Анализ временных рядов

Учебно-методическое пособие

Белгород 2017

Составители: Ерина Т.А., Кузьмичева Т.Г.

В учебно-методическом пособии излагаются методические указания по изучению тем одного из разделов эконометрики – временные ряды и прогнозирование. Также даны указания для изучения статистических моделей с панельными данными.

Приведено достаточное количество примеров по построению и анализу различных моделей временных рядов. Представлены варианты индивидуальных заданий для студентов.

Для студентов, бакалавров и магистров экономических направлений и специальностей вузов, а также для специалистов по прикладной экономике и финансам.

Содержание

МЕТОДИЧЕСКИЕ УКАЗАНИЯ ПО ИЗУЧЕНИЮ ТЕМ

1. Методологические аспекты финансовой математики и экономико-математического моделирования

Любую сложную систему можно охарактеризовать многими одновременно присущими ей специфическими чертами. Выбор и использование при моделировании объекта тех или иных его сторон зависит от цели исследования, от того, какие конкретные ответы должны быть получены в результате решения задачи. В данном курсе финансовая система будет рассматриваться как сложная система, которой присущи прежде всего признаки экономических систем: динамичность и инерционность развития, саморегулирование и адаптация, стохастичность и детерминированность и др., а также специфические характеристики, наиболее важными из которых являются три взаимосвязанные величины: размер платежа, срок (время) и процентная ставка.

Моделирование каждой из этих величин, их взаимосвязи и влияние на них других финансово-экономических факторов позволяет решить следующие задачи:

разрабатывать прогнозы изменения каждого из финансовых показателей для различных периодов упреждения;

составлять финансовые планы, в том числе и оптимальные; измерять конечные финансовые результаты операций для всех участвующих в ней сторон;

выявлять зависимость конечных результатов от основных параметров операций или сделки, измерять взаимосвязь этих параметров, определять их допустимые значения и т.д.

Методологической основой решения указанных задач служат методы количественного финансового анализа, а также традиционные экономико-математические методы и модели: методы выявления тенденций, экстраполяции и адаптивные модели, корреляционно-регрессионный анализ, математического программирования и т.д.

2. Математические методы анализа финансово-экономических показателей

В настоящее время особенностью математико-статистического и количественно-финансового анализа, а также прогнозирования экономических показателей является учет временного фактора. Этот фактор, особенно при планировании и прогнозировании долгосрочных операций, играет не меньшую роль, чем размеры денежных сумм. В процессе инвестирования и кредитования это связано прежде всего с тем, что теоретически любая сумма денег может быть инвестирована под проценты и принести доход.

При математическом анализе экономических процессов и их прогнозировании, когда в качестве исходной информации используются временные (динамические) ряды, фактор времени определяется особенностями временных рядов, их отличием от простых статистических выборок в фиксированный момент времени. Эти особенности состоят в следующем:

последовательные по времени уровни временных рядов являются взаимозависимыми, особенно это относится к близко расположенным наблюдениям;

в зависимости от момента наблюдения уровни во временных рядах обладают разной информативностью: информационная ценность наблюдений убывает по мере их удаления от текущего момента времени;

с увеличением количества уровней временного ряда точность статистических характеристик не будет увеличиваться пропорционально числу наблюдений, а при появлении новых закономерностей развития она может даже уменьшаться.

Таким образом, учет фактора времени в финансово-экономическом анализе и прогнозировании является непременным условием, а его игнорирование может привести к неправильным оценкам и принятию ошибочных решений.

Анализ временных рядов, отражающих развитие экономических процессов, начинается с **оценки данных.** Уровни исследуемого показателя обязательно должны быть сопоставимы и однородны, а для выявления тенденций, кроме этого, устойчивы и количество наблюдений достаточно велико. **Сопоставимость** предполагает формирование всех уровней по одной и той же методике, использование одинаковой единицы измерения и шага наблюдений.

Требование **однородности** данных предполагает отсутствие сильных изломов тенденций, а также нетипичных, аномальных наблюдений. При поиске тенденций бывает целесообразно от

бросить часть прошлых данных, если они отражают уже утратившую силу закономерность прошлого развития. Наличие аномальных (резко выделяющихся) наблюдений приводит к искажению результатов. Формально она проявляется как сильный скачок (спад) с последующим приблизительным восстановлением предыдущего уровня. Примером такого наблюдения может служить значение курса доллара, зафиксированное в "черный вторник". Для диагностики аномальных наблюдений и их устранения существуют формальные критерии, некоторые из них изложены в теме 4.

Устойчивость характеризует преобладание закономерности над случайностью в изменении уровней ряда. На графиках устойчивых временных рядов даже визуально прослеживается закономерность, на графиках неустойчивых рядов изменения последовательных уровней представляются хаотичными, и поэтому поиск закономерностей в формировании значений уровней таких рядов лишен смысла.

Требование **полноты данных** обусловливается тем, что закономерность может обнаружиться лишь при наличии минимально допустимого объема наблюдений. Если целью исследования является построение модели динамики, то число уровней ряда наблюдений должно быть не меньше 7-10.

Следующим этапом анализа экономических процессов является **выявление тенденций** в развитии исследуемого показателя. Тенденция проявляется не только в увеличении или уменьшении численного среднего текущего значения временного ряда, она присуща и другим его характеристикам: дисперсии, автокорреляции, корреляции с другими показателями и т.д. Наиболее важными из перечисленных тенденций являются тенденции среднего значения и дисперсии. Первую из них можно определить из графика исходных данных, а более точно — с использованием метода Фостера-Стьюарта - метода проверки существенности разности средних и других методов, подробное описание которых дано в [1].

Наличие тенденции среднего уровня на графике становится более ясным, когда на нем отражены сглаженные значения исходных данных. Процедура сглаживания полезна также при построении некоторых математических моделей. Наиболее распространенными методами сглаживания являются использование простой скользящей средней, взвешенной скользящей средней, метод экспоненциального сглаживания. Алгоритмы их применения, преимущества и недостатки подробно рассмотрены в работах [2,3].

Традиционными показателями, характеризующими развитие экономических процессов, были и **остаются показатели роста и прироста.** Следует обратить особое внимание на область их рационального применения и технику расчета, особенно средних показателей как абсолютных, так и относительных.

Для характеристики динамики изменения экономических показателей все чаще используется понятие **автокорреляции,** которое характеризует не только взаимозависимость уровней одного и того же ряда, относящихся к разным моментам наблюдений, но и степень устойчивости развития процесса во времени, величину оптимального периода прогнозирования и т.п. Например, коэффициент автокорреляции процентных ставок депозитов, межбанковского кредита и кредитования предприятий, вычисленный для временных рядов, уровни которых представлены месячными данными, при лаге (сдвиге), равном единице, равняется 0,70-0,85, при лаге в два шага его значения колеблются в пределах 0,30-0,55, а при больших сдвигах значения быстро приближаются к нулю. Следовательно, прогноз этих показателей в настоящее время следует ограничить лишь одним-двумя месяцами, поскольку на более длительный период взаимосвязь не прослеживается.

При использовании месячных наблюдений, например 12 уровней, некоторые исследователи считают возможным давать прогноз на 2-3 шага, так как объем наблюдений кажется достаточно большим, а горизонт прогнозирования не слишком далеким. Однако это возможно лишь для тех экономических показателей, которые не подвержены внутригодичным устойчивым колебаниям (сезонности). В противном случае необходимо использовать методы и модели, способные наряду с тенденцией улавливать сезонные колебания. Эти методы и способы их реализации на ПЭВМ рассмотрены в работе [2].

3. Методологические вопросы экономического прогнозирования

Экономическое прогнозирование (ЭП) — это процесс разработки экономических прогнозов, основанных на научных методах познания экономических явлений и использовании всей совокупности методов, средств и способов экономической про

гностики. В то же время ЭП является частью **прогностики -** прикладной научной дисциплины, изучающей закономерности и способы разработки прогнозов развития объектов любой природы. ЭП рассматривает в качестве объекта процесс конкретного расширенного воспроизводства, а в качестве предмета - познание возможных состояний функционирующих экономических объектов в будущем, исследование закономерностей и способов разработки экономических прогнозов.

Основным содержанием ЭП является качественный и количественный анализ реальных экономических (в том числе финансово-кредитных) процессов, выявления объективных условий, факторов и тенденций развития. Содержание ЭП предопределяет основные *принципы разработки прогнозов*, к числу которых относятся системность, адекватность и альтернативность.

Системность ЭП означает, что явление рассматривается, с одной стороны, как единое целое, а с другой - как совокупность относительно самостоятельных направлений прогнозирования. Практическая реализация этого принципа предполагает создание моделей, которые соответствовали бы содержанию каждого отдельного блока и одновременно позволяли бы построить целостную картину возможного развития объекта в будущем.

Адекватность означает максимальное приближение теоретической модели к устойчивым, существенным закономерностям. Она предполагает учет реальных процессов, т.е. необходимость оценю! сложившихся и возможных отклонений от господствующих тенденций, определение возможной области рассеивания, что эквивалентно оценке вероятностной реализации выявленной тенденции. Практическое использование этого принципа означает, что построенные модели должны быть сначала проверены с точки зрения их способности имитировать уже сложившиеся тенденции.

При переходе от модельной имитации сложившихся закономерностей к предвидению процессов и тенденций будущего развития возникает необходимость построения альтернатив, т.е. определения возможных путей развития исследуемого объекта. **Альтернативность** прогнозирования связана с возможностью развития экономического объекта по разным траекториям, при разных взаимосвязях и структурных соотношениях. Главная проблема практической реализации этого принципа состоит в том, чтобы отделить те варианты развития, которые осуществимы, от вариантов, которые при сложившихся предвидимых условиях не могут быть реализованы. Среди всех вариантов наибольшую вероятность

реализации имеет обычно экстраполяционный, поскольку он исходит из сохранения сложившихся устойчивых условий и тенденций. Построение экстраполяционных вариантов позволяет установить "точку отсчета" для других вариантов.

Названные принципы лежат в основе конкретных методов ЭП. **Методом прогнозирования** называется способ исследования объекта прогнозирования, направленный на разработку прогнозов. **Модель прогнозирования** представляет собой модель исследуемого объекта, записанную в математической форме. Она должна позволить получить информацию о возможных состояниях объекта в будущем и (или) путях и сроках их осуществления. При построении прогнозной модели могут быть использованы один или несколько методов. Например, при построении линейной модели (уравнение прямой) могут быть использованы следующие методы: средних, двух точек, наименьших квадратов, экспоненциального сглаживания, гармонических весов, эволюции и др.

В большинстве методов, используемых в настоящее время в практической деятельности, заложены следующие **основные предположения:**

основные факторы, тенденции и зависимости, наблюдавшиеся в прошлом, сохранятся либо можно будет предвидеть и определить направление их изменения в прогнозируемом периоде;

развитие экономических процессов может быть представлено в виде плавной траектории, т.е. оно должно обладать некоторой инерционностью;

экономические процессы имеют вероятностный характер, а развитие исследуемого объекта определяется суммарным влиянием закономерности и случайности.

Последнее свойство обусловливает целесообразность использования статистических методов прогнозирования, которые при необходимости могут дополняться и другими методами (аналогий, экспертных оценок и т.д.). Эти предположения вполне разумны и находят подтверждение на практике. Так, например, в банковской деятельности ежедневные денежные потоки формируются под влиянием определенных закономерностей (запланированные платежи), а также необязательных, а порой и непредвиденных поступлений или платежей.

Процесс прогнозирования, опирающийся на статистические методы, распадается на **два основных этапа.** Первый (индуктивный) - *обобщение данных*, наблюдаемых за достаточно продолжительный период, и представление статистических закономерностей в виде модели, которая выражается либо аналитической функцией тенденции развития, либо в виде зависимости от нескольких факторов — аргументов. Второй этап — *собственно прогноз* — является дедуктивным. На основе выявленных закономерностей определяют ожидаемые значения прогнозируемого показателя, которые должны быть критически осмыслены с содержательной точки зрения. Указанные этапы конкретизируются в определенной последовательности шагов.

Обоснование теоретических предположений, являющихся исходными для составления прогноза, и выбор системы показателей, адекватно отображающих развитие объекта. Каждый показатель должен иметь экономическое содержание, отражать конкретный процесс и быть количественно измеримым.

Разработка системы моделей, отображающих развитие отдельных сторон, отдельных показателей исследуемого объекта, а также взаимосвязей между показателями.

Сбор и предварительный анализ данных.

Количественная оценка взаимосвязей показателей и параметров моделей. В зависимости от выбираемого критерия и численного метода оценки получаются разные результаты. Наибольшее распространение получил метод наименьших квадратов из-за простоты реализации и надежности результатов.

Определение прогнозных значений на основе построенной модели. Экстраполяция выявленных тенденций (продление на будущее) позволяет получить точечный прогноз. Однако вероятность точного попадания в эту точку практически равна нулю. Отсюда следует необходимость вычисления перспективных оценок в виде "вилки" через доверительные интервалы.

Методы ЭП, в частности, статистического прогнозирования, "осмысливают" лишь часть формализованной информации о прошлом развитии, в то время как специалист-аналитик владеет еще и значительным объемом слабо формализуемой, но очень важной информации о текущем и будущем развитии исследуемого объекта, а также его внешней среды. Поэтому полученную прогностическую информацию следует рассматривать как дополнительную, которая призвана помочь аналитику прояснить ситуацию и принять правильное решение.

Таким образом, ЭП требует глубокого анализа и учета объективных экономических законов общественного развития, конкретного экономического анализа исследуемого объекта, наличия достоверной количественной информации о прошлом развитии, технических и программных средств реализации методов прогнозирования, а также практических навыков по их использованию.

4. Экстраполяция тенденций и динамики развития финансово-экономических показателей

Основной формой представления информации о динамике финансово-экономических показателей являются временные ряды (ВР) наблюдений, т.е. ряды динамики, у которых в качестве признака упорядочения берется время. ВР, состоящий из N уровней $y(1)$, $y(2)$, $y(N)$, может быть записан в компактной форме: $Y(t)$, t= 1,2,..., N, где t- порядковый номер наблюдения.

Формально задача прогнозирования сводится к получению оценок значений ряда на некотором периоде будущего, т.е. к получению $Y_p(t)$ в момент времени $t=N+1,\ N+2...$ При использовании методов экстраполяции исходят из предположения о сохранении закономерностей прошлого развития на периоде прогнозирования.

Простейшим способом прогнозирования является подход от фактически достигнутого уровня $Y(N)$ при помощи среднего абсолютного прироста (САП), в соответствии с которым прогноз на $к$ шагов вперед на момент времени $t=N+k$ получается по формуле:

$$Y_p(N+k)=Y(N)+k*\text{САП}, \qquad (4.1)$$

$$\text{САП}=(Y(N)-Y(1))/(N-1) \qquad (4.2)$$

Этот способ является очень привлекательным из-за простоты и легкости реализации. Однако он имеет несколько существенных недостатков, один из которых состоит в невозможности сформировать интервал, внутрь которого попадет прогнозируемая величина, и указать степень уверенности в этом. В этой связи данный подход используется лишь как начальный ориентир будущего развития или же в условиях очень малого объема наблюдений, когда использовать статистические методы невозможно.

Статистические методы исследования исходят из предположения о возможности представления уровней ряда в виде суммы компонент, отражающих закономерность и случайность развития:

$$Y(t)=f(t)+E(t), \qquad (4.3)$$

где $f(t)$— тренд (долговременная тенденция) развития;

$E\ (t)$ — остаточная компонента.

Основной целью статистического анализа временных рядов является изучение соотношения между закономерностью и случайностью в формировании значений уровней ряда, оценка количственной меры их влияния. Закономерности, объясняющие динамику показателя в прошлом, могут быть использованы для прогнозирования его значений в будущем, а учет случайности позволяет определить вероятность отклонения от закономерного развития и их возможную величину.

Формирование уровней ряда определяется закономерностями трех основных типов: тенденцией среднего, тенденцией взаимосвязи между последовательными уровнями ряда и тенденцией взаимосвязи между исследуемым показателем и показателями-факторами, оказывающими на него причинное воздействие. Соответственно различают задачи анализа и моделирования тенденций, взаимосвязи между последовательными уровнями ряда, причинных взаимодействий между исследуемым показателем и показателями-факторами. Первая из них решается с помощью моделей кривых роста, вторая - с помощью адаптивных методов и моделей, третья - на основе эконометрического моделирования, базирующегося на методах корреляционно-регрессионного анализа.

Анализ и моделирование тенденций

Для отражения тенденции изменения f(t) исследуемого показателя в моделях кривых роста используются разнообразные математические функции, в которых задействован только один фактор - время (t). Из большого числа кривых роста воспользуемся простейшей линейной моделью вида:

$$Y_р(t)=a_0+a_1t \quad (t=1,2,\dots,N), \qquad (4.4)$$

Параметры кривых роста оцениваются по методу наименьших квадратов (МНК), т.е. подбираются таким образом, чтобы график функции кривой роста располагался на минимальном удалении от точек исходных данных. Математически критерий оценки параметров линейной модели записывается так:

$$\sum\{Y(t)-(a_0+a_1t)\}^2 \rightarrow min$$

Приравняв нулю частные производные по a_0 и a_1, можно получить систему двух уравнений с двумя неизвестными:

$$\begin{cases} a_0 N + a_1 \Sigma t = \Sigma Y(t) \\ a_0 \Sigma t + a_1 \Sigma t^2 = \Sigma(Y(t)t). \end{cases}$$

Суммирование проводится по всем уровням ряда. Для упрощения расчетов используется центрирование переменных относительно их среднего значения. Решением системы являются значения, вычисляемые по формулам:

$$a_1=\Sigma[(t-t_{cp})-(Y(t)-Y_{cp})] : \Sigma(t-t_{cp})^2, \qquad (4.5)$$

$$a_0=Y_{cp}-a_1 t_{cp},$$

где t_{cp} - среднее значение фактора времени;

Y_c - среднее значение исследуемого показателя.

Возьмем конкретные данные о курсе акций за девять недель и построим для них линейную модель. В табл. 4.1 приведены промежуточные вычисления и результаты использования линейной модели. В нижней строке записаны суммы значений в колонках. Исходя из первых двух, легко определить: Y_{cp} = 504:9 = 56, t_{cp} = 45 : 9 = 5. Подставляя рассчитанные суммы в вышеприведенную формулу, получаем:

a_1 = 429 : 60 = 7,2,

a_0 = 56 - 7,2 x 5 = 20,0.

Т а б л и ц а 4.1. Оценка параметров уравнения прямой

t	Факт $Y(t)$	$t-t_{cp}$	$(t-t_{cp})^2$	Y_t-Y_{cp}	$(t-t_{cp})*(Y_t-Y_{cp})$	Расчет $Y_p(t)$	Отклонение $E(t)$
1	**25**	**-4**	**16**	**-31**	**124**	**27,2**	**-2,2**
2	**34**	**-3**	**9**	**-22**	**66**	**34,4**	**-0,4**
3	**42**	**-2**	**4**	**-14**	**28**	**41,6**	**0,4**
4	**51**	**-1**	**1**	**-5**	**5**	**48,8**	**2,2**
5	**55**	**0**	**0**	**-1**	**0**	**56,0**	**-1,0**
6	**67**	**1**	**I**	**12**	**12**	**63,2**	**3,8**
7	**73**	**2**	**4**	**17**	**34**	**70,4**	**2,6**
8	**76**	**3**	**9**	**20**	**60**	**77,6**	**-1.6**
9	**81**	**4**	**16**	**25**	**100**	**84,8**	**-3,8**
45	**504**	**0**	**60**	**0**	**429**	**504**	**0**

Таким образом, линейная модель имеет вид:

$Y_p(t)$=20,0+7,2t (4.6)

Последовательно подставляя в модель вместо фактора t его значения от 1 до N, получаем расчетные значения уровней $Y_p(t)$:

$Y_p(1)$=20,0+7,2*1=27,2 (t=1),

$Y_p(2)$=20,0+7,2*2=34,4 и т.д. (t=2).

Вычислим отклонения расчетных значений от фактических наблюдений: $E(t) = Y(t)- Y_p(t)$ (t= 1,2,...,9):

$E(1)=Y(1)-Y_p(1)$=25,0-27,2=-2,2,

$E(2)=Y(2)-Y_p(2)$=34,0-34,4=-0,4 и т.д.

Отражает ли эта модель закономерность изменения исследуемого показателя, т.е. можно ли полученные значения)$Y_p(t)$ рассматривать как тенденцию? Для ответа на этот вопрос оценим ее качество.

Качество модели определяется ее **адекватностью** исследуемому процессу, которая характеризуется выполнением определенных статистических свойств, и **точностью,** т.е. степенью близости к фактическим данным. Модель считается хорошей со статистической точки зрения, если она адекватна и достаточно точна.

Модель является адекватной, если ряд остатков обладает свойствами случайности, независимости последовательных уровней и нормальности распределения. Результаты исследования адекватности отражены в табл. 4.2.

Т а б л и ц а 4.2. Оценка адекватности модели

t	Отклонение $E(t)$	Точки поворота	$E(t)^2$	$E(t+1)$	$E(t)-E(t+1)$	$[E(t)-E(t+1)^2]$	$E(t)$* $E(t+1)$	$E(t)$: $Y(t)$*100 (+)
	-2,2	—	4,84	-0,4	-1,8	3,24	0,88	8,8
2	-0,4	0	0,16	0,4	-0,8	0,64	-0,16	1,2
3	0,4	0	0,16	2,2	-1,8	3,24	0,88	1,0
4	2,2	I	2,20	-1,0	3,2	10,24	-2,20	4,3
5	-1,0	1	1,00	3,8	-4,8	23,04	-3,80	1,8
6	3,8	1	14,44	2,6	1,2	1,44	9.88	5.7
7	2,6	0	6,76	-1,6	4.2	17.64	-4,16	3.6
8	-1,6	0	2,56	-3,8	2,2	4,84	6,08	2,1
9	-3,8	—	14,44	—	-	—	-	4,7
Σ	0	3	26,56	-	-	64,32	7,40	33.2

Проверку **случайности** уровней ряда остатков проведем на основе **критерия поворотных точек.** В соответствии с ним каждый уровень ряда сравнивается с двумя рядом стоящими. Если он больше или меньше их, то эта точка считается поворотной. Далее подсчитывается сумма поворотных точек $"p"$. В случайном ряду чисел должно выполняться строгое неравенство:

$$P>[2(N\text{-}2)/3–\sqrt{(16N-29)/90}] \tag{4.7}$$

Квадратные скобки здесь означают, что от результата вычислений берется целая часть числа (не путать с процедурой округления!). При N =9 в правой части неравенства имеем:

$$[2*7/3\text{-}2\sqrt{115/90}] = [4{,}66\text{-}2*1{,}13]=[2{,}4]=2.$$

В табл. 4.2 в третьей графе для первого и последнего наблюдения проставим прочерк, ноль - если точка неповоротная, и единицу, если она поворотная. В нашем примере количество поворотных точек равно трем *(p*= 3), неравенство (4.7) выполняется, следовательно, свойство случайности выполняется.

При проверке **независимости** (отсутствия автокорреляции) определяется отсутствие в ряду остатков систематической составляющей. Это проверяется с помощью **«d-критерия Дарбина-Уотсона»,** в соответствии с которым вычисляется коэффициент d.

$$d=\sum_{t=2}^{N}[E(t)-E(t+1)]^2:\sum_{t=1}^{N}E(t)^2 \tag{4.8}$$

Вычисленная величина этого критерия сравнивается с двумя табличными уровнями (нижним d_1и верхним d_2). Если d- коэффициент (или $(d`))$ находится в интервале от нуля до d_1, то уровни рада остатков сильно автокоррелированы, а модель неадекватна. Если его значение попадает в интервал от d_2 до *2*, то уровни ряда являются независимыми. Если d-коэффициент превышает 2, то это свидетельствует об отрицательной корреляции и перед входом в таблицу его величину надо преобразовать: $d`$=4-$d$. В нашем примере $d$ = 64,32 : 25,56 = 2,52. Модифицируем полученную величину: $d`$=4 -2,52 = 1,48. Для линейной модели при 9 наблюдениях можно взять в качестве критических табличных уровней величины d_1= 1.08 и d_2= 1.36. Следовательно, расчетное значение d попало в зону подтверждения гипотезы о независимости уровней.

Если рассчитанная величина d попадает в зону между d_1 и d_2, то однозначного вывода сделать нельзя и необходимо применение других критериев, например, **первого коэффициента автокорреляции** $r(1)$, который вычисляется по формуле:

$$r(1)=\sum_{t=2}^{N-1} E(t) * E(t-1) : \sum_{t=1}^{N} E(t)^2. \qquad (4.9)$$

Если $|r(1)| > r$ (табл) (при N <15 r(табл) = 0,36), то присутствие в остаточном ряду существенной автокорреляции подтверждается. Для контроля сделанного по d-критерию вывода воспользуемся этим коэффициентом: $r(1)$= 7,40 : 25,56 = 0,29. Следовательно, по этому критерию также подтверждается выполнение свойства независимости уровней остаточной компоненты.

Соответствие ряда **остатков нормальному закону распределения** определим при помощи RS-критерия:

$$RS=(E_{max}-E_{min}):S, \qquad (4.10)$$

где E_{max}- максимальный уровень ряда остатков;

E_{min}- минимальный уровень ряда остатков;

S- среднее квадратическое отклонение.

Если значение этого критерия попадает между табулированными границами с заданным уровнем вероятности, то гипотеза о нормальном распределении ряда остатков принимается. Для N =10 и 5%-ного уровня значимости этот интервал равен (2,7-3.7).

В нашем примере: E_{max}=3,8 и E_{min} = -3,8, а размах 7,6.

$$S=\frac{\sqrt{E(t)^2}}{N-1} = \sqrt{26{,}56:(9-1)} = \sqrt{3{,}32 = 1{,}82,}$$

RS=(3,8-)-3,8):1,82=4,18.

Расчетное значение не попадает в интервал. Следовательно, свойство нормальности распределения не выполняется, что не позволяет строить доверительный интервал прогноза.

Для характеристики точности воспользуемся среднеквадратическим отклонением и средней относительной ошибкой:

$\bar{E}_{отн}=1/N*\sum_{t=1}^{N}\{|E(t)|:Y(t)*100\%\}=33{,}2:9=3{,}7\%$ (4.11)

Ее величина менее 5% свидетельствует об удовлетворительном уровне точности модели (ошибка в 10 и более процентов является очень большой).

Точечный прогноз на k шагов вперед получается путем подстановки в модель параметра $t=N+1,...,N+k$. При прогнозировании на два шага имеем:

$Y_p(10)=20{,}0+7{,}2*10=92{,}0$ ($k=1$, $t=10$),

$Y_p(11)=20{,}0+7{,}2*11=99{,}2$ ($k=2$, $t=11$).

Доверительный интервал прогноза будет иметь следующие границы:

Верхняя граница прогноза = $Y_p(N+k)+U(k)$ (4.12)

Нижняя граница прогноза = $Y_p(N+k)-U(k)$

Величина $U(k)$для линейной модели имеет вид:

$U(k)=S*K_p\sqrt{1+\frac{1}{N}+\left(N+k-t_{\text{ср}}\right)^2/\sum_{t=1}^{N}\left(t-t_{\text{ср}}\right)^2}.$ (4.13)

Коэффициент K_p является табличным значением t-статистики Стьюдента. Если исследователь задаст уровень вероятности попадания

прогнозируемой величины внутрь доверительного интервала, равный 70%, то K_p= 1,05.

$U(1)=1{,}82*1{,}05\sqrt{1+1/9+(9+1-5)^2:60}=2{,}4,$

$U(2)=1{,}82*1{,}05\sqrt{1+1/9+(9+2-5)^2:60}=2{,}4.$

В табл. 4.3 сведены результаты расчетов прогнозных оценок по формуле (4.12).

Т а б л и ц а 4 . 3 Прогнозные оценки по линейной модели

Время t	Шаг k	Прогноз $Y_p(t)$	Нижняя граница	Верхняя граница
10	**1**	**92,0**	**89,6**	**94,4**
11	**2**	**99,2**	**96,7**	**101,7**

Если построенная модель адекватна, то с выбранной пользователем вероятностью можно утверждать, что при сохранении сложившихся закономерностей развития прогнозируемая величина попадет в интервал, образованный нижней и верхней границами. В нашем случае такое утверждение не совсем правомерно из-за неполной адекватности модели.

Прогнозирование на основе адаптивных моделей

Для лучшего отображения особенностей изменения исследуемого показателя на конце периода наблюдения целесообразно использовать адаптивные модели [3,2,1], каждая из которых имеет определенный механизм приспособления к новым условиям. Общим для всех моделей этой группы является придание наибольшего веса последним наблюдениям при оценке параметров.

Для исследования динамики развития воспользуемся одной из таких моделей - моделью Брауна.

Расчетное значение в момент времениt получается по формуле:

$$Y_p(t) = a_0(t\text{-}1)+a_1(t\text{-}1)*k, \qquad (4.14)$$

где k- количество шагов прогнозирования (обычно k = 1).

Это значение сравнивается с фактическим уровнем и полученная ошибка прогноза $E(t) = Y(t)\text{-}Y_p(t)$ используется для корректировки модели. Корректировка параметров осуществляется по формулам:

$$a_1(t)=a_1(t\text{-}1)+a_0(t\text{-}1)+E(t)(1–B^2), \qquad (4.15)$$

$$a_0(t)=a_0(t\text{-}1)+ E(t) (1\text{-}B^2),$$

где B- коэффициент дисконтирования данных, отражающий большую степень доверия к более поздним данным. Его значение должно быть в интервале от 0 до 1.

Такой процесс модификации модели (t = 1,2,...,N) в зависимости от се текущих прогнозных качеств обеспечивает адаптацию к новым закономерностям развития. Для прогнозирования используется модель, полученная на последнем шаге (при $t= N$).

Воспользуемся этой схемой адаптивного прогнозирования. Начальные оценки параметров получим по первым пяти точкам при помощи МНК.

Т а б л и ц а 4.4. Оценка начальных значений параметров модели

t	Факт Y_t	t-t_{cp}	$(t$-$t_{cp})^2$	Y_t- t_{cp}	$(t$-$t_{cp})$*$(Y_t$-$Y_{cp})$
1	**25**	**-2**	**4**	**- 15,6**	**31,2**
2	**34**	**-1**	**1**	**-7,4**	**7,4**
3	**42**	**0**	**0**	**-0,6**	**0,0**
4	**51**	**I**	**1**	**9,6**	**9,6**
5	**55**	**2**	**4**	**13,6**	**27,2**
15	**207**	**0**	**10**	**0**	**75,4**

Используя эти результаты, получим: Y_{cp} = 207 : 5 = 41,4, t_{cp} = 3.

$a_1(0)$ = 75,4 : 10 = 7,5,

$a_o(0)$ = 41,4 - 7,5 * 3 = 19,9.

Возьмем k = 1 и B = 0,6. Подробно покажем расчет на первых двух шагах, а остальные отразим в табл. 4.5.

$t=1.$ $Y_p(1) = a_0(0) + a_1(0)$ *k= 19,9 + 7,5*1 = 27,4,

$E(1) = Y(1) - Y_p(1)$ = 25 - 27,5 = -2,4,

$a_0(1) = Y_p(1) + E(1)\ (1 - B^2)$ = 27,4 - 2,4 * 0,64 = 25,9,

$a_1(1) = a_1(0) + E(1)\ (1 - B)^2$ = 7,5 - 2,4 * 0,16 = 7,1.

$t=2.$ $Y_p(2)$ =$a_0(1)$+$a_1(1)$*k= 25,9 + 7,1*1 = 33,0,

$E(2)= Y(2) - Y_p(2)$ = 34 - 33,0 = 1,0.

$a_0(2) = Y_p(2)+ E(2)\ (1 - B^2)$ = 33.0 + 1,0 *0,64 = 33,6.

$a_1(2) = a_1(1) + E(2)(1- B)^2$ = 7,1 + 1,0 *0,16 = 7,3.

Таким образом, на последнем шаге получена модель:

$$Y_p\ (N+\kappa) = 82{,}3 + 6{,}3\ k. \qquad (4.16)$$

Т а б л и ц а 4.5 Оценка параметров модели

t	Факт $Y(t)$	$a_0(t)$	$a_1(t)$	Расчет $Y_p(t)$	Отклонение $E(t)$
0	**-**	**19,9**	**7,5**	—	—
1	**25**	**25,9**	**7,1**	**27,4**	***-2,4***
2	**34**	**33,6**	**7,3**	**33.0**	**1,0**
3	**42**	**41,6**	**7,5**	**40,9**	**1,1**
4	**51**	**50,0**	**7,7**	**49,1**	**1,9**
5	**55**	**56,7**	**7,3**	**57,7**	**-2,7**
6	**67**	**65,9**	**7,8**	**64,0**	**3,0**
7	**73**	**73,3**	**7,7**	**73,7**	**-0,7**
8	**76**	**77.8**	**6,9**	**81,0**	**-5,0**
9	**81**	**82,3'**	**6,3**	**84,7**	**-3,7**

Оценка ее качества на основе остаточной компоненты $E(t)$ дало следующие результаты: p = 4; d = 1,49; R_s = 2,8. Сопоставив эти значения с критическими уровнями, можно констатировать, что все свойства выполняются и, следовательно, построенная модель адекватна. Она имеет следующие точностные характеристики: S = 2,9; $E_{отн}$ = 4,5%.

Прогнозные оценки по модели (4.16) получаются путем подстановки в нее значения $k=1$ и k = 2, а интервальные — по тем же формулам, что и для кривых роста.

Y_p(10)=82,3+6,3*1=88,4,

Y_p(11)=82,3+6,3*2=94,5,

$U(1)=2{,}9*1{,}05\sqrt{1+1}/9+(9+1-5)^2:60=3{,}7$,

$U(2)=2{,}9*1{,}05\sqrt{1+1}/9+(9+2-5)^2:60=4{,}0$.

k = 1 (t = 10) k = 2 (t= 11)

Нижняя граница: 88,4 - 3,7 = 84,1 94,5 - 4,0 = 90,1.

Верхняя граница: 88,4 + 3,7 = 92,1 94,5 + 4,0 = 98,5.

Результаты вычислений сведены в табл. 4.

Т а б л и ц а 4 . 6 Прогнозные оценки по модели Брауна

Время t	Шаг k	Прогноз $Y_p(t)$	Нижняя граница	Верхняя граница
10	**1**	**88,4**	**84,1**	**92,1**
11	**2**	**94,5**	**90,1**	**98,5**

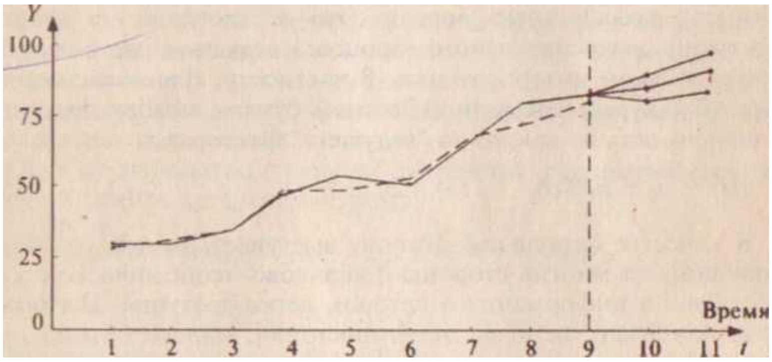

Рис. 4.1. Результаты аппроксимации и прогнозирования по адаптивной модели Брауна

Учитывая адекватность построенной модели, можно утверждать, что при сохранении сложившихся закономерностей динамики развития прогнозируемая величина с вероятностью 70% попадет в интервал, образованный нижней и верхней границами. На рис. 4.1 представлены результаты аппроксимации и прогнозирования по этой модели.

5. Анализ и прогнозирование финансово-экономических показателей с учетом ведущих факторов

В экономической деятельности часто требуется не только получить прогнозные оценки исследуемого показателя, но и количественно охарактеризовать степень влияния на него других факторов, а также возможные последствия их изменения в будущем. Для решения этой задачи используется аппарат корреляционно-регрессионного анализа.

Связь между зависимой переменной $Y(t)$ и m независимыми факторами можно охарактеризовать функцией регрессии $Y(t)=(X_1,X_2,\ldots,X_m)$, которая показывает, каково будет в среднем значение переменной Y, если переменные X примут конкретное значение. Это обстоятельство позволяет использовать модель регрессии не только для анализа, но и для прогнозирования. В качестве зависимой переменной может выступать практически любой финансово-экономический показатель, характеризующий, например, деятельность банка как предприятия или курс ценной бумаги, которая обращается на рынке достаточно долгое время.

В практике финансово-экономического анализа наибольшее распространение получили самые простые однофакторные линейные регрессионные модели. Это обусловлено не столько простотой вычислительного процесса, сколько ясностью их экономической интерпретации. В частности, для исследования динамики курса конкретной ценной бумаги можно построить ее зависимость от какого-то "ведущего" фактора:

$$Y(t)= a_o+ a_1X(t), \quad t=1,2,\ldots,N. \qquad (5.1)$$

В качестве "ведущего" обычно выступает фактор, серьезно влияющий на многие стороны финансово-экономической жизни страны и информация о котором легко доступна. Им может быть, например, цена на энергоносители, индекс объема продукции отрасли, взвешенная с учетом капитала сумма эффективности всех рисковых ценных бумаг, фигурирующих на рынке, т.е. различного рода индексы типа Доу-Джонса, Standard and Poorse SP-500 Index и др.

Отметим, что многие работники финансово-экономической сферы, используя некоторые результаты математико-аналитических исследований, не всегда имеют возможность достаточно точно оценить их степень надежности. В этой связи выполним основные этапы исследования взаимосвязи двух переменных и получим прогнозные оценки на два шага вперед, обратив особое внимание на вопросы статистической оценки достоверности результатов.

Итак, мы имеем данные о динамике изменения двух показателей за девять периодов, числовые значения которых приведены в первых трех графах табл. 5.1. Для конкретности будем рассматривать зависимую переменную как характеристику эффективности интересующей нас ценной бумаги, а в качестве фактора - показатель эффективности рынка ценных бумаг.

Прежде всего, оценим величину влияния фактора на исследуемый показатель при помощи коэффициента парной корреляции (результаты вычислений отражены в таблице):

$$r_{y,x} = \frac{\sum[(Y(t)-Y_{cp})*(X(t)-X_{cp})]}{\sqrt{\sum(Y(t)-Y_{cp})^2\sum(X(t)-X_{cp})^2}} = \frac{1049}{\sqrt{386*3125}} = 0.955.$$

Значение коэффициентов парной корреляции лежит в интервале от -1 до +1. Положительное его значение свидетельствует о прямой связи, отрицательное - об обратной, т.е. когда растет одна переменная, другая уменьшается. Чем ближе его значение к единице, тем теснее связь. Считается, что связь достаточно сильная, если коэффициент корреляции по абсолютной величине превышает 0,7, и слабая, если он меньше 0,3. При равенстве его нулю связь полностью отсутствует. В нашем случае можно утверждать о сильной прямой зависимости двух исследуемых показателей.

Оценка параметров модели регрессии осуществляется по МНК на основе следующих формул:

$$a_1=\Sigma[(X(t)-X_{cp})(Y(t)-Y_{cp})]/\Sigma(x(t)-X_{cp})^2, \qquad (5.2)$$

$$a_0=Y_{cp}-a_1*X_{cp}$$

Воспользуемся этими формулами, а необходимые промежуточные результаты вычислений приведем в табл. 5.1.

Таблица 5.1 Оценка параметров уравнения регрессии

t	$Y(t)$	$X(t)$	X_t-X_{cp}	$(X(t)$-$X_{cp})^2$	Y_t-Y_{cp}	$(Y(t)$-$Y_{cp})^2$	$(X_t$-$X_{cp})(Y_t$-$Y_{cp})$	Расчет $Y\,p(t)$	Отклонение $E(t)$
1	**25**	**45**	**-9**	**81**	**-31**	**961**	**276**	**31,7**	**-6,7**
2	**34**	**47**	**-7**	**49**	**-22**	**484**	**154**	**37,1**	**-3,1**
3	**42**	**50**	**-4**	**16**	**-14**	**196**	**56**	**45,2**	**-3,2**
4	**51**	**48**	**-6**	**36**	**-5**	**25**	**30**	**39,8**	**11,2**
5	**55**	**54**	**0**	**0**	**-1**	**1**	**0**	**56,0**	**-1,0**
6	**67**	**57**	**3**	**9**	**12**	**141**	**36**	**64,1**	**2,9**
7	**73**	**61**	**7**	**49**	**17**	**289**	**119**	**74,9**	**-1,9**
8	**76**	**59**	**5**	**25**	**20**	**400**	**100**	**69,5**	**6,5**
9	**81**	**65**	**11**	**121**	**25**	**625**	**275**	**85,7**	**-4,7**
Σ	**504**	**486**	**0**	**386**	**0**	**3125**	**1049**	**-**	**0**

a_1 =1049 : 386 = 2,7,

a_o = 56 - 2,7 х 54 = -89,8,

$Y_p(t)$ = -89,8 + 2,7 * $X(t)$, t=1,…,9. (5.2`)

Расчетные значения определяются путем последовательной подстановки в эту модель значения фактора:

$Y_p(1)$ = -89,8 + 2,7 * 45 - 31,7,

$Y_p(2)$ =-89,8 + 2,7*47 = 37,1 и т.д.

Оценка ее качества на основе остаточной компоненты $E(t)$ дала следующие результаты: p = 7; d = 2,26 (d' = 1,74); R_s = 3,1. Сопоставив эти значения с критическими уровнями, можно констатировать, что все свойства выполняются и, следовательно, построенная модель **адекватна.** Однако она имеет не очень хорошие точностные характеристики: S= 5,8; $E_{отн}$ = 8,8%. Следовательно, данную модель можно использовать для анализа, но она эффективна для получения прогнозных оценок.

Для расширенной характеристики модели регрессии вычислим несколько дополнительных показателей, первым из которых является коэффициент детерминации:

$R^2 = 1 - S^2_e : S^2_y = 1 - \Sigma E(t)^2 / \Sigma(Y(t) - Y_{cp})^2$ =1 - 267,4 : 3125 = 0,914. (5.3

Он показывает долю вариации результативного признака под воздействием изучаемых факторов. Следовательно, более 91% вариации зависимой переменной учтено в модели и обусловлено влиянием включенного фактора.

Учитывая, что коэффициенты регрессии нельзя использовать для непосредственной оценки влияния факторов на зависимую переменную из-за различия единиц измерения, используем **коэффициенты эластичности** *Э* и **бета-коэффициент,** которые рассчитываются соответственно по формулам:

$Э = a_1 * X_{ср} : Y_{ср} = 2{,}7 * 54 : 56 = 2{,}6,$ (5.4)

$\beta = a_1 * S_x : S_y = 2{,}7 * 386 : 3125 = 0{,}95.$ (5.5)

Коэффициент эластичности показывает, на сколько процентов изменяется зависимая переменная при изменении фактора на один процент. Следовательно, при изменении эффективности рынка на один процент эффективность нашей ценной бумаги увеличится на 2,6%.

Бета-коэффициент с математической точки зрения показывает, на какую часть величины среднего квадратического отклонения меняется среднее значение зависимой переменной с изменением независимой переменной на одно среднеквадратическое отклонение при фиксированном на постоянном уровне значении остальных независимых переменных. С экономической точки зрения он определяет влияние общей ситуации на рынке в целом на судьбу конкретной ценной бумаги. Если его величина положительна, то эффективность ценной бумаги аналогична эффективности рынка. Отрицательная величина бега- коэффициента свидетельствует о снижении эффективности ценной бумаги при повышении эффективности рынка. Этот коэффициент можно рассматривать как количественную меру риска инвестиций в ценные бумаги. При $\beta>1$ риск инвестиций выше, чем в среднем по рынку, а при $\beta<1$ - наоборот.

В нашем случае бета-коэффициент равен 0,95. Это свидетельствует о том, что при возрастании эффективности рынка будет возрастать эффективность исследуемой ценной бумаги, но риск инвестиций в нее несколько меньше среднерыночного.

Прогнозные значения $X_p(10)$ и $X_p(11)$ можно определить экспертно или вычислить на основе экстраполяционных методов. Получим прогнозные оценки фактора на основе величины его среднего прироста по соотношению (4.1):

САП=(65 -45):(9 - 1) =20 : 8 = 2,5,

$X_p(10) = X(9) + 2{,}5 * 1 = 65 + 2{,}5 = 67{,}5$,

$X_p(11) = X(9) + 2{,}5 * 2 = 65 + 5.0 = 70{,}0$.

Для получения прогнозных оценок зависимой переменной по модели (5.2) подставим в нее найденные прогнозные значения фактора:

$Y_p(10) = -89{,}8 + 2{,}7 * 67{,}5 = 92{,}5$ (t=10),

$Y_p(11) = -89{,}8 + 2{,}7 * 70.0 = 99{,}2$ (t=11).

Доверительный интервал прогноза будет иметь следующие границы:

Верхняя граница прогноза: $Y_p(N+k) + U(k)$,

Нижняя граница прогноза: $Y_p(N+k) - U(k)$.

Величина $U(k)$ для линейной модели регрессии имеет вид:

$$U(k) = S * k_p\sqrt{1 + 1/N + (X_p(N+k) - X_{cp})^2 / \sum X(t) - X_{cp})^2}$$

Для прогноза на два шага имеем:

$$U(1) = 5{,}8 \bullet 1{,}05\sqrt{1 + \frac{1}{9} + (67{,}5 - 54)^2 : 386} = 7{,}7,$$

$$U(2) = 5{,}8 \bullet 1{,}05\sqrt{1 + \frac{1}{9} + (70{,}0 - 54)^2 : 386} = 8{,}1.$$

Результаты вычисления прогнозов представим в табл. 5.2и для сравнения сведем их в табл**.** 5.3.

Т а б л и ц а 5.2 **Прогнозные оценки по модели регрессии**

Время t	Шаг k	Прогноз $Y_p(t)$	Нижняя граница	Верхняя фаница
10	**1**	**92,5**	**84,8**	**100,2**
11	**2**	**99,2**	**91,1**	**107,3**

Сравнивая точечные прогнозные оценки модели регрессии с оценками по линейной временной модели, можно отметить их явную близость, однако доверительный интервал регрессионной модели заметно шире, что снижает ее практическую значимость.

Адаптивная модель статистически полностью адекватна и имеет достаточно высокие точностные характеристики. Ее результаты можно взять в качестве прогноза.

Т а б л и ц а 5.3 Сводная таблица результатов исследования

Модель	Остаточная компонента			Адекватность	Средне-квадра-тическое	Среднее по модулю отклонение, %
	Независимость	Случайность	Нормальность			
$Y(t)$ = 20,0+7,2t	**да**	**да**	**нет**	**не полностью**	**1,8**	**3,7**
$Y(t)$ = 82,3+6,3k	**да**	**да**	**да**	**да**	**2,9**	**4,5**
Y(t)= -89,8+2,7X_t	**да**	**да**	**да**	**да**	**5,8**	**8,8**

6 . Статистические модели с панельными данными

Стандартные модели эконометрики с непрерывными данными в виде пространственной выборки или временного ряда не всегда позволяют адекватно оценить представляющие интерес параметры. В ряде случаев получаются неэффективные, иногда смещенные и несостоятельные оценки. В некоторых случаях применение стандартной модели и вовсе оказывается невозможным. В настоящей главе мы рассмотрим другие типы выборочных данных и связанные с ними экономические модели.

Данная тема посвящена панельным данным, которые в некотором роде оказываются объединением пространственных данных и временных рядов. Также здесь рассматриваются ограничения на возможные значения зависимой переменной и отбор наблюдений в выборку.

В экономических исследованиях нередко приходится рассматривать большие массивы данных в последовательные моменты времени. Это могут быть индивидуумы, домашние хозяйства, регионы, страны, у которых отслеживается динамика некоторых количественных показателей. Такие данные, сочетающие в себе пространственные выборки и временные ряды, называются панельными данными.

Пусть имеется n объектов, T временных наблюдений. Рассматривается зависимая переменная Y со значениями y_{ti}, где $t = 1,..., T$, $i = 1,\dots n$; и k - регрессоров X со значениями X_{ij}, $j = 1,..., k$, $t = 1,..., T$, $i = 1,\dots n$.

Во многих случаях допустимо предположение, что наблюдения в различные моменты времени не коррелируют друг с другом. Соответствующие модели называются *статическими*.

Возможны две крайние точки зрения на такую структуру наблюдений: их можно рассматривать как единый временной ряд с «толстым сечением», в котором наблюдения представляют собой п значений объясняемой переменной. В этом случае модель принимает вид

$$y_{ij} = \alpha + \sum_{j=1}^{k} \beta_j x_{j,ti} + \varepsilon_{ti} \tag{6.1}$$

и содержат $k+1$ параметр. По сути эта модель является классической, и ее следует оценивать обычным методом наименьших квадратов. В этой модели объекты панели лишены какой-либо индивидуальности и представляют собой наблюдения однородной выборки.

Противоположная точка зрения заключается в том, чтобы рассматривать данные как п различных временных рядов. То есть все объекты считаются абсолютно индивидуальными, и действия регрессоров на них различными. Тогда уравнения модели имеет вид

$$y_{ij} = \alpha_i + \sum_{j=1}^{k} \beta_{ji} x_{j,ti} + \varepsilon_{ti}$$

и содержат $kn + n$ параметров.

Модели с панельными данными позволяют сочетать оба этих подхода: объекты считаются индивидуальными, но изучается общее действие регрессоров на эти объекты. Один из возможных подходов заключается в следующем: регрессоры действуют на все объекты одинаково, но кроме рассматриваемых величин Xj, есть еще - как правило, ненаблюдаемые - факторы Q_a, которые, собственно, и обуславливают различия между объектами. Если считать факторы Q_a постоянными по времени, можно рассматривать уравнения

$$y_{ij} = \alpha + \sum_{j=1}^{k} \beta_j x_{j,ti} + \sum_{a=1}^{A} \Upsilon_a q_{ai} + \varepsilon_{ti} .$$

Обозначив $\alpha + \sum_{a=1}^{A} \Upsilon_a q_{ai_i} = \alpha_i$,

получим модель с *панельными данными* в виде

$$y_{ij} = \alpha_i + \sum_{j=1}^{k} \beta_j x_{j,ti} + \varepsilon_{ti} \qquad (6.2)$$

Модели с фиксированным и случайным эффектами

Индивидуальность объектов в панели данных может носить различный характер. Объекты могут быть существенно различными, и их нельзя рассматривать как выборку из некоторой генеральной совокупности. Таковыми, например, могут быть государства, регионы, крупные предприятия. В этом случае величины α_i, следует считать постоянными параметрами. Такие модели называются *моделями с фиксированным эффектом.*

В других случаях индивидуальность объектов носит случайный характер, то есть объекты панели считаются выборкой из некоторой генеральной совокупности. В этом случае характеризующие индивидуальность величины α_i, рассматриваются в виде

$$\alpha_i = \alpha + u_i,$$

где u_i— случайные величины, некоррелирующие с ε_{ti}, и регрессорами, и $M(u_i) = 0$.

Такие модели называются *моделями со случайным эффектом*.

С математической точки зрения типы эффектов различаются следующим образом: в моделях со случайным эффектом считается, что индивидуальные факторы не коррелируют с регрессорами, т.е.

факторы не коррелируют с регрессорами, т.е.

$$\rho(u_i, X_{j,i}) = 0 , \qquad (6.3)$$

между тем как в моделях с фиксированным эффектом равенство (6.3) не выполняется.

Рассмотрим следующий пример.

Оценивается модель производственной функции Кобба-Дугласа $Q = AK^{\alpha}L^{\beta}$ топливно-энергетических предприятий: Q - выпуск, K - капиталовложения, L- трудозатраты, $\alpha,\beta -$ соответствующие эластичности, которые и требуется оценить.

Для этого следует рассмотреть уравнения

$$\ln Q_{ti} = \ln A_i + \alpha \ln K_{ti} + \beta \ln L_{ti} \tag{6.4}$$

как регрессию на объясняемую переменную $\ln Q$ и объясняющие переменные InK, InL. При этом панельные данные содержат данные $t = 1, \cdots, 8, i = 1, \cdots, 48466$.

Если считать A - постоянной для всех предприятий величиной, то модель (6.4) оценивается с помощью обычного метода наименьших квадратов. При этом получается следующее уравнение

$$\text{In}Q = -2{,}48807 + 0{,}329571 \ln K + 0{,}92838 \ln L. \tag{6.5}$$

Все регрессоры значимы и $R^2 = 0{,}5805$.

Однако предприятия, очевидно, существенно различаются и по размерам, и по эффективности, и по качеству управления. Так что представление об однородности объектов панели скорее всего неправомочно, и модель следует оценивать с учетом индивидуальных эффектов.

Оценивание модели с фиксированным эффектом

Введем обозначения

$$
\mathrm{Y} = \begin{pmatrix} y_{11} \\ \cdots \\ \underline{y_{1n}} \\ y_{21} \\ \cdots \\ \underline{y_{2n}} \\ \underline{\cdots} \\ y_{T1} \\ \cdots \\ y_{Tn} \end{pmatrix}; \quad \mathrm{X} = \begin{pmatrix} x_{1,11} | & .. & .. & x_{k,T1} \\ .. & .. & .. & .. \\ .. & .. & .. & .. \\ x_{1,Tn} & .. & .. & x_{k,Tn} \end{pmatrix}
$$

$$
I = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ \dots & \dots & 1 & \dots \\ 0 & 0 & 0 & 1 \end{pmatrix}; \quad \varepsilon = \begin{pmatrix} \varepsilon_{11} \dots \varepsilon_{1n} \\ \varepsilon_{21} \dots \varepsilon_{2n} \\ \dots \\ \varepsilon_{T1} \dots \varepsilon_{Tn} \end{pmatrix}; \quad \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} \alpha_1 \dots \alpha_n \\ \beta_1 \dots \beta_n \end{pmatrix}. \qquad (6.6)
$$

Здесь $\mathrm{Y}(\mathrm{Tn} \times 1)$, $\mathrm{X}(\mathrm{Tn} \times \mathrm{k})$, $\mathrm{I}(\mathrm{Tn} \times \mathrm{n})$, $\varepsilon(\mathrm{Tn} \times 1)$, $\begin{pmatrix} \alpha \\ \beta \end{pmatrix}(\mathrm{n} + \mathrm{k}, 1)$.

Тогда *модель с фиксированным эффектом* может быть записана в виде

$$
\mathrm{Y} = \widetilde{\mathrm{X}} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} + \varepsilon \qquad (6.7)
$$

где $\widetilde{\mathrm{X}} = (\mathrm{I}|\mathrm{X}) - \big(\mathrm{nT} \times (\mathrm{n} + \mathrm{k})\big)$ матрица.

Применяя метод наименьших квадратов к уравнению (6.7), получим:

$$
\begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \widetilde{\mathrm{X}}' \big(\widetilde{\mathrm{X}}' \widetilde{\mathrm{X}}\big)^{-1} \mathrm{Y}. \qquad (6.8)
$$

Соответствующие оценки называются *оценками с фиксированным эффектом*. Как правило, наибольший интерес представляют оценки $\widehat{\beta}_{\mathrm{FE}}$ параметров β.

Непосредственным вычислением можно проверить утверждение:

$$
\widehat{\beta}_{\mathrm{FE}} = \widehat{\beta}_{\mathrm{W}} \qquad (6.9)
$$

Таким образом, использование внутригруппового усреднения позволяет получать оценки параметров модели с фиксированным эффектом. Параметры α_{i}, в свою очередь, могут быть оценены следующим образом:

$$
\widehat{\alpha}_{\mathrm{i}} = \bar{y}_{\mathrm{i}} - \sum_{\mathrm{j}=1}^{\mathrm{k}} \widehat{\beta}_{\mathrm{jFE}} \bar{x}_{\mathrm{ji}}.
$$

Подчеркнем, как правило, наибольший интерес представляют именно оценки параметров β.

Оценим уравнения (6.4), используя модель с фиксированным эффектом. Получим:

$$\ln Q = 0{,}75995 + 0{,}11421 \ln K + 0{,}60393 \ln L. \qquad (6.10)$$

Оценивание модели со случайным эффектом

Модель со случайным эффектом имеет следующий вид:

$$y_{ti} = \alpha + \sum_{j=1}^{k} \beta_j x_{j,ti} + u_i + \varepsilon_{ti}. \qquad (6.11)$$

Обозначим

$$U = \begin{pmatrix} u_1 \dots u_n \\ u_1 \dots u_n \\ \dots \\ u_1 \dots u_n \end{pmatrix} \text{– столбец } (nT \times 1).$$

Тогда в обозначениях (6.6) уравнение (6.11) записывается в виде:

$$Y = \alpha + X\beta + (u + \varepsilon). \qquad (6.12)$$

Уравнение (6.12) представляет собой обобщенную модель линейной регрессии: матрица ковариации ошибок регрессии $u + \varepsilon$имеет вид:

$$\Omega = \begin{pmatrix} (\sigma_u^2 + \sigma_\varepsilon^2)E_T & \sigma_u^2 E_T & \dots & \sigma_u^2 E_T \\ \sigma_u^2 E_T & (\sigma_u^2 + \sigma_\varepsilon^2)E_T & \dots & \sigma_u^2 E_T \\ \dots & \dots & \dots & \dots \\ \sigma_u^2 E_T & \sigma_u^2 E_T & \dots & (\sigma_u^2 + \sigma_\varepsilon^2)E_T \end{pmatrix} \qquad (6.13)$$

Здесь E_T— единичная матрица порядка T; $\Omega(nT \times nT)$. Для оценивания модели (6.12) можно применить метод наименьших квадратов. Приведем лишь окончательный качественный результат: оценка параметра β в модели со случайным эффектом имеет вид

$$\hat{\beta}_{RE} = W_{\hat{\beta}_B} + (E_k - W)\hat{\beta}_W \qquad (6.14)$$

где $W -$ матрица, пропорциональная обратной матрице ковариацией оценки $\hat{\beta}_B$.

Эта оценка представляет «средневзвешенное» межгрупповой и внутригрупповой оценок.

Для практического использования формулы (6.14) следует оценить матрицу Ω, т.е. параметры σ_u^2 и σ_ε^2. Как обычно, для оценок ошибок регрессии используются суммы квадратов остатков:

$$\hat{\sigma}_\varepsilon^2 = \frac{1}{nT - n - k}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(y_{ti} - \bar{y}_\imath - (x_{ti} - \bar{x})'\hat{\beta}_{FE}\right)^2,$$

$$\hat{\sigma}_B^2 = \frac{1}{n - k - 1}\sum_{i=1}^{n}\left(\bar{y}_\imath - \hat{\alpha}_B - \bar{x}_{ti}'\hat{\beta}_B\right)^2,$$

$$\hat{\sigma}_u^2 = \hat{\sigma}_B^2 - \frac{1}{T}\hat{\sigma}_\varepsilon^2.$$

Оценим теперь уравнение (6.4) с помощью модели со случайным эффектом. Получим:

$$\ln Q = -1{,}13749 + 0{,}24671 \ln K + 0{,}77544 \ln L. \qquad (6.15)$$

Как видно, получается существенное различие как с уравнением (6.10), так и с уравнением (6.5).

Проблема выбора модели с панельными данными

Рассмотренный выше пример показывает, что результат оценивания в значительной мере зависит от выбора модели, т.е. от структуры выборки.

При выборе статической модели панельных данных исследователь должен определить, считать ли объекты существенно индивидуальными, случайно выбранными из однородной генеральной совокупности, либо же лишенными всякой индивидуальности. Неверный выбор может привести либо к искажению результатов, либо к значительной потере эффективности.

Сразу следует заметить, что оценки модели с фиксированным эффектом $\left(\hat{\beta}_{BE}\right)$ в любом случае являются состоятельными и несмещенными. Поэтому в случае достаточно большой выборки (при большом T) их можно использовать всегда. Однако, если регрессоры не коррелируют с индивидуальными эффектами - так обстоит дело в моделях со случайным эффектом - оценки $\hat{\beta}_{RE}$ оказываются неэффективными, и это очень существенно для выборок небольшого объема. В то же время, если в моделях присутствует фиксированный эффект, оценки $\hat{\beta}_{RE}$ оказываются несостоятельными.

Проблему выбора модели исследователь решает индивидуально, часто на основании интуиции. При этом, однако, можно использовать статистические тесты.

Итак, пусть рассматривается модель

$$y_{ti} = \alpha_i + \sum_{j=1}^{k} \beta_j x_{j,ti} + \varepsilon_{ti}$$

и требуется сделать выбор между моделью с фиксированным и случайным эффектом (заметим, что классическую модель можно рассматривать и как частный случай модели с фиксированным эффектом, если все α_i, равны, и как частный случай модели со случайным эффектом, если $\sigma_u^2 = 0$).

Выбор делается в пользу случайного эффекта, если принимается гипотеза H_0:

$$r(\alpha_i, X_{j,ti}) = 0$$

Проверку гипотезы H_0 можно осуществить с помощью *теста Хаусмана*. Его идея основана на том, что в случае справедливости H_0обе оценки $\hat{\beta}_{FE}$и $\hat{\beta}_{RE}$ являются состоятельными, а, следовательно, не должны существенно различаться, т.е. матрица $\mathrm{Cov}(\hat{\beta}_{FE} - \hat{\beta}_{RE})$ не должна быть слишком большой. Статистика

$$(\hat{\beta}_{FE} - \hat{\beta}_{RE})' \mathrm{Cov}(\hat{\beta}_{FE} - \hat{\beta}_{RE})^{-1} (\hat{\beta}_{FE} - \hat{\beta}_{RE}) \qquad (6.16)$$

имеетχ^2 — распределение с k степенями свободы.

Для использования статистики (6.16) необходимо получить оценку матрицы $\mathrm{Cov}(\hat{\beta}_{FE} - \hat{\beta}_{RE})$. Можно показать, что при выполнении нулевой гипотезы выполняется асимптотическое равенство

$$\mathrm{Cov}(\hat{\beta}_{FE} - \hat{\beta}_{RE}) \approx \mathrm{Cov}(\hat{\beta}_{FE}) - \mathrm{Cov}(\hat{\beta}_{RE}).$$

При этом:

$$\mathrm{C\hat{o}v}(\hat{\beta}_{RE}) = \hat{\sigma}_{\varepsilon}^{2}\left(\sum_{i=1}^{n}\sum_{t=1}^{T}(x_{ti} - \bar{x}_i)(x_{ti} - \bar{x}_i)' + \frac{\hat{\sigma}_{\varepsilon}^{2}T}{\hat{\sigma}_{\varepsilon}^{2} + T\hat{\sigma}_{u}^{2}}\sum_{i=1}^{n}(\bar{x}_i - \bar{x})(\bar{x}_i - \bar{x})'\right)^{-1},$$

$$\mathrm{C\hat{o}v}(\hat{\beta}_{FE}) = \hat{\sigma}_{\varepsilon}^{2}\left(\sum_{i=1}^{n}\sum_{t=1}^{T}(x_{ti} - \bar{x})(x_{ti} - \bar{x}_i)'\right)^{-1},$$

где

$$\hat{\sigma}_{\varepsilon}^{2} = \frac{1}{nT - n - k} - \sum_{i=1}^{n}\sum_{t=1}^{T}\left(y_{ti} - \bar{y}_i - (\bar{x}_i - \bar{x})'\hat{\beta}_{FE}\right)^{2}.$$

Тест Хаусмана реализован в большинстве современных эконометрических программ.

Применим тест Хаусмана для оценки эластичностей производственной функции (уравнения (6.4)). Вычисленное значение статистики (6.16) равно $\chi^2 = 141{,}01$. При этом $\chi^2_{0,05;2}(X) = 5{,}99$, $\chi^2_{0,01;2}(X) = 9{,}21$, т.е. гипотеза отвергается при любом разумном уровне значимости. Это означает, что при оценивании уравнения (6.4) следует выбирать модель с фиксированным эффектом и применять в качестве оценки уравнение (6.10).

В данной теме рассматривались только статические модели. Динамические модели, учитывающие корреляцию между временными наблюдениями, оказываются значительно более сложными, и их рассмотрение выходит за рамки большинства стандартных курсов.

7. Вопросы для самопроверки

1. Назовите основные предпосылки и принципы прогнозирования на основе статистических экстраполяционных методов.

2. Какие достоинства и недостатки присущи традиционному способу прогнозирования на основе средних темпов роста и прироста?

3. Почему показатель среднего уровня редко используется для характеристики временных рядов?

4. Укажите особенности временных рядов наблюдений и требования к ним при использовании математических методов.

5. Каким образом влияют аномальные наблюдения на выявление закономерностей развития исследуемого показателя?

6. На какие компоненты можно разложить уровни временных радов годовых и месячных данных? Каким образом их можно исследовать?

7. Назовите и кратко охарактеризуйте основные этапы разработки прогноза с использованием статистических методов.

8. Опишите технологию разработки прогнозов на ЭВМ с использованием программ статистической обработки данных.

9. Каким образом учитывается информационная ценность данных при оценке параметров на основе МНК и адаптивных методов.

10. Дайте общую характеристику кривых роста и адаптивных моделей. Укажите их общие свойства и отличия.

11. Охарактеризуйте свойства полиномиальных и экспоненциальных временных моделей. Дайте содержательную интерпретацию их коэффициентов.

12. В чем заключается методика оценки качества математических моделей?

13. Что включает в себя понятие адекватности математических моделей прогнозирования? Какова методика ее определения?

14. Что такое точность математических моделей? Назовите ее основные характеристики. Может ли модель быть достаточно точной, но неадекватной?
15. Что представляет собой ретроспективный прогноз? Какова его роль для оценки точности математических моделей?
16. Какие факторы влияют на доверительный интервал прогноза? Как доверительная вероятность прогноза соотносится с практической значимостью результатов прогнозирования?
17. Какие ограничения накладываются на количество факторов, включаемых в регрессионную модель и чем они вызваны?
18. В каких случаях целесообразно включать в модель регрессии фактор "время"?
19. Какими статистическими средствами характеризуется взаимосвязь переменных?
20. Какое влияние на вычислительный процесс и результаты оказывает изменение статуса фактора с "независимой переменной" на "независимую обязательную переменную"?
21. Является ли высокое значение парного коэффициента корреляции свидетельством тесной взаимосвязи переменных?
22. Охарактеризуйте назначение коэффициентов регрессии, эластичности, бетта- и дельта-коэффициентов и их роль в содержательном анализе.
23. Какими средствами оценивается качество построенных регрессионных моделей?
24. Почему регрессионные модели, являющиеся более мощным инструментом исследования, на практике не всегда дают лучшие результаты по сравнению с временными экстраполяционными моделями?
25. Каким образом на основе регрессионной модели получается прогноз зависимой переменной? Какие факторы влияют на ширину доверительного интервала?

26. Что такое «панельные данные»?

27. Охарактеризуйте модели со случайным эффектом.

8. Индивидуальные задания

Номер Вашего варианта соответствует номеру по списку группы. В соответствии с ним из таблицы выберите показатель $Y(t)$, а данные фактора $X(t)$ возьмите из следующей по порядку строки.

На основе исходных данных об объеме производства продукции $Y(t)$ и производственных фондов $X(t)$ за семимесячный период наблюдения построить точечный и интервальный прогнозы на два шага вперед (для вероятности P= 70% используйте коэффициент K_p = 1,05) и сформулируйте свой вывод о выполненных расчетах.

1. Для зависимой переменной $Y(t)$постройте:

линейную модель $Y(t)=a_o+a_1*t$, параметры которой оцените с помощью МНК;

адаптивную модель Брауна $Y(t)=a_0+a_1k$,

линейную однопараметрическую модель регрессии $Y(t)=a_0+a_1*X(t)$.

2. Оцените качество построенных моделей, исследовав их адекватность и точность.

а) Адекватность моделей определите на основе исследования: случайности остаточной компоненты по критерию пиков; независимости уровней ряда остатков по d-критерию (в качестве критических используйте уровни d_1= 1,08 и d_2= 1.36) или по первому коэффициенту корреляции, критический уровень которого $r(1)$ = 0,36;

нормальности распределения остаточной компоненты по R/S- критерию (с критическими уровнями 2,5 - 3,4).

б) Для оценки точности модели используйте среднее квадратическое отклонение и среднюю по модулю ошибку.

3. Для модели регрессии рассчитайте парный коэффициент корреляции переменных, коэффициент эластичности и бета-коэффициент. Прогнозные оценки фактора *X(t)* на два шага вперед получите на основе среднего прироста от фактически достигнутого уровня.

4. Отобразите на графике фактические данные, результаты аппроксимации и прогнозирования по лучшей модели.

Вычисления проведите с одним знаком в дробной части. Основные промежуточные результаты вычислений представьте в таблицах.

Номер варианта	Значения Y(*t*) при *t*								
	1	2	3	4	5	6	7	8	9
1	10	14	21	24	33	41	44	47	49
2	43	47	50	48	54	57	61	59	65
3	3	7	10	11	15	17	21	25	23
4	30	28	33	37	40	42	44	49	47
5	5	7	10	12	15	18	20	23	26
6	12	15	16	19	17	20	24	25	28
7	20	27	30	41	45	51	53	55	61
8	8	13	15	19	25	27	33	35	40
9	45	43	40	36	38	34	31	28	25
10	33	35	40	41	45	47	45	51	53
11	10	15	21	23	25	34	32	37	41
12	16	20	22	20	25	23	25	28	30
13	12	17	20	21	25	27	24	28	31
14	20	22	24	26	25	29	35	38	43
15	25	30	36	41	38	43	47	45	50
16	80	75	78	72	69	70	64	61	59
17	24	22	26	29	33	31	28	33	36
18	30	34	40	38	42	48	50	52	53
19	88	85	84	86	81	80	83	78	76
20	50	52	54	59	57	60	63	68	70
21	25	27	30	31	35	41	42	45	47
22	75	77	73	70	66	63	67	63	61
23	28	34	32	36	39	42	45	41	46
24	15	20	24	30	33	37	36	40	42
25	70	74	76	75	78	78	83	85	87
26	82	79	78	72	69	70	64	61	59
27	25	27	26	29	32	32	30	33	35
28	32	34	38	40	42	46	50	52	53
29	90	87	85	86	82	80	81	78	76
30	55	57	54	59	57	60	63	66	64
31	12	15	18	22	25	31	32	37	41
32	26	30	32	30	35	33	35	38	40
33	62	67	80	81	85	87	84	88	91
34	18	21	24	26	25	29	34	38	41
35	28	32	36	40	38	43	45	48	50
36	82	77	78	72	69	70	67	64	62
37	28	24	26	29	33	31	28	33	35
38	32	34	41	38	42	48	50	52	55
39	90	88	84	86	82	80	81	78	76
40	56	58	60	63	67	66	70	72	74
41	35	37	40	41	45	51	52	55	57
42	65	67	63	60	56	53	57	53	51
43	29	33	32	36	38	41	44	42	46
44	35	40	44	50	53	57	56	60	62
45	72	74	76	75	79	78	82	85	89
46	85	81	78	72	69	70	64	61	56
47	23	27	26	29	32	34	36	41	45
48	30	34	36	40	41	46	49	52	53
49	95	89	85	86	82	80	81	76	73
50	52	54	55	59	60	62	63	66	70

Литература

1. Орлова И.В., Половников В.А., Федосеев В.В. Курс лекций по экономико-математическому моделированию.- М.: ВЗФИ, 1993.-148 с.
2. Горчаков А.А, Орлова И.В. Компьютерное экономико-математическое моделирование: Учебное пособие.- М.: Компьютер, ЮНИТИ, 1995.-170 с.
3. Карасев А.И. и др. Математические методы и модели в планировании. М.: Экономика, 1987. – 240 с.
4. Кремер Н.Ш., Путко Б.А., Эконометрика: учебник для студентов вузов.- М.: ЮНИТИ-ДАНА, 2010. – 328 с.