

Лекция 6. Эконометрика ВРЕМЕННЫХ РЯДОВ

Вопросы лекции:

1. Понятие временного ряда
2. Трендовые модели генерации значений временного ряда
3. Фильтрация и сглаживание временного ряда
4. Прогнозирование и его виды
5. Сезонная компонента в прогнозировании
6. Модели авторегрессии
7. Стационарные ряды. Модель авторегрессии и проинтегрированного скользящего среднего

Вопрос 1. Понятие временного ряда

Эконометрические модели, характеризующие протекание процесса во времени или состояние одного объекта в последовательные моменты времени (или периоды времени), представляют **модели временных рядов**.

Временным рядом называется последовательность значений признака (набор) данных $x_1, \dots, x_n$, принимаемых в течение нескольких последовательных моментов времени или периодов.

Каждая единица статистического материала называется **измерением** или **отсчётом**, также допустимо называть его **уровнем** на указанный с ним момент времени. Во временном ряде каждому отчету должно быть указано время измерения или номер измерения по порядку.

Временной ряд существенно отличается от простой выборки данных, так как при анализе учитывается взаимосвязь измерений со временем, а не только статистическое разнообразие и статистические характеристики выборки.

Таблица 6.1 - Пример временного ряда

	2006	2007	2008	2009	2010	2011	2012
Номинальный ВВП (млрд. руб)	26917,20	33247,50	41428,60	38808,70	44491,40	55798,70	62356,90

Между уровнями временного ряда, или ряда динамики, может иметься зависимость. Значения каждого последующего уровня ряда зависят от предыдущих. Корреляционную зависимость между последовательными уровнями ряда динамики называют **автокорреляцией уровней ряда**.

Существует две основные цели анализа временных рядов: определение природы ряда и прогнозирование.

Когда возникает необходимость предсказания будущих значений в регистрируемой последовательности, адекватным является понятие временного ряда (модель).

Предсказание будущих значений ряда может обоснованно осуществить только, если будет выявлено некоторая тенденция их поведения. Именно в силу наличия тенденции, оказывается важным порядок следствия значений временного ряда.



Рис.6.1. Пример определения тенденции поведения временного ряда

Таким образом, важнейшей задачей анализа временного ряда является: определение схемы генерации их значений, которые описывают искомые тенденции.

Вопрос 2. Трендовые модели генерации значений временного ряда.

Тренд – неслучайная функция времени, которая описывает основную тенденцию поведения временного ряда.

Обычное математическое представление трендовых моделей имеет вид:

$$x_t = f(t) + \varepsilon(t) \quad (6.1)$$

где $f(t)$ – тренд; $\varepsilon(t)$ – «невязка», характеризующая неточность совпадения значений временного ряда со значениями тренда.

Предполагаем, что ε_t – имеет независимые значения.



Рис.6.2. Пример построения тренда для временного ряда

Задача заключается в определении $f(t)$ – как некой функции от времени, так как в этом случае появляется возможность осуществить предсказание значений временного ряда за пределами его интервалов регистрации.

$$\hat{x}_{n+k} = f(t_n + t_k) \quad (6.2)$$

Возникает проблема определения вида функции f , которая наилучшим образом позволяет экстраполировать данные за пределы их регистрации.

Иногда (чаще всего) в качестве меры отклонения временного анализа от гипотетического тренда используется оценка среднеквадратического отклонения

Основным принципом подбора функций тренда является принцип наименьших квадратов. **лекция 5 вопрос 2**

Для реализации этого принципа необходимо задавать конкретный вид функциональной зависимости $f(t)$. **лекция 1 вопрос 3**

Для разных временных рядов вид этой функциональной зависимости – разный, поэтому прежде чем переходить к минимизации стандартного отклонения, необходимо провести разведочный анализ данных, которые могут дать ответ о действительном наличии некоторого тренда, и о приблизительном виде соответствующей функциональной зависимости. Например,

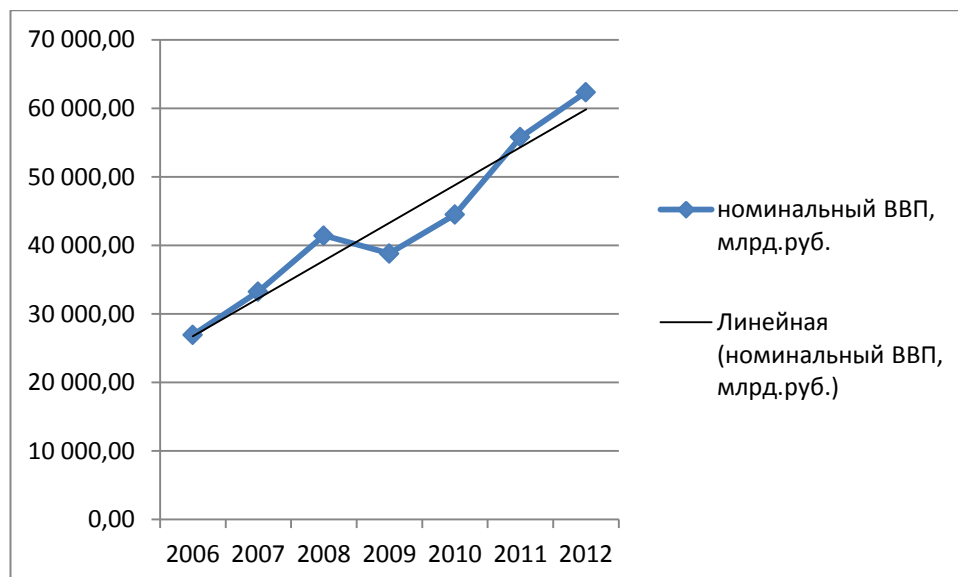


Рис.6.3. Тренд временного ряда динамики ВВП

В данном случае неровная линия – это линия временного ряда, то есть реальные значения наблюдаемых показателей за определенные периоды времени (например, годы), которые откладываются на оси x , а по оси y отложены наблюдаемые значения показателя. Функция $f(t)$ (линейная) является линией тренда для данного временного ряда и позволяет экстраполировать данные для будущих временных промежутков.

Вопрос 3. Фильтрация и сглаживание временного ряда

3.1. Медианная фильтрация (сглаживание)

Если тренд является монотонным (устойчиво возрастает или устойчиво убывает), то анализировать такой ряд обычно нетрудно. Если временные ряды содержат значительную ошибку, то первым шагом выделения тренда является сглаживание.

Фильтрация – это удаление чего-то, что является ненужным или разделение на компоненты.

Сглаживание - это процедура исключения или уменьшения размаха колебаний значений временного ряда.

Сглаживание всегда включает некоторый способ локального усреднения данных, при котором несистематические компоненты взаимно погашают друг друга.

Самый общий метод сглаживания - скользящее среднее, в котором каждый член ряда заменяется простым или взвешенным средним n соседних членов, где n - ширина "кармана". Вместо среднего можно использовать медиану значений. Если в данных имеются выбросы (связанные, например, с ошибками измерений), то сглаживание медианой обычно приводит к более достоверным линиям тренда, по сравнению со скользящим средним.

Основной недостаток медианного сглаживания в том, что при отсутствии явных выбросов, он приводит к менее гладким и достоверным линиям тренда, а также не позволяет использовать веса.

Медианная фильтрация получила развитие в последние 20 лет.

Y_{med} – является медианой последовательности из $2m+1$ значений ($y_{k-1}, \dots, y_k, y_{k+1}$), если выполняется :

$$Y_{med} = Z(k), \quad (6.3)$$

Где

$Z(i) \in (y_{k-m}, \dots, y_{k+m}), i=1 \dots 2m+1$

$Z(1) \leq Z(2) \leq \dots \leq Z(2m+1)$

То есть последовательность $Z(i)$ представляет собой упорядоченную по возрастанию последовательность значения “у”

Пример: пусть $m=1$, тогда $2m+1 = 3$

Последовательность значений у следующая: 0.1; -3; 4

Тогда

Z_1	Z_2	Z_3
-	0,1	4

Тогда $Y_{med} = 0.1$

В зависимости от значения $2m+1$, $m=1$ – медиана по тройкам, $m=2$ – медиана по пятеркам и т.д.

В рассмотрение вовлекаются: отрезок исходного временного ряда, содержащий $2m+1$ подряд идущих значений. Процедура фильтрации осуществляется для всех значений временного ряда, кроме первого и последнего.

Эта фильтрация позволяет исключить выбросы, то есть значения временного ряда, которые «сильно» отличаются от остальных. При этом эти выбросы заменяются соседними значениями той же последовательности. Количество вовлекаемых в рассмотрение значений исходного временного ряда называется **апертурой фильтра**.

При апертуре $2m+1$ можно исключить m – подряд идущих выбросов.

Медианная фильтрация не искажает тенденции монотонного возрастания или убывания отрезков временного ряда.

3.2. Фильтрация методом скользящего среднего

В основном методы алгоритмического сглаживания – это различные модификации скользящего среднего.

В основе этих методов лежит следующий постулат: если индивидуальный разброс значений временного ряда $x(t)$ вокруг тренда $f(t)$ характеризуется дисперсией σ^2 , то разброс среднего из n членов временного

ряда $(x(1) + x(2) + \dots + x(n))/n$ около того же значения будет характеризоваться дисперсией σ^2/n .

Алгоритм скользящего среднего заключается в следующем:

$$\hat{f}(t) = \sum_{k=-m}^m w_k x(t + k), \quad (6.4)$$

Где $t=m+1, m+2, \dots, n-m$,

w_k – некоторые весовые коэффициенты, причем

$$\sum_{k=-m}^m w_k = 1 \quad (6.5)$$

Поскольку, диапазон суммирования (при изменении t от значения $m+1$ до значения $n-m$) скользит по временному ряду (при переходе от t к $t+1$ в составе слагаемых происходит замена слагаемого $x(t - m)$ слагаемым $x(t + m + 1)$), то данный метод называется методом скользящего среднего (МСС).

Когда ошибка измерения очень большая, используется метод сглаживания методом наименьших квадратов, взвешенных относительно расстояния или метод отрицательного экспоненциально взвешенного сглаживания. Все эти методы отфильтровывают шум и преобразуют данные в относительно гладкую кривую. Ряды с относительно небольшим количеством наблюдений и систематическим расположением точек могут быть сглажены с помощью бикубических сплайнов.

МСС образуют группу методов, отличаясь друг от друга параметрами m и w_k .

Метод простого скользящего среднего исходит из простого предположения, что следующий во времени показатель по своей величине равен средней, рассчитанной за последние три месяца.

Пример практического применения метода скользящего среднего в экономических расчетах

Выполнить выравнивание динамического ряда и определить прогнозное значение, исходя из имеющихся данных о динамике цены квартир в Ивановской области. Данные приведены в таблице 6.2.

Таблица 6.2– Исходные данные о динамике цены 1м² жилья в Ивановской области

год	квартал	Цена, руб.
2011	I квартал	40448,94
	II квартал	37664
	III квартал	36223,04
	IV квартал	36558,71
2012	I квартал	37261,51
	II квартал	36996,2

	<i>III квартал</i>	35727,18
	<i>IV квартал</i>	36777,6
2013		прогноз

Выравнивание проводится путем нахождения средних значений трех предыдущих временных периодов. Поэтому выравнивание начинается с 3 квартала 2010года.

Прогнозное значение цены для 3 квартала равно $(40448,94+37664+36223,04)/3 = 38111,99$, значение цены для 4 квартала $(37664+36223,04+36558,71)/3$ и так далее.

Расчет в третьем столбце таблицы 6.3.

Таблица 6.3 – Расчет методом скользящего среднего

год	квартал	Цена, руб.	Цена МСС, руб
2011	<i>I квартал</i>	40448,94	
	<i>II квартал</i>	37664	
	<i>III квартал</i>	36223,04	38111,99
	<i>IV квартал</i>	36558,71	36815,25
2012	<i>I квартал</i>	37261,51	36681,08
	<i>II квартал</i>	36996,2	36938,80
	<i>III квартал</i>	35727,18	36661,63
	<i>IV квартал</i>	36777,6	36500,32
2013			36252,39

Реализация метода скользящего среднего может быть проведена с помощью специальных компьютерных программ, начиная с электронных таблиц Microsoft Excel. *практич задание*

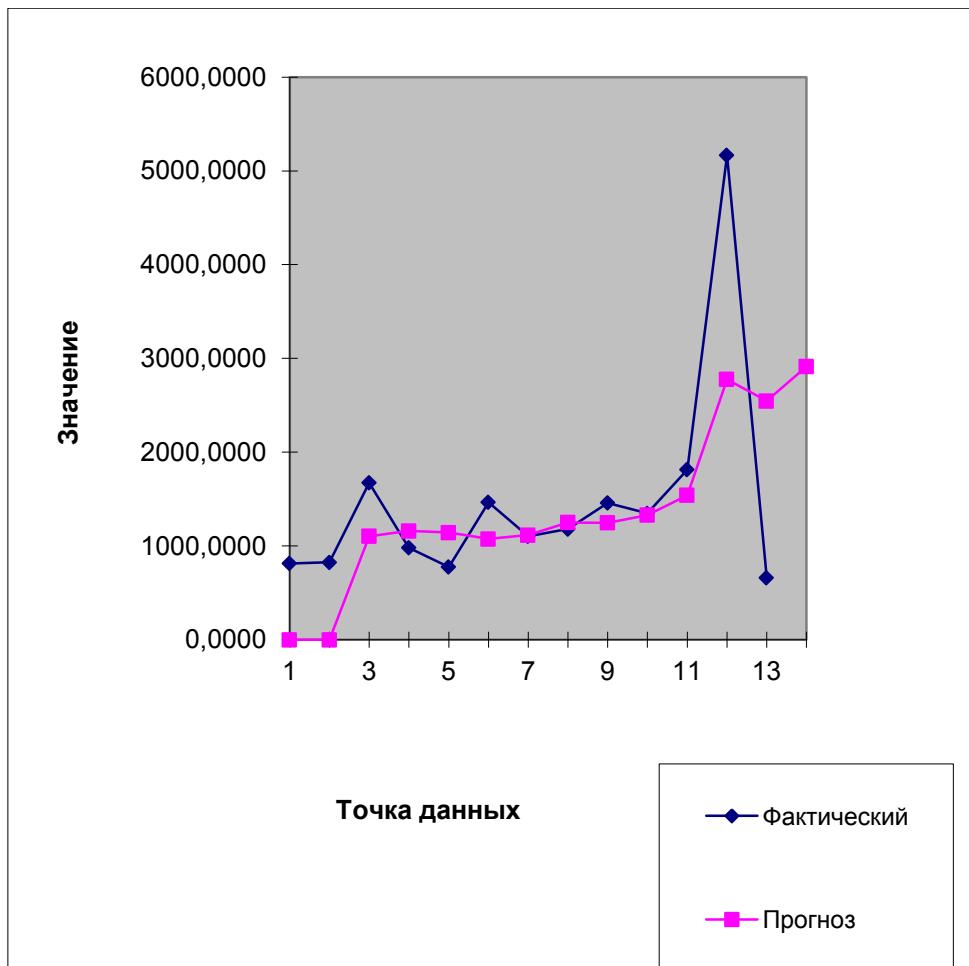


Рис.6.4. Пример прогноза методом скользящего среднего в Microsoft Excel объема ввода в действие жилья, тыс.м²

3.3. Метод экспоненциально взвешенного скользящего среднего (метод Брауна)

Методы скользящего среднего основываются на том, что все значения временного ряда имеют одинаковую информационную ценность. Однако в задачах прогноза, в которых сглаженная функция $f(t)$ используется обычно для формирования прогнозов на несколько тактов вперёд, недавние значения $x(t)$ очевидно ценнее, чем значения ряда в далёком прошлом, так как ряд далее будет вести себя так, какова сформировавшаяся тенденция в настоящем и недалёком прошлом. Этот постулат лег в основу метода экспоненциально взвешенного скользящего среднего Брауна.

$$\hat{f}(t) = \frac{1 - \alpha}{1 - \alpha^t} \sum_{k=0}^{t-1} \alpha^k x(t - k) \quad (6.6)$$

$$\hat{f}(t) = \frac{1 - \alpha}{1 - \alpha^t} (x(t) + \alpha x(t - 1) + \alpha^2 x(t - 2) + \dots + \alpha^{t-1} x(1)) \quad (6.7)$$

$$0 < \alpha < 1 \quad (6.8)$$

Таким образом, значения временного ряда тоже сглаживаются, однако существуют следующие отличия от обычного МСС:

- 1) скользит только правый край интервала усреднения (левый край закреплён в точке $t = 1$)
- 2) веса при $x(t - k)$ экспоненциально уменьшаются по мере «удаления в прошлое»,
- 3) формула даёт оценку сглаженного значения временного ряда не в средней, а в правой, конечной точке усреднения,
- 4) нет проблемы крайних значений.

Метод экспоненциального сглаживания представляет прогноз показателя на будущий период в виде суммы фактического показателя за данный период и прогноза на данный период, взвешенных при помощи специальных коэффициентов.

Представим, что составляется прогноз определенной экономической величины на следующий месяц. Тогда:

$$\hat{D}_{i+1} = a\tilde{D}_i + [1-a]*\hat{D}_i \quad (6.9)$$

где

$\hat{D}_{i+1}$ – прогноз на месяц $i+1$;

X_m – значение исследуемой величины в месяце (фактические данные);

$\hat{D}_i$ – прогноз на месяц i ,

a – специальный коэффициент, определяемый статистическим путем.

Фактором затухания $(1-a)$ называется корректировочный фактор, минимизирующий нестабильность данных генеральной совокупности. Значение фактора по умолчанию равно 0,3.

Для константы сглаживания наиболее подходящими являются значения от 0,2 до 0,3. Эти значения показывают, что ошибка текущего прогноза установлена на уровне от 20 до 30 процентов ошибки предыдущего прогноза. Более высокие значения константы ускоряют отклик, но могут привести к непредсказуемым выбросам. Низкие значения константы могут привести к большим промежуткам между предсказанными значениями.

Пример практического применения метода скользящего среднего в экономических расчетах

Составить прогноз цены жилья методом экспоненциального сглаживания. Исходные данные приведены в таблице 6.4.

Таблица 6.4– Исходные данные для прогноза цены жилья, руб/м²

год	квартал	цена, руб.
2010	I	
	квартал	5713,59
	II	5684,38

	<i>квартал</i>	
	<i>III</i>	
	<i>квартал</i>	6190,21
	<i>IV</i>	
2011	<i>квартал</i>	6879,61
	<i>I</i>	
	<i>квартал</i>	7793,15
	<i>II</i>	
2012	<i>квартал</i>	7629,06
	<i>III</i>	
	<i>квартал</i>	8073,86
	<i>IV</i>	
2013	<i>квартал</i>	9753,7
	<i>I</i>	
	<i>квартал</i>	10355,29
	<i>II</i>	
2014	<i>квартал</i>	10876,03
	<i>III</i>	
	<i>квартал</i>	11251,26
	<i>IV</i>	
2015	<i>квартал</i>	12407,03
	<i>I</i>	
	<i>квартал</i>	
	<i>II</i>	

Специальный коэффициент α стандартно принимается равным 0,7.
Тогда решение приведено в таблице 6.5.

Таблица 6.5 – Расчет методом экспоненциального сглаживания

<i>год</i>	<i>квартал</i>	<i>цена, руб.</i>	<i>Прогноз, руб.</i>
2010	<i>I</i>	5713,59	
	<i>квартал</i>		
	<i>II</i>	5684,38	5713,59
	<i>квартал</i>		
2011	<i>III</i>	6190,21	5693,143
	<i>квартал</i>		
	<i>IV</i>	6879,61	6041,0899
	<i>квартал</i>		
2012	<i>I</i>	7793,15	6628,05397
	<i>квартал</i>		
	<i>II</i>	7629,06	7443,621191
	<i>квартал</i>		
2013	<i>III</i>	8073,86	7573,428357
	<i>квартал</i>		
	<i>IV</i>	9753,7	7923,730507
	<i>квартал</i>		
2014	<i>I</i>	10355,29	9204,709152
	<i>квартал</i>		

II квартал	10876,03	10010,11575
III квартал	11251,26	10616,25572
IV квартал	12407,03	11060,75872
		13998,21954

Например, прогноз цены на 3 квартал 2010 года $=0,7*5684,38 + 0,3*5713,59 = 5693,143$

И т.д.

Прогноз на 1 квартал 2013 года $=0,7*12407,03 + 0,3*11060,759 = 13998,22$

Реализация метода экспоненциального сглаживания может быть проведена с помощью специальных компьютерных программ, начиная с электронных таблиц Microsoft Excel. *практич задание*

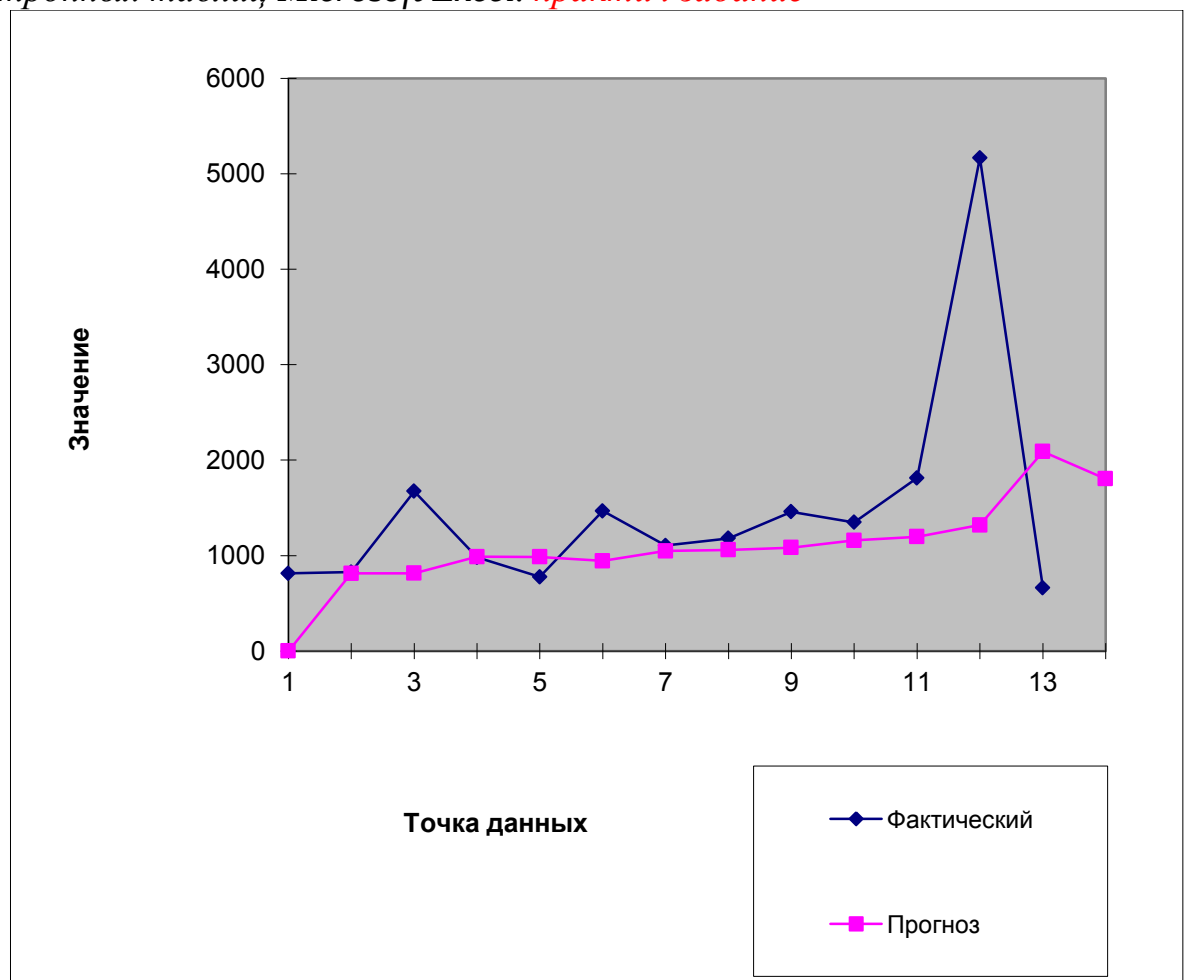


Рис.6.5. Пример прогноза методом экспоненциального сглаживания в Microsoft Excel объема ввода в действие жилья, тыс.м²

Самый прямой способ оценки прогноза, полученного на основе определенного значения α - построить график наблюдаемых значений и прогнозов на один шаг вперед.

Вопрос 4. Прогнозирование и его виды

Говоря о динамических рядах и методах их построения, необходимо уточнить само понятие прогнозирования и методы его осуществления.

Прогнозирование – это научно обоснованные суждения о возможном состоянии и будущем развитии того или иного объекта или отдельных его элементов, а также об альтернативных путях, методах, сроках достижения определённого состояния этого объекта.

Принципами прогнозирования являются:

1. *системный подход;*
2. *научная обоснованность;*
3. *альтернативность и вариантность;*
4. *выделение генеральной цели, ведущего звена;*
5. *отражение объективных закономерностей;*
6. *достоверность.*

Прогнозирование и планирование экономики представляет собой сложный многоступенчатый и итеративный процесс, в ходе которого должен решаться обширный круг различных социально-экономических и научно-технических проблем, для чего необходимо использовать в сочетании самые разнообразные методы.

Метод прогнозирования – это способ исследования объекта прогнозирования, направленный на разработку прогноза. Совокупность специальных правил, приемов и методов составляет методику прогнозирования.

Классификационная схема методов прогнозирования

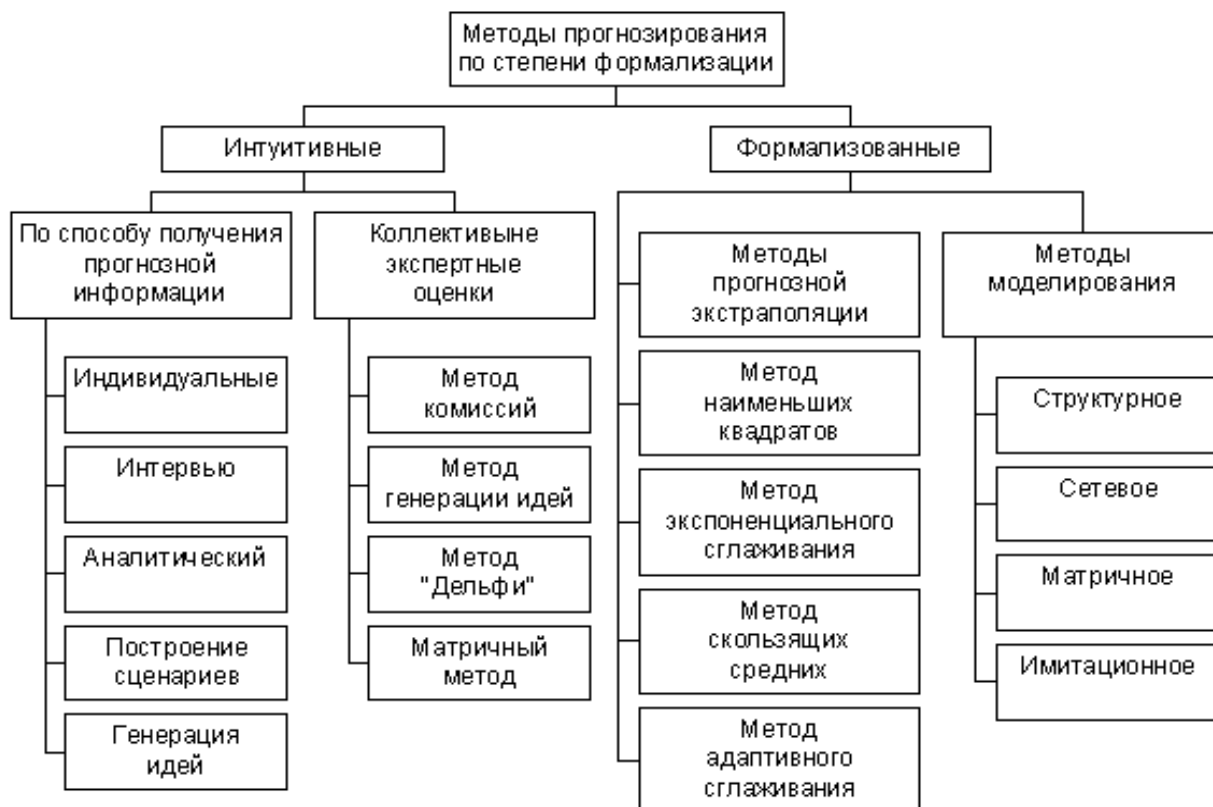


Рис.6.6. Классификация методов прогнозирования

По степени формализации методы экономического прогнозирования можно подразделить на интуитивные и формализованные. Среди интуитивных методов широкое распространение получили методы экспертных оценок. Применяются также методы исторических аналогий и прогнозирования по образцу.

К формализованным методам относятся методы экстраполяции и методы моделирования.

Методы моделирования предполагают использование в процессе прогнозирования и планирования различного рода экономико-математических моделей, представляющих собой формализованное описание исследуемого экономического процесса (объекта) в виде математических зависимостей и отношений. Различают следующие модели: матричные, оптимального планирования, экономико-статистические (трендовые, факторные, эконометрические), имитационные, принятия решений.

В зависимости от уровня управления экономическими и социальными процессами различают макроэкономические, межотраслевые, межрайонные, отраслевые, региональные модели и модели микроуровня (модели развития фирмы).

При разработке управленческих решений широкое распространение нашел метод сценариев, также дающий возможность оценить наиболее вероятный ход развития событий и возможные последствия принимаемых

решений. Разрабатываемые специалистами сценарии развития анализируемой ситуации позволяют с, тем или иным уровнем достоверности определить возможные тенденции развития, взаимосвязи между действующими факторами, сформировать картину возможных состояний, к которым может прийти ситуация под влиянием тех или иных воздействий. Профессионально разработанные сценарии позволяют более полно и отчетливо определить перспективы развития ситуации, как при наличии различных управляющих воздействий, так и при их отсутствии. С другой стороны, сценарии ожидаемого развития ситуации позволяют своевременно осознать опасности, которыми чреваты неудачные управленческие воздействия или неблагоприятное развитие событий.

В системе электронных таблиц Microsoft Excel существует несколько функций, с помощью которых можно выполнить прогнозирование.

На практике при выборе аналитической функции рекомендуется подбирать функцию $\tilde{y}_x$ с таким расчетом, чтобы ее конструктивные элементы, коэффициенты и константы поддавались экономической интерпретации, а линия тренда отображала наиболее характерное изменение признака.

Наиболее широко используются следующие функции:

- 1) линейная: $\tilde{y}_x = a + bx$, где a и b - константы;
- 2) параболическая: $\tilde{y}_x = a + bx + cx^2$, где a , b , c - константы;
- 3) степенная: $\tilde{y}_x = ax^b$;
- 4) экспоненциальная: $\tilde{y}_x = bt^x$;
- 5) гиперболическая: $\tilde{y}_x = a + b/x$;
- 6) простая модифицированная экспоненциальная функция: $\tilde{y}_x = a - b^x$
- 7) логистическая: $\tilde{y}_x = 1/(a + b)^x$.

Для визуального выбора наиболее точной аппроксимации можно воспользоваться построением точечной диаграммы исходных данных с последующим добавлением линии тренда.

Пример практической реализации простейших методов прогнозирования в практике экономических расчетов

Рассмотрим пример выбора аппроксимации данных статистики ВВП за ряд лет.

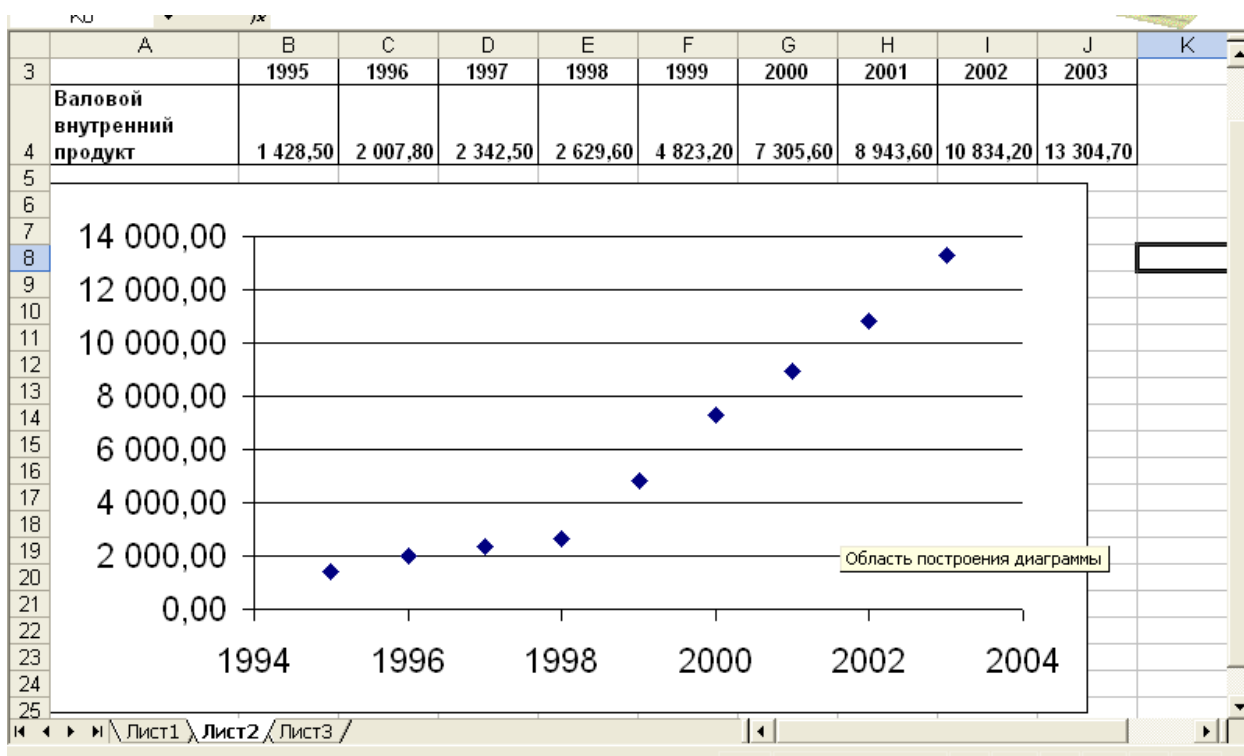


Рис. 6.7. Точечная диаграмма динамики ВВП

Существует возможность создания линии тренда на диаграмме без создания данных для линии тренда.

1. Выберите ряд данных, к которому нужно добавить линию тренда.
2. Выберите команду **Добавить линию тренда** в меню **Диаграмма**.
3. На вкладке **Тип** выберите нужный тип регрессионной линии тренда.

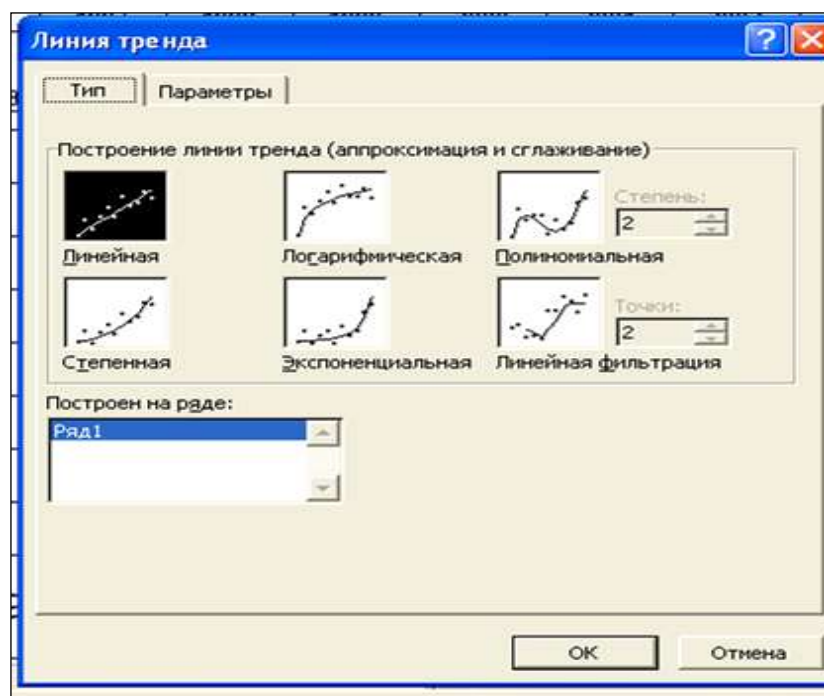


Рис. 6.8. Выбор линии тренда

Методом подбора определяется наиболее приемлемая форма линии тренда для аппроксимации фактических данных. Для аппроксимации фактических данных динамики ВВП выбраны линейная и экспоненциальная линии тренда (рис.6.9, 6.10).

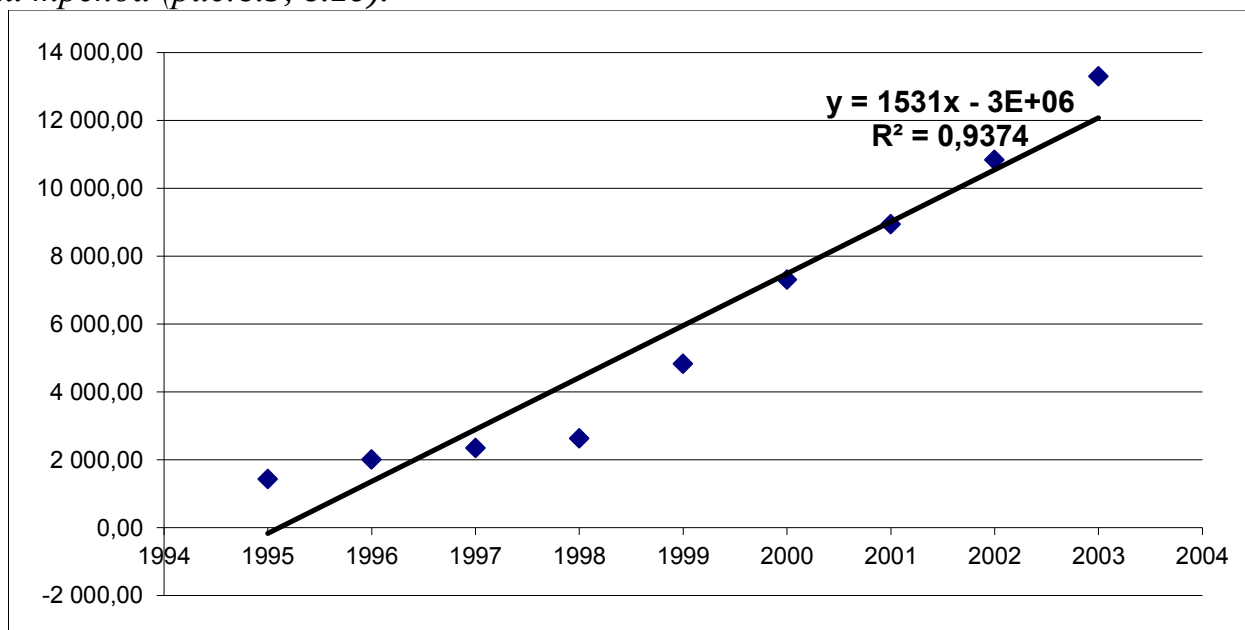


Рис. 6.9. Линейная аппроксимация динамики ВВП

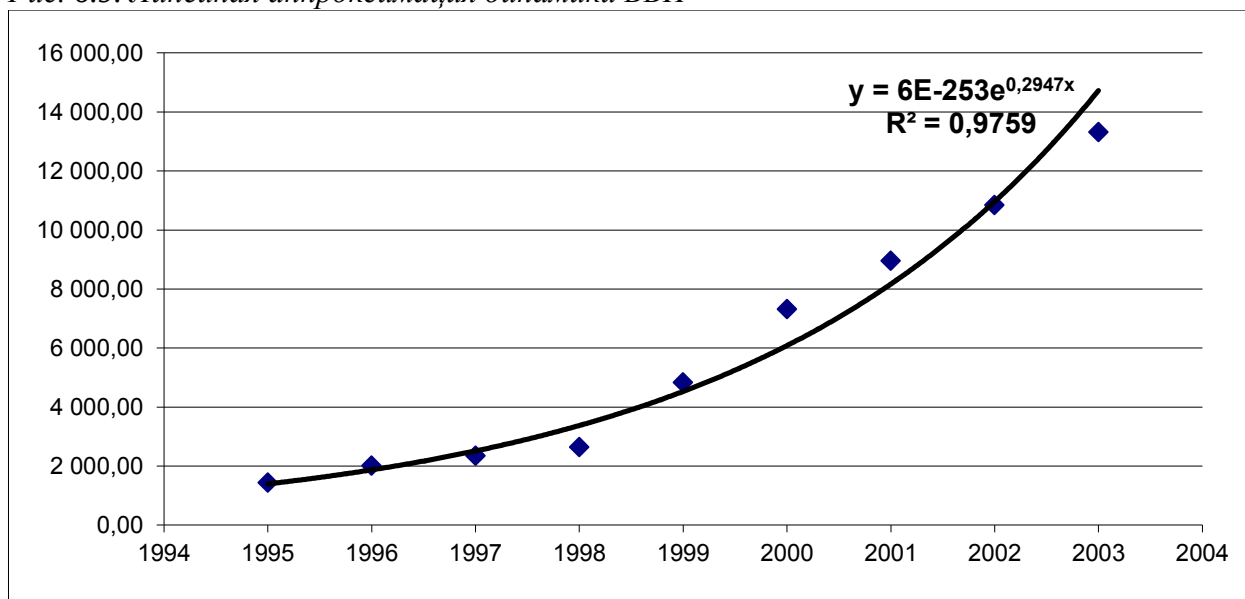


Рис. 6.10. Экспоненциальная линия тренда ряда динамики ВВП

Функция «Тенденция» возвращает значения в соответствии с линейным трендом, исходя из допущения о неизменности прогнозных тенденций в будущем.

Аппроксимирует прямой линией (по методу наименьших квадратов) массивы «известные значения y» и «известные значения x». Возвращает значения y, в соответствии с этой прямой для заданного массива «новые значения x».

Синтаксис

ТЕНДЕНЦИЯ (известные значения y; известные значения x; новые значения x; конст)

«Известные значения y » - это множество значений y , которые уже известны для соотношения $y = tx + b$.

«Известные значения x » - это необязательное множество значений x , которые уже известны для соотношения $y = tx + b$.

Массив «известные значения x » может содержать одно или несколько множеств переменных. Если используется только одна переменная, то «известные значения y » и «известные значения x » могут иметь любую форму, при условии, что они имеют одинаковую размерность. Если используется более одной переменной, то «известные значения y » должны быть вектором (то есть интервалом высотой в одну строку или шириной в один столбец).

Если «известные значения x » опущены, то предполагается, что это массив $\{1;2;3;\dots\}$ такого же размера как и «известные значения y ».

«Новые значения x » - это новые значения x , для которых ТЕНДЕНЦИЯ возвращает соответствующие значения y .

«Конст» - это логическое значение, которое указывает, требуется ли, чтобы константа b была равна 1.

- Если конст имеет значение ИСТИНА или опущено, то b вычисляется обычным образом.

- Если конст имеет значение ЛОЖЬ, то b полагается равным 0, и значения t подбираются таким образом, чтобы выполнялось соотношение $y = tx$.

Замечания

- Формулы, которые возвращают массивы, должны быть введены как формулы массива.

- При вводе константы массива для аргумента, такого как «известные значения x », следует использовать точку с запятой для разделения значений в одной строке и двоеточие для разделения строк.

Рассмотрим пример построения прогноза значений валового внутреннего продукта РФ. Исходные данные для построения прогноза заносят в систему электронных таблиц Excel (рис.6.13).

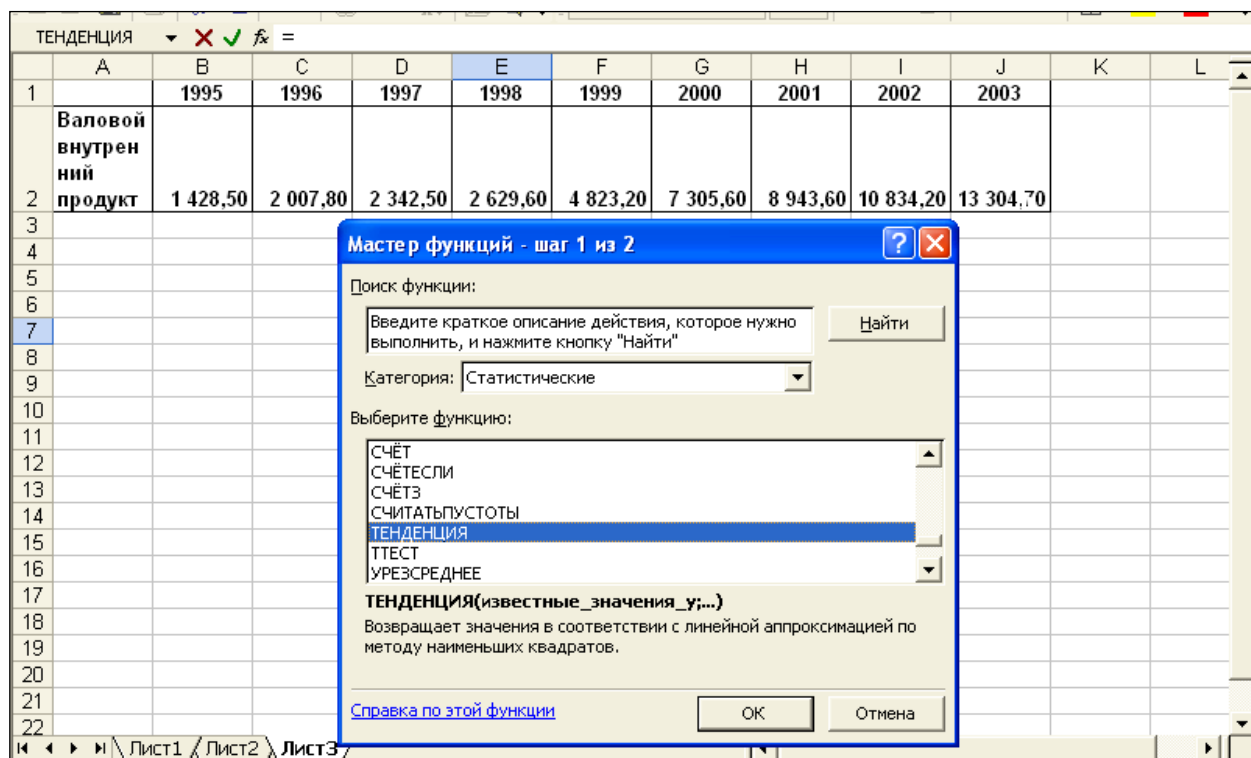


Рис. 6.13 Оформление исходных данных и выбор функции прогнозирования
Далее необходимо заполнить поля диалогового окна функции.

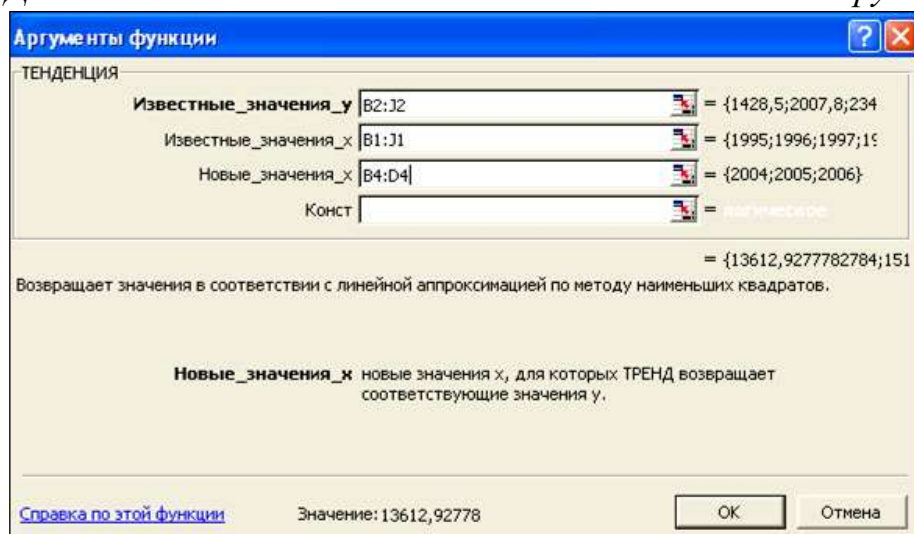


Рис.6.14. Диалоговое окно функции «Тенденция»

Формулу в этом примере необходимо ввести как формулу массива. Выделите диапазон C2:C13 или B15:B19, начиная с ячейки, содержащей формулу. Нажмите клавишу F2, а затем нажмите клавиши CTRL+SHIFT+ENTER.

	A	B	C	D	E	F	G	H	I	J	K
1		1995	1996	1997	1998	1999	2000	2001	2002	2003	
2	Валовой внутренний продукт	1 428,50	2 007,80	2 342,50	2 629,60	4 823,20	7 305,60	8 943,60	10 834,20	13 304,70	
3	прогноз(соответствующий ВВП)	-166,402	1364,634	2895,671	4426,708	5957,744	7488,781	9019,818	10550,85	12081,89	
4	прогнозВВП	2004	2005	2006							
5		13612,93	15143,96	16675							
6											
7											
8											
9											
10											
11											
12											
13											
14											
15											
16											
17											
18											
19											
20											
21											

Рис.6.15. Результаты прогнозирования объема ВВП с использованием линейной аппроксимации (функции «Тенденция»)

По такой системе действуют и другие функции прогнозирования в Microsoft Excel. Различие состоит в способах аппроксимации значений.

Функция «Лгрфприбл» в регрессионном анализе вычисляет экспоненциальную кривую, аппроксимирующую данные и возвращает массив значений, описывающий эту кривую.

Функция «Рост» рассчитывает прогнозируемый экспоненциальный рост на основании имеющихся данных. Функция РОСТ возвращает значения y для последовательности новых значений x , задаваемых с помощью существующих x - и y -значений. Функция рабочего листа РОСТ может применяться также для для аппроксимации существующих x - и y -значений экспоненциальной кривой.

Синтаксис

РОСТ(известные значения y ; известные значения x ; новые значения x ; конст)

«Известные значения y » - это множество значений y , которые уже известны для соотношения $\tilde{y}_x = b m^x$.

«Известные значения x » - это необязательное множество значений x , которые уже известны для соотношения $\tilde{y}_x = b m^x$.

Если «известные значения x » опущены, то предполагается, что это массив $\{1;2;3;\dots\}$ такого же размера как и «известные значения y ».

«Новые значения x » - это новые значения x , для которых РОСТ возвращает соответствующие значения y .

«Конст» - это логическое значение, которое указывает, требуется ли, чтобы константа b была равна 1.

- Если конст имеет значение ИСТИНА или опущено, то b вычисляется обычным образом.

- Если конст имеет значение ЛОЖЬ, то b полагается равном 1, а значения t подбираются так, чтобы $\hat{y}_x = t^x$.

Замечания

- Если какие-либо числа в массиве «известные значения y » равны 0 или отрицательны, то функция РОСТ возвращает значение ошибки #ЧИСЛО!.

- При вводе константы массива для аргумента, такого как «известные значения x », следует использовать точку с запятой для разделения значений в одной строке и двоеточие для разделения строк.

Рассмотрим пример построения прогноза значений валового внутреннего продукта РФ. В этом примере используются те же данные, что и в примере для функции «Тенденция». Диалоговое окно функции «Рост» выглядит следующим образом (рис.6.16).

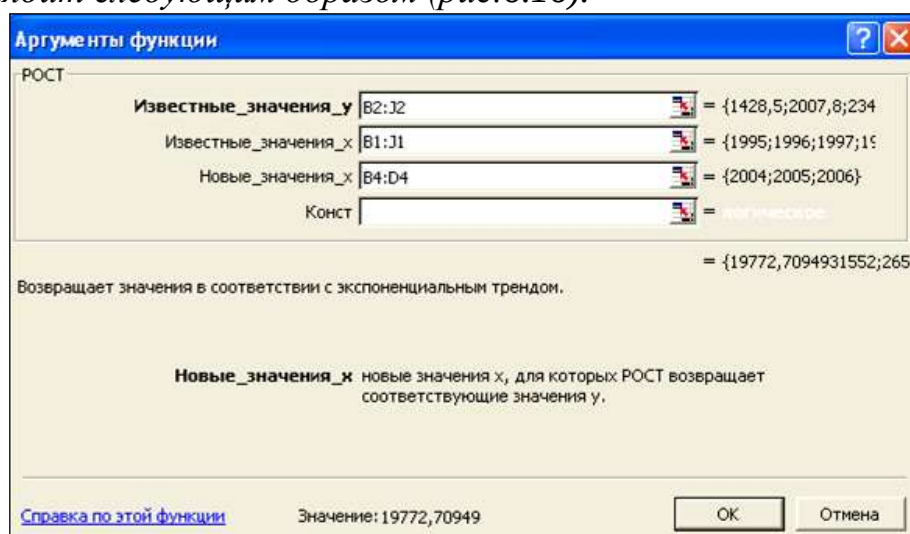


Рис. 6.16. Параметры диалогового окна функции «Рост»

Прогноз с использованием данной функции приведен на рис.6.17.

	A	B	C	D	E	F	G	H	I	J	K
1		1995	1996	1997	1998	1999	2000	2001	2002	2003	
2	Валовой внутренний продукт	1 428,50	2 007,80	2 342,50	2 629,60	4 823,20	7 305,60	8 943,60	10 834,20	13 304,70	
3	прогноз (соответств ующий ВВП)	1393,29199	1870,875	2512,161	3373,263	4529,527	6082,127	8166,918	10966,32	14725,28	
4	прогноз	2004	2005	2006							
5		19772,7095	26550,26	35650,98							
6											

Рис.6.17. Результаты прогнозирования объема ВВП с использованием экспоненциальной линии тренда (функции «Рост»)

Следует отметить, что значение y , предсказанное с помощью уравнения регрессии, может быть недостоверным, если оно находится вне диапазона значений y , которые использовались для определения уравнения.

Многофакторные модели позволяют одновременно учитывать воздействие нескольких факторов на уровень прогнозируемого показателя. Данные модели

используются при прогнозировании макроэкономических показателей, показателей спроса на продукцию, себестоимости, цен, прибыли и др.

Применение уравнений регрессии в прогнозировании связано, как правило, с предположением о том, что тенденции, сложившиеся в прошлом, в основном сохранятся и в будущем. Поскольку такой гарантии нет, к таким моделям следует относиться с большой осторожностью. Однако любые исследования, обращенные в будущее, исходят из информации о прошлом и настоящем.

Приведем пример построения прогноза оценки комплексного состояния региона (города) на основе данных статистики по регионам России на примере Центрального федерального округа.

Результаты обработки исходных данных приведены в таблице 6.6.
лекция 4 вопрос 2

Таблица 6.6 – Значения коэффициентов парной корреляции

	ВРП на душу населения	Объем инвестиций на душу населения	внешнеторгового оборота на душу населения	Финансовая обеспеченность	Доля занятых на малых предприятиях	Уровень безработицы	Соотношение	Доля населения	Общий объем розничного товарооборота	ОФ на душу населения	Коэфф. Энгеля	Социальный уровень
ВРП на душу населения	1,00											
Объем инвестиций на душу населения	0,66	1,00										
Объем внешнеторгового оборота на душу населения	0,79	0,87	1,00									
Финансовая обеспеченность региона на душу населения	0,87	0,84	0,96	1,00								
Доля занятых на малых предприятиях	0,63	0,72	0,87	0,86	1,00							
Уровень регистрируемой безработицы	-0,32	-0,48	-0,40	-0,38	-0,30	1,00						
Соотношение среднедушевых доходов и среднедушевого прожиточного минимума	-0,19	-0,09	-0,12	-0,16	-0,17	0,32	1,00					
Доля населения с доходами ниже прожиточного	-0,47	-0,40	-0,39	-0,47	-0,34	0,41	0,20	1,00				

минимума													
Общий объем розничного товарооборота и платных услуг на душу населения	0,65	0,89	0,94	0,90	0,87	-	0,27	0,00	-	0,35	1,00		
ОФ на душу населения	0,53	0,68	0,59	0,57	0,35	-	0,19	0,09	-	0,15	0,58	1,00	
Коэффициент плотности автомобильных дорог	0,58	0,76	0,63	0,65	0,45	-	0,32	0,30	-	0,51	0,66	0,51	1,00
Сводный показатель уровня развития отраслей социальной инфраструктуры	0,02	0,32	0,29	0,21	0,46	-	0,18	0,04	0,02	0,31	0,38	0,04	1,00
Комплексная оценка	0,85	0,79	0,84	0,89	0,72	-	0,48	0,16	-	0,62	0,73	0,58	0,30

Проведя статистический анализ данных Центрального федерального округа, можно сказать, что основными факторами, влияющими на комплексную оценку, являются валовой региональный продукт, объем инвестиций на душу населения, уровень безработицы, основные фонды на душу населения, общий объем розничного товарооборота, коэффициент плотности автодорог, обеспеченность социальной инфраструктурой. Остальные факторы обладают достаточной степенью корреляционной зависимости от вышеперечисленных, поэтому исключаются из рассмотрения.

Анализ формы взаимосвязи между факторами проводится с помощью функции «Регрессия» пункта меню «Сервис», «Анализ данных». Результаты обработки исходных данных приведены на рисунке 6.18.

	A	B	C	D	E	F	G	H	I
1	ВЫВОД ИТОГОВ								
2									
3	Регрессионная статистика								
4	Множественный R	0,868773601							
5	R-квадрат	0,754767569							
6	Нормированный R-квадрат	0,530095334							
7	Стандартная ошибка	6,542829309							
8	Наблюдения	18							
9									
10	Дисперсионный анализ								
11		df	SS	MS	F	Значимость F			
12	Регрессия	7	1449,303	207,0433	4,836486	0,012847			
13	Остаток	11	470,8948	42,80862					
14	Итого	18	1920,198						
15									
16	Коэффициенты Значения t-статистики P-Значения нижние 95% верхние 95% нижние 95,0% верхние 95,0%								
17	Y-пересечение	0	#N/D	#N/D	#N/D	#N/D	#N/D	#N/D	#N/D
18	ВРП на душу населения	0,227199172	0,097103	2,339785	0,039187	0,013478	0,440921	0,013478	0,440921
19	Объем инвестиций на душу населения	0,768434555	0,725266	1,05952	0,312074	-0,82787	2,364736	-0,82787	2,364736
20	Уровень регистрируемой безработицы	-2,704785962	1,93056	-1,40104	0,188778	-6,95392	1,54435	-6,95392	1,54435
21	Общий объем розничного товарооборота	0,069296985	0,130356	0,531598	0,605573	-0,21761	0,356209	-0,21761	0,356209
22	ОФ на душу населения	0,009781965	0,066963	0,146081	0,8865	-0,1376	0,157166	-0,1376	0,157166
23	Коэффициент плотности автомобильных дорог	-0,585147453	0,264023	-2,21627	0,048682	-1,16626	-0,00404	-1,16626	-0,00404
24	Сводный показатель уровня социального развития	-0,040035279	0,022676	-1,76556	0,105174	-0,08994	0,009874	-0,08994	0,009874

Рис.6.18 Результаты регрессионного анализа

Проведя регрессионный анализ влияющих на комплексную оценку факторов, можно получить уравнение для определения зависимости комплексной оценки социально-экономического развития региона от характеристик региона (города).

$$K = 0.23\hat{A} + 0.79\hat{E} - 2.7\hat{O} + 0.07\hat{U} + 0.01\hat{F} - 0.59\hat{P} - 0.04\hat{C} \quad (6.10)$$

где K – комплексная оценка, баллы;

B – валовой региональный продукт (с учетом паритета покупательной способности) на душу населения, тыс.руб.;

I – объем инвестиций в основной капитал на душу населения, тыс.руб.;

U – уровень регистрируемой безработицы, % к экономически активному населению;

T – общий объем розничного товарооборота и платных услуг (с учетом паритета покупательной способности) на душу населения, тыс.руб.;

Φ – основные фонды отраслей экономики (по полной балансовой стоимости, с учетом степени удорожания капитальных затрат) на душу населения, тыс.руб.;

P – плотность дорожной сети (коэффициент Энгеля);

C – сводный показатель уровня социального развития города (региона).

Значимость и достоверность полученной регрессионной модели обеспечивается расчетным значением F -критерия Фишера, равным 4,84, что больше необходимого табличного значения. **лекция 5 вопрос 6**

Отдельными расчетами (или взаимосвязанными с помощью системы уравнений) на прогнозируемый период определяются значения B, I, U, T, Φ, P и C .

Тогда уравнение регрессии позволяет дать на тот же период прогноз величины комплексной оценки.

Вопрос 5. Сезонная компонента в прогнозировании

Сезонная составляющая - составляющая временного ряда, описывающая регулярные изменения его значений в пределах некоторого периода и представляющая собой последовательность почти повторяющихся циклов

Сезонная составляющая имеет строгую периодичность, тогда как длительность отдельных циклов циклической составляющей - величина переменная.

В общем, периодическая зависимость может быть формально определена как корреляционная зависимость порядка k между каждым i -м элементом ряда и $(i-k)$ -м элементом. Ее можно измерить с помощью автокорреляции (т.е. корреляции между самими членами ряда); k обычно называют лагом (иногда используют эквивалентные термины: сдвиг, запаздывание). Если ошибка измерения не слишком большая, то сезонность можно определить визуально, рассматривая поведение членов ряда через каждые k временных единиц. При изучении коррелограмм следует помнить, что автокорреляции последовательных лагов формально зависимы между собой.

Один из способ описания временного ряда – разложение его на компоненты: тренд, периодическую, иррегулярную. Когда временная ось связана с датами, а период – с месяцами или кварталами, периодическую компоненту называют сезонной.

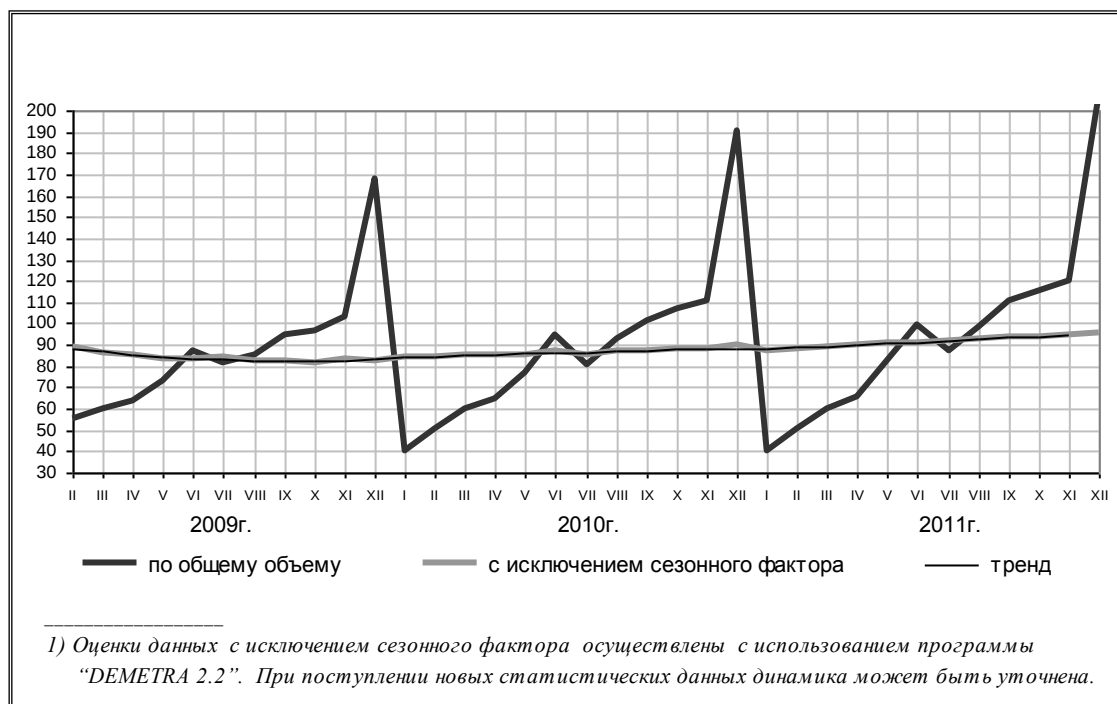


Рис.6.19. Динамика (временной ряд) инвестиций в основной капитал (в % к среднемесячному значению 2008г) по фактическим данным и с исключением сезонного фактора.

Уровень временного ряда представляется в виде суммы тренда, циклической и случайной компонент или в виде произведения этих компонент.

Согласно аддитивной модели временной ряд с сезонными колебаниями представляется в виде:

$$y_t = T_t + S_t + \varepsilon_t \quad (6.11)$$

где: y_t – значение прогнозируемой переменной для t -го момента времени; T_t – трендовая составляющая y_t ; S_t – сезонная составляющая y_t ; ε_t – случайная ошибка (невязка).

Согласно мультипликативной модели временной ряд с сезонными колебаниями имеет в вид:

$$y_t = T_t \cdot S_t \cdot \varepsilon_t \quad (6.12)$$

Применение мультипликативных моделей обусловлено тем, что в некоторых временных рядах значение сезонной компоненты представляет собой определенную долю трендового значения.

Отличить аддитивную модель от мультипликативной можно по величине сезонной вариации. Аддитивной модели присуща практически постоянная сезонная вариация, тогда как у мультипликативной она возрастает или убывает.

Процедуру построения тренд-сезонных моделей можно описать в виде следующей последовательности шагов.

1. Оценивание сезонной составляющей с учетом характера сезонности (аддитивной или мультипликативной)
2. Сезонная корректировка (десезонализация)
3. Расчет параметров тренда на основе применения временного ряда, полученного на втором шаге.
4. Моделирование динамики исходного ряда с учетом сезонной и трендовой составляющих.
5. Оценка точности и адекватности полученной модели.
6. Использование построенной модели для прогнозирования.

Удаление сезонных составляющих делает ряд стационарным, что необходимо для применения ARIMA и других методов, например, спектрального анализа.

Исходный ряд можно скорректировать, вычитая из него (аддитивная модель) или деля его значения на (мультипликативная модель) значения сезонной составляющей. Получающийся в результате ряд называется сезонной корректировкой ряда (из ряда убрана сезонная составляющая).

Пример практического применения прогнозирования с учетом сезонности в экономических расчетах

Рассмотрим возможный алгоритм построения прогноза объема реализации продукции с сезонным характером продаж. В качестве такого продукта будут выступать изделия из оцинкованной стали - воздуховоды и фасонные изделия для систем вентиляции. Данные по объемам продаж за 2010-2011 годы предоставлены ООО "Мосрегионвент", занимающемуся торговлей продукцией данного вида. Т.к. монтаж инженерных систем характеризуется сезонностью наряду с иными видами работ в строительстве, то прогнозирование сбыта продукции для данного вида работ является более чем актуальным.

Аддитивную модель прогнозирования можно представить в виде формулы:

$$F = T + S + E \quad (6.13)$$

где: F – прогнозируемое значение; T – тренд; S – сезонная компонента; E – ошибка прогноза.

Таблица 6.7 - Фактические объемы реализации продукции

<i>№ п/п</i>	<i>Месяц</i>	<i>Объем продаж, тыс.руб.</i>
1	Декабрь 2009 г.	2 345,25
2	Январь 2010 г.	1 326,81
3	Февраль 2010 г.	2 796,54
4	Март 2010 г.	2 754,70
5	Апрель 2010 г.	3 458,45

6	Май 2010 г.	3 224,55
7	Июнь 2010 г.	4 263,84
8	Июль 2010 г.	3 685,03
9	Август 2010 г.	3 311,00
10	Сентябрь 2010 г.	3 193,71
11	Октябрь 2010 г.	3 524,47
12	Ноябрь 2010 г.	2 497,52
13	Декабрь 2010 г.	2 461,99
14	Январь 2011 г.	1 427,99
15	Февраль 2011 г.	2 807,81
16	Март 2011 г.	2 999,47
17	Апрель 2011 г.	3 208,47
18	Май 2011 г.	3 184,05
19	Июнь 2011 г.	4 793,98
20	Июль 2011 г.	4 239,78
21	Август 2011 г.	3 845,48
22	Сентябрь 2011 г.	3 291,03
23	Октябрь 2011 г.	3 443,84
24	Ноябрь 2011 г.	3 344,12

1. Определяем тренд, наилучшим образом аппроксимирующий фактические данные. Для этого используем полиномиальный тренд, позволяющий сократить ошибку прогнозной модели. **лекция 5**

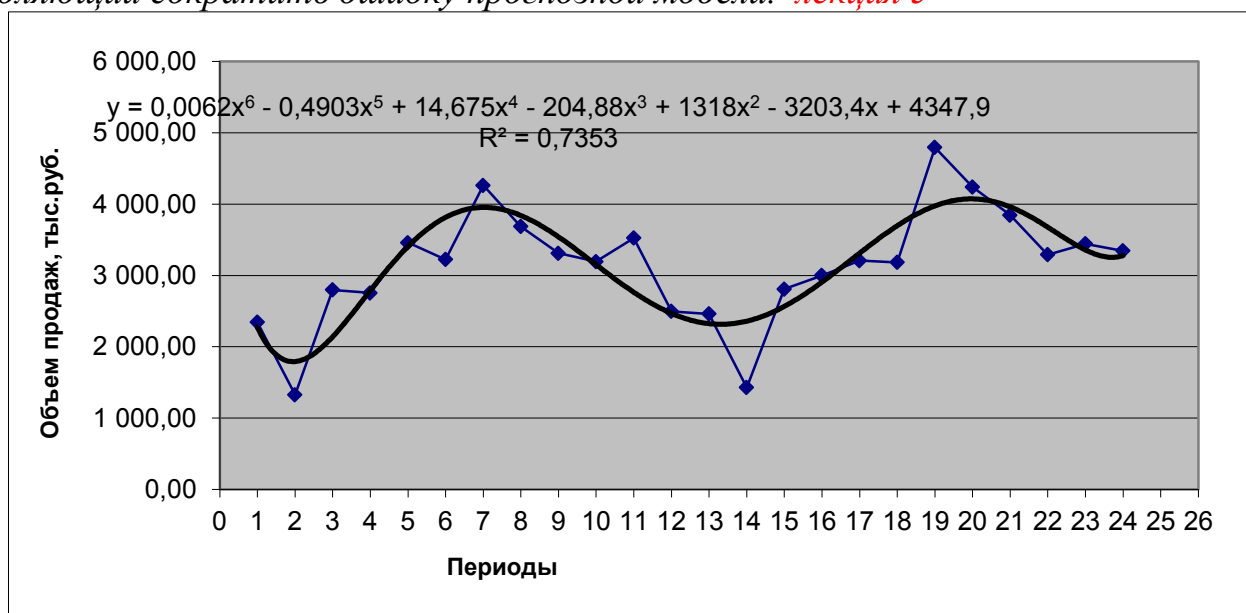


Рис. 6.20. Построение тренда

$$y = 0,0062x^6 - 0,4903x^5 + 14,675x^4 - 204,88x^3 + 1318x^2 - 3203,4x + 4347,9 \quad (6.14)$$

x — номер периода наблюдений.

Достоверность аппроксимации: $R^2 = 0,735$

2. Рассчитаем значения тренда, исходя из полученной модели (подставляя значения в вышеприведенную формулу).

Таблица 6.8- Значение тренда

№ n/n	Месяц	Объем продаж, тыс.руб.	Значение тренда
1	Декабрь 2009 г.	2 345,25	2 271,81
2	Январь 2010 г.	1 326,81	1 793,57
3	Февраль 2010 г.	2 796,54	2 141,99
4	Март 2010 г.	2 754,70	2 790,11
5	Апрель 2010 г.	3 458,45	3 407,46
6	Май 2010 г.	3 524,55	3 816,91
7	Июнь 2010 г.	4 263,84	3 955,89
8	Июль 2010 г.	3 685,03	3 842,08
9	Август 2010 г.	3 311,00	3 543,66
10	Сентябрь 2010 г.	3 193,71	3 153,90
11	Октябрь 2010 г.	3 524,47	2 770,27
12	Ноябрь 2010 г.	2 497,52	2 478,03
13	Декабрь 2010 г.	2 261,99	2 338,27
14	Январь 2011 г.	1 427,99	2 380,40
15	Февраль 2011 г.	2 507,81	2 599,09
16	Март 2011 г.	2 999,47	2 955,75
17	Апрель 2011 г.	2 808,47	3 384,38
18	Май 2011 г.	3 384,05	3 801,94
19	Июнь 2011 г.	3 793,98	4 123,18
20	Июль 2011 г.	4 239,78	4 279,90
21	Август 2011 г.	3 845,48	4 244,72
22	Сентябрь 2011 г.	3 291,03	4 059,30
23	Октябрь 2011 г.	3 443,84	3 866,95
24	Ноябрь 2011 г.	3 344,12	3 949,88

Например, значение тренда за период $1 = 0,0062 \cdot 1^6 - 0,4903 \cdot 1^5 + 14,675 \cdot 1^4 - 204,88 \cdot 1^3 + 1318 \cdot 1^2 - 3203,4 \cdot 1 + 4347,9 = 2271,81$;

Для тренда периода 2:

$x=2$, подставляем в формулу, и т.д.

3. Вычитая из фактических значений объемов продаж значения тренда, определим величины сезонной компоненты.

Таблица 6.9- Сезонная компонента

№ n/n	Месяц	Объем продаж, тыс.руб.	Значение тренда	Сезонная компонента
1	Декабрь 2009 г.	2 345,25	2 271,81	73,44
2	Январь 2010 г.	1 326,81	1 793,57	-466,75
3	Февраль 2010 г.	2 796,54	2 141,99	654,55
4	Март 2010 г.	2 754,70	2 790,11	-35,41
5	Апрель 2010 г.	3 458,45	3 407,46	50,99
6	Май 2010 г.	3 524,55	3 816,91	-292,37
7	Июнь 2010 г.	4 263,84	3 955,89	307,95
8	Июль 2010 г.	3 685,03	3 842,08	-157,06
9	Август 2010 г.	3 311,00	3 543,66	-232,66
10	Сентябрь 2010 г.	3 193,71	3 153,90	39,81
11	Октябрь 2010	3 524,47	2 770,27	754,20

	г.			
12	Ноябрь 2010 г.	2 497,52	2 478,03	19,49
1	Декабрь 2010 г.	2 261,99	2 338,27	-76,29
2	Январь 2011 г.	1 427,99	2 380,40	-952,40
3	Февраль 2011 г.	2 507,81	2 599,09	-91,28
4	Март 2011 г.	2 999,47	2 955,75	43,72
5	Апрель 2011 г.	2 808,47	3 384,38	-575,91
6	Май 2011 г.	3 384,05	3 801,94	-417,88
7	Июнь 2011 г.	3 793,98	4 123,18	-329,20
8	Июль 2011 г.	4 239,78	4 279,90	-40,12
9	Август 2011 г.	3 845,48	4 244,72	-399,25
10	Сентябрь 2011 г.	3 291,03	4 059,30	-768,26
11	Октябрь 2011 г.	3 443,84	3 866,95	-423,11
12	Ноябрь 2011 г.	3 344,12	3 949,88	-605,77

Например, для периода 1 сезонная компонента равна $2345,25 - 2271,81 = 73,44$ и т.д.

4. Скорректируем значения сезонной компоненты таким образом, чтобы их сумма была равна нулю. Для этого произведем корректировку среднего по двум сезонам на расчетную величину: итоговое среднее значение по сезонам/12.

Таблица 6.10- Корректировка сезонной компоненты

№ п/п	1 сезон	2 сезон	Итого	Среднее	Сезонная компонента
1	73,44	-76,29	-2,84	-1,42	161,89
2	-466,75	-952,40	-1 419,15	-709,58	-546,26
3	654,55	-91,28	563,27	281,64	444,95
4	-35,41	43,72	8,31	4,16	167,47
5	50,99	-575,91	-524,92	-262,46	-99,14
6	-292,37	-417,88	-710,25	-355,12	-191,81
7	307,95	-329,20	-21,25	-10,62	152,69
8	-157,06	-40,12	-197,18	-98,59	64,73
9	-232,66	-399,25	-631,91	-315,95	-152,64
10	39,81	-768,26	-728,45	-364,23	-200,91
11	754,20	-423,11	331,09	165,55	328,86
12	19,49	-605,77	-586,28	-293,14	-129,82
Итого			-1 959,78	-1 959,78	0,0000

Делим все данные за 2 года на 2 сезона, записываем определенные в предыдущей таблице значения сезонной компоненты в столбцы «1 сезон» и «2 сезон». Далее определяем итог как сумму значений столбцов за соответствующий месяц. Так, 1 месяц $\text{итого} = 73,44 + (-76,29) = -2,84$ и т.д.

Затем определяем среднее значение как $\text{итого} / 2$. Так, для 1 месяца $\text{среднее} = -2,84 / 2 = -1,42$

Итого среднее определяется как сумма всех значений по столбцу.

Далее находим расчетную величину для корректировки сезонной компоненты как $\text{итого среднее} / 12$: $-1959,78 / 12 = -163,31$.

Теперь рассчитываем последний столбец. Сезонная компонента за каждый месяц равна соответствующее среднее минус расчетная величина корректировки. Так, сезонная компонента 1 месяца= $-1,42 - (-163,31) = 161,89$ и т.д.

5. Рассчитываем ошибки модели как разности между фактическими значениями и значениями модели.

Таблица 6.11- Расчет ошибок

№ п/п	Объем продаж, тыс.руб.	Значение модели	Отклонение
1	2 345,25	2 433,70	-88,45
2	1 326,81	1 247,30	79,51
3	2 796,54	2 586,94	209,60
4	2 754,70	2 957,58	-202,88
5	3 458,45	3 308,32	150,13
6	3 524,55	3 625,10	-100,56
7	4 263,84	4 108,58	155,26
8	3 685,03	3 906,81	-221,78
9	3 311,00	3 391,02	-80,02
10	3 193,71	2 952,99	240,72
11	3 524,47	3 099,13	425,34
12	2 497,52	2 348,21	149,31
13	2 261,99	2 500,17	-238,18
14	1 427,99	1 834,13	-406,14
15	2 507,81	3 044,04	-536,23
16	2 999,47	3 123,22	-123,75
17	2 808,47	3 285,23	-476,76
18	3 384,05	3 610,13	-226,07
19	3 793,98	4 275,87	-481,89
20	4 239,78	4 344,63	-104,85
21	3 845,48	4 092,09	-246,61
22	3 291,03	3 858,38	-567,35
23	3 443,84	4 195,81	-751,97
24	3 344,12	3 820,06	-475,94

Значение модели = значение тренда (таб. 6.8) + скорректированная сезонная компонента (таб. 6.10, последний столбец)

Так, для 1 периода значение модели= $2271,81 + 161,89 = 2433,7$ и т.д.

Отклонение = объем продаж - значение модели= $2345,25 - 2433,7 = -88,45$ и т.д.

Так, для 1 периода Отклонение = $2345,25 - 2433,7 = -88,45$ и т.д.

Находим среднеквадратическую ошибку модели (E) по формуле:

$$E = \Sigma O^2 : \Sigma (T+S*2)^2 \quad (6.15)$$

где: T- трендовое значение объёма продаж; S – сезонная компонента;

O- отклонения модели от фактических значений

$$E = (-88,45 + 79,51 + \dots + (-751,97) + (-475,94))^2 / (((2271,81 + \dots + 3949,88) + (161,89 + \dots + -129,82)*2)^2) = 0,0025$$

Величина полученной ошибки (0,25%) позволяет говорить, что построенная модель хорошо аппроксимирует фактические данные, т.е. она вполне отражает экономические тенденции, определяющие объём продаж, и является предпосылкой для построения прогнозов высокого качества. Полученный прогноз представлен на рисунке 6.21.

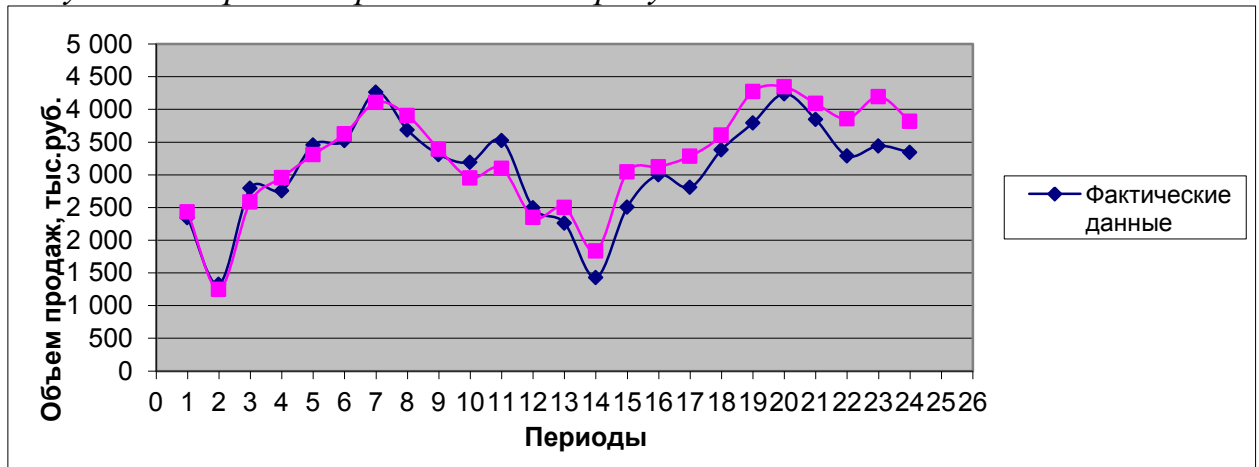


Рис.6.21. Прогноз согласно построенной модели с учетом сезонности

Кроме того, прогноз может содержать циклические компоненты. Циклическая компонента отличается от сезонной тем, что она обычно имеет большую временную протяженность и проявляется через неравные промежутки времени. Как и в предыдущих случаях, такая циклическая компонента может изменять прогнозируемую величину аддитивно, либо мультипликативно.

Вопрос 6. Модели авторегрессии

Прогнозирование с использованием модели авторегрессии опирается на предыдущие значения экономических величин. Слово авторегрессия означает зависимость последующего значения от предыдущих. **лекция 4 вопрос3.**

Зависимость в случае авторегрессии предполагается линейная, то есть прогноз представляет собой сумму значений за предыдущие дни с некоторыми коэффициентами, которые являются постоянными и определяют параметры модели авторегрессии. Сколько периодов из прошлого мы будем брать, чтобы попытаться спрогнозировать будущие значения, называется порядком модели авторегрессии p .

Авторегрессионная (AR-) модель — модель временных рядов, в которой значения временного ряда в данный момент линейно зависят от предыдущих значений этого же ряда.

Модели авторегрессии и скользящего среднего являются основными линейными моделями стационарных временных рядов.

Если значения уровней временного ряда точно определены какой-либо математической функцией, то данный временной ряд называется детерминированным, а сама функция называется реализацией исследуемого процесса. Если же уровни временного ряда могут быть описаны с помощью функции распределения вероятностей, то такой временной ряд называется случайным.

Уровни случайного временного ряда могут представлять собой непрерывные и дискретные случайные величины. Случайная величина называется дискретной, если множество ее возможных значений является конечным или счетным. Случайная величина называется непрерывной, если она может принимать любое значение из конечного или бесконечного интервала. Примером случайного временного ряда с дискретными уровнями может служить фиксация значений ежемесячной выдачи стипендий студентам. Пример случайного временного ряда с непрерывными уровнями — регистрация с определенной периодичностью температуры воздуха, которая во времени изменяется непрерывно.

Процесс, развивающийся во времени в соответствии с законами теории вероятностей, называется **стохастическим процессом**. Стохастический процесс называется стационарным, если его основные свойства остаются неизменными во времени:

1) постоянное математическое ожидание стационарного ряда $E(y_t)$, т. е. среднее значение временного ряда, вокруг которого изменяются уровни, является величиной неизменной;

2) постоянная дисперсия стационарного ряда, определяющая размах его колебаний относительно среднего значения σ^2 :

3) постоянство коэффициентов автокорреляция стационарного ряда с лагом l .

Временной ряд, не удовлетворяющий вышеперечисленным свойствам, называется нестационарным временным рядом.

В этом случае уровень временного ряда представляется в виде:

$$y_t = c + a_1 y_{t-1} + a_2 y_{t-2} + \dots + a_p y_{t-p} + \varepsilon_t \quad (6.16)$$

где c — постоянная величина (часто принимается равной 0), p — порядок авторегрессии; $a_1, \dots, a_p$ — коэффициенты авторегрессии, ε_t — случайная компонента.

Каждое наблюдение есть сумма случайной компоненты (случайное воздействие) и линейной комбинации предыдущих наблюдений.

На практике чаще всего используются модели первого, второго, максимум третьего порядков. Авторегрессионная модель порядка p обозначается как $AR(p)$.

Модель авторегрессии первого порядка $AR(1)$ называется марковским процессом, так как значения переменной y в текущий момент времени i

зависят только от значений переменной y в предыдущий момент времени ($t - 1$). (рис.6.22)

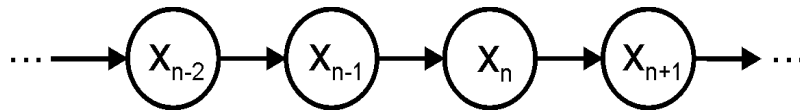


Рис. 6.22. Графическое представление авторегрессии 1го порядка

Модель авторегрессии второго порядка $AR(2)$ имеет название **процесс Юла**.

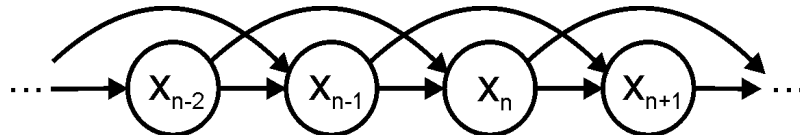


Рис. 6.23. Графическое представление авторегрессии 2го порядка

На коэффициенты авторегрессионной модели второго порядка накладываются следующие дополнительные ограничения:

$$\begin{aligned} (a_1 + a_2) &< 1 \\ (a_2 - a_1) &< 1 \\ |a_2| &< 1 \end{aligned} \quad (6.17)$$

Например, зная статистику продаж за предыдущие три дня, возможно прогнозировать, сколько будет продано в следующий день. Тогда, порядок модели $p = 3$. Кратко это часто записывается как $AR(3)$. Формула будет иметь вид

$$y_t = c + a_1 y_{t-1} + a_2 y_{t-2} + a_3 y_{t-3} + \varepsilon_t \quad (6.18)$$

где c — постоянная величина, a_1, a_2, a_3 — коэффициенты авторегрессии, ε_t — случайная компонента.

Для предварительного изучения особенностей автокорреляционного взаимодействия элементов ряда целесообразно проводить графический анализ исходных данных. **лекция 4 вопрос 3**

Интервалы времени, характеризующие удаленность сопоставляемых уровней ряда друг от друга, называются **периодом запаздывания**. Он показывает, через какой промежуток времени изменение переменной X_{tk} окажет воздействие на X_t . Изучение графических построений для различных значений k позволяет приблизительно оценить направление и силу связи между близлежащими членами ряда.

Для оценки тесноты связи используется коэффициент автокорреляции. **лекция 4 вопрос 3**

Автокорреляционная функция $AR(p)$ -процесса экспоненциально затухает с возможной осцилляцией (колебанием аналогично маятника). При этом частная автокорреляционная функция при $k > p$ равна нулю. Это

свойство используется для идентификации порядка AR-модели по выборочной частной автокорреляционной функции временного ряда.

Оценка параметров уравнений авторегрессии выполняется методом наименьших квадратов. **лекция 5 вопрос 2**

Прогнозирование на основе авторегрессионной модели представляет многоэтапную процедуру, каждая стадия которой позволяет определить величину показателя на очередной единичный отрезок времени.

Построение данного класса моделей реализовано в эконометрических программных комплексах (например, Statistica, SPSS и др.)

С помощью AR-моделей можно моделировать сезонность. **лекция 6 вопрос 5.** Такие модели обозначают SAR (Seasonal AR).

Другой разновидностью линейных моделей стационарных временных рядов является **модель скользящего среднего**.

Общая модель скользящего среднего порядка q имеет вид:

$$y_t = v_t - \varphi_1 v_{t-1} - \varphi_2 v_{t-2} - \dots - \varphi_q v_{t-q} \quad (6.19)$$

где q — это порядок скользящего среднего; φ_i — неизвестные коэффициенты, подлежащие оцениванию; v_t — белый шум.

То есть, текущее наблюдение ряда представляет собой сумму случайной компоненты в данный момент и линейной комбинации случайных воздействий в предыдущие моменты времени.

Модель скользящего среднего порядка q обозначается $MA(q)$.

Как и в случае авторегрессионных моделей, наиболее широко применяются модели скользящего среднего первого и второго порядков.

Для достижения большей гибкости при построении модели изучаемых временных рядов полезно включать в нее и авторегрессионные члены, и члены скользящего среднего. Такая модель называется смешанной **моделью авторегрессии - скользящего среднего**.

Модель авторегрессии - скользящего среднего (ARMA) обобщает две более простые модели временных рядов - модель авторегрессии (AR) и модель скользящего среднего (MA).

Она обозначается как $ARCC(p,q)$ или $ARMA(p,q)$. *Модель $ARCC(p,q)$ также является линейной моделью стационарных временных рядов.*

$$y_t = \delta y_{t-1} + v_t - \varphi v_{t-1} \quad (6.20)$$

где δ — параметр авторегрессии; φ — параметр скользящего среднего; v_t — белый шум.

Для того, чтобы определить необходимый порядок разности, нужно исследовать график ряда и автокоррелограмму. Сильные изменения уровня обычно требуют взятия несезонной разности первого порядка (лаг=1). Сильные изменения наклона требуют взятия разности второго порядка. Сезонная составляющая требует взятия соответствующей сезонной разности.

Однако, чрезмерное количество взятых разностей приводит к менее стабильным оценкам коэффициентов.

Модель, используемая при моделировании нестационарных временных рядов, называется **моделью авторегрессии и проинтегрированного скользящего среднего - ARIMA**.

В основе данной модели лежат два процесса:

- 1) процесс авторегрессии (формула (6.16);
- 2) процесс скользящего среднего (формула (6.19))

Дополнительно модели ARIMA могут содержать константу, интерпретация которой зависит от подгоняемой модели. Если в модели нет параметров авторегрессии, то константа c есть среднее значение ряда, если параметры авторегрессии имеются, то константа представляет собой свободный член. Если бралась разность ряда, то константа представляет собой среднее или свободный член преобразованного ряда.

*В оригинальных обозначениях Бокса и Дженкинса модель ARIMA записывается как **ARIMA (p, d, q)**, где p — параметры авторегрессии; d — порядок разностного оператора; q — параметры скользящего среднего.*

*Для рядов с периодической сезонной компонентой применяется ARIMA (АРПСС) с сезонностью, которая в обозначениях Бокса и Дженкинса записывается как **ARIMA (p, d, q) (ps, ds, qs)**, где параметры во второй скобке соответственно называются «сезонная авторегрессия», «сезонная разность» и «сезонное скользящее среднее».*

Аппроксимация временного ряда с помощью модели АРПСС происходит в несколько этапов.

1. Проверка ряда на стационарность.

Ряд является стационарным, если он имеет постоянные по времени средние, дисперсию и автокорреляцию с удаленными сезонной, циклической и трендовой составляющей. Такой ряд называется стационарным рядом случайных остатков. Проверить исходный ряд на стационарность можно с помощью автокорреляционной функции (АКФ) и частной автокорреляционной функции (ЧАКФ) остатков. Остатки представляют собой разности наблюдаемого временного ряда и значений, вычисленных с помощью модели.

2. Идентификация порядка модели и оценивание ее параметров.

На этом этапе необходимо решить, как много параметров p и q должно присутствовать в модели процесса.

Большинство встречающихся на практике временных рядов можно с достаточной степенью точности аппроксимировать одной из 5 основных моделей, которые можно идентифицировать по виду автокорреляционной (АКФ) и частной автокорреляционной функции (ЧАКФ).

- Один параметр (p): АКФ - экспоненциально убывает; ЧАКФ - имеет резко выделяющееся значение для лага 1, нет корреляций на других лагах.

- Два параметра авторегрессии (p): АКФ имеет форму синусоиды или экспоненциально убывает; ЧАКФ имеет резко выделяющиеся значения на лагах 1, 2, нет корреляций на других лагах.

- Один параметр скользящего среднего (q): АКФ имеет резко выделяющееся значение на лаге 1, нет корреляций на других лагах. ЧАКФ экспоненциально убывает.

- Два параметра скользящего среднего (q): АКФ имеет резко выделяющиеся значения на лагах 1, 2, нет корреляций на других лагах. ЧАКФ имеет форму синусоиды или экспоненциально убывает.

- Один параметр авторегрессии (p) и один параметр скользящего среднего (q): АКФ экспоненциально убывает с лага 1; ЧАКФ - экспоненциально убывает с лага 1.

Во время оценивания порядка модели используется квазиньютоновский алгоритм максимизации правдоподобия наблюдения значений ряда по значениям параметров (реализован в программных продуктах)

Для оценки параметров используется t -статистика Стьюдента. **лекция 5**
вопрос 5.

3. **Прогноз.** Полученные оценки параметров используются на последнем этапе для вычисления нового значения ряда и определения доверительного интервала для прогноза.

Показатели качества модели АРПСС

Для каждой из выбранных моделей оцениваются значения параметров и находятся остатки.

Проверяется, насколько модель соответствует данным.

Выбирается та модель, которая описывается меньшим количеством параметров.

Качество построенной модели осуществляется через проверку автокорреляции ее остатков. В этом случае применяют **общий критерий множителей Лагранжа (LM test)**, позволяющий обнаруживать в регрессионных остатках автокорреляцию более высоких порядков, чем первый. Но он применим только для больших выборок.

Контрольные вопросы по теме 6.

1. Дайте понятие временного ряда. Перечислите его основные характеристики.
2. Что такое автокорреляция уровней временного ряда и как ее можно оценить количественно?
3. Перечислите основные виды трендов.
4. Выпишите общий вид аддитивной и мультипликативной моделей временного ряда.
5. Перечислите этапы построения модели временного ряда.
6. С какими целями проводится выявление и устранение сезонного эффекта?