Лекция 5
Автокорреляция

1. Автокорреляция и её последствия
2. Обнаружение автокор-реляции
3. Оценка коэффициен-тов при автокорреляции

1. Автокорреляция и её последствия

В классической модели считается, что выполняется предпосылка 4° МНК и значение случайной величины ε_i не зависит от значений возмущений ε_j в других наблюдениях, т.е. $M(\varepsilon_i \cdot \varepsilon_j) = 0$ при $i \neq j$.

В практических задачах это условие может не выполняться.

Автокорреляция определяется как корреляционная зависимость между значениями одного показателя, упорядоченными в пространстве или во времени (временные ряды).

Автокорреляция случайной составляющей ε_i - это корреляционная зависимость ε_i со значениями этой же составляющей ε в других наблюдениях ε_{i-l}. Величину l (величину сдвига) во *временных* рядах называют *лагом.* При $l=1$ речь идёт о соседних наблюдениях.

Различают *положительную* и *отрицательную* автокорреляцию.

В качестве примера проанализируем модель зависимости спроса на мороженное y (по ежемесячным данным) от доходов x в предположении, что состояние погоды будет единственным фактором, "скрытым" в переменной ε .

Очевидно, что будет несколько последовательных наблюдений, когда теплая погода способствует увеличению спроса на мороженое и, следовательно, значения ε будут положительные.

После этого будут несколько последовательных наблюдений, когда при холодной погоде ситуация будет складываться противоположным образом (рис. 1).

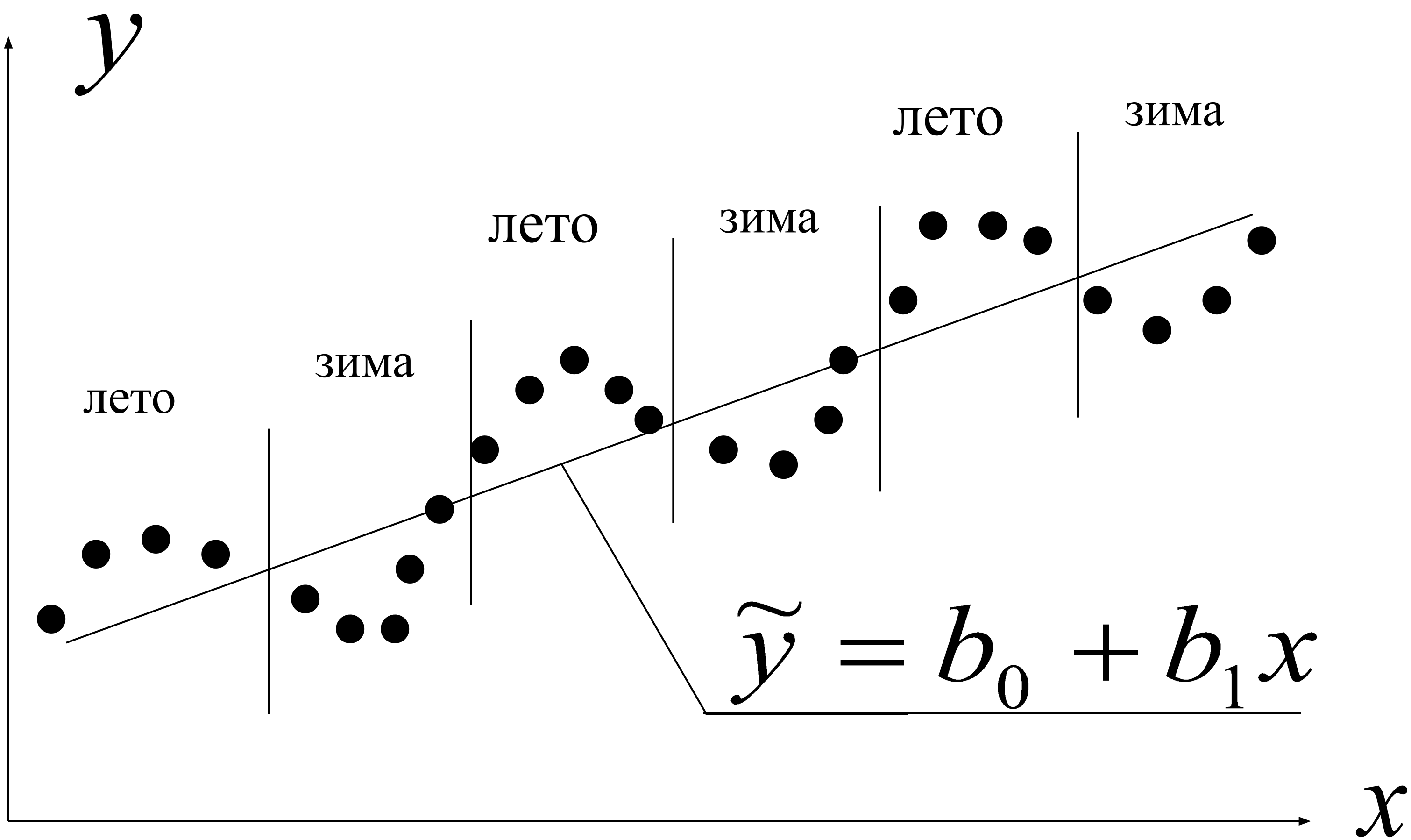


Рис. 1

Из рисунка видно, что имеются зоны, где наблюдаемые значения y_i оказываются выше объясненного значения $\tilde{y}_i$ (лето) и зоны, где они располагаются ниже линейной линии регрессии (зима).

Такие чередующиеся зоны графически выражают *положительную* автокорреляцию.

Отрицательная автокорреляция встречается в тех случаях, когда последовательные наблюдения действуют друг на друга по принципу "маятника" - завышенные значения y_i по сравнению с $\tilde{y}_i$ в предыдущем наблюдении приводят к занижению их в последующем наблюдении и наоборот (рис. 2).

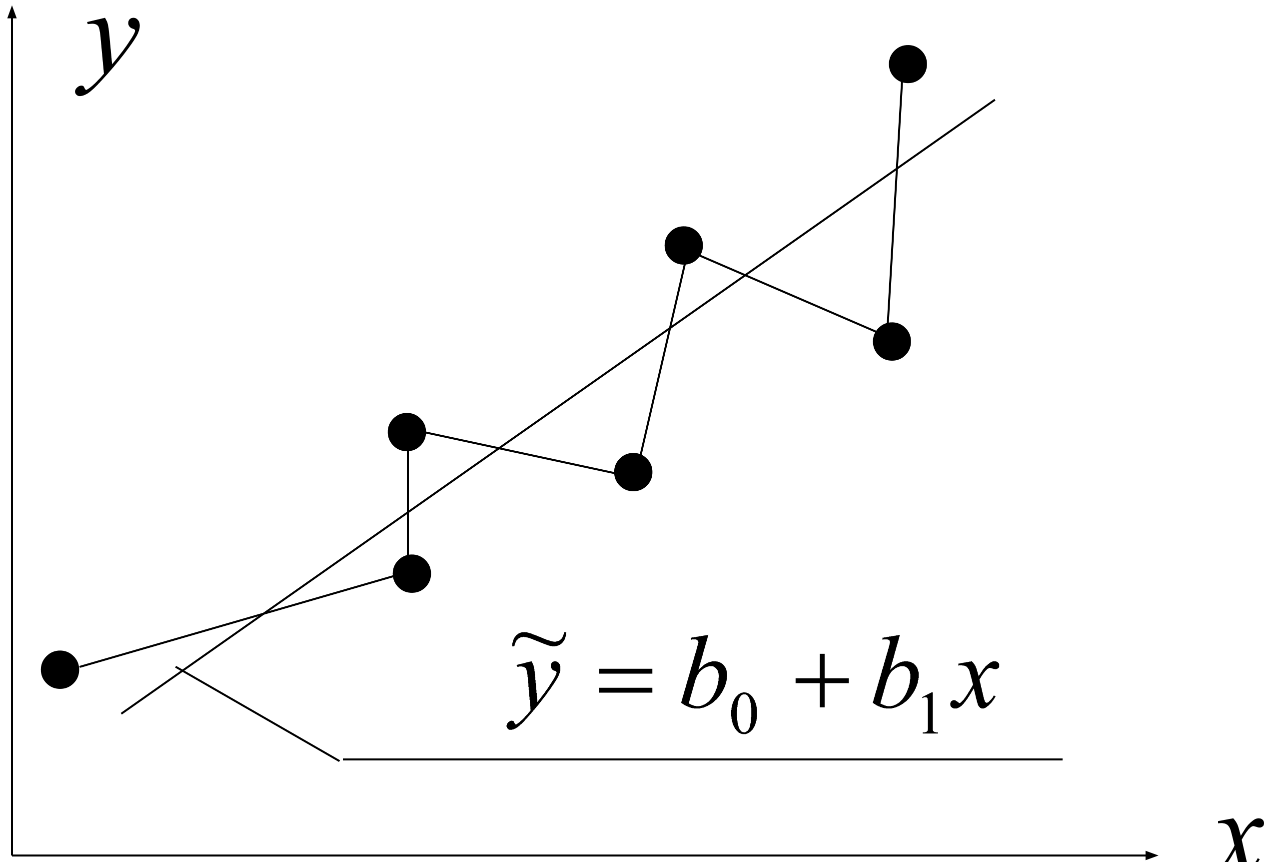


Рис. 2

В экономике отрицательная корреляция встречается достаточно редко.

Последствия автокорреляции в некоторой степени сходны с последствиями гетеро-скедастичности: оценки b_j коэффициентов модели перестают быть эффективными, значения t –статистик – завышенными, дисперсии оценок $D(b_j)$ являются смещенными.

2. Обнаружение автокорреляции

Для обнаружения автокорреляции в первую очередь можно использовать наиболее простой **графический** способ. Оценкой составляющей ε_i является остаток $e_i = y_i - \tilde{y}_i$.

Отсюда если корреляция ошибок регрессии не равна нулю, то она присутствует и в остатках регрессии.

В соответствии с предпосылками МНК остатки e_i должны быть случайными, что на графике координатной плоскости $i0e$ (i –номер наблюдения) выглядит "облаком, проткнутым " осью абсцисс i (рис. 3).

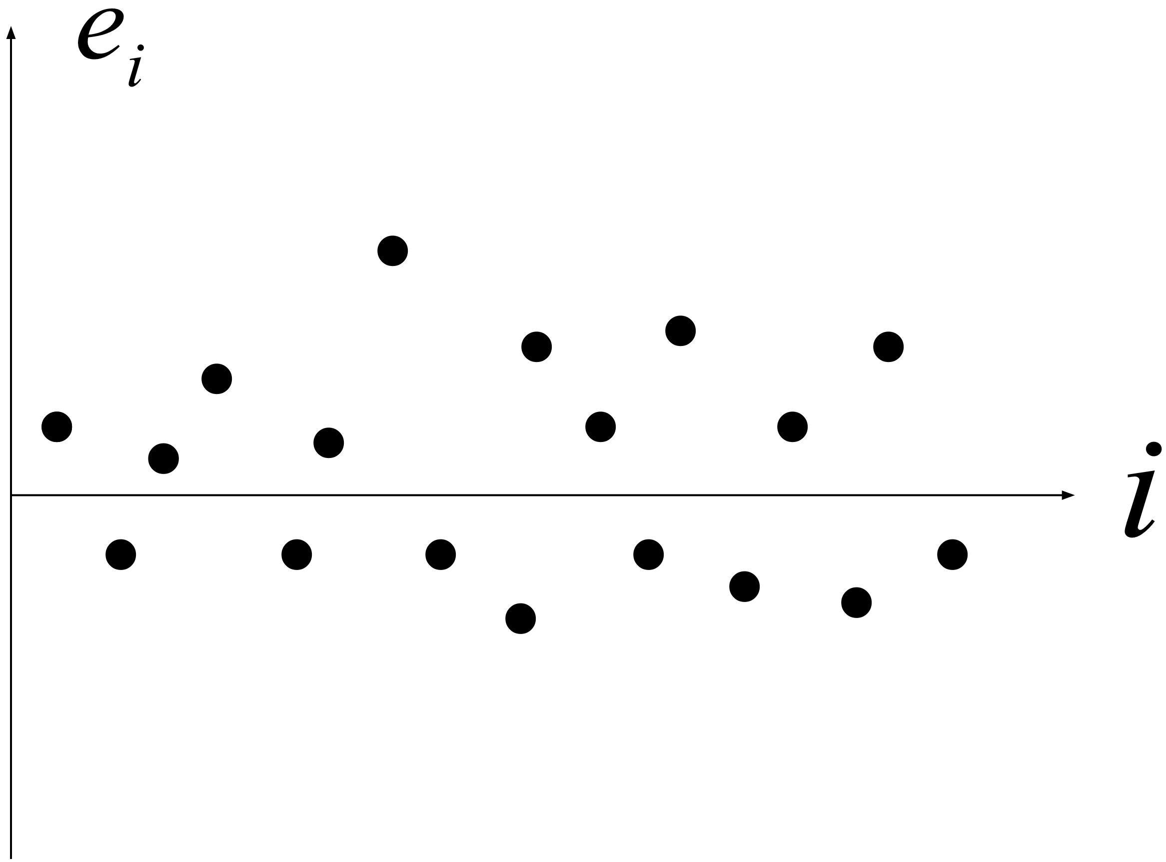


Рис. 3

Когда остатки содержат тенденцию (возрастающую (рис. 4) или убывающую) или циклические колебания (рис. 5), то это говорит о наличии автокорреляции.

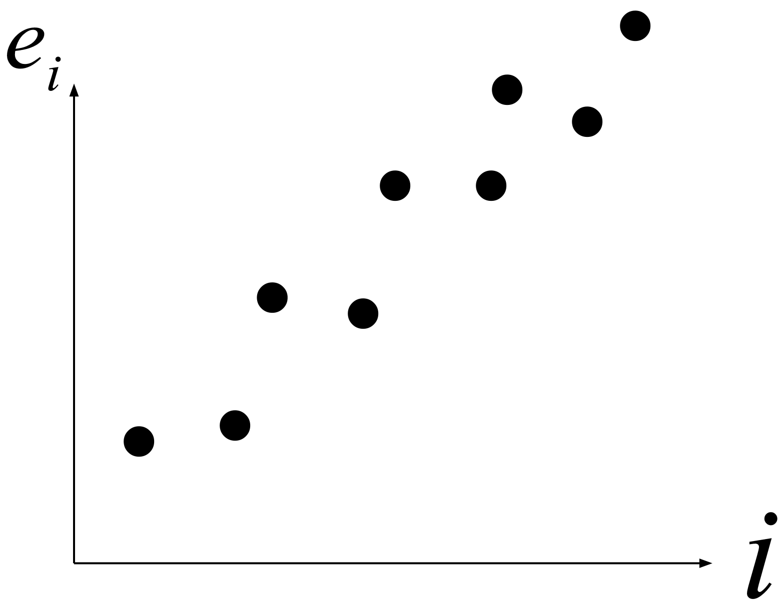


Рис. 4

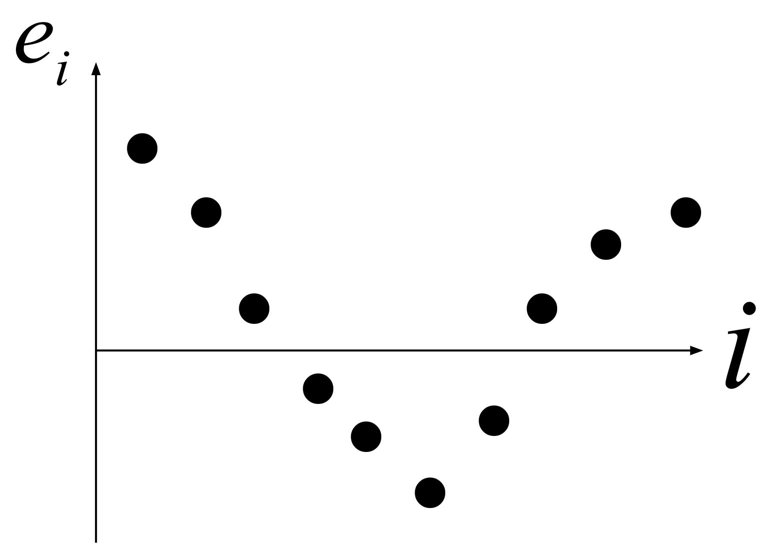


Рис. 5

Для обнаружения автокорреляции используют также статистические тесты. Наиболее распространенным является критерий *Дарбина-Уотсона.* Этот тест используется для обнаружения автокорреляции первого порядка, когда автокорреляция подчиняется уравнению

$$\varepsilon_i = \rho\varepsilon_{i-1} + \zeta_i \qquad (1)$$

где ζ_i – случайная составляющая, удовлетворяющая предпосылкам МНК.

Это означает, что величина случайного возмущения в любом наблюдении равна его значению в предыдущем наблюдении, умноженному на ρ, плюс новое возмущение ζ_i. Данная схема называется *авторегрессией*, поскольку определяется значением той же самой величины ε_i с запаздыванием. Так как запаздывание здесь равно 1, то уравнение (1) *называют авторегрессией 1-го порядка.*

Если $\rho > 0$, то автокорреляция положительна, а при $\rho < 0$ - отрицательна. При $\rho = 0$ автокорреляция отсутствует.

Критерий Дарбина-Уотсона сводится к проверке гипотезы $H_0 : \rho = 0$.

Для проверки гипотезы используется статистика

$$d = \frac{\sum_{i=2}^{n} (e_i - e_{i-1})^2}{\sum_{i=1}^{n} e_i^2}, \qquad (2)$$

которую так и называют *статистикой Дарбина-Уотсона.*

Нетрудно показать, что для больших выборок справедливо равенство

$$d \cong 2(1 - r_1), \qquad (3)$$

где r_1 – выборочный коэффициент корреляции между соседними возмущениями, определяемый по формуле

$$r_1 = \frac{\sum_{i=2}^{n} e_i e_{i-1}}{\sum_{i=1}^{n} e_i^2}.$$

Если корреляция отсутствует, то $r_1 = 0$, и согласно выражению (3) величина d должна быть близка к 2. При $r_1 \to 1$ величина $d \to 0$ и это означает положи-тельную автокорреляцию. Если же $r_1 \to -1$, то $d \to 4$ и это говорит об отрицательной автокорреляции.

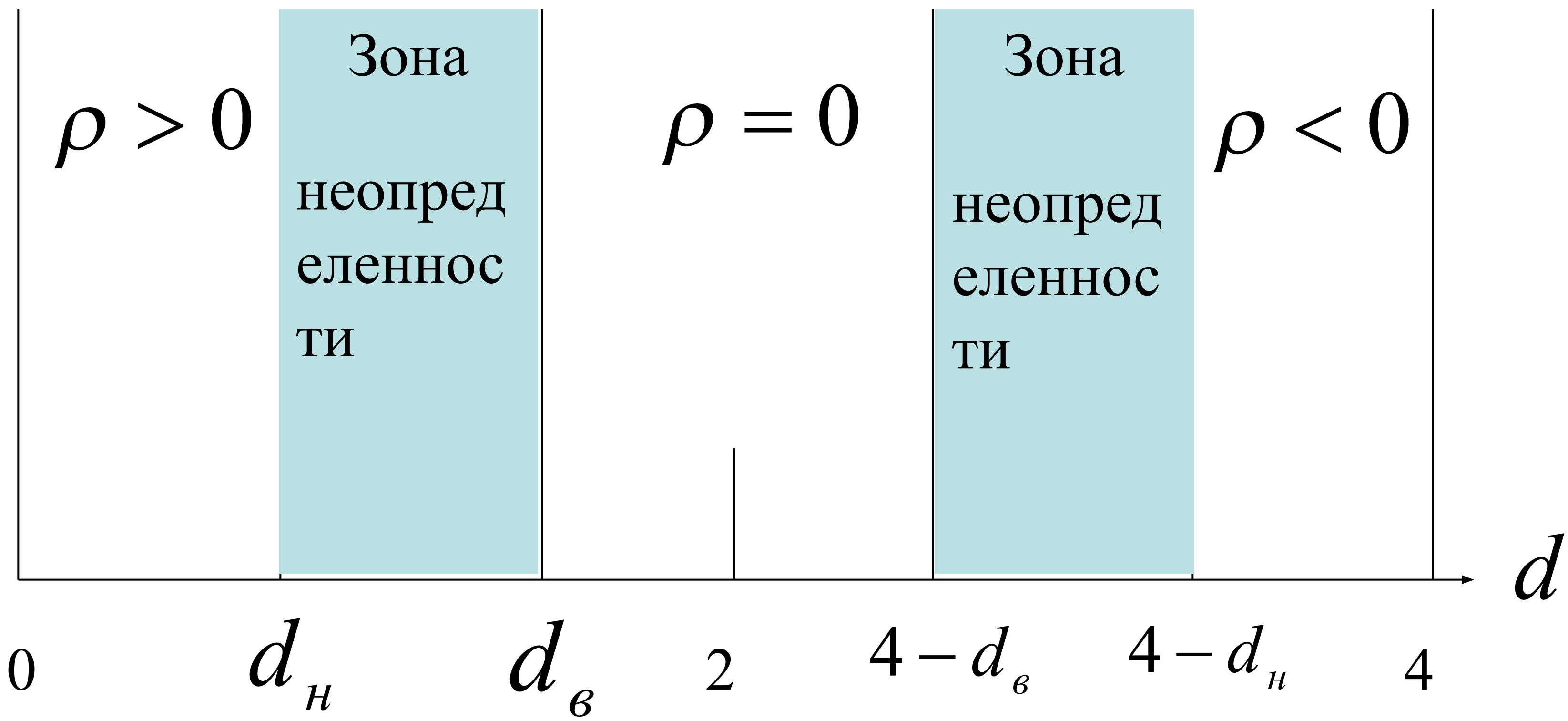
Критическое значение статистики d зависит от числа p факторов модели, от объема выборки n и, к сожалению, от конкретных значений объясняющих переменных. Поэтому невозможно составить таблицу критических точек статистики d, как это можно было сделать для $t-$ и $F-$статистик.

Но можно вычислить критическую *верхнюю* $d_в$ и критическую *нижнюю* $d_н$ границы для критерия d, которые зависят только от объёма выборки n, числа p факторов модели и уровня значимости α.

В итоге алгоритм применения теста Дарбина-Уотсона следующий:

- выдвигается гипотеза H_0 об отсутствии автокорреляции;

- по формуле (2) или (3) находится фактическое значение d ;
- по специальным таблицам Дарбина-Уотсона определяются значения $d_в$ и $d_н$ по известным числам n , p и α ;
- по этим значениям числовой промежуток $[0,4]$ изменения d разбивается на пять интервалов (рис. 6):



Зона
неопределенности
Зона
неопределенности
$\rho > 0$
$\rho = 0$
$\rho < 0$
d
0
$d_н$
$d_в$
2
$4 - d_в$
$4 - d_н$
4

Рис. 6

- в зависимости от того, в какой интервал попадает фактическое значение d делается вывод об автокорреляции. Если d попадает в зону неопределенности, то тест не даёт ответа об автокорреляции.

Описанный тест обладает следующими недостатками:

- тест можно использовать только для тех моделей, которые имеют свободный член;

- тест проверяет только автокорреляцию первого порядка;
- тест даёт достоверные результаты только при больших выборках;
- наличие в тесте зон неопределенности.

3. Оценка коэффициентов при автокорреляции

Рассмотрим основной подход к оценке параметров регрессии для случая, когда имеется автокорреляция на примере парной линейной регрессии вида

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i. \quad (4)$$

Для предыдущего наблюдения модель запишется

$$y_{i-1} = \beta_0 + \beta_1 x_{i-1} + \varepsilon_{i-1}. \quad (5)$$

Будем считать автокорреляцию первого порядка (1) и коэффициент корреляции ρ известным.

Умножим на него уравнение (5) и вычтем результат почленно из уравнения (4)

$$y_i - \rho \cdot y_{i-1} = (1-\rho)\beta_0 + \beta_1(x_i - \rho \cdot x_{i-1}) + (\varepsilon_i - \rho \cdot \varepsilon_{i-1})$$

Сделав замену переменных

$$y_i' = y_i - \rho \cdot y_{i-1}, \quad x_i' = x_i - \rho \cdot x_{i-1}, \quad (6)$$

$$\beta_0' = \beta_0(1 - \rho), \quad \varepsilon_i' = \varepsilon_i - \rho \cdot \varepsilon_{i-1},$$

получим в силу (1)

$$y_i' = \beta_0' + \beta_1 x_i' + \varsigma_i \qquad (7)$$

где ζ_i – случайная составляющая, удовлетворяющая предпосылкам МНК. Поэтому оценка параметров уравнения (7) может быть выполнена обычным МНК

Отсюда схема метода:

- преобразовать исходные переменные по формулам (6) для $i = \overline{2, n}$;
- применив обычный МНК к уравнению (7), определить оценки b_0', b_1 коэффициентов этого уравнения;
- вычислить оценку b_0 по формуле

$$b_0 = \frac{b_0'}{1 - \rho};$$

- записать оценку исходного уравнения

$$\tilde{y} = b_0 + b_1 x.$$

Видно, что такой способ преобразования переменных приводит к потере первого наблюдения. Это уменьшает на единицу число степеней свободы и при малых объёмах представляет чувствительную потерю. В этих случаях первое наблюдение восстанавливают с помощью *поправки Прайса-Уинстона*:

$$x_1' = \sqrt{1-\rho^2}\,x_1, \quad y_1' = \sqrt{1-\rho^2}\,y_1. \quad (8)$$

Основная идея метода основывается на знании ρ. Но на практике этот коэффициент неизвестен. Существуют различные методы его оценки. Например, оценку можно найти с использованием статистики Дарбина-Уотсона

$$r = 1 - d/2.$$

Другой подход, называемый методом *Кохрана-Оркатта*, представляет следующий итерационный процесс:

1. Оценивается регрессия (4) с исход-ными *непреобразованными* данными.

2. Вычисляются остатки $e_i = y_i - \tilde{y}_i$

3. Оценивается регрессионная зависимость e_i от e_{i-1}, соответствующая формуле

$$e_i = r \cdot e_{i-1} + \xi_i$$

где r – оценка ρ.

4. С найденным значением r уравнение (4) преобразуется в уравнение (7), оценивание которого позволяет получить пересмотренные оценки b_0', b_1 коэффициентов $\beta_0', \beta_1,$ а затем вычислить и b_0, b_1.

5. Повторно вычисляются остатки e_i и процесс возвращается к пункту 3 алгоритма. Чередование этапов пересмотра оценок b_0, b_1 и оценки r продолжается до тех пор, пока оценки на последнем и предпоследнем цикле не совпадут с заданной степенью точности.