ЛЕКЦИЯ № 9.3

Тема: Ряды динамики. Анализ временных рядов.

1. Ряды динамики

Происходящие в экономических системах процессы в основном проявляются как ряд расположенных в хронологическом порядке значений определенного показателя, который в своих изменениях отражает развитие изучаемого явления.

Ряд наблюдений за значениями определенного показателя, упорядоченный в зависимости от возрастающих или убывающих значений другого показателя, называют *динамическим рядом, временным рядом, рядом динамики.* Отдельные наблюдения временного ряда называются *уровнями* этого ряда.

Временные ряды бывают моментные, интервальные и производные. *Моментные ряды* характеризуют значения показателя на определенные моменты времени. *Интервальные ряды* характеризуют значения показателя за определенные интервалы времени. *Производные ряды* получаются из средних или относительных величин показателя.

Уровни ряда могут иметь детерминированные или случайные значения. Ряд последовательных данных о количестве дней в месяце, квартале, годе являются примерами рядов с детерминированными значениями.

Прогнозированию подвергаются ряды со случайными значениями уровней. Каждый показатель таких рядов может иметь дискретную или непрерывную величину.

Важное значение для прогнозирования имеет выбор интервалов между соседними уровнями ряда. При слишком большом интервале времени могут быть упущены некоторые закономерности в динамике показателя. При слишком малом – увеличивается объем вычислений, могут появляться ненужные детали в динамике процесса.

Поэтому выбор интервала времени между уровнями ряда должен решаться конкретно для каждого процесса, причем удобнее иметь равноотстоящие друг от друга уровни.

Важным условием правильного отражения временным рядом реального процесса развития является *сопоставимость уровней ряда.* Несопоставимость чаще всего встречается в стоимостных характеристиках, изменениях цен, территориальных изменениях, при укрупнении предприятий и др. Для несопоставимых величин показателя неправомерно проводить прогнозирование.

Для успешного изучения динамики процесса необходимо, чтобы информация была полной, временной ряд имел достаточную длину, отсутствовали пропущенные наблюдения. Уровни временных рядов могу иметь аномальные значения. Появление таких значений может быть вызвано ошибками при сборе, записи или передачи информации. Это ошибки технического порядка или ошибки первого рода. Однако аномальные значения могут отражать реальные процессы, например, скачок курса доллара и др. Такие аномальные значения относят к ошибкам второго рода, они устранению не подлежат.

Для выявления аномальных уровней временных рядов можно использовать метод Ирвина.

Метод предполагает использование следующей формулы:

$$\lambda_t = \frac{|y_t - y_{t-1}|}{\sigma_y},\ t = 1,2,\dots,n$$

где σ_y, среднеквадратическое отклонение временного ряда $y_1, y_2, \dots, y_i, y_n$.

$$\sigma_y = \sqrt{\frac{\sum(y_t-\bar{y})^2}{n-1}},\ \bar{y} = \frac{\sum y_t}{n}$$

Расчетные значения λ_2, λ_3,... сравниваются с табличными значениями критерия Ирвина λ_a, и если оказываются больше табличных, то соответствующее значение y_t уровня ряда считается аномальным.

Значения критерия Ирвина для уровня значимости α = 0,05 приведены в табл. 1.

Таблица 1

n	2	3	10	20	30	50	100
λ_a	2,8	2,3	1,5	1,3	1,2	1,1	1,0

После выявления аномальных уровней необходимо определение причин их возникновения. Если они вызваны ошибками технического порядка, то устраняются или заменой аномальных уровней соответствующими значениями по кривой, аппроксимирующей временной ряд, или заменой уровней средней арифметической двух соседних уровней ряда.

Ошибки, возникающие из-за воздействия факторов, имеющих объективный характер, устранению не подлежат.

Если во временном ряду проявляется длительная тенденция изменения экономического показателя, то в этом случае говорят, что имеет место тренд.

Пусть дан временной ряд $y_1, y_2, \ldots, y_i, y_n$, $t = 1,2,\ldots,n$.

Считают, что значения уровней временных рядов экономических показателей складываются из следующих составляющих (компонент): тренда, сезонной, циклической и случайной.

Под *трендом* понимают изменение, определяющее общее направление развития или основную тенденцию временного ряда. Тренд относят к систематической составляющей долговременного действия. Во временных рядах часто происходят регулярные колебания, которые относятся к периодическим составляющим рядов экономических процессов.

Если период колебаний не превышает 1 года, то их называют *сезонными,* более 1 года – *циклическими составляющими.* Чаще всего причиной сезонных колебаний являются природные, климатические условия, циклических – демографические циклы др. Тренд, сезонная и циклическая составляющая называются *регулярными,* или *систематическими компонентами временного ряда.* Если из временного ряда удалить регулярную компоненту, то останется *случайная компонента.*

Если временной ряд представлен в виде суммы составляющих компонент, то модель называется *аддитивной,* если в виде произведения, то *мультипликативной* или *смешанного* типа.

$y_t = u_t + s_t + v_t + e_t$ — аддитивная форма,

$y_t = u_t\, s_t\, v_t\, e_t$ — мультипликативная форма,

$y_t = u_t\, s_t\, v_t + e_t$ — смешанная форма,

где y_t — уровни временного ряда;

u_t — временной тренд;

s_t — сезонная компонента;

v_t — циклическая составляющая;

e_t — случайная компонента.

Прогнозирование временных рядов целесообразно начинать с построения графика исследуемого показателя. Однако в нем не всегда прослеживается присутствие тренда. Поэтому в этих случаях необходимо выяснить, существует ли тенденция во временном ряду или она отсутствует.

Дан временной ряд: $y_1, y_2, \ldots, y_i, y_n$, $t = 1,2,\ldots,n$.

Рассмотрим *критерий «восходящих и нисходящих» серий,* согласно которому тенденция определяется по следующему алгоритму.

1.Для исследуемого временного ряда определяется последовательность знаков, исходя из условий

+, если $y_{t+1} - y_t > 0$,

- , если $y_{t+1} - y_t < 0$.

При этом, если последующее наблюдение равно предыдущему, то учитывается только одно наблюдение.

2. Подсчитывается число серий $v(n)$. Под серией понимается последовательность подряд расположенных плюсов или минусов, причем один плюс или один минус считается серией.

3. Определяется протяженность самой длинной серии $l_{max}(n)$

4. По табл. 2 находится значение $l(n)$.

Таблица 2

Длина ряда (n)	$n<=26$	$26<n<=153$	$153<n<=170$
Значение $l(n)$	5	6	7

5. Если нарушается хотя бы одно из следующих неравенств, то гипотеза об отсутствии тренда отвергается с доверительной вероятностью 0,95.

$$v(n) > \left[(2n-1)/3 - 1{,}96\sqrt{16n - 29/90}\right],\ l_{max}(n) \le l(n)$$

Квадратные скобки неравенства означают целую часть числа. Под целой частью числа понимают целое число, не превосходящее само это число. Например, [2, 4] = 2.

2. Основные показатели динамики экономических процессов

Для количественной оценки динамики эколого-экономических процессов применяют такие статистические показатели, как ***абсолютные приросты, темпы роста и прироста***. Они подразделяются на ***цепные, базисные*** и ***средние***.

Если сравнение уровней временного ряда осуществляется с одним и тем же уровнем, принятым за базу, то показатели называются ***базисными***. Если сравнение осуществляется с переменной базой, и каждый последующий уровень сравнивается с предыдущим, то вычисленные показатели называются ***цепными***.

Формулы для вычисления цепных, базисных и средних абсолютных приростов, темпов роста:

Обозначения	Абсолютный прирост	Темп роста	Темп прироста
Цепной	$\Delta y_t = y_t - y_{t-1}$	$T_t = y_t / y_{t-1} \cdot 100\%$	$K_t = T_t - 100\%$
Базисный	$\Delta y_{бt} = y_t - y_б$	$T_{бt} = y_t / y_б \cdot 100\%$	$K_{бt} = T_{бt} - 100\%$
Средний	$\Delta y_t = (y_n - y_1)/(n-1)$	$T = (y_n / y_1)^{1/(n-1)} \cdot 100\%$	$K = T - 100\%$

Описание динамики ряда средним приростом соответствует его представлению в виде прямой, проходящей через две данные точки. Для получения прогнозного значения на один шаг вперёд достаточно к последнему наблюдению добавить ***значение среднего абсолютного прироста***:

$$y_{n+1} = y_n + \Delta y,$$

где y_n - значение показателя в n точке временного ряда;

y_{n+1} - прогнозное значение показателя в точке $n+1$;

Δy - значение среднего прироста временного ряда.

Получение прогнозного значения по этой формуле корректно, если динамика близка к линейной. На такой равномерный характер развития динамики указывают примерно одинаковые цепные абсолютные приросты.

Использование среднего темпа роста (среднего темпа прироста) для описания динамики развития ряда соответствует его представлению в виде показательной или экспоненциальной кривой, проведенной через две крайние точки, и характерно для процессов, изменение динамики которых происходит с постоянным темпом роста.

Прогнозное значение на i шагов вперед определяется по формуле

$$y_{n+i} = y_n \cdot \bar{T}^i,$$

где y_{n+i} - прогнозная оценка значения показателя в точке $n+i$;

$\bar{T}$ - средний темп прироста, выраженный не в %.

Недостатком прогнозирования с использованием среднего прироста и среднего темпа роста является то, что они учитывают начальный и конечный уровни ряда, исключая влияния промежуточных уровней. тем не менее они используются как простейшие, приближенные способы прогнозирования.

Предварительный анализ временных рядов экономических показателей заключается в выявлении аномальных значений уровней ряда, которые не соответствуют возможностям рассматриваемой экономической системы, а также в определении наличия тренда. Наиболее распространенным приемом для устранения аномальных значений показателей и отсутствия тенденции временного ряда является сглаживание временного ряда. При этом производится замена фактических уровней временного ряда расчетными, что способствует более четкому проявлению тенденции.

3. Анализ дискретных временных рядов

Определим *дискретный временной ряд* как последовательность измерений значений переменной (процесса) за определенный период через одинаковые промежутки времени:

$$Z_1, Z_2, Z_3, \dots, Z_t, \dots, Z_n, \qquad (1)$$

Последовательные наблюдения в (1) обычно зависимы. С детерминистской точки зрения (1) можно представить как:

$$Z_t = f(t) + \varepsilon_t,$$

где t =1,2,...,n;

f – гладкая (непрерывная и дифференцируемая) функция, характеризующая долгосрочное движение в зависимости от времени - тренд;

ε_t – случайный ряд возмущений, наложенный на систематическую часть.

При наличии во временном ряду тенденции и циклических колебаний значения каждого последующего уровня ряда зависят от предыдущих. Корреляционная зависимость между последовательными уровнями временного ряда называют *автокорреляцией уровней ряда.* Количественно ее можно измерить с помощью линейного коэффициента корреляции между уровнями исходного временного ряда и уровнями этого ряда, сдвинутыми на один или несколько шагов во времени, называемого коэффициентом автокорреляции.

Коэффициент автокорреляции уровней ряда первого порядка, смещенных на одну единицу времени, определяется по формуле:

$$r_1 = \frac{\sum_{t=2}^{n}(y_t - \bar{y}_1)\cdot(y_{t-1} - \bar{y}_2)}{\sqrt{\sum_{t=2}^{n}(y_t - \bar{y}_1)^2 \cdot \sum_{t=2}^{n}(y_{t-1} - \bar{y}_2)^2}}.$$

Коэффициент автокорреляции уровней ряда второго порядка:

$$r_2 = \frac{\sum_{t=3}^{n}(y_t - \bar{y}_3)\cdot(y_{t-2} - \bar{y}_4)}{\sqrt{\sum_{t=3}^{n}(y_t - \bar{y}_3)^2 \cdot \sum_{t=3}^{n}(y_{t-2} - \bar{y}_4)^2}}, \text{где } \bar{y}_3 = \frac{\sum_{t=3}^{n} y_t}{n-2}; \;\; \bar{y}_4 = \frac{\sum_{t=3}^{n} y_{t-2}}{n-2}.$$

Аналогично можно определить коэффициенты автокорреляции более высоких порядков.

Так как коэффициент автокорреляции строится по аналогии с линейным коэффициентом корреляции, то по нему можно судить о наличии линейной или близкой к линейной тенденции. Чем ближе коэффициент автокорреляции первого порядка к единице, тем более выражена линейная тенденция. Для некоторых временных рядов, имеющих сильную нелинейную тенденцию, коэффициент автокорреляции уровней исходного ряда может приближаться к нулю.

Последовательность коэффициентов автокорреляции уровней первого, второго и т. д. порядков называют ***автокорреляционной функцией временного ряда.*** Если наиболее высоким оказался коэффициент автокорреляции первого порядка, исследуемый ряд содержит только тенденцию. Если наиболее высоким оказался коэффициент автокорреляции порядка τ, то ряд содержит циклические, или сезонные колебания с периодичностью в τ моментов времени. Если ни один коэффициент не является значимым, можно сделать вывод о том, что-либо ряд не содержит тенденции и циклических колебаний, либо содержит сильную нелинейную тенденцию.

Число периодов иди. моментов времени, по которым рассчитывается коэффициент автокорреляции, называют ***лагом.***

Построение аналитической функции для моделирования тенденции (тренда) временного ряда называют аналитическим выравниванием временного ряда. Тенденция во времени может принимать разные формы, для ее формализации используют функции.

Такой подход, несмотря на заслуженную критику, используется и в настоящее время.

Второй подход (стохастический) заложил Эдни Юл в 1927 году. Он предложил для его объяснения пример, ставший классическим: «Если рассматривать свободное качание маятника, отклоняющегося на малый угол под действием силы тяжести, то хорошо известно, что его движение является гармоническим, т. е. оно может быть представлено синусоидальной и косинусоидальной волной с постоянными амплитудами и периодами колебаний. Но если маленький мальчик обстреливает маятник горохом нерегулярным образом, то его движение будет возмущено. Маятник будет качаться, но с нерегулярными амплитудами и периодами колебаний. Фактически вместо такого поведения, при котором расхождение между теорией и наблюдением можно отнести за счет незначительной ошибки, горох вызывает ряд толчков, *воздействующих на будущее движение системы.* Эта концепция приводит к теории *стохастических процессов,* важнейшим разделом которой является теория стохастических временных рядов».

Третий подход к анализу временных рядов - это спектральный анализ в частотной области. В частном случае можно получить выравнивание по ряду Фурье (при этом обычно рассматривается не более 5 гармоник (j= 1,2,3,4,5)):

$$z_t = a_0 + \sum_{j=1}^{k}\left(a_j \cos jt + b_j \sin jt\right). \qquad (2)$$

Параметры a_j и b_j - находятся с помощью МНК, в результате применения которого, получим:

$$a_0 = \frac{1}{n}\sum_{t=1}^{n} z_t\,,\ \ a_j = \frac{2}{n}\sum_{t=1}^{n} z_t \cos jt\,,\ \ b_j = \frac{2}{n}\sum_{t=1}^{n} z_t \sin jt. \qquad (3)$$

Анализ временных рядов преследует целый ряд целей, например:

1) описание поведения ряда;
2) построение модели для объяснения наблюдений;
3) пункты 1) и 2) используют для прогноза, исходя из предположения о сохранении тенденции развития в будущем.

Для достижения поставленных целей используют модели, основанные на перечисленных выше детерминистском, стохастическом, спектральном и других подходах. В общем случае можно предположить в модели наличие следующих компонент:

1) тренд или долгосрочное колебание;
2) регулярное движение относительно тренда;
3) сезонная компонента;
4) остаток.

Модель, в которой временной ряд представлен как сумма перечисленных выше компонент, называется *аддитивной моделью* временного ряда. Если временной ряд представлен как произведение компонент, то она называется *мультипликативной моделью* временного ряда.

Отделить тренд и сезонность в общем случае невозможно, так как они взаимно проникают друг в друга. При выделении тренда и сезонности остается колеблющийся ряд. Удаление тренда (сглаживание временного ряда) можно осуществить с помощью скользящей средней (СС). Скользящая средняя, в отличие от простой средней для всей выборки, содержит сведения о тенденциях изменения данных.

Для этого к первым (2m+1) точкам ряда подбирают полином

$$Q_p(t) = a_p t^p + a_{p-1} t^{p-1} + a_{p-2} t^{p-2} + \ldots + a_1 t + a_0$$

(для определения значения тренда в (m+1) точке) и минимизируют:

$$\sum_{t=-m}^{m} \left(z_t - a_p t^p - a_{p-1} t^{p-1} - a_{p-2} t^{p-2} - \ldots - a_1 t + a_0\right)^2 \qquad (4)$$

Затем подбирают полином того же порядка для второго, третьего,... (2*m*+2) наблюдения. Эта процедура продолжается вдоль всего ряда до последней группы из (2*m*+1) точек. На самом деле нет необходимости подбирать полином каждый раз. Покажем, что эта процедура соответствует некоторой линейной комбинации наблюдений с постоянными коэффициентами.

Например, пусть 2*m*+1=5 или *t:* -2,-1,0,1,2; $p = 3$ $\left(Q_3(t) = a_3 t^3 + a_2 t^2 + a_1 t' + a_0\right)$, тогда (4) примет вид:

$$\sum_{t=-m}^{m} (z_t - a_3 t^3 - a_2 t^2 - a_1 t - a_0)^2 \qquad (5)$$

После дифференцирования (5) по a_j и преобразований (так как суммы *t* с нечётной степенью равны 0), получим систему уравнений

$$\begin{cases} \sum z_t = 5a_0 + \qquad +10a_2 \quad , \\ \sum t z_t = \quad 10a_1 + \qquad +34a_3, \\ \sum t^2 z_t = 10a_0 + \qquad +34a_2, \\ \sum t^3 z_t = \quad 34a_1 + \qquad +130a_3. \end{cases} \qquad (6.)$$

Из (1) и (3) уравнения системы (6) следует, что:

$$a_0 = \frac{1}{35}(-3z_{-2} + 12z_{-1} + 17z_0 + 12z_1 - 3z_2),$$

но $z_0 = a_0$. Итак, значение тренда в какой-либо точке равно средневзвешенному значению пяти точек с данной точкой в качестве центральной и весами $\frac{1}{35}$[-3,12,17,12,-3]. Для пяти точек и *p* = 1 получаем простую скользящую среднюю (СС):

Кроме рассмотренного подхода к выводу формул взвешенных СС существуют другие способы определения СС: использование простых СС, формулы Спенсера и т. д.

При рассмотрении СС в рамках примера (*p*=3, *m*=2) следует отметить проблему крайних двух точек - они не оцениваются, хотя её можно решить, определив из (6) коэффициенты a_1, a_2, a_3.

Для прогноза в следующей точке следует определить Q_3 (3).

Рассмотренные выше СС (их называют иногда линейными фильтрами) являются симметрическими (т. е. коэффициенты (веса) симметричны относительно среднего).

Для прогнозирования в статистике используют асимметричные фильтры. Так, в Excel простая *скользящая средняя* заменяет не средний, а последний уровень ряда в промежутке сглаживания, она (СС) используется для расчета значений в прогнозируемом периоде, на основе среднего значения переменной для указанного числа предшествующих периодов, по формуле:

$$F_{t+1} = \frac{1}{N} \sum_{h=0}^{N} Z_{t-h+1}, \qquad (7)$$

где N – число предшествующих периодов, входящих в СС;

Z_h, – фактическое значение в момент времени h;

F_h – прогнозируемое значение в момент времени h.

Асимметричные СС иногда могут учитывать степень «устаревания данных», т. е. каждое новое наблюдение будет иметь вес больше предыдущих, например:

$$F_{t+1} = (1-\alpha)(A_{t+1} + \alpha A_t + \alpha^2 A_{t-1} + \ldots) = (1-\alpha)\sum_{i=0}^{\infty} \alpha' A_{t-i+1},\ 0 < \alpha < 1 \quad (8)$$

Рассмотренный подход к определению асимметричных СС в Excel носит название *экспоненциальное сглаживание* (ЭС) (или экспоненциальных средних). ЭС предназначается для предсказания значения F_{t+1} на основе прогноза для предыдущего периода F_t, скорректированного с учетом погрешностей в этом прогнозе (A_t - F_t), из (8) можно получить, что:

$$F_{t+1} = F_t + \alpha\,(A_t - F_t) = \alpha A_t + (1-\alpha)F_t.$$

Существует ещё целый ряд методов сглаживания и экстраполяции: модель Хольта - Уинтерса (содержит три параметра α_1; α_2, α_3 и позволяет учесть сезонность), модель Харрисона является модификацией предыдущей и выражает сезонность через гармоники. Указанные методы были разработаны для анализа экономических процессов. Широко известны также модель Бокса - Дженкинса, фильтры Калмана и Бюсси.

Практически все рассмотренные методы содержат предположения относительно исходных данных, генерирующих временной ряд. Критерием адекватности той или иной модели может служить только практическое достижение первоначальных целей анализа временных рядов (описание поведения ряда, объяснение изменения наблюдений, прогноз и т. д.).

Замечание. Тренд даёт возможность прогнозировать основную тенденцию изменения явления во времени. Значения тренда $\hat{z}_t$, обычно сопровождаются ошибками. Хотя при анализе временных рядов обычно нельзя утверждать нормальность распределения отклонений относительно тренда, но, тем не менее, в качестве одного из возможных подходов это предположение используется. Доверительный интервал для тренда определяется как

$$\hat{z}_t \pm t_{\text{дв.}\alpha,v} \cdot S_{\hat{z}},$$

где z_t - значение тренда в момент времени t, $t_{\text{дв.}\alpha,v}$ - квантиль распределения Стьюдента для двусторонней критической области при уровне значимости α с $v = n - d$ степенями свободы (n — число наблюдений, d - число оцениваемых параметров, например, для уравнения прямой d = 2);

$S_{\hat{z}} = \sqrt{\frac{\sum(z_t - \hat{z}_t)^2}{v}}$ – среднеквадратическое отклонение членов ряда от тренда.

Стационарные временные ряды. Временной ряд, не имеющий тренда (либо с исключённым трендом), называется *стационарным,* или иначе - если его свойства не зависят от начала отсчёта времени (механизм, генерирующий ряд, не меняется со временем, хотя и носит вероятностный характер). Поэтому перечисленные ниже параметры для данного ряда являются постоянными:

$M(z_t) = M, M(z_t - M)^2 = \sigma^2 = D(z_t);\ \ M[(z_t - M)(z_{t+k} - M)] = c_k$ - k-ая автоковариация, $\rho_k = \rho_{-k} = \frac{c_k}{\sigma^2} -$ соответствующая автокорреляция.

Иногда совокупность значений р* представляется на графике и называется коррелограммой (но не каждая последовательность констант является коррелограммой).

Для стационарного процесса рассматривают 3 основных типа моделей (соответствующих определённым типам стационарных стохастических процессов):

1. Авторегрессии (АР) порядка p. В этой модели текущее значение t выражается через линейную комбинацию p предыдущих значений процесса плюс случайный импульс ε_t:

$$z_t = a_1 z_{t-1} + a_2 z_{t-2} + a_3 z_{t-3} + \ldots + a_{t-p} z_{t-p} + \varepsilon_t \quad (9).$$

Важными частными случаями являются:

а) при p=1, модель процесса Маркова (процесс с отсутствием последействия - каждое следующее значение зависит только от предыдущего):

$$z_t = a_1 z_{t-1} + \varepsilon_t; \quad (10)$$

b) при p=2, модель процесса Юла:

$$z_t = a_1 z_{t-1} + a_2 z_{t-2} + \varepsilon_t \tag{11}$$

2. Скользящего среднего (СС):

$$z_t = b_1 \varepsilon_{t-1} + b_2 \varepsilon_{t-2} + b_3 \varepsilon_{t-3} + \ldots + b_{t-q} \varepsilon_{t-q} + \varepsilon_t \tag{12}$$

(термин СС не означает, что сумма весов при ъ равна 1).

3. Авторегрессии - скользящего среднего (АРСС):

$$z_t = a_1 z_{t-1} + a_2 z_{t-2} + a_3 z_{t-3} + \ldots + a_{t-p} z_{t-p} + b_1 \varepsilon_{t-1} + b_2 \varepsilon_{t-2} + b_3 \varepsilon_{t-3} + \ldots + b_{t-q} \varepsilon_{t-q} + \varepsilon_t$$

Обычно на практике достаточно рассматривать модели АР, СС, АРСС при p и q, не превышающих 2.

Для описания *нестационарных* процессов пользуются экспоненциально взвешенными средними, а в более общем случае моделями авторегрессии - проинтегрированного скользящего среднего (АРПСС).

Замечание. 1. Альтернативным методом построения модели прогноза является метод группового учета аргумента (МГУА), позволяющий с помощью ЭВМ определить в данном классе функций оптимальную структуру искомой зависимости и идентифицировать параметры по внешним критериям, установленным человеком. В средине XX века Кенуй и Тьюки предложили метод складного ножа (бут-стреп метод), согласно которому вся совокупность данных разбивается на части и проводится статистический анализ всех частей для достижения несмещённости оценок. Этот метод и был положен в основу МГУА. Человек задает среду решения (список переменных и перечень опорных функций), а также критерий селекции (отбора). Для проверки адекватности последовательность наблюдений разбивается на части (последоваельности). Первая часть (обучающая последовательность) используется для определения оценок коэффициентов по методу наименьших квадратов. Вторая часть (проверочная последовательность) используется для селекции моделей по внешнему критерию, например, построения лучшей прогнозирующей модели.

2. С точки зрения современной науки развитие сложных процессов является детерминированным только между определёнными точками структурных изменений - точками бифуркаций, появление которых случайно. Поэтому можно говорить о существовании пределов предсказуемости, прогнозировать далее которых принципиально невозможно из-за случайного появления бифуркационных точек.

При стандартных подходах (детерминистском, стохастическом и спектральном) наличие неучтенных факторов описывается присутствием случайной составляющей (ε_t), имеющей тот или иной закон распределения (обычно нормальный). Линейные модели окружающего мира, к которым сводятся перечисленные выше методы давно перестали удовлетворять потребностям человека. В настоящее время назрела необходимость описания сложных объектов (процессов) с помощью нелинейных методов, дающих возможность учитывать реальную жизнь (в которой нелинейные связи преобладают), а также учитывать ненаблюдаемые и неконтролируемые факторы. Для этого совокупность данных разбивается на две части: по первой строятся различные модели (обучающая последовательность), по второй отбираются лучшие модели (проверочная последовательность), дающие наилучший прогноз.

Следует отметить, что предлагаемая идеология соответствует принципу Soft Computing - «мягких вычислений» (терпимость к нечеткости и частичной истинности используемых данных для достижения интерпретируемости, гибкости и низкой стоимости решений), сформулированному известным математиком Лотфи Заде (1994).

Согласно идеологии DATA MINING главным является практическая применимость метода, а не его математическое обоснование, (которое присутствует, но по принципиальным моментам не может быть аксиоматизировано. Например, из теоремы К. Гёделя о неполноте формальной арифметики следует принцип внешних дополнений, согласно которому целью построения математической модели должно быть не достижение формального критерия (минимизация квадратов ошибок, F-Фишера и т. д.), а решение поставленной задачи, например, получение прогнозирующей модели.

Литература

1. *Кендэл М.* Временные ряды. / Пер. с англ. Ю.П. Лукашинз. М.: Финансы и статистика, 1981. -199 с, с илл.

Представлены основные идеи и методы анализа временных рядов. Элементарное изложение и достаточное количество примеров делает эту книгу незаменимой при практических исследованиях.

2. *Четыркин Е.М., Калихман И.Л.* Вероятность и статистика. - М.: Финансы и статистика, 1982. - 319 с: с илл.

Рассматриваются основные идеи теории вероятностей и на их основе описываются методы математической статистики, имеющие наибольшее применения на практике. Изложены основы анализа временных рядов