**Анализ временных рядов**

В трех предыдущих заметках описаны регрессионные модели, позволяющие прогнозировать отклик по значениям объясняющих переменных. В настоящей заметке мы покажем, как с помощью этих моделей и других статистических методов анализировать данные, собранные на протяжении последовательных временных интервалов. В соответствии с особенностями каждой компании, упомянутой в сценарии, мы рассмотрим три альтернативных подхода к анализу временных рядов.[[1]](#footnote-1)

Материал будет проиллюстрирован сквозным примером: **прогнозирование доходов трех компаний**. Представьте себе, что вы работаете аналитиком в крупной финансовой компании. Чтобы оценить инвестиционные перспективы своих клиентов, вам необходимо предсказать доходы трех компаний. Для этого вы собрали данные о трех интересующих вас компаниях — Eastman Kodak, Cabot Corporation и Wal-Mart. Поскольку компании различаются по виду деловой активности, каждый временной ряд обладает своими уникальными особенностями. Следовательно, для прогнозирования необходимо применять разные модели. Как выбрать наилучшую модель прогнозирования для каждой компании? Как оценить инвестиционные перспективы на основе результатов прогнозирования?

Обсуждение начинается с анализа ежегодных данных. Демонстрируются два метода сглаживания таких данных: скользящее среднее и экспоненциальное сглаживание. Затем демонстрируется процедура вычисления тренда с помощью метода наименьших квадратов и более сложные методы прогнозирования. В заключение, эти модели распространяются на временные ряды, построенные на основе ежемесячных или ежеквартальных данных.

**Прогнозирование в бизнесе**

Поскольку экономические условия с течением времени изменяются, менеджеры должны прогнозировать влияние, которое эти изменения окажут на их компанию. Одним из методов, позволяющих обеспечить точное планирование, является прогнозирование. Несмотря на большое количество разработанных методов, все они преследуют одну и ту же цель — предсказать события, которые произойдут в будущем, чтобы учесть их при разработке планов и стратегии развития компании.

Современное общество постоянно испытывает необходимость в прогнозировании. Например, чтобы выработать правильную политику, члены правительства должны прогнозировать уровни безработицы, инфляции, промышленного производства, подоходного налога отдельных лиц и корпораций. Чтобы определить потребности в оборудовании и персонале, директора авиакомпаний должны правильно предсказать объем авиаперевозок. Для того чтобы создать достаточное количество мест в общежитии, администраторы колледжей или университетов хотят знать, сколько студентов поступят в их учебное заведение в следующем году.

Существуют два общепринятых подхода к прогнозированию: качественный и количественный. Методы качественного прогнозирования особенно важны, если исследователю недоступны количественные данные. Как правило, эти методы носят весьма субъективный характер. Если статистику доступны данные об истории объекта исследования, следует применять методы количественного прогнозирования. Эти методы позволяют предсказать состояние объекта в будущем на основе данных о его прошлом. Методы количественного прогнозирования разделяются на две категории: анализ временных рядов и методы анализа причинно-следственных зависимостей.

*Временной ряд* — это набор числовых данных, полученных в течение последовательных периодов времени. Метод анализа временных рядов позволяет предсказать значение числовой переменной на основе ее прошлых и настоящих значений. Например, ежедневные котировки акций на Нью-Йоркской фондовой бирже образуют временной ряд. Другим примером временного ряда являются ежемесячные значения индекса потребительских цен, ежеквартальные величины валового внутреннего продукта и ежегодные доходы от продаж какой-нибудь компании.

*Методы анализа причинно-следственных зависимостей* позволяют определить, какие факторы влияют на значения прогнозируемой переменной. К ним относятся методы множественного регрессионного анализа с запаздывающими переменными, эконометрическое моделирование, анализ лидирующих индикаторов, методы анализа диффузионных индексов и других экономических показателей. Мы расскажем лишь о методах прогнозирования на основе анализа временн*ы*х рядов.

**Компоненты классической мультипликативной модели временн*ы*х рядов**

Основное предположение, лежащее в основе анализа временных рядов, состоит в следующем: факторы, влияющие на исследуемый объект в настоящем и прошлом, будут влиять на него и в будущем. Таким образом, основные цели анализа временных рядов заключаются в идентификации и выделении факторов, имеющих значение для прогнозирования. Чтобы достичь этой цели, были разработаны многие математические модели, предназначенные для исследования колебаний компонентов, входящих в модель временного ряда. Вероятно, наиболее распространенной является классическая мультипликативная модель для ежегодных, ежеквартальных и ежемесячных данных. Для демонстрации классической мультипликативной модели временных рядов рассмотрим данные о фактических доходах компании Wm.Wrigley Jr. Company за период с 1982 по 2001 годы (рис. 1).

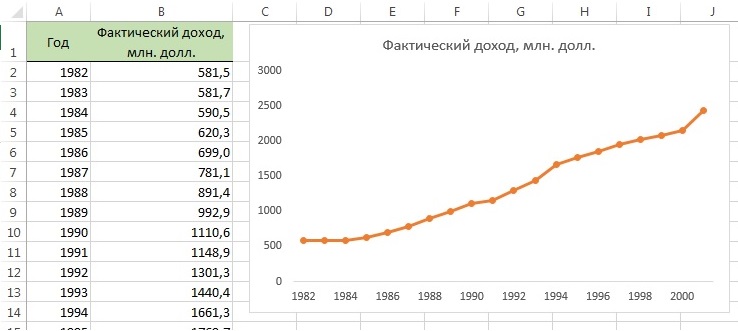


Рис. 1. График фактического валового дохода компании Wm.Wrigley Jr. Company (млн. долл. в текущих ценах) за период с 1982 по 2001 годы

Как видим, на протяжении 20 лет фактический валовой доход компании имел возрастающую тенденцию. Эта долговременная тенденция называется трендом. *Тренд* — не единственный компонент временного ряда. Кроме него, данные имеют циклический и нерегулярный компоненты. *Циклический* *компонент* описывает колебание данных вверх и вниз, часто коррелируя с циклами деловой активности. Его длина изменяется в интервале от 2 до 10 лет. Интенсивность, или амплитуда, циклического компонента также не постоянна. В некоторые годы данные могут быть выше значения, предсказанного трендом (т.е. находиться в окрестности пика цикла), а в другие годы — ниже (т.е. быть на дне цикла). Любые наблюдаемые данные, не лежащие на кривой тренда и не подчиняющиеся циклической зависимости, называются иррегулярными или *случайными компонентами*. Если данные записываются ежедневно или ежеквартально, возникает дополнительный компонент, называемый *сезонным*. Все компоненты временных рядов, характерных для экономических приложений, приведены на рис. 2.

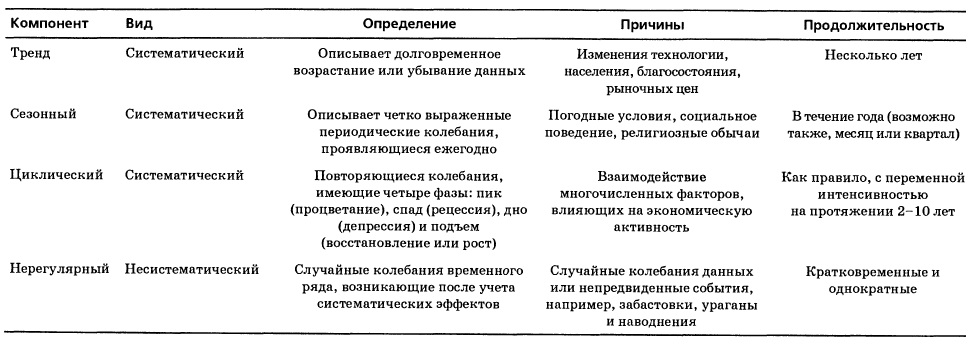


Рис. 2. Факторы, влияющие на временные ряды

Классическая мультипликативная модель временного ряда утверждает, что любое наблюдаемое значение является произведением перечисленных компонентов. Если данные являются ежегодными, наблюдение *Yi*, соответствующее *i*-му году, выражается уравнением:

*(1) Yi = Ti\*Ci\*Ii*

где *Ti* — значение тренда, *Ci* — значение циклического компонента в *i*-ом году, *Ii* — значение случайного компонента в *i*-ом году.

Если данные измеряются ежемесячно или ежеквартально, наблюдение *Yi*, соответствующее i-му периоду, выражается уравнением:

(2) *Yi = Ti\*Si\*Ci\*Ii*

где *Ti* — значение тренда, *Si* — значение сезонного компонента в *i*-ом периоде, *Ci* — значение циклического компонента в *i*-ом периоде, *Ii* — значение случайного компонента в *i*-ом периоде.

На первом этапе анализа временных рядов строится график данных и выявляется их зависимость от времени. Сначала необходимо выяснить, существует ли долговременное возрастание или убывание данных (т.е. тренд), или временной ряд колеблется вокруг горизонтальной линии. Если тренд отсутствует, то для сглаживания данных можно применить метод скользящих средних или экспоненциального сглаживания.

**Сглаживание годовых временных рядов**

В сценарии мы упомянули о компании Cabot Corporation. Имея штаб-квартиру в Бостоне, штат Массачусеттс, она специализируется на производстве и продаже химикатов, строительных материалов, продуктов тонкой химии, полупроводников и сжиженного природного газа. Компания имеет 39 заводов в 23 странах. Рыночная стоимость компании составляет около 1,87 млрд. долл. Ее акции котируются на Нью-Йоркской фондовой бирже под аббревиатурой СВТ. Доходы компании за указанный период приведены на рис. 3.

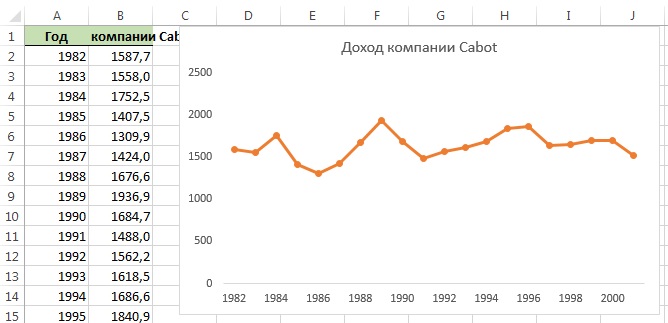


Рис. 3. Доходы компании Cabot Corporation в 1982–2001 годах (млрд. долл.)

Как видим, долговременная тенденция повышения доходов затемнена большим количеством колебаний. Таким образом, визуальный анализ графика не позволяет утверждать, что данные имеют тренд. В таких ситуациях можно применить методы скользящего среднего или экспоненциального сглаживания.

**Скользящие средние.** Метод скользящих средних весьма субъективен и зависит от длины периода *L*, выбранного для вычисления средних значений. Для того чтобы исключить циклические колебания, длина периода должна быть целым числом, кратным средней длине цикла. Скользящие средние для выбранного периода, имеющего длину *L*, образуют последовательность средних значений, вычисленных для последовательностей длины *L*. Скользящие средние обозначаются символами *MA(L)*.

Предположим, что мы хотим вычислить пятилетние скользящие средние значения по данным, измеренным в течение *n* = 11 лет. Поскольку *L* = 5, пятилетние скользящие средние образуют последовательность средних значений, вычисленных по пяти последовательным значениям временного ряда. Первое из пятилетних скользящих средних значений вычисляется путем суммирования данных о первых пяти годах с последующим делением на пять:



Второе пятилетнее скользящее среднее вычисляется путем суммирования данных о годах со 2-го по 6-й с последующим делением на пять:



Этот процесс продолжается, пока не будет вычислено скользящее среднее для последних пяти лет. Работая с годовыми данными, следует полагать число *L* (длину периода, выбранного для вычисления скользящих средних) нечетным. В этом случае невозможно вычислить скользящие средние для первых (*L* – 1)/2 и последних (*L* – 1)/2 лет. Следовательно, при работе с пятилетними скользящими средними невозможно выполнить вычисления для первых двух и последних двух лет. Год, для которого вычисляется скользящее среднее, должен находиться в середине периода, имеющего длину *L*. Если *n* = 11, a *L* = 5, первое скользящее среднее должно соответствовать третьему году, второе — четвертому, а последнее — девятому. На рис. 4 показаны графики 3- и 7-летних скользящих средних, вычисленные для доходов компании Cabot Corporation за период с 1982 по 2001 годы.

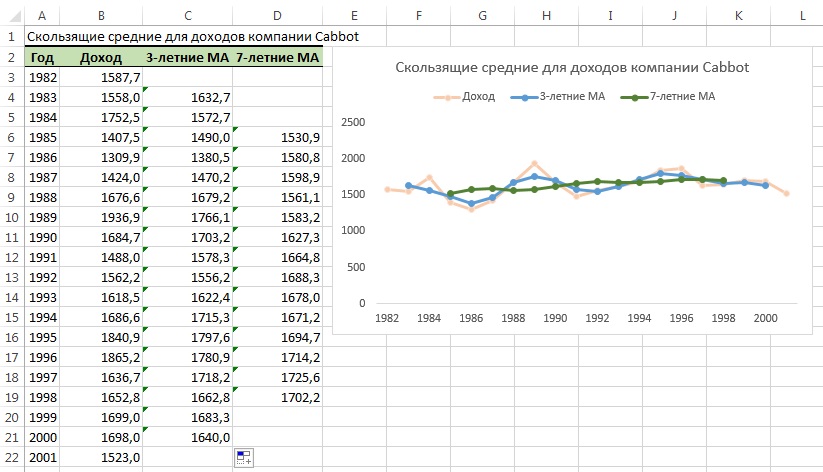


Рис. 4. Графики 3- и 7-летних скользящих средних, вычисленные для доходов компании Cabot Corporation

Обратите внимание на то, что при вычислении трехлетних скользящих средних проигнорированы наблюдаемые значения, соответствующие первому и последнему годам. Аналогично при вычислении семилетних скользящих средних нет результатов для первых и последних трех лет. Кроме того, семилетние скользящие средние намного больше сглаживают временной ряд, чем трехлетние. Это происходит потому, что семилетним скользящим средним соответствует более долгий период. К сожалению, чем больше длина периода, тем меньшее количество скользящих средних можно вычислить и представить на графике. Следовательно, больше семи лет для вычисления скользящих средних выбирать нежелательно, поскольку из начала и конца графика выпадет слишком много точек, что исказит форму временного ряда.

**Экспоненциальное сглаживание.** Для выявления долговременных тенденций, характеризующих изменения данных, кроме скользящих средних, применяется метод экспоненциального сглаживания. Этот метод позволяет также делать краткосрочные прогнозы (в рамках одного периода), когда наличие долговременных тенденций остается под вопросом. Благодаря этому метод экспоненциального сглаживания обладает значительным преимуществом над методом скользящих средних.

Метод экспоненциального сглаживания получил свое название от последовательности экспоненциально взвешенных скользящих средних. Каждое значение в этой последовательности зависит от всех предыдущих наблюдаемых значений. Еще одно преимущество метода экспоненциального сглаживания над методом скользящего среднего заключается в том, что при использовании последнего некоторые значения отбрасываются. При экспоненциальном сглаживании веса, присвоенные наблюдаемым значениям, убывают со временем, поэтому после выполнения вычислений наиболее часто встречающиеся значения получат наибольший вес, а редкие величины — наименьший. Несмотря на громадное количество вычислений, Excel позволяет реализовать метод экспоненциального сглаживания.

Уравнение, позволяющее сгладить временной ряд в пределах произвольного периода времени *i*, содержит три члена: текущее наблюдаемое значение *Yi*, принадлежащее временному ряду, предыдущее экспоненциально сглаженное значение *Ei–1* и присвоенный вес *W*.

*(3) E1 = Y1 Ei = WYi + (1 – W)Ei–1, I = 2, 3, 4, …*

где *Ei* – значение экспоненциально сглаженного ряда, вычисленное для *i*-го периода, *Ei–1* – значение экспоненциально сглаженного ряда, вычисленное для (*i* – 1)-гo периода, *Yi* – наблюдаемое значение временного ряда в *i*-ом периоде, *W* – субъективный вес, или сглаживающий коэффициент (0 < *W* < 1).

Выбор сглаживающего коэффициента, или веса, присвоенного членам ряда, является принципиально важным, поскольку он непосредственно влияет на результат. К сожалению, этот выбор до некоторой степени субъективен. Если исследователь хочет просто исключить из временного ряда нежелательные циклические или случайные колебания, следует выбирать небольшие величины *W* (близкие к нулю). С другой стороны, если временной ряд используется для прогнозирования, необходимо выбрать большой вес *W* (близкий к единице). В первом случае четко проявляются долговременные тенденции временного ряда. Во втором случае повышается точность краткосрочного прогнозирования (рис. 5).

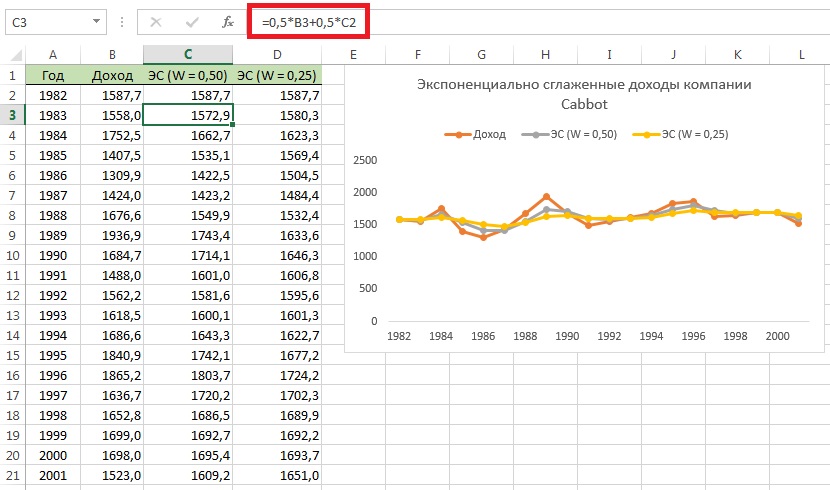


Рис. 5 Графики экспоненциально сглаженного временного ряда (W=0,50 и W=0,25) для данных о доходах компании Cabot Corporation за период с 1982 по 2001 годы; формулы расчета см. в файле Excel

Экспоненциально сглаженное значение, полученное для *i*-го временного интервала, можно использовать в качестве оценки предсказанного значения в (*i*+1)-м интервале:

*(4) Ŷi+1 = Ei*

Для предсказания доходов компании Cabot Corporation в 2002 году на основе экспоненциально сглаженного временного ряда, соответствующего весу *W* = 0,25, можно использовать сглаженное значение, вычисленное для 2001 года. Из рис. 5 видно, что эта величина равна 1651,0 млн. долл. Когда станут доступными данные о доходах компании в 2002 году, можно применить уравнение (3) и предсказать уровень доходов в 2003 году, используя сглаженное значение доходов в 2002 году:



*Пакет анализа* Excel способен построить график экспоненциального сглаживания в один клик. Пройдите по меню *Данные* → *Анализ данных* и выберите опцию *Экспоненциальное сглаживание* (рис. 6). В открывшемся окне *Экспоненциальное сглаживание* задайте параметры. К сожалению, процедура позволяет построить только один сглаженный ряд, поэтому, если вы хотите «поиграть» с параметром *W*, повторите процедуру.

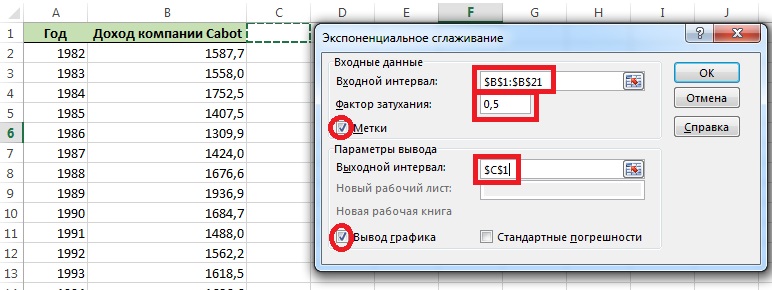


Рис. 6. Построение графика экспоненциального сглаживания с помощью Пакета анализа

**Вычисление трендов с помощью метода наименьших квадратов и прогнозирование**

Среди компонентов временного ряда чаще других исследуется тренд. Именно тренд позволяет делать краткосрочные и долгосрочные прогнозы. Для выявления долговременной тенденции изменения временного ряда обычно строят график, на котором наблюдаемые данные (значения зависимой переменной) откладываются на вертикальной оси, а временные интервалы (значения независимой переменной) — на горизонтальной. В этом разделе мы опишем процедуру выявления линейного, квадратичного и экспоненциального тренда с помощью метода наименьших квадратов.

**Модель линейного тренда** является простейшей моделью, применяемой для прогнозирования: *Yi = β0 + β1Xi +* εi. Уравнение линейного тренда:

*(5) Ŷi = b0 + b1Xi*

Напомним, что метод [линейного регрессионного анализа](http://baguzin.ru/wp/?p=6078) используется для вычисления выборочного наклона *b1* и сдвига *b0*. Вычислив уравнение *Ŷi = b0 + b1Xi*, в него можно подставлять значения *X*, чтобы определять отклик *Y*.

Если при аппроксимации временного ряда с помощью метода наименьших квадратов первое наблюдение расположить в начале координат, поставив его в соответствие значению *X* = 0, интерпретация коэффициентов упрощается. Все последующие наблюдения получают целочисленные номера: 1, 2, 3, так что *n*-е (последнее) наблюдение будет иметь номер *n* – 1. Например, если временной ряд записывается на протяжении 20 лет, первый год обозначается цифрой 0, второй— цифрой 1, третий — цифрой 2 и так далее, а последний (20-й) год — числом 19.

В сценарии была упомянута компания Wm. Wrigley Jr. Company, являющаяся крупнейшим производителем жевательной резинки в США. Акции компании котируются на Нью-Йоркской фондовой бирже под аббревиатурой WWY. Рыночная стоимость компании составляет 13 млрд. долл. Фактические доходы компании Wm. Wrigley Jr. Company в 1982-2001 годах приведены на рис. 7. Затем с помощью индекса потребительских цен (Consumer Price Index — CPI), вычисляемого Бюро статистики Министерства труда США, фактические доходы были преобразованы в реальные. Для этого следует умножить величину фактического дохода на коэффициент 100/CPI.

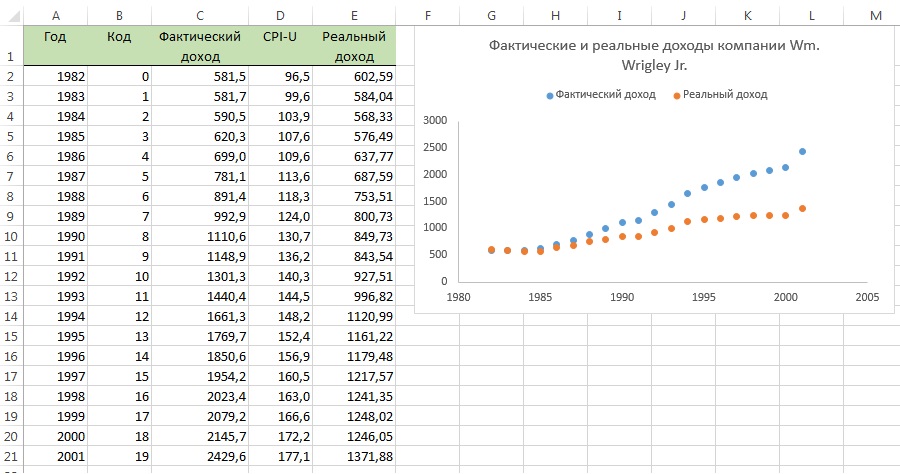


Рис. 7. Фактические и реальные доходы компании Wm. Wrigley Jr. Company в 1982-2001 годах

Обозначим последовательные значения переменной *X* с помощью целых чисел от 0 до 19, а затем выполним регрессионный анализ с помощью *Пакета анализа* (рис. 8).

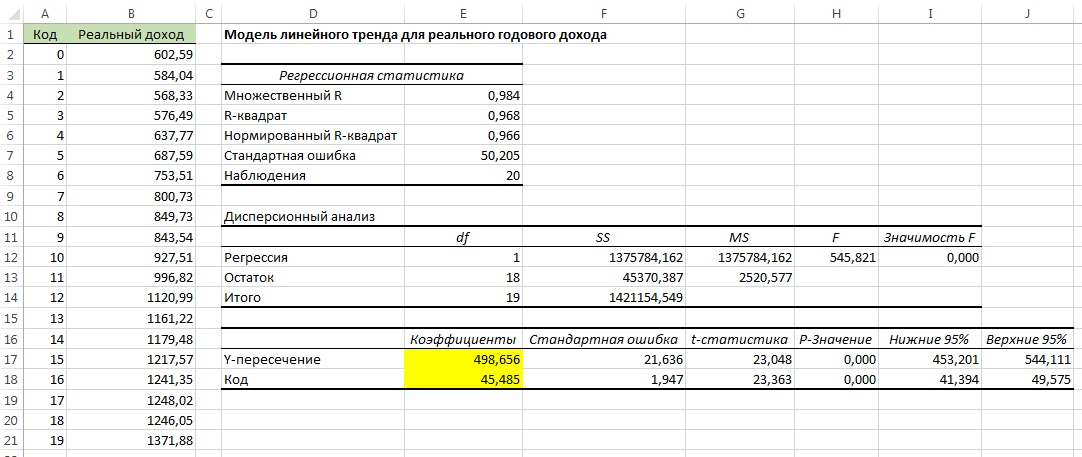


Рис. 8. Модель линейной регрессии для предсказания реального дохода компании Wm. Wrigley Jr.; построена с помощью Пакета анализа Excel

Уравнение линейной регрессии имеет следующий вид (см. ячейки Е17, Е18 на рис. 8): *Ŷi* = 498,656 + 45,485*Хi*, где началом координат является 1982 год, а шаг переменной *X* равен одному году. Регрессионные коэффициенты интерпретируются следующим образом:

* Сдвиг *b0* = 498,656 представляет собой предсказанное среднее значение реальных доходов компании Wm. Wrigley Jr. Company в 1982 году.
* Наклон *b1* = 45,485 представляет собой предсказанное увеличение реальных доходов компании в среднем на 45,485 млрд. долл. в год.

Линия тренда и временной ряд реальных доходов показаны на рис. 9. График можно построить на основании уравнения линейной регрессии (колонка D; рис. 9а) или простым добавлением линии тренда на ранее построенный график доходов (рис. 9б). Видно, что на протяжении ряда лет доходы компании линейно возрастали. Скорректированный коэффициент *r2* равен 0,966 (ячейка Е6 на рис. 8). Следовательно, все изменения реальных доходов хорошо описываются линейным трендом. Возникает вопрос: а нельзя ли выбрать еще более точную модель? Для ответа на него рассмотрим еще две модели — квадратичную и экспоненциальную.

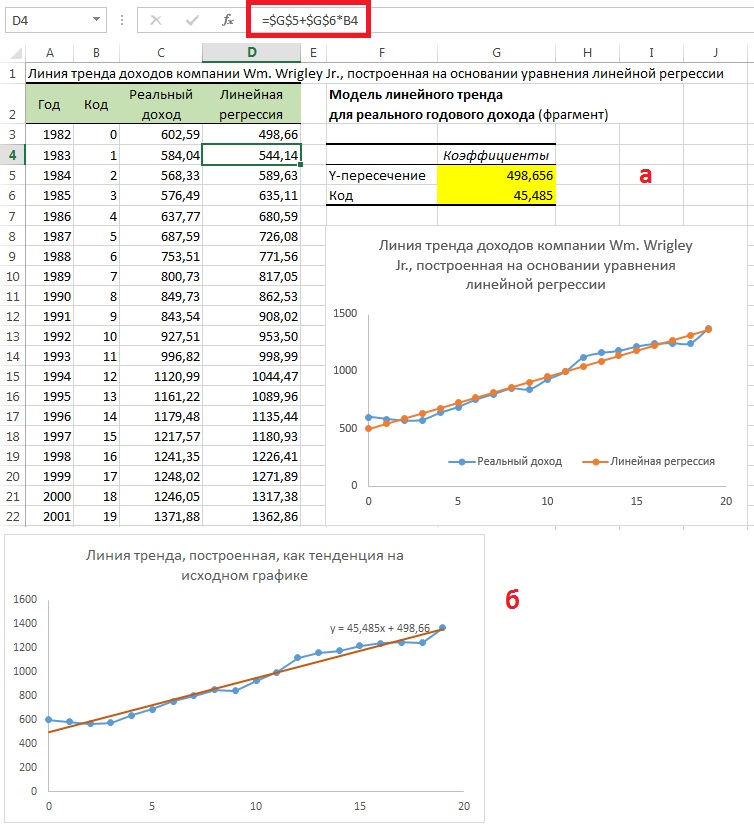


Рис. 9. Линия тренда реальных доходов компании Wm. Wrigley Jr., вычисленная с помощью метода наименьших квадратов, построенная: (а) на основании уравнения линейной регрессии; (б) добавлением линии тренда на график

**Модель квадратичного тренда,** или полиномиальная модель второй степени является простейшей нелинейной моделью, применяемой для прогнозирования: . Уравнение квадратичного тренда:



где *b0* – оценка сдвига отклика *Y*, *b1* – оценка линейного эффекта, *b2* – оценка квадратичного эффекта.

Построим квадратичный тренд путем добавления линии тренда на график с исходными данными (рис. 10).

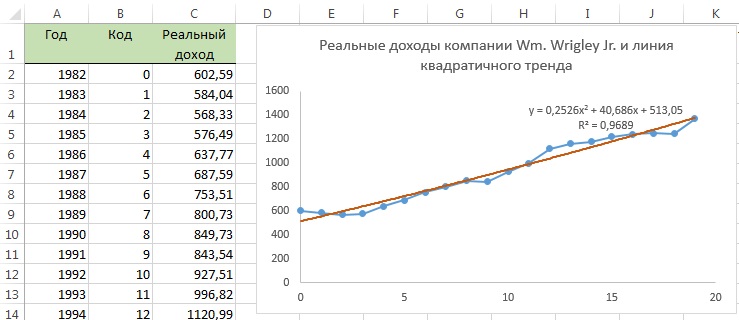


Рис. 10. График квадратичного тренда для предсказания реальных доходов компании Wm. Wrigley Jr.

Как показано на рис. 10, уравнение линейной регрессии: *Y* =513,05 + 40,686*Х* + 0,2526*Х2*, где началом координат является 1982 год, а шаг переменной *X* равен одному году. Этот график аппроксимирует временной ряд почти так же, как и линейный тренд. Скорректированный коэффициент *r2* равен 0,965 (рис. 11), а *t*-статистика, учитывающая вклад квадратичного эффекта, равна 0,656 (соответствующее *р*-значение равно 0,521).

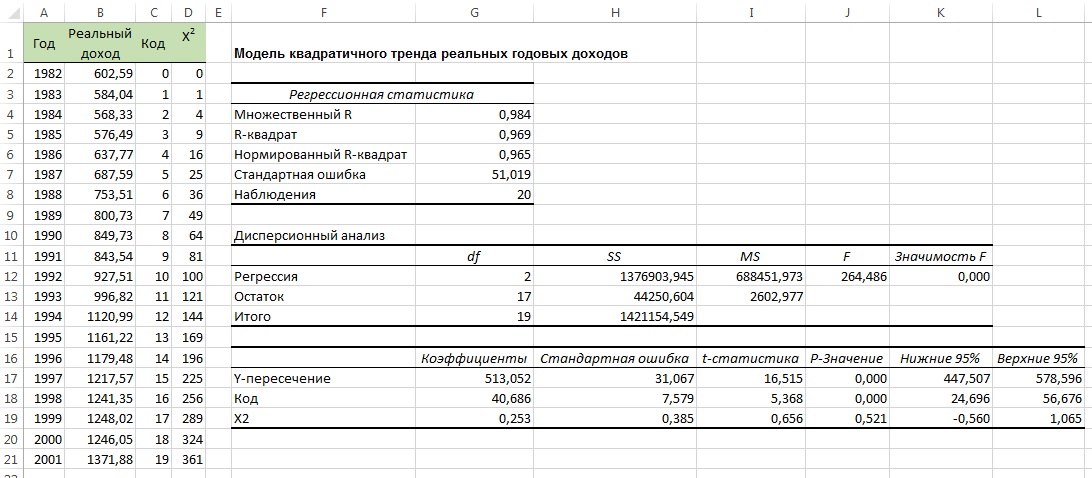


Рис. 11. Модель квадратичного тренда реальных годовых доходов

**Модель экспоненциального тренда.** Если временной ряд является возрастающим, а относительное изменение данных — постоянным, можно применять модель экспоненциального тренда. Учитывая сложность формул ограничимся только графическим анализом. Для этого на график с исходными данными добавим линию экспоненциального тренда, а также выведем на график уравнение тренда и параметр *r2* (рис. 12).

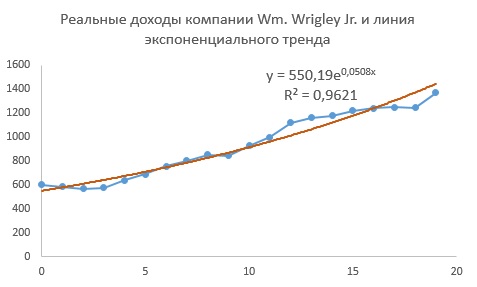


Рис. 12. График экспоненциального тренда для предсказания реальных доходов компании Wm. Wrigley Jr.

Экспоненциальная модель аппроксимирует временной ряд почти так же, как линейная и квадратичная модель. Скорректированный коэффициент *r2* равен 0,960, в то время как для линейной модели этот коэффициент равен 0,966, а для квадратичной – 0,965.

**Выбор модели на основе разностей первого и второго порядка, а также относительных разностей**

Для аппроксимации данных о реальном доходе компании Wm. Wrigley Jr. Company мы применили три модели: линейную, квадратичную и экспоненциальную. Какая из этих моделей лучше? Кроме визуального впечатления и сравнения скорректированных коэффициентов *r2*, в качестве инструмента для оценки качества модели применяются разности первого, второго и третьего порядка.

Выбор модели на основе анализа разностей первого и второго порядка, а также относительных разностей:

* Если исходные данные хорошо аппроксимируются линейной моделью, разность первого порядка должна быть постоянной. Иначе говоря, разности между двумя последовательными значениями одинаковы: *Y2 – Y1 = Y3 – Y2 = … = Yn – Yn–1*
* Если исходные данные хорошо аппроксимируются квадратичной моделью, разность второго порядка должна быть постоянной. Иначе говоря, разности между двумя последовательными разностями первого порядка одинаковы: *(Y3 – Y2) – (Y2 – Y1) = (Y4 – Y3) – (Y3 – Y2) = … = (Yn – Yn–1) – (Yn–1 – Yn–2)*
* Если исходные данные хорошо аппроксимируются экспоненциальной моделью, относительная разность должна быть постоянной. Иначе говоря, относительные разности, вычисленные по двум последовательным наблюдениям, одинаковы:

Не следует ожидать, что модель будет идеально аппроксимировать конкретный набор данных. Несмотря на это, при выборе подходящей модели необходимо анализировать разности первого и второго порядка, а также относительные разности. Для нашего примера с компанией Wm. Wrigley Jr. результаты такого анализа приведены на рис. 13.

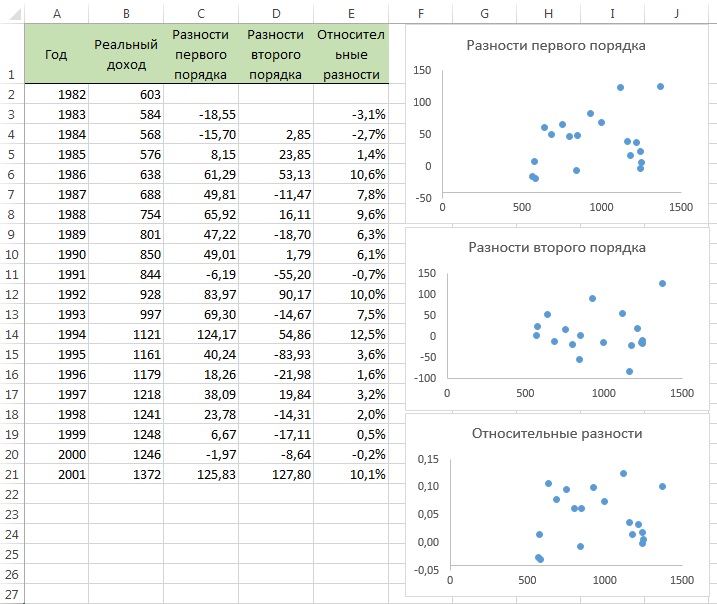
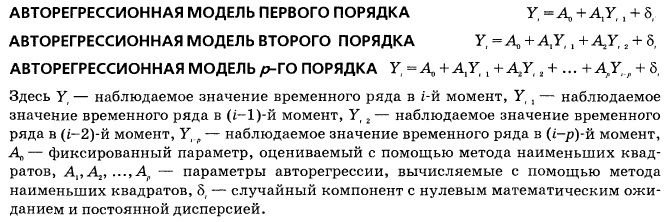


Рис. 13. Разности первого и второго порядка, а также относительные разности, вычисленные на основе данных о реальных доходах компании Wm. Wrigley Jr.

Анализ рис. 13 показывает, что разности первого и второго порядка, а также относительные разности не остаются постоянными. Итак, несмотря на то, что скорректированный коэффициент *r2* у всех трех моделей, рассмотренных выше, одинаков и приближенно равен 0,96, возможно, существуют более точные модели.

**Вычисление тренда с помощью авторегрессии и прогнозирование**

Другой подход к прогнозированию основан на авторегрессионной модели. Часто значения временного ряда в какой-то момент времени сильно коррелируют как с предшествующими, так и с последующими значениями. Автокорреляция первого порядка оценивает степень зависимости между последовательными значениями временного ряда. Автокорреляция второго порядка оценивает силу связи между значениями, разделенными двумя временными интервалами. Автокорреляция *р*-го порядка представляет собой величину корреляции между значениями, разделенными *р* временными интервалами. Авторегрессионная модель позволяет лучше оценить предысторию и получить более точный прогноз.



Авторегрессионная модель первого порядка внешне напоминает модель простой линейной регрессии, а авторегрессионные модели второго и *р*-го порядков похожи на модель множественной регрессии. В регрессионных моделях параметры регрессии обозначаются символами β0, β1, …, βk а их оценки — символами *b0*, *b1*, …, *bk*. В авторегрессионных моделях аналогичные параметры обозначаются символами *А0*, *А1* ..., *Аp*, а их оценки — символами *а0*, *а0*, ..., *аp*.

В авторегрессионной модели первого порядка рассматриваются лишь соседние значения временного ряда. В авторегрессионной модели второго порядка оценивается зависимость и корреляция как между соседними, так и между последовательными значениями временного ряда, разделенными двумя временными интервалами. В авторегрессионной модели *р*-го порядка оценивается зависимость и корреляция между соседними значениями, последовательными значениями временного ряда, разделенными двумя временными интервалами, и так далее вплоть до последовательных значений временного ряда, разделенных *р* временными интервалами.

Выбор подходящей авторегрессионной модели представляет собой нелегкую задачу. В процессе ее решения необходимо оценить простоту модели и возможные потери вследствие игнорирования автокорреляции между данными. С другой стороны, модели высоких порядков сопряжены с оценками многочисленных параметров, которые могут оказаться бесполезными, особенно если длина *n* временного ряда не очень велика. Это происходит потому, что при вычислении параметра *Ар* каждое значение временного ряда *Yi* сравнивается с его ближайшими соседями, расположенными не далее, чем через *р* временных интервалов (т.е. величина *Yi* сравнивается со значениями *Yi–1*, *Yi–2*, …, *Yi–p*). Иными словами, чем выше порядок авторегрессионной модели, тем больше первых ее членов теряется.

Выбрав модель и применив метод наименьших квадратов для вычисления оценок регрессионных параметров, необходимо оценить ее адекватность. Для этого можно использовать либо авторегрессионную модель конкретного порядка, которую уже применяли для похожих данных, либо сразу построить модель с несколькими параметрами, а затем последовательно исключать из нее параметры, не имеющие статистически значимого вклада. В последнем случае применяется *t*-критерий значимости параметра *Аp*, имеющего наивысший порядок в данной авторегрессионной модели. Нулевая и альтернативная гипотезы формулируются следующим образом: *H0*: *Ар* = 0, *H1*: *Ар* ≠ 0.

Использование *t*-критерия значимости параметра авторегрессии *Ар*, имеющего наивысший порядок:

*(7) t =*

где *Аp* — гипотетическое значение параметра, имеющего наивысший порядок в регрессионной модели, *ар* — оценка параметра авторегрессии *Аp*, имеющего наивысший порядок, Sap — стандартная ошибка оценки *ар*. Тестовая *t*-статистика имеет *t*-распределение с *n–2р–1* степенями свободы.[[2]](#footnote-2)

При заданном уровне значимости α нулевая гипотеза отклоняется, если тестовая *t*-статистика больше верхнего или меньше нижнего критического уровня *t*-распределения. Иначе говоря, решающее правило формулируется следующим образом: если *t* > *tU* или *t* < *tL*, нулевая гипотеза *Н0* отклоняется, в противном случае нулевая гипотеза не отклоняется (рис. 14).



Рис. 14. Области отклонения гипотезы для двустороннего критерия значимости параметра авторегрессии *Ар*, имеющего наивысший порядок

Если нулевая гипотеза (*Ар* = 0) не отклоняется, значит, выбранная модель содержит слишком много параметров. Критерий позволяет отбросить старший член модели и оценить авторегрессионную модель порядка *р–1*. Эту процедуру следует продолжать до тех пор, пока нулевая гипотеза *Н0* не будет отклонена.

Авторегрессионная модель является весьма полезным инструментом для аппроксимации и предсказания значений временного ряда. Этапы авторегрессионного моделирования годовых временных рядов:

1. Выберите порядок *р* оцениваемой авторегрессионной модели с учетом того, что *t*-критерий значимости имеет *n–2р–1* степеней свободы.
2. Сформируйте последовательность переменных *р* «с запаздыванием» так, чтобы первая переменная запаздывала на один временной интервал, вторая — на два и так далее. Последнее значение должно запаздывать на *р* временных интервалов (см. рис. 15).
3. Примените *Пакет анализа* Excel для вычисления регрессионной модели, содержащей все *р* значений временного ряда с запаздыванием.
4. Оцените значимость параметра *АР*, имеющего наивысший порядок: а) если нулевая гипотеза отклоняется, в авторегрессионную модель можно включать все *р* параметров; б) если нулевая гипотеза не отклоняется, отбросьте *р*-ю переменную и повторите п.3 и 4 для новой модели, включающей *р–1* параметр. Проверка значимости новой модели основана на *t*-критерии, количество степеней свободы определяется новым количеством параметров.
5. Повторяйте п.3 и 4, пока старший член авторегрессионной модели не станет статистически значимым.

Чтобы продемонстрировать авторегрессионное моделирование, вернемся к анализу временного ряда реальных доходов компании Wm. Wrigley Jr. На рис. 15 показаны данные, необходимые для построения авторегрессионных моделей первого, второго и третьего порядка. Для построения модели третьего порядка необходимы все столбцы этой таблицы. При построении авторегрессионной модели второго порядка последний столбец игнорируется. При построении авторегрессионной модели первого порядка игнорируются два последних столбца. Таким образом, при построении авторегрессионных моделей первого, второго и третьего порядка из 20 переменных исключаются одна, две и три соответственно.

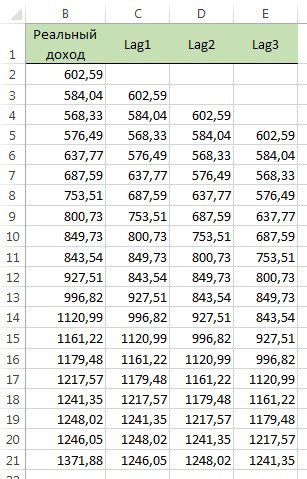


Рис. 15. Исходные данные для построения авторегрессионных моделей первого, второго и третьего порядка для реальных доходов компании Wm. Wrigley Jr.

Выбор наиболее точной авторегрессионной модели начинается с модели третьего порядка. Для корректной работы *Пакета анализа* следует в качестве входного интервала *Y* указать диапазон В5:В21, а входного интервала для *Х* – С5:Е21. Данные анализа приведены на рис. 16.

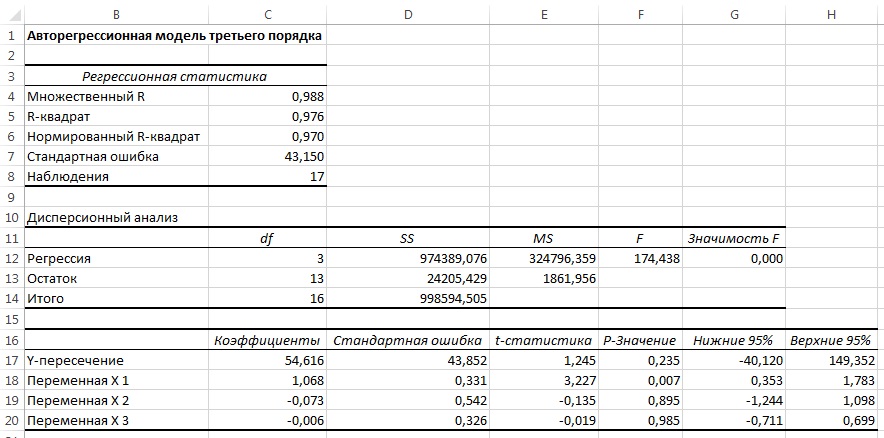


Рис. 16. Авторегрессионная модель третьего порядка для реальных доходов компании Wm. Wrigley Jr.

Проверим значимость параметра *А3*, имеющего наивысший порядок. Его оценка *а3* равна –0,006 (ячейка С20 на рис. 16), а стандартная ошибка равна 0,326 (ячейка D20). Для проверки гипотез Н0: А3 = 0 и Н1: А3 ≠ 0 вычислим *t*-статистику:

*t =*  = –0,019

При уровне значимости α = 0,05, критические величины двухстороннего *t*-критерия с n–2p–1 = 20–2\*3–1 = 13 степенями свободы равны: *tL* =СТЬЮДЕНТ.ОБР(0,025;13) = ­–2,160; *tU* =СТЬЮДЕНТ.ОБР(0,975;13) = +2,160. Поскольку –2,160 < *t* = –0,019 < +2,160 и *р* = 0,985 > α = 0,05, нулевую гипотезу *Н0* отклонять нельзя. Таким образом, параметр третьего порядка не имеет статистической значимости в авторегрессионной модели и должен быть удален.

Повторим анализ для авторегрессионной модели второго порядка (рис. 17). Оценка параметра, имеющего наивысший порядок, *а2* = –0,205, а ее стандартная ошибка равна 0,276. Для проверки гипотез Н0: А2 = 0 и Н1: А2 ≠ 0 вычислим *t*-статистику:

*t =*  = –0,744

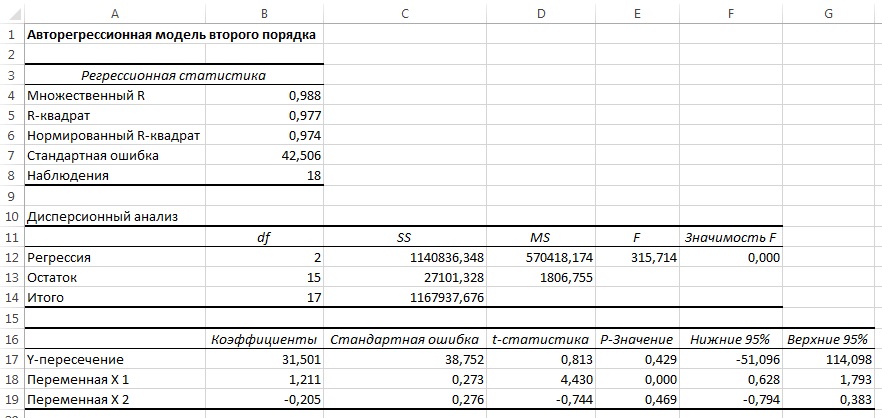


Рис. 17. Авторегрессионная модель второго порядка для реальных доходов компании Wm. Wrigley Jr.

При уровне значимости α = 0,05, критические величины двухстороннего *t*-критерия с n–2p–1 = 20–2\*2–1 = 15 степенями свободы равны: *tL* =СТЬЮДЕНТ.ОБР(0,025;15) = ­–2,131; *tU* =СТЬЮДЕНТ.ОБР(0,975;15) = +2,131. Поскольку –2,131 < *t* = –0,744 < –2,131 и *р* = 0,469 > α = 0,05, нулевую гипотезу *Н0* отклонять нельзя. Таким образом, параметр второго порядка не является статистически значимым, и его следует удалить из модели.

Повторим анализ для авторегрессионной модели первого порядка (рис. 18). Оценка параметра, имеющего наивысший порядок, *а1* = 1,024, а ее стандартная ошибка равна 0,039. Для проверки гипотез Н0: А1 = 0 и Н1: А1 ≠ 0 вычислим *t*-статистику:

*t =*  = 26,393

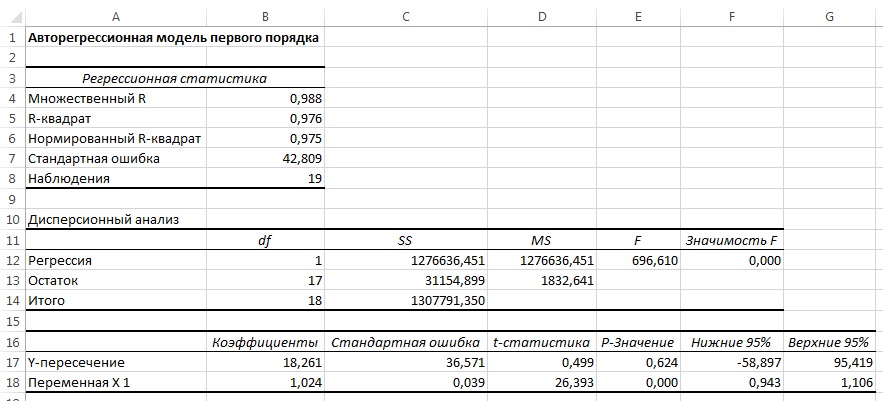


Рис. 18. Авторегрессионная модель первого порядка для реальных доходов компании Wm. Wrigley Jr.

При уровне значимости α = 0,05, критические величины двухстороннего *t*-критерия с n–2p–1 = 20–2\*1–1 = 17 степенями свободы равны: *tL* =СТЬЮДЕНТ.ОБР(0,025;17) = ­–2,110; *tU* =СТЬЮДЕНТ.ОБР(0,975;17) = +2,110. Поскольку –2,110 < *t* = 26,393 < –2,110 и *р* = 0,000 < α = 0,05, нулевую гипотезу *Н0* следует отклонить. Таким образом, параметр первого порядка является статистически значимым, и его нельзя удалять из модели. Итак, модель авторегрессии первого порядка лучше других аппроксимирует исходные данные. Используя оценки *а0* = 18,261, *а1* = 1,024 и значение временного ряда за последний год — Y20 = 1 371,88, можно предсказать величину реальных доходов компании Wm. Wrigley Jr. Company в 2002 г.:



**Выбор адекватной модели прогнозирования**

Выше были описаны шесть методов прогнозирования значений временного ряда: модели линейного, квадратичного и экспоненциального трендов и авторегрессионные модели первого, второго и третьего порядков. Существует ли оптимальная модель? Какую из шести описанных моделей следует применять для прогнозирования значения временного ряда? Ниже перечислены четыре принципа, которыми необходимо руководствоваться при выборе адекватной модели прогнозирования. Эти принципы основаны на оценках точности моделей. При этом предполагается, что значения временного ряда можно предсказать, изучая его предыдущие значения.

Принципы выбора моделей для прогнозирования:

* Выполните анализ остатков.
* Оцените величину остаточной ошибки с помощью квадратов разностей.
* Оцените величину остаточной ошибки с помощью абсолютных разностей.
* Руководствуйтесь принципом экономии.

**Анализ остатков.** Напомним, что остатком называется разность между предсказанным и наблюдаемым значением. Построив модель для временного ряда, следует вычислить остатки для каждого из *n* интервалов. Как показано на рис. 19, панель А, если модель является адекватной, остатки представляют собой случайный компонент временного ряда и, следовательно, распределены нерегулярно. С другой стороны, как показано на остальных панелях, если модель не адекватна, остатки могут иметь систематическую зависимость, не учитывающую либо тренд (панель Б), либо циклический (панель В), либо сезонный компонент (панель Г).

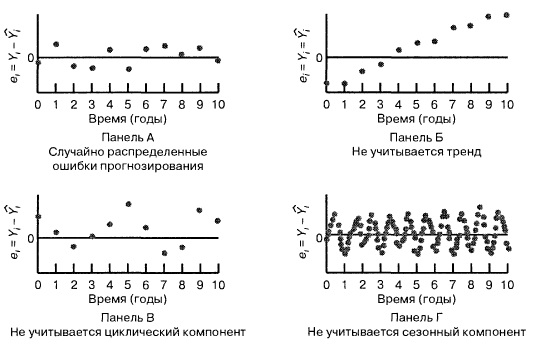


Рис. 19. Анализ остатков

**Измерение абсолютной и среднеквадратичной остаточных погрешностей.** Если анализ остатков не позволяет определить единственную адекватную модель, можно воспользоваться другими методами, основанными на оценке величины остаточной погрешности. К сожалению, статистики не пришли к консенсусу относительно наилучшей оценки остаточных погрешностей моделей, применяемых для прогнозирования. Исходя из принципа наименьших квадратов, можно сначала провести регрессионный анализ и вычислить стандартную ошибку оценки *SXY*. При анализе конкретной модели эта величина представляет собой сумму квадратов разностей между фактическим и предсказанным значениями временного ряда. Если модель идеально аппроксимирует значения временного ряда в предыдущие моменты времени, стандартная ошибка оценки равна нулю. С другой стороны, если модель плохо аппроксимирует значения временного ряда в предыдущие моменты времени, стандартная ошибка оценки велика. Таким образом, анализируя адекватность нескольких моделей, можно выбрать модель, имеющую минимальную стандартную ошибку оценки SXY.

Основным недостатком такого подхода является преувеличение ошибок при прогнозировании отдельных значений. Иначе говоря, любая большая разность между величинами *Yi* и *Ŷi* при вычислении суммы квадратов ошибок SSE возводится в квадрат, т.е. увеличивается. По этой причине многие статистики предпочитают применять для оценки адекватности модели прогнозирования среднее абсолютное отклонение (mean absolute deviation — MAD):

*(8) MAD =*

При анализе конкретных моделей величина MAD представляет собой среднее значение модулей разностей между фактическим и предсказанными значениями временного ряда. Если модель идеально аппроксимирует значения временного ряда в предыдущие моменты времени, среднее абсолютное отклонение равно нулю. С другой стороны, если модель плохо аппроксимирует такие значения временного ряда, среднее абсолютное отклонение велико. Таким образом, анализируя адекватность нескольких моделей, можно выбрать модель, имеющую минимальное среднее абсолютное отклонение.

**Принцип экономии.** Если анализ стандартных ошибок оценок и средних абсолютных отклонений не позволяет определить оптимальную модель, можно воспользоваться четвертым методом, основанным на принципе экономии. Этот принцип утверждает, что из нескольких равноправных моделей следует выбирать простейшую.

Среди шести рассмотренных в главе моделей прогнозирования наиболее простыми являются линейная и квадратичная регрессионные модели, а также авторегрессионная модель первого порядка. Остальные модели намного сложнее.

**Сравнение четырех методов прогнозирования.** Для иллюстрации процесса выбора оптимальной модели вернемся к временному ряду, состоящему из величин реального дохода компании Wm. Wrigley Jr. Company. Сравним четыре модели: линейную, квадратичную, экспоненциальную и авторегрессионную модель первого порядка. (Авторегрессионные модели второго и третьего порядка лишь незначительно улучшают точность прогнозирования значений данного временного ряда, поэтому их можно не рассматривать.) На рис. 20 показаны графики остатков, построенные при анализе четырех методов прогнозирования с помощью *Пакета анализа* Excel. Делая выводы на основе этих графиков, следует быть осторожным, поскольку временной ряд содержит только 20 точек. Методы построения см. соответствующий лист Excel-файла.

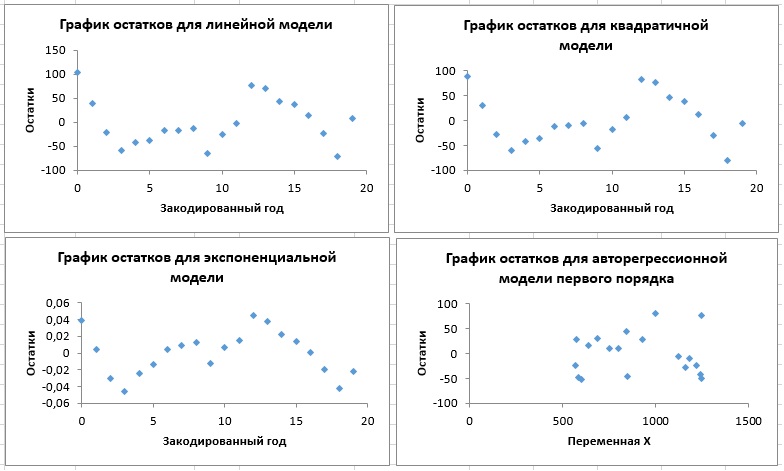


Рис. 20. Графики остатков, построенные при анализе четырех методов прогнозирования с помощью *Пакета анализа* Excel

Ни одна модель, кроме авторегрессионой модели первого порядка, не учитывает циклический компонент. Именно эта модель лучше других аппроксимирует наблюдения и характеризуется наименее систематической структурой. Итак, анализ остатков всех четырех методов показал, что наилучшей является авторегрессионная модель первого порядка, а линейная, квадратичная и экспоненциальная модели имеют меньшую точность. Чтобы убедиться в этом, сравним величины остаточных погрешностей этих методов (рис. 21). С методикой расчетов можно ознакомиться, открыв Excel-файл. На рис. 21 указаны фактические значения *Yi* (колонка *Реальный доход*), предсказанные значения *Ŷi*, а также остатки *еi* для каждой из четырех моделей. Кроме того, показаны значения *SYX* и *MAD*. Для всех четырех моделей величинs *SYX* и *MAD* примерно одинаковые. Экспоненциальная модель является относительно худшей, а линейная и квадратичная модели превосходят ее по точности. Как и ожидалось, наименьшие величины *SYX* и *MAD* имеет авторегрессионная модель первого порядка.

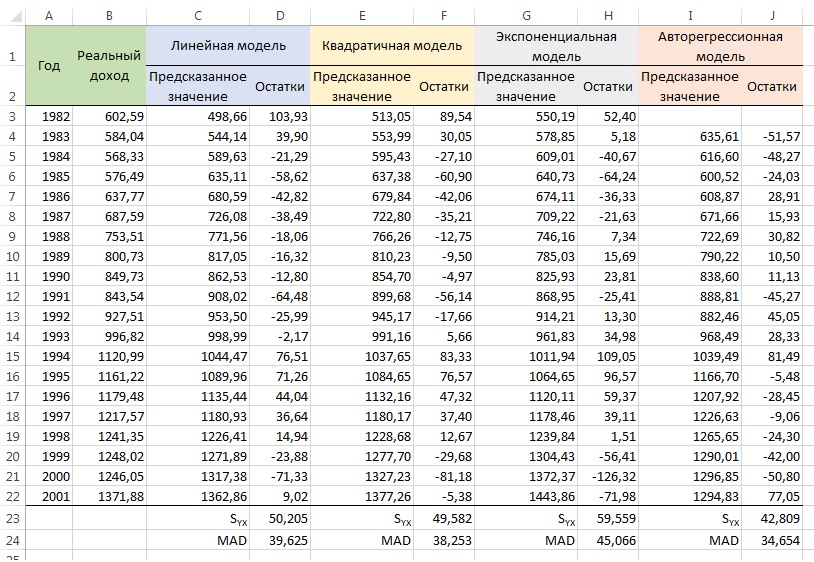


Рис. 21. Сравнение четырех методов прогнозирования с помощью показателей SYX и MAD

Выбрав конкретную модель прогнозирования, необходимо внимательно следить за дальнейшими изменениями временного ряда. Помимо всего прочего, такая модель создается, чтобы правильно предсказывать значения временного ряда в будущем. К сожалению, такие модели прогнозирования плохо учитывают изменения в структуре временного ряда. Совершенно необходимо сравнивать не только остаточную погрешность, но и точность прогнозирования будущих значений временного ряда, полученную с помощью других моделей. Измерив новую величину *Yi* в наблюдаемом интервале времени, ее необходимо тотчас же сравнить с предсказанным значением. Если разница слишком велика, модель прогнозирования следует пересмотреть.

**Прогнозирование временн*ы*х рядов на основе сезонных данных**

До сих пор мы изучали временные ряды, состоящие из годовых данных. Однако многие временные ряды состоят из величин, измеряемых ежеквартально, ежемесячно, еженедельно, ежедневно и даже ежечасно. Как показано на рис. 2, если данные измеряются ежемесячно или ежеквартально, следует учитывать сезонный компонент. В этом разделе мы рассмотрим методы, позволяющие прогнозировать значения таких временных рядов.

В сценарии, описанном в начале главы, упоминалась компания Wal-Mart Stores, Inc. Рыночная капитализация компании 229 млрд. долл. Ее акции котируются на Нью-Йоркской фондовой бирже под аббревиатурой WMT. Финансовый год компании заканчивается 31 января, поэтому в четвертый квартал 2002 года включаются ноябрь и декабрь 2001 года, а также январь 2002 года. Временной ряд квартальных доходов компании приведен на рис. 22.

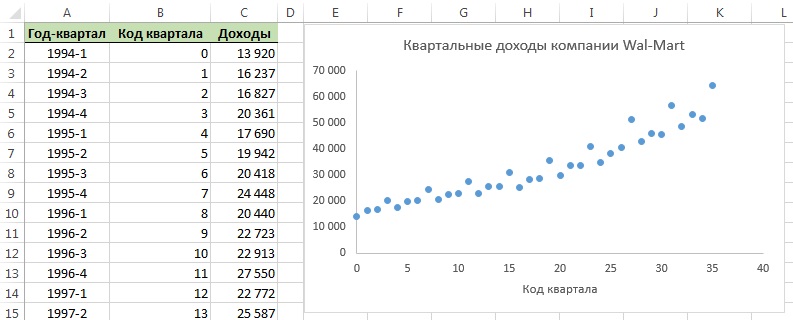


Рис. 22. Квартальные доходы компании Wal-Mart Stores, Inc. (млн. долл.)

Для таких квартальных рядов, как этот, классическая мультипликативная модель, кроме тренда, циклического и случайного компонента, содержит сезонный компонент: *Yi = Ti\*Si\*Ci\*Ii*

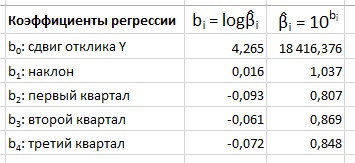
**Прогнозирование месячных и временн*ы*х рядов с помощью метода наименьших квадратов.** Регрессионная модель, включающая сезонный компонент, основана на комбинированном подходе. Для вычисления тренда применяется метод наименьших квадратов, описанный ранее, а для учета сезонного компонента — категорийная переменная (подробнее см. [Введение в множественную регрессию](http://baguzin.ru/wp/?p=6136) раздел *Регрессионные модели с фиктивной переменной и эффекты взаимодействия*). Для аппроксимации временных рядов с учетом сезонных компонентов используется экспоненциальная модель. В модели, аппроксимирующей квартальный временной ряд, для учета четырех кварталов нам понадобились три фиктивные переменные *Q1*, *Q2* и *Q3*, а в модели для месячного временного ряда 12 месяцев представляются с помощью 11 фиктивных переменных. Поскольку в этих моделях в качестве отклика используется переменная log*Yi*, а не *Yi*, для вычисления настоящих регрессионных коэффициентов необходимо выполнить обратное преобразование.

Чтобы проиллюстрировать процесс построения модели, аппроксимирующей квартальный временной ряд, вернемся к доходам компании Wal-Mart. Параметры экспоненциальной модели, полученные с помощью *Пакета анализа* Excel, показаны на рис. 23.



Рис. 23. Регрессионный анализ квартальных доходов компании Wal-Mart Stores, Inc.

Видно, что экспоненциальная модель довольно хорошо аппроксимирует исходные данные. Коэффициент смешанной корреляции *r2* равен 99,4% (ячейки J5), скорректированный коэффициент смешанной корреляции — 99,3% (ячейки J6), тестовая *F*-статистика — 1 333,51 (ячейки M12), а *р*-значение равно 0,0000. При уровне значимости α = 0,05, каждый регрессионный коэффициент в классической мультипликативной модели временного ряда является статистически значимым. Применяя к ним операцию потенцирования, получаем следующие параметры:



Коэффициенты интерпретируются следующим образом.

* Параметр = 18 416,376, сдвиг зависимой переменной *Y*, является значением некорректированного тренда квартальных доходов в первом квартале 1994 года, т.е. в первом временном периоде.
* Величина (1) х 100% = 3,66% оценивает темп роста квартальных доходов.
* Величина = 0,807 представляет собой сезонный множитель для первого квартала по отношению к четвертому кварталу. Это число означает, что доходы, полученные в первом квартале, на 19,3% меньше, чем доходы, полученные в четвертом квартале.
* Величина = 0,869 представляет собой сезонный множитель для второго квартала по отношению к четвертому. Это число означает, что доходы, полученные во втором квартале, на 13,1% меньше, чем доходы, полученные в четвертом квартале.
* Величина = 0,848 представляет собой сезонный множитель для третьего квартала по отношению к четвертому. Это число означает, что доходы, полученные в третьем квартале, на 15,2% меньше, чем доходы, полученные в четвертом квартале.

Используя регрессионные коэффициенты *bi*, можно предсказать доход, полученный компанией в конкретном квартале. Например, предскажем доход компании для четвертого квартала 2002 года (*Xi* = 35):

log*Ŷi* = *b0* + *b1Хi* = 4,265 + 0,016\*35 = 4,825

*Ŷi* = 104,825 = 66 834

Таким образом, согласно прогнозу в четвертом квартале 2002 года компания должна была получить доход, равный 67 млрд. долл. (вряд ли следует делать прогноз с точностью до миллиона). Для того чтобы распространить прогноз на период времени, находящийся за пределами временного ряда, например, на первый квартал 2003 года (*Xi* = 36, *Q1* = 1), необходимо выполнить следующие вычисления:

log*Ŷi* = *b0* + *b1Хi* + *b2Q1* = 4,265 + 0,016\*36 – 0,093\*1 = 4,748

*Ŷi* = 104,748 = 55 976

**Индексы**

Индексы используются в качестве индикаторов, реагирующих на изменения экономической ситуации или деловой активности. Существуют многочисленные разновидности индексов, в частности, индексы цен, количественные индексы, ценностные индексы и социологические индексы. В данном разделе мы рассмотрим лишь индекс цен. *Индекс* — величина некоторого экономического показателя (или группы показателей) в конкретный момент времени, выраженный в процентах от его значения в базовый момент времени.

**Индекс цен.** Простой индекс цен отражает процентное изменение цены товара (или группы товаров) в течение заданного периода времени по сравнению с ценой этого товара (или группы товаров) в конкретный момент времени в прошлом. При вычислении индекса цен прежде всего следует выбрать базовый промежуток времени — интервал времени в прошлом, с которым будут производиться сравнения. При выборе базового промежутка времени для конкретного индекса периоды экономической стабильности являются более предпочтительными по сравнению с периодами экономического подъема или спада. Кроме того, базовый промежуток не должен быть слишком удаленным во времени, чтобы на результаты сравнения не слишком сильно влияли изменения технологии и привычек потребителей. Индекс цен вычисляется по формуле:

*(9) Ii = \* 100*

где *Ii* — индекс цен в *i*-м году, *Рi* — цена в *i*-м году, *Рбаз* — цена в базовом году.

Индекс цен — процентное изменение цены товара (или группы товаров) в заданный период времени по отношению к цене товара в базовый момент времени. В качестве примера рассмотрим индекс цен на неэтилированный бензин в США в промежутке времени с 1980 по 2002 г. (рис. 24). Например:

*I2002 = \* 100 = \* 100 = 104,8*



Рис. 24. Цена галлона неэтилированного бензина и простой индекс цен в США с 1980 по 2002 г. (базовые годы — 1980 и 1995)

Итак, в 2002 г. цена неэтилированного бензина в США была на 4,8% больше, чем в 1980 г. Анализ рис. 24 показывает, что индекс цен в 1981 и 1982 гг. был больше индекса цен в 1980 г., а затем вплоть до 2000 года не превышал базового уровня. Поскольку в качестве базового периода выбран 1980 г., вероятно, имеет смысл выбрать более близкий год, например, 1995 г. Формула для пересчета индекса по отношению к новому базовому промежутку времени:

*(10) Iновый = \* 100*

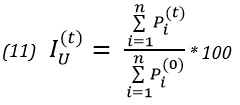
где *Iновый* — новый индекс цен, *Iстарый* — старый индекс цен, *Iновая* база – значение индекса цен в новом базовом году при расчете для старого базового года.

Предположим, что в качестве новой базы выбран 1995 год. Используя формулу (10), получаем новый индекс цен для 2002 года:

*Iновый = \* 100 = \* 100 = 113,9*

Итак, в 2002 г. неэтилированный бензин в США стоил на 13,9% больше, чем в 1995 г.

**Невзвешенные составные индексы цен.** Несмотря на то что индекс цен на любой отдельный товар представляет несомненный интерес, более важным является индекс цен на группу товаров, позволяющий оценить стоимость и уровень жизни большого количества потребителей. Невзвешенный составной индекс цен, определенный формулой (11), приписывает каждому отдельному виду товаров одинаковый вес. Составной индекс цен отражает процентное изменение цены группы товаров (часто называемой потребительской корзиной) в заданный период времени по отношению к цене этой группы товаров в базовый момент времени.



где *t* — период времени (0, 1, 2, ...), *i* — номер товара (1, 2, …, *n*), *n* — количество товаров в рассматриваемой группе,  — сумма цен на каждый из *n* товаров в период времени *t*,  — сумма цен на каждый из *n* товаров в нулевой период времени,  — величина невзвешенного составного индекса в период времени *t*.

На рис. 25 представлены средние цены на три вида фруктов за период с 1980 по 1999 гг. Для вычисления невзвешенного составного индекса цен в разные годы применяется формула (11), считая базовым 1980 год.

Итак, в 1999 г. суммарная цена фунта яблок, фунта бананов и фунта апельсинов на 59,4% превышала суммарную цену на эти фрукты в 1980 г.

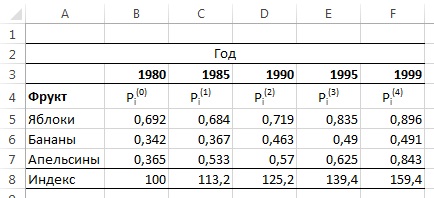
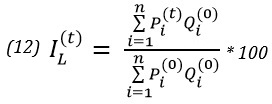


Рис. 25. Цены (в долл.) на три вида фруктов и невзвешенный составной индекс цен

Невзвешенный составной индекс цен выражает изменения цен на всю группу товаров с течением времени. Несмотря на то что этот индекс легко вычислять, у него есть два явных недостатка. Во-первых, при вычислении этого индекса все виды товаров считаются одинаково важными, поэтому дорогие товары приобретают излишнее влияние на индекс. Во-вторых, не все товары потребляются одинаково интенсивно, поэтому изменения цен на мало потребляемые товары слишком сильно влияют на невзвешенный индекс.

**Взвешенные составные индексы цен.** Из-за недостатков невзвешенных индексов цен более предпочтительными являются взвешенные индексы цен, учитывающие различия цен и уровней потребления товаров, образующих потребительскую корзину. Существуют два типа взвешенных составных индексов цен. *Индекс цен Лапейрэ*, определенный формулой (12), использует уровни потребления в базовом году. Взвешенный составной индекс цен позволяет учесть уровни потребления товаров, образующих потребительскую корзину, присваивая каждому товару определенный вес.



где *t* — период времени (0, 1, 2, ...), *i* — номер товара (1, 2, …, *n*), *n* — количество товаров в рассматриваемой группе,  — количество единиц товара *i* в нулевой период времени,  — значение индекса Лапейрэ в период времени *t*.

Вычисления индекса Лапейрэ показаны на рис. 26; в качестве базового используется 1980 год.

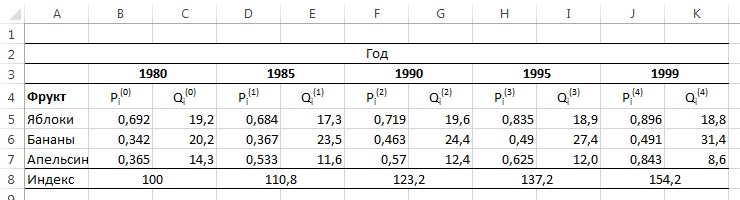
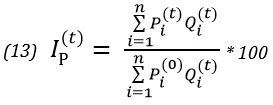


Рис. 26. Цены (в долл.), количество (потребление в фунтах на душу населения) трех видов фруктов и индекс Лапейрэ

Итак, индекс Лапейрэ в 1999 г. равен 154,2. Это свидетельствует от том, что в 1999 году эти три вида фруктов были на 54,2% дороже, чем в 1980 году. Обратите внимание на то, что этот индекс меньше невзвешенного индекса, равного 159,4, поскольку цены на апельсины — фрукты, потребляемые меньше остальных, — выросли больше, чем цена яблок и бананов. Иначе говоря, поскольку цены на фрукты, потребляемые наиболее интенсивно, выросли меньше, чем цены на апельсины, индекс Лапейрэ меньше невзвешенного составного индекса.

*Индекс цен Пааше* использует уровни потребления товара в текущем, а не базовом периоде времени. Следовательно, индекс Пааше более точно отражает полную стоимость потребления товаров в заданный момент времени. Однако этот индекс имеет два существенных недостатка. Во-первых, как правило, текущие уровни потребления трудно определить. По этой причине многие популярные индексы используют индекс Лапейрэ, а не индекс Пааше. Во-вторых, если цена некоторого конкретного товара, входящего в потребительскую корзину, резко возрастает, покупатели снижают уровень его потребления по необходимости, а не вследствие изменения вкусов. Индекс Пааше вычисляется по формуле:



где *t* — период времени (0, 1, 2, ...), *i* — номер товара (1, 2, …, *n*), *n* — количество товаров в рассматриваемой группе,  — количество единиц товара *i* в нулевой период времени,  — значение индекса Пааше в период времени *t*.

Вычисления индекса Пааше показаны на рис. 27; в качестве базового используется 1980 год.

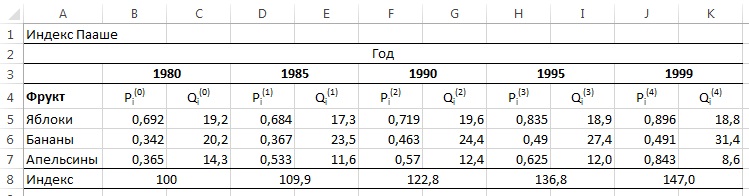


Рис. 27. Цены (в долл.), количество (потребление в фунтах на душу населения) трех видов фруктов и индекс Пааше

Итак, индекс Пааше в 1999 г. равен 147,0. Это свидетельствует от том, что в 1999 году эти три вида фруктов были на 47,0% дороже, чем в 1980 году.

**Некоторые популярные индексы цен.** В бизнесе и экономике используется несколько индексов цен. Наиболее популярным является индекс потребительских цен (Consumer Index Price — CPI). Официально этот индекс называется CPI-U, чтобы подчеркнуть, что он вычисляется для городов (urban), хотя, как правило, его называют просто CPI. Этот индекс ежемесячно публикуется Бюро статистики труда (U. S. Bureau of Labor Statistics) в качестве основного инструмента для измерения стоимости жизни в США. Индекс потребительских цен является составным и взвешенным по методу Лапейрэ. При его вычислении используются цены 400 наиболее широко потребляемых продуктов, видов одежды, транспортных, медицинских и коммунальных услуг. В данный момент при вычислении этого индекса в качестве базового используется период 1982–1984 гг. (рис. 28). Важной функцией индекса CPI является его использование в качестве дефлятора. Индекс CPI используется для пересчета фактических цен в реальные путем умножения каждой цены на коэффициент 100/CPI. Расчеты показывают, что за последние 30 лет среднегодовые темпы инфляции в США составили 2,9%.

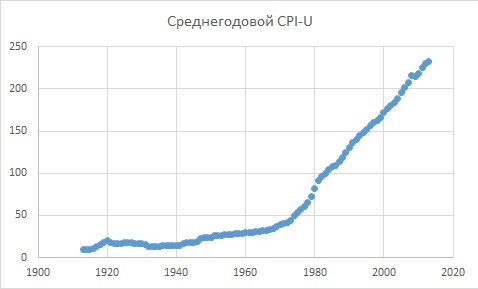


Рис. 28. Динамика Consumer Index Price; полные данные см. Excel-файл

Другим важным индексом цен, публикуемым Бюро статистики труда, является индекс цен производителей (Producer Price Index — PPI). Индекс PPI является взвешенным составным индексом, использующим метод Лапейрэ для оценки изменения цен товаров, продаваемых их производителями. Индекс PPI является лидирующим индикатором для индекса CPI. Иначе говоря, увеличение индекса PPI приводит к увеличению индекса CPI, и наоборот, уменьшение индекса PPI приводит к уменьшению индекса CPI. Финансовые индексы, такие как индекс Доу-Джонса для акций промышленных предприятий (Dow Jones Industrial Average — DJIA), S&P 500 и NASDAQ, используются для оценки изменения стоимости акций в США. Многие индексы позволяют оценить прибыльность международных фондовых рынков. К таким индексам относятся индекс Nikkei в Японии, Dax 30 в Германии и SSE Composite в Китае.

**Ловушки, связанные с анализом временн*ы*х рядов**

Значение методологии, использующей информацию о прошлом и настоящем для того, чтобы прогнозировать будущее, более двухсот лет назад красноречиво описал государственный деятель Патрик Генри: «У меня есть лишь одна лампа, освещающая путь, — мой опыт. Только знание прошлого позволяет судить о будущем».

Анализ временных рядов основан на предположении, что факторы, влиявшие на деловую активность в прошлом и влияющие в настоящем, будут действовать и в будущем. Если это правда, анализ временных рядов представляет собой эффективное средство прогнозирования и управления. Однако критики классических методов, основанных на анализе временных рядов, утверждают, что эти методы слишком наивны и примитивны. Иначе говоря, математическая модель, учитывающая факторы, действовавшие в прошлом, не должна механически экстраполировать тренды в будущее без учета экспертных оценок, опыта деловой активности, изменения технологии, а также привычек и потребностей людей. Пытаясь исправить это положение, в последние годы специалисты по эконометрии разрабатывали сложные компьютерные модели экономической активности, учитывающие перечисленные выше факторы.

Тем не менее, методы анализа временных рядов представляют собой превосходный инструмент прогнозирования (как краткосрочного, так и долгосрочного), если они применяются правильно, в сочетании с другими методами прогнозирования, а также с учетом экспертных оценок и опыта.

**Резюме.** В заметке с помощью анализа временных рядов разработаны модели для прогнозирования доходов трех компаний: Wm. Wrigley Jr. Company, Cabot Corporation и Wal-Mart. Описаны компоненты временного ряда, а также несколько подходов к прогнозированию годовых временных рядов — метод скользящих средних, метод экспоненциального сглаживания, линейная, квадратичная и экспоненциальная модели, а также авторегрессионная модель. Рассмотрена регрессионная модель, содержащая фиктивные переменные, соответствующие сезонному компоненту. Показано применение метода наименьших квадратов для прогнозирования месячных и квартальных временных рядов (рис. 29).

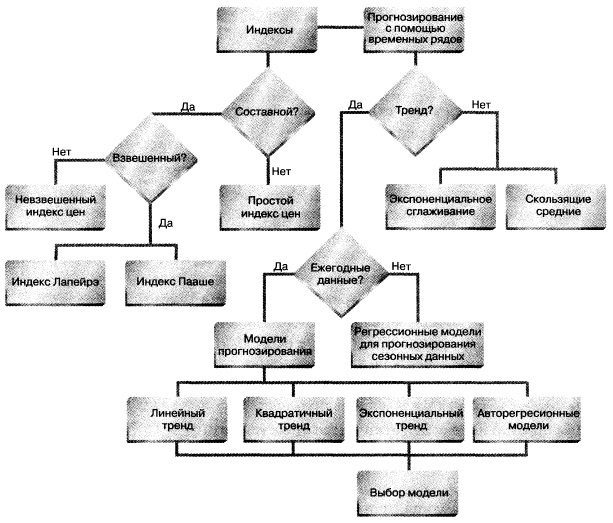


Рис. 29. Структурная схема заметки

Предыдущая заметка [Построение моделей множественной регрессии](http://baguzin.ru/wp/?p=6198)

Следующая заметка

К оглавлению [Статистика для менеджеров с использованием Microsoft Excel](http://baguzin.ru/wp/?p=5285)

1. Используются материалы книги Левин и др. Статистика для менеджеров. – М.: Вильямс, 2004. – с. 983–1074 [↑](#footnote-ref-1)
2. Тестовая *t*-статистика теряет *р* степеней свободы при оценке наклона и одну при оценке сдвига генеральной совокупности откликов. Еще *р* степеней свободы утрачиваются при сравнении значений временного ряда. [↑](#footnote-ref-2)