Линейные регрессионные модели с гетероскедастичными и автокоррелированными остатками

1. Предпосылки метода наименьших квадратов
2. Гетероскедостичность, выявление и устранение
3. Автокорреляция, выявление и устранение
4. Мультиколлениарность, выявление и устранение
5. Проблемы спецификации модели

Автокорреляция, выявление и устранение

Автокорреляция (последовательная корреляция) определяется как корреляция между наблюдаемыми показателями, упорядоченными во времени (временные ряды) или в пространстве (перекрестные данные).

Основных причин, вызывающих появление автокорреляции

- Ошибки спецификации
- Инерция
- Эффект паутины
- Сглаживание данных

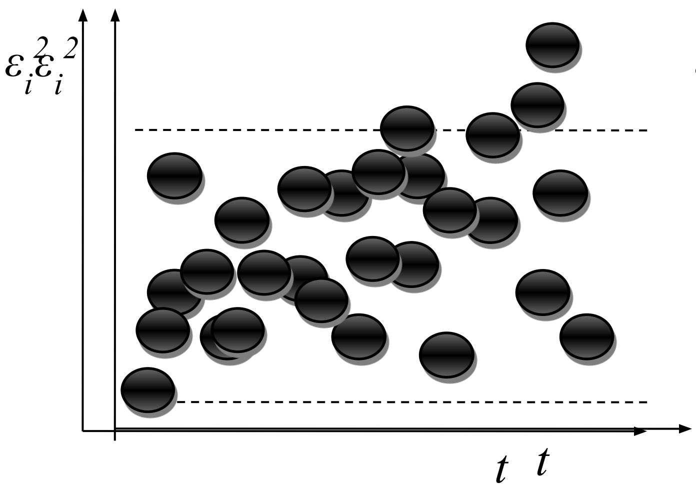
Последствия автокорреляции:

1. Оценки параметров, оставаясь линейными и несмещенными, перестают быть эффективными
2. Дисперсии оценок являются смещенными
3. Оценка дисперсии регрессии σ^2 является смещенной оценкой истинного значения генеральной дисперсии, во многих случаях занижая его
4. Выводы по *t*- и *F*-статистикам, определяющим значимость коэффициентов регрессии и коэффициента детерминации, возможно, будут неверными случаях занижая его

Методы обнаружения автокорреляции:

1. графический анализ остатков
2. метод рядов
3. критерий Дарбина-Уотсона
4. тест серий Бреуша-Годфри
5. Q-тест Льюинга-Бокса
6. тест Льюинга-Бокса

Графический анализа остатков



Возрастающая
тенденция в остатках

Циклические колебания в остатках

наличие автокорреляции

Метод рядов

Последовательно определяются знаки отклонений ε_t

(------) (++++++++) (---) (++++) (-)

6 «-», 8 «+», 3 «-», 4 «+», 1 «-» при T=22

Если при достаточно большом количестве наблюдений (T_1>10, T_2>10) количество рядов k лежит в пределах

$$M(k) - D(k) < k < M(k) + D(k)$$

$$M(k) = \frac{2T_1T_2}{T_1 + T_2} + 1 \qquad D(k) = \frac{2T_1T_2(2T_1T_2 - T_1 - T_2)}{(T_1 + T_2)^2(T_1 + T_2 - 1)}$$

то гипотеза об отсутствии автокорреляции не отклоняется

В нашем случае T_1=12, T_2=10, k=5, $M(k)$=11,91, $D(k)$=5,148 →

$6{,}75<5<17{,}06$

Критерий Дарбина-Уотсона

1 этап: Строят уравнение регрессии и находят отклонения ε_t

2 этап: Рассчитывают величину:

$$DW = \frac{\sum_{t=2}^{T}(\varepsilon_t - \varepsilon_{t-1})^2}{\sum_{t=1}^{T}\varepsilon_t^2}$$

3 этап: Проверяют выполняемость условий:

$DW<d_1$ – в ряду автокорреляция есть;

$DW>d_2$ – в ряду автокорреляции нет;

$d_1<DW<d_2$ – гипотеза о независимости выполняется условно и необходимо дальнейшее исследование границ критерия

Механизм проверки гипотезы о наличии автокорреляции остатков

Есть положительная автокорреляция остатков. H0 отклоняется с вероятностью *P=(1-α)* принимается H1	Зона неопределенности	Нет оснований отклонять H_0 (автокорреляция остатков отсутствует)	Зона неопределенности	Есть отрицательная автокорреляция остатков. H_0 отклоняется с вероятностью *P=(1-α)* принимается H1
0 d_1		d_2 2 $4-d_1$		$4-d_2$ 4

Тест серий Бреуша-Годфри

1 шаг: вычисляем регрессионное уравнение и находим отклонения;

2 шаг: строим уравнение: $\tilde{\varepsilon}_t = \rho\varepsilon_{t-1} + u_t$

3 шаг: на основе *t*-критерия Стьюдента проверяют статистическую значимость параметра ρ.

Если $t_{факт} > t_{табл}$ (параметр статистически значим), то в анализируемом ряду наблюдается автокорреляции.

Авторегрессионная схема первого порядка

Строят парное линейное уравнение регрессии:

$$\tilde{y}_t = a_0 + a_1 x_t + \varepsilon_t$$

Наблюдению с индексом *t* соответствует выражение:

$$y_t = a_0 + a_1 x_t + \varepsilon_t$$

Наблюдению с индексом *t-1* соответствует выражение:

$$y_{t-1} = a_0 + a_1 x_{t-1} + \varepsilon_t$$

Отклонения подвержены воздействию авторегрессии первого порядка:

$$\varepsilon_t = \rho\varepsilon_{t-1} + u_t$$

$$y_t - \rho y_{t-1} = a_0(1-\rho) + a_1(x_t - \rho x_{t-1}) + (\varepsilon_t - \rho\varepsilon_{t-1})$$

Последовательно заменяя

$$y'_t = y_t - \rho y_{t-1} \qquad a'_0 = a_0(1-\rho)$$

$$x'_t = x_t - \rho x_{t-1} \qquad u_t = \varepsilon_t - \rho\varepsilon_{t-1}$$

получим:

$$\tilde{y}'_t = a'_0 + a_1 x'_t + u_t$$

методы оценивания коэффициента ρ

- на основе статистики Дарбина-Уотсона
- на основе метода Кохрана-Орката
- на основе метода Хилдрета-Лу
- на основе метода первых разностей

Метод Кохрана-Орката

1 этап: Оценивается по МНК регрессия и для нее определяются оценки отклонений ε_t;

2 этап: Оценивается регрессионная зависимость: $\varepsilon_t = \rho\varepsilon_{t-1} + u_t$

3 этап: На основе данной оценки строится уравнение:

$$y_t - \rho y_{t-1} = a_0(1-\rho) + a_1(x_t - \rho x_{t-1}) + (\varepsilon_t - \rho\varepsilon_{t-1})$$

4 этап: Значения $a_0' = a_0(1-\rho)$ $a_1' = a_1$ подставляются в уравнение регрессии:

$$\tilde{y}_t = a_0' + a_1' x_t + \varepsilon_t$$

Затем вновь вычисляются оценки ε_t отклонений и возвращаются ко второму этапу.

Определение ρ на основе статистики Дарбина-Уотсона

Определение ρ на основе метода Хилдрета-Лу

$$DW \approx 2(1 - r_{\varepsilon_t \varepsilon_{t-1}})$$

$$r = r_{\varepsilon_t \varepsilon_{t-1}} \Rightarrow r \approx 1 - DW/2$$

$$y_t - \rho y_{t-1} = a_0(1-\rho) + a_1(x_t - \rho x_{t-1}) + (\varepsilon_t - \rho\varepsilon_{t-1})$$

оценивается для каждого возможного значения ρ из отрезка [-1,1]

Метод первых разностей

полагают $\rho = 1$

$$y_t - \rho y_{t-1} = a_0(1 - \rho) + a_1(x_t - \rho x_{t-1}) + (\varepsilon_t - \rho\varepsilon_{t-1})$$

$$y_t - y_{t-1} = a_1(x_t - x_{t-1}) + (\varepsilon_t - \varepsilon_{t-1})$$

Последовательно заменяя

$$\Delta y_t = y_t - y_{t-1} \quad \Delta x_t = x_t - x_{t-1} \quad u_t = \varepsilon_t - \rho\varepsilon_{t-1}$$

$$\Delta y_t = a_1 \Delta x_t + u_t$$