В первом случае, если опущены переменные, которые должны быть включены в регрессию, оценки коэффициентов $\hat{\beta}_j$, j=1,…,k являются, вообще говоря, смещенными (но обладают меньшей дисперсией) за исключением двух случаев, когда $\hat{\gamma}_j$=0, j=1,…,l или регрессоры $X_1,\ldots,X_k$ и $Z_1,\ldots,Z_l$ ортогональны.

Смещенной является и оценка дисперсии случайной ошибки σ_u^2, а, следовательно, стандартные ошибки и многие статистические тесты, в которых используется значение σ_u^2, становятся некорректными.

Во втором случае, если включены переменные, которые не должны присутствовать в модели, оценки коэффициентов $\hat{\beta}_j$, j=1,…,k будут несмещенными, но неэффективными. Поскольку несмещенность оценок и величины дисперсии σ_u^2 сохраняется, возникает иллюзия, что надо включать в модель как можно больше регрессоров. Но в этом случае падает точность оценок, и может возникнуть проблема мультиколлинеарности объясняющих переменных.

На практике, однако, нам неизвестен процесс, порождающий данные, т.е. мы не знаем истинную модель. Поэтому, как правило, возникает проблема – какую модель выбрать: короткую или длинную, т.е. включать дополнительные регрессоры в модель или не включать: в первом случае мы получим смещенные оценки коэффициентов регрессии, а во втором случае – неэффективные оценки. Решение этой проблемы может быть найдено на основе критерия минимума среднеквадратичного отклонения значений коэффициентов, см. [5, с. 112-114].

Часто случается также, что исследователь не может использовать данные по переменным, которые включены в модель. Некоторые переменные, например, невозможно измерить, другие поддаются измерению, но это достигается большими затратами времени и ресурсов. В таких случаях вместо отсутствующих переменных полезно использовать некоторые их заменители (proxy).

Например, если вы не имеете данных о качестве образования, вы можете использовать показатель качества образования как отношение числа преподавателей к числу студентов или денежные расходы на одного студента.

Причин использования "прокси"-переменных две: во-первых, если пропущена важная для модели переменная, то оценки будут смещены (случай 1 выше), а, во-вторых, результаты оценки регрессии с включением замещающих переменных могут дать косвенную информацию о тех переменных, которые замещены данными переменными.

4.2. Обобщенный метод наименьших квадратов

Обобщим КЛММР вида (3.1). Пусть по-прежнему мы располагаем выборочными наблюдениями над k переменными Y_i и X_{ji}, j=1,..., k, i=1,2,…,n и строим регрессию:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \ldots + \beta_k X_{ki} + u_i, \quad i = 1,2,\ldots,n \qquad (4.3)$$

Откажемся от предположения КЛММР о некоррелированности и гомоскедастичности случайной ошибки (3.3). То есть относительно переменных модели в уравнении (4.3) примем следующие основные гипотезы:

$$\mathrm{E}(u_i)=0; \qquad (4.4)$$

$$\mathrm{E}(u_i u_j) = \begin{cases} \sigma_i^2 & \text{при} \quad i = j, \\ \sigma_{ij} & \text{при} \quad i \neq j, \end{cases} \qquad (4.5)$$

$X_1, X_3, \ldots, X_k$ – неслучайные переменные; (4.6)

Не должно существовать строгой линейной зависимости между переменными $X_1, X_3, \ldots, X_k$. (4.7)

Суть гипотезы (4.5) в том, что все случайные ошибки u_i имеют непостоянную дисперсию, то есть не выполняется условие гомоскедастичности дисперсии – имеет место гетероскедастичность дисперсии ошибок. Кроме того, ковариации остатков могут быть произвольными и отличными от нуля (вторая строчка соотношения (4.5)).

Модель вида (4.3)-(4-7) называется обобщенной линейной моделью множественной регрессии (ОЛММР). Отличие ОЛММР от КЛММР состоит в изменении предположений о поведении случайной ошибки (4.5).

К ОЛММР может быть применен метод наименьших квадратов, однако (3.6) оказывается неприменимой к модели (4.3)-(4-7) в силу потери свойства оптимальности оценок. Но МНК к ОЛММР может быть применен.

Критерий минимизации суммы квадратов ошибок МНК в силу условия (4.5) заменяется на другой – минимизация обобщенной суммы квадратов отклонений (с учетом ненулевых ковариаций случайной ошибки для разных наблюдений и непостоянной дисперсии ошибки) и соответственно усложняется вид системы уравнений для определения оценок коэффициентов по сравнению с системой (3.6) для МНК. После решения полученной системы линейных алгебраических уравнений получим линейные несмещенные оценки коэффициентов ОЛММР, которые будут эффективными. Указанный метод получения оценок называется обобщенным методом наименьших квадратов (ОМНК) или методом Айткена.

Обозначим[6]:

[6] Этот абзац может быть опущен без ущерба для дальнейшего усвоения материала пособия.

$$\mathbf{y}=\begin{bmatrix}Y_1\\Y_2\\...\\Y_n\end{bmatrix};\ \mathbf{X}=\begin{bmatrix}1 & X_{21} & ... & X_{k1}\\1 & X_{22} & ... & X_{k2}\\... & & ... & \\1 & X_{2n} & ... & X_{kn}\end{bmatrix};\ \boldsymbol{\beta}=\begin{bmatrix}\beta_0\\\beta_2\\...\\\beta_k\end{bmatrix};\ \mathbf{u}=\begin{bmatrix}\mathrm{u}_1\\\mathrm{u}_2\\...\\\mathrm{u}_n\end{bmatrix};\Omega=\begin{bmatrix}\omega_{11} & ... & \omega_{1n}\\\omega_{21} & ... & \omega_{2n}\\... & ... & ...\\\omega_{n1} & ... & \omega_{nn}\end{bmatrix}.$$

Тогда модель (4.3)-(4.7) запишется в матричном виде:

$\boldsymbol{y}=\mathbf{X}\boldsymbol{\beta}+\mathbf{u}$,

при условиях

$\mathrm{E}(\mathbf{u})=\mathbf{0}$;

$\mathrm{E}(\mathbf{u}\mathbf{u}^{\mathrm{T}})=\sigma^2\boldsymbol{\Omega}$;

$\mathbf{X}$ – не из случайных чисел;

rank($\mathbf{X}$)=k+1<n.

Оценки МНК получаются по формуле $\hat{\boldsymbol{\beta}}=\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y}$. Оценки ОМНК получаются по формуле $\hat{\boldsymbol{\beta}}=\left(\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{y}$.

Подчеркнем, что для применения ОМНК в (4.5) необходимо знать значения в правой части равенства (в частности элементы матрицы $\boldsymbol{\Omega}$), что на практике случается крайне редко. Поэтому каким-либо способом оценивают величины σ_i^2, σ_{ij}, $i, j=1,\ldots,n$. А затем используют эти оценки в расчетах коэффициентов модели. Этот подход составляет суть так называемого <u>доступного обобщенного метода наименьших квадратов</u>. Конкретные способы оценки неизвестных ковариаций будут рассмотрены ниже.

4.3 Линейная модель множественной регрессии с гетероскедастичными остатками

Довольно часто при построении регрессии анализируемые объекты неоднородны, например, при исследовании структуры потребления домохозяйств естественно ожидать, что колебания в структуре будут выше для богатых, чем для бедных домохозяйств. В этой ситуации предположение (3.3) о постоянстве дисперсии случайной ошибки (имеется в виду возможное поведение случайного члена до того, как сделана выборка) оказывается не соответствующим действительности. В случаях, когда дисперсия *u* одинакова в каждый момент времени или для каждого значения *X*, существуют определенные ограничения (в некоторой полосе) для расположения точек на графике *X* и *Y*, согласно которым отчетливой тенденции к увеличению или уменьшению дисперсии σ_u^2 по мере роста *X* не наблюдается.

На рис. 4.1 приводятся примеры изменения разброса (гетероскедастичности) случайной ошибки регрессии.

На рис. 4.1а изображена ситуация, когда значения дисперсии σ_u^2 растут по мере увеличения значений регрессора *X*. На рис. 4.1б дисперсия ошибки достигает максимальной величины при средних значениях *X*, уменьшаясь по мере приближения к крайним значениям. Наконец, на рис. 4.1в дисперсия ошибки оказывается наибольшей при малых значениях *X*, быстро уменьшается и становится однородной по мере увеличения независимой переменной *X*.

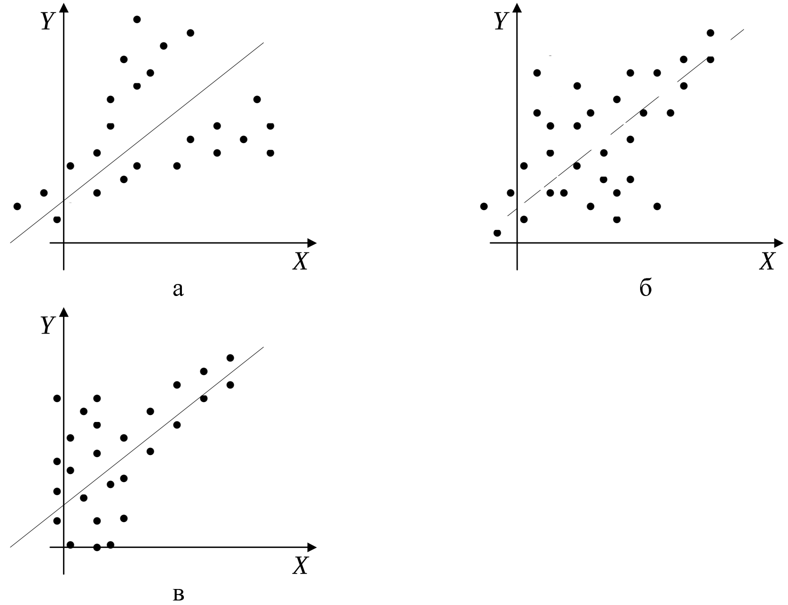


Рис. 4.1. Примеры гетероскедастичности

<u>Гетероскедастичность дисперсии</u> случайного члена означает, что

$$\mathrm{E}\left(u_i u_j\right)=\begin{cases}\sigma_i^2 & \text{при} \quad i=j,\\ 0 & \text{при} \quad i\neq j\end{cases}, \tag{4.8}$$

т.е. нарушается предположение (3.3) в КЛММР, и мы должны рассматривать ОЛММР с нулевой ковариацией случайных ошибок (ср. (4.5) и (4.8)).

<u>Основные последствия гетероскедастичности</u> проявляются в получении неэффективных оценок МНК и занижении стандартных ошибок коэффициен-

тов регрессии, что завышает *t*-статистику и дает неправильное представление о точности уравнения регрессии.

Поэтому для оценивания регрессии с гетероскедастичными случайными ошибками применяется ОМНК.

Предположим, что нам известны значения величин σ_i^2, $i=1,\ldots,n$. Тогда уравнение (4.3) разделим на σ_i:

$$\frac{Y_i}{\sigma_i}=\frac{\beta_0}{\sigma_i}+\beta_1\frac{X_{1i}}{\sigma_i}+\ldots+\beta_k\frac{X_{ki}}{\sigma_i}+\frac{u_i}{\sigma_i},\quad i=1,2,\ldots,n,$$

и получим регрессию с постоянной (гомоскедастичной) дисперсией случайного члена, действительно $\mathrm{V}\left(\frac{u_i}{\sigma_i}\right)=\frac{1}{\sigma_i^2}\mathrm{V}(u_i)=1,\ i=\overline{1,n}$.

Для получения оценок неизвестных дисперсий σ_i^2, $i=1,\ldots,n$ будем предполагать, что они пропорциональны некоторым числам, т.е. $\mathrm{V}(u_i)=\mathrm{E}(u_iu_i)=\frac{\sigma^2}{\lambda_i},\ i=\overline{1,n}$, где σ^2 – некоторая константа.

Принимая различные гипотезы относительно характера гетероскедастичности, будем иметь соответствующие значения λ_i.

Если дисперсия случайного члена пропорциональна квадрату регрессора *X*, так что $\mathrm{E}(u_i^2)=\sigma^2X_i^2,\ i=\overline{1,n}$, то $\lambda_i=\frac{1}{X_i^2}$, $i=1,\ldots,n$.

Если дисперсия случайного члена пропорциональна *X*, так что $\mathrm{E}(u_i^2)=\sigma^2X_i,\ i=\overline{1,n}$, то $\lambda_i=\frac{1}{X_i}$, $i=1,\ldots,n$. Например, для случая одной объясняющей переменной имеем в этом случае систему уравнений ОМНК вида:

$$\begin{cases}\beta_0\sum_i\lambda_i+\beta_1\sum_i\lambda_iX_i=\sum_i\lambda_iY_i;\\ \beta_0\sum_i\lambda_iX_i+\beta_1\sum_i\lambda_iX_i^2=\sum_i\lambda_iX_iY_i.\end{cases}$$

Поскольку значения λ_i, $i=1,\ldots,n$ являются фактически весами, которые устраняют неоднородность дисперсии, то ОМНК для системы с гетероскедастичностью часто называют <u>методом взвешенных наименьших квадратов</u>.

Существуют также и другие методы коррекции модели на гетероскедастичность, в частности состоятельное оценивание стандартных ошибок. Известны способы коррекции стандартных ошибок Уайта и Невье-Веста [5, с. 144-146].

<u>О проверке выборки на гомоскедастичность</u>.

Рассмотрим вопрос тестирования выборки на наличие гомоскедастичности. Возможности такой проверки зависят от природы исходных данных.

Если имеется обширная выборка, то можно воспользоваться стандартным <u>критерием однородности дисперсии Бартлетта</u>.

Расчленяя выборку на *m* независимых групп (каждой из них соответствует единственное значение переменной *X*), вычислим величины:

$$Q_1=n\ln\left(\sum_{i=1}^{m}\frac{n_i}{n}s_i^2\right)-\sum_{i=1}^{m}n_i\ln s_i^2,\ Q_2=1+\frac{1}{3(m-1)}\left(\sum_{i=1}^{m}\frac{1}{n_i}-\frac{1}{n}\right),$$

причем $\sum n_i=n$, здесь n_i - число наблюдений в *i* группе, s_i^2 - дисперсия ошибки в *i* группе. Величина Q_1/Q_2 будет приближенно удовлетворять распределению χ^2 с (*m*-1) степенями свободы. Если вычисленное по выборке значение χ^2 меньше критического, то гипотеза об однородности выборочной дисперсии принимается, в противном случае отклоняется.

В случаях малого количества наблюдений в выборке, когда группировка данных невозможна, используется <u>тест Голдфелда и Куандта</u>. Он предусматривает осуществление следующих шагов:

1. Упорядочить наблюдения по убыванию той независимой переменной, относительно которой есть подозрение на гетероскедастичность.

2. Опустить *v* наблюдений, оказавшихся в центре (*v* должно быть примерно равно четверти общего количества наблюдений *n*).

3. Оценить отдельно обыкновенным методом наименьших квадратов регрессии на первых (*n*-*v*)/2 наблюдениях и на последних (*n*-*v*)/2 наблюдениях при условии, что (*n*-*v*)/2 больше числа оцениваемых параметров *k*.

4. Пусть e_1 и e_2 - суммы квадратов остатков от первой и второй регрессий соответственно. Тогда статистика $Q=e_1/e_2$ будет удовлетворять *F* - распределению с ((*n*-*v*-2*k*)/2; (*n*-*v*-2*k*)/2) степенями свободы. При $Q < F_\alpha$ гипотеза об однородности выборочной дисперсии принимается, в противном случае (с ростом величины *Q*) отклоняется.

Очевидно, что решающим для этого теста является выбор величины *v*. Слишком большое значение *v* уменьшает надежность теста. Экспериментально авторами теста установлено, что для одной объясняющей переменной оптимальное *v*=8 при *n*=30 и *v*=16 при *n*=60.

Кроме перечисленных, могут использоваться тесты на гетероскедастичность Уайта, Бреуша-Пагана и др.

<u>Пример</u>. Проверим по критерию Бартлетта данные из примера 1 раздела 3. Будем иметь табл. 4.1. В табл. 4.1 учтено, что среднее значение e_i равно 0, а значит, $(e_i-\bar{e}_i)^2=e_i^2$. Примем *m*=2. Тогда:

Q_1=20 ·ln(10/20 ·167,41 + 10/20 ·59,69) – (10 ·ln(167,41)+10 · ln(59,69))=2,55; Q_2=1+1/3 ·(1/10+1/10-1/20)=1,05; Q_1/Q_2=2,43.

При одной степени свободы критическое значение χ^2 при 5% уровне значимости равно 3,84, а следовательно, гипотеза об однородности выборочной дисперсии принимается.

Для тех же данных применим тест Гольдфельда и Куандта. В нашем случае число объясняющих переменных k=2, количество исходных данных в выборке n=20. Упорядочим наблюдения по убыванию независимой переменной X_2 – расстояние перевозки, относительно которой есть подозрение на гетероскедастичность. Опустим 4 наблюдения, оказавшихся в центре, т.е. v=4. При значении v=4 получим суммы квадратов остатков от первой и второй регрессий соответственно e_1=1167,38 и e_2=31,49. Статистика $Q=e_1/e_2$=1167,38/31,49 = 37,07 удовлетворяет F-распределению с (6; 6) степенями свободы. $F_{0,05}(6, 6)$ = 4,28, $Q > F$ и гипотеза об однородности выборочной дисперсии должна быть отвергнута.

Поскольку тесты дают противоположные результаты (что не редкость в эконометрике), то лучше согласиться с наихудшим вариантом, т.е. предположить наличие гетероскедастичности и предпринять соответствующие корректирующие меры. В частности, скорректировать стандартные ошибки по формуле Невье-Веста. В таблице 4.2 представлены результаты регрессии до корректировки и после корректировки на гетероскедастичность. Видно, что на величине коэффициентов регрессии корректировка на гетероскедастичность не отражается, а стандартные ошибки и значения статистик были пересчитаны. ∇

Таблица 4.1

Проверка гомоскедастичности дисперсии по критерию Бартлетта

Y	Ошибка e_i	e_i^2	Y	Ошибка e_i	e_i^2
51	-2,49	6,20	26	-0,68	0,46
16	-1,86	3,46	6	5,27	27,72
74	31,93	1019,21	5,8	-5,29	27,93
7,5	-3,18	10,11	13,8	-16,74	280,23
33	-2,17	4,71	6,2	8,94	79,87
26	-18,38	337,64	7,9	-3,57	12,74
11,5	-3,45	11,90	5,4	5,18	26,79
52	5,58	31,14	56	7,72	59,60
15,8	-3,11	9,67	25,5	-0,85	0,72
8	-8,72	76,04	7,1	4,85	23,47
		s_1^2=167,41			s_2^2=59,69

Таблица 4.2

Переменные	Коэффициент		Стандартная ошибка		Значение t-статистики		Значение критерия Фишера $F(2,17)$		R^2	
	до	после	до	после	до	после	до	после	до	после
X_1	1,156	1,156	0,246	0,251	4,694	4,588	24,17	20,87	0,73	0,73
X_2	15,104	15,104	3,352	4,112	4,505	3,673				
Константа	-17,313	-17,313	6,447	5,297	-2,685	-3,268				

4.4. Линейная модель множественной регрессии с автокорреляцией остатков

Вернемся еще раз к предположению (3.3). Из него, в частности, следует, что ковариации случайной ошибки для разных наблюдений равны нулю. Если к тому же случайные ошибки распределены нормально, то это означает их попарную независимость.

Однако регрессионные модели в экономике часто содержат стохастические зависимости между значениями случайных ошибок – автокорреляцию ошибок. Ее причинами являются: во-первых, влияние некоторых случайных факторов или опущенных в уравнении регрессии важных объясняющих переменных, которое не является однократным, а действует в разные периоды времени; во-вторых, случайный член может содержать составляющую, учитывающую ошибку измерения объясняющей переменной.

Применение к модели с автокорреляцией остатков обыкновенного МНК приведет к следующим последствиям:

1. Выборочные дисперсии полученных оценок коэффициентов будут больше по сравнению с дисперсиями по альтернативным методам оценивания, т.е. оценки коэффициентов будут неэффективны.

2. Стандартные ошибки коэффициентов будут оценены неправильно, чаще всего занижены, иногда настолько, что нет возможности воспользоваться для проверки гипотез соответствующими точными критериями – мы будем чаще отвергать гипотезу о незначимости регрессии, чем это следовало бы делать в действительности.

3. Прогнозы по модели получаются неэффективными.

На практике исследователь в этом случае поставлен перед проблемой тестирования наличия в модели автокорреляции, а также выявления причины автокорреляции при ее обнаружении: или в модели опущена существенная переменная, или структура ошибок зависит от времени. То есть, исследование остатков позволяет судить о правильности модели и ее пригодности для прогнозирования.

Простейшим способом проверки наличия автокорреляции является графическое изображение остатков e_i. Возможно построение:

- графика временной последовательности, если остатки получены в разные моменты времени;
- графика зависимости остатков от значений $\hat{Y}_i$, полученных по регрессии;
- графиков зависимости остатков от объясняющих переменных.

Если изображение остатков представляет собой горизонтальную полосу, это указывает на отсутствие каких-либо проблем, связанных с моделью. В противном случае в зависимости от вида и типа графика можно получить информацию о: неадекватности модели, ошибочности расчетов, необходимости включения в модель линейного или квадратичного члена от времени; наконец о непостоянстве дисперсии.

Ясно, что ошибки могут коррелировать по-разному, однако без нарушения общности можно рассматривать так называемую сериальную корреляцию (автокорреляцию), когда зависимость между ошибками, отстоящими на некоторое количество шагов s, называемое порядком корреляции (в частности, на один шаг, s=1), остается одинаковой, что хорошо проявляется визуально на графике в системе координат (e_i; $e_{i\text{-}s}$). Например, для s=1 на рис. 4.2 показаны отрицательная (слева) и положительная (справа) автокорреляция остатков. В экономических исследованиях чаще всего встречается положительная автокорреляция.

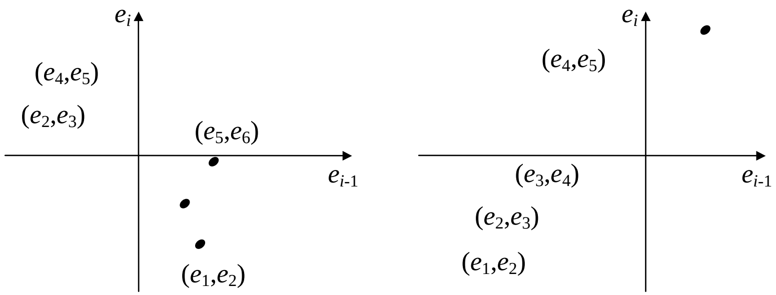


Рис. 4.2. Автокорреляция остатков

Более достоверным способом проверки существования автокорреляции является применение статистических критериев. Хорошо известны два – критерий знаков (относится к непараметрическим критериям) и критерий Дарбина-Уотсона.

Для проведения проверки по критерию знаков необходимо расположить остатки e_i во временной последовательности, выписать их знаки, подсчитать число образующихся при этом серий n_u из одинаковых знаков, а также n_1 – число остатков со знаком плюс и n_2 – число остатков со знаком минус. Далее определяется вероятность $Pr(n_u)$ появления n_u групп при нулевой гипотезе – последовательность остатков полностью случайна (автокорреляция отсутствует). Если $Pr(n_u) < 1-\alpha$, где α – уровень доверия, то нулевая гипотеза отвергается.

Для ускорения расчетов для выборок с n_1, n_2 не больше 20 составлены таблицы с критическими значениями n_u при уровне доверия α=0,05.

Для больших выборок истинное распределение ошибок достаточно точно аппроксимируется нормальным со средним $\mu=2n_1n_2/(n_1+n_2)+1$ и дисперсией $\sigma^2=2n_1n_2(2n_1n_2 - n_1 - n_2)/(n_1 + n_2)^2/(n_1 + n_2 - 1)$, а величина $z=(u - \mu + 0{,}5)/\sigma$ подчиняется нормированному нормальному распределению, следовательно, критические значения n_u могут быть вычислены по формулам $(\mu + z_\alpha\sigma)$ и $(\mu - z_\alpha\sigma)$, где z_α определяется из условия $\Phi_0(z_\alpha)=(1-\alpha)/2$ (значения $\Phi_0(x)=\frac{1}{\sqrt{2\pi}}\int\limits_0^x e^{-u^2/2}du$ даны в справочниках).

Пример. Получены остатки 0,6; 1,9; –1,8; –2,7; –2,9; 1,4; 3,3; 0,3; 0,8; 2,3; –1,4; –1,1, которые обнаруживают следующую последовательность знаков + + – – – + + + + + – –. Имеем n_u=4, n_1=7, n_2=5. По таблице находим критические значения для n_u: 3 и 11. Так как $3 < n_u < 11$, то нулевая гипотеза принимается, то есть остатки независимы и автокорреляция отсутствует. ∇

Критерий знаков достаточно прост и не использует информацию о величине e_i, и поэтому недостаточно эффективен.

Для проверки гипотезы о существовании линейной автокорреляции первого порядка, которая чаще всего имеет место на практике, предпочтителен критерий Дарбина-Уотсона, основанный на статистике:

$$d=\frac{\sum\limits_{i=2}^{n}(e_i-e_{i-1})^2}{\sum\limits_{i=1}^{n}e_i^2} \qquad (4.9)$$

Значения первых разностей ошибки в (4.9) будут обнаруживать тенденцию к уменьшению по абсолютной величине по сравнению с абсолютными значениями e_i при положительной автокорреляции и к увеличению при отрицательной автокорреляции.

Для статистики d имеются верхний d_{U} и нижний d_{L} пределы уровня значимости. Различные статистические решения для нулевой гипотезы H_0: автокорреляция равна нулю, даны в табл. 4.3. При этом появляются области неопределенности, так как величина e_i зависит не только от значений u, но и от значений последовательных X.

Следует отметить, что критерий Дарбина-Уотсона предназначен для моделей с детерминированными (нестохастическими) регрессорами X и не применим, например, в случаях, когда среди объясняющих переменных есть лаговые значения переменной Y.

Таблица 4.3

Области статистических решений для критерия Дарбина-Уотсона

$d<d_L$	$d_L<d<d_U$	$d_U<d<2$; $2<d<(4–d_U)$	$(4–d_U)<d<(4–d_L)$	$d>(4–d_L)$
Отвергаем H_0 в пользу гипотезы о положительной автокорреляции	H_0 не принимается и не отвергается	Принимается H_0	H_0 не принимается и не отвергается	Отвергаем H_0 в пользу гипотезы об отрицательной автокорреляции

Пример. Для примера 1 из п. 3.2 n=20, k=2 имеем табл. 4.4.

Далее по формуле (4.9) d=4397,66/2050,37=2,14.

Значения d_L и d_U при уровне значимости 5% получим из справочника при n=20 и k=2: d_L=1,10, d_U=1,54.

Так как d>2, то вычисляем $4–d_U$=2,46 и $4–d_L$=2,90 и $2<d<4–d_U$.

Согласно табл. 4.3 гипотеза о равенстве нулю автокорреляции принимается. ∇

Какой бы тест на автокорреляцию не использовался, необходимо помнить, что рекомендуется в случаях неопределенности (см. табл. 4.3) принимать гипотезу о наличии автокорреляции, поскольку это гарантирует от отрицательных последствий автокорреляции. В случаях же некорректного принятия гипотезы о равенстве нулю автокорреляции получаем модель, которая не может иметь удовлетворительного применения, хотя формально проходит все проверки.

Таблица 4.4

Вычисление значения статистики d

Ошибка e_i	e_i^2	e_{i-1}	$(e_i-e_{i-1})^2$	Ошибка e_i	e_i^2	e_{i-1}	$(e_i-e_{i-1})^2$
1	2	3	4	5	6	7	8
-2,49	6,20			-0,68	0,46	-8,72	64,64
-1,86	3,46	-2,49	0,40	5,27	27,72	-0,68	35,40
31,93	1019,21	-1,86	1141,76	-5,29	27,93	5,27	111,51
-3,18	10,11	31,93	1232,71	-16,74	280,23	-5,29	131,10
-2,17	4,71	-3,18	1,02	8,94	79,87	-16,74	659,46
-18,38	337,64	-2,17	262,76	-3,57	12,74	8,94	156,50

Продолжение таблицы 4.4

1	2	3	4	5	6	7	8
-3,45	11,90	-18,38	222,90	5,18	26,79	-3,57	76,56
5,58	31,14	-3,45	81,54	7,72	59,60	5,18	6,45
-3,11	9,67	5,58	75,52	-0,85	0,72	7,72	73,44
-8,72	76,04	-3,11	31,47	4,85	23,47	-0,85	32,49
Сумма					2050,37		4397,66

Рассмотрим методы оценивания уравнения регрессии при наличии автокорреляции остатков.

Пусть имеем обобщенную линейную модель множественной регрессии в виде (4.3)-(4.7) с гомоскедастичными остатками $\mathrm{E}(u_i u_i)=\sigma_u^2$.

Предположим, что остатки u_i удовлетворяют следующему уравнению:

$$u_i=\rho u_{i-1}+\varepsilon_i,\ i=2,...,n, \qquad (4.10)$$

представляющему собой авторегрессионную модель первого порядка, для которой выполнено $|\rho|\leq 1$, а ε_i удовлетворяют условиям:

$$\mathbf{E}(\varepsilon_i)=0;\ \mathbf{E}(\varepsilon_i\varepsilon_{i+s})=\begin{cases}\sigma_\varepsilon^2,\ s=0;\\ 0,\ s\neq 0.\end{cases} \qquad (4.11)$$

Тогда несложно показать, что будет выполняться:

$$\mathbf{E}(u_i u_j)=\begin{cases}\sigma_u^2,\ i=j\\ \rho^{|j-i|},\ i\neq j\end{cases}. \qquad (4.12)$$

Условие (4.12) является аналогом (4.5) и фактически означает гомоскедастичность дисперсии случайного члена (первая строчка) и автокорреляцию первого порядка (вторая строчка). Ясно, что если бы было известно значение ρ в (4.10) и затем в (4.12), то можно было бы применить ОМНК (элементы матрицы $\mathbf{\Omega}$ в этом случае вычисляются согласно (4.12)) и получить эффективные оценки коэффициентов регрессии. Однако на практике значение ρ в большинстве случаев не известно, поэтому используются следующие методы оценивания регрессионной модели.

Метод 1. Отказавшись от определения величины ρ, являющейся узким местом модели, статистически, можно положить ρ=0,5; 1 или -1. Однако даже грубая статистическая оценка будет, видимо, более эффективной, поэтому другой способ определения ρ с помощью статистики Дарбина-Уотсона $\rho\approx 1–0,5d$. Применяя затем непосредственно ОМНК, получим оценки коэффициентов.

Метод 2. Если значение ρ в (4.12) задано, то альтернативная схема отыскания оценок коэффициентов модели множественной регрессии суть (в целях упрощения, не нарушая общности, иллюстрация метода дана для случая парной регрессии):

а) Запишем уравнение модели для случая i и i–1:

$$Y_i = \beta_0 + \beta_1 X_i + u_i$$
$$Y_{i-1} = \beta_0 + \beta_1 X_{i-1} + u_{i-1}.$$

Вычтем из обеих частей первого уравнения умноженное на ρ второе уравнение:

$$Y_i - \rho Y_{i-1} = \beta_0(1-\rho) + \beta_1(X_i - \rho X_{i-1}) + u_i - \rho u_{i-1}$$

или переобозначив:

$$\tilde{Y}_i = Y_i - \rho Y_{i-1},\ \tilde{\beta}_0 = \beta_0(1-\rho),\ \tilde{X}_i = X_i - \rho X_{i-1}$$

с учетом (4.10) $\varepsilon_i = u_i - \rho u_{i-1}$, получим модель

$$\tilde{Y}_i = \tilde{\beta}_0 + \beta_1 \tilde{X}_i + \varepsilon_i, \qquad (4.13)$$

для случайного члена которой выполняется условие (4.11), т.е. автокорреляция отсутствует. При указанном преобразовании первое наблюдение умножается на $\sqrt{1-\rho^2}$, т.е. $\tilde{Y}_1 = \sqrt{1-\rho^2}Y_1$, $\tilde{X}_1 = \sqrt{1-\rho^2}X_1$.

б) Применяем обыкновенный МНК к модели (4.13).

В общем случае мы не располагаем информацией о порядке автокорреляции и значениях параметров в авторегрессионном уравнении, а значит, и методы 1 и 2 не дадут искомого результата.

Тем не менее, оценки коэффициентов можно найти приближенно с помощью следующих методов (опять в целях упрощения, не нарушая общности, иллюстрация методов дана для случая парной регрессии).

Метод 3. Итеративная процедура Кохрейна-Оркатта.

а) Оценивается регрессия $Y_i = \beta_0 + \beta_1 X_i + u_i$ с исходными не преобразованными данными с помощью обыкновенного МНК.

б) Вычисляются остатки e_i.

в) Оценивается регрессия $e_i = \rho e_{i-1} + \varepsilon_i$, и коэффициент при e_{i-1} дает оценку ρ.

г) С учетом полученной оценки ρ уравнение $Y_i = \beta_0 + \beta_1 X_i + u_i$ преобразовывается к виду (4.13), оценивание которого позволяет получить пересмотренные оценки коэффициентов β_0 и β_1.

д) Вычисляются остатки регрессии (4.13) и процесс выполняется снова, начиная с этапа в).

Итерации заканчиваются, когда абсолютные разности последовательных значений оценок коэффициентов β_0, β_1 и ρ будут меньше заданного числа (точности).

Подобная процедура оценивания порождает проблемы, касающиеся сходимости итерационного процесса и характера найденного минимума: локальный или глобальный.

Метод 4. Метод Хилдрета-Лу основан на тех же принципах, что и рассмотренный метод 3, но использует другой алгоритм вычислений. Здесь регрессия (4.13) оценивается МНК для каждого значения ρ из диапазона [-1, 1] с некоторым шагом внутри него. Значение, которое дает минимальную стандартную ошибку для преобразованного уравнения (4.13), принимается в качестве оценки ρ, а коэффициенты регрессии определяются при оценивании уравнения (4.13) с использованием этого значения.

Метод 5. Дарбиным была предложена простая схема, дающая эффективные оценки коэффициентов:

а). Подставляя (4.10) в модель $Y_i=\beta_0+\beta_1 X_i+u_i$, получим с учетом $u_{i-1} = Y_{i-1} - \beta_0 - \beta_1 X_{i-1}$:

$$Y_i=\beta_0(1-\rho)+\rho Y_{i-1}+\beta_1(X_i - \rho X_{i-1}) + \varepsilon_i,$$

где ошибка ε_i удовлетворяет (4.11). Применяя обыкновенный МНК к последней модели, получаем оценку ρ как коэффициента при Y_{i-1}.

б). Вычисляем значения преобразованных переменных $\tilde{Y}_i = Y_i - \rho Y_{i-1}$, $\tilde{\beta}_0 = \beta_0(1-\rho)$, $\tilde{X}_i = X_i - \rho X_{i-1}$ и применяем к ним обыкновенный МНК. Получаем искомые оценки коэффициентов регрессии.

Достоинством метода является простота его распространения на случай автокорреляции более высокого порядка.

Как показывают эксперименты, проведенные для малых выборок, лучшим является двухшаговый метод 2, использующий оценку ρ, полученную по методу, предложенному Дарбиным (метод 5 шаг а)).

4.5. Фиктивные переменные. Тест Чоу

Факторы (объясняющие переменные), применяемые в задаче регрессии до сих пор, принимали значения из некоторого непрерывного интервала. Иногда может понадобиться ввести в модель переменные, значения которых детерминированы и дискретны. Например, данные получены для трех разных районов, или на двух фабриках, или на разных машинах и т.п. Переменные такого типа обычно называют фиктивными или искусственными. Эти переменные позволяют отразить в модели эффекты сдвига во времени или в пространстве, воздействия качественных переменных. Пример фиктивной переменной - это переменная X_0 при свободном члене β_0 в уравнении регрессии (3.1), которая принята равной 1. Эту переменную необязательно вводить в модель, но ее использование обеспечивает некоторое удобство в обозначениях. Во многих других случаях введение фиктивных переменных диктуется необходимостью.

Пример. Допустим, мы хотим отразить в модели разное происхождение куриных окорочков (исходные данные[7] - таблица 4.5), часть из которых получены в Америке, а часть в Канаде, при построении регрессионной зависимости веса окорочков Y от возраста кур X. Для этого в модель включим фиктивную переменную Z: Z=0 для Америки, Z=1 для Канады:

$$Y=\beta_0 + \beta_1 X + \alpha Z.$$

Таблица 4.5

Данные для расчета модели с фиктивной переменной

X	28	20	32	22	29	27	28	26	21	27	29
Y	13,3	8,9	15,1	10,4	13,1	12,4	13,2	11,8	11,5	14,2	15,4
Z	1	1	1	1	1	0	0	0	0	1	0

Если бы мы построили регрессию Y на X, то получили бы такое уравнение

$$Y=0{,}442+0{,}465X.$$

Воспользовавшись моделью с фиктивной переменной получим

$$Y=0{,}643+0{,}466X-0{,}422Z$$

или для различных стран:

$Y_{\text{К}}$ =0,221+0,466X для Канады и $Y_{\text{А}}$=0,643+0,466X для Америки.

Экспериментальные данные и три прямые, подобранные методом наименьших квадратов, приведены на рис. 4.3. Все три линии практически параллельны.

Дисперсионный анализ показывает значимость полученных зависимостей, причем уравнение (как с фиктивной переменной, так и без фиктивной переменной) объясняет до 80% вариации относительно среднего.

Вывод, который можно сделать в этом случае – введение фиктивной переменной не дает весомого улучшения модели в смысле дополнительно объясненной вариации. ∇

Ясно, что для какой-либо задачи существует не единственный способ выбора фиктивных переменных, а в большинстве случаев путей их представления много. Это обстоятельство оказывается выгодным, поскольку в некоторых случаях можно угодить в ловушку, когда существует линейная зависимость между введенными фиктивными переменными.

Чтобы избежать ловушки, необходимо выбрать одну из категорий в качестве эталонной и определять фиктивные переменные для остальных возможных

[7] Пример взят из [4]

категорий, причем выбор эталонной категории не влияет на сущность регрессии.

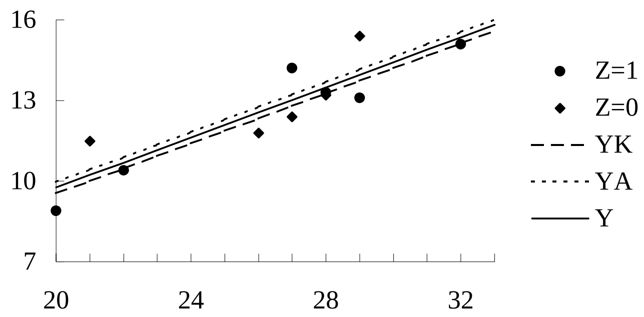


Рис. 4.3

Может потребоваться включение в модель более одной совокупности фиктивных переменных. Это особенно часто встречается при работе с перекрестными выборками. Поясним такую процедуру – множественных совокупностей фиктивных переменных – на примере[8].

Пример. Предположим, что исследуется зависимость между весом новорожденного и семейным положением матери, а также рожала ли она раньше.

Введем фиктивную переменную M, которая принимает значения 1, если мать одинока, и 0 – в остальных случаях.

Введем также фиктивную переменную числа родов в прошлом D, равную 1 для матерей, которые рожали в прошлом, и 0 для матерей, которые ранее не рожали.

При этом двойном наборе фиктивных переменных имеется четыре возможных случая с соответствующими комбинациями значений фиктивных переменных:

1. Замужняя мать, первые роды M=0, D=0.
2. Одинокая мать, первые роды M=1, D=0.
3. Замужняя мать, не первые роды M=0, D=1.
4. Одинокая мать, не первые роды M=1, D=1.

Первый случай по смыслу является основной совместной эталонной категорией. Коэффициент при M будет представлять оценку разности веса новорожденных, если мать одинока (ожидаем отрицательный знак коэффициента). Ко-

[8] Пример из [3].

эффициент при D будет представлять оценку дополнительного веса при рождении, если ребенок не является первенцем. Ребенок для четвертой категории матерей будет подвержен обоим воздействиям. ∇

Фиктивные переменные могут быть введены не только в правую часть регрессионного соотношения, но и зависимая переменная может быть представлена в такой форме. Это возможно в тех случаях, когда в качестве зависимой переменной мы рассматриваем ответы на вопросы, пользуется ли человек собственной машиной, имеет ли счет в банке и т.п., причем во всех случаях зависимая переменная принимает дискретные значения.

Фиктивные переменные могут быть использованы для учета взаимодействия между различными группами факторов.

Пример. Проиллюстрируем сказанное на примере с окорочками. Для построения двух прямых рассмотрим модель:

$$Y=\beta_0+\beta_1X+Z(\gamma_1+\gamma_2X)+u \text{ или } Y=\beta_0+\beta_1X+\gamma_1Z+\gamma_2XZ+u.$$

Такой подход позволяет проверить различные варианты гипотез:

1. Гипотеза H_0: $\gamma_1=\gamma_2=0$ против альтернативы H_1: что это не так. Если гипотеза H_0 будет отвергнута, то мы придем к выводу, что модели не одинаковы, а если нет, то можно пользоваться одной моделью независимо от происхождения окороков.

2. Если гипотеза H_0 в предыдущем пункте будет отвергнута, то можно проверить гипотезу H_0: $\gamma_2=0$. Если H_0 принимается, то мы заключаем, что имеющиеся два набора данных отличаются только уровнем, имея одинаковые углы наклона.

При необходимости могут быть выбраны и другие варианты проверок, если это разумно для задачи. Получим для указанной выше модели уравнение МНК:

$$Y=2{,}974+0{,}377X-3{,}649Z+0{,}123(XZ),$$

причем $R^2=0{,}82$.

Два отдельных уравнения для $Z=1$: $Y=-0{,}675+0{,}5X$;

и для $Z=0$: $Y=2{,}974+0{,}377X$.

Как видно, уравнения несколько отличаются от тех линий, что приведены на рис. 4.3.

Для проверки гипотезы H_0: $\gamma_1=\gamma_2=0$ составим таблицу дисперсионного анализа (табл. 4.6). Значение $F=3{,}399/0{,}983=3{,}458$, что меньше $F_{0,05}(2;7)=4{,}74$, а, следовательно, гипотеза H_0 принимается, то есть можно пользоваться одной моделью как для окороков из Америки, так и из Канады. Последнее подтверждается ранее полученными результатами.

Как показывает пример, использование взаимодействия с фиктивными переменными упрощает построение подходящих критериев и получение правильных статистик для проверки гипотез. ∇

Таблица 4.6

Источник вариации	Сумма квадратов	Степени свободы	Средний квадрат
X	24,447	1	10,414
Z, XZ	6,797	2	3,399
Остаток	6,881	7	0,983
Всего	38,125	10	

Часто эконометрист сталкивается с ситуацией, когда к уже имеющейся выборке он хочет присоединить небольшую дополнительную порцию данных, но не знает, можно ли считать выборки регрессионно однородными.

Если необходимо выяснить, можно ли использовать одну и ту же модель для двух разных выборок данных или следует оценивать отдельные регрессии для каждой выборки, то можно воспользоваться тестом Чоу.

Рассмотрим модели:

$$Y_i=\beta_0'+\beta_1'X_{1i}+\ldots+\beta_k'X_{ki}+u_i',\quad i=1,2,\ldots,n_1 \tag{4.14}$$

$$Y_i=\beta_0''+\beta_1''X_{1i}+\ldots+\beta_k''X_{ki}+u_i'',\quad i=1,2,\ldots,n_2 \tag{4.15}$$

Мы хотим проверить гипотезу

$$H_0:\ \beta_j'=\beta_j'',\ j=\overline{0,k},\ \mathrm{V}u_i'=\mathrm{V}u_i''=\sigma_u^2,$$

которая содержательно означает, что для двух имеющихся выборок из n_1 и n_2 наблюдений можно использовать одну и ту же регрессионную модель, т.е. выборки можно объединить.

Процедура Чоу для статистической проверки гипотезы H_0 суть:

1. Строим МНК оценки регрессии (4.14) и вычисляем сумму квадратов остатков, которую обозначим e_{ur}'. Строим МНК оценки регрессии (4.15) и вычисляем сумму квадратов остатков, которую обозначим e_{ur}''.

2. Строим МНК оценки регрессии по объединенной (общей) выборке, содержащей в себе все наблюдения (числом n_1+n_2) обеих выборок и вычисляем сумму квадратов остатков, которую обозначим e_r.

3. Критическая статистика F вычисляется по формуле:

$$F=\frac{(e_r-e_{ur}'-e_{ur}'')/(k+1)}{(e_{ur}'+e_{ur}'')/(n_1+n_2-2k-2)}$$

и имеет распределение Фишера с $(k+1)$ и (n_1+n_2-2k-2) степенями свободы. Если $F>F_\alpha$, то нулевая гипотеза отвергается, и в этом случае мы не можем объединить две выборки в одну.

5. Временные ряды

5.1.Специфика временных рядов

Часто исследователь имеет дело с данными в виде временных рядов.

Совокупность наблюдений $y(t_1), y(t_2),..., y(t_n)$ анализируемой величины $Y(t)$, произведенных в последовательные моменты времени $t_1, t_2,..., t_n$, называется временным рядом.

Иначе говоря, временной ряд – это упорядоченная во времени последовательность наблюдений.

Среди временных рядов выделяют одномерные, полученные в результате наблюдения одной, фиксированной характеристики исследуемого объекта, и, многомерные временные ряды как результат наблюдений нескольких характеристик одного исследуемого объекта в течение ряда моментов времени.

По времени наблюдения временные ряды делятся на дискретные и непрерывные. Дискретные ряды, в свою очередь, разделяются на ряды с равноотстоящими и произвольными моментами наблюдения.

Временные ряды бывают детерминированными и случайными: первые получены как значения некоторой неслучайной функции, а вторые - как реализации случайной величины.

Стохастические временные ряды подразделяются на стационарные и нестационарные. Ряд $y(t)$ называется *стационарным* (в узком смысле), если среднее, дисперсия и ковариации $y(t)$ не зависят от t.

В дальнейшем, если не оговорено иначе, будем рассматривать одномерные, дискретные с равноотстоящими моментами наблюдений случайные временные ряды.

Природа временных рядов существенно отличается от природы пространственных данных, что проявляется в весьма специфических свойствах временных рядов. В своей работе исследователь должен учитывать эти особенности, основные из которых отображены в таблице 5.1.

Таблица 5.1

Особенности временных рядов

Характеристики наблюдений	Тип данных	
	Пространственные данные	Временные ряды
Порядок	Не существенен	Существенен
Статистическая независимость	Независимы	Не являются статистически независимыми
Функция распределения	Распределены одинаково	Распределены неодинаково
Количество	Как правило, большое	Как правило, небольшое

Наличие автокорреляции	Встречается нечасто	Встречается часто

Значения элементов временного ряда формируются под воздействием ряда факторов, среди которых выделяют:

- долговременные, формирующие в длительной перспективе общую тенденцию анализируемого признака. Эта тенденция описывается с помощью некоторой функции, называемой трендом (T);
- сезонные, формирующие периодически повторяемые в определенное время года колебания анализируемого признака (S);
- циклические, формирующие изменения анализируемого в результате воздействия циклов экономической, демографической или астрофизической природы (C);
- случайные, не поддающиеся учету и регистрации, как результат воздействия случайных, внешних факторов (U).

Первые три составляющие часто объединяют в одну детерминированную и рассматривают модель ряда в виде $y_t=f(t)+u_t$, $\forall t$. Изменение уровня $f(t)$ со временем называют при этом трендом.

Предметом анализа временного ряда является выделение и изучение указанных компонент ряда, как правило в рамках одной из моделей ряда: либо аддитивной $Y=T+C+S+U$, либо мультипликативной $Y=T \cdot C \cdot S \cdot U$.

Некоторые составляющие могут отсутствовать в тех или иных рядах.

В результате анализа временного ряда необходимо определить, какие из неслучайных составляющих присутствуют в разложении ряда, построить для них хорошие оценки, подобрать модель, описывающую поведение остатков и оценить ее параметры.

5.2. Проверка гипотезы о существовании тренда

Для выявления факта наличия или отсутствия неслучайной составляющей $f(t)$, то есть для проверки гипотезы о существовании тренда - H_0: $\mathbf{E}y(t)=a=\text{const}$, используют следующие критерии.

I. *Критерий серий*. Упорядочим члены ряда по возрастанию: $y_1, y_2, ..., y_t, ..., y_n$. Определим медиану ряда:

$$y_{med} = \begin{cases} y_{\frac{n+1}{2}}, \text{ если } n \text{ нечетно,} \\ \frac{1}{2}\left(y_{\frac{n}{2}} + y_{\frac{n}{2}+1}\right), \text{ если } n \text{ четно.} \end{cases}$$

Образуем последовательность плюсов и минусов, соответствующую исходному ряду, по правилу: если $y_t>y_{med}$, то y_t соответствует плюс, если $y_t<y_{med}$,

то – минус. Под серией понимается последовательность подряд идущих плюсов и подряд идущих минусов. Подсчитаем общее число серий ν и протяженность самой длинной серии τ.

Если хотя бы одно из неравенств:

$$\nu > \left[\frac{1}{2}\left(n+2-1{,}96\sqrt{n-1}\right)\right],$$

$$\tau < \left[1{,}43\ln(n+1)\right]$$

окажется нарушенным, то гипотеза H_0 отвергается с вероятностью ошибки α, заключенной между 0,05 и 0,0975.

II. *Критерий "восходящих" и "нисходящих" серий.* Аналогично предыдущему критерию исследуется последовательность плюсов и минусов. Правило построения последовательности: если y_{t+1}-y_t>0, то y_t соответствует плюс, если y_{t+1}-y_t<0, то – минус (если подряд идут несколько равных наблюдений, то во внимание принимается одно из них).

Если хотя бы одно из неравенств:

$$\nu > \left[\frac{1}{3}(2n-1)-1{,}96\sqrt{\frac{16n-29}{90}}\right],$$

$$\tau < \tau_0,$$

окажется нарушенным, то гипотеза H_0 отвергается с вероятностью ошибки α, заключенной между 0,05 и 0,0975. Величина τ_0 определяется в зависимости от n:

n	$n \le 26$	$26 < n \le 153$	$153 < n \le 1170$
τ_0	$\tau_0=5$	$\tau_0=6$	$\tau_0=7$

III. Критерий квадратов последовательных разностей (критерий Аббе). Если есть основания полагать, что разброс наблюдений y_t относительно своих средних значений подчиняется нормальному закону распределения вероятностей, то применяется критерий Аббе - см. [1], с. 801-802.

5.3. Аналитическое выравнивание временных рядов, оценка параметров уравнения тренда

Метод обработки временных рядов, целями которого является устранение случайных колебаний и построение аналитической функции, характеризующей зависимость уровней ряда от времени – тренда, называется аналитическим выравниванием временного ряда.

Суть метода аналитического выравнивания состоит в том, чтобы заменить фактические уровни временного ряда $y(t_1), y(t_2),..., y(t_n)$ на теоретические

$\hat{y}(t_1), \hat{y}(t_2),..., \hat{y}(t_n)$. Расчет $\hat{y}(t_1), \hat{y}(t_2),..., \hat{y}(t_n)$ осуществляется по некоторому формализованному уравнению, принятому за математическую модель тренда. Для построения трендов чаще всего применяют такие функции, как:

- линейная: $\hat{y}_t = a + b \cdot t$;
- степенная: $\hat{y}_t = a \cdot t^b$;
- гиперболическая: $\hat{y}_t = a + \frac{b}{t}$;
- экспоненциальная: $\hat{y}_t = e^{a+b\cdot t}$;
- полиномы второго и более высоких порядков: $\hat{y}_t = a_0 + a_1 \cdot t + a_2 \cdot t^2 + ... + a_p \cdot t^p$.

Расчет параметров тренда производится методом МНК. В качестве зависимой переменной выступают фактические уровни ряда $y(t_1), y(t_2),..., y(t_n)$, а независимой переменной является время $t = 1,2,...,n$. Заметим, что для нелинейных трендов необходима процедура линеаризации, аналогичная рассмотренной в разделе 3.

Выбор функции тренда может быть осуществлен несколькими способами. Наиболее простым считается тот, в ходе которого анализируют цепные абсолютные приросты (первые разности уровней ряда) Δ_t, абсолютные ускорения уровней ряда (вторые разности ряда) Δ_Δ и цепные коэффициенты роста K_t.

Если примерно одинаковы Δ_t, то ряд имеет линейный тренд, если же примерно постоянны Δ_Δ, то для описания тенденции временного ряда следует выбрать параболу второго порядка, и, если примерно равны K_t, необходимо использовать экспоненциальную или степенную функции.

Пример 1.[9] Рассчитаем параметры уравнения тренда по следующим данным:

Таблица 5.2

Темпы роста номинальной месячной заработной платы (за 10 месяцев 1999г., % к уровню декабря 1998г.)

Месяц	Темп роста номинальной заработной платы	Месяц	Темп роста номинальной заработной платы
Январь	82,9	Июнь	121,6
Февраль	87,3	Июль	118,6
Март	99,4	Август	114,1
Апрель	104,8	Сентябрь	123,0
Май	107,2	Октябрь	127,3

[9] См. [7], с. 235-238.

Для выявления тенденции временного ряда рассчитаем цепные абсолютные приросты (первые разности уровней ряда) Δ_t, абсолютные ускорения уровней ряда (вторые разности ряда) Δ_Δ и цепные коэффициенты роста K_t.

Таблица 5.3

Месяц	t	y_t	$\Delta_t = y_t - y_{t-1}$	$\Delta_\Delta = \Delta_t - \Delta_{t-1}$	$K_t = \frac{y_t}{y_{t-1}}$
Январь	1	82,9	-	-	-
Февраль	2	87,3	4,4	-	1,053
Март	3	99,4	12,1	7,7	1,139
Апрель	4	104,8	5,4	-6,7	1,054
Май	5	107,2	2,4	-3,0	1,023
Июнь	6	121,6	14,4	12,0	1,134
Июль	7	118,6	-3,0	-17,4	0,975
Август	8	114,1	-4,5	-1,5	0,962
Сентябрь	9	123,0	8,9	13,4	1,078
Октябрь	10	127,3	3,7	-5,2	1,035

Наибольшей стабильностью отличаются цепные коэффициенты роста. Для описания тенденции временного ряда используем степенной или экспоненциальный тренд. Для того чтобы убедиться в этом, рассчитаем уравнение тренда и коэффициенты детерминации уравнения для наиболее часто применяемых функций, применяя МНК. Получим табл. 5.4. Коэффициенты детерминации рассчитаны по линеаризованным уравнениям регрессии.

Как мы и предполагали, степенной тренд лучше всего описывает тенденцию анализируемого временного ряда, что подтверждается высоким значением коэффициента детерминации. ∇

Таблица 5.4

Уравнения трендов

Тип тренда	Уравнение	$R^2_{скорр}$
Линейный	$\hat{y}_t = 82{,}66 + 4{,}72 \cdot t$	0,873
Парабола второго порядка	$\hat{y}_t = 72{,}9 + 9{,}599 \cdot t - 0{,}444 \cdot t^2$	0,920
Степенной	$\ln \hat{y}_t = 4{,}39 + 0{,}193 \cdot \ln t$	0,931
Экспоненциальный	$\ln \hat{y}_t = 4{,}43 + 0{,}045 \cdot t$	0,856
Гиперболический	$\hat{y}_t = 122{,}57 - 47{,}63 / t$	0,728

Интерпретация параметров тренда существенно зависит от его типа.

Если тренд имеет линейную форму, то a - начальный уровень временного ряда в период времени t=0 и b - средний за период абсолютный прирост уровней ряда.

Если же ряд имеет, например, экспоненциальный тренд, то a - начальный уровень временного ряда в период времени t=0 и e^b - средний за единицу времени коэффициент роста уровней ряда.

Трактовка параметров степенного тренда аналогична трактовке параметров экспоненциального тренда.

Пример (продолжение примера 1). Согласно уравнению линейного тренда $\hat{y}_t = 82{,}66 + 4{,}72 \cdot t$ темпы роста заработной платы за 10 месяцев 1999 г. изменялись от начального уровня 82,66% со средним за месяц абсолютным приростом в 4,72 процентных пункта.

Мы можем заменить фактические уровни временного ряда $y(t_1), y(t_2),..., y(t_n)$ на теоретические $\hat{y}(t_1), \hat{y}(t_2),..., \hat{y}(t_n)$, подставляя значения t в уравнение тренда:

$$\hat{y}_1^{лин} = 82{,}66 + 4{,}72 \cdot 1 = 87{,}38;$$

$$\hat{y}_2^{лин} = 82{,}66 + 4{,}72 \cdot 2 = 92{,}10, \quad ...$$

Уравнение экспоненциального тренда в исходной форме имеет вид:

$$\hat{y}_t = e^{4{,}43} \cdot e^{0{,}045 \cdot t}, \Rightarrow$$

$$\hat{y}_t = 83{,}96 \cdot 1{,}046^t.$$

Таким образом, начальный уровень ряда в начальный период времени равен 83,96, а средний цепной коэффициент роста - 1,045. Следовательно, темпы роста заработной платы за 10 месяцев 1999 г. изменялись от начального уровня 83,96% со средним за месяц цепным коэффициентом роста в 104,5%. Теоретические значения временного ряда рассчитываются как:

$$\hat{y}_1^{эксп} = \hat{y}_0^{эксп} \cdot 1{,}045 = 83{,}96 \cdot 1{,}045 = 87{,}74;$$

$$\hat{y}_2^{эксп} = \hat{y}_1^{эксп} \cdot 1{,}045 = 87{,}74 \cdot 1{,}045 = 91{,}82, \quad ...$$

Уравнение тренда параболы второго порядка имеет вид:

$$\hat{y}_t = 72{,}9 + 9{,}599 \cdot t - 0{,}444 \cdot t^2.$$

Следовательно, темпы роста заработной платы за 10 месяцев 1999 г. изменялись от начального уровня 72,9% со среднемесячным абсолютным приростом, описываемым зависимостью вида $(9{,}599 \cdot t - 0{,}444 \cdot t^2)$. Теоретические значения уровней ряда могут быть рассчитаны как:

$$\hat{y}_1^{параб} = 72{,}9 + 9{,}599 \cdot 1 - 0{,}444 \cdot 1 = 82{,}055;$$

$$\hat{y}_2^{параб} = 72{,}9 + 9{,}599 \cdot 2 - 0{,}444 \cdot 4 = 90{,}322, \quad ...$$

∇

5.4. Метод последовательных разностей

Часто при аналитическом выравнивании ряда используется модель тренда в виде полинома.

Для определения порядка аппроксимирующего полинома в этом случае выделения тренда широко используется метод последовательных разностей членов анализируемого временного ряда.

Метод основан на следующем математическом факте: если временной ряд $y_1, y_2, ..., y_t, ..., y_n$ содержит в качестве своей неслучайной составляющей алгебраический полином $f(t)=a_0+a_1t+...+a_pt^p$ порядка p, то переход к последовательным разностям $y(1), y(2), \dots, y(n)$, повторенный $p+1$ раз (то есть переход к последовательным разностям порядка $p+1$), исключает неслучайную составляющую (включая константу a_0), оставляя элементы, выражающиеся только через остаточную случайную компоненту $u(t)$.

Алгоритм метода. Последовательно для $k=1,2,\dots$ вычисляем разности $\Delta^k y(t)$ ($t=1,2,\dots, n-k$). Анализируем поведение разностей в зависимости от их порядка k. Начиная с некоторого k разности стабилизируются, оставаясь приблизительно на одном уровне при дальнейшем росте k. Это значение k и будет давать порядок сглаживающего полинома, то есть p.

При применении метода следует иметь в виду, что стабилизация разностей не доказывает, что ряд первоначально состоял из полинома плюс случайный остаток, а только то, что он может быть приближенно представлен таким образом.

Пример. Имеются данные о базисных темпах роста среднедушевого дохода населения области за 10 месяцев (в % к январю). Расчет первых и вторых разностей показывает, что для ряда y_t тренд может быть адекватно описан полиномом второй степени. ∇

Таблица 5.5

Расчет последовательных разностей

Месяц	Темпы роста среднедушевого дохода (%), y_t	$\Delta y_t=y_t - y_{t-1}$	$\Delta^2 y_t=\Delta y_t - \Delta y_{t-1}$
Февраль	102	-	-
Март	103	1	-
Апрель	107	4	3
Май	114	7	3
Июнь	125	11	4
Июль	139	14	3
Август	157	18	4
Сентябрь	178	21	3
Октябрь	201	23	2
Ноябрь	227	26	3

5.5. Аддитивная и мультипликативная модели временного ряда

Простейшим подходом к моделированию временных рядов, содержащих сезонные колебания, является построение аддитивной или мультипликативной моделей временного ряда.

Выбор одной из этих моделей основывается на анализе структуры временного ряда.

Если амплитуда сезонных колебаний примерно постоянна, то строят аддитивную модель. Если же амплитуда колебаний непостоянна, то есть возрастает или уменьшается, то строят мультипликативную модель.

Процесс построения модели ряда в этом случае включает следующие этапы:

1. Выравнивание исходного ряда методом скользящей средней. Расчет значений сезонной компоненты S.
2. Устранение сезонной компоненты из исходных уровней ряда и получение выравненных данных $(T+U)$ в аддитивной или $(T \cdot U)$ в мультипликативной модели.
3. Аналитическое выравнивание уровней $(T+U)$ или $(T \cdot U)$ и расчет значений T с использованием полученного уравнения тренда.
4. Расчет полученных по модели значений $(T+S)$ или $(T \cdot S)$
5. Расчет абсолютных и/или относительных ошибок.

Рассмотрим процесс построения аддитивной модели на примере.

Пример. Имеются данные о количестве продукции (тыс.шт.), проданной фирмой «Вега» в течение последних 20 кварталов.

Квартал	Объем продаж	Квартал	Объем продаж	Квартал	Объем продаж	Квартал	Объем продаж
1	8,4	6	9,1	11	10,1	16	12,2
2	8,6	7	9,2	12	10,8	17	11,9
3	8,8	8	9,9	13	10,5	18	12,3
4	9,5	9	9,7	14	10,7	19	12,5
5	8,5	10	9,9	15	11	20	13,2

Этап 1. Проведем выравнивание ряда методом скользящей средней. Для этого просуммируем уровни ряда по 4 кварталам последовательно. Далее разделим полученные суммы на 4 и найдем скользящие средние, уже не содержащие сезонной компоненты. Найдем центрированные скользящие средние, для чего вычислим средние значения из двух последовательных скользящих средних. Вычислим оценки сезонной компоненты как разность между фактическим уровнем продаж и центрированными скользящими средними.

Таблица 5.6

Расчет оценок сезонной компоненты

Квартал	Объем продаж, тыс.шт.	Итого за 4 квартала	Скользящая средняя за 4 квартала	Центрированная скользящая средняя	Оценка сезонной компоненты
1	2	3	4	5	6
1	8,4				
2	8,6				
		35,3	8,825		
3	8,8			8,8375	-0,0375
		35,4	8,85		
4	9,5			8,9125	0,5875
		35,9	8,975		
5	8,5			9,025	-0,525
		36,3	9,075		
6	9,1			9,125	-0,025
		36,7	9,175		
7	9,2			9,325	-0,125
		37,9	9,475		
8	9,9			9,575	0,325
		38,7	9,675		
9	9,7			9,7875	-0,0875
		39,6	9,9		
10	9,9			10,0125	-0,1125
		40,5	10,125		
11	10,1			10,225	-0,125
		41,3	10,325		
12	10,8			10,425	0,375
		42,1	10,525		
13	10,5			10,6375	-0,1375
		43	10,75		
14	10,7			10,925	-0,225
		44,4	11,1		
15	11			11,275	-0,275
		45,8	11,45		
16	12,2			11,65	0,55
		47,4	11,85		
17	11,9			12,0375	-0,1375
		48,9	12,225		

Квартал	Объем продаж, тыс.шт.	Итого за 4 квартала	Скользящая средняя за 4 квартала	Центрированная скользящая средняя	Оценка сезонной компоненты
18	12,3			12,35	-0,05
		49,9	12,475		
19	12,5				
20	13,2				

Используем полученные оценки сезонной компоненты для расчета сезонности S. Для этого найдем средние квартальные оценки сезонной компоненты, использовав данные всех кварталов. Заметим, что сумма значений сезонной компоненты по всем кварталам должна быть равна нулю, поэтому значения сезонной компоненты корректируются на величину, полученную как частное от деления суммы оценок сезонных компонент на число сезонов.

Таблица 5.7

Корректировка значений сезонной компоненты

Показатели	Год	Квартал			
		1	2	3	4
	1	-	-	-0,0375	0,5875
	2	-0,525	-0,025	-0,125	0,325
	3	-0,0875	-0,1125	-0,125	0,375
	4	-0,1375	-0,225	-0,275	0,55
	5	-0,1375	-0,05	-	-
Итого за квартал		-0,8875	-0,4125	-0,5625	1,8375
Средняя оценка сезонной компоненты для квартала		-0,2218	-0,1031	-0,1406	0,4593
Скорректированная оценка сезонной компоненты		-0,2203	-0,1015	-0,1390	0,4609

Рассчитаем корректирующий коэффициент:

$$k=[(-0,22188)+(-0,10313)+(-0,14063)+ 0,459375]/4=-0,00625/4= -0,00156.$$

Скорректированные оценки сезонной компоненты определяются путем вычитания из средней оценки сезонной компоненты для квартала корректирующего коэффициента. Полученные таким образом значения занесены в таблицу 5.7.

Этап 2. Устраним сезонную компоненту из исходных уровней ряда и получим выравненные данные $T+U=y_i-S$ (столбец 4).

Таблица 5.8

Расчет выравненных значений T и ошибок E в аддитивной модели

t	y_i	S_i	$T+U=y_i-S$	T	$T+S$	$U=y_i-(T+S)$	U^2
1	2	3	4	5	6	7	8
1	8,4	-0,2203	8,6203	8,1545	7,9341	0,6861	0,4707
2	8,6	-0,1015	8,7015	8,3845	8,2829	0,4185	0,1751

t	y_i	S_i	$T+U=y_i-S$	T	$T+S$	$U=y_i-(T+S)$	U^2
3	8,8	-0,1390	8,9390	8,6146	8,4755	0,4635	0,2148
4	9,5	0,46093	9,0390	8,8446	9,3056	-0,2666	0,0710
5	8,5	-0,2203	8,7203	9,0747	8,8544	-0,1344	0,0179
6	9,1	-0,1015	9,2015	9,3047	9,2032	-0,0016	0,0000
7	9,2	-0,1390	9,3390	9,5348	9,3957	-0,0566	0,0032
8	9,9	0,46093	9,4390	9,7648	10,2258	-0,7867	0,6189
9	9,7	-0,2203	9,9203	9,9949	9,7746	0,1457	0,0212
10	9,9	-0,1015	10,0010	10,2249	10,1234	-0,1218	0,0148
11	10,1	-0,1390	10,2390	10,4550	10,3159	-0,0769	0,0059
12	10,8	0,46093	10,3390	10,6850	11,1460	-0,8069	0,6511
13	10,5	-0,2203	10,7203	10,9151	10,6948	0,0254	0,0006
14	10,7	-0,1015	10,8015	11,1451	11,0436	-0,2420	0,0585
15	11	-0,1390	11,1390	11,3752	11,2361	-0,0971	0,0094
16	12,2	0,46093	11,7390	11,6052	12,06622	-0,3271	0,1070
17	11,9	-0,2203	12,1203	11,8353	11,6150	0,5052	0,2553
18	12,3	-0,1015	12,4015	12,0653	11,9638	0,4377	0,1916
19	12,5	-0,1390	12,6390	12,2954	12,1563	0,4826	0,2329
20	13,2	0,46093	12,7390	12,5254	12,9864	-0,2473	0,0611

Этап 3. Определим компоненту T. Для этого проведем аналитическое выравнивание ряда ($T+U$) с помощью линейного тренда. Имеем линейный тренд вида:

$$T = 7{,}9244+0{,}2301t.$$

Стандартная ошибка коэффициента регрессии 0,293. R^2=0,95.

Подставляя в уравнение тренда последовательно t= 1,…,20, получим значения тренда для каждого уровня временного ряда (столбец 5, табл. 5.8).

Этап 4. Найдем значения уровней ряда, полученные по аддитивной модели как ($T+S$) (столбец 6, табл. 5.8).

Этап 5. Рассчитаем абсолютную ошибку как $U=y_i-(T+S)$, (столбец 7, табл. 5.8). Качество полученной модели можно проверить, используя сумму квадратов абсолютных ошибок (столбец 8). Сумма квадратов абсолютных ошибок равна 3,18. По отношению к сумме квадратов отклонений исходных уровней ряда от его среднего уровня, равной 40,32, эта величина составит 7,89%.

Следовательно, аддитивная модель объясняет 92,11% общей вариации объема продаж за 20 кварталов. ∇

Рассмотрим построение мультипликативной модели на примере.

Пример. Имеются поквартальные данные об объеме экспорта одной из областей РФ за 5 лет (млн. долл.).

Таблица 5.9

Квартал	Объем экспорта, млн.долл.	Квартал	Объем экспорта, млн.долл.	Квартал	Объем экспорта, млн.долл.	Квартал	Объем экспорта, млн.долл.
1	19,3	6	15,8	11	20,3	16	25,4
2	12,3	7	17,2	12	22,3	17	31,8
3	13,2	8	19,9	13	29,7	18	23,9
4	15,6	9	26,3	14	21,1	19	25,8
5	21,5	10	19,1	15	23,7	20	27,4

Этап 1. Проведем выравнивание ряда методом скользящей средней. Для этого просуммируем уровни ряда по 4 кварталам последовательно. Далее разделим полученные суммы на 4 и найдем скользящие средние, уже не содержащие сезонной компоненты. Найдем центрированные скользящие средние, для чего вычислим средние значения из двух последовательных скользящих средних. Вычислим оценки сезонной компоненты как частное от деления фактического уровня экспорта на центрированные скользящие средние.

Таблица 5.10

Расчет оценок сезонной компоненты

Квартал	Объем продаж, тыс.шт.	Итого за 4 квартала	Скользящая средняя за 4 квартала	Центрированная скользящая средняя	Оценка сезонной компоненты
1	2	3	4	5	6
1	19,3				
2	12,3				
		60,4	15,1		
3	13,2			15,375	0,858537
		62,6	15,65		
4	15,6			16,0875	0,969697
		66,1	16,525		
5	21,5			17,025	1,262849
		70,1	17,525		
6	15,8			18,0625	0,87474
		74,4	18,6		
7	17,2			19,2	0,895833
		79,2	19,8		
8	19,9			20,2125	0,984539
		82,5	20,625		
9	26,3			21,0125	1,251636
		85,6	21,4		
10	19,1			21,7	0,880184
		88	22		
11	20,3			22,425	0,90524
		91,4	22,85		
12	22,3			23,1	0,965368
		93,4	23,35		
13	29,7			23,775	1,249211
		96,8	24,2		
14	21,1			24,5875	0,85816
		99,9	24,975		
15	23,7			25,2375	0,939079
		102	25,5		

Квартал	Объем продаж, тыс.шт.	Итого за 4 квартала	Скользящая средняя за 4 квартала	Центрированная скользящая средняя	Оценка сезонной компоненты
16	25,4			25,85	0,982592
		104,8	26,2		
17	31,8			26,4625	1,201701
		106,9	26,725		
18	23,9			26,975	0,886006
		108,9	27,225		
19	25,8				
20	27,4				

Используем полученные оценки сезонности для расчета сезонной компоненты S. Для этого найдем средние квартальные оценки сезонной компоненты, используя данные всех кварталов.

Таблица 5.11

Расчет значений сезонной компоненты

Показатели	Год	Квартал			
		1	2	3	4
	1	-	-	0,8585	0,9696
	2	1,2628	0,8747	0,8958	0,9845
	3	1,2516	0,8801	0,9052	0,9653
	4	1,2492	0,8581	0,9390	0,9825
	5	1,2017	0,8860	-	-
Итого за квартал		4,9653	3,4990	3,5986	3,9021
Средняя оценка сезонной компоненты для квартала		1,2413	0,8747	0,8996	0,9755
Скорректированная оценка сезонной компоненты		1,2440	0,876	0,9016	0,9776

Заметим, что сумма значений сезонной компоненты по всем кварталам должна быть равна числу периодов в цикле. В нашем примере, цикл – год, в котором соответственно 4 квартала. Поэтому окончательный вариант сезонной компоненты будет получен корректировкой, заключающейся в умножении средней оценки сезонной компоненты для квартала на коэффициент k:

$$k=4/(1,2413+0,8747+0,8996+0,9755)=4/3,9913=1,0021.$$

Полученные таким образом значения были занесены в табл. 5.11 (строка 3).

Этап 2. Устраним сезонную компоненту из исходных уровней ряда и получим выравненные данные $T \cdot U=y_i/S$ (столбец 4, табл. 5.12).

Таблица 5.12

Расчет выравненных значений T и ошибок U в мультипликативной модели

t	y_i	S	$T \cdot U=y_i/S$	T	$T \cdot U$	$U=y_i-(T \cdot S)$	U^2
1	2	3	4	5	6	7	8

t	y_i	S	$T \cdot U=y_i/S$	T	$T \cdot U$	$U=y_i-(T \cdot S)$	U^2
1	19,3	1,2440	15,5139	14,2959	17,7847	0,8723	0,7609
2	12,3	0,8766	14,0303	15,0690	13,2105	1,0620	1,1279
3	13,2	0,9016	14,6402	15,8421	14,2836	1,0249	1,0505
4	15,6	0,9776	15,9563	16,6151	16,2440	0,9822	0,9648
5	21,5	1,2440	17,2823	17,3882	21,6317	0,7989	0,6383
6	15,8	0,8766	18,0227	18,1613	15,9214	1,1319	1,2813
7	17,2	0,9016	19,0767	18,9344	17,0717	1,1174	1,2486
8	19,9	0,9776	20,3546	19,7074	19,2673	1,0564	1,1160
9	26,3	1,2440	21,1407	20,4805	25,4786	0,8297	0,6884
10	19,1	0,8766	21,7869	21,2536	18,6324	1,1693	1,3672
11	20,3	0,9016	22,5149	22,0266	19,8597	1,1336	1,2852
12	22,3	0,9776	22,8094	22,7997	22,2905	1,0232	1,0471
13	29,7	1,2440	23,8738	23,5728	29,3255	0,8140	0,6627
14	21,1	0,8766	24,0683	24,3459	21,3433	1,1276	1,2716
15	23,7	0,9016	26,2859	25,1189	22,6478	1,1606	1,3470
16	25,4	0,9776	25,9802	25,8920	25,3137	1,0263	1,0533
17	31,8	1,2440	25,5618	26,6651	33,1725	0,7705	0,5937
18	23,9	0,8766	27,2622	27,4381	24,0542	1,1333	1,2845
19	25,8	0,9016	28,6150	28,2112	25,4359	1,1249	1,2655
20	27,4	0,9776	28,0259	28,9843	28,3369	0,9890	0,9781

Этап 3. Определим компоненту T. Для этого проведем аналитическое выравнивание ряда $(T \cdot E)$ с помощью линейного тренда. Имеем линейный тренд вида:

$$T = 13,5229+0,7730t.$$

Стандартная ошибка коэффициента регрессии 0,735. R^2=0,97.

Подставляя в уравнение тренда последовательно t= 1,…,20, получим значения тренда для каждого уровня временного ряда (столбец 5, табл. 5.12).

Этап 4. Найдем значения уровней ряда, полученные по мультипликативной модели как $(T \cdot S)$ (столбец 6, табл. 5.12).

Этап 5. Рассчитаем абсолютную ошибку как $U=y_i-(T \cdot S)$, (столбец 7, табл. 5.12). Качество полученной модели можно проверить, используя сумму квадратов абсолютных ошибок (столбец 8). Общая сумма квадратов абсолютных ошибок равна 21,033. По отношению к сумме квадратов отклонений исходных уровней ряда от его среднего уровня, равной 530,072, эта величина составит 3,9681%:

$$(21,03378/530,072) \cdot 100=3,97 \%.$$

Следовательно, мультипликативная модель объясняет 96,03% общей вариации экспорта. ∇

5.6. Модели стационарных и нестационарных временных рядов и их идентификация

<u>Модели авторегрессии порядка p (AutoRegressive - $AR(p)$ models).</u>

Достаточно часто экономические показатели, представленные в виде временного ряда, имеют сложную структуру. Моделирование таких рядов путем построения модели тренда, сезонности и периодической составляющей не приводит к удовлетворительным результатам. Ряд остатков часто имеет статистические закономерности. Наиболее распространенными моделями стационарных рядов являются модели авторегрессии и модели скользящего среднего.

Будем рассматривать класс стационарных временных рядов. Задача состоит в построении модели остатков временного ряда u_t и прогнозирования его значений.

Авторегрессионная модель предназначена для описания стационарных временных рядов. Стационарный процесс удовлетворяет уравнению авторегрессии бесконечного порядка с достаточно быстро убывающими коэффициентами. В частности поэтому авторегрессионная модель достаточно высокого порядка может хорошо аппроксимировать почти любой стационарный процесс. В связи с этим модель авторегрессии часто применяется для моделирования остатков в той или иной параметрической модели, например регрессионной модели или модели тренда.

Модель авторегрессии порядка 1 $AR(1)$ *(марковский процесс).*

Марковскими называются процессы, в которых состояние объекта в каждый следующий момент времени определяется только состоянием в настоящий момент и не зависит от того, каким путем объект достиг этого состояния. В терминах корреляционного анализа для временных рядов марковский процесс можно описать следующим образом: существует статистически значимая корреляционная связь исходного ряда с рядом, сдвинутым на один временной интервал, и отсутствует с рядами, сдвинутыми на два, три и т. д. временных интервала. В идеальном случае эти коэффициенты корреляции равны нулю.

Авторегрессионная модель первого порядка определяется соотношением:

$$u(t)=\mu\, u(t-1)+\varepsilon(t)\,, \qquad (5.1)$$

где μ - числовой коэффициент $|\mu|<1$, $\varepsilon(t)$ – последовательность случайных величин, образующих «белый шум» ($E(\varepsilon(t))=0$, $E(\varepsilon(t)\,\varepsilon(t+\tau))=\begin{cases}\sigma^2, \tau=0\\ 0, \ \tau\neq 0\end{cases}$).

Модель (5.1) называется также марковским процессом.

Имеем:

$$\mathbf{E}(u(t))\equiv 0. \qquad (5.2)$$

$$r(u(t)u(t\pm\tau))=\mu^{\tau}. \qquad (5.3)$$

$$\mathbf{D}u(t)=\sigma^2/(1-\mu^2). \qquad (5.4)$$

$$\mathrm{cov}(u(t)u(t\pm\tau))=\mu^{\tau}\mathbf{D}u(t). \qquad (5.5)$$

Из (5.3) следует, что при $|\mu|$ близком к единице дисперсия $u(t)$ будет намного больше дисперсии ε_t. Это значит (учитывая (5.2) $\mu=r(u(t)u(t\pm 1))=r(1)$, т.е. параметр μ может быть интерпретирован как значение автокорреляции первого порядка), что в случае сильной корреляции соседних значений ряда $u(t)$ ряд слабых возмущений ε_t будет порождать размашистые колебания остатков $u(t)$.

Условие стационарности ряда (5.1) определяется требованием $|\mu|<1$.

Автокорреляционная функция (АКФ) $r(\tau)$ марковского процесса определяется соотношением (5.3).

Частная автокорреляционная функция

$$r_{\text{част}}(\tau)=r(u(t)u(t+\tau)) \;/\; u(t+1)=u(t+2)=\ldots=u(t+\tau-1)=0$$

может быть вычислена по формуле: $r_{\text{част}}(2)=(r(2)-r^2(1))/(1-r^2(1))$. Для второго и выше порядков (см. [1], с. 413, 414) должно быть $r_{\text{част}}(\tau)=0\ \forall\tau=2,3,\ldots$. Это удобно использовать для подбора модели (5.1): если вычисленные по оцененным невязкам $u(t)=y_t-\hat{f}_t$ выборочные частные корреляции статистически незначимо отличаются от нуля при $\tau=2,3,\ldots$, то использование модели $AR(1)$ для описания случайных остатков не противоречит исходным данным.

Идентификация модели. Требуется статистически оценить параметры μ и σ^2 модели (5.1) по имеющимся значениям исходного ряда y_t.

Выделяем неслучайную составляющую $\hat{f}_t$ и получаем невязки $\hat{u}_t=y_t-\hat{f}_t$. Находим дисперсию невязок $\hat{\gamma}=\frac{1}{n}\sum_{t=1}^{n}(\hat{u}_t-\bar{u})^2$, где $\bar{u}=\frac{1}{n}\sum_{t=1}^{n}\hat{u}_t$ (для большинства методов выделения $\hat{f}_t$ автоматически $\bar{u}=0$). Далее с учетом (5.2), (5.3) получим формулы для оценки параметров модели (5.1):

$$\hat{\mu}=\frac{\frac{1}{n-1}\sum_{t=1}^{n-1}(\hat{u}_t-\bar{u})(\hat{u}_{t+1}-\bar{u})}{\hat{\gamma}},$$

$$\hat{\sigma}^2=(1-\hat{\mu}^2)\hat{\gamma}.$$

Модели авторегрессии p порядка – $AR(p)$ при $p\geq 2$ см. в [1], с. 834-837:

$$u(t)=\mu_1 u(t-1)+\mu_2 u(t-2)+\ldots+\varepsilon(t). \qquad (5.6)$$

<u>Пример</u>. График первой разности ряда, хорошо описывающейся моделью $AR(1)$, представлен на рис. 5.1; график выборочной автокорреляционной функции (АКФ) первой разности этого ряда представлен на рис. 5.2. ∇

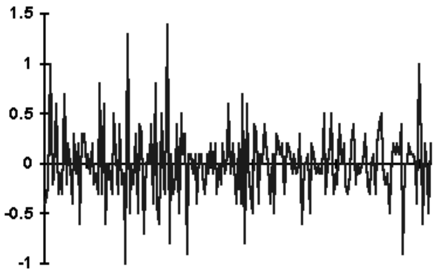


Рис. 5.1

Рис. 5.2

Модели скользящего среднего порядка q (Moving Average - $MA(q)$ models).

Часто на показатель в текущий момент времени оказывает воздействие значение показателя в предыдущие моменты. Хотя воздействие отдаленных элементов незначительно, в сумме оно может оказывать существенное влияние на модель. Учесть это воздействие возможно в модели скользящего среднего. Моделирование воздействия всех предшествующих элементов ряда на показатель в текущий момент основано на предпосылке о том, что в ошибках модели за несколько предшествующих периодов сосредоточена информация о всей предыстории ряда.

Моделью скользящего среднего порядка q называется процесс:

$$u(t)=\varepsilon(t)-\theta_1\,\varepsilon(t-1)-\theta_2\,\varepsilon(t-2)-\ldots-\theta_q\,\varepsilon(t-q). \tag{5.7}$$

В частности, модели порядка 1 и 2 соответственно имеют вид:

$$u(t)=\varepsilon(t)-\theta\,\varepsilon(t-1), \tag{5.8}$$

$$u(t)=\varepsilon(t)-\theta_1\,\varepsilon(t-1)-\theta_2\,\varepsilon(t-2). \tag{5.9}$$

Переход от формы (5.6) к форме (5.7) осуществляется с помощью последовательной подстановки в правую часть формулы (5.6) вместо $u(t-1)$, $u(t-2)$, … их выражений, вычисленных по формуле (5.6) для моментов времени t-1, t-2, …. Это означает двойственность в представлении анализируемого временного ряда – две эквивалентные формы линейного процесса - и обратимость *AR* и *MA* моделей.

В качестве примера рассмотрим *модель скользящего среднего первого порядка* – $MA(1)$. Данная модель описывается соотношением (5.8). Можно показать, что стационарность $u(t)$ обеспечивается при любом значении параметра θ. Модель обратима (представима в виде модели авторегрессии бесконечного порядка) при условии $|\theta|<1$.

Автокорреляционная функция:

$$r(\tau)=\begin{cases}\dfrac{-\theta}{1+\theta^2}, & \tau=1,\\[2mm] 0, & \tau\ge 2.\end{cases}$$

Частная корреляционная функция процесса $MA(1)$, определяющая степень тесноты корреляционной связи между $u(t)$ и $u(t\pm\tau)$, $\tau=1,2,\ldots$ при фиксированных значениях всех промежуточных элементов этого ряда задается выражением:

$$r_{частн}(\tau)=-\theta^{\tau}\,\frac{1-\theta^2}{1-\theta^{2(\tau+1)}}.$$

Идентификация модели $MA(1)$. Требуется статистически оценить параметры θ и σ^2 модели (5.8) по имеющимся значениям исходного ряда y_t. Выделяем неслучайную составляющую $\hat f_t$ и получаем невязки $\hat u_t=y_t-\hat f_t$. Находим оценку автокорреляции $\hat r(1)$:

$$\hat r(1)=\frac{1}{n-1}\sum_{t=1}^{n-1}(\hat u_t-\bar u)(\hat u_{t+1}-\bar u)\Big/\frac{1}{n}\sum_{t=1}^{n}(\hat u_t-\bar u)^2.$$

Подставляя $\hat r(1)$ в выражение для автокорреляционной функции, имеем квадратное уравнение для θ:

$$\theta^2+(1/\hat r(1))\,\theta+1=0.$$

Из двух решений приведенного квадратного уравнения ($\theta_1\,\theta_2=1$) одно будет меньше единицы – его и выбираем в качестве искомой оценки параметра в модели $MA(1)$.

Оценка σ^2 получается по формуле: $\hat\sigma^2=\dfrac{\frac{1}{n}\sum_{t=1}^{n}(\hat u_t-\bar u)^2}{1+\hat\theta^2}$.

Модель скользящего среднего второго порядка – $MA(2)$ отличается более сложным построением - см. [1], с. 843-845.

Важное практическое значение имеют процессы, первая (или более высокая) разность которых стационарна и является процессом $MA(q)$. Подобные процессы устроены как случайные колебания с непостоянным средним уровнем, или (для второй разности) непостоянным углом наклона.

Модели авторегрессии-скользящего среднего (AutoRegressive - Moving Average - $ARMA(p, q)$ models).

На практике для экономичной параметризации анализируемого процесса иногда бывает необходимо включить в модель как члены, описывающие авторегрессию, так и члены, моделирующие остаток в виде скользящего среднего. Такой линейный процесс имеет вид:

$$u(t)=\mu_1 u(t\text{-}1)+\ldots+\mu_p u(t\text{-}p)+\varepsilon(t)\text{-}\ \theta_1\,\varepsilon(t\text{-}1)\text{-}\ldots\text{-}\ \theta_q\,\varepsilon(t\text{-}q) \qquad (5.10)$$

и называется процессом авторегрессии - скользящего среднего порядка (p, q) – *ARMA*(p, q).

Рассмотрим в качестве примера модель *ARMA*(1, 1). В соответствии с моделью (5.10) процесс *ARMA*(1, 1) описывается формулой:

$$u(t)=\mu u(t\text{-}1)+\varepsilon(t)\text{-}\ \theta\varepsilon(t\text{-}1) \text{ или } u(t)\text{-}\mu u(t\text{-}1)=\varepsilon(t)\text{-}\ \theta\,\varepsilon(t\text{-}1).$$

Процесс *ARMA*(1, 1) стационарен, если корень характеристического уравнения *AR*(1) модели $1\text{-}\mu z=0$ по модулю больше единицы. То есть должно быть $|\mu|<1$. Обратимость процесса *ARMA*(1, 1) обеспечивается требованием, чтобы корень характеристического уравнения *MA*(1) модели $1\text{-}\theta z=0$ по модулю был больше единицы. То есть должно быть $|\theta|<1$. АКФ:

$$r(\tau)=\begin{cases}\dfrac{(1-\mu\theta)(\mu-\theta)}{1+\theta^2-2\mu\theta}, \tau=1,\\ \mu r(\tau-1)+\mu^{\tau-1}r(1), \tau\ge 2.\end{cases}$$

Автокорреляционная функция экспоненциально убывает от начального значения $r(1)$, причем это убывание монотонно, если μ положительно, и колебательно (знакопеременно), если μ отрицательно.

Из последнего равенства и условий стационарности и обратимости следует, что $r(1)$ и $r(2)$ должны удовлетворять условиям:

$$\begin{cases}|r(2)|<|r(1)|,\\ r(2)>r(1)(2r(1)+1),\ r(1)<0,\\ r(2)>r(1)(2r(1)-1),\ r(1)>0.\end{cases}$$

Эти условия бывают полезными при проверке гипотезы (по выборочным значениям коэффициентов автокорреляции) о том, что анализируемый процесс может быть описан *ARMA*(1, 1) моделью.

Идентификация модели *ARMA*(1, 1). Требуется статистически оценить параметры μ, θ и σ^2 модели по имеющимся значениям исходного ряда y_t.

Этап 1. $\hat{\mu}=\dfrac{\hat{r}(2)}{\hat{r}(1)}$.

Этап 2. Из уравнения модели несложно получить систему уравнений вида:

$$\begin{cases}\hat{\gamma}\left(1+\hat{\mu}^2\right)-2\hat{\mu}\hat{\gamma}(1)=\sigma^2\left(1+\theta^2\right),\\ \hat{\gamma}(1)\left(1+\hat{\mu}^2\right)-\hat{\mu}\left(\hat{\gamma}+\hat{\gamma}(2)\right)=-\theta\sigma^2.\end{cases}$$

Поделив первое уравнение системы на второе, получим квадратное уравнение относительно θ:

$$A=-(1+\theta^2)/\ \theta, \text{ где } A=\frac{\hat{\gamma}\left(1+\hat{\mu}^2\right)-2\hat{\mu}\hat{\gamma}(1)}{\hat{\gamma}(1)\left(1+\hat{\mu}^2\right)-\hat{\mu}\left(\hat{\gamma}+\hat{\gamma}(2)\right)}.$$

Из двух корней уравнения выбираем тот, который удовлетворяет условию обратимости $|\theta|<1$. Оценку σ^2 определяем из любого уравнения системы.

<u>Модель авторегрессии - проинтегрированного скользящего среднего (AutoRegressive Integrated Moving Average - *ARIMA*(p, q, k) models).</u>

Модель впервые была предложена Дж.Боксом и Г.Дженкинсом и поэтому известна как модель Бокса-Дженкинса. Это одна из наиболее популярных моделей для построения краткосрочных прогнозов значений временных рядов.

Будем рассматривать нестационарные, однородные временные ряды. То есть ряды, для которых случайный остаток $u(t)$, получающийся после вычитания из ряда $y(t)$ его неслучайной составляющей $f(t)$, представляет нестационарный временной ряд. Модель Бокса-Дженкинса предназначена для описания нестационарных временных рядов со следующими свойствами:

а) в рамках аддитивной модели $y(t)$ включает $f(t)$, имеющий вид алгебраического полинома от t степени k-1, причем коэффициенты полинома могут быть как стохастические, так и нестохастические,

б) ряд $y_k(t)$, t=1,2,…, n-k, получившийся из $y(t)$ после применения к нему метода последовательных разностей, может быть описан моделью *ARMA*(p, q).

Следовательно, модель Бокса-Дженкинса имеет вид:

$$y_k(t)=\mu_1 y_k(t\text{-}1)+\ldots+\mu_p y_k(t\text{-}p)+\varepsilon(t)\text{-}\ \theta_1\,\varepsilon(t\text{-}1)\text{-}\ldots\text{-}\ \theta_q\,\varepsilon(t\text{-}q), \qquad (5.11)$$

где $y_k(t)=\Delta^k y(t)=y(t)\text{-}C_k^1 y(t\text{-}1)+C_k^2 y(t\text{-}2)\text{-}\ldots+(-1)^k y(t\text{-}k)$, $t=k+1, k+2, \ldots, n$. Здесь Δ^k – k-я последовательная разность анализируемого процесса $y(t)$ ($\Delta=y(t)\text{-}y(t\text{-}1)$, $\Delta^2=\Delta y(t)\text{-}\Delta y(t\text{-}1)$ и т.п.).

Введем операторы сдвига во времени:

$$F_+y_t=y_{t+1} \text{ и } F_-y_t=y_{t-1}.$$

Причем $F_+F_-=1$, $F^k_-y_t=y_{t-k}$, $F^k_+y_t=y_{t+k}$, $\Delta=1-F_-$.

Тогда оператор авторегрессии порядка p *AR*(p) имеет вид:

$$A_p(F_-,\ \mu)=1\text{-}\mu_1F_-\text{-}\mu_2F^2_-\text{-}\ \ldots\ \text{-}\mu_pF^p_-,$$

а оператор скользящего среднего порядка q *MA*(q):

$$B_q(F_-,\ \theta)=1\text{-}\ \theta_1F_-\text{-}\ \theta_2F^2_-\text{-}\ \ldots\ \text{-}\ \theta_qF^q_-.$$

Модель *ARIMA*(p, q, k) будет с учетом формулы (5.11) и введенных операторов иметь вид:

$$A_p(F_-,\ \mu)\Delta^k y(t)=B_q(F_-,\ \theta)\varepsilon(t). \qquad (5.11^a)$$

На практике применяются модели *ARIMA*(p, q, k), в которых p, q, k не превышают 2. Например, *ARIMA*(1, 1, 1):

$$A_1(F_-,\ \mu)\Delta y(t) = B_1(F_-,\ \theta)\varepsilon(t) \Rightarrow (1-\mu F_-)(y_t - y_{t-1}) = (1-\theta F_-)\ \varepsilon_t \Rightarrow$$
$$y_t - y_{t-1} - \mu y_{t-1} + \mu y_{t-2} = \varepsilon_t - \theta\ \varepsilon_{t-1} \Rightarrow$$
$$y(t) = (1+\mu)y(t-1) - \mu y(t-2) + \varepsilon(t) - \theta\,\varepsilon(t-1).$$

Частным случаем модели *ARIMA* является модель авторегрессии *AR*(p), для которой $q=k=0$. Другой частный случай - модель скользящего среднего *MA*(q), для которой $p=k=0$.

Важные специальные классы моделей - модели *ARIMA*(0, q, k), и модели *ARMA*(p, q) = *ARIMA*(p, q, 0).

Модель *AR*(1) при положительном коэффициенте автокорреляции представляет собой колебательный процесс с преобладанием длинных волн. Если коэффициент корреляции отрицателен, процесс является сильно осциллирующим. Модель *ARIMA*(0, 1, 1) описывает случайный процесс с непостоянным уровнем. Аналогичное утверждение справедливо для модели *ARIMA*(0, 2, 2), описывающей случайный процесс с переменным уровнем и углом наклона.

Идентификация *ARIMA* моделей.

Структура модели ARIMA описывается тремя параметрами (p, q, k). Кроме того, разные по форме модели могут быть довольно близки друг другу. Поэтому весьма важно по возможности правильно определить структуру модели. Рассмотрим этапы идентификации.

1. Подбирается порядок модели k. Для этого используется либо метод последовательных разностей, либо анализ автокорреляционных функций процессов $\Delta y(t)$, $\Delta^2 y(t)$, … - пока не достигнем быстрого затухания (стационарности) автокорреляционной функции для некоторого k. Дж.Бокс и Г.Дженкинс предлагают взять за визуальный критерий стационарности быстрое убывание значений выборочной АКФ. Использование завышенного порядка разности приводит к росту дисперсии ошибок и к заметному росту дисперсии прогноза.

2. Находим $y_k(t)=\Delta^k y(t)$ и идентифицируем *ARMA*(p, q) модель.

<u>Пример</u>. Для определения порядков авторегрессии и скользящего среднего продемонстрируем вид и свойства теоретических АКФ и частной АКФ простейших моделей.

Пример АКФ и частной АКФ для модели *AR*(1) представлен на рис. 5.3; 5.4. Пример АКФ и частной АКФ для модели *AR*(2) содержится на рис. 5.5; 5.6. Из содержания рис. 5.3-5.6 следует, что все значения частной АКФ для лагов, больших порядка авторегрессии, статистически незначимы. Пример АКФ и частной АКФ для модели *MA*(1) изображен на рис. 5.7; 5.8.

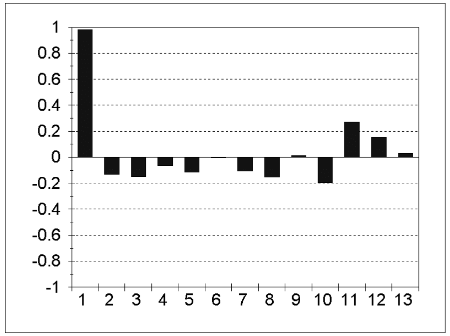


Рис. 5.3

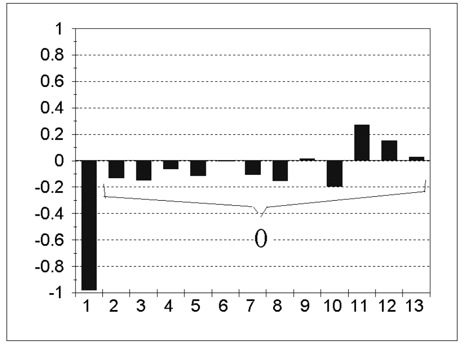


Рис. 5.4

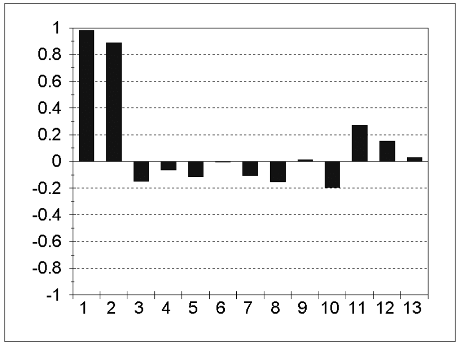


Рис. 5.5

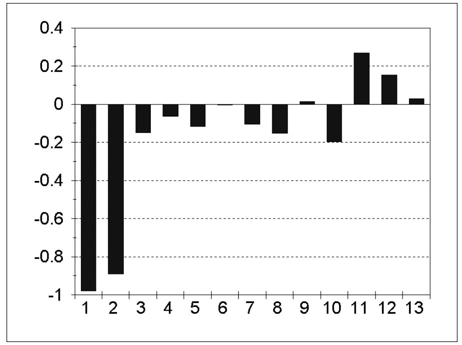


Рис. 5.6

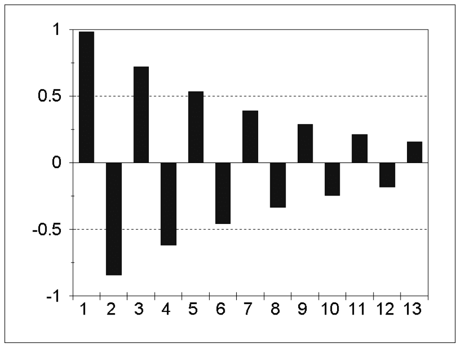


Рис. 5.7.

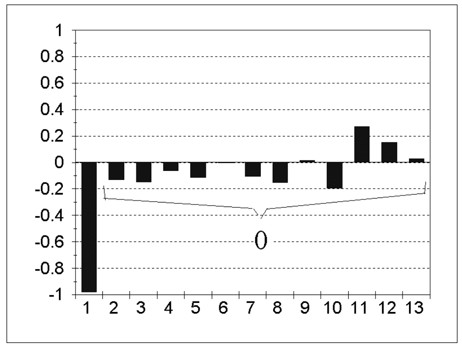


Рис. 5.8.

Пример АКФ и частной АКФ для модели *MA*(2) представлен на рис. 5.9; 5.10. Для модели *MA*(q) все значения АКФ для лагов, больших q, равны нулю. Для модели *ARMA*(p, q) значения АКФ после лага p-q представляют собой смесь затухающих синусоид и экспонент, а значения частной АКФ ведут себя аналогично после лага *q-p*. ∇

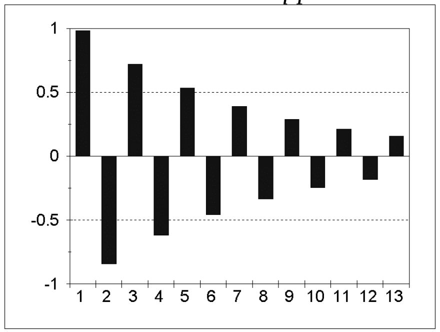


Рис. 5.9

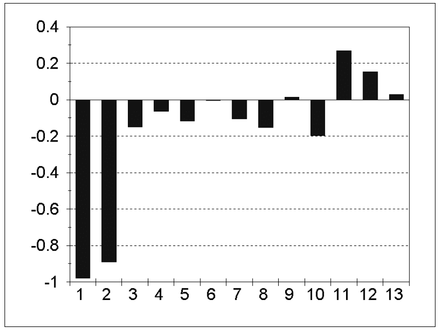


Рис. 5.10

Общий подход Бокса-Дженкинса к анализу временных рядов показан на рис. 5.11. Схема процесса выбора модели временного ряда показана на рис. 5.12.

Если процесс выбора модели успешно осуществлен, возникает проблема оценки качества построенной модели. Для «хорошей» модели остатки должны быть «белым шумом», т.е. их выборочные автокорреляции не должны значимо отклоняться от нуля. Кроме того, модель не должна содержать лишних параметров, т.е. нельзя уменьшить число параметров без появления значимой автокорреляции остатков. Для диагностики модели необходимо попытаться модифицировать ее, меняя порядки авторегрессии и скользящего среднего. Одновременно повышать оба порядка не рекомендуется ввиду опасности вырождения модели.

5.7. Тестирование стационарности временного ряда

Как было отмечено выше, стационарные временные ряды имеют следующие отличительные черты: значения ряда колеблются вокруг постоянного среднего значения с постоянной дисперсией, которая не зависит от времени, АКФ затухает с увеличением лага. При анализе экономических явлений чаще приходится иметь дело с нестационарными временными рядами, которые не имеют постоянного среднего, дисперсия которых зависит от времени, а АКФ затухает очень медленно. Для подбора модели ряда и прогнозирования его значений необходимо уметь распознавать тип временного ряда.

Рассмотрим процесс авторегрессии первого порядка

$$y(t)=\mu y(t-1)+\varepsilon(t).$$

Ряд $y(t)$ является стационарным рядом, если $-1<\mu<1$. Если $\mu=1$, то $y(t)$ – нестационарный временной ряд – случайное блуждание со сдвигом: в этом слу-

чае считают, что временной ряд $y(t)$ имеет *единичный корень*.

Вычтем $y(t-1)$ из обеих частей модели: $\Delta y(t)=\gamma y(t-1)+\varepsilon(t)$, где $\gamma=\mu-1$.

Дики и Фуллер рассмотрели три регрессии:

$$\Delta y(t)=\gamma y(t-1)+\varepsilon(t),$$

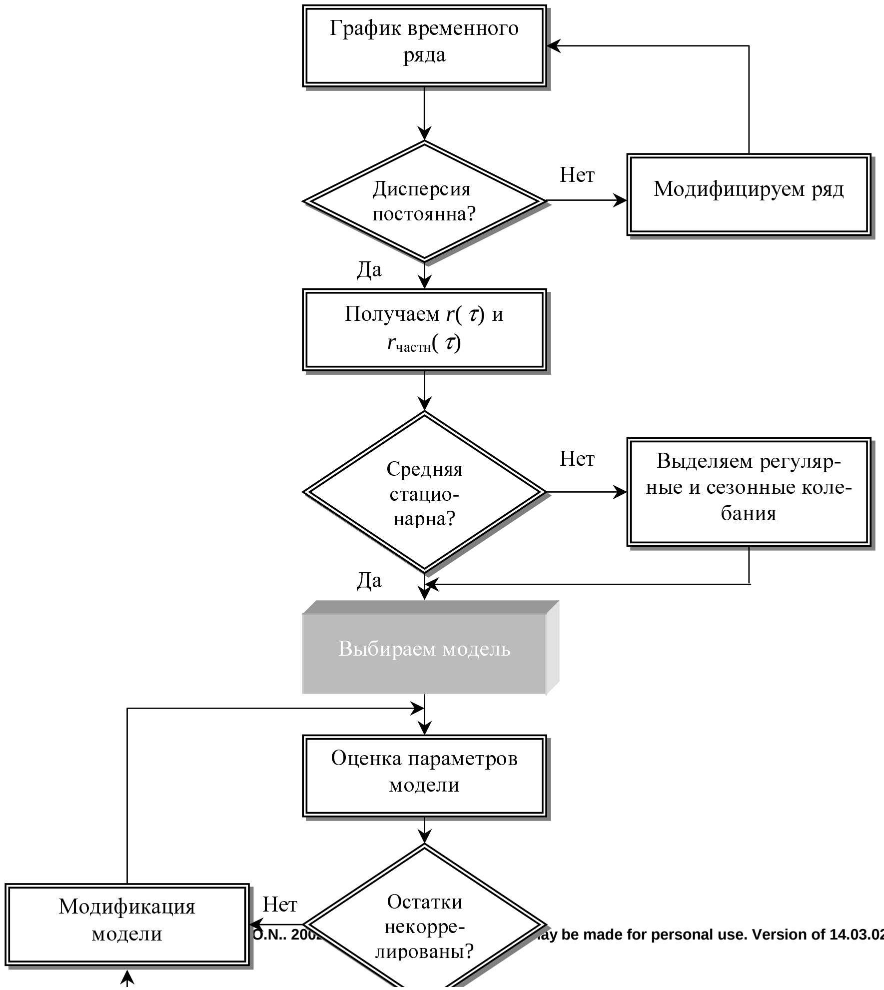
$$\Delta y(t)=\mu_0+\gamma y(t-1)+\varepsilon(t),$$

$$\Delta y(t)=\mu_0+\gamma y(t-1)+\mu_2 t+\varepsilon(t).$$

Вторая регрессия содержит постоянный элемент μ_0, а третья, кроме этого, и линейный временной тренд. Во всех трех регрессиях интересующий параметр γ.

Нулевая гипотеза H_0: $\gamma=0$ против альтернативы H_1: $\gamma<0$.

Тест Дики-Фуллера (Dickey-Fuller) состоит в следующем. Оцениваются методом наименьших квадратов одно из указанных выше уравнений.



Прогнозирование

Рис. 5.11. Подход Бокса-Дженкинса

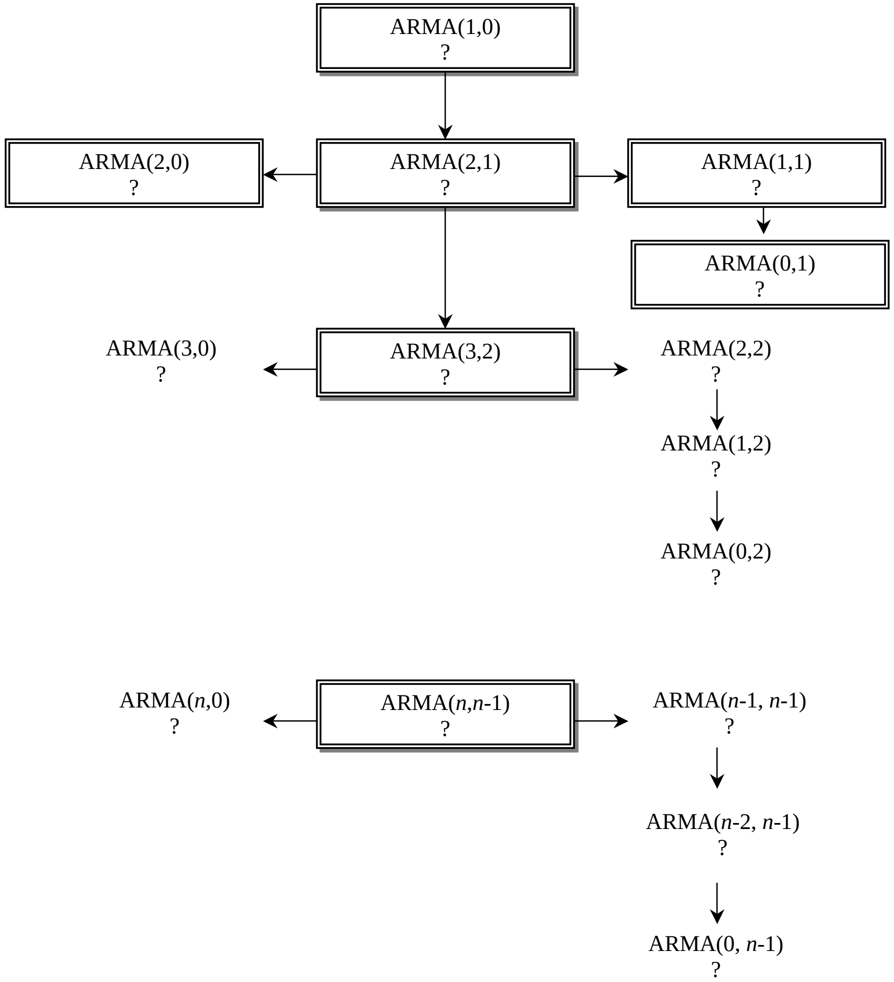


Рис. 5.12. Процесс выбора *ARIMA* модели

Получают оценку γ, стандартную ошибку и соответствующее значение t – статистики. Сравнивая значение t-статистики с табличным, определяют, принять или отклонить H_0. Критическое значение t-статистики имеет нестандартное распределение и зависит от формы регрессии и объема выборки – см в [5].

Критические значения не изменятся, если указанные выше модели заменить авторегрессионным процессом произвольного порядка:

$$\Delta y(t)=\gamma y(t\text{-}1)+\sum_{i=2}^{p}\mu_i \Delta y_{t-i+1}+\varepsilon(t),$$

$$\Delta y(t)=\mu_0+\gamma y(t\text{-}1)+\sum_{i=2}^{p}\mu_i \Delta y_{t-i+1}+\varepsilon(t),$$

$$\Delta y(t)=\mu_0+\gamma y(t\text{-}1)+\mu_2 t+\sum_{i=2}^{p}\mu_i \Delta y_{t-i+1}+\varepsilon(t).$$

Для последних моделей Дики и Фуллер предложили три дополнительные статистики для тестирования обобщенных гипотез о коэффициентах:

ϕ_1: H_0: $\gamma=\mu_0=0$.

ϕ_2: H_0: $\gamma=\mu_0=\mu_2=0$.

ϕ_3: H_0: $\gamma=\mu_2=0$.

Статистики ϕ_i конструируются как F тест: $\phi_i=\dfrac{(RSS_r-RSS_{ur})/g}{RSS_{ur}/(n-k)}$, i=1,2,3,

где RSS_r и RSS_{ur} – квадраты ошибок короткой и длинной регрессий, g – число исключенных переменных, n – число наблюдений, k – число параметров в длинной регрессии. Большие значения ϕ_i ведут к отклонению нулевой гипотезы. Критические значения статистик вычислены Дики и Фуллером и затабулированы.

5.8. Эконометрический анализ взаимосвязанных временных рядов

Коинтеграция и мнимая регрессия.

Рассмотрим два временных ряда y_t и x_t. Предположим, что оба ряда имеют единичные корни, то есть являются нестационарными. Предположим далее, что исследователь не знает механизмов, порождающих y_t и x_t, и оценивает регрессию:

$$y_t=\beta x_t+\varepsilon_t,\ t=1,\dots,n. \tag{5.12}$$

Если $\varepsilon_t=y_t-\beta x_t$, t=1,…,n является стационарным временным рядом, то временные ряды y_t и x_t называются коинтегрированными, а вектор $(1\ -\beta)$ называется коинтегрирующим вектором.

Примеры.

1. Длинная ставка процента R, короткая ставка процента r: $\varepsilon_t=R_t-r_t$, вектор коинтеграции (1 –1).

2. Логарифм потребления C_t, логарифм дохода y_t: $\varepsilon_t=C_t-y_t$, вектор коинтеграции (1 –1).

3. Логарифм обменного курса D_t, логарифм внутренней цены P_t, логарифм цен мирового рынка P_t^*: $\varepsilon_t=D_t-P_t+P_t^*$, вектор коинтеграции (1 –1 1). ∇

В случае коинтегрируемости временных рядов говорят о долгосрочном динамическом равновесии. Если y_t и x_t коинтегрированы, то y_t и βx_t содержат

общую нестационарную компоненту – долговременную тенденцию, а разность $y_t-\beta x_t$ стационарна и совершает флуктуации около нуля.

Таким образом, коинтеграция временных рядов – причинно-следственная зависимость в уровнях временных рядов, которая выражается в совпадении или противоположной направленности их тенденций и случайной колеблемости.

Возможен случай, когда ошибка $\varepsilon_t=y_t-\beta x_t$, t=1,…,n в регрессии (5.12) является нестационарным временным рядом. Тогда условия классической регрессионной модели (п. 3) не выполняются, в частности дисперсия ε_t не является постоянной. Кроме того, МНК оценка параметра β не состоятельна, поэтому с ростом объема выборки увеличиваются шансы получения ложных выводов о взаимосвязи y_t и x_t. Такая ситуация называется ложной (мнимой) регрессией. На практике признаками мнимой регрессии являются высокое значение R^2 и малое значение статистики Дарбина-Уотсона.

Для проверки рядов на коинтеграцию используются тесты Энгеля-Гранжера или Йохансена.

Пример. Рассмотрим временные ряды логарифмов доходов и расходов на потребление с августа 1990 г. по январь 1992 г. в России. Графический анализ – рис. 5.1 показывает, что тенденции этих рядов совпадают.

Расчет параметров уравнения регрессии логарифма расходов y_t на логарифм доходов x_t обычным МНК дает следующие результаты:

$$\hat{y}_t=0{,}9x_t+\varepsilon_t,$$

n=25, R^2=0,80, критерий Дарбина-Уотсона 1,85, стандартная ошибка коэффициента регрессии 0,009.

Для тестирования рядов на коинтеграцию определим оценки остатков $\hat{\varepsilon}_t=\hat{y}_t-0{,}9x_t$ и построим регрессию первых разностей $\Delta\hat{\varepsilon}_t$ на $\hat{\varepsilon}_{t-1}$:

$$\Delta\hat{\varepsilon}_t=-0{,}95\hat{\varepsilon}_{t-1}.$$

Фактическое значение t-критерия для коэффициента последней регрессии равно –4,46, что превышает по абсолютной величине критическое значение 1,94, рассчитанное Энгелем и Гранжером, при уровне значимости 5%, т.е. с вероятностью 0,95 можно утверждать, что временные ряды логарифмов доходов и расходов на потребление коинтегрированы. ∇

При изучении двух взаимосвязанных временных рядов на предварительной стадии регрессионного анализа рекомендуется устранить сезонные или циклические колебания, если они имеются в исследуемых временных рядах, в соответствии с принятой аддитивной или мультипликативной моделями рядов.

Если рассматриваемые временные ряды y_t и x_t содержат тенденцию, то коэффициент корреляции, характеризующий степень зависимости между y_t и x_t будет иметь высокое значение. Такая же ситуация будет иметь место тогда, когда y_t и x_t зависят от переменной времени t. Как в первом, так и во втором случае имеет место ложная корреляция, которая приводит при построении регрессии y_t на x_t вида (5.12) к автокорреляции в остатках и нестационарности ряда остатков регрессии (ложная регрессия), то есть к нарушению предпосылок МНК.

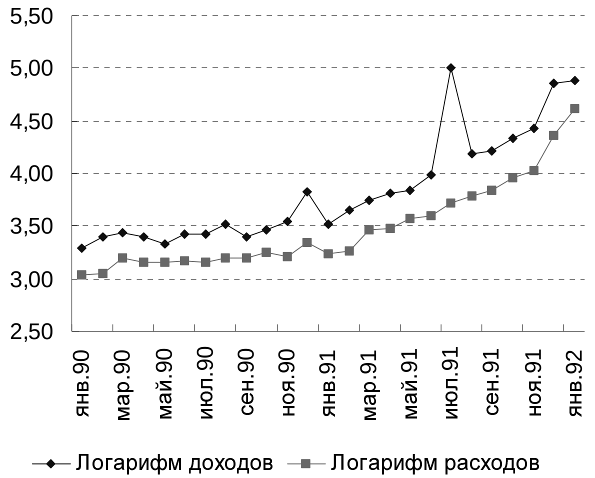


Рис. 5.13.

Для получения регрессии со стационарным временным рядом остатков ε_t, как уже указывалось ранее, может быть использован метод последовательных разностей, когда переход к некоторым k-м разностям уровней ряда позволяет получить стационарный ряд остатков.

Другими методами исключения тренда из анализируемой модели (5.12) являются методы включения фактора времени и отклонений от тренда.

Метод включения фактора времени.

Для устранения влияния времени на результат и факторы при изучении взаимосвязанных рядов динамики используется прием включения времени t в качестве независимой переменной в модель регрессии, что позволяет зафиксировать воздействие фактора t. Достоинством такого подхода является использование всей имеющейся выборки в отличие от метода последовательных разностей, который приводит к потере некоторого числа наблюдений.

Рассмотрим, например, модель вида:

$$y_t = \alpha + \beta_1 x_t + \beta_2 t + \varepsilon_t,$$

которая относится к моделям с включенным фактором времени. Параметры модели определяются обычным МНК.

Пример. Потребительские расходы и доходы населения (тыс. у. е.) за ряд лет характеризуются следующими данными (табл. 5.13).

Таблица 5.13

Показатель	Год								
	1	2	3	4	5	6	7	8	9
Потребительские расходы	46	50	54	59	62	67	75	86	100
Доходы	59	63	64	66	71	78	89	101	114

Оценим уравнение регрессии потребительских расходов y_t на доходы x_t вида:

$$y_t = \alpha + \beta x_t + \varepsilon_t.$$

Получим, применяя МНК:

$$y_t = -5{,}38 + 0{,}92 x_t + \varepsilon_t,$$

причем R^2=0,98, стандартная ошибка коэффициента β_1 при x_t 0,04, статистика Дарбина-Уотсона 0,86. Т.е. имеем случай мнимой регрессии, когда статистика Дарбина-Уотсона показывает наличие положительной автокорреляции остатков ε_t, а коэффициент детерминации близок к единице.

Применяя метод включения фактора времени, оценим регрессию вида:

$$y_t = \alpha + \beta_1 x_t + \beta_2 t + \varepsilon_t.$$

Получим, применяя МНК:

$$y_t = 3{,}88 + 0{,}69 x_t + 1{,}65 t + \varepsilon_t,$$

причем R^2=0,99, стандартная ошибка коэффициента β_1 при x_t 0,11, статистика Дарбина-Уотсона 1,3.

Полученное уравнение имеет следующую интерпретацию. Значение параметра β_1=0,69, говорит о том, что при увеличении дохода на 1 тыс. у.е., потребительские расходы возрастут в среднем на 0,69 тыс. у.е., если существующая тенденция будет неизменна. Значение β_2=1,65 свидетельствует о том, что без учета роста доходов населения ежегодный средний абсолютный прирост потребительских расходов составит 1,65 тыс. у.е. ∇

Метод отклонения уровней ряда от основной тенденции.

Если каждый из рядов y_t и x_t содержит тренд, то аналитическим выравниванием по каждому из рядов можно найти параметры тренда и определить расчетные по тренду уровни рядов $\hat{y}_t$ и $\hat{x}_t$. Влияние тенденции можно устранить путем вычитания расчетных значений тренда из фактических. Дальнейший регрессионный анализ проводят с отклонениями от тренда $y_t - \hat{y}_t$ и $x_t - \hat{x}_t$.

Пример. Потребительские расходы и доходы населения (тыс. у.е.) за ряд лет характеризуются данными табл. 5.13.

Рассчитаем линейные тренды по каждому из временных рядов методом МНК:

$\hat{y}_t$ =35,39+6,23t , R^2=0,93 стандартная ошибка коэффициента при t 0,63,

$\hat{x}_t$ =45,33+6,60t , R^2=0,89 стандартная ошибка коэффициента при t 0,85.

По трендам определим расчетные значения $\hat{y}_t$ и $\hat{x}_t$ и отклонения от трендов $y_t - \hat{y}_t$ и $x_t - \hat{x}_t$.

Таблица 5.14

Тренды и отклонения от трендов для временных рядов доходов и потребительских расходов

Время, t	y_t	x_t	$\hat{y}_t$	$\hat{x}_t$	$y_t - \hat{y}_t$	$x_t - \hat{x}_t$
1	46	59	41,62	51,93	4,38	7,07
2	50	63	47,86	58,53	2,14	4,47
3	54	64	54,09	65,13	-0,09	-1,13
4	59	66	60,32	71,73	-1,32	-5,73
5	62	71	66,56	78,33	-4,56	-7,33
6	67	78	72,79	84,93	-5,79	-6,93
7	75	89	79,02	91,53	-4,02	-2,53
8	86	101	85,26	98,13	0,74	2,87
9	100	114	91,49	104,73	8,51	9,27

Проверим полученные отклонения от трендов на автокорреляцию. Коэффициенты автокорреляции первого порядка составляют:

$$r_{\Delta x_t}(1) = 0{,}56,\ r_{\Delta y_t}(1) = 0{,}67,$$

в то время как для исходных рядов $r_{x_t}(1)$ =0,99, $r_{y_t}(1)$ =0,99.

Таким образом, полученные ряды отклонений от трендов можно использовать для получения количественной характеристики связи исходных временных рядов потребительских расходов и доходов населения. Коэффициент корреляции по отклонениям от трендов равен 0,93, тогда как этот же показатель по начальным уровням ряда был равен 0,99. Связь между потребительскими расходами и доходами населения прямая и сильная.

Результаты построения модели регрессии по отклонениям от трендов следующие:

Константа	0,00
Коэффициент регрессии	0,69
Стандартная ошибка коэффициента регрессии	0,09
R^2	0,88
Статистика Дарбина-Уотсона	1,30

Содержательная интерпретация модели в отклонениях от трендов затруднительна, но она может быть использована для прогнозирования. ∇

Библиографический список

1. Айвазян С.А., Мхитарян В.С. Прикладная статистика и основы эконометрики. Учебник для вузов. М.: ЮНИТИ, 1998. 1022 с.
2. Джонстон Дж. Эконометрические методы.- М.: Статистика, 1980. 432 с.
3. Доугерти К. Введение в эконометрику. М.: ИНФРА-М, 2001. 402 с.
4. Дрейпер Н., Смит Г. Прикладной регрессионный анализ. М.: Финансы и статистика, 1986. 392 с.
5. Магнус Я.Р, Катышев П.К., Пересецкий А.А. Эконометрика. Начальный курс. М.: Дело, 2000. 400 с.
6. Практикум по эконометрике/Под ред. И.И.Елисеевой. М.: Финансы и статистика, 2001. 192 с.
7. Эконометрика/Под ред. И.И. Елисеевой. М.: Финансы и статистика, 2001. 344 с.
8. Кремер Н., Путко Б. Эконометрика. М.: ЮНИТИ-ДАНА, 2002. 311 с.

Приложение

Статистические таблицы

Критерий Дарбина-Уотсона (d). Значения d_L и d_U при 5% уровне значимости.

n	k=1		k=2		k=3		k=4		k=5	
	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U
15	0,95	1,23	0,83	1,40	0,71	1,61	0,59	1,84	0,48	2,09
16	0,98	1,24	0,86	1,40	0,75	1,59	0,64	1,80	0,53	2,03
17	1,01	1,25	0,90	1,40	0,79	1,58	0,68	1,77	0,57	1,98
18	1,03	1,26	0,93	1,40	0,82	1,56	0,72	1,74	0,62	1,93
19	1,06	1,28	0,96	1,41	0,86	1,55	0,76	1,72	0,66	1,90
20	1,08	1,28	0,99	1,41	0,89	1,54	0,79	1,70	0,70	1,87
21	1,10	1,30	1,01	1,41	0,92	1,54	0,83	1,69	0,73	1,84
22	1,12	1,31	1,04	1,42	0,95	1,54	0,86	1,68	0,77	1,82
23	1,14	1,32	1,06	1,42	0,97	1,54	0,89	1,67	0,80	1,80
24	1,16	1,33	1,08	1,43	1,00	1,54	0,91	1,66	0,83	1,79
25	1,18	1,34	1,10	1,43	1,02	1,54	0,94	1,65	0,86	1,77
26	1,19	1,35	1,12	1,44	1,04	1,54	0,96	1,65	0,88	1,76
27	1,21	1,36	1,13	1,44	1,06	1,54	1,99	1,64	0,91	1,75
28	1,22	1,37	1,15	1,45	1,08	1,54	1,01	1,64	0,93	1,74
29	1,24	1,38	1,17	1,45	1,10	1,54	1,03	1,63	0,96	1,73
30	1,25	1,38	1,18	1,46	1,12	1,54	1,05	1,63	0,98	1,73
31	1,26	1,39	1,20	1,47	1,13	1,55	1,07	1,63	1,00	1,72
32	1,27	1,40	1,21	1,47	1,15	1,55	1,08	1,63	1,02	1,71
33	1,28	1,41	1,22	1,48	1,16	1,55	1,10	1,63	1,04	1,71
34	1,29	1,41	1,24	1,48	1,17	1,55	1,12	1,63	1,06	1,70
35	1,30	1,42	1,25	1,48	1,19	1,55	1,13	1,63	1,07	1,70
36	1,31	1,43	1,26	1,49	1,20	1,56	1,15	1,63	1,09	1,70
37	1,32	1,43	1,27	1,49	1,21	1,56	1,16	1,62	1,10	1,70
38	1,33	1,44	1,28	1,50	1,23	1,56	1,17	1,62	1,12	1,70
39	1,34	1,44	1,29	1,50	1,24	1,56	1,19	1,63	1,13	1,69
40	1,35	1,45	1,30	1,51	1,25	1,57	1,20	1,63	1,15	1,69
45	1,39	1,48	1,34	1,53	1,30	1,58	1,25	1,63	1,21	1,69
50	1,42	1,50	1,38	1,54	1,34	1,59	1,30	1,64	1,26	1,69
55	1,45	1,52	1,41	1,56	1,37	1,60	1,33	1,64	1,30	1,69
60	1,47	1,54	1,44	1,57	1,40	1,61	1,37	1,65	1,33	1,69
65	1,49	1,55	1,46	1,59	1,43	1,62	1,40	1,66	1,36	1,69
70	1,51	1,57	1,48	1,60	1,45	1,63	1,42	1,66	1,39	1,70
75	1,53	1,58	1,50	1,61	1,47	1,64	1,45	1,67	1,42	1,70
80	1,54	1,59	1,52	1,62	1,49	1,65	1,47	1,67	1,44	1,70
85	1,56	1,60	1,53	1,63	1,51	1,65	1,49	1,68	1,46	1,71
90	1,57	1,61	1,55	1,64	1,53	1,66	1,50	1,69	1,48	1,71
95	1,58	1,62	1,65	1,65	1,54	1,67	1,52	1,69	1,50	1,71
100	1,59	1,63	1,67	1,65	1,55	1,67	1,53	1,70	1,51	1,72

n - число наблюдений, k - число объясняющих переменных

Таблица критических величин n_u критерия последовательности знаков

n_1	n_2																		
	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
2											2	2	2	2	2	2	2	2	2
3					2	2	2	2	2	2	2	2	2	3	3	3	3	3	3
4				2	2	2	3	3	3	3	3	3	3	3	4	4	4	4	4
5			2	2	3	3	3	3	3	4	4	4	4	4	4	4	5	5	5
6		2	2	3	3	3	3	4	4	4	4	5	5	5	5	5	5	6	6
7		2	2	3	3	3	4	4	5	5	5	5	5	6	6	6	6	6	6
8		2	3	3	3	4	4	5	5	5	6	6	6	6	6	7	7	7	7
9		2	3	3	4	4	5	5	5	6	6	6	7	7	7	7	8	8	8
10		2	3	3	4	5	5	5	6	6	7	7	7	7	8	8	8	8	9
11		2	3	4	4	5	5	6	6	7	7	7	8	8	8	9	9	9	9
12	2	2	3	4	4	5	6	6	7	7	7	8	8	8	9	9	9	10	10
13	2	2	3	4	5	5	6	6	7	7	8	8	9	9	9	10	10	10	10
14	2	2	3	4	5	5	6	7	7	8	8	9	9	9	10	10	10	11	11
15	2	3	3	4	5	6	6	7	7	8	8	9	9	10	10	11	11	11	12
16	2	3	4	4	5	6	6	7	8	8	9	9	10	10	11	11	11	12	12
17	2	3	4	4	5	6	7	7	8	9	9	10	10	11	11	11	12	12	13
18	2	3	4	5	5	6	7	8	8	9	9	10	10	11	11	12	12	13	13
19	2	3	4	5	6	6	7	8	8	9	10	10	11	11	12	12	13	13	13
20	2	3	4	5	6	6	7	8	9	9	10	10	11	12	12	13	13	13	14

n_1	n_2																		
	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
2																			
3																			
4				9	9														
5			9	10	10	11	11												
6			9	10	11	12	12	13	13	13	13								
7				11	12	13	13	14	14	14	14	15	15	15					
8				11	12	13	14	14	15	15	16	16	16	16	17	17	17	17	17
9					13	14	14	15	16	16	16	17	17	18	18	18	18	18	18
10					13	14	15	16	16	17	17	18	18	18	19	19	19	20	20
11					13	14	15	16	17	17	18	19	19	19	20	20	20	21	21
12					13	14	16	16	17	18	19	19	20	20	21	21	21	22	22
13						15	16	17	18	19	19	20	20	21	21	22	22	23	23
14						15	16	17	18	19	20	20	21	22	22	23	23	23	24
15						15	16	18	18	19	20	21	22	22	23	23	24	24	25
16							17	18	19	20	21	21	22	23	23	24	25	25	25
17							17	18	19	20	21	22	23	23	24	25	25	26	26
18							17	18	19	20	21	22	23	24	25	25	26	26	27
19							17	18	20	21	22	23	23	24	25	26	26	27	27
20							17	18	20	21	22	23	24	25	25	26	27	27	28

Двусторонние квантили t - распределения Стьюдента

m	α						
	0,10	0,05	0,025	0,020	0,010	0,005	0,001
1	6,314	12,706	25,452	31,821	63,657	127,3	636,6
2	2,920	4,303	6,205	6,965	9,925	14,089	31,598
3	2,353	3,182	4,177	4,541	5,841	7,453	12,941
4	2,132	2,776	3,495	3,747	4,604	5,597	8,610
5	2,015	2,571	3,163	3,365	4,032	4,773	6,859
6	1,943	2,447	2,969	3,143	3,707	4,317	5,959
7	1,895	2,365	2,841	2,998	3,499	4,029	5,405
8	1,860	2,306	2,752	2,896	3,355	3,833	5,041
9	1,833	2,262	2,685	2,821	3,250	3,690	4,781
10	1,812	2,228	2,634	2,764	3,169	3,581	4,587
12	1,782	2,179	2,560	2,681	3,055	3,428	4,318
14	1,761	2,145	2,510	2,624	2,977	3,326	4,140
16	1,746	2,120	2,473	2,583	2,921	3,252	4,015
18	1,734	2,101	2,445	2,552	2,878	3,193	3,922
20	1,725	2,086	2,423	2,528	2,845	3,153	3,849
22	1,717	2,074	2,405	2,508	2,819	3,119	3,792
24	1,711	2,064	2,391	2,492	2,797	3,092	3,745
26	1,706	2,056	2,379	2,479	2,779	3,067	3,707
28	1,701	2,048	2,369	2,467	2,763	3,047	3,674
30	1,697	2,042	2,360	2,457	2,750	3,030	3,646
∞	1,645	1,960	2,241	2,326	2,576	2,807	3,291

m - число степеней свободы

Квантили распределения χ^2

Число степеней свободы	Уровень значимости					
	0,50	0,30	0,20	0,10	0,05	0,01
1	0,455	1,074	1,642	2,706	3,841	6,635
2	1,386	2,408	3,219	4,605	5,991	9,210
3	2,366	3,665	4,642	6,251	7,815	11,341
4	3,357	4,878	5,989	7,779	9,488	13,277
5	4,351	6,064	7,289	9,236	11,070	15,086
6	5,348	7,231	8,558	10,645	12,592	16,812
7	6,346	8,383	9,803	12,017	14,067	18,475
8	7,344	9,524	11,030	13,362	15,507	20,090
9	8,343	10,656	12,242	14,684	16,919	21,666
10	9,342	11,781	13,442	15,987	18,307	23,209
11	10,341	12,899	14,631	17,272	19,675	24,725
12	11,340	14,011	15,812	18,549	21,026	26,217
13	12,340	15,119	16,985	19,812	22,362	27,688
14	13,339	16,222	18,151	21,064	23,685	29,141
15	14,339	18,322	19,311	22,307	24,996	30,578
16	15,338	18,418	20,465	23,542	26,296	32,000
18	17,338	20,601	22,760	25,989	28,869	34,805
20	19,337	22,775	25,038	28,412	31,410	37,566
24	23,337	27,096	29,553	33,196	36,415	42,980
30	29,336	33,530	36,250	40,256	43,773	50,892

Если число степеней свободы больше 30, то выражение $\sqrt{2\chi^2} - \sqrt{2n-1}$ можно рассматривать как переменную со стандартным нормальным распределением, где n - число степеней свободы.

95% квантили распределения Фишера $F(n_1, n_2)$

n_2	n_1																		
	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	161	200	216	225	230	234	237	239	241	242	244	246	248	249	250	251	252	253	254
2	18,5	19,0	19,2	19,2	19,3	19,3	19,4	19,4	19,4	19,4	19,4	19,4	19,4	19,5	19,5	19,5	19,5	19,5	19,5
3	10,1	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79	8,74	8,70	8,66	8,64	8,62	8,59	8,57	8,55	8,53
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,91	5,86	5,80	5,77	5,75	5,72	5,69	5,66	5,63
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74	4,68	4,62	4,56	4,53	4,50	4,46	4,43	4,40	4,37
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,00	3,94	3,87	3,84	3,81	3,77	3,74	3,70	3,67
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64	3,57	3,51	3,44	3,41	3,38	3,34	3,30	3,27	3,23
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35	3,28	3,22	3,15	3,12	3,08	3,04	3,01	2,97	2,93
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14	3,07	3,01	2,94	2,90	2,86	2,83	2,79	2,75	2,71
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98	2,91	2,85	2,77	2,74	2,70	2,66	2,62	2,58	2,54
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85	2,79	2,72	2,65	2,61	2,57	2,53	2,49	2,45	2,40
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,75	2,69	2,62	2,54	2,51	2,47	2,43	2,38	2,34	2,30
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67	2,60	2,53	2,46	2,42	2,38	2,34	2,30	2,25	2,21
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60	2,53	2,46	2,39	2,35	3,31	2,27	2,22	2,18	2,13
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54	2,48	2,40	2,33	2,29	2,25	2,20	2,16	2,11	2,07
16	4,49	3,53	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,42	2,35	2,28	2,24	2,19	2,15	2,11	2,06	2,01
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41	2,34	2,27	2,19	2,15	2,11	2,06	2,02	1,97	1,92
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35	2,28	2,20	2,12	2,08	2,04	1,99	1,95	1,90	1,84
22	4,30	3,44	3,05	2,82	2,66	2,55	2,46	2,40	2,34	2,30	2,23	2,15	2,07	2,03	1,98	1,94	1,89	1,84	1,78
24	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,36	2,30	2,25	2,18	2,11	2,03	1,98	1,94	1,89	1,84	1,79	1,73
25	4,24	3,39	2,99	2,76	2,60	2,49	2,40	2,34	2,28	2,24	2,16	2,09	2,01	1,96	1,92	1,87	1,82	1,77	1,71
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21	2,16	2,09	2,01	1,93	1,89	1,84	1,79	1,74	1,68	1,62
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	2,08	2,00	1,92	1,84	1,79	1,74	1,69	1,64	1,58	1,51
60	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10	2,04	1,99	1,92	1,84	1,75	1,70	1,65	1,59	1,53	1,47	1,39
120	3,92	3,07	2,68	2,45	2,29	2,18	2,09	2,02	1,96	1,91	1,83	1,75	1,66	1,61	1,55	1,50	1,43	1,35	1,25
∞	3,84	3,00	2,60	2,37	2,21	2,10	2,01	1,94	1,88	1,83	1,75	1,67	1,57	1,52	1,46	1,39	1,32	1,22	1,00

n_1 – число степеней свободы числителя, n_2 – число степеней свободы знаменателя

Эконометрика

Учебное пособие

Арженовский Сергей Валентинович
Федосова Оксана Николаевна

Директор издательства В.Е. Смейле
Редактор О.Н. Шимко
Корректор Е.В. Барыбин
Компьютерная верстка и макетирование авторов

Изд.№ 47/5577 Подписано к печати 14.03.2002 Бумага офсетная
Печать офсетная Формат 60 •84/16 Объем 6,38 уч.- изд.л.
Заказ № Тираж 200 экз. "С" 47

344007, г. Ростов-на-Дону, ул. Б. Садовая, 69. РГЭУ. Издательство.

Отпечатано в отделе оперативной полиграфии РГЭУ «РИНХ».