Автокорреляция

- **Автокорреляция** *представляет собой зависимость последовательных элементов временного и (или) пространственного рядов данных.*

$$\operatorname{cov}(u_j, u_l) \neq 0, \quad j \neq l$$

Причины автокорреляции

- временной разрез выборки (инерционность, цикличность многих экономических процессов и явлений)
- неверная спецификация модели (в уравнение не включена существенно влияющая переменная и ее влияние отражается на величинах остатков)

Простейший вид автокорреляции остатков

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

$|\rho| < 1$ –автокорреляционный коэффициент первого порядка

t – период времени

Ошибки ε_t обладают свойствами

$$M(\varepsilon_t) = 0$$

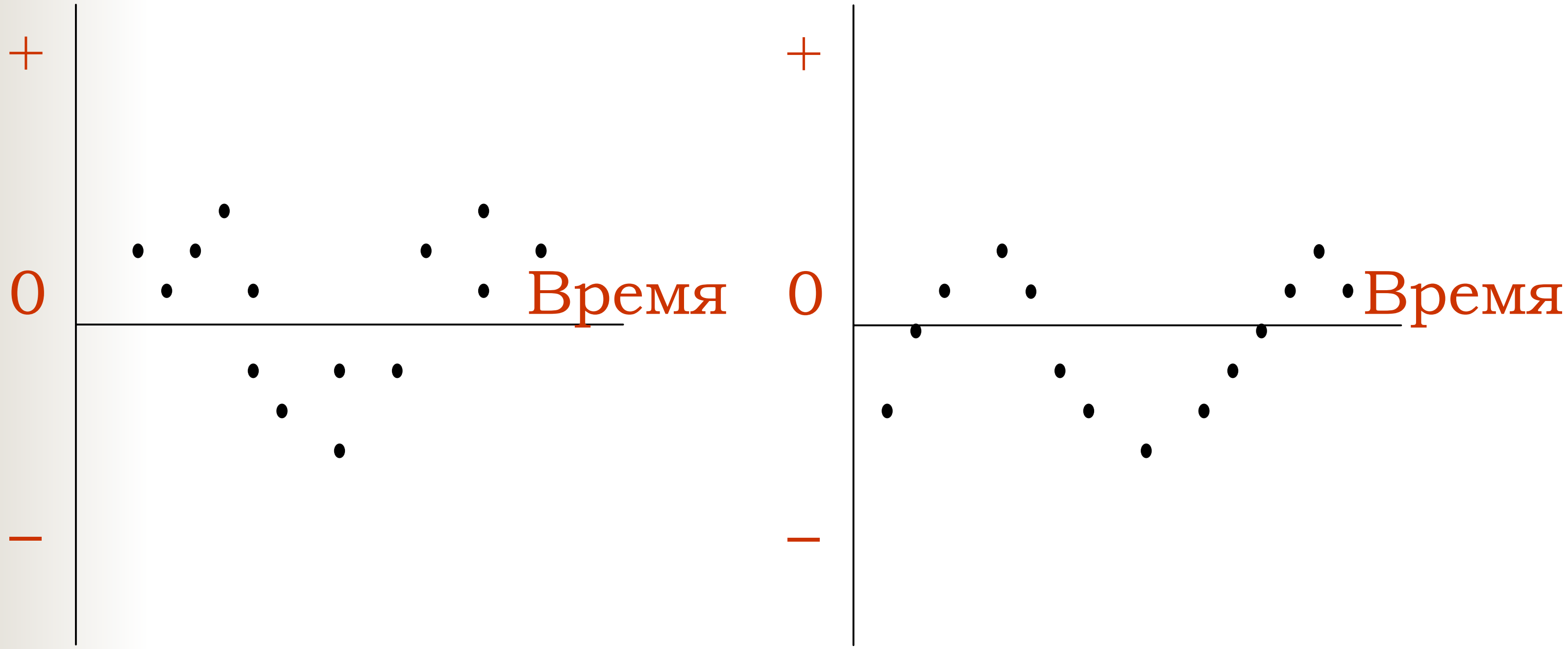
$$\begin{cases} Var(\varepsilon_t) = \sigma_\varepsilon^2 = const \\ \operatorname{cov}(\varepsilon_t, \varepsilon_{t+s}) = 0, \ s \neq 0 \end{cases}$$

Если

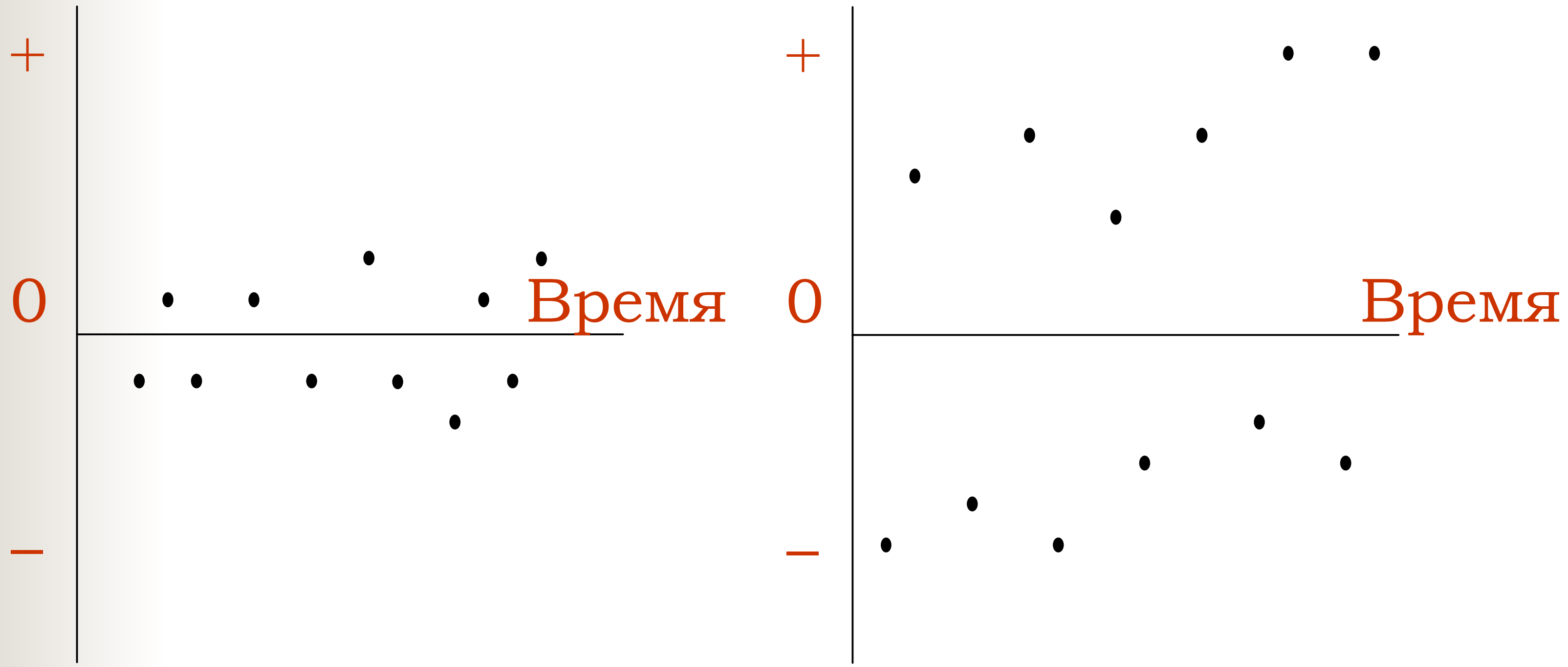
$$\rho = 0,$$

то автокорреляция отсутствует

Автокорреляция может быть положительной и отрицательной (определяется знаком ρ)



Графическая иллюстрация положительной автокорреляции



Графическая иллюстрация отрицательной автокорреляции

Более длительные запаздывания

$$u_t = \rho u_{t-4} + \varepsilon_t$$

Запаздывания, учитывающие несколько периодов времени

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \varepsilon_t$$

(автокорреляции второго порядка)

Преобразуем автокорреляционное уравнение первого порядка

$$u_t = \rho u_{t-1} + \varepsilon_t = \rho\left(\rho u_{t-2} + \varepsilon_{t-1}\right) + \varepsilon_t = \ldots =$$

$$= \varepsilon_t + \rho\varepsilon_{t-1} + \rho^2\varepsilon_{t-2} + \ldots$$

$$\left|\rho^{\infty}\right| \to 0 \qquad \Longrightarrow \qquad \underset{t,r\to\infty}{\rho^r u_{t-r}} = 0$$

$$u_t = \sum_{r=0}^{\infty} \rho^r \varepsilon_{t-r}$$

Рассмотрим дисперсию остатков

$$Var(u_t) = M(u_t^2) = M(\varepsilon_t^2) + \rho^2 M(\varepsilon_{t-1}^2) + \rho^4 M(\varepsilon_{t-2}^2) + \ldots$$

$$Var(u_t) = (1 + \rho^2 + \rho^4 + \ldots)\sigma_\varepsilon^2$$

Так как

$$|\rho| < 1,$$

то

$$\sigma_u^2 = \frac{\sigma_\varepsilon^2}{1 - \rho^2}$$

Ковариация последовательных значений остатков u_t

$$\operatorname{cov}(u_t, u_{t-1}) = \rho\sigma_u^2$$

$$\operatorname{cov}(u_t, u_{t-2}) = \rho^2\sigma_u^2$$

$$\vdots$$

$$\operatorname{cov}(u_t, u_{t-s}) = \rho^s\sigma_u^2$$

Таким образом, при наличии автокорреляции остатков имеет место выражение

$$Var(u) = \sigma_u^2 Z$$

или

$$Var(u) = V$$

$$Var(u) = V = \sigma_u^2 \cdot \begin{pmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{n-1} \\ \rho & 1 & \rho & \cdots & \rho^{n-2} \\ \rho^2 & \rho & 1 & \cdots & \rho^{n-3} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \cdots & 1 \end{pmatrix}$$

Последствия автокорреляции

- Неэффективность оценок параметров модели (выборочные дисперсии вектора оценок могут быть неоправданно большими);

- Невозможность использования статистических критериев Стьюдента и Фишера для проверки значимости параметров модели, так как выборочные дисперсии рассчитываются по неуточненным формулам;

- Неэффективность прогнозов (с большой выборочной дисперсией), получаемых на основе неэффективных оценок параметров.

Критерий Дарбина-Уотсона

- 1-й шаг. Рассчитываем статистику Дарбина-Уотсона

$$DW = \frac{\sum_{t=2}^{n} (u_t - u_{t-1})^2}{\sum_{t=1}^{n} u_t^2}$$

- 2-й шаг. Для заданного уровня значимости α , числа степеней свободы, равного числу факторов, включенных в модель, и числа наблюдений находим значения

$$DW_1 \qquad DW_2$$

– верхнюю и нижнюю границы.

Зоны автокорреляции

При $\rho = 1$

значение критерия $DW = 0$

при $\rho = 0$ $DW = 2$

при $\rho = -1$ $DW = 4$

Пример.

Имеется выборка из 12 наблюдений

Номер наблюдения (магазин)	Зависимая переменная (товарооборот), y	Независимая переменная (торговая площадь), x
1	2,93	0,31
2	5,27	0,98
3	6,85	1,21
4	7,01	1,29
5	7,02	1,12
6	8,35	1,49
7	4,33	0,78
8	5,77	0,94
9	7,68	1,29
10	3,16	0,48
11	1,52	0,24
12	3,15	0,55

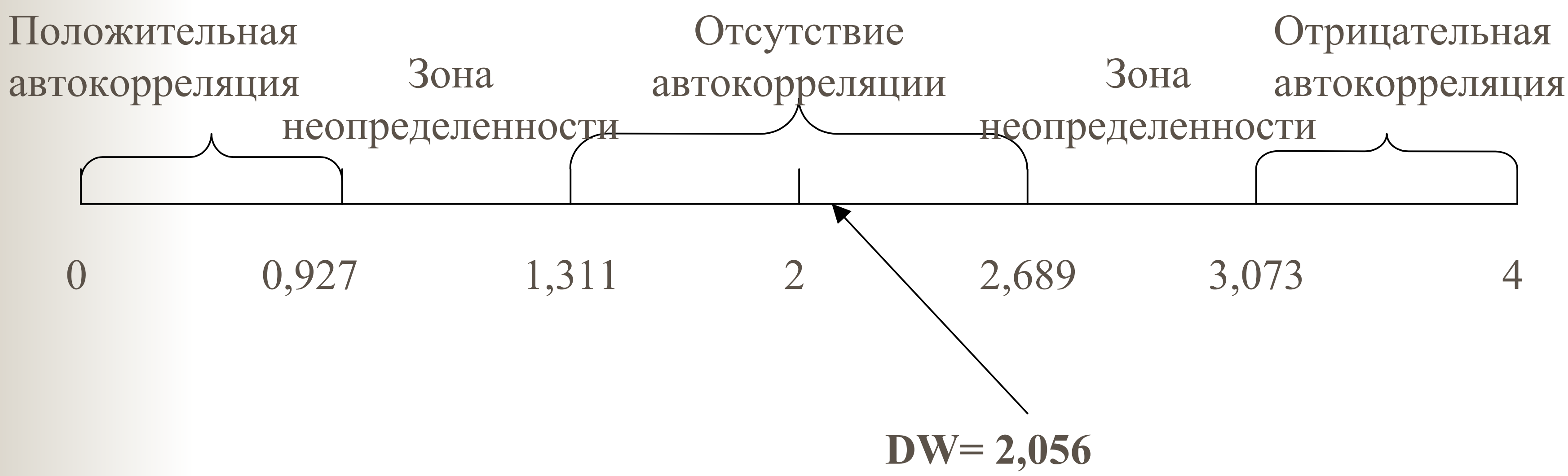
Необходимо оценить наличие автокорреляции остатков

Уравнение имеет вид

$$\widehat{y} = 0{,}6057 + 5{,}222x$$

$$DW = \frac{\sum_{t=2}^{n} \left(u_t - u_{t-1}\right)^2}{\sum_{t=1}^{n} u_t^2} = \frac{3{,}425}{1{,}666} = 2{,}056$$

$$DW_1 = 0{,}927 \qquad DW_2 = 1{,}311$$



Автокорреляция отсутствует

Критерий фон Неймана

$$Q=\frac{\dfrac{\sum_{t=2}^{n}\left(u_t-u_{t-1}\right)^2}{n-1}}{\dfrac{\sum_{t=1}^{n}u_t^2}{n}}$$

Сопоставление с критерием Дарбина-Уотсона

$$\boxed{Q = DW \frac{n}{n-1}}$$

Полученное значение критерия фон Неймана сравнивается с табличным для заданного уровня значимости α и числа наблюдений.

Если $Q_{расч.} < Q_{табл.}$,

то имеет место положительная автокорреляция

Нециклический и циклический коэффициенты автокорреляции

- С помощью нециклического коэффициента устанавливается степень взаимосвязи каждого последующего значения величины остатка с предыдущим, т.е. между двумя рядами чисел

$$u_1, u_2, \ldots, u_{n-1}$$

и

$$u_2, u_3, \ldots, u_n$$

Для этого используется формула

$$r_{нц}=\frac{\sum_{t=2}^{n}u_t u_{t-1}-\frac{1}{n-1}\left(\sum_{t=1}^{n}u_t\right)\left(\sum_{t=2}^{n}u_{t-1}\right)}{\sqrt{\left[\sum_{t=1}^{n}u_t^2-\frac{1}{n-1}\left(\sum_{t=1}^{n}u_t\right)^2\right]\left[\sum_{t=2}^{n}u_{t-1}^2-\frac{1}{n-1}\left(\sum_{t=2}^{n}u_{t-1}\right)^2\right]}}$$

- Данный коэффициент изменяется в пределах от –1 до +1 . Значение коэффициента, близкое к нулю, свидетельствует об отсутствии автокорреляции остатков.

Циклический коэффициент автокорреляции

- Циклический коэффициент устанавливает степень связи между следующими рядами чисел

$$u_1, u_2, \ldots, u_{n-1}, u_n$$

и

$$u_2, u_3, \ldots, u_n, u_1$$

Для этого используется формула

$$r_{ц} = \frac{\sum_{t=2}^{n} u_t u_{t-1} + u_n u_1 - \frac{1}{n}\left(\sum_{t=1}^{n} u_t\right)^2}{\sum_{t=1}^{n} u_t^2 - \frac{1}{n}\left(\sum_{t=1}^{n} u_t\right)^2}$$

- Если имеет место равенство $u_1 = u_n$, то нециклический и циклический коэффициенты автокорреляции совпадают

- Для проверки наличия автокорреляции остатков рассчитанное значение циклического коэффициента сравнивают с табличным при заданном уровне значимости α и числе наблюдений n.

- Если

$$r_{расч.} \geq r_{табл.},$$

то автокорреляция существует

- Если принять следующее допущение, что

$$\sum_{t=1}^{n} u_t \approx \sum_{t=2}^{n} u_{t-1} \approx 0$$

то циклический коэффициент автокорреляции можно рассчитать по формуле

$$\boxed{r_{ц} = \frac{\sum_{t=2}^{n} u_t u_{t-1}}{\sum_{t=1}^{n} u_t^2}}$$

или

$$\boxed{r_{ц} = \frac{\dfrac{1}{n-1}\sum_{t=2}^{n} u_t u_{t-1}}{\dfrac{1}{n}\sum_{t=1}^{n} u_t^2}}$$

Оценка параметров модели с автокоррелированными остатками Метод Эйткена

- Если в эконометрической модели

$$y_t = a_0 + a_1 x_t + u_t$$

остатки обладают свойством автокорреляции и изменяются в соответствии с моделью

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

- Где величины ε_t распределены нормально, то для того чтобы избежать автокорреляции, исходная модель преобразуется к виду, когда ее остатки представлены величинами ε_t Остатки ε_t могут быть выражены через остальные элементы модели

$$\varepsilon_t = u_t - \rho u_{t-1}$$

- Запишем уравнение регрессии для $t-1$ наблюдения

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

- Умножим обе части уравнения на автокорреляционный коэффициент ρ и вычтем из уравнения для периода времени t. В результате получим:

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

К преобразованным данным можно применить метод наименьших квадратов (1МНК).

- Параметр ρ приблизительно равен циклическому коэффициенту автокорреляции с поправкой на величину смещения.
- Применение метода Эйткена предполагает нахождение матриц Z или V

$$Z = \begin{pmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{n-1} \\ \rho & 1 & \rho & \cdots & \rho^{n-2} \\ \rho^2 & \rho & 1 & \cdots & \rho^{n-3} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \cdots & 1 \end{pmatrix}$$

- Коэффициенты ρ^{s} определяют порядок автокорреляции остатков u_t .

- Для автокорреляции нулевого порядка

$$\rho^{0} = 1$$

$$Z^{-1} = \frac{1}{1-\rho^2}\begin{pmatrix} 1 & -\rho & 0 & \cdots & 0 \\ -\rho & 1+\rho^2 & -\rho & \cdots & 0 \\ 0 & -\rho & 1+\rho^2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix}$$

- Для расчета величины ρ циклический коэффициент корректируется на величину смещения

$$r_{ц}^{скор.} = \frac{n}{n-1} \cdot \frac{\sum_{t=2}^{n} u_t u_{t-1}}{\sum_{t=1}^{n} u_t^2} + \frac{m+1}{n}$$

m – число независимых переменных

или

$$\boxed{r_{ц}^{скор.} = r_{ц} + \frac{r_{ц} + \lambda}{n - 1 - \dfrac{1 + r_{ц}\lambda}{1 - r_{ц}\lambda}}}$$

- Параметр λ рассчитывается с помощью формулы

$$\lambda = \frac{\sum_{t=2}^{n} x_t x_{t-1}}{\sum_{t=2}^{n} x_{t-1}^2}$$

x_t – отклонение от своего среднего значения

- Для матрицы $V = \sigma_u^2 Z$ остаточная дисперсия вычисляется как

$$\sigma^2_{ош.} = \frac{u^{/}u}{n-k}$$

- С учетом смещения

$$\sigma^2_{ош.} = \frac{u^{/}u}{n-k}\left[n - \frac{1+\lambda\rho}{1-\lambda\rho}\right]$$

Алгоритм расчета параметров модели методом Эйткена

- <u>1-й шаг</u> Расчет параметров уравнения регрессии методом 1МНК и расчет остатков

- <u>2-й шаг</u> Исследование остатков на наличие автокорреляции

- <u>3-й шаг</u> Формирование матрицы ковариации остатков Z или V
- <u>4-й шаг</u> Обращение матрицы Z или V

- 5-й шаг   Оценка параметров модели методом Эйткена, используя выражения

$$\widehat{A} = \left(X^{/} Z^{-1} X\right)^{-1} X^{/} Z^{-1} Y$$

или

$$\widehat{A} = \left(X^{/} V^{-1} X\right)^{-1} X^{/} V^{-1} Y$$

Метод преобразования исходной информации

- 1-й шаг   Преобразование исходной информации с помощью параметра ρ
- 2-й шаг   Использование метода 1МНК для получения оценок параметров уравнения регрессии

- Матрица T преобразования исходной информации выглядит следующим образом

$$T = \begin{pmatrix} \sqrt{1-\rho^2} & 0 & 0 & \cdots & 0 \\ -\rho & 1 & 0 & \cdots & 0 \\ 0 & -\rho & 1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix}$$

- Уравнение регрессии преобразовывается к виду

$$TY = TX\widehat{A} + Tu$$

- В этом случае дисперсионная матрица будет отвечать необходимым требованиям постоянства дисперсий остатков

$$M\left(Tu u' T'\right) = \sigma_u^2 E$$

Преобразованные данные имеют вид

$$TY = \begin{pmatrix} \sqrt{1-\rho^2}\, y_1 \\ y_2 - \rho y_1 \\ y_3 - \rho y_2 \\ y_4 - \rho y_3 \\ \vdots \\ y_n - \rho y_{n-1} \end{pmatrix}$$

$$TX = \begin{pmatrix} \sqrt{1-\rho^2} & \sqrt{1-\rho^2}\, x_1^1 & \dots & \sqrt{1-\rho^2}\, x_1^m \\ 1-\rho & x_2^1 - \rho x_1^1 & \dots & x_2^m - \rho x_1^m \\ \dots & \dots & \dots & \dots \\ 1-\rho & x_n^1 - \rho x_{n-1}^1 & \dots & x_n^m - \rho x_{n-1}^m \end{pmatrix}$$

- В ряде случаев используется матрица T_1 размерностью $(n-1)\times n$, которая получается из матрицы T путем вычеркивания из нее первой строки

$$T_1 = \begin{pmatrix} -\rho & 1 & 0 & \dots & 0 \\ 0 & -\rho & 1 & \dots & 0 \\ 0 & 0 & -\rho & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$