождение автокорелляций и автоковариаций

for i = 1:m+1

sum = 0;

for j = 1:n-(i-1)

sum = sum + (x(j) - x\_mean) \* (x(j+(i-1)) - x\_mean);

end

c\_cd(i) = sum / n;

end

for i = 2:m+1

r\_cd1(i) = c\_cd(i) / c\_cd(1);

end

r\_cd1(1)=1;

% r1<1/2 поэтому модель ARMA(p,1) статистически значима

%проверим статистику Бокса-Пирса

gamma\_bp1=0;

for k=1:m-1

r\_ch=0;

r\_zn=0;

for j=1:n

r\_zn=r\_zn+(x(j))^2;

end

for i=k+1:n

r\_ch=r\_ch+(x(i))\*(x(i-k));

end

gamma\_bp1=gamma\_bp1+(r\_ch/r\_zn)^2;

end

gamma\_bp1=gamma\_bp1\*n;

%по статистике Бокса-Пирса ряд стационарен, но проверим еще и

%статистику Льюнга-Бокса

sum=0;

for i=1:m-1

sum=sum+(r\_cd1(i+1)^2)/(n-i);

end

LB=n\*(n+2)\*sum

%По статистике Льюнга\_Бокса ряд тоже стационарен

% Построим прогноз на 25 день торгов

teta=(-1+sqrt(1-4\*(r\_cd1(2))^2))/(2\*r\_cd1(2))

y\_pred=0;

for i=2:n+1

y\_pred(i) = a\_AIC(1,1)\*x(i-1)+mean(y)-teta;

end

y\_pred(1)=y(1);

create\_ARMA\_plot

function create\_ARMA\_plot(y, y\_pred)

t = 1:length(y);

figure('Color', 'w')

hold on

plot(t, y, Color='red')

t(25)=25;

plot(t, y\_pred, Color='blue')

hold off

grid on

grid minor

title('Модель ARMA(1,1)');

legend('Исходные данные', 'модель ARMA(1,1)');

xlabel('t')

ylabel('y')

end

**Приложение С (задание 3)**

main

clc, clearvars, close all, format compact

y= [18.58,18.75,18.83,18.61,18.62,18.50,18.32,18.43,18.37,18.67,18.78,18.72,18.66,18.81,18.98,19.15,19.21,19.14,19.17,18.99,18.83,19.00,19.07,18.93];

n=24;

t=1:n;

[u,E,D,cor\_us,cor\_u,gamma\_u,gamma\_us,u\_sq,u\_sigma\_sq,sigma\_sq]=find\_characteristics(n,y);

[y\_pred]=predict\_GARCH(y,sigma\_sq,u);

predict\_GARCH\_plot(y,y\_pred);

model\_GARCH(y,u);

find\_characteristics

function [u,E,D,r,r\_u,gamma\_u,gamma\_us,u\_sq,u\_sigma\_sq,sigma\_sq]=find\_characteristics(n,y)

%%%Переход к относительным приращениям

u=0;

for i=2:n

u(i-1)=(y(i)-y(i-1))/y(i-1);

end

%%%Найдем выборочные мат ожидание и дисперсию

E(1)=mean(u(19:23));

E(2)=mean(u(14:23));

E(3)=mean(u(9:23));

D(1)=0;D(2)=0;D(3)=0;

for i=1:5

D(1)=D(1)+(u(i+18)-E(1))^2;

end

D(1)=D(1)/5;

for i=1:10

D(2)=D(2)+(u(i+13)-E(2))^2;

end

D(2)=D(2)/10;

for i=1:15

D(3)=D(3)+(u(i+8)-E(3))^2;

end

D(3)=D(3)/15;

%%%%Пункт 3

%%%%model = garch('garchlags',1,'archlags',1); [estM1,H,logL] = estimate(model,u)

%%Пункт четвертый

%Найдем ряды сигма квадрат и ...

N=8;

sigma\_sq(1:N)=0;

k=15;

for i=16:length(u)

for j=1:k

sigma\_sq(i-k)=sigma\_sq(i-k)+(u(i-j)-mean(u(i-k:i)))^2;

end

sigma\_sq(i-k)=sigma\_sq(i-k)/(k-1);

end

u\_sq(1:N)=0;

for i=k+1:length(u)

u\_sq(i-k)=u(i)^2;

end

u\_sigma\_sq(1:N)=0;

for i=1:N

u\_sigma\_sq(i)=u\_sq(i)/sigma\_sq(i);

end

%%%Найдем первых пять автокорреляций

m=5;

for i = 1:m+1

sum = 0;

for j = 1:N-(i-1)

sum = sum + (u\_sigma\_sq(j) - mean(u\_sigma\_sq)) \* (u\_sigma\_sq(j+(i-1)) - mean(u\_sigma\_sq));

end

c(i) = sum / N;

end

for i = 2:m+1

r(i) = c(i) / c(1);

end

r(1)=1;

%%%%%% для u\_sq

for i = 1:m+1

sum = 0;

for j = 1:N-(i-1)

sum = sum + (u\_sq(j) - mean(u\_sq)) \* (u\_sq(j+(i-1)) - mean(u\_sq));

end

c(i) = sum / N;

end

for i = 2:m+1

r\_u(i) = c(i) / c(1);

end

r\_u(1)=1;

%%%проверим статистику Льюнга-Бокса

gamma\_u=0;

sum=0;

for i=1:m

sum=sum+r\_u(i)/(N-i);

end

gamma\_u=(N+2)\*N\*sum;

gamma\_us=0;

sum=0;

for i=1:m

sum=sum+r(i)/(N-i);

end

gamma\_us=(N+2)\*N\*sum;

model\_GARCH

function frcst = model\_GARCH(y,u)

model = garch('garchlags', 1, 'archlags', 1);

[estM1, H, logL, info] = estimate(model, transpose(u), 'Display', {'params'});

frcst = forecast(estM1, 1, transpose(u))\*y(length(y))\*normrnd(0,1)+y(length(y));

predict\_GARCH\_plot(y, frcst);

end

predict\_GARCH

function y\_pred = predict\_GARCH(y, sigma\_sq, u)

model = garch('garchlags', 1, 'archlags', 1);

[estM1, H, logL, info] = estimate(model, transpose(u), 'Display', {'params'});

sigma\_n = sum(sigma\_sq/length(sigma\_sq)) \* info.X(1) + u(23)^2 \* info.X(3) + sigma\_sq(1) \* info.X(2)

new\_u\_n = sigma\_n \* normrnd(0, 1);

y\_pred = new\_u\_n \* y(length(y)) + y(length(y));

end

predict\_GARCH\_plot

function predict\_GARCH\_plot(y, y\_pred)

t = 1:length(y);

figure('Color', 'w')

hold on

plot(t, y, Color='red')

plot(25, y\_pred, 'bo')

hold off

grid on

grid minor

title('Предсказание модели');

legend('Исходные данные', 'Предсказанное значение GARCH(1,1) модели');

xlabel('t')

ylabel('y')

end