САРАТОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИМ. Н.Г. ЧЕРНЫШЕВСКОГО

Механико-математический факультет

Кафедра теории вероятностей, математической статистики
и управления стохастическими процессами

Агафонова Н.Ю.

АНАЛИЗ ВРЕМЕННЫХ РЯДОВ

Саратов 2010

УДК 330.4(075.3)

Агафонова Н.Ю.
Анализ временных рядов. Учебно-методическое пособие.

Учебно-методическое пособие подготовлено в соответствии с положениями и требованиями Государственного образовательного стандарта высшего профессионального образования, в соответствии с требованиями, предъявляемыми к подготовке специалистов по специальности «Прикладная информатика в экономике». Учебно–методическое пособие содержит изложение основных понятий теории временных рядов в их приложении к анализу экономических процессов, а также тестовые задания для проверки знаний и подготовки к экзамену.

Для студентов специальности «Прикладная информатика в экономике».

Составитель

Старший преподаватель Н.Ю. Агафонова

УДК 330.4(075.3)

ВВЕДЕНИЕ

Модели временных рядов применяются в исследованиях динамики значительного числа реальных процессов различной природы. Они часто используются в исследованиях динамики пассажиропотоков, складских запасов, спроса на различные виды продукции, миграционных процессов в человеческом и биологических сообществах, в радиотехнике, анализе химических процессов, моделировании природных событий: динамики числа солнечных пятен, природных катастроф и многих других процессов.

Самое широкое применение модели временных рядов нашли в исследованиях финансовых рынков, в анализе динамики финансовых показателей, прогнозировании цен на различные товары, курсов акций, соотношений курсов валют и т.п.

Таким образом, в образовании студента специальности «Прикладная информатика в экономике» изучение курса «Анализ временных рядов» должно занимать существенное место.

Следует заметить, что среди изучаемых на данной специальности предметов есть курс «Эконометрика», в который анализ временных рядов входит как составная часть, однако не рассматривается там в должном объеме. Кроме того, в общем курсе эконометрики не уделяется внимания (в силу ограниченности во времени), построению стохастических моделей временных рядов.

Настоящее учебное пособие содержит как изложение основных понятий теории временных рядов, так и построение основных моделей временных рядов с использованием методов теории вероятностей.

ГЛАВА 1. ОСНОВНЫЕ ОПРЕДЕЛЕНИЯ И ПОДХОДЫ К ПОСТРОЕНИЮ ПРОГНОЗА.

Широкий круг социально-экономических, технических и естественно научных процессов часто представляется набором последовательных значений показателя $y_1, y_2, \ldots, y_t, \ldots, y_T$, зафиксированных в равноотстоящие друг от друга моменты времени $t = 1, 2, \ldots, T$, так, что длина интервала $(t, t+1)$ является постоянной. Этот набор значений y_t, $t = 1, 2, \ldots$ обычно называется временным рядом (временной серией). Такой ряд представляет собой дискретный временной процесс.

Изменения значений y_t во времени в реальной жизни обычно происходят под воздействием каких-либо причин, факторов. Однако в силу их многочисленности, сложности измерения, неразработанности теоретических предположений относительно взаимосвязей с переменной y и т.п. обосновать и построить «подходящую» для описания процесса y_t, $t = 1, 2, \ldots$ многофакторную эконометрическую модель классического типа не всегда представляется возможным. В результате в отношении ряда y_t часто выдвигается предположение, что совокупное влияние этих факторов формирует как бы внутренние закономерности в развитии процесса y_t, что дает возможность применить для его описания эконометрическую модель из специфического класса моделей временных рядов.

При решении экономических задач часто возникает необходимость построения прогноза. Существуют два подхода: качественный и количественный.

<u>Качественный</u> подход применяется, если о некотором явлении имеется недостаточная информация, либо это явление новое.

Количественный подход применяется в тех случаях, когда исследователь обладает информацией о прошедших периодах времени (в достаточном объёме).

При построении количественного прогноза данные рассматриваются как динамические или как временные ряды.

Динамическим называется ряд значений некоторого явления (показатель), расположенного в порядке возрастания другого явления (признак). Например, рост, вес; доход в семье, расходы на питание.

Ряд называется *временным,* если в качестве признака используется время. (время, доход семьи).

Временные ряды возникают при изучении различных экономических явлений. Например, запас товара на складе, цена на некоторые товары, цена на акцию. Предполагают, что закономерности и особенности экономического явления можно установить по наблюдаемым данным.

Временные ряды обозначают $Y(t)$ или y_t, где t – время (момент времени), y_t – уровень ряда (показатель).

Задача прогнозирования состоит в том, чтобы по наблюдениям $y_1, y_2,..., y_t$ получить y_{t+1}, y_{t+2},... Здесь пользуются предположением о том, что закономерности, присущие явлению в прошлом, сохранятся и в будущем. Такое предположение является верным при построении краткосрочных либо оперативных прогнозов.

Статистический подход к изучению временных рядов состоит в том, что в развитии процесса можно выделить составляющие части:

$$Y(t)=f(t)+s(t)+u(t)+e(t),$$

где $f(t)$ – функция тренда (тенденция развития), $s(t)$ – сезонная компонента, $u(t)$ – циклическая компонента, $e(t)$ – остаточная компонента.

ГЛАВА 2. ХАРАКТЕРИСТИКА СОСТАВЛЯЮЩИХ ВРЕМЕННОГО РЯДА

Итак, при построении временного ряда его разделяют на составляющие части

$$Y(t)=f(t)+s(t)+u(t)+e(t),$$

где $f(t)$ – функция тренда, $s(t)$ – сезонная компонента, $u(t)$ – циклическая компонента, $e(t)$ – остаточная компонента. Охарактеризуем их более подробно.

- *Трендом* характеризуют долговременную тенденцию развития некоторого явления. При этом она выражается некоторой монотонной функцией. В качестве примера трендов можно указать изменение демографических характеристик, рост экономических показателей, рост потребления.
- *Сезонная компонента* характеризует воздействие факторов, возникающих с определённой периодичностью. Особенностью является то, что их действие заканчивается в течение года. Например, загруженность трассы в течение суток, повышение спроса на товары для школьников в конце августа.
- *Циклическая компонента* – это функция, описывающая явление, действующее с длительным периодом. Особенностью является то, что для выявления циклической компоненты обычно недостаточно только наблюдаемых данных, а требуется анализ общей, экономической, социальной и даже исторической ситуации. Например, демографические ямы.
- *Остаточная компонента* выражает воздействие случайных факторов и при изучении этой компоненты требуется изучение статистического и вероятного анализа. В зависимости от характера выделяют: «белый шум», авторегрессию, скользящее среднее и смешанную случайную компоненту.

Временной ряд можно считать состоящим из двух частей:

Временной ряд

Детерминированная составляющая			Случайная составляющая $e(t)$			
Тренд $f(t)$	Циклическая компонента $u(t)$	Сезонная компонента $s(t)$	«Белый шум»	Авторегрессия	Скользящее среднее	Смешанная

Задача анализа временных рядов состоит в том, чтобы с помощью детерминированной компоненты предсказывать прогнозное значение временного ряда, а с помощью случайной компоненты предсказывать величину возможного отклонения и вероятность такого отклонения.

Требования к исходной информации

Для того, чтобы анализ временного ряда обладал в нужной степени достоверностью в первую очередь необходимо обеспечить качество исходной информации:

1. Данные должны быть сопоставимы;
2. Данные должны быть однородными;
3. Данные должны быть устойчивыми;
4. Необходим достаточно большой объём данных

ГЛАВА 3. ЭТАПЫ ПОСТРОЕНИЯ ПРОГНОЗА ПО ВРЕМЕННЫМ РЯДАМ

Прогнозирование экономических процессов, которые могут быть представлены одномерными временными рядами, сводится к выполнению следующих этапов:

1. Предварительный анализ данных;

2. Построение модели, т.е. выбор кривых, описывающих явление, и численное оценивание параметров модели;

3. Проверка адекватности моделей и оценка их точности;

4. Выбор лучшей модели;

5. Расчёт точечного и интервального прогноза.

Охарактеризуем некоторые из этих этапов.

При *предварительном анализе данных* происходит выявление аномальных отклонений, проверка наличия тренда, сглаживание временных рядов и расчёт показателей развития динамики экономических процессов.

Выявление аномальных отклонений производится с помощью критерия Ирвина:

Пусть $y_1, y_2,..., y_n$ – временной ряд. Вычислим значение критерия

$$\lambda_t = \frac{|y_t - y_{t-1}|}{S}, \text{ где } S = \sqrt{\frac{1}{n-1}\sum_{t=1}^{n}\left(y_t - \bar{y}\right)^2}.$$

S – несмещённое среднеквадратическое отклонение данного ряда

$\bar{y}$ – выборочное среднее ряда или средний уровень ряда.

Далее проверяется гипотеза H_0: «аномальные данные отсутствуют». По таблицам для критерия Ирвина находится $\lambda_{критическое}$, если $\lambda_t < \lambda_{критическое}$, то y_t считается нормальным; если $\lambda_t > \lambda_{критическое}$, то y_t считается аномальным. В этом случае аномальное значение исключается из ряда данных и вместо него подставляют обычно среднеарифметическое из двух соседних значений.

<u>Выявление тренда</u> обычно происходит визуальным образом и зависит от опытности исследователей. Основная цель – угадать функцию, по которой развивается процесс. Здесь требуется информация о самом явлении.

При построении модели используется понятие кривых роста. Обычно используют кривые роста, стараясь выбрать кривую максимально простого вида:

$Y(t) = a_0 + a_1 t$ – линейный тренд;

$Y(t) = a_0 + a_1 t + a_2 t^2$ – квадратичный тренд;

$Y(t) = a_0 e^{-a_1 t}$ – экспоненциальный тренд.

Подборка коэффициентов и выбор моделей осуществляется на основании метода МНК (метод наименьших квадратов). Пусть $\hat{Y}_t$ – прогнозное (вычисление по модели) значение. y_t – наблюдавшийся уровень ряда. Коэффициенты модели подбираются таким образом, чтобы сумма $\sum_{t=1}^{n}(y_t - \hat{Y}_t)^2 \to \min$. Модель, для которой достигнут такой минимум, считается наилучшей. Например, при $Y_t = a_0 + a_1 t$ (уравнение регрессии).

<u>Оценка качества построенной модели</u>

Пусть построена некоторая модель, на основании которой вычислено значение

$\hat{Y}_t$ – значение, вычисленное по модели;

y_t – наблюдение;

$e_t = y_t - \hat{Y}_t$ – остаток в определении y_t.

Для того, чтобы **модель** могла считаться **хорошей**, необходимо доказать, что **ряд остатков** e_t является **случайным и подчиняется нормальному закону распределений**. Для доказательства используются:

1. *Проверка равенства нулю математического ожидания*. Выдвигается основная гипотеза

$H_0: Me_t = 0$, то есть проверяется $\bar{e} = 0$

Для проверки этой гипотезы строится критерий или случайная величина Стьюдента:

$$t = \frac{|\bar{e}|}{S}\sqrt{n}$$

Здесь S_e – несмещённое среднеквадратичное отклонение ряда остатков. По таблице распределения Стьюдента при $\alpha = 0{,}05$ (заданном) находим $t_{кр} = t(\alpha, n)$ Если $t < t_{кр}$ – считают, что $Me_t = 0$, если $t > t_{кр}$ – считают, что $Me_t \neq 0$. Вероятность $\alpha = 0{,}05$ – является вероятностью того, что мы ошибёмся, приняв гипотезу H_0, то есть вероятность ошибки первого рода.

2. *Проверка условия случайности возникновения отдельных отклонений от тренда*. Например, метод поворотной точки: точка y_t считается поворотной, если она одновременно меньше или больше двух соседних значений. Считается, что остатки случайны, если поворот точки приходится на каждые 1,5 значения. Если больше или меньше, то остатки нельзя считать случайными и модель следует уточнить.

3. *Проверка наличия или отсутствия автокорреляции в отклонении модели*. Автокорреляция означает, что некоторое значение e_t зависит от одного или нескольких своих предыдущих значений. Наличие автокорреляции проверяют с помощью критерия Дарбина-Уотсона:

$$d = \frac{\sum_{t=2}^{n} (e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^2}$$

Для статистики Дарбина-Уотсона также существуют таблицы критических значений. В этих таблицах указывается интервал (d_1, d_2), при попадании в который, принимается то или иное решение. В частности, если $d_2 < d < 2$, то ряд остатков не коррелирован, если $d < d_1$, то имеется корреляция. При $d_1 < d < d_2$ имеется неопределённость, и сделать вывод о наличии или отсутствии корреляции нельзя. В случае, когда $d > 2$, говорят о наличии отрицательной автокорреляции. Если не удалось решить вопрос о наличии или отсутствии автокорреляции, то необходимо использовать более точные методы.

4. *Соответствие ряда остатков нормальному закону распределения*. Используется RS-критерий:

$RS = \frac{(e_{\max} - e_{\min})}{S_e}$, где $S_e = \sqrt{\frac{1}{n-1}\sum_{t=1}^{n} e_t^2}$ – несмещённое среднеквадратичное отклонение.

Если значения RS согласно таблице критерия попадает между критическими значениями $r_1 < RS < r_2$, то гипотезу о нормальном распределении принимают, в противном случае ряд из остатков нельзя считать нормально распределённым, а модель – хорошей.

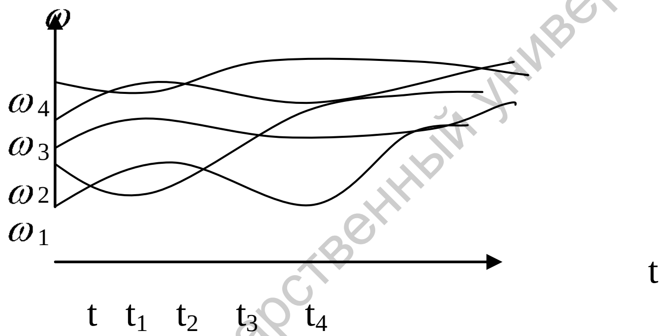
Модель считается адекватной, если выполняются **все** пункты.

ГЛАВА 4. НЕКОТОРЫЕ ОПРЕДЕЛЕНИЯ ИЗ ТЕОРИИ СЛУЧАЙНЫХ ПРОЦЕССОВ

Перейдём к изучению случайной составляющей временного ряда и построению математических моделей этой составляющей. Введём основные определения.

<u>*Определение*</u>***.*** ***Случайным процессом*** называется функция двух переменных $\xi = \xi(\omega, t)$, где ω – элементарный исход эксперимента, t – время. Таким образом $\xi : \Omega \times T \to R$ - функция двух переменных, принимающая вещественное значение.

Пусть зафиксирован элементарный исход эксперимента $\xi = \xi(t)$. Тогда эта функция выражает развития случайного явления во времени. Рассмотрим



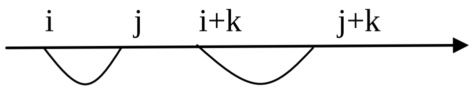
$\xi(t)$ называется траекторией процесса, а набор значений $\xi(t_1)$, $\xi(t_2)$, ... называется рядом.

<u>*Определение*</u>***.Временным рядом*** называется последовательность значений случайного процесса, взятых в некоторые моменты времени $y_1, y_2, ..., y_n$.

То есть временной ряд можно рассматривать как последовательность случайных величин.

<u>*Определение*</u>***.Последовательность случайных величин*** называется ***стационарной***, если

$\forall i=\overline{1,n}\quad My_t=a<+\infty$ и $\operatorname{cov}(y_i,y_j)=\operatorname{cov}(y_{i+k},y_{j+k})$ (*) $i+k>j$.



Выражение (*) означает, что величины на любых двух непересекающихся промежутках времени одинаковой длины одинаково коррелированны.

__Определение__. ***Ковариацией*** случайных величин X и Y называется число $\operatorname{cov}(X,Y)=M\left((X-MX)(Y-MY)\right)$.

__Определение__. ***Коэффициентом корреляции*** случайных величин X и Y называется величина $r=\dfrac{\operatorname{cov}(X,Y)}{\sqrt{DX}\sqrt{DY}}$.

__Основные свойства коэффициента корреляции__.

1. $-1\le r\le 1$, т.е. коэффициент корреляции принимает значения от -1 до 1.
2. Если случайные величины X и Y независимы, то $r=0$. (обратное неверно!)
3. Если величины X и Y линейно зависимы, то $|r|=1$. и наоборот.

__Определение__. ***Коэффициентом автокорреляции*** называется величина $r=\dfrac{\operatorname{cov}(y_i,y_j)}{\sqrt{Dy_i}\sqrt{Dy_j}}$, то есть коэффициент корреляции между уровнями временного ряда y_i и y_j.

Здесь вводится понятие порядка автокорреляции. Например, коэффициент автокорреляции первого порядка имеет вид $r_1=\dfrac{\operatorname{cov}(y_i,y_{i+1})}{\sqrt{Dy_i}\sqrt{Dy_{i+1}}}=\dfrac{\operatorname{cov}(y_i,y_{i+1})}{\sigma^2}$. Говорят, что это коэффициент с лагом 1.

Коэффициент $r_k=\dfrac{\operatorname{cov}(y_i,y_{i+k})}{\sigma^2}$ – коэффициент автокорреляции с лагом k.

__Определение__. Последовательность случайных величин ε_t называется ***белым шумом***, если $\forall t=\overline{1,n}$, $M\varepsilon_t=0$, $D\varepsilon_t=\sigma^2=const$ и $\operatorname{cov}(\varepsilon_t,\varepsilon_{t+k})=0\ \ \forall k>1$.

Таким образом, «белый шум» – это последовательность некоррелированных случайных величин, одинаково распределённых и имеющих нулевое математическое ожидание и постоянную дисперсию σ^2. Иногда добавляется следующее требование: $\varepsilon_t \propto N(0,\sigma^2)$, которое означает, что величины ε_t распределены по нормальному закону с математическим ожиданием, равным нулю и дисперсией, равной σ^2. В этом случае говорят, что шум – гауссовский «белый шум».

Основным предположением построения модели является то, что текущее значение определяется некоторой предысторией и случайной ошибкой. В основном строятся модели не более чем второго порядка.

Рассмотрим **ПРИМЕР ВЫЧИСЛЕНИЯ КОЭФФИЦИЕНТА КОРРЕЛЯЦИИ** между случайными величинами X и Y.

Пусть даны случайные величины X и Y и их совместное распределение.

$x\downarrow,y\rightarrow$	2	3	4
0	0,1	0,2	0,1
1	0,3	0,1	0,2

Требуется рассчитать коэффициент корреляции и ответить, являются ли данные величины зависимыми?

1. Восстановим частные распределения X и Y. Для этого суммируем вероятности в столбцах для X и в строках для Y.

x	0	1
p	0,4	0,6

y	2	3	4
p	0,4	0,3	0,3

$\Sigma = 1$

2. Определим числовые характеристики величин *x* и *y*:

$Mx = 0 \cdot 0{,}4 + 1 \cdot 0{,}6 = 0{,}6$

$Mx^2 = 0^2 \cdot 0{,}4 + 1^2 \cdot 0{,}6 = 0{,}6$

$Dx = Mx^2 - (Mx)^2 = 0{,}6 - 0{,}36 = 0{,}24$

$My = 2 \cdot 0{,}4 + 3 \cdot 0{,}3 + 4 \cdot 0{,}3 = 0{,}8 + 0{,}9 + 1{,}2 = 2{,}9$

$My^2 = 2^2 \cdot 0{,}4 + 3^2 \cdot 0{,}3 + 4^2 \cdot 0{,}3 = 1{,}6 + 2{,}7 + 4{,}8 = 9{,}1$

$Dy = My^2 - (My)^2 = 9{,}1 - 8{,}41 = 0{,}69$

3. Вычислим совместное математическое ожидание этих случайных величин x и y:

$$Mxy = \sum_{i=1}^{n} \sum_{j=1}^{m} x_i y_j p_{ij}$$

$$Mxy = 0 \cdot 2 \cdot 0{,}1 + 0 \cdot 3 \cdot 0{,}2 + 0 \cdot 4 \cdot 0{,}1 + 1 \cdot 2 \cdot 0{,}3 + 1 \cdot 3 \cdot 0{,}1 + 1 \cdot 4 \cdot 0{,}2 = 0{,}6 + 0{,}3 + 0{,}8 = 1{,}7$$

4. Вычисление коэффициента ковариации:

$\operatorname{cov}(x, y) = Mxy - Mx \cdot My = 1{,}7 - 0{,}6 \cdot 2{,}9 = -0{,}04$

$$r = \frac{-0{,}04}{\sqrt{0{,}24}\sqrt{0{,}69}} = \frac{-0{,}04}{0{,}5 \cdot 0{,}8} = \frac{-0{,}04}{0{,}4} = -0{,}1$$

Напомним, что коэффициент корреляции $-1 \le r \le 1$.

Так как в данном случае r близко к нулю, величины X и Y слабо коррелированны и их совместное распределение плохо оценивается при помощи линейной функции. Так как коэффициент cov отрицательный, то величины X и Y отрицательно коррелированны, то есть с ростом одной величины другая уменьшается..

ГЛАВА 5. МОДЕЛИ ВРЕМЕННЫХ РЯДОВ

5.1. Преобразование нестационарных рядов в стационарные

В предыдущей главе было отмечено, что наличие стационарности у случайного процесса вообще и у временного ряда в частности, значительно облегчает его анализ. Поэтому если в исходном ряде не наблюдается стационарность, то её пытаются установить для некоторого преобразованного ряда данных.

<u>*Определение*</u>***. Интеграцией*** называется процесс приведения нестационарного ряда к стационарному.

Выполнение стационарности обозначается $I(p)$, где р - порядок интеграции. $I(0)$ – стационарный ряд. Чаще всего используется интеграция с помощью взятия разностей.

Методы интеграции

1. Взятие конечных разностей. Пусть исходный ряд $y_1, y_2, \ldots, y_n$ не является стационарным. Построим ряд $x_1, x_2, \ldots, x_{n-1}$, где $x_t = y_t - y_{t-1}$. Если этот ряд удовлетворяет условиям стационарности, то исходный ряд обозначают $I(1)$, и делают вывод, что исходный закон y_t близок к линейному. В противном случае переходят к ряду $z_1, z_2, \ldots, z_{n-1}$, где $z_t = x_t - x_{t-1} = y_t - 2y_{t-1} + y_{t-2}$.Аналогично, обозначают $I(2)$, и делают вывод, что закон y_t близок к квадратичному.

2. Логарифмирование цепных индексов, то есть $x_t = \ln\dfrac{y_t}{y_{t-1}}$. Имеет место, если временной ряд y_t близок к экспоненте $y_t \approx a_0 e^{-a_1 t}$.

Процедуру интеграции приходится применять довольно часто, но после этого в стационарном ряде можно говорить о постоянстве коэффициентов автокорреляции второго порядка. В зависимости от характера поведения этих коэффициентов разделяют следующие процессы:

AR(k) –(autoregressive model)- модель авторегрессии порядка k ; MA(m) – Moving Average – скользящее среднее порядка m; ARMA(k,m) – модель авторегрессии-скользящего среднего порядка k ; ARIMA(k,d,m) –

интегрированная модель авторегрессии-скользящего среднего, k-порядок авторегрессии, m- порядок скользящего среднего, d-порядок интеграции.
Перейдем к рассмотрению основных моделей.

5.2 Модель авторегрессии порядка k - AR(k)

Пусть имеется временной ряд $y_1, y_2, \ldots, y_n$, или y_t $t = 1,2,\ldots,n$, где y_t – текущее значение уровня.

Основное предположение состоит в том, что текущее значение уравнения ряда y_t является линейной комбинацией *k* предыдущих значений и случайной ошибки.

Общая модель авторегрессии:

$y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \ldots + \alpha_k y_{t-k} + \varepsilon_t$,

где $k \le t$ и $\alpha_1, \ldots, \alpha_k$ – параметры модели или коэффициенты модели, ε_t – случайная ошибка или «белый шум». При построении модели AR необходимо решить две задачи: какой порядок модели следует выбрать и чему равны коэффициенты модели. Решение этих вопросов называется процедурой оценки моделей.

$$y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \ldots + \alpha_k y_{t-k} + \varepsilon_t \qquad \textbf{(1).}$$

Без ограничений предполагаем, что $My_t = 0$.

{Действительно, предположим $My_t \neq 0$, тогда $y_t' = y_t - \bar{y}$,

$My_t' = M\left(y_t - \bar{y}\right) = My_t - M\left(\frac{1}{n}\sum_{t=1}^{n} y_t\right) = My_t - \frac{1}{n}\sum_{t=1}^{n} My_t$, $\forall t, j$, $My_t = My_j$,

$My_t - \frac{1}{n} n My_t = 0$. }

Умножим выражение (1) на величину y_{t-i} и возьмём от полученного выражения математическое ожидание

$$M(y_t y_{t-i}) = \alpha_1 M(y_{t-1} y_{t-i}) + \alpha_2 M(y_{t-2} y_{t-i}) + \ldots + \alpha_i M(y_{t-i}^2) + \ldots + \alpha_k M(y_{t-k} y_{t-i}) + M \quad (2)$$

Рассмотрим выражение ковариации $\operatorname{cov}(y_t y_{t-1}) = My_t y_{t-1} - My_t My_{t-i} = My_t y_{t-i}$. Введём обозначение $\gamma_i = \operatorname{cov}(y_t y_{t-i})$ - коэффициент автоковариации. Тогда выражение (2) может быть записано в следующем виде:

$$\gamma_i = \alpha_1 \gamma_{i-1} + \alpha_2 \gamma_{i-2} + \ldots + \alpha_k \gamma_{i-k} \qquad \textbf{(2')}$$

$$My_{t-i}\varepsilon_t = My_{t-i} \cdot M\varepsilon_t = 0.$$

Независимость следует из того, что ошибка текущего время не зависит от того, какие значения были получены до текущего момента времени $\gamma_0 = M(y_{t-i} y_{t-i}) = D(y_{t-i}) = Dy_t = \sigma^2$. Разделим уравнение 2’ на γ_0, получим

$$\rho_i = \alpha_1 \rho_{i-1} + \alpha_2 \rho_{i-1} + \ldots + \alpha_k \rho_{i-k}, \qquad \textbf{(3)}$$

где $\rho_i = \dfrac{\gamma_i}{\gamma_0} = \dfrac{\operatorname{cov}(y_t, y_{t-i})}{Dy_t}$.

ρ_i – теоретический коэффициент автокорреляции порядка i.

Таким образом, получили выражение для коэффициентов автокорреляции.

Нахождение коэффициентов модели. Для модели *AR(k)* имеет место (3). При помощи этого уравнения можно получить оценки коэффициентов $\alpha_1, \ldots, \alpha_k$. Используется следующая идея: по результатам наблюдения $y_1, \ldots, y_k$ получим выборочные коэффициенты корреляции $r_1, \ldots, r_n$ $r = \dfrac{\frac{1}{n}\sum x_i y_i - \bar{x}\,\bar{y}}{S_x S_y}$.

Затем вместо теоретических коэффициентов автокорреляции подставляют в **(3)** выборочные значения и получают

систему уравнений Юла-Уокера для определения оценок коэффициентов модели:

$$\begin{cases} r_1 = a_1 + a_2 r_1 + \ldots + a_k r_{k-1} \\ r_2 = a_1 r_1 + a_2 + \ldots + a_k r_{k-2} \\ \ldots \\ r_k = a_1 r_{k-1} + a_2 r_{k-2} + \ldots + a_k \end{cases} \qquad \textbf{(4)}$$

$a_1, \ldots, a_k$ - оценки коэффициентов $\alpha_1, \ldots, \alpha_k$

В этой системе известны $r_1, \ldots, r_k$, неизвестны $a_1, \ldots, a_k$. Заметим, что с теоретической точки зрения оценки Юла-Уокера должны быть состоятельными и несмещёнными(см. курс теории вероятностей и математической статистики). Однако на практике это не всегда выполняется, наиболее сильно нарушается требование несмещённости. Причина заключается в том, что ошибки ε_t в действительности зависят от предыдущих значений $y_{t-1}, y_{t-2}, \ldots$, но так слабо, что их полагают белым шумом. Однако это все-таки отражается на модели и снижает ее точность. Наиболее точно оценки коэффициентов модели можно получить для *AR(1), AR(2)*

Обоснование применения модели AR

Для того, чтобы обосновать применение модели авторегрессии, обычно рассматривается соотношение между двумя дисперсиями. σ_y^2 – дисперсия исходного процесса; σ_ε^2 – дисперсия ошибки.

Рассмотрим выражение (2) в случае $i = 0$.

$$M(y_t y_t) = \alpha_1 M(y_t y_{t-1}) + \alpha_2 M(y_t y_{t-2}) + \ldots + \alpha_k M(y_t y_{t-k}) + M(y_t \varepsilon_t).$$

Рассмотрим $M(y_t \varepsilon_t) = M(\alpha_1 y_{t-1}\varepsilon_t + \alpha_2 y_{t-2}\varepsilon_t + \ldots + \alpha_k y_{t-k}\varepsilon_t + \varepsilon_t^2)$ (=)

Так как уровень ряда y_{t-1} и текущий уровень ошибки ε_t являются независимыми случайными величинами, то $My_{t-1}\varepsilon_t = My_{t-1}M\varepsilon_t = 0$. Поэтому, (=)$M\varepsilon_t^2 = \sigma_\varepsilon^2 = \sigma^2$

Получаем, что $\gamma_0 = \alpha_1\gamma_1 + \alpha_2\gamma_2 + \sigma_\varepsilon^2$.

Рассмотрим γ_i. Было введено, что коэффициент автокорреляции $\beta_i = \dfrac{\gamma_i}{\gamma_0}$ => $\gamma_i = \rho_i \gamma_0$. Подставим полученное выражение в γ_0:

$\gamma_0 = \alpha_1 \gamma_0 \rho_1 + \alpha_2 \gamma_0 \rho_2 + \ldots + \alpha_k \gamma_0 \rho_k + \sigma_\varepsilon^2$,

$\gamma_0 = M(y_t y_t) = M y_t^2 = D y_t = \sigma_y^2$

$\gamma_0 = \sigma_y^2 = \dfrac{\sigma_\varepsilon^2}{1 - \alpha_1 \beta_1 - \alpha_2 \beta_2 - \ldots - \alpha_k \beta_k}$ – выражение дисперсии ряда через дисперсию ошибки.

Пусть по результатам наблюдений получены выборочные коэффициенты $r_1, r_2, \ldots, r_k$, Тогда можно оценить отношение $\dfrac{r_y^2}{r_\varepsilon^2} = \dfrac{1}{1 - a_1 \sqrt{1} - a_2 \sqrt{2} - \ldots - a_k r_k}$.

Модель считается хорошей, если $\dfrac{r_y^2}{r_\varepsilon^2} >> 1$ (гораздо больше).

Примеры анализа моделей:

1. Модель AR(1).

Вид $y_t = \alpha_1 y_{t-1} + \varepsilon_t$.

Система уравнений Юла-Уокера $r_1 = a_1$.

Оценка качества. Соотношение: $\dfrac{r_y^2}{r_\varepsilon^2} = \dfrac{1}{1 - a_1 r_1} = \dfrac{1}{1 - r_1^2}$. Видно, что чем ближе коэффициент r_1 к 1, тем выше качество модели $-1 \le r \le 1$.

Например, пусть $r = 0{,}9$, тогда $\dfrac{r_y^2}{r_\varepsilon^2} = \dfrac{1}{1 - 0{,}81} = \dfrac{1}{0{,}19} \approx \dfrac{1}{0{,}2} = 5$. При таком значении коэффициента корреляции дисперсия ошибки в 5 раз меньше дисперсии уровня ряда.

2. Модель AR(2).

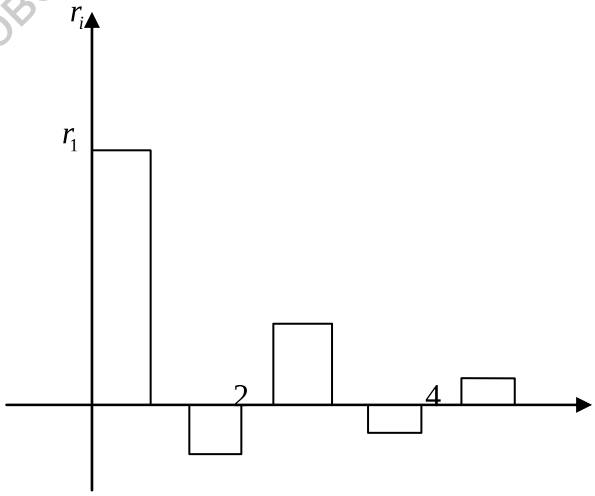
Вид $y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \varepsilon_t$.

Система Юла Уокера принимает вид: $\begin{cases} r_1 = a_1 + a_2 r_1 \\ r_2 = a_1 r_1 + a_2 \end{cases}$

$a_1 = r_1(1 - a_2) = r_1\left(1 - \frac{r_2 - r_1^2}{1 - r_1^2}\right)$ $r_2 = r_1^2 - r_1^2 a_2 + a_2 = \frac{r_1(1 - r_2)}{1 - r_1^2}$. Решение данной системы имеет вид: $\begin{cases} a_1 = \frac{r_1(1 - r_2)}{1 - r_1^2} \\ a_2 = \frac{r_2 - r_1^2}{1 - r_1^2} \end{cases}$.

Оценка качества. Подставим оценки коэффициентов в выражение для дисперсии: $\frac{r_y^2}{r_\varepsilon^2} = \frac{1}{1 - a_1 r_1 - a_2 r_2}$. После решения системы можно определить для полученной модели соотношение между дисперсиями. Анализ показывает, что при значении $r_1, r_2 \approx 0{,}9$ соотношение между дисперсиями также равно 5:1. Для модели *AR(1), AR(2)* можно наблюдать соотношение между дисперсиями 5:1.

Автокорреляционная функция – это последовательность коэффициентов автокорреляции r_i. Графическое изображение этих коэффициентов носит название корреллограмма. С её помощью идентифицируют модели авторегрессии скользящего среднего.



i

r_2 1 3

Для каждого процесса характерно своё поведение автокорреляционной функции.

5.3. Модели скользящего среднего MA(m)

Эти модели строят на основании предположения о том, что текущее значение уровня ряда представляется в виде линейной комбинации текущей и прошлых значений ошибки, то есть

$$y_t = \varepsilon_t + \beta_1\varepsilon_{t-1} - \beta_2\varepsilon_{t-2} - \ldots - \beta_m\varepsilon_{t-m},$$

где β_i – параметры модели, ε_t – белый шум, $m \le t$ – порядок модели.

Напомним, $M\varepsilon_t = 0$, $D\varepsilon_t = \sigma_\varepsilon^2 = const$, $\operatorname{cov}(\varepsilon_t, \varepsilon_{t-i}) = \gamma_i = \begin{cases} \sigma_\varepsilon^2, i = 0 \\ 0, i \ne 0 \end{cases}$.

Автокорреляционная функция имеет вид: $\beta_i() = 1 \; i = 0$, $\beta_i() = 0 \;\; \forall i \ge 1$.

Найдём коэффициенты автокорреляции для модели *MA(m)*

$$\gamma_i = \operatorname{cov}(y_t, y_{t-i}) = My_ty_{t-i} - My_tMy_{t-i} = My_ty_{t-i}.$$

Вспомогательный результат

$$My_t = M(\varepsilon_t - \beta_i\varepsilon_{t-1} - \ldots - \beta_m\varepsilon_{t-m}) = M\varepsilon_t - \beta_1 M\varepsilon_{t-i} - \ldots - \beta_m M\varepsilon_{t-m} = 0,$$

$$My_{t-i} = 0.$$

Рассмотрим его выражение при $i = 0$,

$$\gamma_0 = \operatorname{cov}(y_t, y_t) = M(\varepsilon_t - \beta_1\varepsilon_{t-i} - \ldots - \beta_m\varepsilon_{t-m}) \times (\varepsilon_t - \beta_i\varepsilon_{t-1} - \ldots - \beta_m\varepsilon_{t-m}) =$$
$$= M(\varepsilon_{t-1}\varepsilon_t) = M(\varepsilon_i\varepsilon_j) = \operatorname{cov}(\varepsilon_i, \varepsilon_j) = 0$$
$$M\varepsilon_t\varepsilon_t = M\varepsilon_{t-1}\varepsilon_{t-1} = \ldots = M\varepsilon_t^2 = \sigma_\varepsilon^2, \;\; \sigma_\varepsilon^2(1 + \beta_1^2 + \ldots + \beta_m^2)\gamma_0 = \sigma_y^2.$$

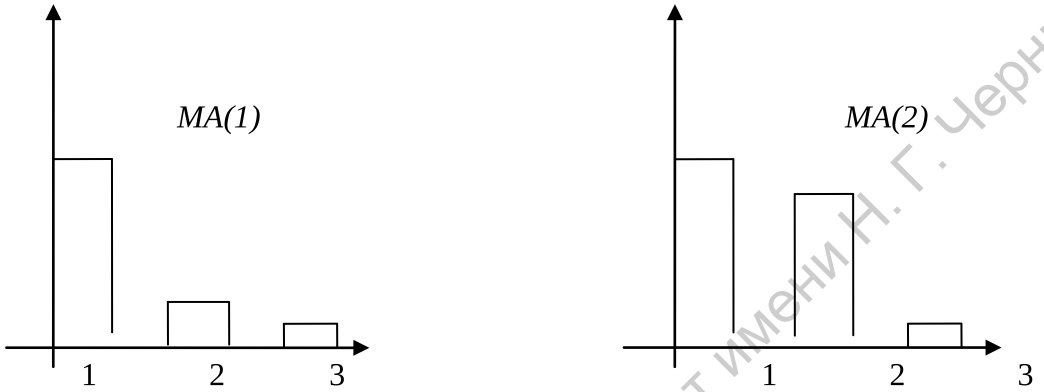
Пусть теперь $i = 1$, тогда

$$M(y_t, y_{t-1}) = \gamma_1 = M(\varepsilon_t - \beta_1\varepsilon_{t-1} - \ldots - \beta_m\varepsilon_{t-m}) \times (\varepsilon_{t-1} - \beta_1\varepsilon_{t-2} - \ldots - \beta_m\varepsilon_{t-m}) =$$
$$= -\beta_1\sigma_\varepsilon^2 + \beta_1\beta_2\sigma_\varepsilon^2 + \ldots + \beta_m\beta_{m-1}\sigma_\varepsilon^2.$$

Аналогично для любого i можно записать, что

$$\gamma_i = \begin{cases} (-\beta_i + \beta_1\beta_{i+1} + \ldots + \beta_{m-i}\beta_m)\sigma_\varepsilon^2, i = \overline{1,m} \\ 0, i > m \end{cases}.$$

Полученное выражение используется при идентификации моделей. Автокорреляционная функция модели *MA(m))* обрывается после момента m:



Получим ***выражение для коэффициентов автокорреляции.***

Для этого разделим γ_i на γ_0: $\rho_i = \begin{cases} \dfrac{-\beta_i + \beta_1\beta_{i+1} + \ldots + \beta_{m-1}\beta_m}{1 + \beta_1^2 + \ldots + \beta_m^2} \\ 0 \end{cases}$.

Аналогично решению системы Юла-Уокера, для получения оценок коэффициентов модели *M*А необходимо вычислить выборочные коэффициенты автокорреляции r_1, r_m, подставить полученные выражения и получить оценки $b_1, \ldots, b_m$. Однако такую систему нужно решать численными методами.

Примеры нахождения оценок MA(1).

Вид модели $y_t = \varepsilon_t - \beta\varepsilon_{t-1}$.

Выражение для дисперсии $\sigma_y^2 = \gamma_0 = (1 + \beta_1^2)\sigma_\varepsilon^2$, $\rho_1 = \dfrac{-\beta_1}{1 + \beta_1^2}$.

Пусть получено значение выборочного коэффициента автокорреляции, тогда получаем $r_1 = \frac{-b_1}{1+b_1^2}$, $r_1 b_1^2 + b_1 + r_1 = 0$, $(ax^2 + x + c = 0)$,

$$b_1 = \frac{-\frac{1}{r_1} + \sqrt{\left(\frac{1}{r_1}\right)^2 - 4}}{2},\; b_2 = \frac{-\frac{1}{r_1} - \sqrt{\left(\frac{1}{r_1}\right)^2 - 4}}{2}.\;\; (b_1)_1 \cdot (b_1)_2 = 1,\;\; (b_1)_1 = \frac{1}{(b_1)_2}.$$

Отсюда следует, что из двух корней полученного уравнения, один из корней всегда >1, а другой <1. Согласно теории стационарных процессов необходимым условием стационарности является $b_1 < 1$. Также необходимо, чтобы $\left(\frac{1}{r_1}\right) - 4 \ge 0$, $|r_1| \le 0{,}5$.

Вывод: модели скользящего среднего порядка 1 могут применяться только для описания процесса с автокорреляционной функции, обрывающейся после первой задержки и таких, что $r < 0{,}5$.

Оценим дисперсию для процесса *MA(1)*, получаем: $\frac{\sigma_y^2}{\sigma_\varepsilon^2} = (1 + b_1^2) \le 1{,}25$.

Это означает, что точность модели *MA* равна 1,25 и относительный выигрыш в точности составляет менее 25%.

5.4. Модели ARMA(p,q) или ARIMA(p,d,q)

Обобщение модели ARMA(p,q) или ARIMA(p,d,q) – модели авторегрессии скользящего среднего *ARMA(p,q)=ARIMA(p,0,q)*.

Эти модели основаны на предположении о том, что текущий уровень ряда y_t является линейной комбинацией *p* своих предыдущих уровней и *q* своих предыдущих ошибок. При идентификации модели *ARMA(p,q)* пользуются тем, что их автокорреляционные функции затухают плавно по экспоненте или синусоиде.

Общий вид модели:

$$y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \ldots + \alpha_p y_{t-p} + \varepsilon_t - \beta_1 \varepsilon_{t-1} - \beta_2 \varepsilon_{t-2} - \ldots - \beta_q \varepsilon_{t-q}.$$

ГЛАВА 6. ИССЛЕДОВАНИЕ ВРЕМЕННОГО РЯДА «УРОВЕНЬ БЕЗРАБОТИЦЫ В США С 1969 ПО 1996 ГОД»

Рассмотрим основные методы анализа временных рядов на примере анализа данных об уровне безработицы в США по данным за 1969-1996 гг. Данные взяты из [1] и представлены в таблице 1.

Таблица 1.

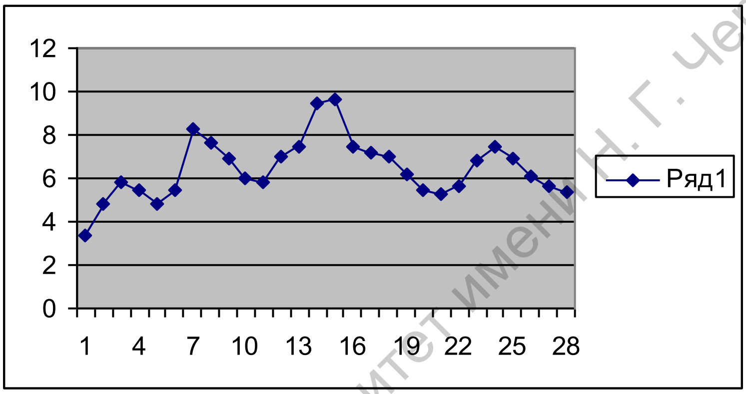
№	Год	Уровень
1	1969	3,4
2	1970	4,8
3	1971	5,8
4	1972	5,5
5	1973	4,8
6	1974	5,5
7	1975	8,3
8	1976	7,6
9	1977	6,9
10	1978	6
11	1979	5,8
12	1980	7
13	1981	7,5
14	1982	9,5
15	1983	9,6
16	1984	7,5
17	1985	7,2
18	1986	7
19	1987	6,2
20	1988	5,5
21	1989	5,3
22	1990	5,6
23	1991	6,8
24	1992	7,5
25	1993	6,9
26	1994	6,1

27	1995	5,6
28	1996	5,4

Проверим этот ряд на наличие тренда.

Построим график по исходным данным с помощью пакета статистического анализа Excel.

Рисунок 1. Изменение уровня безработицы.



Визуальный анализ позволяет предположить отсутствие тренда. Проведем соответствующий тест, который состоит в том, что сравниваются средние уровни ряда. Разобьем ряд на две части, и проверим гипотезу о равенстве дисперсий с помощью F-критерия Фишера. Воспользуемся средствами анализа данных в Excel. Получим следующие результаты:

Таблица 2

Двухвыборочный F-тест для дисперсии

	Переменная 1	*Переменная 2*
Среднее	6,314	6,586
Дисперсия	2,538	1,386
Наблюдения	14	14
df	13	13
F	1,831	
P(F<=f) одностороннее	0,144	

F критическое одностороннее	2,577

Так как расчетное значение F-теста (1.831) меньше критического (2.577) при уровне значимости 0.05, то нет оснований отвергать гипотезу о равенстве средних и на основании этого критерия делаем вывод о том, что тренд в данном ряду действительно отсутствует.

Построим линейную однопараметрическую модель регрессии. Для этого снова воспользуемся надстройкой *Анализ данных*, инструментом *Регрессия*.

Получаем следующие результаты:

Таблица 3

	Коэффициенты	*Стандартная ошибка*	*t-статистика*
Y-пересечение	6,017	0,538	11,184
Переменная X 1	0,030	0,032	0,922

Во втором столбце табл.3 содержатся коэффициенты уравнения регрессии, в третьем столбце – стандартные ошибки коэффициентов уравнения регрессии, а в четвертом –*t*-статистика, используемая для проверки значимости коэффициентов уравнения регрессии.

Уравнение регрессии зависимости уровня безработицы Y_t от времени имеет вид

$$Y_t = 6.07 + 0,30t$$

Оценим качество построенной модели. Для этого проверим остатки регрессии на равенство математического ожидания нулю, отсутствие автокорреляции и соответствие нормальному закону.

Таблица 4.

Наблюдение	*Предсказанное*	*Остатки*	*et^2*	*(et-et-*	*(et*et-1)*
1	6,05	-2,65	7,00		
2	6,08	-1,28	1,63	1,88	3,38

3	6,11	-0,31	0,09	0,94	0,39
4	6,14	-0,64	0,40	0,11	0,19
5	6,17	-1,37	1,87	0,53	0,87
6	6,20	-0,70	0,48	0,45	0,95
7	6,23	2,07	4,30	7,67	-1,44
8	6,26	1,34	1,81	0,53	2,79
9	6,29	0,61	0,38	0,53	0,83
10	6,32	-0,32	0,10	0,86	-0,19
11	6,35	-0,55	0,30	0,05	0,17
12	6,38	0,62	0,39	1,37	-0,34
13	6,41	1,09	1,20	0,22	0,68
14	6,44	3,06	9,39	3,88	3,36
15	6,46	3,14	9,83	0,00	9,61
16	6,49	1,01	1,01	4,54	3,15
17	6,52	0,68	0,46	0,11	0,68
18	6,55	0,45	0,20	0,05	0,30
19	6,58	-0,38	0,15	0,69	-0,17
20	6,61	-1,11	1,24	0,53	0,43
21	6,64	-1,34	1,81	0,05	1,50
22	6,67	-1,07	1,15	0,07	1,44
23	6,70	0,10	0,01	1,37	-0,10
24	6,73	0,77	0,59	0,45	0,07
25	6,76	0,14	0,02	0,40	0,10
26	6,79	-0,69	0,48	0,69	-0,09
27	6,82	-1,22	1,50	0,28	0,85
28	6,85	-1,45	2,11	0,05	1,78

График остатков регрессии имеет вид:

Необходимые коэффициенты имеют значения:

DW	0,57
R1	0,62

RS	4,25

По таблицам критерия Дарбина-Уотсона определяем значения границ интервала для принятия решения: d_L=1.302, d_U=1.461. Таким образом, вычисленное значение критерия $DW < d_L$, и следовательно, в ряду остатков присутствует автокорреляция. Таким образом, по данному критерию модель нельзя считать адекватной. Этот же вывод можно сделать, рассмотрев коэффициент автокорреляции 1-го порядка, который по абсолютной величине превышает табличное значение R1>r(1), где R1=0.62, r(1)=0.271.

Проверим соответствие ряда остатков нормальному распределению. По таблицам RS-критерия находим, что для данного объема выборки n=28 значение критерия должно принадлежать интервалу (3.41-4.79). Так как значение RS=4.25 находится внутри этого интервала модель по этому критерию можно считать адекватной, а ряд остатков - соответствующим нормальному закону.

Вычисление среднего значения для ряда остатков дало результат $\bar{e} = 0.0$ (значащие цифры появляются в 15 разряде). Таким образом, гипотеза о равенстве математического ожидания нулю выполняется.

В итоге, приходим к выводу, что модель нельзя считать статистически адекватной, что означает, что точечные и интервальные оценки, полученные с ее помощью будут недостаточно надежны.

Список рекомендуемой литературы

1. Бабешко Л.О. Основы эконометрического моделирования: Учебное пособие. Изд. 2-е, испр. М.: КомКнига, 2006.
2. Магнус Я.Р., Катышев П.К., Пересецкий А.А. Эконометрика: Начальный курс: Учебное пособие для вузов. М.: Дело, 2005.
3. Орлова И.В. Экономико-математическое моделирование: Практическое пособие по решению задач.—М.: Вузовский учебник, 2005.
4. Тихомиров Н.П., Дорохина Е.Ю. Эконометрика: М.: Экзамен, 2003.