Временные ряды и их предварительный анализ

Задачи

1. Виды временных рядов.
2. Этапы предварительного анализа временных рядов.
3. Описательные характеристики временных рядов.
4. Простейшие приемы прогнозирования.

Основные определения

Прогнозирование – это научно обоснованное, основанное на системе установленных причинно-следственных связей и закономерностей, выявление состояния и вероятных путей развития явления или процесса.

Прогноз – это количественное вероятностное утверждение в будущем о состоянии объекта с относительно высокой степенью достоверности на основе анализа тенденций и закономерностей прошлого и настоящего.

Основные определения

Особенности прогноза:

- является следствием действительности, итогом выводов, эмпирических данных и обоснованных предположений
- имеет элемент случайности
- носит многовариантный характер
- его точность проверяется временем
- требует соблюдения объективности

Прогноз позволяет:

- оценить состояние объекта и при необходимости найти возможные управляющие решения
- выявить проблемы, возможные в будущем
- моделировать варианты развития событий

Основные определения

- **Горизонт прогнозирования (**иногда период упреждения**)** - предельный срок, в пределах которого прогноз выполняется с заданной точностью (иногда просто число периодов в будущем, которые покрывает прогноз).
- **Период прогнозирования** - это основная единица времени, на которую делается прогноз.
- **Интервал прогнозирования** - частота, с которой делается новый прогноз. (часто совпадает с периодом прогнозирования; это означает, что прогноз пересматривается каждый период).
- **Методы прогнозирования** - совокупность приемов и способов мышления, позволяющих построить прогноз.

Основные группы методов прогнозирования

- экстраполяция
- нормативные расчеты
- экспертные оценки
- аналогии
- математическое моделирование

Основные этапы прогнозирования социально-экономических процессов

- постановка задачи и сбор необходимой информации
- данные должны быть:
  - достоверными и точными (достоверный источник)
  - значимыми (отражать исследуемые явления)
  - согласованными
  - собраны через определенные интервалы времени
- первичная обработка исходных данных
- определение круга возможных моделей прогнозирования оценка параметров модели
- исследование качества выбранных моделей, проверка их адекватности реальному процессу, выбор лучшей модели
- построение прогноза
- содержательный анализ полученного прогноза

Основные этапы прогнозирования социально-экономических процессов: замечания

- Более сложная методика необязательно дает лучший результат - выбранная модель изменяется со временем.
- А вот возможности создающего прогноз играют большую роль в получении хорошего прогноза.
- Одни методы подходят для краткосрочных прогнозов, другие – для долгосрочных.
- Комбинирование нескольких методов дает лучший результат, чем применение их по отдельности.
- содержательный анализ полученного прогноза

Рисунок 1 – Классификация прогнозов

Типология прогнозов: критерии и признаки

1) масштаб прогнозирования

-макро (страна)-, микро (предприятиие)-, мезо (отрасль, регион, комплекс)-экономический прогноз;

-структурный (межотраслевой и межрегиональный) прогноз;

2) горизонт прогнозирования

-оперативные (до 1 месяца);

-краткосрочные (от 1 месяца до 1 года) – для разработки безотлагательных решений

-среднесрочные (от 1 года до 5 лет);

-долгосрочные (от 5 лет до 15-20 лет) – чтобы наметить основной курс развития предприятия

-дальнесрочные (свыше 20 лет).

Типология прогнозов: критерии и признаки

Применительно к комплексным национальным экономическим прогнозам принята следующая классификация:

-краткосрочные прогнозы до 2-3 лет,

-среднесрочные до 5-7 лет,

- долгосрочные до 15-20 лет.

Типология прогнозов: критерии и признаки

3) характер объекта

-научно-технический (развитие НТП, техническое прогнозирование)

-демографический

-использования или количества природных ресурсов

-военно-политический

- динамики народного хозяйства и др.

4) цели

- поисковый: строится на основе продолжения в будущем тенденций развития изучаемого явления; «что произойдет, если сохранится соответствующая тенденция»

- нормативный (программный) – определяет пути и сроки достижения возможного (желаемого) состояния объекта.

Типология прогнозов: критерии и признаки

5) степень информационной обеспеченности объектов прогнозирования

- объекты с полным обеспечением количественной информацией, для которых имеется в наличии ретроспективная количественная информация в объеме достаточном для реализации метода экстраполяции, либо статистического метода;

-объекты с неполным обеспечением количественной информацией;

-объекты с наличием качественной ретроспективной информацией;

-объекты с полным отсутствием ретроспективной информации (как правило, это проектируемые и строящиеся объекты).

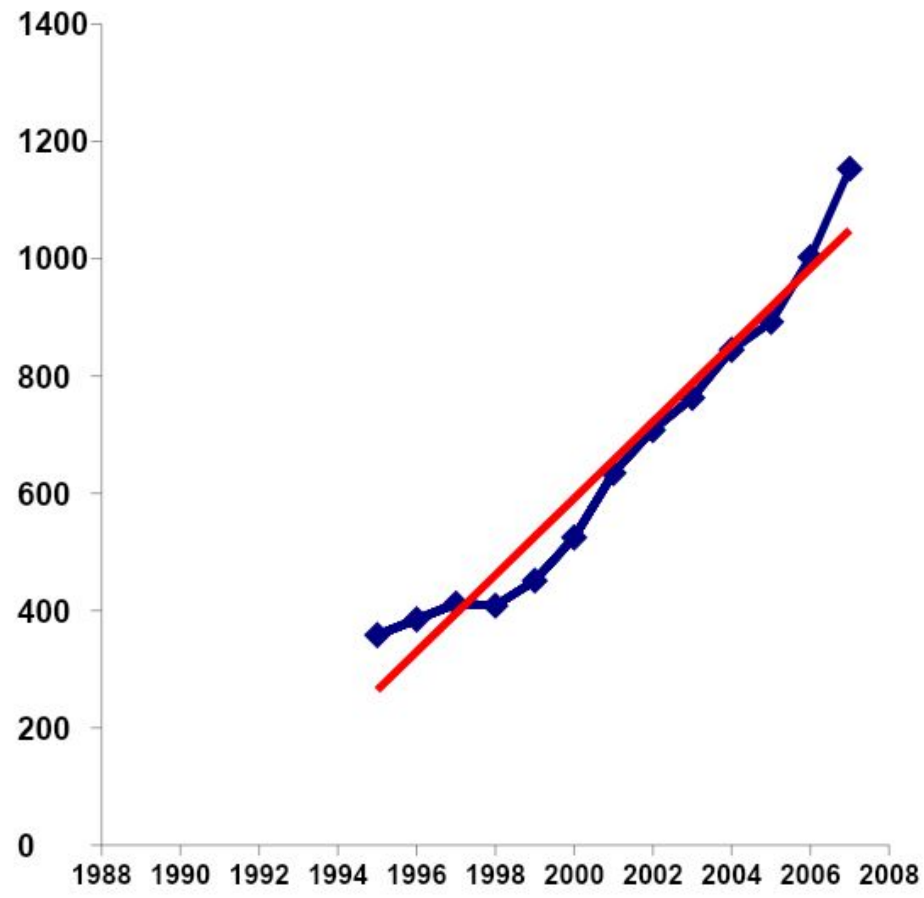
Основные определения

Временной ряд (y_t, $t=1,\ldots T$) – это последовательность упорядоченных во времени числовых показателей, характеризующих уровень состояния и изменения изучаемого явления.

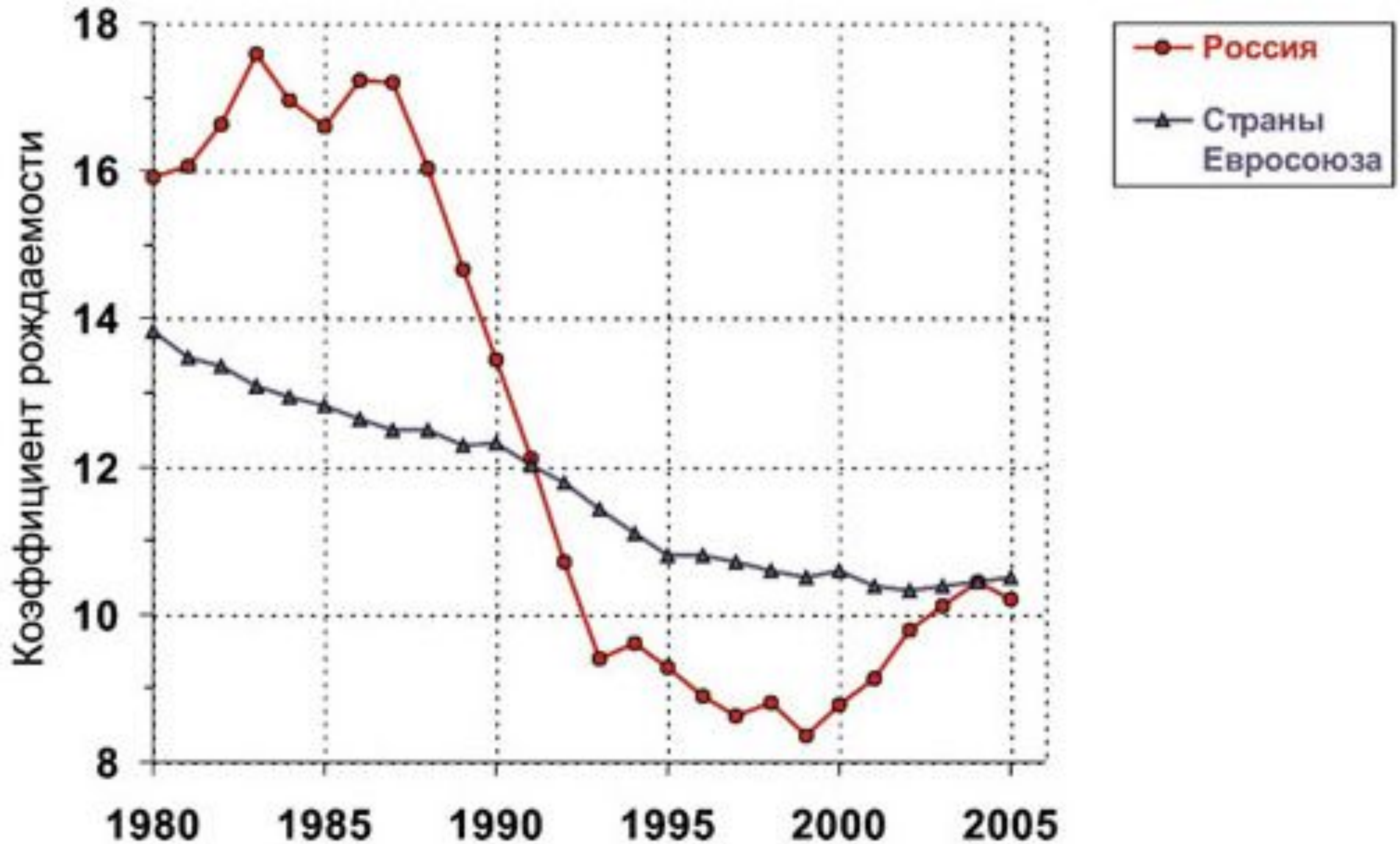
Каждый уровень временного ряда формируется под воздействием большого числа факторов и при изучении ВР предполагается, что совокупное влияние этих факторов формирует общие закономерности в развитии процесса.

Моделирование продаж объёма пива в РФ

Годы	Продажа пива
1995	358
1996	384,85
1997	411,7
1998	408,2
1999	451,1
2000	524,6
2001	634,6
2002	707,8
2003	762,5
2004	844,7
2005	892,1
2006	1002,8
2007	1153,3



Динамика коэффициентов рождаемости (на 1000 человек населения)

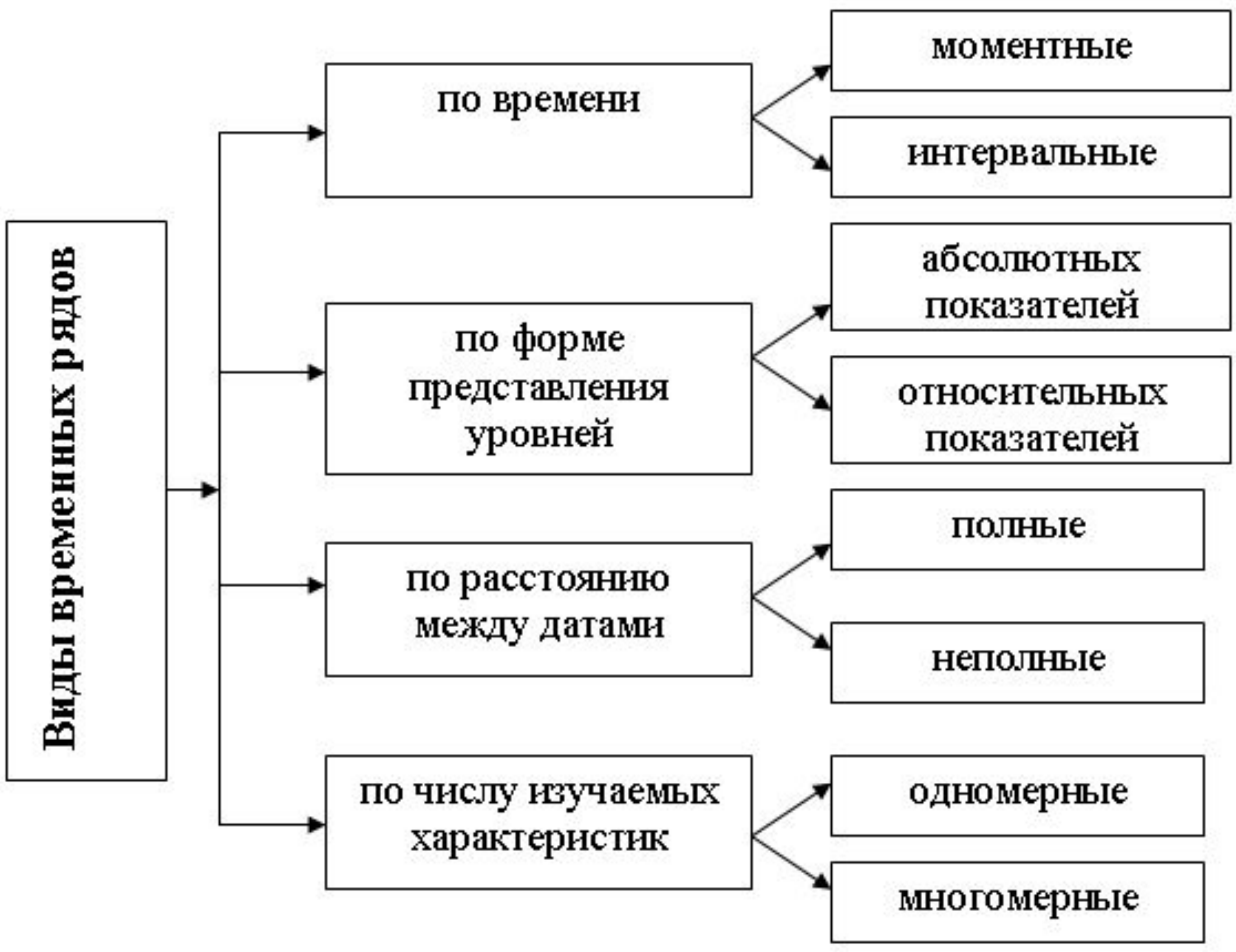


Русский крест - динамика общих коэффициентов рождаемости и смертности (на 1000 человек населения)

Особенности исследования временных рядов

- существенность порядка наблюдений;
- существование временных лагов между временными рядами;
- наличие долговременных тенденций (трендов) во временных рядах;
- малый объем выборки;
- автокоррелированность уровней временных рядов;
- развитие социально-экономических процессов и явлений происходит непрерывно, но реально исследовать можно лишь дискретные по времени значения процесса;
- уровни временного ряда не являются одинаково распределенными.

Классификация временных рядов



Классификация временных рядов

- **по времени** – моментные и интервальные.
  - *В интервальном ряду* уровень ряда характеризует результат, накопленный или вновь произведенный за определенный интервал времени.
  - *В моментном ряду* уровень ряда характеризует изучаемое явление в конкретный момент времени.
- **по форме представления уровней** – ряды абсолютных, относительных и средних величин.
- **по расстоянию между датами или интервалами** выделяют полные (измерения сделаны в равноотстоящие моменты времени) и неполные (в неравноотстоящие) временные ряды.
- **по количеству фиксируемых характеристик** изучаемого явления выделяют одномерные ВР (одна характеристика/один объект), и многомерные временные ряды (при наблюдении нескольких характеристик выделенного объекта).

Условия правильного формирования временных рядов

- Важное значение для исследования процесса имеет выбор ширины интервалов между соседними членами ряда.
- Если выбрать слишком большой интервал, можно упустить существенные закономерности в динамике показателей, в то же время слишком малый интервал может привести к появлению ненужных деталей, то есть к засорению общей тенденции.

Условия правильного формирования временных рядов

- Важнейшим условием правильного формирования временных рядов является сопоставимость уровней, образующих ряд.
- Уровни ряда, подлежащие изучению, должны быть однородны по экономическому содержанию, и учитывать существо изучаемого явления и цель исследования.
- Должна быть:

- сопоставимость уровней по экономическому содержанию

- сопоставимость по территории,

- сопоставимость уровней по кругу охватываемых объектов,

- сопоставимость уровней по единицам измерения, моменту регистрации,

- сопоставимость уровней по методике расчета,

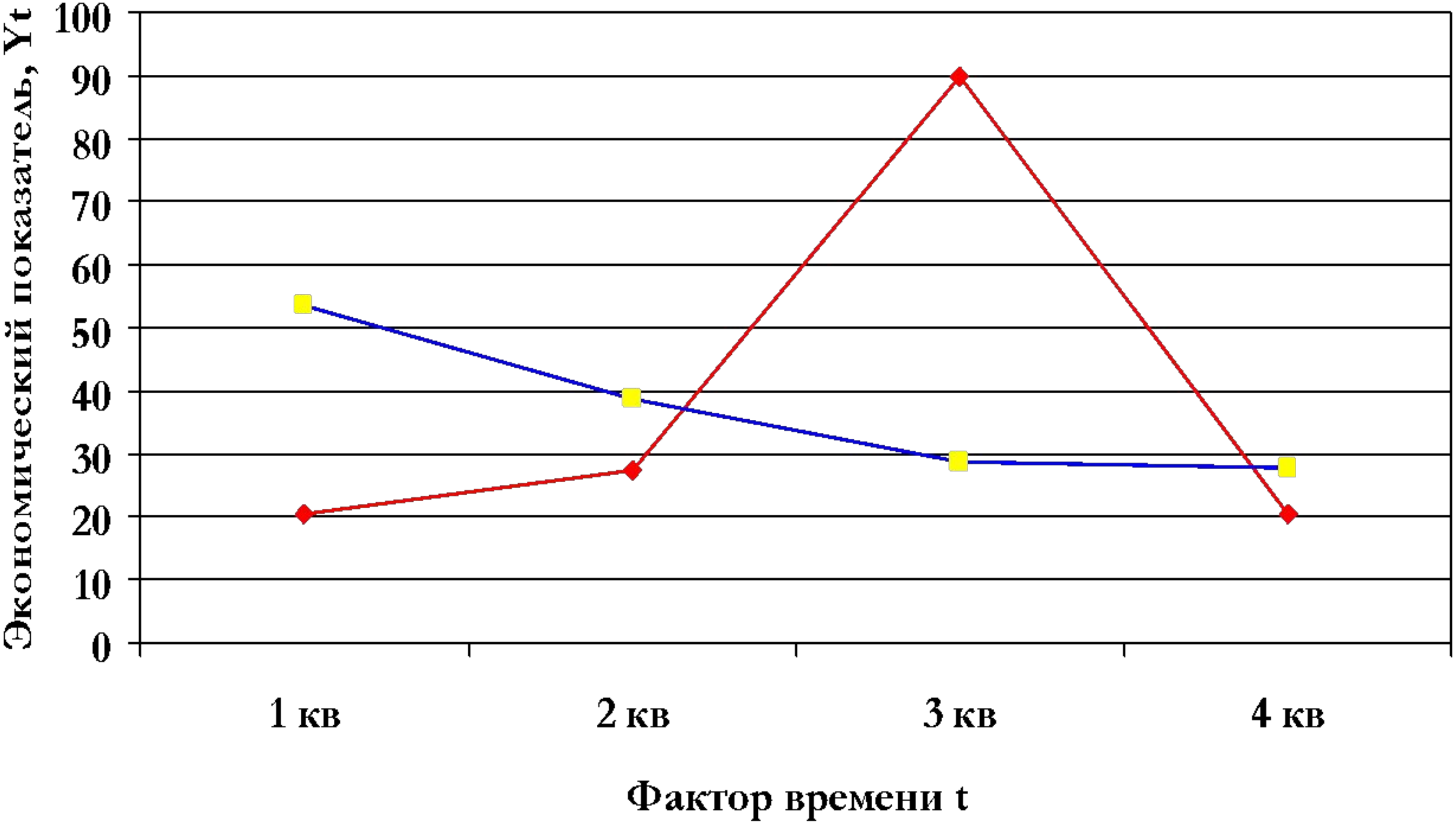
- сопоставимость уровней по ценам,

- сопоставимость уровней по достоверности.

Этапы предварительного анализа временных рядов

1. Построение графика исходных данных
2. Расчет и анализ основных статистических показателей: среднее, дисперсия, размах вариации и др.
3. Расчет абсолютных и относительных показателей динамики
4. Оценивание автокорреляционной и частной автокорреляционной функции исходного временного ряда

– Графическая форма представления ВР



Абсолютные показатели динамики (на цепной и базисной основе)

- Абсолютный прирост

$$\Delta_i = y_i - y_{i-1} \qquad \Delta_i = y_i - y_0$$

- Темп роста

$$T_{Рцеп_i} = \frac{y_i}{y_{i-1}} \cdot 100 \qquad T_{p_i} = \frac{y_i}{y_0} \cdot 100$$

- Темп прироста

$$\mathrm{T_{пр}} = \mathrm{T_{р}} - 100\%.$$

- Ускорение

$$\mathbf{a}_t = \Delta_t - \Delta_{t-1}$$

Средние показатели динамики

Средний абсолютный прирост показывает, на сколько в среднем в единицу времени изменяется уровень ряда

$$\Delta\bar{y}_{цеп} = \frac{\sum \Delta y_{цеп}}{T-1}, \qquad \Delta\bar{y}_{баз} = \frac{y_T - y_0}{T-1},$$

Средний темп роста – показывает, во сколько в среднем за единицу времени изменился уровень динамического ряда

$$\bar{T}_{p_{цеп}} = \sqrt[T-1]{T_{p_{цеп1}} \cdot T_{p_{цеп2}} \cdot \ldots \cdot T_{p_{цепn}}} \cdot 100 \qquad \bar{T}_{P_{баз}} = \sqrt[T-1]{\frac{y_T}{y_0}} \cdot 100$$

Средний темп прироста

$$\bar{T}_{пр} = \bar{T}_{р} - 100\,\%.$$

Автокорреляционная функция (АКФ)

Одно из главных отличий последовательности наблюдений, образующих временной ряд, от случайной выборки заключается в том, что члены временного ряда являются, вообще говоря, статистически взаимозависимыми.

Степень тесноты статистической связи между двумя случайными величинами может быть измерена парным коэффициентом корреляции.

Поскольку в нашем случае коэффициент измеряет корреляцию, существующую между членами одного и того же временного ряда, его принято называть коэффициентом автокорреляции.

При анализе изменения величины $r(\tau)$ в зависимости от значения τ принято говорить об автокорреляционной функции $r(\tau)$.

Автокорреляционная функция (АКФ)

Оценка автокорреляционной функции:

$$\hat{r}(\tau) = \frac{\gamma(\tau)}{\gamma(0)} = \frac{\sum_{t=\tau+1}^{T}(y_t - \bar{y}_t)(y_{t-\tau} - \bar{y}_t)}{\sum_{t=1}^{T}(y_t - \bar{y}_t)^2}$$

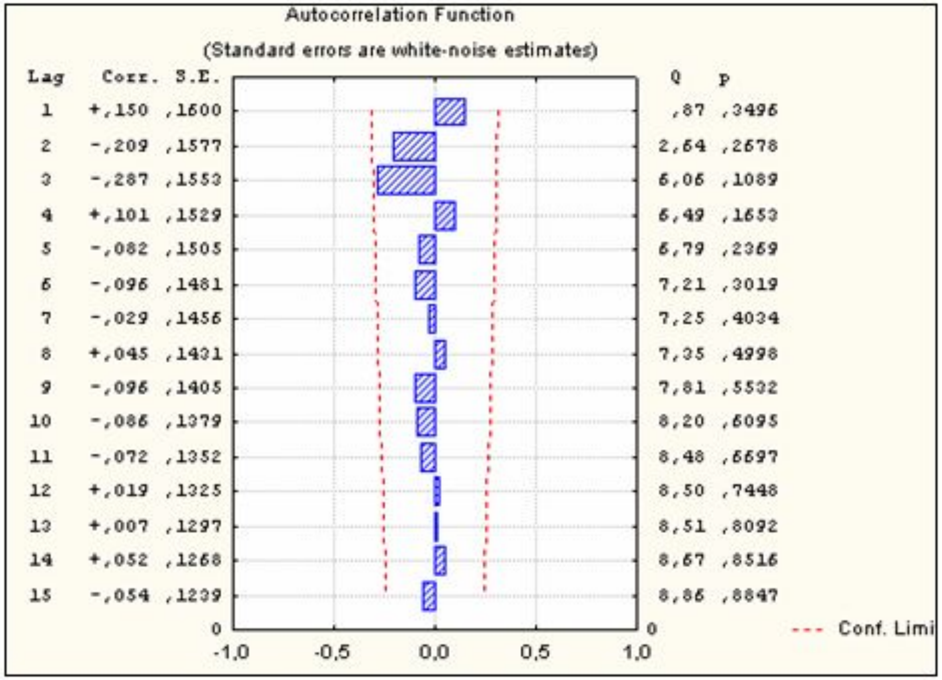
где $\gamma(0)$ - выборочная дисперсия;

$\gamma(\tau)$ - выборочная автоковариация;

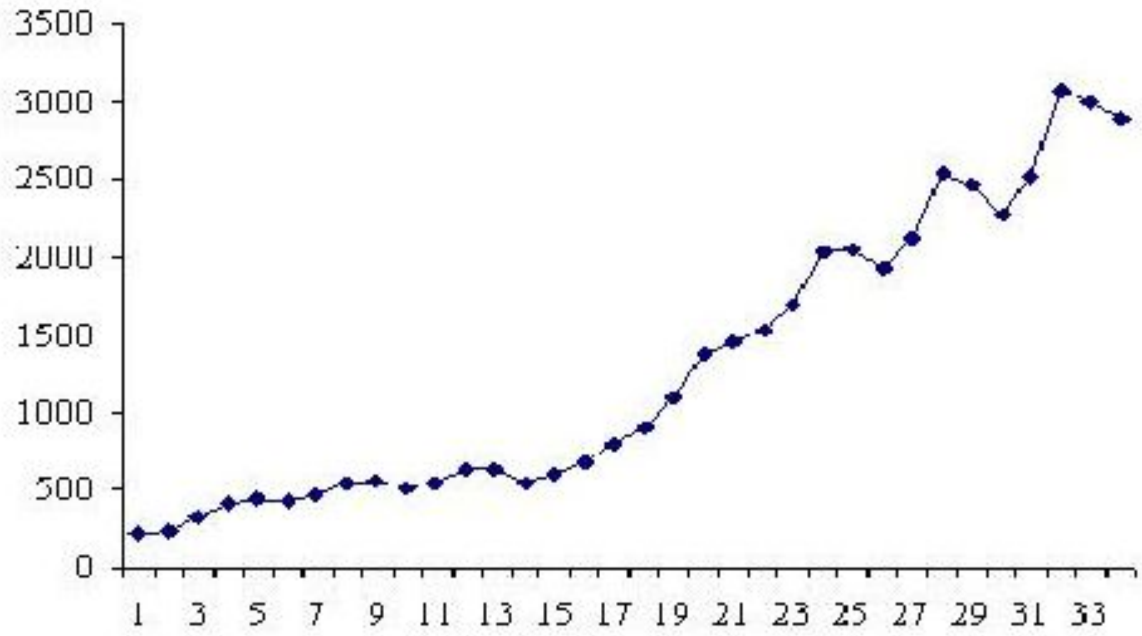
τ - лаг (порядок автокорреляции).

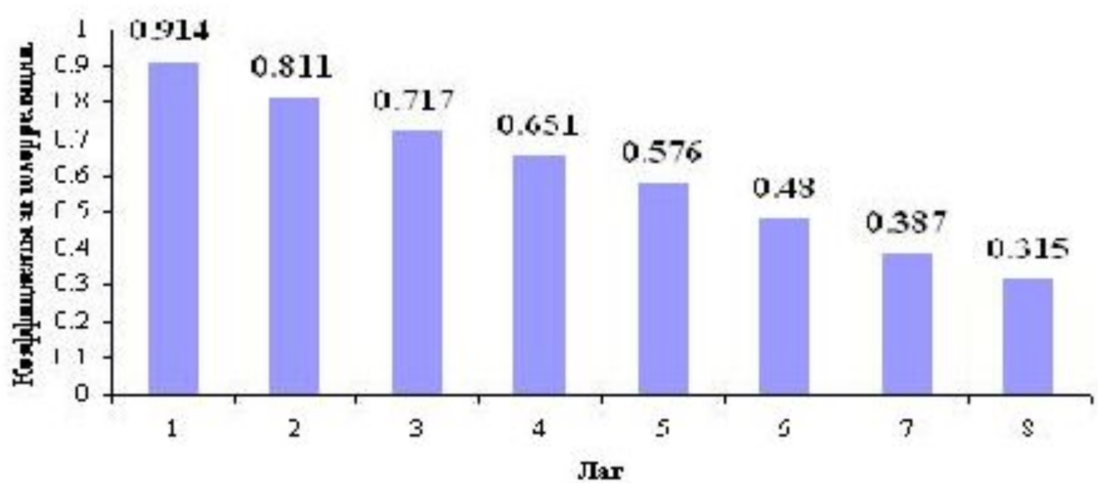
Пример коррелограммы автокорреляционной функции (АКФ) в ППП Statistica

График автокорреляционной функции называют коррелограммой.



Коррелограмма автокорреляционной функции ВВП





Частная автокорреляционная функция (ЧАКФ)

С помощью этой функции реализуется идея измерения автокорреляции, существующей между разделенными τ тактами времени членами временного ряда x_t и $x_{t+\tau}$, при устраненном опосредованном влиянии на эту взаимозависимость всех промежуточных членов этого временного ряда.

- Частная автокорреляция 1-го порядка может быть подсчитана с использованием соотношения:

$$r_{\text{част}}(2) = r\left(x_t, x_{t+2} \middle| x_{t+1} = \mu\right) = \frac{r(2) - r^2(1)}{1 - r^2(1)},$$

- где μ − среднее значение анализируемого стационарного процесса.

Частная автокорреляционная функция (ЧАКФ)

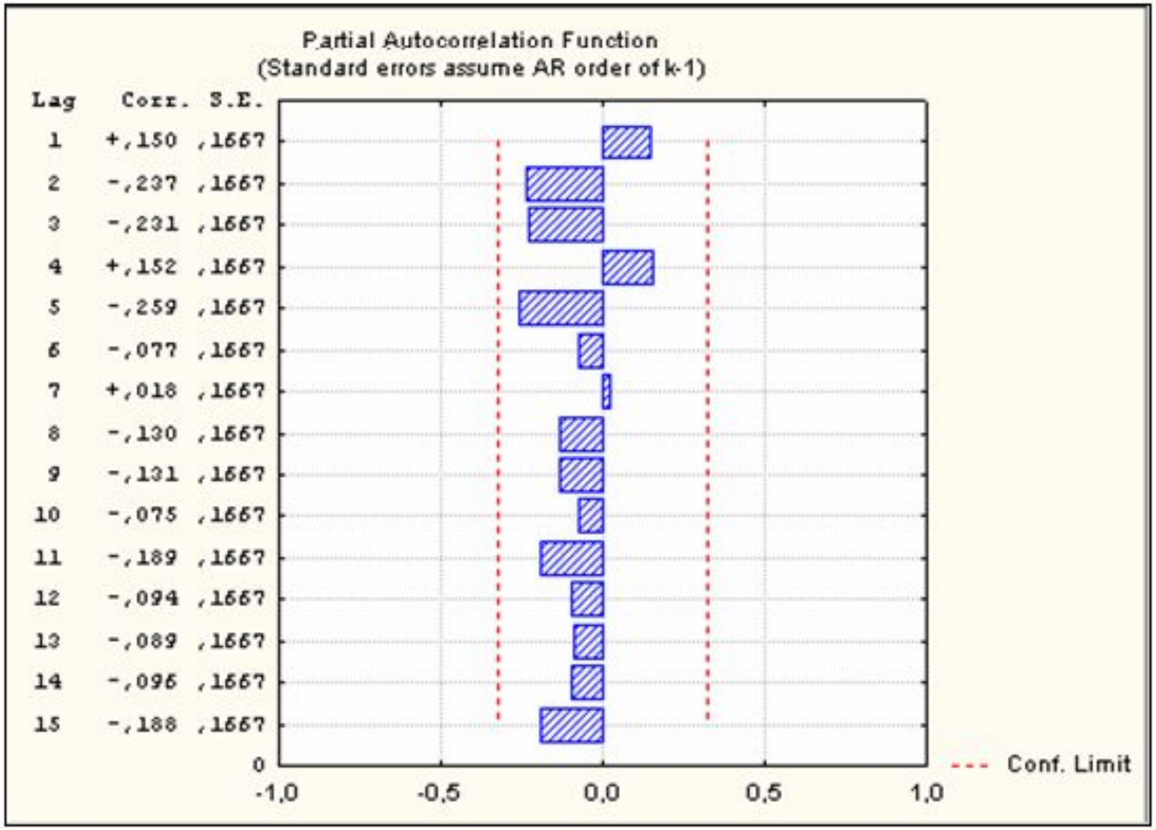
Оценка частной автокорреляционной функции:

$$r_{\text{частн}}(\tau) = r_{t,t-\tau/(t-1,t-2,\dots)} = -\frac{A_{t,t-\tau}}{\sqrt{A_{tt}A_{t-\tau,t-\tau}}}$$

где $A_{t,t-\tau}$ - алгебраические дополнения матрицы R

$$R = \begin{pmatrix} 1 & r(1) & \dots & r(\tau) \\ r(1) & 1 & \dots & r(\tau-1) \\ \dots & \dots & \dots & \dots \\ r(\tau) & r(\tau-1) & \dots & 1 \end{pmatrix}$$

Пример коррелограммы частной автокорреляционной функции (АКФ) в ППП Statistica



Простейшие методы прогнозирования

- прогнозирование в предположении абсолютной неизменности значений предшествующих уровней в будущем;
- метод среднего уровня ряда;
- метод абсолютного прироста;
- метод среднего темпа прироста.

Прогнозирование в предположении абсолютной неизменности значений предшествующих уровней

Прогнозирование в предположении абсолютной неизменности значений предшествующих уровней в будущем исходит из утверждения, что каждое следующее прогнозное значение будет равно предыдущему значению признака, то есть

$$\hat{y}^*_{t+L} = \hat{y}^*_{t+L-1},$$

где t - количество периодов времени,

$\hat{y}^*_{t+L}$ - прогнозное значение на период упреждения,

$\hat{y}^*_{t+L-1}$ - прогнозное значение, предшествующего периоду упреждения.

Метод прогнозирования на основе среднего уровня ряда

Метод прогнозирования на основе среднего уровня ряда используется для тех случаев, когда изменение уровней временных рядов носит стационарный характер

$$\hat{y}^{*}_{t+L} = \bar{y}, \quad \bar{y} = \frac{1}{T}\sum_{t=1}^{T} y_t .$$

Интервальный прогноз:

$$\hat{y}^{*}_{t+L} = \bar{y} \pm t_{\alpha} \cdot \sigma_{\bar{y}}, \quad \sigma_{\bar{y}} = \frac{\sigma_y}{\sqrt{T}},$$

где t_{α} - табличное значение критерия Стьюдента с $\nu = T - 1$ и уровнем значимости α,

σ_y - среднее квадратическое отклонение,

T - число уровней ряда.

Метод среднего абсолютного прироста

Метод среднего абсолютного прироста предполагает, что общая тенденция развития изучаемого социально экономического явления наилучшим образом аппроксимируется линейной формой аналитического выражения и выполняются следующие условия:

1. абсолютные приросты приблизительно одинаковы;
2. выполняется неравенство $\sigma^2_{ост} \le \rho^2$

где $\sigma^2_{ост} = \frac{1}{T}\sum\left(y_t - \bar{y}_{\bar{\Delta}}\right)^2$, $\bar{y}_{\bar{\Delta}}$ - теоретические значения уровней ряда, выровненные методом среднего абсолютного прироста;

$$\rho^2 = \frac{1}{2T}\sum_{t=1}^{T}\Delta_t^2 .$$

Общая модель прогноза имеет вид

$$\hat{y}^*_{t+L} = y_t + \bar{\Delta}_y \cdot L, \qquad \bar{\Delta}_t = \frac{y_t - y_1}{T-1} = \frac{\sum \Delta_{y_t}}{T-1}$$

где y_t - последний уровень исходного временного ряда или уровень, принятый за базу,

L - период упреждения,

$\bar{\Delta}_y$ - средний абсолютный прирост.

Метод среднего темпа роста

Метод среднего темпа роста применяется в случае, если цепные темпы роста $T_{y_t} = \frac{y_t}{y_{t-1}}$, рассчитанные по данным исходного ряда динамики за исследуемый период времени, имеют примерно одинаковые цифровые значения, а тенденция развития явления подчиняется геометрической прогрессии и может быть описана показательной кривой. Модель прогноза методом среднего темпа роста имеет вид

$$\hat{y}^*_{t+L} = y_t \bar{T}^L_y, \qquad \bar{T}^L_y = \sqrt[T-1]{\frac{y_T}{y_1}} = \sqrt[n-1]{\prod_{t=1}^{T-1} T_{y_t}},$$

где y_t - последний уровень исходного временного ряда или уровень, принятый за базу,

$\bar{T}^L_y$ - средний темп роста.

Применение простейших методов прогнозирования

- при краткосрочном прогнозировании
- при малом объем выборки
- при равномерном увеличении или уменьшении значений признака

Показатели точности

В качестве характеристики точности моделей могут быть использованы следующие показатели:

1) Коэффициент детерминации;

2) Средняя ошибка аппроксимации: $A = \frac{1}{T}\frac{\sum_{t=1}^{T}\left|y_t - \hat{y}_t\right|}{y_t} * 100\%$

Если A не превышает 8-10% модель считается обладающей хорошими статистическим свойствами.

3) Коэффициент несоответствия: $KH_3 = \sqrt{\frac{\sum_{t=T_1}^{T}\left(\hat{y}_t - y_t\right)^2}{\sum_{t=T_1}^{T} y_t^2}}$

4) Средняя квадратическая ошибка прогноза: $\sigma = \sqrt{\frac{\sum_{t=T_1}^{T}(y_t - \hat{y}_t)^2}{T-(T_1-1)}}$,

Для оценки модели используются наблюдения от 1 до T_1-1, а для определения ее прогностической способности – наблюдения от T_1 до T, то есть объем контрольной выборки составляет T-(T_1-1).

Показатели точности

5) средняя абсолютная ошибка:

$$d = \frac{1}{T - (T_1 - 1)} \sum_{t=T_1}^{T} \left| y_t - \hat{y}_t \right|$$

6) средняя абсолютная процентная ошибка:

$$d^* = \frac{100}{T - (T_1 - 1)} \sum_{t=T_1}^{T} \frac{\left| y_t - \hat{y}_t \right|}{y_t}$$

Недостаток этого показателя: поскольку в знаменателе дроби находится y_t, то общий результат будет изменяться в зависимости от того, что больше по абсолютной величине y_t или $\hat{y}_t$.

6) Для исправления вышеописанного недостатка была создана скорректированная версия этого показателя:

$$ad = \frac{100}{T - (T_1 - 1)} \sum_{t=T_1}^{T} \frac{\left| y_t - \hat{y}_t \right|}{\left| y_t + \hat{y}_t \right|}$$

Показатели точности

7) U- статистика Тейла. Данная статистика позволяет сравнивать прогностические способности моделей. Если значение статистики < 1 то анализируемая модель лучше по прогностическим способностям первоначальной «эталонной» модели, если > 1- то хуже, если =1, то анализируемая модель эквивалентна по прогностическим способностям первоначальной «эталонной» модели. Расчет статистики производится по формуле:

$$U = \sqrt{\sum_{t=T_1}^{T} \frac{\left(\dfrac{y_t - \hat{y}_t}{y_t} \right)^2}{\left(\dfrac{y_t - \hat{y}_t^{*}}{y_t} \right)^2}},$$

где $У_t$ - наблюдения по контрольной выборке;

$У_t$ - прогнозные значения по анализируемой модели;

$\hat{У}^{*}_{T+L}$ - прогнозные значения по предварительной «эталонной» модели.

Критерии наличия тренда, основанные на знаках разностей

Критерий «восходящих» и «нисходящих» серий

Критерий «восходящих» и «нисходящих» серий реализуется в виде следующей последовательности шагов:

- Определяется последовательность исходя из следующих условий:

$$\delta_i = \begin{cases} +, y_t - y_{t-1} > 0 \\ -, y_t - y_{t-1} < 0 \end{cases}$$

- подсчитывается ν_T - число серий в совокупности δ_i, где под серией понимается подряд идущие «+» или «-»
- определяется продолжительность самой длинной серии $\tau_{max}(T)$:
- проверка гипотез основана на том, что при условии случайности ряда протяженность самой длинной серии не должна быть слишком большой, а общее число серий не должно быть слишком маленьким.

$$\begin{cases} \nu(T) > [\frac{1}{3}(2T-1) - 1.96\sqrt{\frac{16T-29}{90}}] \\ \tau_{max}(T) \le \tau_0(T) \end{cases},$$

где $\tau_0(T)$ находится по таблице:

T	n<26	26-153	153-170
$\tau_0(T)$	5	6	7

Если не выполняется одно из условий данной системы, следовательно, H_0 отвергается, то есть существует трендовая составляющая.

Критерий серий основанный на медиане выборки

$$H_0 : My_t = const$$

$$H_1 : My_t \neq const$$

1. Определение Me ряда.
2. $y_t > y_{med}$ =>"+"

   $y_t < y_{med}$ =>"–"

   $y_t = y_{med}$ =>"0"
3. $K_{\max(n)} < [3{,}31 \cdot \ln(n+1)]$
4. $V(n) > [0{,}5 \cdot (n + 1 - 1{,}96 \cdot \sqrt{n-1})$

Если хотя бы одно из условий (3-4) нарушено, то гипотеза отвергается => в ряду есть тренд.

Исходные данные

	Yt
июн.06	85
июл.06	76
авг.06	107
сен.06	94
окт.06	120
ноя.06	103
дек.06	126
янв.07	88
фев.07	80
мар.07	142
апр.07	181
май.07	166
июн.07	139
июл.07	94
авг.07	110
сен.07	127
окт.07	208
ноя.07	171
дек.07	245
янв.08	160
фев.08	180
мар.08	211
апр.08	257
май.08	240
июн.08	244

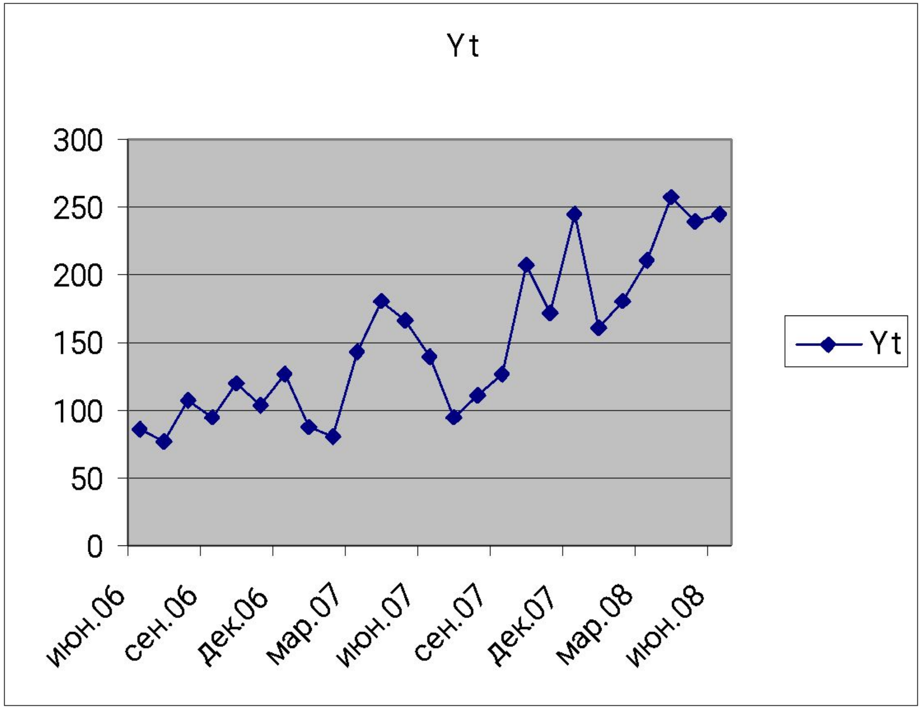


Рисунок 1 – График временного ряда

Критерий серий основанный на медиане выборки

Дата	Yt	Yme
июн.06	76	-
июл.06	80	-
авг.06	85	-
сен.06	88	-
окт.06	94	-
ноя.06	94	-
дек.06	103	-
янв.07	107	-
фев.07	110	-
мар.07	120	-
апр.07	126	-
май.07	127	-
июн.07	139	0
июл.07	142	+
авг.07	160	+
сен.07	166	+
окт.07	171	+
ноя.07	180	+
дек.07	181	+
янв.08	208	+
фев.08	211	+
мар.08	240	+
апр.08	244	+
май.08	245	+
июн.08	257	+

$$H_0 : My_t = const$$

$$H_1 : My_t \neq const$$

$$n = 25$$

$$Me = 139$$

$K_{\max(n)} < [3{,}31 \cdot \ln(n+1)]$ -длина серии

$$K_{\max(25)} = 12$$

$$K_{\max(25)} > [8{,}49]$$

$V(n) > [0{,}5 \cdot (n+1-1{,}96 \cdot \sqrt{n-1})]$ -число серий

$$V(25) = 2$$

$$V(25) < [8{,}2]$$

Гипотеза H_0 отвергается → есть тренд

Проверка гипотез об отсутствии сезонности

Критерий «пиков и ям»

y_j , j=1,...,T $y_{k-1}<y_k$, $y_{k+1}<y_k$ -пик,

$y_{k-1}>y_k$, $y_{k+1}>y_k$ -яма

Для трех последовательных значений y_i, y_{i+1}, y_{i+2}, $i=1,\ldots,(T-2)$ определим

$$x_i = \begin{cases} 1, & \text{если среди них есть экстремальная точка} \\ 0 & \text{иначе} \end{cases}$$

Тогда число экстремальных точек:

$$e = \sum_{i=1}^{T-2} x_i$$

H_0: случайный характер временного ряда, нет сезонности

H_1: присутствие сезонности

Алгоритм проверки гипотезы:

1. Определяется число e экстремальных точек временного ряда по T имеющимся в распоряжении значениям.
2. Вычисляется t- статистика

$$t = \frac{3e - 2T + 4}{\sqrt{16T - 29}}\sqrt{10}$$

В случае справедливости нулевой гипотезы статистика t распределена по стандартному нормальному закону. Расчетное значение t сравнивается с двусторонней критической точкой стандартного нормального распределения.

Если $|t| < t_{\frac{\alpha}{2}}$ то гипотеза H_O не отвергается.