Тема 3. Моделирование временных рядов

Модели, построенные по временным данным, представляют *модели временных рядов.* Временной ряд – ряд значений какого-либо показателя, характеризующий один и тот же объект за несколько последовательных моментов или периодов времени. Уровень временного ряда складывается из следующих основных компонентов:

- трендовой компоненты, характеризующей основную тенденцию ряда (*T*);
- циклической компоненты, характеризующей циклические колебания изучаемого явления (*S*);
- случайной компоненты, которая является результатом воздействия множества случайных факторов (*E*).

При различных сочетаниях этих факторов зависимость уровней ряда от времени может принимать разные формы. Во-первых, большинство временных экономических показателей имеют тенденцию, которая характеризует совокупное долговременное воздействие множества факторов на динамику изучаемого показателя. По всей видимости, эти факторы, взятые в отдельности, могут оказывать разнонаправленное воздействие на исследуемый показатель. Однако в совокупности они формируют его возрастающую или убывающую тенденцию (рис. 3.1).

Рис. 3.1 Возрастающая тенденция временного ряда

Во-вторых, изучаемый показатель может быть подвержен циклическим колебаниям. Эти колебания могут носить сезонный характер, поскольку экономическая деятельность ряда отраслей зависит от времени года (например, цены на сельскохозяйственную продукцию в летний период выше, чем в зимний; уровень безработицы в курортных городах в зимний период выше по сравнению с летним). При наличии больших массивов данных за длительные промежутки времени можно выявить циклические колебания, связанные с общей динамикой конъюнктуры рынка, а также с фазой бизнес-цикла, в которой находится страна.

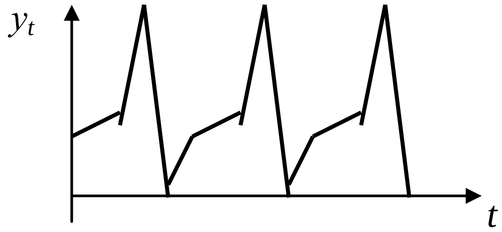


Рис. 3.2. Сезонная компонента временного ряда

Некоторые временные ряды не содержат тенденции и циклическую компоненту, а каждый следующий их уровень образуется как сумма среднего уровня ряда и некоторой (положительной или отрицательный) случайной компоненты.

Рис. 3.3. Случайная компонента временного ряда

Очевидно, что реальные данные не соответствуют полностью ни одной из описанных выше моделей. Чаще всего они содержат все три компоненты. Каждый их уровень формируется под воздействием тенденции, сезонных и случайной компоненты.

Фактический уровень временного ряда можно представить как функцию от этих компонент: $Y = f(T, S, E)$. В зависимости от вида связи между этими компонентами может быть построена либо аддитивная модель: $Y = T + S + E$, либо мультипликативная модель: $Y = T \cdot S \cdot E$ ряда динамики.

Основная задача эконометрического исследования отдельного временного ряда – выявление и придание количественного выражения каждой из перечисленных выше компонент, с тем чтобы использовать полученную информацию для прогнозирования будущих значений ряда или построении моделей взаимосвязи двух или более временных рядов.

При наличии тенденции и циклических колебаний значения каждого последующего уровня ряда зависят от предыдущих значений. *Автокорреляция уровней ряда* – корреляционная связь между последовательными уровнями одного и того же ряда динамики. Количественно ее можно измерить с помощью линейного коэффициента корреляции между уровнями исходного временного ряда и уровнями этого ряда, сдвинутыми на несколько шагов.

Пример. Пусть имеются следующие условные данные предприятия о средних расходах на конечное потребление за 8 лет (табл. 3.1).

Таблица 3.1 – Данные предприятия о средних расходах на конечное потребление за 8 лет

Год, t	1	2	3	4	5	6	7	8
Расходы на потребление, y_t	7	8	8	10	11	12	14	16

Можно предположить, что расходы на конечное потребление в текущем году зависят от расходов на конечное потребление предыдущих лет.

Определим коэффициент корреляции между рядами y_t и y_{t-1}.

Одна из рабочих формул для расчета коэффициента корреляции имеет вид:

$$r_{yx} = \frac{\sum (x_j - \bar{x}) \cdot (y_j - \bar{y})}{\sqrt{\sum (x_j - \bar{x})^2 \cdot \sum (y_j - \bar{y})^2}} .$$

В качестве переменной x рассмотрим ряд $y_2,\ y_3, \ldots, y_8$; в качестве переменной y – $y_1,\ y_2, \ldots, y_7$. Тогда приведенная выше формула примет вид:

$$r_1 = \frac{\sum_{t=2}^{n} (y_t - \bar{y}_1) \cdot (y_{t-1} - \bar{y}_2)}{\sqrt{\sum_{t=2}^{n} (y_t - \bar{y}_1)^2 \cdot \sum_{t=2}^{n} (y_{t-1} - \bar{y}_2)^2}} ,$$

где

$$\bar{y}_1 = \frac{\sum_{t=2}^{n} y_t}{n-1} ; \quad \bar{y}_2 = \frac{\sum_{t=2}^{n} y_{t-1}}{n-1} .$$

Эту величину называют коэффициентом автокорреляции уровней ряда первого порядка, так как он измеряет зависимость между соседними уровнями ряда y_t и y_{t-1}, т.е. при лаге 1.

Таблица 3.2 – **Расчетные данные**

t	y_t	y_{t-1}	$y_t - \bar{y}_1$	$y_{t-1} - \bar{y}_2$	$(y_t - \bar{y}_1)*(y_{t-1} - \bar{y}_2)$	$(y_t - \bar{y}_1)^2$	$(y_{t-1} - \bar{y}_2)^2$
1	7	-	-	-	-	-	-
2	8	7	-3,29	-3	9,87	10,8241	9
3	8	8	-3,29	-2	6,58	10,8241	4
4	10	8	-1,29	-2	2.58	1,6641	4
5	11	10	-0,29	0	0	0,0841	0
6	12	11	0,71	1	0,71	0,5041	1
7	14	12	2,71	2	5,42	7,3441	4
8	16	14	4,71	4	18,84	22,1841	16
Итого	86	70			44	53,4287	38

Для данного примера средние значения составят:

$$\bar{y}_1 = \frac{\sum_{t=2}^{n} y_t}{n-1} = \frac{8+8+10+11+12+14+16}{8-1} = \frac{79}{7} = 11{,}29;$$

$$\bar{y}_2 = \frac{\sum_{t=2}^{n} y_{t-1}}{n-1} = \frac{7+8+8+10+11+12+14}{8-1} = \frac{70}{7} = 10.$$

$$r_1 = \frac{\sum_{t=2}^{n}(y_t - \bar{y}_1)\cdot(y_{t-1} - \bar{y}_2)}{\sqrt{\sum_{t=2}^{n}(y_t - \bar{y}_1)^2 \cdot \sum_{t=2}^{n}(y_{t-1} - \bar{y}_2)^2}} = \frac{44}{\sqrt{53{,}43\cdot 38}} = 0{,}976.$$

Полученное значение свидетельствует об очень тесной зависимости между расходами на конечное потребление текущего и непосредственно предшествующих годов и, следовательно, о наличие во временном ряде расходов на конечное потребление сильной линейной тенденции.

Аналогично можно определить коэффициенты второго и более высоких порядков. Так, коэффициент автокорреляции второго порядка характеризует тесноту связи между уровнями ряда y_t и y_{t-2}, и определяется по формуле:

$$r_2 = \frac{\sum_{t=3}^{n}(y_t - \bar{y}_3)\cdot(y_{t-1} - \bar{y}_4)}{\sqrt{\sum_{t=3}^{n}(y_t - \bar{y}_3)^2 \cdot \sum_{t=2}^{n}(y_{t-2} - \bar{y}_4)^2}},$$

где

$$\bar{y}_3 = \frac{\sum_{t=3}^{n} y_t}{n-2}; \quad \bar{y}_2 = \frac{\sum_{t=3}^{n} y_{t-2}}{n-2}.$$

Число периодов, по которым рассчитываются коэффициент автокорреляции, называется ***лагом***. С увеличением лага число пар значений, по которым рассчитывается коэффициент автокорреляции, уменьшается.

Свойства коэффициента автокорреляции:

- так как он строится по аналогии с линейным коэффициентом корреляции, то он характеризует тесноту только линейной связи текущего и предыдущего уровней связи; по данному коэффициенту можно судить о наличии линейной тенденции; для некоторых временных рядов, имеющих сильную нелинейную тенденцию коэффициент может приближаться к нулю;

- по знаку коэффициента автокорреляции сделать вывод о тенденции ряда невозможно.

Последовательность коэффициентов автокорреляции уровней первого, второго и т.д. порядков называют ***автокорреляционной функцией временного***

ряда. График зависимости ее значений от величины лага называется ***коррелограммой***.

Рассчитав несколько коэффициентов автокорреляции можно определить лаг, при котором автокорреляция наиболее высока, выявив тем самым *структуру временного ряда*. Если наиболее высоким оказывается значение $r_{t,t-1}$, то исследуемый ряд содержит тенденцию; если наиболее высоким оказался $r_{t,t-L}$, то ряд содержит циклические колебания с периодом *L*. Если ни один из коэффициентом не является значимым, можно сделать одно из двух предположений:

- либо ряд не содержит тенденции и циклических колебаний, а его уровень определяется только случайной компонентой;

- либо ряд содержит сильную нелинейную тенденцию.

Построение аналитической функции для моделирования тенденции (тренда) временного ряда называют *аналитическим выравниванием временного ряда*. Для этого чаще всего применяются следующие функции:

- линейная $\hat{y}_t = a + b \cdot t$;
- гипербола $\hat{y}_t = a + b / t$;
- экспонента $\hat{y}_t = e^{a+b*t}$;
- степенная функция $\hat{y}_t = a * t^b$;
- парабола второго и более высоких порядков $\hat{y}_t = a + b_1 \cdot t + b_2 \cdot t^2 + ... + b_k \cdot t^k$.

Параметры трендов определяются МНК, в качестве независимой переменной выступает время, а в качестве зависимой переменной – фактические уровни временного ряда. Для нелинейных трендов предварительно проводят стандартную процедуру их линеаризации.

Критерием отбора наилучшей формы тренда является наибольшее значение скорректированного коэффициента детерминации.

Пример. Определим МНК параметры линейного тренда на основе данных табл. 3.3.

Таблица 3.3 – **Расчет теоретических уровней линейного тренда**

Год	Производство мяса, тыс. т.	Условное обозначение времени, *t*	t^2	yt	$\hat{y}_t$	$(y-\hat{y}_t)^2$
2019	9,4	1	1	9,4	9,28	0,0144
2020	8,3	2	4	16,6	8,43	0,0169
1921	7,5	3	9	22,5	7,58	0,0064
2022	6,8	4	16	27,2	6,73	0,0049
2023	5,9	5	25	29,5	5,88	0,0004
n=5	$\sum y = 37{,}9$	$\sum t = 15$	$\sum t^2 = 55$	$\sum yt = 105{,}2$	$\sum \hat{y}_t = 37{,}9$	0,043

Приняв в качестве гипотетической функции теоретических уровней прямую $\hat{y}_t = a + b \cdot t$, определим параметры последней:

$$\begin{cases} na + b_1 \sum t = \sum y, \\ a\sum t + b_1 \sum t^2 = \sum yt \end{cases}$$

$$\begin{cases} 5a + 15b_1 = 37{,}9 \\ 15a + 55b_1 = 105{,}2 \end{cases}$$

Решение этой системы можно осуществить по формулам:

$$b_1 = \frac{n\sum yt - \sum t \sum y}{n\sum t^2 - (\sum t)^2}, \qquad a = \frac{\sum y \sum t^2 - \sum t \sum yt}{n\sum t^2 - (\sum t)^2} \qquad \text{или} \quad a = \bar{y} - b_1 \bar{t}\,.$$

$$b_1 = \frac{5*105{,}2 - 15*37{,}9}{5*55 - 15^2} = -0{,}85, \qquad a = \frac{\sum y}{n} - b_1 \frac{\sum t}{n} = \frac{37{,}9}{5} - (-0{,}85)\frac{15}{5} = 10{,}13.$$

Отсюда искомое уравнение тренда: $\hat{y}_t = 10{,}13 - 0{,}85t$. Подставляя в полученное уравнении значения 1, 2, 3, 4, 5, определяем теоретические уровни ряда (см. предпоследнюю графу табл. 3.3). Сравнивая значения эмпирических и теоретических уровней, видим, что они близки, т.е. можно сказать, что найденное уравнение весьма удачно характеризует основную тенденцию изменения уровней именно как линейную функцию.

Система нормальных уравнений упрощается, если отсчет времени ведется от середины ряда. Например, при *нечетном числе уровней* серединная точка (год, месяц) принимается за нуль. Тогда предшествующие периоды обозначаются соответственно -1, -2, -3 и т.д., а следующие за средним – соответственно +1, +2, +3 и т.д. При четном числе уровней два срединных момента (периода) времени обозначают -1 и +1, а все последующие и предыдущие, соответственно, через два интервала: ± 3, ± 5, ± 7 и т.д.

При таком порядке отсчета времени (от середины ряда) $\sum t = 0$, система нормальных уравнений упрощается до следующих двух уравнений, каждое из которых решается самостоятельно:

$$\begin{cases} na = \sum y \quad \Rightarrow \quad a = \dfrac{\sum y}{n}, \\ b_1 \sum t^2 = \sum yt \quad \Rightarrow \quad b_1 = \dfrac{\sum yt}{\sum t^2} \end{cases}$$

Важное значение при построении модели временного ряда имеет учет сезонных и циклических колебаний. Простейшим подходом, позволяющим учесть в модели сезонные и циклические колебания, является расчет значений сезонной/циклической компоненты и построение аддитивной и мультипликативной модели временного ряда.

Общий вид аддитивной модели следующий: $Y=T+S+E$. Эта модель предполагает, что каждый уровень временного уровня ряда может быть представлен как сумма трендовой T, сезонной S и случайной компонент. Общий вид мультипликативной модели выглядит как: $Y=T \cdot S \cdot E$.

Выбор одной из двух моделей проводится на основе анализа структуры сезонных колебаний. Если амплитуда колебаний приблизительно постоянна, строят аддитивную модель временного ряда, в которой значения сезонной компоненты предполагаются постоянными для различных циклов. Если амплитуда сезонных колебаний возрастает или уменьшается, строят мультипликативную модель временного ряда, которая ставит уровни ряда в зависимость от значений сезонной компоненты.

Построение аддитивной и мультипликативной моделей сводится к расчету T, S, E для каждого уровня ряда. Этапы построения модели включают в себя следующие шаги:

1. Выравнивание исходного ряда методом скользящей средней
2. Расчет значений сезонной компоненты S.
3. Устранение сезонной компоненты из исходных уровней ряда и получение выровненных данных в аддитивной $(T+E)$ или мультипликативной $(T \cdot E)$ модели.
4. Аналитическое выравнивание уровней $(T+E)$ или $(T \cdot E)$ и расчет значений T с использованием полученного уравнения тренда.
5. Расчет полученных по модели значений $(T+E)$ или $(T \cdot E)$.
6. Расчет абсолютных и/или относительных ошибок. Если полученные значения не содержат автокорреляции, ими можно заменить исходные уровни ряда и в дальнейшем использовать

временной ряд ошибок E для анализа взаимосвязи исходного ряда и других временных рядов.

Рассмотрим другие методы анализа взаимосвязи, предположив что изучаемые временные ряды не содержат периодических колебаний. Допустим, что изучается зависимость между рядами x и y. Для количественной характеристики этой зависимости используется линейный коэффициент корреляции. Если рассматриваемые временные ряды имеют тенденцию, коэффициент корреляции по абсолютной величине будет высоким. Однако это не говорит о том, что x причина y. Высокий коэффициент корреляции в данном случае – это результат того, что x и y зависят от времени, или содержат тенденцию. При этом одинаковую или противоположную тенденцию могут иметь ряды, совершенно не связанные друг с другом причинно-следственной зависимостью. Например, коэффициент корреляции между численностью выпускников вузов и числом домов отдыха в РФ в период с 1970-1990 г. составил 0,8. Однако, это не говорит о том, что количество домов отдыха способствует росту числа выпускников или наоборот.

Для того чтобы получить коэффициенты корреляции, характеризующие причинно-следственную связь между изучаемыми рядами, следует избавиться от так называемой ложной корреляции, вызванной наличием тенденции в каждом ряду, которую устраняют одним из методов.

Предположим, что по двум временным рядам x_t и y_t строится уравнение парной регрессии линейной регрессии вида: $\hat{y}_t = a + b_1 \cdot x_t + \varepsilon_t$. наличие тенденции в каждом из этих временных рядов означает, что на зависимую y_t и независимую x_t переменные модели оказывает воздействие фактор времени, который непосредственно в модели не учтен. Влияние фактора времени будет выражено в корреляционной зависимости между значениями остатков ε_t за текущий и предыдущие моменты времени, которая получила название автокорреляции в остатках.

Автокорреляция в остатках – это нарушение одной из основных предпосылок МНК – предпосылки о случайности остатков, полученных по уравнению регрессии. Один из возможных путей решения этой проблемы состоит в применении обобщенного МНК.

Для устранения тенденции используются две группы методов:

- методы, основанные на преобразовании уровней исходного ряда в новые переменные, не содержащие тенденции (метод последовательных разностей и метод отклонения от трендов);

- методы, основанные на изучении взаимосвязи исходных уровней временных рядов при элиминировании воздействия фактора времени на зависимую и независимую переменные модели (включение в модель регрессии по временным рядам фактора времени).

Пусть имеются два временных ряда x_t и y_t, каждый из которых содержит трендовую компоненту T и случайную составляющую ε_t. Аналитическое выравнивание каждого из этих рядов позволяет найти параметры соответствующих уравнений трендов и определить расчетные по тренду уровни $\hat{y}_t$ и $\hat{x}_t$ соответственное. Эти расчетные значения можно принять за оценку трендовой компоненты T каждого ряда. Поэтому влияние тенденции можно устранить путем вычитания расчетных значений уровней ряда из фактических. Эту процедуру проделывают для каждого временного ряда в модели. Дальнейший анализ взаимосвязи рядов проводят с использованием не исходных уровней, а отклонений от тренда $x_t - \hat{x}_t$ и $y_t - \hat{y}_t$. Именно в этом и заключается *метод отклонений от тренда*.

В ряде случаев вместо аналитического выравнивания временного ряда с целью устранения тенденции можно применить более простой метод – *метод последовательных разностей.* Если временной ряд содержит ярко выраженную линейную тенденцию, ее можно устранить путем замены исходных уровней ряда цепными абсолютными приростами (первыми разностями).

Пусть $y_t = \hat{y}_t + \varepsilon_t$, $\hat{y}_t = a + b \cdot t$.

Тогда $\Delta_t = y_t - y_{t-1} = a + b \cdot t + \varepsilon_t - (a + b \cdot (t-1) + \varepsilon_{t-1}) = b + (\varepsilon_t - \varepsilon_{t-1})$.

Коэффициент b – константа, которая не зависит от времени. При наличии сильной линейной тенденции отставки достаточно малы и в соответствии с предпосылками МНК носят случайный характер. Поэтому первые разности уровней ряда Δ_t не зависят от переменной времени, их можно использовать для дальнейшего анализа.

Если временной ряд содержит тенденцию в форме параболы второго порядка, то для ее устранения можно заменить исходные уровни ряда на вторые разности: $\Delta_t^2 = \Delta_t - \Delta_{t-1} = 2b + (\varepsilon_t - 2\varepsilon_{t-1} + \varepsilon_{t-2})$.

Если тенденции временного ряда соответствует экспоненциальной, или степенной, тренд, метод последовательных разностей следует применять не к исходным уровням ряда, а к их логарифмам.

Модель вида: $\hat{y}_t = a + b_1 \cdot x_t + b_2 \cdot t + \varepsilon_t$ также относится к группе моделей, включающих фактор времени. Преимущество данной модели перед методами отклонений от трендов и последовательных разностей состоит в том, что она позволяет учесть всю информацию, содержащуюся в исходных данных, поскольку значения x_t и y_t – это уровни исходных временных рядов. Кроме того, модель строится по всей совокупности данных за рассматриваемый период в отличие от метода последовательных разностей, который приводит к потере числа наблюдений. Параметры этой модели определяются обычным МНК.

Пример. Построим уравнение тренда по исходным данным таблицы 3.4.

Таблица 3.4– **Расходы на конечное потребление и совокупный доход (усл. ед.)**

Год	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Расходы на конечное потребление, y_t	7	8	8	10	11	12	14	16
Совокупный доход, x_t	10	12	11	12	14	15	17	20

Система нормальных уравнений имеет вид:

$$\begin{cases} na + b_1\sum x_t + b_2\sum t = \sum y_t, \\ a\sum x_t + b_1\sum x_t^2 + b_2\sum tx_t = \sum x_t y_t, \\ a\sum t + b_1\sum tx_t + b_2\sum t^2 = \sum ty_y. \end{cases}$$

По исходным данным рассчитаем необходимые величины и подставим в систему:

$$\begin{cases} 8a + 111b_1 + 36b_2 = 86 \\ 111a + 1619b_1 + 554b_2 = 1266 \\ 36a + 554b_1 + 204b_2 = 440 \end{cases}$$

Уравнение регрессии имеет вид: $y_t = 1{,}15 + 0{,}49 \cdot x_t + 0{,}63 \cdot t + \varepsilon_t$.

Интерпретация параметров уравнения следующая: $b_1 = 0{,}49$ характеризует, что при увеличении совокупного дохода на 1 д.е. расходы на конечное потребление возрастут в среднем на 0,49 д.е в условиях существования неизменной тенденции. Параметр $b_2 = 0{,}63$ означает, что воздействие всех факторов, кроме совокупного дохода, на расходы на конечное потребление приведет к его среднегодовому абсолютному приросту на 0,63 д.е.

Рассмотрим уравнение регрессии вида: $y_t = a + \sum b_j x_{jt} + \varepsilon_t$. Для каждого момента времени значение компоненты ε_t определяются как $\varepsilon_t = y_t - \hat{y}_t$ или $\varepsilon_t = y_t - (a + \sum b_j x_{jt})$. Рассматривая последовательность остатков как временной ряд, можно построить график их зависимости от времени. В соответствии с предпосылками МНК остатки должны быть случайными (рис. 3.4).

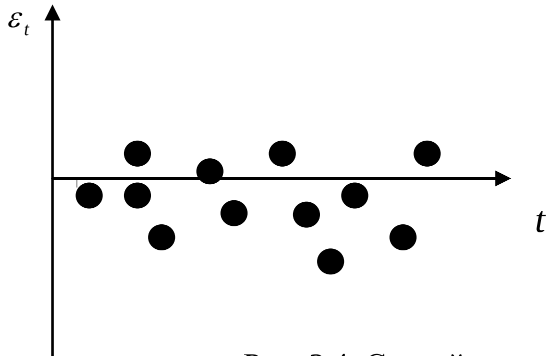
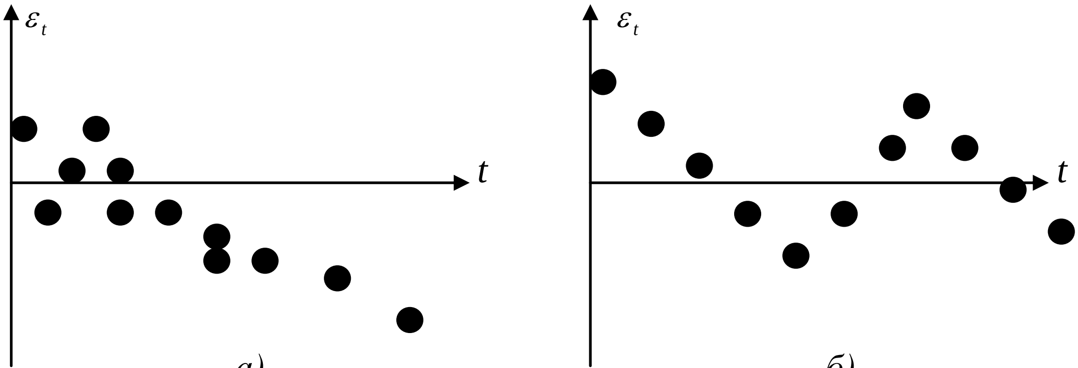


Рис. 3.4 Случайные остатки

Однако при моделировании временных рядов нередко встречаются ситуации, когда остатки содержат тенденцию или циклические колебания (рис. 3.5). Это говорит о том, что каждое следующее значение остатков зависит от предшествующих. В этом случае говорят о наличии автокорреляции в остатках.



а) *б)*

Рис. 3.5 Убывающая тенденция (*а*) и циклические колебания (*б*) в остатках

Автокорреляция случайной составляющей ε - корреляционная зависимость текущих и предыдущих значений случайной составляющей. Последствия автокорреляции случайной составляющей:

- коэффициенты регрессии становятся неэффективными;

- стандартные ошибки коэффициентов регрессии становятся заниженными, а значения t–критерия завышенными.

Для определения автокорреляции остатков известны два наиболее распространенных метода определения автокорреляции остатков. Первый метод – это построение графика зависимости остатков от времени и визуальное определение наличия или отсутствия автокорреляции. Второй метод – это использование критерия Дарбина-Уотсона, который сводится к проверке гипотезы:

- Н0 (основная гипотеза): автокорреляция отсутствует;

- Н1 и Н2 (альтернативные гипотезы): присутствует положительная или отрицательная автокорреляция в остатках соответственно.

Для проверки основной гипотезы используется статистика критерия Дарбина-Уотсона:

$$d = \frac{\sum_{t=2}^{n}(\varepsilon_t - \varepsilon_{t-1})^2}{\sum_{t=1}^{n}\varepsilon_t^2}, \text{ где } \varepsilon_t = y_t - \hat{y}_t .$$

На больших выборках $d \approx 2(1 - r_1^{\varepsilon})$, где r_1^{ε} - коэффициент автокорреляции 1-го порядка.

$$r_1^{\varepsilon} = \frac{\sum_{t=2}^{n}(\varepsilon_t - \overline{\varepsilon_1})\cdot(\varepsilon_{t-1} - \overline{\varepsilon_2})}{\sqrt{\sum_{t=2}^{n}(\varepsilon_t - \overline{\varepsilon_1})\cdot\sum_{t=2}^{n}(\varepsilon_{t-1} - \overline{\varepsilon_2})}} \approx \frac{\sum_{t=2}^{n}\varepsilon_t\cdot\varepsilon_{t-1}}{\sqrt{\sum_{t=2}^{n}\varepsilon_t^2\cdot\sum_{t=2}^{n}\varepsilon_{t-1}^2}} \approx \frac{\sum_{t=2}^{n}\varepsilon_t\cdot\varepsilon_{t-1}}{\sum_{t=2}^{n}\varepsilon_t^2} .$$

Если в остатках существует полная положительная автокорреляция и $r_1^{\varepsilon}=1$, то $d=0$; если в остатках есть полная отрицательная автокорреляция, то $r_1^{\varepsilon} = -1$ и $d=4$; если автокорреляция остатков отсутствует, то $r_1^{\varepsilon} = 0$, то $d=2$. Следовательно, $0 \le d \le 4$.

Существуют специальные статистические таблицы для определения нижней и верхней критических границ d-статистики – d_L и d_U. Они определяются в зависимости от n, числа независимых переменных k и уровня значимости α.

Если $d_{набл} < d_L$, то принимается гипотеза Н1: положительная автокорреляция.

Если $d_u < d_{набл} < 2$, то принимается гипотеза Н0: автокорреляции нет.

Если $2 < d_{набл} < 4 - d_u$, то принимается гипотеза Н0: автокорреляции нет.

Если $d_{набл} > 4 - d_L$, то принимается гипотеза Н2: отрицательная автокорреляция.

Если $4 - d_u < d_{набл} < 4 - d_L$, и $d_L < d_{набл} < d_u$, то имеет место случай неопределенности.

Рис. 3.6 Алгоритм проверки гипотезы о наличии автокорреляции остатков

Для применения критерия Дарбина-Уотсона есть ограничения. Он неприменим для моделей, включающих в качестве независимых переменных лаговые значения результативного признака, т.е. к моделям авторегрессии. Методика направлена только на выявление автокорреляции остатков первого порядка. Результаты являются более достоверными при работе с большими выборками.

В тех случаях, когда имеет место автокорреляция остатков, для определения оценок параметров *a*, *b* используют обобщенный метод МНК, который заключается в последовательности следующих шагов:

1. Преобразовать исходные переменные y_t и x_t к виду

$$y'_t = y_t - r_1^{\varepsilon} \cdot y_{t-1}$$

$$x'_t = x_t - r_1^{\varepsilon} \cdot x_{t-1}$$

2. Применив обычный МНК к уравнению $y'_t = a' + b \cdot x'_t + u_t$, где $u_t = \varepsilon_t - r_1^{\varepsilon} \cdot \varepsilon_{t-1}$ определить оценки параметров a' и *b*.

3. Рассчитать параметр *a* исходного уравнения из соотношения как $a = \dfrac{a'}{1 - r_1^{\varepsilon}}$.

4. Выписать исходное уравнение $\hat{y}_t = a + b_1 \cdot x_t + \varepsilon_t$.

Среди эконометрических моделей, построенных по временным данным, выделяют динамические модели.

Эконометрическая модель является ***динамической***, если в данный момент времени *t* она учитывает значения входящих в нее переменных, относящихся как к текущему, так и к предыдущим моментам времени, т.е. эта модель отражает динамику исследуемых переменных в каждый момент времени.

Существует два основных типа динамических эконометрических моделей. К моделям первого типа относятся модели авторегрессии и модели с распределенным лагом, в которых значение переменной за прошлые периоды времени (лаговые переменные) непосредственно включены в модель. Модели второго типа учитывают динамическую информацию в неявном виде. В эти модели включены переменные, характеризующие ожидаемый и желаемый уровень результата, или один из факторов в момент времени *t*.

Модель с распределенным лагом имеет вид:

$$\hat{y}_t = a + b_0 \cdot x_t + b_1 \cdot x_{t-1} + b_2 \cdot x_{t-2} + \varepsilon_t$$

Модель авторегрессии имеет вид:

$$y_t = a + b_0 * x_t + c_1 * y_{t-1} + \varepsilon_t$$

Построение моделей с распределенным лагом и моделей авторегрессии имеет свою специфику. Во-первых, оценка параметров моделей авторегрессии, а в большинстве случаев и моделей распределенным лагом не может быть проведена с помощью обычного МНК ввиду нарушения его предпосылок и требует специальных статистических методов. Во-вторых, исследователям приходится решать проблемы выбора оптимальной величины лага и определения его структуры. Наконец, в третьих, между моделями с распределенным лагом и моделями авторегрессии имеется определенная взаимосвязь, и в некоторых случаях необходимо осуществить переход от одноного типа моделей к другому.

Рассмотрим модель с распределенным лагом в предположении, что максимальная величина лага конечна:

$$\hat{y}_t = a + b_0 \cdot x_t + b_1 \cdot x_{t-1} + ... + b_l \cdot x_{t-l} + \varepsilon_t .$$

Даная модель говорит о том, что если в некоторый момент времени t происходит изменение независимой переменной x, то это изменение будет влиять на значения переменной y в течение l следующих моментов времени.

Коэффициент регрессии b_0 при переменной x_t характеризует среднее абсолютное изменение y_t при изменении x_t на 1 ед. своего измерения в некоторый фиксированный момент времени t, без учета воздействия лаговых значений фактора x. Этот коэффициент называется *краткосрочным мультипликатором.*

В момент $t+1$ воздействие факторной переменной x_t на результат y_t составит (b_0+b_1) условных единиц; в момент времени $t+2$ это воздействие можно охарактеризовать суммой $(b_0+b_1+b_2)$ и т.д. Полученные таким образом суммы называются *промежуточными мультипликаторами*.

С учетом конечной величины лага можно сказать, что изменение переменной x_t в момент времени t на 1 условную единицу приведет к общему изменению результата через l моментов времени $(b_0+b_1+b_2+...+b_l)$.

Введем следующее обозначение: $b=(b_0+b_1+b_2+...+b_l)$. Величину b называется *долгосрочным мультипликатором*, который показывает абсолютное изменение в долгосрочном периоде $t+l$ результата y под влиянием изменения на 1 ед. фактора x.

Величины $\beta_j = b_j / b, \quad j = 0,l$ называются *относительными коэффициентами* модели с распределенным лагом. Если все коэффициенты b_j имеют одинаковые знаки, то $\sum_{j=0}^{l} \beta_j = 1$. Относительные коэффициенты β_j являются весами для соответствующих коэффициентов b_j. Каждый из них измеряет долю общего изменения результативного признака в момент времени $t+j$.

Зная величины β_j, с помощью стандартных формул можно определить еще две важные характеристики модели множественной регрессии: величину среднего и медианного лагов.

Средний лаг рассчитывается по формуле средней арифметической взвешенной:

$$\bar{L} = \sum_{j=0}^{L} j \cdot \beta_j$$

и представляет собой средний период, в течение которого будет происходить изменение результата под воздействием изменения фактора x в момент t. Если значение среднего лага небольшое, то это говорит о довольно быстром реагировании *y* на изменение *x*. Высокое значение среднего лага говорит о том, что воздействие фактора на результат будет сказываться в течение длительного периода времени.

Медианный лаг (L_{Me}) – это величина лага, для которого период, в течение которого $\sum_{j=0}^{l_{Me}} \beta_j \approx 0{,}5$. Это тот период времени, в течение которого с момента времени *t* будет реализована половина общего воздействия фактора на результат.

Изложенные выше приемы анализа параметров модели с распределенным лагом действительны только в предположении, что все коэффициенты при текущем и лаговых значениях исследуемого фактора имеют одинаковые знаки. Это предположение вполне оправдано с экономической точки зрения: воздействие одного и того же фактора на результат должно быть однонаправленным независимо от того, с каким временным лагом измеряется сила или теснота связи между этими признаками. Однако на практике получить статистически значимую модель, параметры которой имели бы одинаковые знаки, особенно при большой величине лага *l*, чрезвычайно сложно.

Применение обычного МНК к таким моделям в большинстве случаев затруднительно по следующим причинам:

- текущие и лаговые значения независимой переменной, как правило, тесно связаны друг с другом, тем самым оценка параметров модели проводится в условиях высокой мультиколлинеарности;

- при большой величине лага снижается число наблюдений, по которому строится модель, и увеличивается число ее факторных признаков, что ведет к потере числа степеней свободы в модели;

- в моделях с распределенным лагом часто возникает проблема автокорреляции остатков.

Рассмотрим модель авторегрессии например: $y_t = a + b_0 * x_t + c_1 * y_{t-1} + \varepsilon_t$.

Как и в модели с распределенным лагом, b_0 в этой модели характеризует краткосрочное изменение y_t под воздействием изменения x_t на 1 ед. Однако промежуточные и долгосрочный мультипликаторы в модели авторегрессии несколько иные. К моменту времени $t+1$ результат y_t изменился под воздействием изменения изучаемого фактора в момент времени t на b_0 единиц, а y_{t+1} – под воздействием своего изменения в непосредственно предшествующим момент времени на c_1 единиц. Таким образом, общее абсолютное изменение результата в момент $t+1$ составит b_0c_1. Аналогично в момент времени $t+2$ абсолютное изменение результата составит $b_0c_1^2$ единиц и т.д. Следовательно, долгосрочный мультипликатор в модели авторегрессии можно рассчитать как сумму краткосрочного и промежуточного мультипликаторов:

$$b = b_0 + b_0 * c_1 + b_0 * c_1^2 + b_0 * c_1^3 + ... = b_0 * (1 + c_1 + c_1^2 + c_1^3 + ...) = \frac{b_0}{1 - c_1},$$

где $|c_1|\langle 1$.

Такая интерпретация коэффициентов модели авторегрессии и расчет долгосрочного мультипликатора основаны на предпосылке о наличии бесконечного лага в воздействии текущего значения зависимой переменной на ее будущие значения.

Пример. Предположим, по данным о динамике показателей потребления и дохода в регионе была получена модель авторегрессии, описывающая зависимость среднедушевого объема потребления за год (С, тыс. руб.) от среднедушевого совокупного годового дохода (Y, тыс. руб.) и объема потребления предшествующего года:

$$\hat{C}_t = 3 + 0{,}85 \cdot Y_t + 0{,}1 \cdot C_{t-1}.$$

Краткосрочный мультипликатор равен 0,85. В этой модели он представляет собой предельную склонность к потреблению в краткосрочном периоде. Следовательно, увеличение среднедушевого совокупного дохода на 1 тыс. руб. приводит к росту объема потребления в тот же год в среднем на 850 руб. Долгосрочную предельную склонность к потреблению в данной модели можно определить как

$$b = \frac{b_0}{1 - c_1} = \frac{0{,}5}{1 - 0{,}1} = 0{,}944.$$

В долгосрочной перспективе рост среднедушевого совокупного дохода на 1 тыс. руб. приведет к росту объема потребления в среднем на 944 руб. Промежуточные показатели предельной склонности к потреблению можно определить, рассчитав необходимые частные суммы за соответствующие периоды времени. Например, для момента времени *t+1* получим:

$$0{,}85 + 0{,}85 \cdot 0{,}1 = 0{,}935.$$

Это означает, что увеличение среднедушевого совокупного дохода в текущем периоде на 1 тыс. руб. ведет к увеличению объема потребления в среднем на 935 руб. в ближайшем следующем периоде.