Часть 1. Построение модели для волатильности цен акций компаний

В рамках первой части домашнего задание №2 нами было построено модели для волатильности цен акций 4 компаний РФ, осуществляющих торговлю потребительскими товарами – Retail:

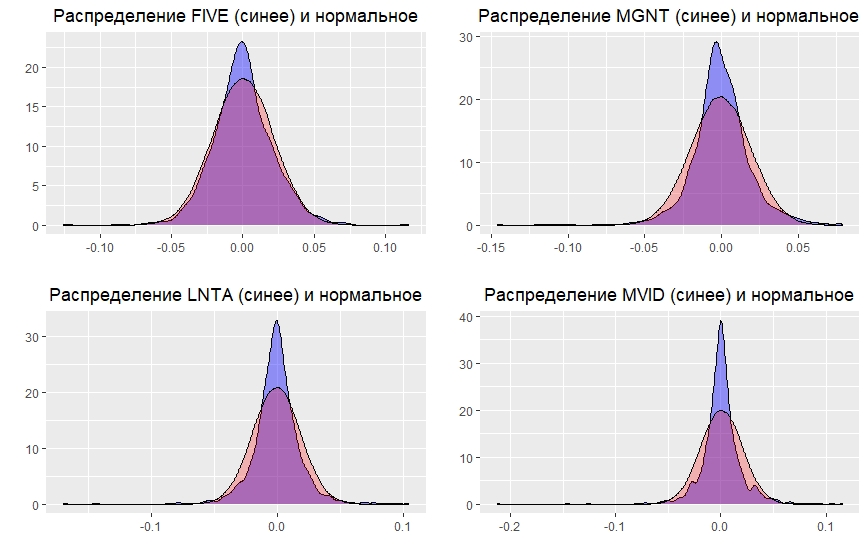
* M.VIDEO
* MAGNIT
* LENTA
* X5 RETAIL GROUP

В качестве анализируемого промежутка времени нами был выбран интервал с 01.09.2014 и по настоящий момент времени. Исключением является лишь модель для компании X5 Retail Group, так как для данного эмитента торговля депозитарными расписками на ММВБ началась лишь с 02.02.2018 г.

Для всех выбранных компаний нами был проведен анализ с целью выбора наилучшей спецификации модели. Тем не менее в последующем тексте нами было принято решение в наибольшей подробности описать весь процесс подбора модели лишь для одного эмитента – M.Video, ввиду потенциально слишком большого объема текста.

1.1 Загрузка и обработка данных, сравнительный анализ доходностей

Для начала проведем первичную загрузку данных, а также проведем конвертацию дат в формат дат ('%Y-%m-%d').

Теперь перейдем к графическому анализу распределений доходностей выбранных эмитентов, а также сравнению их с нормальными. Для этого при помощи команды *qplot* сравним между собой плотности распределений выбранных компаний – ‘FIVE’, ‘MGNT’, ‘LNTA’, ‘MVID’. Полученные графики отражены ниже [Рис. 1]:

**Рис. 1: Плотности распределений выбранных эмитентов - сравнение с нормальными**

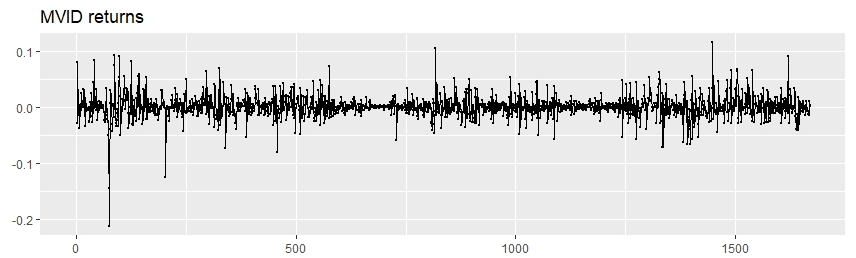
Как видно из полученных иллюстраций, ни одно из полученных распределений доходностей не является нормальным. Тем не менее наиболее приближенно к нему распределение доходностей компании X5 Retail Group, а наименее – распределения доходностей компаний LENTA и M.VIDEO. Так, визуально коэффициенты эксцесса – мера остроты пика у этих компаний велики. Это в свою очередь наталкивает на мысль касательно того, что модификации моделей авторегрессии — скользящего среднего (ARMA), а также обобщенной авторегрессионной гетероскедастичности (GARCH) нужно будет строить в соответствие с предпосылкой об отсутствии нормально распределенных ошибок.

1.2 Построение моделей для компании M.VIDEO

Итак, в наиболее подробном виде нами проведен анализ доходностей и волатильности, а также построение моделей у компании M.VIDEO.

1.2.1 Проверка на стационарность

При построении модели для доходностей рассматриваемого эмитента нами будет выбрана модель интегрированной авторегрессии – скользящего среднего (ARIMA). При выборы спецификации и параметров данной модели мы руководствовались методологией Бокса-Дженкинса (Box, Jenkins). На первоначальном этапе нужно понять, является ли ряд стационарным, необходимо ли взятие разностей некоторого порядка исходного временного ряда. Итак, на первый взгляд, исходя из [Рис. 2], у доходностей данного эмитента отсутствует какой-либо тренд:

Тем не менее для формального обоснования необходимо использовать и количественную методологию. Наиболее широко распространенной методикой в эконометрике для анализа временных рядов для проверки на стационарность является Расширенный тест Дики – Фуллера (ADF-Тест). Как известно, сущность данного теста заключается в проверке нулевой гипотезы о наличии единичного корня. Кроме того, в данном случае в тестовые регрессии происходит включение лагов первых разностей, так как процесс может быть авторегрессией не первого, а более высокого порядка.

**Рис. 2: Доходности компании MVIDEO**

Таким образом, ADF тест применяется к следующей модели:

,

Где - константа, – коэффициент на временной тренд и *p* – порядок лага авторегрессии. (уточнить, нужно ли два первых слагаемых)

Затем происходит тест на наличие единичного корня:

Происходит расчет тестовой статистики в соответствии с формулой, а также сравнение с критическими значениями:

В нащем случае для компании MVIDEO были получены следующая тестовая статистика и уровень значимости:

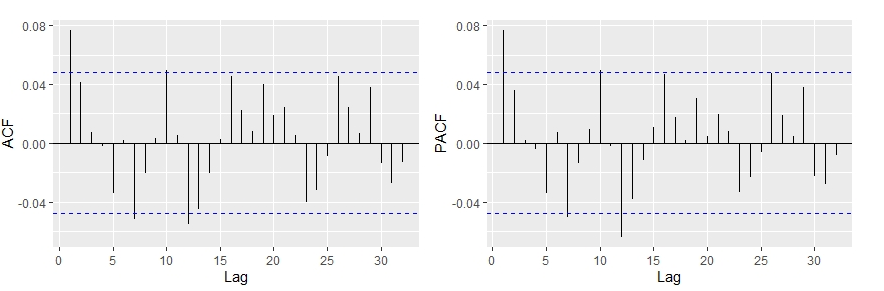
DF = -26.744, p-value = 2.2e-16

Таким образом, нулевая нулевая гипотеза отвергается на любом уровне значимости, что говорит нам об отсутствии единичного корня. А следовательно, можно сделать вывод о том, что процесс является стационарным, а временные ряды подчиняются модели ARIMA следующего порядка: ARIMA (p, 0, q) ~ ARMA (p, q).

1.2.2 Выбор спецификации модели ARIMA

После того, как мы доказали, что временной ряд доходностей рассматриваемого эмитента является стационарным, перейдем непосредственно к выбору спецификации модели ARMA.

Графики автокорреляции (ACF) и частичной автокорреляции (PACF) можно использовать для получения значений p и q для подачи в модель ARMA. Исходя из полученных графиков [Рис. 3], можно сделать вывод о том, что потенциально значимыми лагами в обоих случаях являются лишь первые, остальные лишь незначительно превышают критический уровень. Следовательно, вполне логично было бы предложить модификацию модели ARIMA ~ ARIMA (1, 0, 1).

Тем не менее, формализовано спецификация модели выбирается в соответствии с значениями информационных критериев - Акаике (AIC) и Шварца (BIC). Они показывают меру относительного качества моделей, учитывая степень «подгонки» модели под данные с штрафом на используемое количество оцениваемых параметров. Таким образом, их количественные показатели основаны на компромиссе между точностью и сложностью модели.

**Рис. 3. ACF и PACF для временных рядов**

Исходные расчетные формулы данные критериев выглядят следующим образом:

,

Где *l –* значение логарифмической функции правдоподобия построенной модели, *k* – число использованных параметров, *n* – объем выборки.

Итак, рассматриваются комбинации моделей ARIMA с различными коэффициентами AR и MA. В свою очередь выбирается набор с самыми низкими показателями информационных критериев Акаике (AIC) и Шварца (BIC).

В результате при помощи команды auto.arima нами была найдена наиболее оптимальная спецификация модели ARIMA ~ ARIMA (3, 0, 1) со следующими значениями коэффициентов, стандартных отклонений и значений информационных критериев:

*Coefficients:*

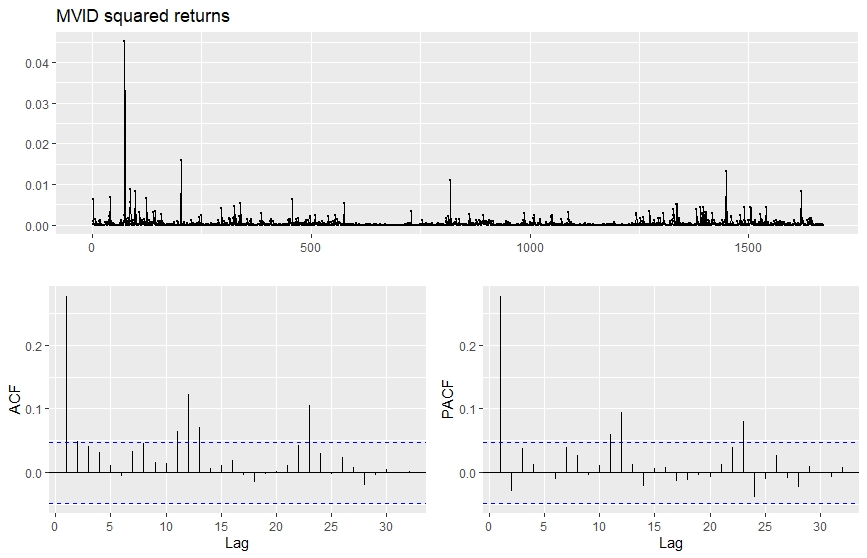
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | *AR1* | *AR2* | *AR3* | *MA1* |
|  | *-0.8048* | *0.1046* | *0.0201* | *0.8811* |
| *s.e.* | *0.1374* | *0.0328* | *0.0275* | *0.1348* |

*AIC=-8353.42*

*BIC=-8326.32*

Тем не менее можно заметить, что коэффициент при *AR3* не является значимым на уровне значимости в 5%, так как если взять разность между коэффициентом и двумя стандартными отклонениями, 0 войдет в доверительный интервал (0.0201 – 2\*0.0275 < 0). Именно поэтому мы посчитали спецификацию модели ARIMA ~ ARIMA (2, 0, 1) наиболее оптимальной.

В заключении проверим визуально наличие условной гетероскедастичности. Для этого изобразим графически квадрат доходностей эмитента M.VIDEO, так как по сути это является оценкой дисперсии случайной ошибки [Рис. 4] (.



**Рис. 4: Квадрат доходностей для MVIDEO; ACF и PACF**

Очевидно, что графики автокорреляции (ACF) и частичной автокорреляции (PACF) демонстрируют наличие гетероскедастичности. Именно поэтому необходимо перейти непосредственно к построению ARIMA-GARCH-модели.

1.2.3 Выбор спецификации модели GARCH

Изначально при помощи пакета *rugarch* оценим самую адекватную модель, ARIMA (2, 0, 1)-GARCH(1, 1) модель, учитывая, что математическое ожидание (µ) равно нулю. Таким образом, в соответствии с предыдущей оценкой, уравнение доходности имеет вид:

*,*

А уравнение для волатильности, , имеет вид:

,

Построим модель ARIMA(2, 0, 1)-GARCH(1,1) и посмотрим на результаты оценивания:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | Std. Error | t value | Pr(>|t|) |
| ***ar1*** | -0.806386 | 0.220522 | -3.65671 | 0.000255 |
| ***ar2*** | -0.021513 | 0.043705 | -0.49223 | 0.622559 |
| ***ma1*** | 0.723081 | 0.219181 | 3.29901 | 0.000970 |
| ***omega*** | 0.000011 | 0.000002 | 6.07240 | 0.000000 |
| ***alpha1*** | 0.184577 | 0.017162 | 10.75504 | 0.000000 |
| ***beta1*** | 0.814423 | 0.010029 | 81.20770 | 0.000000 |

Как можно заметить из результатов оценивания, все переменные, кроме ar2, являются значимыми на любом адекватном уровне значимости. Кроме того, были получены следующие показатели лог функции правдоподобия и информационных критериев:

**LogLikelihood : 4370.576**

**Akaike -5.2270**

**Bayes -5.2076**

Теперь перейдем непосредственно к проведению тестов для оценки качества модели.

**1.2.4 Тесты для полученной ARIMA-GARCH модели**

**Weighted Ljung-Box Test on Standardized Residuals**

Данный тест используется для того, чтобы определить, имеет ли место автокорреляции между стандартизированными остатками. Рассматриваются следующие гипотезы:

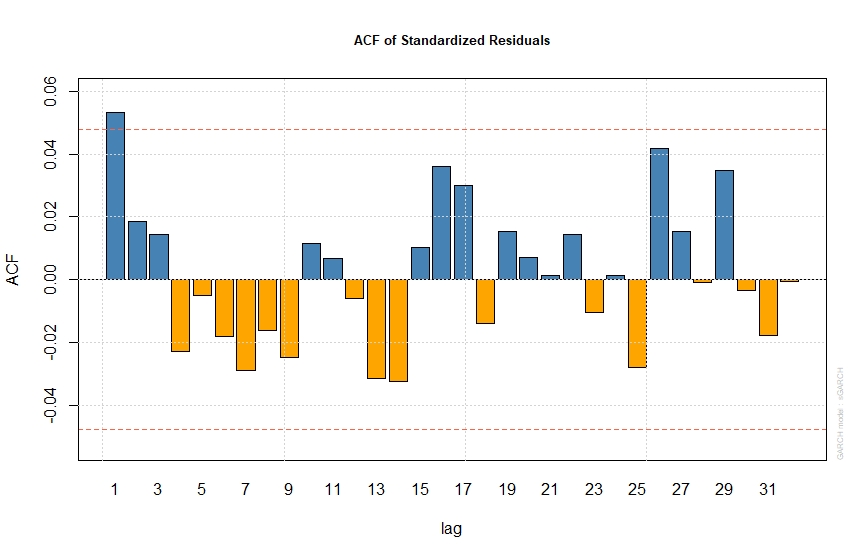
А тестовая статистика вычисляется в соответствии во следующей формулой:

, где

n – размер выборки, - автокорреляция при лаге k, h – число тестируемых лагов.

Получаем следующие тестовые статистики и уровни значимости:

|  |  |  |
| --- | --- | --- |
|  | statistic | p-value |
| Lag[1] | 4.758 | 0.0291599 |
| Lag[2\*(p+q)+(p+q)-1][8] | 6.704 | 0.0007323 |
| Lag[4\*(p+q)+(p+q)-1][14] | 8.617 | 0.2733576 |

Таким образом, для всех тестируемых лагов нулевая гипотеза отвергается в пользу альтернативной, а значит остатки в модели коррелированы между собой. Следовательно, построенная модель среднего недостаточно хороша. Аналогичные результаты продемонстрированы на [Рис. 5] – для первого лага явно видна значимая автокорреляция:

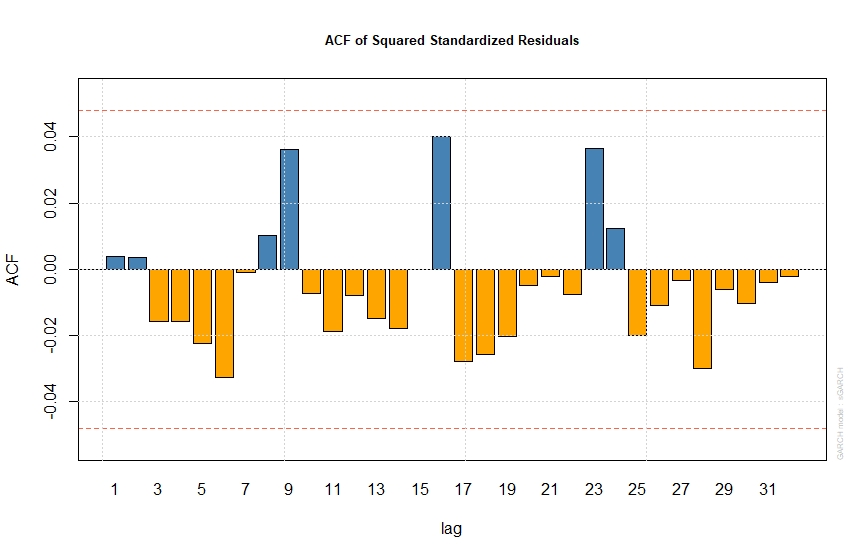
**Рис. 5: Автокорреляция остатков**

**Weighted Ljung-Box Test on Standardized Squared Residuals**

Тот же самый тест проводится для квадратов остатков, полученных из модели. Если модель достаточно хорошо описывает гетескедастичность, то квадраты остатков не должны быть коррелированы между собой, а представлять собой лишь случайные шоки:

Получаем следующие тестовые статистики и уровни значимости:

|  |  |  |
| --- | --- | --- |
|  | statistic | p-value |
| Lag[1] | 0.02376 | 0.8775 |
| Lag[2\*(p+q)+(p+q)-1][5] | 0.61933 | 0.9378 |
| Lag[4\*(p+q)+(p+q)-1][9] | 2.17784 | 0.8827 |

 Итак, согласно полученным тестовым статистикам и уровням значимости, для всех лагов нет основания отвергать основную гипотезу, а значит квадраты остатков не связаны между собой – модель волатильности построена корректно. Такие же результаты можно увидеть и графически [Рис. 6] – отсутствует значимая автокорреляция для всех лагов:

**Рис. 6: Автокорреляция квадратов остатков**

**Sign Bias**

Данный тест позволяет узнать, реагирует ли волатильность в различной мере для отрицательных и положительных шоков. Если получается значимая статистика, то в этом случае имеет смысл построить модель EGARCH. Имеем следующие гипотезы:

Получаем следующие тестовые статистики и уровни значимости:

|  |  |  |
| --- | --- | --- |
|  | statistic | prob sig |
| Sign Bias | 0.09209 | 0.9266 |
| Negative Sign Bias | 0.11580 | 0.9078 |
| Positive Sign Bias | 0.13611 | 0.8918 |
| Joint Effect | 0.03265 | 0.9984 |

Таким образом, во всех случаях мы получаем незначимую тестовую статистику, а значит нет оснований отвергать основную гипотезу об отсутствии асимметрии. Следовательно, необходимости строить модификацию для нашей модели в виде модели EGARCH нет.

**1.2.5 Модификация ARIMA-GARCH модели**

Исходя из полученного графика плотности распределения доходностей компании M.VIDEO [Рис. 1], очевидным становится то, что распределение сильно отличается от нормального. Следовательно полученную модель необходимо модифицировать с целью получения большего числа наблюдения «в хвостах». Нами было решено построить новую модель и ввести предпосылку о том, что остатки распределены в соответствии с распределением Стьюдента:

Построим новую модель ARIMA(2, 0, 1)-GARCH(1,1) и посмотрим на результаты оценивания:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | Std. Error | t value | Pr(>|t|) |
| ***ar1*** | -0.849403 | 0.179547 | -4.7308 | 0.000002 |
| ***ar2*** | -0.032167 | 0.029643 | -1.0852 | 0.277852 |
| ***ma1*** | 0.788020 | 0.179064 | 4.4008 | 0.000011 |
| ***omega*** | 0.000017 | 0.000007 | 2.3598 | 0.018287 |
| ***alpha1*** | 0.280179 | 0.041069 | 6.8221 | 0.000000 |
| ***beta1*** | 0.718821 | 0.055177 | 13.0276 | 0.000000 |
| ***shape*** | 3.277136 | 0.206636 | 15.8595 | 0.000000 |

Как и в предыдущем случае все переменные, кроме ar2, являются значимыми на любом адекватном уровне значимости. Estimate ~ 3.27 у переменной shape, а также p-value -> 0 означают то, что распределение шоков не является нормальным, и оценивание и использованием предпосылки о распределении остатков по Стьюденту действительно имело смысл. Кроме того, были получены следующие показатели лог функции правдоподобия и информационных критериев:

**LogLikelihood : 4605.822**

**Akaike -5.5076**

**Bayes -5.4849**

Таким образом, при сопоставлении с информационными критериями предыдущей модели можно сделать вывод, что новая модель лучше, так как их величины меньше по абсолютному значению.

Теперь перейдем непосредственно к проведению тестов для оценки качества модели.

**Weighted Ljung-Box Test on Standardized Residuals**

Получены следующие показатели:

|  |  |  |
| --- | --- | --- |
|  | statistic | p-value |
| Lag[1] | 2.160 | 0.1417 |
| Lag[2\*(p+q)+(p+q)-1][8] | 4.715 | 0.3482 |
| Lag[4\*(p+q)+(p+q)-1][14] | 6.929 | 0.5743 |

Можно сделать вывод, что на уровне значимости в 10% нет основания отвергать основную гипотезу об отсутствии автокорреляции остатков в пользу альтернативной. А значит, модель стала лучше.

**Weighted Ljung-Box Test on Standardized Squared Residuals**

|  |  |  |
| --- | --- | --- |
|  | statistic | p-value |
| Lag[1] | 0.3852 | 0.5348 |
| Lag[2\*(p+q)+(p+q)-1][5] | 1.3842 | 0.7683 |
| Lag[4\*(p+q)+(p+q)-1][9] | 3.0353 | 0.7531 |

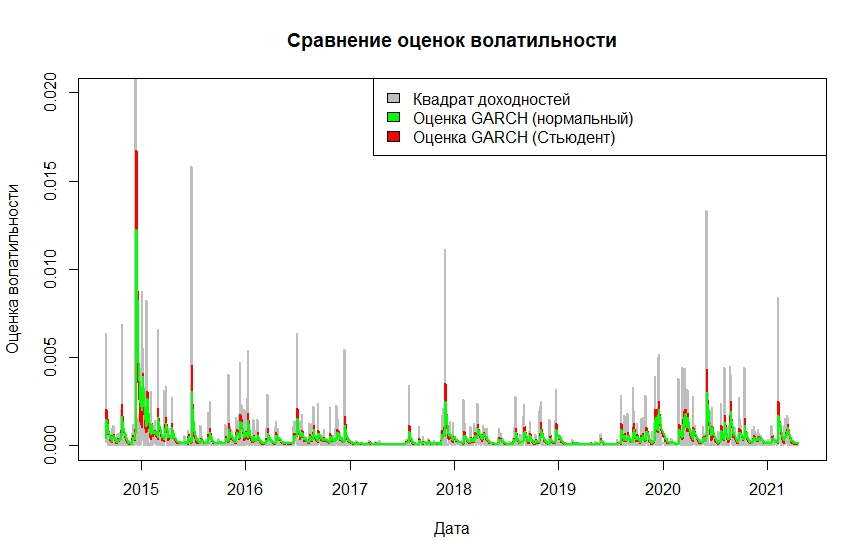
Как и в предыдущем случае, в соответствии в полученными тестовыми статистиками нет оснований отвергать основную гипотезу о наличии автокорреляции между квадратами остатков, модель волатильности построена корректно.

**Sign Bias Test**

|  |  |  |
| --- | --- | --- |
|  | t-value | prob sig |
| Sign Bias | 0.2242 | 0.8226 |
| Negative Sign Bias | 0.9942 | 0.3202 |
| Positive Sign Bias | 0.6477 | 0.5173 |
| Joint Effect | 1.4229 | 0.7002 |

Полученные тестовые статистики позволяют сделать вывод об отсутствии основания для отвержения основной гипотезы. А значит в новой модели отсутствует асимметрия волатильности.

**1.2.6 Сравнение оценок волатильности**

Посмотрим графически, какая из моделей лучше предсказывает всплески волатильности. Для этого на графике построим квадрат доходностей (несмещенная оценка волатильности), а также оценки волатильности, полученные в моделях ARIMA(2, 0, 1)-GARCH(1, 1) с нормальным распределением ошибок и распределением ошибок по Стьюденту [Рис. 7]:

**Рис. 7: Сравнение оценок волатильности**

Несложно заметить, что оценка модели GARCH с распределением шоков по Стьюеднту лучше предсказывает всплески волатильности, потому что распределение имеет более тяжелые «хвосты».

Таким образом, построенная новая модель, где берется предпосылка о распределении шоков по Стьюденту лучше в соответствии со следующими критериями:

1. Информационные критерии у новой модели имеют более низкие показатели по абсолютному значению
2. Weighted Ljung-Box Test on Standardized Residuals показал отсутствие автокорреляции стандартизированных остатков, распределение шоков не является нормальным
3. Визуально можно убедиться в том, что новая модель лучше предсказывает всплески волатильности [Рис. 7]

**1.2.7 Прогноз волатильности**

 В соответствии с построенной моделью ARIMA(2, 0, 1)-GARCH(1,1) построим прогноз волатильности на 10 дней и сравним с полученной оценкой за последний квартал [Рис. 8]:

**Рис. 8: Оценка волатильности и построение прогноза на 10 дней**

Таким образом, построенная нами модель четко уловила то, что за первичным ростом волатильности идет еще больший ее рост.

ДАЛЕЕ – ОБОБЩЕННО ДЛЯ ТРЕХ ЭМИТЕНТОВ