Оценка качества регрессии. Автокорреляция остаточной компоненты модели

ПРОБЛЕМА АВТОКОРРЕЛЯЦИИ ОСТАТКОВ РЕГРЕССИОННОЙ МОДЕЛИ

Автокорреляция – взаимосвязь между последовательными уровнями временного ряда.

Виды автокорреляции:

- автокорреляция между уровнями ряда;
- автокорреляция в отклонениях от тренда, или автокорреляция ошибок модели.

Автокорреляция ошибок модели означает корреляционную зависимость ошибок модели в текущий момент времени e_t и ошибок в момент (t – L) : e_{t-L}.

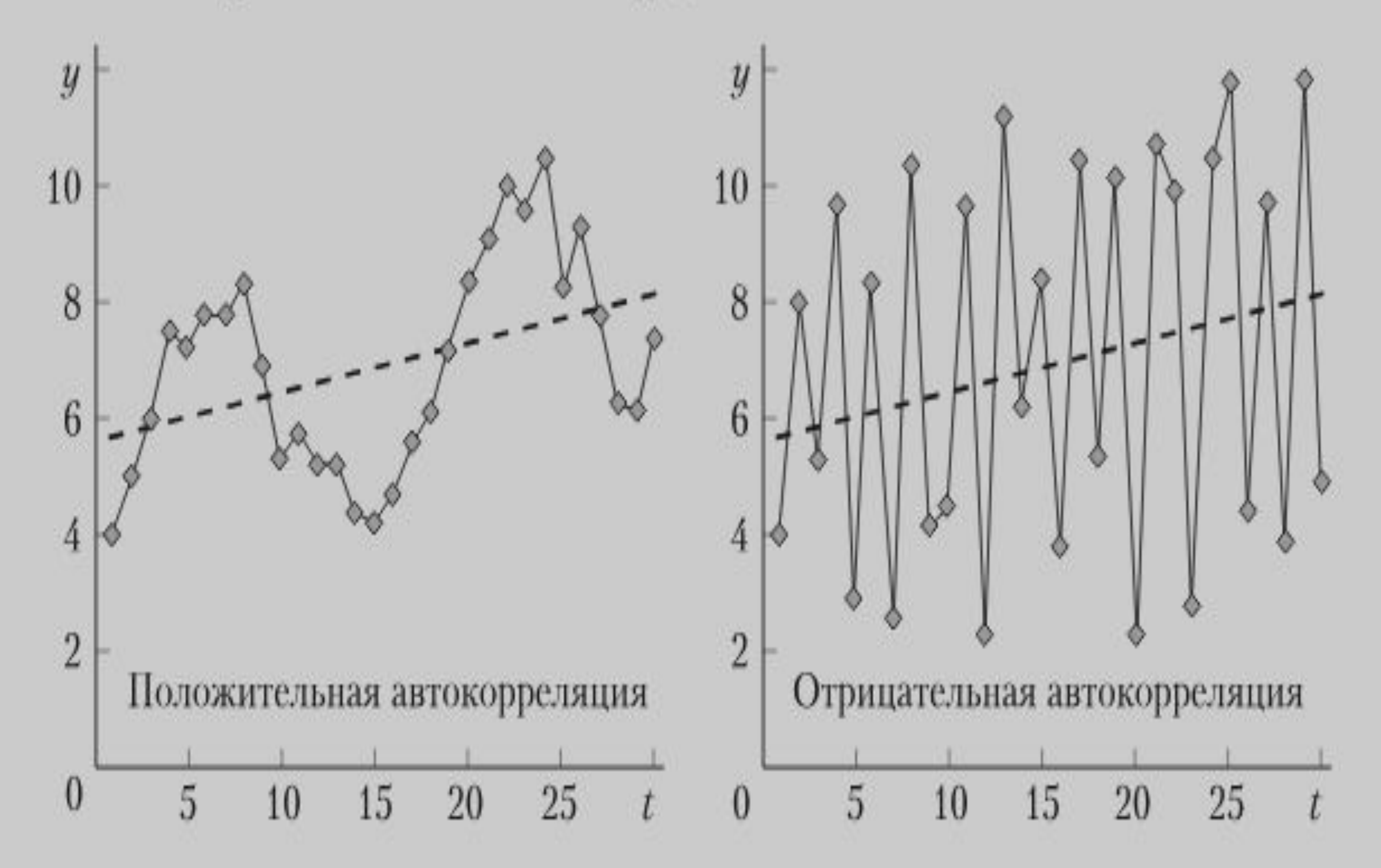
Величина L , называемая запаздыванием или лагом, определяет порядок автокорреляции.

Автокорреляция ошибок возникает, как правило, в случае, когда исходными данными для построения модели являются временные ряды исследуемых показателей.

Основные причины появления автокорреляции остатков модели:

- Неправильная спецификация модели;
- Инерция (запаздывание в изменении экономических показателей;
- Влияние одних экономических показателей на другие;
- Сглаживание данных.

Возможна положительная и отрицательная автокорреляция.
Положительная автокорреляция означает постоянное *однонаправленное* действие неучтенных факторов на результирующий показатель.
Отрицательная автокорреляция означает *разнонаправленное*
действие неучтенных факторов на результирующий показатель.


y
10
8
6
4
2
0
5
10
15
20
25
t
Положительная автокорреляция
Отрицательная автокорреляция

Автокорреляция остатков модели означает нарушение одной из предпосылок классической регрессионной модели.

Последствия автокореляции остатков:

- Оценки параметров модели не будут эффективными (не будут иметь наименьшую возможную дисперсию);
- Значения стандартных ошибок параметров модели могут быть занижены; значение общей дисперсии модели также может быть заниженным;
- Выводы по t - критерию и F- критерию относительно значимости модели и ее параметров могут быть ошибочными.

ВЫЯВЛЕНИЕ АВТОКОРРЕЛЯЦИИ ОСТАТКОВ

Наиболее распространенные способы выявления автокорреляции остатков:

- Анализ графика изменения остаточной компоненты в зависимости от времени (применяется для моделей парной
регрессии)
- Использование критерия Дарбина – Уотсона .

При использовании критерия Дарбина – Уотсона
выч
выб

$$d_{\text{расч}} = \frac{\sum(e_t - e_{t-1})^2}{\sum e_t^2} \approx 2(1 - r_1)$$

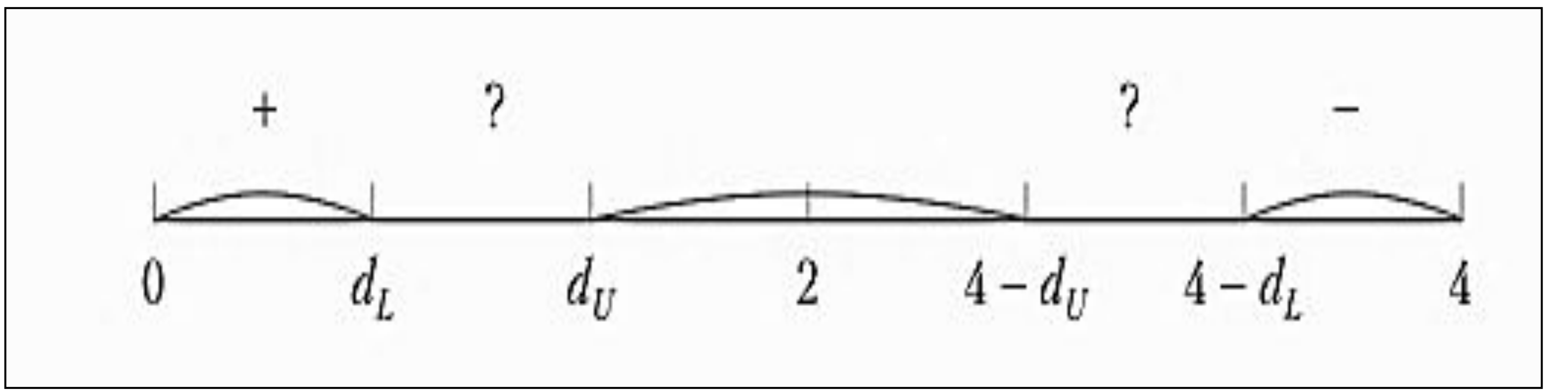
r1 – коэффициент корреляции ошибок в момент времени t и t-1

$$r_1 = \frac{\sum e_t e_{t-1}}{\sum e_t^2}$$

$$0 \leq d_{\text{расч}} \leq 4$$

$$d_{\text{расч}} = \begin{cases} 0 & - \text{положительная автокорреляция} \\ 2 & - \text{автокорреляция отсутствует} \\ 4 & - \text{отрицательная автокорреляция} \end{cases}$$

Критические точки dl и du для наблюдаемой статистики $d_{\text{расч}}$ при данном числе наблюдений и объясняющих переменных и данном уровне значимости заданы таблично.



Критические области статистики Дарбина – Уотсона

(+), (-) – положительная или отрицательная автокорреляция;
? – область неопределенности

Выводы по таблице критических точек Дарбина – Уотсона:

$0 \leq d_{\text{расч}} \leq d_l$ – положительная автокорреляция

$d_l \leq d_{\text{расч}} \leq d_u$ – вывод о наличии автокорреляции не определен

$d_u \leq d_{\text{расч}} \leq 4 - d_u$ – автокорреляция отсутствует

$4 - d_u \leq d_{\text{расч}} \leq 4 - d_l$ – вывод о наличии автокорреляции не определен

$4 - d_l \leq d_{\text{расч}} \leq 4$ – отрицательная автокорреляция

Ограничения при использовании критерия Дарбина – Уотсона:

- Регрессионная модель должна содержать свободный член;
- Случайные отклонения определяются по итерационной схеме:

$$\boldsymbol{e_t = \rho e_{t-1} + v_t}$$

v_t - случайная ошибка, удовлетворяющая предпосылкам МНК;

- Набор независимых переменных не должен содержать лаговых переменных, например y_{t-1}

УСТРАНЕНИЕ АВТОКОРРЕЛЯЦИИ ОСТАТКОВ ПРИ ПОСТРОЕНИИ МОДЕЛИ РЕГРЕССИИ ПО ВРЕМЕННЫМ РЯДАМ

Методы исключения тенденции из уровней временного ряда и построение регрессионной модели по остаточным величинам

А) Метод последовательных разностей

Целесообразно использовать при наличии четко выраженной линейной тенденции во временных рядах исследуемых показателей. Вместо уровней исходных временных рядов Xt, Yt используются цепные абсолютные приросты:

$$\Delta y_t = y_t - y_{t-1};\ \Delta x_t = x_t - x_{t-1}$$

Тогда модель регрессии имеет вид:

$$\Delta y_t = a + b\Delta x_t + u_t$$

u_t - случайная ошибка

Методы исключения тенденции из уровней временного ряда и построение регрессионной модели по остаточным величинам

Б) Метод отклонений от тренда

Временные ряды исследуемых показателей Xt, Yt могут иметь разные тенденции . Для устранения тенденции из уровней исходных временных рядов требуется:

- Для каждого из исходных временных рядов Xt, Yt найти уравнение тренда и по уравнению тренда определить расчетные значения соответствующих показателей: $\hat{y}_{t,} \hat{x}_t$
- По каждому из временных рядов исходных данных найти остаточные величины:

$$dy = y_t - \hat{y}_{t,}\, dx = x_t - \hat{x}_t$$

- Построить модель регрессии:

$$dy = f\,(dx)$$

Непосредственное включение фактора времени в регрессию

Математически доказано, что если временные ряды характеризуются линейной тенденцией, то включение в модель фактора времени t равносильно построению регрессионной модели по отклонениям от трендов с последующим переходом от нее к исходным уровням ряда значений yt.

Уравнение регрессии имеет вид:

$$Y = a + bx + ct$$

Параметр с показывает средний абсолютный прирост зависимой переменной при постоянном уровне независимой переменной.

Производственная функция Кобба – Дугласа также может быть записана с включением фактора времени:

$$Y = aL^{\alpha}K^{\beta}e^{ct}$$

Параметр с – автономный рост объема продукции при неизменных затратах капитала и труда.

Обобщенный МНК при построении модели регрессии по временным рядам

Если имеется модель парной линейной регрессии , построенная по временным рядам,:

$$y = b_0 + b_1 x + e,$$

Для наблюдений в момент t и t-1 справедливы формулы:

$y_t = b_0 + b_1 x_t + e_t$ (1); $y_{t-1} = b_0 + b_1 x_{t-1} + e_{t-1}$ (2)

Пусть случайные ошибки подвержены воздействию автокорреляции первого порядка:

$$e_t = \rho e_{t-1} + v_t$$

Считается, что величина ρ известна.

Вычтем из соотношения (1) соотношение (2), умноженное на ρ:

$$y_t - \rho y_{t-1} = b_0(1-\rho) + b_1(x_t - \rho x_{t-1}) + (e_t - \rho e_{t-1}) \quad (3)$$

Положим:

$$y_t^* = y_t - \rho y_{t-1};\ x_t^* = x_t - \rho x_{t-1}$$

$$b_0^* = b_0(1-\rho)$$

Тогда с учетом соотношения (3) получим модель регрессии:

$$y_t^* = b_0^* + b_1 x_t^* + v_t \quad (4)$$

Ошибки модели удовлетворяют предпосылкам МНК; оценки параметров b_0^* и b_1 будут несмещенными и могут быть получены обычным МНК.

Такое преобразование исходных данных называется авторегрессионной схемой первого прядка. При ее применении теряется первый уровень временного ряда исходных данных.

Для того, чтобы исключить потерю первого наблюдения, применяется поправка Прайса – Уинстена:

$$x_1^* = \sqrt{1-\rho^2}x_1; \quad y_1^* = \sqrt{1-\rho^2}y_1$$

После того, как параметры модели (4) оценены обычным МНК, можно найти свободный член исходного уравнения регрессии с помощью обратного преобразования:

$$b_0 = b_0^*/(1-\rho)$$

ОМНК можно обобщить и на случай множественной регрессии.

На практике величину ρ нельзя точно определить, но можно получить ее оценку разными способами. В этом случае говорят о <u>доступном ОМНК.</u>

Один из распространенных способов оценки величины ρ основан на применении статистики Дарбина – Уотсона:

$$\rho \approx r_1 \approx 1 - \frac{d_{\text{расч}}}{2}$$

В пакетах прикладных статистических программ для оценки величины ρ часто используется процедуры Кохрейна – Оркатта, Хилдрета – Лу и другие, в рамках которых величина ρ оценивается методом последовательных приближений.