Министерство образования и науки Российской Федерации

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИМЕНИ Н. Г. ЧЕРНЫШЕВСКОГО»

Кафедра теории функций и
стохастического анализа

МОДЕЛИРОВАНИЕ ФИНАНСОВЫХ ПРОЦЕССОВ С ЗАВИСИМОЙ ВАРИАЦИЕЙ

АВТОРЕФЕРАТ БАКАЛАВРСКОЙ РАБОТЫ

Студентки 4 курса 412 группы
направления 01.03.02 — Прикладная математика и информатика
механико-математического факультета
Худошиной Анастасии Олеговны

Научный руководитель
доцент,к.ф.-м.н ______________ Н. Ю. Агафонова

Заведующий кафедрой
д. ф.-м. н. ______________ С. П. Сидоров

Саратов 2017

ВВЕДЕНИЕ

Работа посвящена моделированию финансовых процессов с зависимой вариацией. Вопросы, касающиеся прогнозирования и моделирования изменчивости различных показателей на финансовых рынках, считаются очень актуальными на данный момент. Традиционные модели временных рядов Autoregressive moving-average (ARMA) не всегда учитывает характеристики, которыми обладают финансовые временные ряды. Поэтому для прогнозирования изменение цен лучше использовать модели AutoRegressive Conditional Heteroscedasticity (ARCH) и Generalized ARCH (GARCH). Впервые американский экономист Роберт Энгл в 1982 году разработал ARCH модели, а после него Боллерслев 1986 году предложил обобщение этих моделей (GARCH).

Предметом исследования являются модели процесса ARCH и GARCH.

Цель работы: рассмотреть временные ряды, изучить понятие стационарности, описать модели временных рядов AR(p) и MA(q), рассмотреть модели с условной гетероскедастичностью, разработать реализацию процесса ARCH(1), найти прогнозные значения, проанализировать полученные оценки характеристик этой модели. Основные результаты работы носят теоретический характер и могут найти применение в анализе временных рядов.

Работа состоит из введения, трех разделов, заключения и списка использованных источников.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ.

Во введении обоснована актуальность темы выпускной квалификационной работы, определена ее цель, описана структура работы, при этом введены основные определения.

Определение. Исходы $\omega_1, \omega_2, ..., \omega_N$ будем называть элементарными событиями, а их совокупность $\Omega = \{\omega_1, ..., \omega_N\}$ пространством элементарных событий.

Определение. Набор чисел $\{P_\xi(x_1), ..., P_\xi(x_m)\}$ называется распределением вероятностей случайной величины ξ.

Определение. Пусть $x \in R^1$. Функция $F_\xi(x) = P\{\omega : \xi(\omega) < x\}$ называется функцией распределения случайной величины ξ.

Теорема Гаусса-Маркова. В предположении модели 1-3: 1. $Y_t = a + bX_t + \varepsilon_t, \ t = 1, ...n;$
2. X_t-детерминиованная величина;
3. $E\varepsilon_t = 0, \ E(\varepsilon_t^2) = V(\varepsilon_t) = \sigma^2, \ E(\varepsilon_t, \varepsilon_s) = 0, \ t \neq s;$
оценки МНК, имеют наименьшую дисперсию в классе всех линейных несмещенных оценок.

Определение. Функция

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} dt,$$

называется нормальным или гауссовским распределением вероятностей на числовой прямой с нормальной или гауссовской плотностью

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2}$$

.

Определение. Пусть X_t - значение случайного процесса в момент времени t (t может быть вещественным, если процесс непрерывный, или целым, если процесс дискретный). Если X_t имеет среднее значение μ_t и дисперсию σ_t^2, то автокорреляция X_t определяется следующим образом: $R(t, s) = \frac{\mathrm{E}[(X_t - \mu_t)(X_s - \mu_s)]}{\sigma_t \sigma_s}$.

Определение. Коэффициентом корреляции $\rho(\xi, \eta)$ случайных величин

$\xi,\ \eta$, дисперсии которых существуют и отличны от нуля, называется число:

$$\rho(\xi,\eta)=\frac{E((\xi-E\xi)(\eta-E\eta))}{\sqrt{D\xi}\sqrt{D\eta}}$$

.

Определение. Процесс ε_t, удовлетворяющий условиям теоремы Гаусса-Маркова означает, что $E(\varepsilon_t)=0$; $Var(\varepsilon_t)=\sigma^2$; $Cov(\varepsilon_i,\varepsilon_j)=0$ при $i\neq j$.

Определение. Процесс слабо стационарен и случайные величины ε_i распределены в совокупности нормально, то будем называть его гауссовым белым шумом и обозначим $\varepsilon_i\sim WN(0,\sigma^2)$.

Определение. Случайной функцией $\xi(t,\omega),\ t\in T$, называют измеримое отображение $\xi:\Omega\to R^n$ пространства элементарных событий Ω в R^n, зависящее от параметра t. Если $T=[a;b]-$ отрезок числовой оси, а параметр t интерпретируется как время, то случайную функцию называют случайным процессом.

Определение. Процесс случайного блуждания это процесс, который задается следующим образом: $X_t=X_{t-1}+\varepsilon_t$, где ε_t - белый шум.

Первая глава содержит основные определения и понятия, используемые в анализе временных рядов.

Ряды являются распространенной формой описания данных, так как позволяют рассматривать историю изменения интересующих значений, и мы с легкостью можем судить об отклонениях исследуемой величины.

Под временным рядом понимают выборку из последовательности случайных величин X_t, где t принимает целочисленные значения от 1 до T. Дискретным случайным или стохастическим процессом называется совокупность случайных величин $\{X_t,t\in[1,T]\}$.

Случайным процессом назовем любое отображение $X:\Omega\times T\to R$, такое что $\forall t_0\in T\ X(\omega,t_0)$ удовлетворяет определению случайной величины. Реализацией случайного процесса называется некоторая последовательность значений первой, второй и так далее случайных величин, при фиксированном случае. Иногда рассматриваем последовательность $\{X_t,t\in[1,T]\}$ как подпоследовательность бесконечной последовательности $\{X_t,t=0,\pm1,..\}$ и именно эту последовательность назовем стохастическим процессом, порождающим наблюденные данные. Если t пробегает непрерывный отрезок времени, то $X(\omega, t)$ называют случайным процессом с непрерывным временем.

Различают два вида стационарности: строгая и слабая.

Случайный процесс называют строго стационарным, если сдвиг во времени не меняет ни одну из функций плотности распределения, т.е. если ко всем моментам времени прибавить некоторую (целочисленную) величину, то сама функция плотности не изменится, $f_n(x_{t_1}, ...x_{t_n}) = f_n(x_{t_1+\triangle}, ..., x_{t_n+\triangle})$ для всех n, моментов времени $t_1, ...t_n$ целочисленных $\triangle$.

Если случайный процесс строго стационарен, то у него:
1)математическое ожидание не зависит от времени;
2) дисперсия не зависит от времени;
3) автоковариационная и автокорреляционные функции: а) зависят только от сдвига, от разности моментов времени; б) являются четными функциями.
Случайный процесс называют слабо стационарным, если математическое ожидание и дисперсия существуют и не зависят от времени, а автокорреляционная (автоковариационная) функция зависит только от разности значений $(t_1 - t_2)$.

Декомпозиция или теорема Вольда.

Вольд показал, что чисто недетерминированный стационарный в широком смысле случайный процесс может быть представлен в следующем виде:

$$X_t - \mu = \sum_{\tau=0}^{\infty} \psi_\tau \varepsilon_{t-\tau}$$

где μ_t– математическое ожидание этого процесса, ε_j– белый шум с конечными математическим ожиданием и дисперсией. То есть всякий слабо стационарный процесс представляется в виде линейной комбинации белых шумов, с разными весовыми коэффициентами.

Вторая глава состоит из моделей вренных рядов AR(p) и MA(q), и моделей временных рядов с зависимой вариацией.

Стохастический процесс называется процессом скользящего среднего порядка q, если $MA(q) : x_t = \sum_{\tau=0}^{q} \psi_t \varepsilon_{t-\tau}$, что равносильно $x = \varepsilon_t + \psi_1 \varepsilon_{t-1} + ... + \psi_q \varepsilon_{t-q}$. Название «скользящее среднее» обусловлено тем, что текущее значение случайного процесса определяется взвешенным средним q предыдущих значений белого шума. MA(q) применяют для того, чтобы сгладить сильно колеблющиеся данные.

Процесс является частным случаем разложения Вольда и поэтому он стационарный, Математическое ожидание: $E(x_t) = 0$. Дисперсия: $V(x_t) = \sigma^2 \sum_{i=1}^{q} \psi_i^2$.

Процессом авторегрессии порядка p назовем процесс значение которого определяется линейной комбинацией конечного числа его предыдущих значений и добавлением белого шума, и его общее уравнение имеет вид: $X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + ... + \alpha_p X_{t-p} + \varepsilon_t$, где ε_t - белый шум.

Процессы AR(p) и MA(q) схожи, но процесс MA(q) всегда стационарен. Для AR(p) это условие очень жесткое: либо он стационарен и сводится к MA(∞), либо он вообще не стационарен.

Перейдем к рассмотрению модели с условной гетероскедастичностью.

На современных рынках моделирование цен активов часто приводит к тому, что временные ряды могут быть аппроксимированы моделью случайного блуждания. Теории эффективных рынков данный факт не противоречит,но все же допускается более широкий класс моделей. В таких моделях приращения процесса не обязательно должны быть независимы и иметь одинаковое распределение. Подчеркнем, что процесс должен быть таким, чтобы невозможно было достичь гарантированного выигрыша,основываясь на прошлой информации. Введем понятия мартингала для расширение класса процессов случайного блуждания. В переводе с английского языка слово martingale означает - система назначения ставок в азартной игре, при которой ставка удваивается при каждом проигрыше.

Мартингалом назовем стохастический процесс y_t, обладающий следующими свойствами:

1) математическое ожидание процесса конечно для любого момента времени t;

2) Условное математическое ожидание процесса на основе всей информации, известной к моменту s, равно значению процесса в момент $s : E(y_t|\Omega_s) = y_s$ $\forall s \leq t$, где Ω_s — совокупность всей информации к моменту s. Это свойство носит название мартингального.

При анализе одномерного дискретного временного ряда вся информация к некоторому моменту времени s представляет собой просто реализацию процесса на отрезке $[1, s]$, и мартингальное свойство может быть переписано в виде $(t_t|y_1, ..., y_s) = y_s$, $\forall s \leq t$. Отсюда следует, что и прогноз приращения мартингала, обеспечивающий минимальную среднеквадратическую ошибку, равен нулю, и условное математическое ожидание приращения процесса равно нулю.Получили, что случайное блуждание является частным случаем мартингала. При анализе реальных финансовых временных рядов часто встречаются и иные процессы, для которых мартингальное равенство может быть заменено неравенством $(t_t|y_1, ..., y_s) \geq y_s$, $\forall s \leq t$, и его называют субмартингалом. Если же неравенство выполняется с противоположным знаком, процесс называется супермартингалом. Из свойств следует эквивалентная запись мартингала $y_t = y_{t-1} + \eta_t$, где η_t называется мартингальным приращением, или мартингальной разностью.

Некоторые черты реальных процессов на финансовых рынках не очень хорошо описывались уже разработанными моделями. В 1982 г. Энглом была предложена модель, получившая название модели с условной гетероскедастичностью или ARCH-модель. Энгл предложил эту модель, чтобы описать хорошо известную из практики черту поведения многих финансовых рядов: когда периоды большой изменчивости показателя сменяются периодом малой изменчивости. В то же время статистическая проверка постоянства дисперсии (как меры изменчивости) не отвергает гипотезы о гомоскедастичности. Начнем с простой авторегрессионной модели

$$y_t = \lambda y_{t-1} + \varepsilon_t,\ \varepsilon_t \sim SWN(0, \sigma^2)$$

Где $SWN(0, \sigma^2)$ сильный белый шум, у которого свойство некоррелированности заменено на независимости. Рассмотрим $E(y_t|y_{t-1}) = \lambda y_{t-1}$ условное математическое ожидание представляет собой теоретическую регрессию и зависит от времени t. Условная дисперсия $var(y_t|y_{t-1}) = \sigma^2$ от времени не зависит. Соответственно безусловное математическое ожидание равно нулю, а безусловная дисперсия равна $\frac{\sigma^2}{1-\lambda^2}$. Определим эту зависимость в линейном виде:

$$h_t = var(\varepsilon_t) = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2.$$

Из того, что дисперсия случайной ошибки должна быть положительной, следует, что параметр α_0 тоже должен быть положительный, а параметр α_1

неотрицательным. Процесс, определяемый формулой

$$u_t = \varepsilon_t h_t^{1/2} = \varepsilon_t(\alpha_0 + \alpha_1 u_{t-1}^2)^{1/2},\ \varepsilon_t \sim SWN(0, \sigma^2)$$

называется процессом ARCH(1). Параметр в скобках указывает наибольшую длину лага в уравнении для дисперсии.

Процесс ARCH(1) является частным случаем процесса ARCH(p). Этот процесс определяется уравнением типа AR(p) для дисперсии $h_t = var(\varepsilon_t) = \alpha_0 + \alpha_1\varepsilon_{t-1}^2 + ... + \alpha_p\varepsilon_{t-p}^2$.

Имеем, $u_t = \varepsilon_t h_t^{1/2} = \varepsilon_t(\alpha_0 + \alpha_1 u_{t-1}^2 + ... + \alpha_p u_{t-p}^2)^{1/2},\ \varepsilon_t \sim SWN(o, \sigma^2)$. Условное и безусловное математическое ожидание процесса ARCH(p) равно нулю. Условная дисперсия задана $E_{t-1}(u_t^2) = \sigma^2(\alpha_0 + \alpha_1 u_{t-1}^2 + ... + \alpha_p u_{t-p}^2)$, а безусловная дисперсия будет сложной дробно-рациональной функцией от параметров $\alpha_0, \alpha_1, ...$ Важно, что процесс также будет гомоскедастичным, и сохранится свойство некоррелированности, т.к. это свойство не зависит от конкретного уравнения процесса, а лишь от равенства нулю условного математического ожидания. Вывод, процесс ARCH(p) является случайным блужданием и мартингалом и он слабо стационарный. Чтобы протестировать модель на наличие условной гетероскедастичности в остатках нам не требуется разработка новых специальных тестов. Нулевая гипотеза об отсутствии условной гетероскедастичности эквивалентна равенству нулю всех коэффициентов кроме коэффициента α_0 в уравнении дисперсии. Алгоритм проверки состоит из следующих шагов:

1. Строим необходимую модель и получаем ряд остатков e_t. Если исследованию на условную гетероскедастичность подвергается одномерный ряд исходных данных, то этот шаг не нужен.
2. Методом наименьших квадратов оцениваем модель

$$e_t^2 = \alpha_0 + \alpha_1 e_{t-1}^2 + ... + \alpha_p e_{t-p}^2 + \nu_t$$

Проверяем гипотезу об адекватности регрессии, построенной на втором шаге: $H_0 : \alpha_1 = ... = \alpha_p = 0$ против естественной альтернативы $H_1 : \alpha_1^2 + ... + \alpha_p > 0$.

Если нулевая гипотеза отвергается, то не только устанавливается присутствие условной гетероскедастичности, но и определяется длина наибольшего лага,входящего в уравнение (параметр p).

Боллерслевом было предложено обобщение ARCH-модели, которое позволяет достичь большей гибкости, а именно добавить в уравнение для дисперсии зависимость от предыдущих значений дисперсии:

$$h_t = var(\varepsilon_t) = \alpha_0 + \alpha_1\varepsilon_{t-1}^2 + ... + \alpha_p\varepsilon_{t-p}^2 + \gamma_1 h_{t-1} + ... + \gamma_q h_{t-q}$$

Эта модель называет GARCH(p,q) - обобщенная авторегрессионная модель с условной гетероскедастичностью. Здесь условная дисперсия подчиняется модели Autoregressive Distributed Lags (ADL(p,q)) и зависит от квадратов ошибок предыдущих и значений условной дисперсии.

Для оценки процессов GARCH(p,q) используют метод максимального правдоподобия в предположении о гауссовости сильного белого шума ε_t. С ростом порядка процесса GARCH возникают все большие трудности оценки параметров, поэтом в основном используют только процесс GARCH(1,1).

В третьей главе построение и анализ оценок модели ARCH(1). Воспользовавшись выше рассмотренной теорией, на языке C++ реализована подпрограмма (см. Приложение 1). В нашей задачи используются следующие данные: N=1000, L0=0.9, eps[0]=3, y[0]=1, a=0.3, b=0.8. Как сказано выше эти значения должны удовлетворять ограничениям |L1|<1, |b|<1, L0>0, a>0. Далее по набору данных на выходе, мы вычислили оценки модели, сделали график компьютерной реализации последовательности $\varepsilon_t = \xi_t\sigma_t$ (см. рис. 2), построили диаграмму рассеяния (см. рис. 1) и подсчитали прогнозные значения (см. табл. 1). И, видно, что прогнозируемые данные близки к реальным, и этими прогнозными значениями можно воспользоваться при дальнейшем анализе.

ЗАКЛЮЧЕНИЕ

В работе мы рассмотрели основные понятия связанные со случайными процессами, временными рядами, строгую и слабую стационарность, модели временных рядов с зависимой вариацией, модели временных рядов AR(p) и MA(q). В практической части на языке C++ была разработана подпрограмма реализующая модель ARCH(1), найдены ее прогнозные значения, проанализировали полученные оценки характеристик. В ходе чего сделан общий вывод: имеющиеся оценки модели можно в дальнейшем использовать для прогнозирования на финансовых рынках.