

Билет 1. Пространственные данные и временные ряды. Модель парной регрессии. Подгонка кривой. Мера отклонения.

Пространственные данные – совокупность экономической информации, полученной за один и тот же период времени.

Временной ряд – совокупность экономической информации, характеризующая объект в разные периоды времени.

Модель парной регрессии

Рассмотрим модели регрессии. В них присутствует объясняемая (зависимая) переменная y и набор объясняющих (независимых) переменных x_1, x_2, \dots, x_n . Связь между ними выражается следующим образом: $y = f(x_1, x_2, \dots, x_n, \beta_1, \beta_2, \dots, \beta_m)$, где f – некоторая функция, а $\beta_1, \beta_2, \dots, \beta_m$ – набор параметров.

Пусть дан набор значений двух переменных X_t и $Y_t, t = 1 \dots n$. Необходимо “подогнать” функцию $Y = f(X)$ из семейства функций вида $y = f(x, \beta)$ таким образом, чтобы она наилучшим образом описывала зависимость Y от X . Подгонка осуществляется выбором параметров β .

Определим следующие меры отклонения полученных значений от реальных значений:

1. $F = \sum_{t=1}^n (Y_t - f(X_t, \beta))^2$ – сумма квадратов отклонений
2. $F = \sum_{t=1}^n |Y_t - f(X_t, \beta)|$ – сумма модулей отклонений
3. $F = \sum_{t=1}^n g(Y_t - f(X_t, \beta))^2$ – где g - мера, с которой отклонение входит в F

Билет 2. Метод наименьших квадратов. Классическая нормальная линейная регрессионная модель с двумя переменными.

Гомоскедастичность, гетероскедастичность, автокорреляция. Теорема Гаусса-Маркова. Оценка дисперсии ошибок. Оценка максимального правдоподобия коэффициентов регрессии.

Метод наименьших квадратов

Пусть $f(x, \beta)$ – линейная функция с параметрами a и b : $Y = f(X, a, b) = a + bX$. Метод наименьших квадратов позволяет подобрать наилучшие параметры путем решения задачи оптимизации:

$$F = \sum_{t=1}^n (Y_t - f(X_t, \beta))^2 \rightarrow \min$$

Записав необходимые условия экстремума, и решив систему уравнений, получаем оптимальные параметры:

$$\hat{a} = \frac{1}{n} \sum Y_t; \quad \hat{b} = \frac{n \sum X_t Y_t - \sum X_t \sum Y_t}{n \sum X_t^2 - (\sum X_t)^2} = \frac{Cov(X, Y)}{Var(X)}$$

Линейная регрессионная модель с двумя переменными

Добавим к постановке задачи статистические свойства данных. Запишем уравнение зависимости Y_t от X_t , добавив в него случайную величину ε_t . Получим регрессионное уравнение:

$$Y_t = a + bX_t + \varepsilon_t, \quad t = 1, \dots, n$$

В этом уравнении X_t – неслучайная детерминированная величина, называется регрессор. Y_t и ε_t – случайные величины.

Случайная ошибка может возникать:

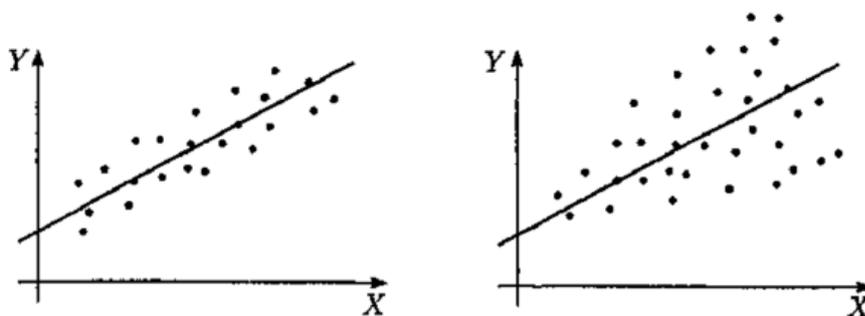
- Если модель является упрощением действительности, и в ней опускаются некоторые параметры.
- Если возможны неточности в измерении данных, используемых в модели.

Можно перечислить основные гипотезы линейной регрессионной модели с двумя переменными:

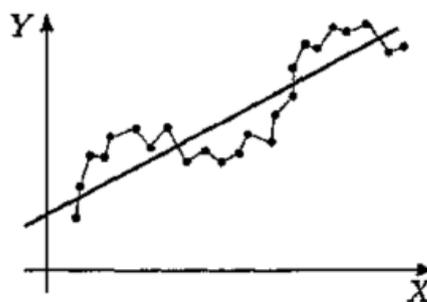
1. $Y_t = a + bX_t + \varepsilon_t, \quad t = 1, \dots, n$
2. X_t – детерминированная величина, причем хотя бы одно из значений вектора от остальных
3. $E\varepsilon_t = 0, Var(\varepsilon_t) = \sigma^2$, не зависит от t
4. Ошибки для разных наблюдений не коррелированы: $E(\varepsilon_s \varepsilon_t) = 0$ при $t \neq s$
5. Зачастую $\varepsilon_t \sim N(0, \sigma^2), t = 1, \dots, n$

Условие независимости дисперсии ошибки от номера наблюдения $(\varepsilon_t)^2 - \text{Var}(\varepsilon_t) = \sigma^2$ называется *гомоскедантностью*. Случай, когда оно не выполнено – *гетероскедантность*.

Примеры гомоскедантности и гетероскедантности ошибок:



Условие 4 зачастую не выполняется для временных рядов и называется *автокорреляцией* ошибок. Автокорреляция выглядит так:



Теорема Гаусса-Маркова

В модели линейной регрессии с двумя параметрами оценки параметров \hat{a} и \hat{b} , полученные методом наименьших квадратов, имеют наименьшую дисперсию в классе всех линейных несмещенных оценок.

Исходя из теоремы, полученные МНК оценки параметров наилучшие для модели. Остается оценить оставшийся параметр σ^2 .

Оценка дисперсии ошибок

Найдем значение регрессионной функции с посчитанными коэффициентами: $\hat{Y}_t = \hat{a} + \hat{b}X_t$. Тогда, остатком регрессии будет называться следующая случайная величина: $e_t = Y_t - \hat{Y}_t$.

Между оценкой σ^2 и суммой квадратов ошибки регрессии e_t существует связь, называемая несмещенной оценкой дисперсии ошибок:

$$s^2 = \hat{\sigma}^2 = \frac{1}{n-2} \sum e_t^2$$

Если дисперсия ошибок σ^2 известна, то значения коэффициентов \hat{a} и \hat{b} можно найти по формулам из теоремы Гаусса-Маркова.

На практике, как правило, σ^2 неизвестна, и, заменив σ^2 на s^2 , можем получить оценки дисперсий коэффициентов \hat{a} и \hat{b} :

$$\widehat{Var}(\hat{b}) = s^2 \frac{1}{\sum (X_t - \bar{X})^2}$$

$$\widehat{Var}(\hat{a}) = s^2 \frac{\sum X_t^2}{n \sum (X_t - \bar{X})^2}$$

Эти формулы используются для нахождения стандартных отклонений оценок коэффициентов регрессии: $s_{\hat{b}} = \sqrt{\widehat{Var}(\hat{b})}$, $s_{\hat{a}} = \sqrt{\widehat{Var}(\hat{a})}$.

3. Статистические свойства МНК оценок параметров регрессии. Проверка гипотезы $H_0: b=b_0$ ($b_0=0, b_0 \neq 0$). Доверительные интервалы для коэффициентов регрессии. Анализ вариации зависимой переменной в регрессии (TSS, ESS, RSS). Коэффициент детерминации R^2 .

Основные гипотезы, лежащие в основе линейной регрессионной модели

1. $Y_t = a + bX_t + \epsilon_t, t = 1, \dots, n$ – спецификация модели;
2. X_t – детерминированная величина; вектор $(X_1, \dots, X_n)'$ не коллинеарен вектору $e = (1, \dots, 1)'$;
3. Остатки $\epsilon_t, i = 1, \dots, n$:
 - 3а. $\mathbb{E}(\epsilon_t \epsilon_s) = 0$ при $t \neq s$;
 - 3б. $\mathbb{E}(\epsilon_t) = 0, \mathbb{E}(\epsilon_t^2) = \mathbb{D}(\epsilon_t) = \sigma^2$ – не зависит от t ;
4. $\epsilon \sim N(0, \sigma^2 I_n)$, т.е. вектор регрессионных остатков ϵ – нормально распределенный случайный вектор со средним 0 и матрицей ковариации $\sigma^2 I_n$ (I_n – единичная матрица размера $n \times n$). В случае выполнения этого условия модель называется **нормальной регрессионной моделью**. Или то же самое, что Y_t имеют совместное нормальное распределение.

Свойства МНК-оценок:

1. МНК-оценки коэффициентов регрессии

МНК-оценки коэффициентов регрессии \hat{a}, \hat{b} имеют совместное нормальное распределение,

$$\hat{a} \sim N\left(a, \sigma^2 \frac{\sum X_t^2}{n \sum x_t^2}\right), \quad \hat{b} \sim N\left(b, \sigma^2 \frac{1}{\sum x_t^2}\right)$$

Если гипотеза нормальности ошибок не выполняется, то при некоторых условиях регулярности на поведение X_t при росте n оценки \hat{a}, \hat{b} имеют асимптотически нормальное распределение, т.е.

$$\hat{a} \rightarrow N\left(a, \sigma^2 \frac{\sum X_t^2}{n \sum x_t^2}\right), \quad \hat{b} \rightarrow N\left(b, \sigma^2 \frac{1}{\sum x_t^2}\right), \quad \text{при } n \rightarrow \infty$$

2. Распределение оценки дисперсии ошибок s^2

$$\frac{(n-2)s^2}{\sigma^2} \sim \chi^2(n-2)$$

3. Независимость s^2 и МНК-оценок \hat{a}, \hat{b}

$$\text{Cov}(\epsilon_t, \hat{a}) = 0$$

$$\text{Cov}(\epsilon_t, \hat{b}) = 0$$

Проверка гипотезы $b = b_0$

Пусть $\sigma_{\hat{a}}^2 = \frac{1}{\sum x_t^2}$, $s_{\hat{a}} = \frac{s\sigma_{\hat{a}}}{\sigma}$, $\sigma_{\hat{b}}^2 = \frac{\sigma^2}{\sum x_t^2}$, $s_{\hat{b}} = \frac{s\sigma_{\hat{b}}}{\sigma}$, тогда

$$t = \frac{\hat{a} - a}{s_{\hat{a}}} \sim t(n - 2), \quad (1)$$

$$t = \frac{\hat{b} - b}{s_{\hat{b}}} \sim t(n - 2) \quad (2)$$

распределены по закону Стьюдента.

Статистику (2) можно использовать для проверки гипотезы $H_0 : b = b_0$ против альтернативной гипотезы $H_1 : b \neq b_0$. Предположим, что верна гипотеза H_0 , тогда

$$t = \frac{\hat{b} - b_0}{s_{\hat{b}}} \sim t(n - 2)$$

Тогда на $100(1 - \alpha)\%$ уровне значимости

$$\mathbb{P}\{-t_{\frac{\alpha}{2}} < t < t_{\frac{\alpha}{2}}\} = 1 - \alpha.$$

Мы отвергаем гипотезу H_0 (и принимаем H_1) на $100(1 - \alpha)\%$ уровне значимости, если $|t| > t_{\frac{\alpha}{2}}$.

Двусторонний $100(1 - \alpha)\%$ -ный доверительный интервал для параметра b имеет следующий вид:

$$[\hat{b} - t_{\frac{\alpha}{2}}(n - 2)s_{\hat{b}}, \hat{b} + t_{\frac{\alpha}{2}}(n - 2)s_{\hat{b}}].$$

Анализ вариации зависимой переменной в регрессии (TSS, ESS, RSS)

Пусть $\hat{Y} = \hat{a} + \hat{b}X_t$, $\bar{Y} = \frac{1}{n} \sum Y_t$, тогда:

$$\text{TSS} = \sum (Y_t - \bar{Y})^2,$$

$$\text{ESS} = \sum (Y_t - \hat{Y})^2,$$

$$\text{RSS} = \sum (\hat{Y} - \bar{Y})^2,$$

$$\text{TSS} = \text{ESS} + \text{RSS}.$$

Коэффициент детерминации R^2

$$R^2 = 1 - \frac{\text{ESS}}{\text{TSS}} = \frac{\text{RSS}}{\text{TSS}}.$$

Если $R^2 = 0$, то это означает, что регрессия ничего не дает, т.е. X_t не улучшает качество предсказания Y_t по сравнению с тривиальным предсказанием $\hat{Y}_t = \bar{Y}$.

Если $R^2 = 1$, то это означает, что $\epsilon_t = 0$, т.е. все наблюдения лежат на регрессионной кривой.

Чем ближе R^2 к 1, тем лучше качество подгонки, \hat{y} более точно аппроксимирует y .

4. Модель множественной регрессии. Метод наименьших квадратов. Теорема Гаусса-Маркова. Статистические свойства МНК-оценок. Оценка дисперсии ошибок. Скорректированный коэффициент детерминации. Проверка статистических гипотез.

Модель множественной регрессии.

$$y_t = \beta_1 + \beta_2 x_{t2} + \dots + \beta_k x_{tk} + \varepsilon_t, t = 1, \dots, n$$

или

$$y_t = \beta_1 x_{t1} + \beta_2 x_{t2} + \dots + \beta_k x_{tk} + \varepsilon_t, t = 1, \dots, n$$

Где x_{tp} – значения регрессора x_p в наблюдении t , а $x_{t1} = 1, t = 1, \dots, n$. С учетом этого замечания различаются модели со свободным членом и без свободного члена.

Основные гипотезы:

1. $y_t = \beta_1 + \beta_2 x_{t2} + \dots + \beta_k x_{tk} + \varepsilon_t, t = 1, \dots, n$ – спецификация модели.
2. x_{t1}, \dots, x_{tk} – детерминированные величины, векторы $x_s = (x_{1s}, \dots, x_{ns})', s = 1, \dots, k$ – линейно независимы в \mathbb{R}^n .
3. $\mathbb{E}\varepsilon_t = 0, \mathbb{E}(\varepsilon_t^2) = V(\varepsilon_t) = \sigma^2$ – не зависит от t .
4. $\mathbb{E}(\varepsilon_t \varepsilon_s) = 0$ при $t \neq s$ – статистическая независимость (некоррелированность) ошибок для разных наблюдений.

Часто добавляется пятое условие:

5. Ошибки $\varepsilon_t, t = 1, \dots, n$ имеют совместное нормальное распределение: $\varepsilon_t \sim N(0, \sigma^2)$.

Метод наименьших квадратов. Теорема Гаусса-Маркова.

Цель - $\hat{\beta}$:

$$e = y - \hat{y} = y - X\hat{\beta}$$

$$ESS = \sum e_t^2 = e'e \rightarrow \min_{\hat{\beta}}$$

$$e'e = y'y - 2\hat{\beta}'X'y + \hat{\beta}'X'X\hat{\beta}$$

Необходимое условие минимума:

$$\frac{\partial ESS}{\partial \hat{\beta}} = -2X'y + 2X'X\hat{\beta} = 0 (*)$$

$$\text{Отсюда: } \hat{\beta}_{OLS} = (X'X)^{-1}X'y$$

Из (*): e ортогонален всем независимым переменным: $X'e = 0$

$$e'e = y'y - \hat{\beta}'X'y$$

Теорема Гаусса-Маркова:

В условиях 1-4 оценки $\hat{\beta}_{OLS}$ оптимальны (обладают наименьшей дисперсией) в классе линейных (по y) несмещенных оценок (матожидание равно истинному значению).

Другой вид:

$$Y_t = a + bX_t + \varepsilon_t, t = 1, \dots, n$$

МНК-оценки:

$$\hat{b} = \frac{n \sum X_t Y_t - (\sum X_t)(\sum Y_t)}{n \sum X_t^2 - (\sum X_t)^2}$$

$$\hat{a} = \frac{1}{n} \sum Y_t - \frac{1}{n} \sum X_t \hat{b}$$

Несмещенная оценка дисперсии ошибок:

$$s^2 = \hat{\sigma}^2 = \frac{1}{n-2} \sum e_t^2$$

Оценки дисперсий \hat{a}, \hat{b} :

$$\hat{V}(\hat{b}) = s^2 \frac{1}{\sum x_t^2} = \frac{s^2}{\sum (X_t - \bar{X})^2}$$

$$\hat{V}(\hat{a}) = s^2 \frac{\sum X_t^2}{n \sum x_t^2} = \frac{s^2 \sum X_t^2}{n \sum (X_t - \bar{X})^2}$$

Скорректированный коэффициент детерминации

При добавлении регрессоров R^2 возрастает.

Скорректированный R^2 :

$$R_{adj}^2 = 1 - \frac{e'e/(n-k)}{y'y/(n-1)}$$

Свойства:

1. $R_{adj}^2 = 1 - (1 - R^2) \frac{(n-1)}{(n-k)}$
2. $R^2 \geq R_{adj}^2, k > 1$
3. $R_{adj}^2 \leq 1$, но может принимать значения < 0 .

Проверка статистических гипотез.

Проверка гипотезы $H_0: \beta_i = \beta_{i0}$.

В рамках нормальной модели $\varepsilon \sim N(0; \sigma^2 I_n)$ регрессор x_i несущественный.

Результаты:

1. $\hat{\beta}_{OLS} \sim N(\beta; \sigma^2 (X'X)^{-1})$
2. Случайная величина $(n-k) \frac{s^2}{\sigma^2} \sim \chi^2(n-k)$
3. Оценки $\hat{\beta}_{OLS}, s^2$ и вектор e независимы.

Отсюда:

$$t = \frac{\hat{\beta}_{OLS,i} - \beta_{i0}}{s_{\hat{\beta}_i}} = \frac{(\hat{\beta}_{OLS,i} - \beta_{i0})/\sigma_{\hat{\beta}_i}}{s_{\hat{\beta}_i}/\sigma_{\hat{\beta}_i}} \sim t(n-k)$$

Где $s_{\hat{\beta}_i}^2 = \sigma_{\hat{\beta}_i}^2 = \hat{\sigma}^2 q^{ii} = s^2 q^{ii}$, где q^{ii} – i -ый диагональный элемент матрицы $(X'X)^{-1}$, $s^2 = \frac{\sum e_t^2}{n-k}$

$[\hat{\beta}_{OLS,i} - t_c s_{\hat{\beta}_i}; \hat{\beta}_{OLS,i} + t_c s_{\hat{\beta}_i}]$ – 95%-доверительный интервал для истинного значения коэффициента β_i , где t_c – двусторонняя 95%-квантиль распределения Стьюдента с $n-k$ степенями свободы.

Нулевая гипотеза $H_0: \beta_i = \beta_{i0}$ отклоняется на 5%-уровне значимости, если:

$$|t| = \left| \frac{\hat{\beta}_{OLS,i} - \beta_i}{s_{\hat{\beta}_i}} \right| > t_c(n - k)$$

Проверка гипотезы $H_0: \beta_2 = \beta_3 = \dots = \beta_k = 0$.

Пусть дана модель регрессии со свободным членом:

$$y_t = \beta_1 + \beta_2 x_{t2} + \dots + \beta_k x_{tk} + \varepsilon_t$$

$$F = \frac{R^2}{1 - R^2} \frac{n - k}{k - 1} = \frac{RSS}{ESS} \frac{n - k}{k - 1} = \frac{\hat{y}'\hat{y}/(k - 1)}{e'e/(n - k)} \sim F(k - 1, n - k)$$

H_0 отвергается на 5%-уровне значимости, если $F > F_c$, где F_c – односторонняя 95%-квантиль распределения Фишера $F(k - 1, n - k)$.

Билет 5. Различные аспекты множественной регрессии. Тест Чоу. Мультиколлинеарность. Фиктивные переменные. Коэффициент частной корреляции. Процедура пошагового отбора переменных. Спецификация модели. Короткая и длинная регрессии. Замещающие переменные.

Модель множественной регрессии

Модель n -мерной регрессии:

$$y_t = \beta_1 + \beta_2 x_{t2} + \dots + \beta_k x_{tk} + \varepsilon_t, \quad t = 1, \dots, n$$

или

$$y_t = \beta_1 x_{t1} + \beta_2 x_{t2} + \dots + \beta_k x_{tk} + \varepsilon_t, \quad t = 1, \dots, n$$

где x_{tp} — значения регрессора x_p в наблюдении t , а $x_{t1} = 1, t = 1, \dots, n$.

Статистики

t-статистику можно использовать для проверки гипотезы $H_0: b = b_0$ против альтернативной гипотезы $H_1: b \neq b_0$. В множественной регрессии как b берем β_i .

Предположим, что верна гипотеза H_0 , и зададимся, например, α -квантилью t -распределения с $(n - 2)$ степенями свободы t_c , т.е. H_0 принимаем, если $|t| \leq t_c$ на заданном уровне значимости $\frac{\alpha}{2}$. Мы отвергаем гипотезу H_0 (и принимаем H_1) на уровне значимости $\frac{\alpha}{2}$, если $|t| > t_c$ («редкое» событие с точки зрения гипотезы H_0).

$$t = \frac{b - b_0}{s_b}, \quad \text{где } s_b = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{n - 2} \frac{\sqrt{\sum x_i^2}}{n\sqrt{DX}}}$$

$t_c(n - 2; \frac{\alpha}{2})$ считаем в Excel или находим здесь: [таблица Стьюдента](#).

Малые значения t -статистики соответствуют отсутствию достоверной статистической связи объясняющей переменной X и зависимой переменной Y .

F-статистику используют для проверки гипотезы H_0 о том, что коэффициенты регрессии одновременно равны нулю, $\beta = (\beta_1, \dots, \beta_k) = 0$. Другими словами, суть расчетов - ответить на вопрос: можно ли анализируемое уравнение регрессии использовать для дальнейшего анализа и прогнозов?

$$F = \frac{R^2}{1 - R^2} \frac{n - k}{k - 1} = \frac{\text{RSS}}{\text{ESS}} \frac{n - k}{k - 1} \sim F(k - 1, n - k)$$
$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2} = 1 - \frac{\text{ESS}}{\text{TSS}} = \frac{\sum (\hat{y}_i - \bar{y})^2}{\sum (y_i - \bar{y})^2} = \frac{\text{RSS}}{\text{TSS}}$$

Гипотеза H_0 отвергается на 5%-уровне значимости, если $F > F_c$, где F_c — односторонняя 95%-квантиль распределения Фишера $F(k - 1, n - k)$. Считаем ее через Excel: ФРАСПОБР(0,95;k-1;n-k).

5.1 Тест Чой (Chow)

Предположим, у нас есть две выборки данных. По каждой выборке мы строим регрессионную модель. Вопрос, который нас интересует: верно ли, что эти две модели совпадают? Рассмотрим модели:

$$y_t = \beta'_1 x_{t1} + \beta'_2 x_{t2} + \dots + \beta'_k x_{tk} + \varepsilon'_t, \quad t = 1, \dots, n, \quad (5.1)$$

$$y_t = \beta''_1 x_{t1} + \beta''_2 x_{t2} + \dots + \beta''_k x_{tk} + \varepsilon''_t, \quad t = n + 1, \dots, n + m, \quad (5.2)$$

в первой выборке n наблюдений и m наблюдений во второй. Например, y — заработная плата, x_t — регрессоры (возраст, стаж, уровень образования и т.п.), и пусть первая выборка относится к женщинам, вторая — к мужчинам. Вопрос: следует ли из оценки моделей (5.1), (5.2), что модель зависимости зарплаты от регрессоров одна и та же для мужчин и женщин?

Сведем эту ситуацию к общей схеме проверки линейных ограничений на параметры модели. Регрессией без ограничений здесь является объединение двух регрессий (5.1), (5.2), т.е. $ESS_{UR} = ESS_1 + ESS_2$, число степеней свободы при этом равно $(n - k) + (m - k) = n + m - 2k$. Предположим теперь, что верна нулевая гипотеза. Тогда регрессия с ограничениями записывается одним уравнением

$$y_t = \beta_1 x_{t1} + \beta_2 x_{t2} + \dots + \beta_k x_{tk} + \varepsilon_t, \quad t = 1, \dots, n + m. \quad (5.3)$$

Оценивая (5.3), получаем ESS_R . Тогда, учитывая, что наложено k ограничений на параметры модели, получаем

$$F = \frac{(ESS_R - ESS_{UR})/k}{ESS_{UR}/(n + m - 2k)} \sim F(k, n + m - 2k). \quad (5.4)$$

Если F -статистика (5.4) больше критического значения F_c , то нулевая гипотеза отвергается. В этом случае мы не можем объединять две выборки в одну.

В противном случае ($F < F_c(k, n + m - 2k)$) \Rightarrow принимается H_0 и можно использовать объединенную регрессию (5.3).

5.2 Мультиколлинеарность

На практике исследователю нередко приходится сталкиваться с ситуацией, когда полученная им регрессия является “плохой”, т.е. t -статистики большинства оценок малы, что свидетельствует о незначимости соответствующих независимых переменных (регрессоров). В то же время F -статистика может быть достаточно большой, что говорит о значимости регрессии в целом. Одна из возможных причин такого явления носит название *мультиколлинеарности* и возникает при наличии высокой корреляции между регрессорами.

Одним из условий классической регрессионной модели является предположение о линейной независимости объясняющих переменных, что означает *линейную независимость столбцов матрицы регрессоров X* или (эквивалентно) что *матрица $(X'X)^{-1}$ имеет полный ранг k* . При нарушении этого условия, т.е. когда один из столбцов матрицы X есть линейная комбинация остальных столбцов, говорят, что имеет место *полная коллинеарность*. В этой ситуации нельзя построить МНК-оценку параметра β , что формально следует из сингулярности матрицы $X'X$ и невозможности решить нормальные уравнения.

Выделим некоторые **наиболее характерные признаки мультиколлинеарности**:

1. Небольшое изменение исходных данных (например, добавление новых наблюдений) приводит к существенному изменению оценок коэффициентов модели.

2. Оценки имеют большие стандартные ошибки, малую значимость, в то время как модель в целом является значимой (высокое значение коэффициента детерминации R^2 и соответствующей F-статистики).
3. Оценки коэффициентов имеют неправильные с точки зрения теории знаки или неоправданно большие значения.

Что можно предпринять в случае мультиколлинеарности?

1. Можно попытаться уменьшить σ_ϵ^2
2. Фактор - число наблюдений
Можно увеличить объем выборки
Можно применить метод группировки. Например, страну можно поделить на области, отдельные города и зоны.
3. Увеличение дисперсии объясняющих переменных.
4. На стадии планирования опроса следует получить такую выборку, в которой объясняющие переменные были бы как можно меньше связаны между собой.
5. Если коррелированные переменные связаны между собой концептуально, то может быть рационально объединить их в совокупный индекс.
6. Можно исключить некоторые из коррелированных переменных если их коэффициенты незначительны. Это может быть опасно, т.к. некоторые из исключаемых коэффициентов могли принадлежать модели и их невключение может служить причиной смещения оценки.
7. Использовать внешнюю информацию относительно коэффициента одной из переменных в уравнении (например, экспертная оценка).

5.3 Фиктивные переменные

Как правило, независимые переменные в регрессионных моделях имеют “непрерывные” области изменения (нац. доход, уровень безработицы, размер зарплаты и т.п.). Однако теория не накладывает никаких ограничений на характер регрессоров, в частности, некоторые переменные могут принимать всего два значения или, в более общей ситуации, дискретное множество значений. Необходимость рассматривать такие переменные возникает довольно часто в тех случаях, когда требуется принимать во внимание какой-либо *качественный признак*.

Следует, ясно понимать, что речь идет о такой же “равноправной” переменной, как и любой из регрессоров $x_j, j = 1, \dots, k$. Ее “фиктивность” состоит только в том, что она количественным образом описывает качественный признак.

Качественное различие можно формализовать с помощью любой переменной, принимающей два значения, а не обязательно значения 0 или 1. Однако на практике почти всегда используют фиктивные переменные типа “0-1”, поскольку в этом случае интерпретация выглядит наиболее просто.

5.4 Частная корреляция

В том случае, когда имеются одна независимая и одна зависимая переменные, естественной мерой зависимости (в рамках линейного подхода) является (выборочный) коэффициент корреляции между ними. Использование множественной регрессии позволяет обобщить это понятие на случай, когда имеется несколько независимых переменных. Корректировка здесь необходима по следующим очевидным соображениям. Высокое значение коэффициента может означать высокую степень зависимости или то, что существует третья переменная, от которой зависят и первая и вторая. Поэтому возникает естественная задача найти “чистую” корреляцию между двумя переменными, исключив (линейное) влияние других факторов. Это можно сделать с помощью коэффициента частной корреляции.

Для простоты предположим, что имеется регрессионная модель

$$y = \beta_0 + x_1\beta_1 + x_2\beta_2 + \epsilon,$$

где, как обычно, y – $n \times 1$ вектор наблюдений зависимой переменной, x_1, x_2 – $n \times 1$ векторы независимых переменных, $\beta_0, \beta_1, \beta_2$ – скалярные параметры, ϵ – $n \times 1$ вектор ошибок. Наша цель – определить корреляцию между y и, например, первым регрессором x_1 после исключения влияния x_2 .

Соответствующая процедура устроена следующим образом:

1. Осуществим регрессию y на x_2 и константу и получим прогнозные значения $\hat{y} = \hat{\alpha}_1 + \hat{\alpha}_2 x_2$.
2. Осуществим регрессию x_1 на x_2 и константу и получим прогнозные значения $\hat{x}_1 = \hat{\gamma}_1 + \hat{\gamma}_2 x_2$.
3. Удалим влияние x_2 , взяв остатки $e_y = y - \hat{y}$ и $e_{x_1} = x_1 - \hat{x}_1$.
4. Определим (выборочный) коэффициент частной корреляции между y и x_1 при исключении влияния x_2 как (выборочный) коэффициент корреляции между e_y и e_{x_1} :

$$r(y, x_1 | x_2) = r(e_y, e_{x_1}). \quad (5.5)$$

Напомним, что из свойств метода наименьших квадратов следует, что e_y и e_{x_1} не коррелированы с x_2 : $x_1^T e_{x_1} = 0$, $x_2^T e_y = 0$. Именно в этом смысле указанная процедура соответствует интуитивному представлению об “исключении (линейного) влияния переменной x_2 ”

Прямыми вычислениями можно показать, что справедлива следующая формула, связывающая коэффициенты частной и обычной корреляции:

$$r(y, x_1 | x_2) = \frac{r(y, x_1) - r(y, x_2)r(x_1, x_2)}{\sqrt{1 - r^2(x_1, x_2)}\sqrt{1 - r^2(y, x_2)}}. \quad (5.6)$$

Значения $r(y, x_1 | x_2)$ лежат в интервале $[-1, 1]$, как у обычного коэффициента корреляции. Равенство коэффициента $r(y, x_1 | x_2)$ нулю означает, говоря нестрого, отсутствие прямого линейного влияния переменной x_1 на y .

Существует тесная связь между коэффициентом частной корреляции $r(y, x_1 | x_2)$ и коэффициентом детерминации R^2 , а именно

$$r^2(y, x_1 | x_2) = \frac{R^2 - r^2(y, x_2)}{1 - r^2(y, x_2)}, \quad (5.7)$$

или

$$1 - R^2 = (1 - r^2(y, x_2))(1 - r^2(y, x_1 | x_2)).$$

5.5 Процедура пошагового отбора переменных

Коэффициент частной корреляции часто используется при решении проблемы *спецификации модели*.

Иногда исследователь заранее знает характер зависимости исследуемых величин, опираясь, например, на экономическую теорию, предыдущие результаты, априорные знания и т.п., и задача состоит лишь в оценивании неизвестных параметров. Однако на практике довольно часто приходится сталкиваться с ситуацией, когда имеется большое число наблюдений различных параметров (независимых переменных), но нет априорной модели изучаемого явления. Возникает естественная проблема, какие переменные включить в регрессионную схему.

В компьютерные пакеты включены различные эвристические процедуры *пошагового отбора регрессоров*. Основными пошаговыми процедурами являются 1) *процедура последовательного присоединения*, 2) *процедура присоединения-удаления* и 3) *процедура последовательного удаления*. Опишем кратко одну из таких процедур, использующую понятие коэффициента частной корреляции.

Процедура присоединения-удаления

На первом шаге из исходного набора объясняющих переменных выбирается (включается в число регрессоров) переменная, имеющая наибольший по модулю коэффициент корреляции с зависимой переменной y .

$$x_i : |r(y, x_i)| = \max_j |r(y, x_j)|.$$

Второй шаг состоит из двух подшагов. На первом из них, который выполняется, если число регрессоров уже больше двух, делается попытка исключить один из регрессоров. Ищется тот регрессор x_s , удаление которого приводит к наименьшему уменьшению коэффициента детерминации. Затем сравнивается значение F-статистики для проверки гипотезы $H_0 : \beta_s = 0$ о незначимости этого регрессора с некоторым заранее заданным пороговым значением $F_{\text{искл}}$. Если $F < F_{\text{искл}}$, то x_s удаляется из списка регрессоров. Второй подшаг состоит в попытке включения нового регрессора из исходного набора предсказывающих переменных. Ищем переменную x_p с наибольшим по модулю частным коэффициентом корреляции (исключается влияние ранее включенных в уравнение регрессоров) и сравниваем значение F-статистики для проверки гипотезы $H_0 : \beta_p = 0$ о незначимости этого регрессора с некоторым заранее заданным пороговым значением $F_{\text{вкл}}$.

$$x_p : |r(y, x_p | x_i)| = \max_j r(y, x_j | x_i).$$

Если $F > F_{\text{вкл}}$, то x_p включается в список регрессоров. Обычно выбирают $F_{\text{искл}} < F_{\text{вкл}}$. Вторым шагом повторяется до тех пор, пока происходит изменение списка регрессоров. Конечно, ни одна из пошаговых процедур не гарантирует получение оптимального по какому-либо критерию набора регрессоров.

5.6 Спецификация модели

		Истинная модель	
		$y = Xb + \varepsilon$	$y = Xb + Z\gamma + \varepsilon$
Оценочная модель	$y = Xb + \varepsilon$	Правильная спецификация. Все оценки и их статистические свойства соответствуют нашим выводам.	Коэффициенты регрессии смещены. Стандартные ошибки рассчитаны некорректно. I
	$y = Xb + Z\gamma + \varepsilon$	Коэффициенты регрессии не смещены. Стандартные ошибки, в целом, рассчитаны корректно. II	Правильная спецификация. Все в порядке.

Случай I. Из модели исключены существенные переменные

- Процесс, порождающий данные ("истинная модель"):
 - (*) $y = Xb + Z\gamma + \varepsilon$ - "длинная" регрессия
- Модель, описывающая данные:
 - (**) $y = Xb + \varepsilon$ - "короткая" регрессия

y – $(n \times 1)$ - вектор наблюдений

X – $(n \times k)$ - матрица

Z – $(n \times l)$ - матрица

b – $(k \times 1)$ - вектор коэффициентов

γ – $(l \times 1)$ - вектор коэффициентов

МНК-оценка вектора параметров b имеет вид для модели (**):

$$\hat{b} = (X^T X)^{-1} X^T y$$

Так как истинная модель имеет вид (*), то

$$E \hat{b} = (X^T X)^{-1} X^T E y = (X^T X)^{-1} X^T (Xb + Z\gamma) = b + (X^T X)^{-1} X^T Z\gamma.$$

$$V(\hat{b}) = V((X^T X)^{-1} X^T y) = (X^T X)^{-1} X^T V(y) (X^T X)^{-1} X^T = \sigma^2 (X^T X)^{-1} X^T X (X^T X)^{-1} = \sigma^2 (X^T X)^{-1}.$$

\Rightarrow оценка \hat{b} смещена, кроме двух случаев:

- 1) $\gamma = 0$.
- 2) $X^T Z = 0$. (ортогональность регрессоров X и Z).

Выводы:

1. Если включить как можно больше переменных в регрессию уравнения, то при этом уменьшится точность оценок, а качество подготовки возрастет.
2. Увеличение числа регрессоров приводит к сильной корреляции между регрессорами и, следовательно, к неустойчивости модели (мультиколлинеарность).

5.7 Замещающие переменные

Часто бывает, что нельзя найти данные по переменной, которую хотелось бы включить в модель, например:

1. Из-за расплывчатого определения переменную нельзя измерить,
2. Измерить можно, но это требует большого количества времени и средств.

В этом случае вместо переменной можно использовать ее заменитель. Это лучше, чем пренебрегать переменной. Например:

1. Качество образования - отношение числа студентов к числу преподавателей,
2. Социально-экономическое положение человека - уровень его дохода.

Причины, по которым следует включить в регрессионное уравнение замещающие переменные:

1. Если опустить важную переменную, то можно получить смещенные оценки,
2. Оценка регрессии можно дать косвенную информацию о замещаемой переменной.

Билет 6. Процесс авторегрессии порядка p . Оператор запаздывания. Условие стационарности процесса авторегрессии

Модель $X_t = aX_{t-1} + \varepsilon_t$ называют **процессом авторегрессии первого порядка**.

Процесс авторегрессии порядка p (в кратком обозначении - AR(p)) определяется соотношениями

$$X_t = a_1X_{t-1} + a_2X_{t-2} + \dots + a_pX_{t-p} + \varepsilon_t, \quad a_p \neq 0,$$

где ε_t - процесс белого шума с $D(\varepsilon_t) = \sigma_\varepsilon^2$.

Для простоты мы будем теперь сразу полагать, что $Cov(X_{t-s}, \varepsilon_t) > 0$ для всех $s > 0$; при этом говорят, что случайные величины ε_t образуют **инновационную (обновляющую) последовательность**, а случайная величина ε_t называется **инновацией** для наблюдения в момент t . Такая терминология объясняется тем, что наблюдаемое значение ряда в момент t получается здесь как линейная комбинация p предшествующих значений этого ряда плюс не коррелированная с этими предшествующими значениями случайная составляющая ε_t , отражающая обновленную информацию, скажем, о состоянии экономики, на момент t , влияющую на наблюдаемое значение X_t .

При рассмотрении процессов авторегрессии и некоторых других моделей удобно использовать **оператор запаздывания L (lag operator)**, который воздействует на временной ряд и определяется соотношением

$$LX_t = X_{t-1}$$

в некоторых руководствах его называют оператором обратного сдвига и используют для него обозначение **B (backshift operator)**.

Если оператор запаздывания применяется k раз, что обозначается как L^k , то это дает в результате

$$L^k X_t = X_{t-k}$$

соотношение, определяющее процесс авторегрессии p – го порядка: $A(L)X_t = \varepsilon_t$.

Для того, чтобы такой процесс был **стационарным**, все корни алгебраического уравнения

$$a(z) = 0$$

т.е. корни $z: |z| > 1$ (вещественные и комплексные) должны лежать вне единичного круга $|z| < 1$.

Билет №7

Процесс скользящего среднего порядка q . Условие стационарности
процесса скользящего среднего
То что мелким - не столь важно

1. Процесс скользящего среднего.

Еще одной простой моделью порождения временного ряда является процесс **скользящего среднего порядка q** ($MA(q)$). Согласно этой модели,

$$X_t = \varepsilon_t + b_1 \varepsilon_{t-1} + b_2 \varepsilon_{t-2} + \dots + b_q \varepsilon_{t-q}, \quad b_q \neq 0,$$

где ε_t — процесс белого шума.

Такой процесс имеет нулевое математическое ожидание. Модель можно обобщить до процесса, имеющего ненулевое математическое ожидание μ , полагая

$$X_t - \mu = \varepsilon_t + b_1 \varepsilon_{t-1} + b_2 \varepsilon_{t-2} + \dots + b_q \varepsilon_{t-q},$$

т.е.

$$X_t = \mu + \varepsilon_t + b_1 \varepsilon_{t-1} + b_2 \varepsilon_{t-2} + \dots + b_q \varepsilon_{t-q}.$$

Для процесса скользящего среднего порядка q используется обозначение $MA(q)$ (скользящее среднее — **moving average**).

При $q = 0$ и $\mu = 0$ получаем процесс белого шума. Если $q = 1$, то

$$X_t - \mu = \varepsilon_t + b \varepsilon_{t-1}$$

— скользящее среднее первого порядка. В последнем случае

$$D(X_t) = (1 + b^2) \sigma_\varepsilon^2,$$

$$\gamma(1) = Cov(X_t, X_{t-1}) = E[(X_t - \mu)(X_{t-1} - \mu)] = b \sigma_\varepsilon^2$$

$$\gamma(k) = Cov(X_t, X_{t-k}) = E[(X_t - \mu)(X_{t-k} - \mu)] = 0, k > 1,$$

так что процесс X_t является стационарным с $E(X_t) = \mu$, $D(X_t) = (1 + b^2) \sigma_\varepsilon^2$,

Автоковариации(для $MA(1)$):

$$\gamma(k) = \begin{cases} (1 + b^2) \sigma_\varepsilon^2, & k = 0 \\ b \sigma_\varepsilon^2, & k = 1 \\ 0, & k > 1 \end{cases}$$

Автокорреляции этого процесса (для $MA(1)$) равны :

$$\rho(k) = \begin{cases} 1, & k = 0 \\ b/(1 + b^2), & k = 1 \\ 0, & k > 1 \end{cases}$$

т.е. коррелограмма процесса имеет весьма специфический вид. Коррелированными оказываются только соседние наблюдения. Корреляция между ними положительна, если $b > 0$, и отрицательна при $b < 0$. Соответственно, процесс $MA(1)$ с $b > 0$ имеет более гладкие, по сравнению с белым шумом, реализации, а процесс $MA(1)$ с $b < 0$ имеет менее гладкие, по сравнению с белым шумом, реализации. Заметим, что для любого процесса $MA(1)$ $\rho(1) \leq 0.5$, т.е. корреляционная связь между соседними наблюдениями невелика, тогда как у процесса $AR(1)$ такая связь может быть сколь угодно сильной (при значениях a , близких к 1). $\rho(k) = a^k$ для $AR(1)$

Модель $MA(q)$ кратко можно записать в виде

$$X_t - \mu = b(L) \varepsilon_t,$$

где

$$b(L) = 1 + b_1 L + \dots + b_q L^q.$$

Для автоковариации имеем:

$$\gamma(k) = E[(X_t - \mu)(X_{t-k} - \mu)] = \begin{cases} \left(\sum_{j=0}^{q-k} b_j b_{j+k} \right) \sigma_\varepsilon^2, & 0 \leq k \leq q, b_0 = 1, \\ 0, & k > q, \end{cases}$$

так что $MA(q)$ является стационарным процессом(определение стационарного процесса - старница 49-50 объединения лекций) с ненулевым математическим ожиданием, дисперсией

$$\sigma_X^2 = (1 + b_1^2 + \dots + b_q^2)\sigma_\varepsilon^2$$

и автокорреляциями

$$\rho_k = \begin{cases} \left(\sum_{j=0}^{q-k} b_j b_{j+k} \right) / \left(\sum_{j=0}^q b_j^2 \right), & k = 0, 1, \dots, q, \\ 0, & k = q + 1, q + 2, \dots \end{cases}$$

Здесь статистическая связь между наблюдениями сохраняется в течение q единиц времени (т.е. "длительность памяти" процесса равна q).

Подобного рода временные ряды соответствуют ситуации, когда некоторый экономический показатель находится в равновесии, но отклоняется от положения равновесия в силу последовательно возникающих непредсказуемых событий, причем система такова, что влияние таких событий отмечается на протяжении некоторого периода времени.

Модели авторегрессии – скользящего среднего ARMA(p,q). Условие стационарности модели авторегрессии – скользящего среднего ARMA(p,q).

Процесс X_t с нулевым математическим ожиданием (МО), принадлежащий такому классу процессов, характеризуется порядками p и q его AR и MA составляющих и обозначается как процесс $ARMA(p, q)$.

Процесс X_t с нулевым МО принадлежит классу $ARMA(p, q)$, если

$$\begin{aligned} X_t &= \sum_{j=1}^p a_j X_{t-j} + \sum_{j=0}^q b_j \varepsilon_{t-j} = \\ &= a_1 X_{t-1} + a_2 X_{t-2} + \dots + a_p X_{t-p} + \varepsilon_t + b_1 \varepsilon_{t-1} + b_q \varepsilon_{t-q}, \\ & \quad a_p \neq 0, \quad b_q \neq 0, \quad \varepsilon_t - \text{белый шум}, \quad (\mathbb{E}X_t = 0). \end{aligned}$$

В операторной форме $a(L)X_t = b(L)\varepsilon_t$, $a(L), b(L)$ – как для $AR(p), MA(q)$. Если процесс X_t имеет постоянное МО μ , то он является процессом типа $ARMA(p, q)$ в широком смысле, если

$$\begin{aligned} X_t - \mu &= \sum_{j=1}^p a_j (X_{t-j} - \mu) + \sum_{j=0}^q b_j \varepsilon_{t-j}, \\ & \quad a_p \neq 0, \quad b_q \neq 0, \quad \varepsilon_t - \text{белый шум}, \quad (\mathbb{E}X_t = \mu \neq 0). \end{aligned}$$

Свойства процесса $ARMA(p, q)$ с $\mathbb{E}(X_t) = \mu$.

1. Процесс стационарен, если все корни уравнения $a(z) = 0$ лежат вне единичного круга $|z| \leq 1$.

Но процесс X_t – типа $ARMA(p, q)$, $a(L)X_t = b(L)\varepsilon_t$, то есть $(\mathbb{E}X_t = 0)$, может быть стационарным и в случае, когда уравнение $a(z) = 0$ имеет корень $z : |z| = 1$.

Пример. Пусть $X_t = \varepsilon_t$. Этот процесс стационарный.

$$X_t - X_{t-1} = \varepsilon_t - \varepsilon_{t-1} \Rightarrow (1 - L)X_t = (1 - L)\varepsilon_t$$

то есть, $a(L)X_t = b(L)\varepsilon_t$, где $a(L) = 1 - L$ и $b(L) = 1 - L$. Иными словами, для процесса X_t получили $ARMA(1, 1)$ представление

$$X_t = X_{t-1} + \varepsilon_t - \varepsilon_{t-1}, \quad (1)$$

для которого уравнение $a(z) = 0$ имеет корень $|z| = 1$.

В общем случае, если у $ARMA(p, q)$ процесса X_t , $a(L)X_t = b(L)\varepsilon_t$, многочлены $a(z)$ и $b(z)$ не имеют общих корней, условие нахождения всех корней уравнения $a(z) = 0$ вне единичного круга $|z| \leq 1$ является *необходимым и достаточным* для стационарности процесса X_t .

2. Если процесс стационарен, то существует эквивалентный ему процесс $MA(\infty)$

$$X_t - \mu = \sum_{j=0}^{\infty} c_j \varepsilon_{t-j}, \quad c_0 = 1, \quad \sum_{j=0}^{\infty} |c_j| < \infty \quad (2)$$

или

$$X_t - \mu = c(L)\varepsilon_t, \quad \text{где} \quad c(z) = \sum_{j=0}^{\infty} c_j z^j = \frac{b(z)}{a(z)} \quad (3)$$

3. Если все корни уравнения $b(z) = 0$ лежат вне единичного круга $|z| \leq 1$ (**условие обратимости**), то существует эквивалентное представление процесса X_t в виде процесса авторегрессии бесконечного порядка $AR(\infty)$

$$X_t - \mu = \sum_{j=1}^{\infty} d_j (X_{t-j} - \mu) + \varepsilon_t, \quad (4)$$

или

$$d(L)(X_t - \mu) = \varepsilon_t, \quad \text{где} \quad d(z) = 1 - \sum_{j=1}^{\infty} d_j z^j = \frac{a(z)}{b(z)} \quad (5)$$

Стационарный процесс $ARMA(p, q)$ всегда можно аппроксимировать процессом скользящего среднего достаточно высокого порядка, а при выполнении условия обратимости его можно также аппроксимировать процессом авторегрессии достаточно высокого порядка.

Специфику формы коррелограммы процесса $ARMA(p, q)$ в общем случае указать труднее, чем для моделей $AR(p)$ и $MA(q)$. Отметим только, что для значений $k > q$ коррелограмма процесса $a(L)X_t = b(L)\varepsilon_t$ выглядит так же, как и коррелограмма процесса авторегрессии $a(L)X_t = \varepsilon_t$.

Если $ARMA(p_1, q_1)$ ряд X_t и $ARMA(p_2, q_2)$ ряд Y_t статистически независимы между собой, и $Z_t = X_t + Y_t$, то типичным является положение, когда Z_t является $ARMA(p, q)$ рядом, у которого:

$$p = p_1 + p_2, \quad (6)$$

$$q = p_1 + q_2, \quad p_1 + q_2 > p_2 + q_1 \quad (7)$$

$$q = p_2 + q_1, \quad p_2 + q_1 > p_1 + q_2 \quad (8)$$

Возможны также ситуации, когда значения p и q оказываются меньше указанных значений. (Такие ситуации возникают в случаях, когда многочлены $a_X(z)$ и $a_Y(z)$, соответствующие авторегрессионным частям процессов X_t и Y_t , имеют общие корни.)

В частном случае, когда оба ряда имеют тип $AR(1)$, но с различными параметрами, их сумма имеет тип $ARMA(2, 1)$.

$$X_t \sim AR(1), Y_t \sim AR(1) \Rightarrow z_t = X_t + Y_t \sim ARMA(2, 1) \quad (9)$$

Из указанного выше факта вытекает, что если каждая из компонент отвечает простой модели AR , то при независимости этих компонент их сумма будет $ARMA$ процессом. Такого же рода процесс мы получим, если часть компонент имеет тип AR , а остальные компоненты имеют тип MA . Единственным исключением является случай, когда все компоненты являются MA процессами – в этом случае в результате получаем MA процесс.

Предположим, наконец, что “истинный” экономический ряд отвечает $AR(p)$ модели, но значения этого ряда измеряются со случайными ошибками, образующими процесс белого шума (т.е. $MA(0)$). Тогда наблюдаемый ряд имеет тип $ARMA(p, p)$.

$$X_t \sim AR(p), Y_t \sim AR(0) \Rightarrow z_t = X_t + Y_t \sim ARMA(p, p) \quad (10)$$

Замечание

$$\sum AR \rightarrow ARMA \quad (11)$$

$$\sum AR + \sum MA \rightarrow ARMA \quad (12)$$

$$\sum MA \rightarrow MA \quad (13)$$

Ранее мы уже говорили о том, что если $ARMA(p, q)$ процесс X_t удовлетворяет условию обратимости, то его можно представить в виде стационарного процесса $AR(\infty)$. Последний, в свою очередь, можно аппроксимировать стационарным процессом $AR(p)$, быть может, достаточно высокого порядка.

Таким образом, в практических задачах можно было бы и вовсе обойтись без использования моделей $ARMA$, ограничиваясь либо AR либо MA моделями. При этом, однако, количество коэффициентов, подлежащих оцениванию, может оказаться слишком большим (что снижает точность оценивания) и даже превосходить количество имеющихся наблюдений. В этом смысле модели $ARMA$ могут быть “**более экономными**”.

1. Обратимость процессов $AR(p)$ и $MA(q)$.

Если влияние прошлых событий ослабевает с течением времени показательным образом, так что $b_j = a^j$, $0 < a < 1$, то искусственное предположение о том, что ряд ε_t начинается в "бесконечном прошлом" приводит к модели **бесконечного скользящего среднего** $MA(\infty)$

$$X_t = \sum_{j=0}^{\infty} a^j \varepsilon_{t-j} = \sum_{j=0}^{\infty} b_j \varepsilon_{t-j}, \quad \text{где } \sum_{j=0}^{\infty} |b_j| < \infty$$

Ранее мы видели, что такое же представление допускает стационарный процесс авторегрессии первого порядка $AR(1)$

$$X_t = aX_{t-1} + \varepsilon_t, \quad a < 1$$

т.е. в рассматриваемом случае процесс $MA(\infty)$ эквивалентен процессу $AR(1)$.

Вообще всякий стационарный процесс $AR(p)$ можно записать в форме процесса $MA(\infty)$:

$$X_t = \mu + \frac{1}{a(L)} \varepsilon_t = \mu + \sum_{j=0}^{\infty} b_j \varepsilon_{t-j} = \mu + b(L) \varepsilon_t,$$

где

$$b(L) = \sum_{j=0}^{\infty} b_j L^j = \frac{1}{a(L)} \quad \text{и} \quad \sum_{j=0}^{\infty} |b_j| < \infty$$

2. Сезонные модели ARMA.

$$X_t = \sum_{j=1}^p a_j X_{t-j} + \sum_{j=0}^q b_j \epsilon_{t-j}, \quad \text{где } b_0 = 1, a_p \neq 0, b_q \neq 0, \epsilon_1 - \text{белый шум} \quad (1)$$

Если наблюдаемый временной ряд обладает выраженной сезонностью, то модель *ARMA*, соответствующая этому ряду, должна содержать составляющие, обеспечивающие проявление сезонности в порождаемой этой моделью последовательности наблюдений. Для квартальных данных чисто сезонными являются стационарные модели **сезонной авторегрессии первого порядка (*SAR(1)*)**.

$$X_t = a_4 X_{t-4} + \epsilon_t, \quad a_4 < 1 \quad (2)$$

и **сезонного скользящего среднего первого порядка (*SMA(1)*)**.

$$X_t = \epsilon_t b_4 \epsilon_{t-4}. \quad (3)$$

Комбинации несезонных и сезонных изменений реализуются, например, в моделях *ARMA*((1, 4), 1) и *ARMA*(1, (1, 4)). Такие модели носят название аддитивных сезонных моделей.

ARMA((1, 4), 1)

$$X_t = a_1 X_{t-1} + a_4 X_{t-4} + \epsilon_t + b_1 \epsilon_{t-1} \quad (4)$$

ARMA(1, (1, 4))

$$X_t = a_1 X_{t-1} + \epsilon_t + b_1 \epsilon_{t-1} + b_4 \epsilon_{t-4}. \quad (5)$$

Кроме рассмотренных примеров аддитивных сезонных моделей, употребляются также и мультипликативные спецификации, например,

$$(1 - a_1 L) X_t = (1 + b_1 L)(1 + b_4 L_4) \epsilon_t, \quad (6)$$

$$(1 - a_1 L)(1 - a_4 L^4) X_t = (1 + b_1 L) \epsilon_t. \quad (7)$$

Первая дает

$$X_t = a_1 X_{t-1} + \epsilon_t + b_1 \epsilon_{t-1} + b_4 \epsilon_{t-4} + b_1 b_4 \epsilon_{t-5}, \quad (8)$$

а вторая

$$X_t = a_1 X_{t-1} + a_4 X_{t-4} - a_1 a_4 X_{t-5} + \epsilon_t + b_1 \epsilon_{t-1} \quad (9)$$

В первой модели допускается взаимодействие составляющих скользящего среднего на лагах 1 и 4 (т.е. значений ϵ_{t-1} и ϵ_{t-4}), а во второй – взаимодействие авторегрессионных составляющих на лагах 1 и 4 (т.е. значений X_{t-1} и X_{t-4}). Конечно, эти две модели являются частными случаями аддитивных моделей

$$X_t = a_1 X_{t-1} + \epsilon_t + b_1 \epsilon_{t-1} + b_4 \epsilon_{t-4} + b_5 \epsilon_{t-5} \quad (10)$$

$$X_t = a_1 X_{t-1} + a_4 X_{t-4} - a_5 X_{t-5} + \epsilon_t + b_1 \epsilon_{t-1} \quad (11)$$

с $b_5 = b_1 b_4$, $a_5 = -a_1 a_4$. При приближенном выполнении последних соотношений (по крайней мере, если гипотезы о наличии таких соотношений не отвергаются), естественно перейти от оценивания аддитивной модели к оцениванию мультипликативной модели, опять следуя *принципу “экономности”* модели (*“parsimony model”*). Впрочем, каких-либо теоретических оснований, ведущих к предпочтению одной формы сезонности перед другой (мультипликативной или аддитивной), не существует.

10. Подбор стационарной модели ARMA. Идентификация стационарной модели ARMA. Поведение автокорреляционных и частных автокорреляционных функций.

Процесс $ARMA(p, q) : x_t = \sum_{j=1}^p \alpha_j x_{t-j} + \sum_{j=0}^q \beta_j \varepsilon_{t-j}, \quad \beta_0 = 1, \quad \alpha_p \neq 0, \quad \beta_q \neq 0, \quad \varepsilon_t$ - белый шум.

Для идентификации стационарной модели ARMA необходимо определить различие поведения **автокорреляционных** и **частных автокорреляционных** функций временных рядов, соответствующих различным моделям ARMA.

В частности, поведение автокорреляционных функций для различных моделей ARMA:

1. автокорреляционные функции ρ_k процесса $AR(p)$, начиная с некоторого $k \geq p$, являются экспоненциально затухающими;
2. автокорреляционные функции ρ_k процесса $MA(q)$, начиная с некоторого $k \geq q$, обрываются ($\rho_k = 0$).

$$\rho_\tau = \frac{1}{Var(X_t)} E\{(X_t - \mu)(X_{t-\tau} - \mu)\} \quad - \text{ автокорреляционная функция}$$

φ_{kk} - k -ое значение частной автокорреляционной функции.

1. AR(p):

Если исследуемый процесс принадлежит к виду $AR(p)$, то $\varphi_{pp} = \alpha_p$. Это означает, что для процесса $AR(p)$ $\varphi_{kk} = 0$ при $k > p$.

В результате частная автокорреляционная функция для процесса AR играет точно такую же качественную роль, как автокорреляционная функция для процесса MA . Она обращается в нуль, как только $k > p$.

2. Комбинация двух различных типов процессов AR(p) и MA(q):

$$x_t = \sum_{j=1}^p \alpha_j x_{t-j} + \sum_{j=0}^q \beta_j \varepsilon_{t-j}, \quad \beta_0 = 1, \quad - \text{ в развернутом виде;}$$

$$\alpha_p(L)x_t = \beta_q(L)\varepsilon_t \quad - \text{ в операторном виде.}$$

Начнем с простой ситуации $ARMA(1, 1)$.

В операторной записи: $(1 - \alpha L)x_t = (1 + \beta L)\varepsilon_t$, или в обычном виде: $x_t = \alpha x_{t-1} + \varepsilon_t + \beta \varepsilon_{t-1}$

Условие стационарности процесса ARMA: $|\alpha| < 1$.

Для вычисления дисперсии процесса удобно использовать $MA(\infty)$ представление, так называемое представление линейного фильтра:

$$\begin{aligned} x_t &= \frac{1 + \beta L}{1 - \alpha L} \varepsilon_t = (1 + \beta L)(1 + \alpha L + \alpha^2 L^2 + \alpha^3 L^3 + \dots) \varepsilon_t = \\ &= [1 + (\alpha + \beta)L + \alpha(\alpha + \beta)L^2 + \alpha^2(\alpha + \beta)L^2 + \dots] \varepsilon_t. \end{aligned}$$

$$x_{t-1} = Lx_t, \quad x_t = \alpha x_{t-1} + \beta \varepsilon_t + \beta \varepsilon_{t-1}$$

$$\begin{aligned} \frac{\gamma(1)}{\gamma(0)} &= \frac{\alpha\gamma(0) + \beta\sigma_\varepsilon^2}{\gamma(0)} = \alpha + \frac{\beta\sigma_\varepsilon^2}{\gamma(0)} = \alpha + \frac{\beta\sigma_\varepsilon^2(1-\alpha^2)}{(1+\beta^2+2\alpha\beta)\sigma_\varepsilon^2} = \frac{\alpha + \alpha\beta^2 + 2\alpha^2\beta + \beta - \alpha^2\beta}{1 + \beta^2 + 2\alpha\beta} = \\ &= \frac{\alpha + \beta + \alpha^2\beta + \beta + \alpha\beta^2}{1 + \beta^2 + 2\alpha\beta} = \frac{(\alpha + \beta) + \alpha\beta(\alpha + \beta)}{1 + \beta^2 + 2\alpha\beta} = \frac{(\alpha + \beta)(1 + \alpha\beta)}{1 + \beta^2 + 2\alpha\beta} \end{aligned}$$

$$\text{Тогда дисперсия } Var(x_t) = \left[1 + (\alpha + \beta)^2 [1 + \alpha^2 + \alpha^4 + \dots] \right] \sigma_\varepsilon^2 = \left[1 + \frac{(\alpha + \beta)^2}{1 - \alpha^2} \right] \sigma_\varepsilon^2 =$$

$$= \frac{1 + \beta^2 + 2\alpha\beta}{1 - \alpha^2} \sigma_\varepsilon^2 \Rightarrow \text{автокорреляционная функция } \gamma(0) = \frac{1 + \beta^2 + 2\alpha\beta}{1 - \alpha^2} \sigma_\varepsilon^2.$$

Умножаем выражение $x_t = \alpha x_{t-1} + \varepsilon_t + \beta \varepsilon_{t-1}$ на x_{t-1} , и берем математическое ожидание от обеих частей. Получаем первую автокорреляцию $\gamma(1) = \alpha\gamma(0) + \beta Cov(\varepsilon_{t-1}, x_{t-1})$.

Умножаем выражение $x_{t-1} = \alpha x_{t-2} + \varepsilon_{t-1} + \beta \varepsilon_{t-2}$ на ε_{t-1} и берем математическое ожидание. Получаем $Cov(\varepsilon_{t-1}, x_{t-1}) = \sigma_\varepsilon^2$. После подстановки получаем $\gamma(1) = \alpha\gamma(0) + \beta\sigma_\varepsilon^2$.

Окончательно:

$$\rho_1 = \frac{\gamma(1)}{\gamma(0)} = \frac{(\alpha + \beta)(1 + \alpha\beta)}{1 + \beta^2 + 2\alpha\beta}.$$

Для всех значений γ с индексом большим, чем на порядок MA -части, получаем, что из равенства $\gamma(k+1) = \alpha\gamma(k)$ следует равенство $\rho_{k+1} = \alpha\rho_k$. То есть мы получили то же самое соотношение, которое имели для "чистой" модели $AR(1)$.

Вывод: начиная со второй, автокорреляции $ARMA(1, 1)$ ведут себя так же, как автокорреляции $AR(1)$, но первые автокорреляции этих процессов различаются. Начиная со второго номера, значения автокорреляционной функции все равно убывают экспоненциально.

Следовательно для процесса $ARMA(p, q)$:

1. Поведение автокорреляционной функции:

Начиная с номера $\max(p, q+1)$, автокорреляционная функция "затухает" в том же смысле, что и для процесса AR . Таким же качественным поведением характеризуется и частная автокорреляционная функция процесса $ARMA(p, q)$, так как для $MA(q)$ автокорреляционная функция для номеров, больших q , равна нулю. Поэтому влияние MA -части при $k > q$ как бы прекращается, и дальше "работает" только $AR(p)$.

2. Поведение частной автокорреляционной функции: Частная автокорреляционная функция для процесса $AR(p)$ для k , больших чем p , равна нулю. То есть $AR(p)$ перестает влиять на частную автокорреляционную функцию, и остается только влияние $MA(q)$.

Вывод: Если процесс относится к типу $ARMA(p, q)$, то, начиная с некоторого номера (причем этот номер важен, он нам говорит о величине p и q), и автокорреляционная, и частная автокорреляционная функции ведут себя как сумма затухающих экспонент, если ряд стационарен.

Пример 1.

Рассмотрим нестационарный ряд $x_t = x_{t-1} + \varepsilon_t$. Если взять первую разность, равную $\Delta x_t = x_t - x_{t-1} = (1 - L)x_t$, то уравнение сведется к следующему: $\Delta x_t = y_t = \varepsilon_t$. То есть в первых разностях ряд станет стационарным.

Билет 11. Нестационарный процесс авторегрессии-интегрированного скользящего среднего ARIMA (p,d,q). Подход Бокса-Дженкинса построения модели типа ARIMA (p,d,q) по реализации временного ряда.

Модель *ARIMA* применяется в тех случаях, когда исследуемый ряд нестационарен. Ряд становится стационарным при помощи оператора разности:

$$\Delta X_t = X_t - X_{t-1}$$

$$\Delta^k X_t = \Delta^{k-1} X_t - \Delta^{k-1} X_{t-1}$$

Модель *ARIMA*(p, d, q) выглядит следующим образом:

$$\Delta^d X_t = \sum_{j=1}^p a_j \Delta^d X_{t-j} + \sum_{j=1}^q b_j \varepsilon_{t-j} + \varepsilon_t$$

Оператор разности степени d позволяет привести к стационарному виду ряд с трендом степени d .

Подход Бокса-Дженкинса

Этот подход служит для выбора типа модели *ARIMA* по временному ряду. Он включает в себя следующие этапы:

1) **Идентификация модели**

Если ряд стационарен, то, начиная с какого-то номера, его автокорреляции начнут убывать. Если этого не происходит, то сначала нужно привести ряд к стационарному, выбрав порядок интеграции d . Обычно $d \leq 2$.

К полученному ряду нужно подобрать параметры p и q модели *ARMA*(p, q).

2) **Оценка модели**

На этом этапе, с уже известными p и q , оцениваются коэффициенты $a_1, \dots, a_p, b_1, \dots, b_q$.

3) **Диагностика и тестирование модели**

4) **Использование модели для прогнозирования будущих значений временного ряда**

Оценивание коэффициентов модели ARMA. Оценивание параметров модели AR(p) методом наименьших квадратов. Оценивание параметров модели MA(q) методом максимального правдоподобия и с помощью процедуры поиска на сетке.

Модель $AR(p) : \alpha_1 x_{t-1} + \dots + \alpha_p x_{t-p} + \varepsilon_t$.

Для нее можно применить обычный метод наименьших квадратов (МНК). Поскольку регрессоры относятся к предыдущим моментам времени ($t - j \neq t$), а ε_t - белый шум, то корреляция регрессоров x_{t-j} со случайным возмущением ε_j отсутствует. Метод наименьших квадратов (МНК) дает:

1. смещенные оценки $E\hat{\theta} \neq \theta$;
2. состоятельные оценки $\lim_{n \rightarrow \infty} \hat{\theta} \rightarrow \theta$, несмотря на то, что присутствует стохастический регрессор.

Нужно проверить по остаткам, действительно ли наши предположения о том, что ε_t - белый шум, выполнены.

Если дополнительно белый шум является гауссовым ($\varepsilon_t \sim \mathcal{N}(0, \sigma_\varepsilon^2)$), то значения x_t распределены нормально, а оценки коэффициентов $\hat{\alpha}_1, \dots, \hat{\alpha}_p$ произведенные МНК, состоятельны и асимптотически нормальны.

Если данные имеют ненулевое выборочное среднее \bar{x} , то можно либо вычесть это среднее из данных и строить регрессию без свободного члена, т.е. рассматривать $x_t - \bar{x}$, либо просто строить регрессию со свободным членом θ .

Другими словами, для оценки математического ожидания процесса $AR(p)$ можно использовать две статистики: $\bar{x} = \frac{1}{T} \sum_{t=1}^T x_t$ и $\hat{\mu} = \frac{\hat{\theta}}{1 - \hat{\alpha}_1 - \dots - \hat{\alpha}_p}$, в которой использованы оценки МНК. Для гауссова процесса ε_t обе оценки состоятельны и асимптотически нормальны. Более того, обе они асимптотически независимы от оценок параметров модели, полученных МНК.

Для моделей скользящего среднего $MA(q) \left(x_t = \sum_{j=1}^q \beta_j \varepsilon_{t-j} \right)$ невозможно аналитически выразить остаточную сумму квадратов через значения реализации x_t и параметры модели, а следовательно, применить МНК.

В предположении нормальности ошибки ε_t выражаем ковариационную матрицу ошибок (ее элементы - значения автокорреляционной функции) через параметры $\beta_1, \beta_2, \dots, \beta_q$ обратимой модели $MA(q)$. Функция правдоподобия для нормального распределенного вектора $\vec{x} = [x_1, x_2, \dots, x_T]'$ имеет вид

$$L(\vec{\beta}, \sigma_\varepsilon^2 \sum_x |\vec{x}) = \frac{1}{(\sqrt{2\pi\sigma_\varepsilon^2})^T} \left(\sqrt{\det \sum_x} \right)^{-1} \exp \left(- \frac{(\vec{x} - \bar{x}) \sum_x^{-1} (\vec{x} - \bar{x})}{2\sigma_\varepsilon^2} \right) \varepsilon \mathcal{N}(0, \sigma_\varepsilon^2 \sum_x)$$

Здесь через \sum_x обозначена ковариационная матрица процесса \vec{x} . Элементы этой матрицы выражаются, согласно свойству обратимости, через параметры модели $\beta_1, \beta_2, \dots, \beta_q$, поэтому процедура численной оптимизации (максимизации функции правдоподобия) позволяет найти оценки метода максимального правдоподобия, которые будут обладать обычными свойствами состоятельности и асимптотической нормальности. Кроме

того, оценки параметров модели $MA(q)$ и оценка дисперсии случайного возмущения асимптотически независимы.

Для оценки параметров модели $ARMA(p, q)$ может применяться комбинация метода наименьших квадратов с поиском на сетке.

Рассмотрим модель $ARMA(2, 2)$.

Пусть уравнение модели имеет вид

$$(1 - \alpha_1 L - \alpha_2 L^2)x_t = (1 + \beta_1 L + \beta_2 L^2)\varepsilon_t$$

Его можно переписать в виде

$$x_t = \frac{(1 + \beta_1 L + \beta_2 L^2)\varepsilon_t}{1 - \alpha_1 L - \alpha_2 L^2}$$

Введем вспомогательный случайный процесс $x_t = \frac{\varepsilon_t}{1 - \alpha_1 L - \alpha_2 L^2}$. Тогда процессы x_t и z_t связаны соотношением $x_t = (1 + \beta_1 L + \beta_2 L^2)z_t$, которое напоминает уравнение $MA(2)$, только вместо ε_t стоит z_t .

Определим "наблюдения" z_t через наблюдения x_t , сконструировав z_t так же, как ранее остатки модели MA , т.е. значения z_t , которые еще не определены (с нулевыми и отрицательными индексами), полагаем равными нулю. Тогда:

$$z_1 = x_1 \quad (x_1 = z_1 + \beta_1 z_0 (= 0) + \beta_2 z_{-1} (= 0))$$

$$z_2 = x_2 - \beta_1 z_1 \quad (x_2 = z_2 + \beta_1 z_1 + \beta_2 z_0 (= 0))$$

$$z_3 = x_3 - \beta_1 z_2 - \beta_2 z_1$$

...

Далее из определения процесса z_t следует, что значения z_t связан с остатками e_t исходной модели следующими соотношениями: $z_t - \alpha_1 z_{t-1} - \alpha_2 z_{t-2} = e_t$. Относительно процесса z_t модель стала $AR(2)$.

Зная "реализацию" z_t для выбранных значений коэффициентов (β_1^0, β_2^0) , можно оценить коэффициенты α_1, α_2 с помощью МНК.

В результате находим оценки коэффициентов $(\tilde{\alpha}_1, \tilde{\alpha}_2, \tilde{\beta}_1, \tilde{\beta}_2)$, но получены они по-разному. Оценки $(\tilde{\beta}_1, \tilde{\beta}_2)$ заданы как начальные значения, т.е. $(\tilde{\beta}_1, \tilde{\beta}_2) = (\beta_1^0, \beta_2^0)$. А потом, исходя из них, построены оптимальные оценки $(\tilde{\alpha}_1, \tilde{\alpha}_2)$.

Применяя численные методы оптимизации (например, поиск на сетке) оцениваем значения параметров, обеспечивающие $\min_{\alpha_1, \alpha_2, \beta_1, \beta_2} \sum e_i^2$.

Если ε_t - белый гауссовский шум, то **для оценивания модели $ARMA(p, q)$** можно также применять **метод максимального правдоподобия**.

Общая схема его применения следующая.

1. Выражаем значения автокорреляционной функции процесса через параметры модели $\alpha_1, \alpha_2, \dots, \alpha_p, \beta_1, \beta_2, \dots, \beta_q$
2. Строим ковариационную матрицу порядка
3. Записываем функцию правдоподобия для имеющейся выборки x_1, \dots, x_T
4. Решаем (как правило, численно) систему уравнений правдоподобия относительно оценок коэффициентов $\hat{\alpha}_1, \hat{\alpha}_2, \dots, \hat{\alpha}_p, \hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_q$

"Диагностика модели типа ARMA"

Два необходимых момента для проверки:

Первый момент - это качество модели. Для проверки качества модели нужен индикатор типа критерия множественной детерминации R^2 .

Второй момент - это некоррелированность остатков. Если остатки коррелированы, то оценивать AR(p)-часть методом наименьших квадратов нельзя, так как получим несостоятельные оценки

Проверка качества модели

Абсолютные критерии, типа стандартного коэффициента множественной детерминации R^2 , не применяются.

Информационный критерий AIC для модели ARMA(p, q) выглядит следующим образом:

$$AIC(p, q) = \ln \hat{\sigma}^2 + 2 \frac{p+q}{T}$$

где T - число наблюдений.

Байесовский информационный критерий BIC

$$BIC(p, q) = \ln \hat{\sigma}^2 + 2 \frac{p+q}{T} \ln T$$

$$\text{Здесь } \hat{\sigma}^2 = \frac{\sum e_t^2}{T-p-q} \text{ (др. вариант } \hat{\sigma}^2 = \frac{\sum e_t^2}{T} \text{)}$$

Структура этих критериев следующая: логарифмы остаточной суммы квадратов плюс штраф за уменьшение числа степеней свободы.

Нужно так подобрать значения параметров p и q, чтобы получить минимальное значение каждого из критериев: $\min_{p,q} AIC(p, q)$, $\min_{p,q} BIC(p, q)$ (R2 мы отдавали предпочтение модели с большим его значением. R^2 для ARMA **НЕ применяем**)

15 вопрос

Проверка автокорреляции остатков модели ARMA(p,q).

Тест Бокса-Пирса (Q-статистика).

Тест Бокса-Льюнга (улучшение Q-статистики).

В 1970 г. Бокс и Пирс предложили статистический критерий для проверки автокорреляции временного ряда, который сегодня принято называть **Q-статистикой**.

Тест Бокса-Пирса проверяет гипотезу о совместном равенстве нулю всех автокорреляций временного ряда до порядка m включительно, т.е. гипотезу

$$H_0: \rho_1 = \rho_2 = \dots = \rho_m = 0$$

против альтернативной гипотезы

$$H_1: \sum_{i=1}^m \rho_i^2 > 0.$$

Бокс и Пирс показали, что при увеличении длины выборки, т.е при больших T статистики

$$Q(m) = T \sum_{k=1}^m r_k^2$$

(r_k – выборочное автокорреляции) имеет асимптотическое распределение χ^2 с m степенями свободы.

Этот тест было предложено применять для проверки наличия автокорреляции остатков типа ARMA(p,q). В этом случае число степеней свободы уменьшается на $p + q$, т.е статистика $Q(m)$ имеет асимптотическое распределение χ^2 с $m - p - q$ степенями свободы, а если при этом ε_t является процессом нормального белого шума, то $Q(m)$ имеет распределение χ^2 с $m - p - q$ степенями свободы.

Так как статистика Бокса-Пирса имеет малую мощность, Бокс и Льюнг в 1978 году предложили использовать для тех же целей **улучшенную Q-статистику**:

$$Q = (T + 2)T \sum_{i=1}^m (T - i)^{-1} r_j^2$$

По сравнению со статистикой Бокса-Пирса различным слагаемым приданы разные веса. Авторы доказали, что эта статистика имеет то же асимптотическое распределение χ_m^2 , но лучше им аппроксимируется при конечном число наблюдений.

Q-статистику Бокса-Пирса и улучшенную Q-статистику Бокса-Льюнга часто называют **портманто - статистикой** (portmanteau-statistics).

Билет 16. Прогнозирование с помощью моделей ARMA. Безусловный прогноз. Условный прогноз и дисперсия ошибки прогноза для модели MA(q).

При прогнозировании при помощи модели ARMA ошибки могут возникать по 2 причинам:

1. Неопределенность будущих значений случайной величины ε
2. Отсутствие точных значений коэффициентов модели $a_1, \dots, a_p, b_1, \dots, b_q$. Имеются только их оценки.

В **условном прогнозе** независимые переменные известны точно. Примером такого прогнозирования может послужить модель множественной регрессии, где регрессоры заданы точно.

В **безусловном прогнозе** независимые переменные точно не известны. Например, в модели ARIMA значение переменной x_t прогнозируется для некоторого будущего значения $t + \tau$. Переменные x_{t-1}, \dots, x_{t-p} , которые являются регрессорами модели, случайны.

Лучше тот прогноз, у которого меньше дисперсия. Поэтому, поскольку условная дисперсия случайной величины не превышает ее безусловную дисперсию, точность условного прогноза всегда не ниже точности безусловного.

Рассмотрим модель MA(q): $x_t = \theta + \varepsilon_t + b_1\varepsilon_{t-1} + \dots + b_q\varepsilon_{t-q}$. В этой модели θ – математическое ожидание ряда. Предположим, что ее коэффициенты b_1, \dots, b_q известны точно, а так же имеются значения $x_t, t \in [1; T]$.

Тривиальным **безусловным** прогнозом для момента времени $T + 1$ будет математическое ожидание:

$$\hat{x}_{T+1} = E\{\theta + \varepsilon_{T+1} + b_1\varepsilon_T + \dots + b_q\varepsilon_{T-q+1}\} = \theta$$

Условным прогнозом для момента времени $T + 1$ является условное математическое ожидание:

$$\hat{x}_{T+1} = E\{\theta + \varepsilon_{T+1} + b_1\varepsilon_T + \dots + b_q\varepsilon_{T-q+1} | x_1, \dots, x_T\}$$

Причем, условные математические ожидания всех слагаемых, кроме ε_{T+1} , не равны 0. Условные математические ожидания всех предыдущих значений равны остаткам $e_t = x_t - \hat{x}_t$. Аналогично составляется прогноз не на 1, а на несколько шагов вперед. Все последующие T компоненты суммы будут равны 0, а предыдущие – остаткам. Таким образом, прогноз зависит от остатков на предыдущих шагах:

$$\hat{x}_{T+1} = \theta + b_1e_T + \dots + b_qe_{T-q+1}$$

Начиная с шага $q + 1$ условный прогноз начинает совпадать с безусловным и равен θ .

Рассмотрев условную дисперсию, получаем, что дисперсия прогноза на τ шагов равна:

$$(1 + b_1^2 + \dots + b_\tau^2)\sigma_\varepsilon^2 \text{ при } \tau < q$$

При $\tau \geq q$ дисперсия ошибки условного прогноза совпадает с дисперсией ошибки безусловного прогноза, то есть с дисперсией случайного процесса x_t .

Билет 17. Условный прогноз и дисперсия ошибки прогноза для модели $AR(p)$. Процедура пост-прогноза.

Рассмотрим модель стационарного процесса $AR(p) : x_t = \theta + \alpha_1 x_{t-1} + \dots + \alpha_p x_{t-p} + \varepsilon_t$.

$$x_{T+1} = \theta + \alpha_1 x_T + \dots + \alpha_p x_{T-p+1} + \varepsilon_{T+1}, \hat{x}_{T+1} = \theta + \alpha_1 x_T + \dots + \alpha_p x_{T-p+1}$$

$$Var\{x_{T+1} - \hat{x}_{T+1} | x_1, \dots, x_T\} = Var\{\varepsilon_{T+1}\} = \sigma_\varepsilon^2$$

Вообще дисперсию можно вычислять через подстановки все новых \hat{x}_{T+n} , но выражение для нее очень быстро становится громоздким, поэтому проще переходить к эквивалентному MA представлению $x_t = \theta + \varepsilon_t + \psi_1 \varepsilon_{t-1} + \dots + \psi_q \varepsilon_{t-q} + \dots$, хотя и с бесконечным числом слагаемых. Тогда дисперсия

ошибки прогноза на τ шагов выражается формулой $\sigma_\varepsilon^2 \sum_{i=0}^{\tau-1} \psi_i^2$.

Рассмотрим, например, модель $AR(2)$ без свободного члена. Тогда $\hat{x}_{T+1} = \alpha_1 x_T + \alpha_2 x_{T-1}$, а $x_{T+1} = \alpha_1 x_T + \alpha_2 x_{T-1} + \varepsilon_{T+1}$. Очевидно, что дисперсия ошибки прогноза на 1 шаг равна σ_ε^2 . Для прогноза на 2 шага соответственно получаем:

$$\tilde{x}_{T+2} = \alpha_1 \tilde{x}_{T+1} + \alpha_2 x_T = \alpha_1(\alpha_1 x_T + \alpha_2 x_{T-1}) + \alpha_2 x_T,$$

$$x_{T+2} = \alpha_1 x_{T+1} + \alpha_2 x_T + \varepsilon_{T+2} = \alpha_1(\alpha_1 x_T + \alpha_2 x_{T-1} + \varepsilon_{T+1}) + \alpha_2 x_T + \varepsilon_{T+2},$$

Дисперсия ошибки прогноза на 2 шага равна: $(1 + \alpha_1^2)\sigma_\varepsilon^2$.

Для прогноза на 3 шага получим:

$$\tilde{x}_{T+3} = \alpha_1 \tilde{x}_{T+2} + \alpha_2 \tilde{x}_{T+1} = \alpha_1(\alpha_1(\alpha_1 x_T + \alpha_2 x_{T-1}) + \alpha_2 x_T) + \alpha_2(\alpha_1 x_T + \alpha_2 x_{T-1}),$$

$$x_{T+3} = \alpha_1 x_{T+2} + \alpha_2 x_{T+1} + \varepsilon_{T+3} = \alpha_1(\alpha_1(\alpha_1 x_T + \alpha_2 x_{T-1} + \varepsilon_{T+1}) + \alpha_2 x_T + \varepsilon_{T+2}) + \alpha_2(\alpha_1 x_T + \alpha_2 x_{T-1} + \varepsilon_{T+1}) + \varepsilon_{T+3}.$$

Дисперсия ошибки прогноза на 3 шага равна $(1 + \alpha_1^2 + \alpha_2^2 + 2\alpha_1^2\alpha_2 + \alpha_1^4)\sigma_\varepsilon^2$.

Что касается матожидания условного прогноза, то для него справедливо следующее:

$$\lim_{\tau \rightarrow \infty} E\{x_{T+\tau} | x_1, \dots, x_T\} = \frac{\theta}{\theta - \alpha_1 - \dots - \alpha_p}$$

Для общей **модели ARMA(p,q)** нужно просто объединить полученные результаты. Если мы хотим получить прогноз с минимальной среднеквадратической ошибкой, то:

- 1) Рассчитываем прогнозные значения $x_{T+\tau}$ по нашей модели, подставляя туда для времени $[1, T]$ - наблюдаемые значения x_1, \dots, x_T и рассчитанные значения остатков e_1, \dots, e_T , а для всех последующих моментов времени $\tau = T+1, T+2, \dots$ - заменяем остатки нулями ($e_{\tau+1} = 0, e_{\tau+2} = 0, \dots$), а для значений x_{T+1}, x_{T+2}, \dots подставляем их прогнозные значения $\hat{x}_{T+1}, \hat{x}_{T+2}, \dots$
- 2) Для получения дисперсии ошибки прогноза переходим от ARMA к MA представлению и пользуемся формулой.

Нестационарные временные ряды

Рассмотрим только частные виды нестационарных временных рядов. Один из таких видов - $ARIMA(p, d, q)$. По определению модели $ARIMA(p, d, q)$, d - это степень интеграции ряда, т.е. ряд становится стационарным после применения d раз операции взятия последовательной разности.

Ниже нестационарные ряды, которые могут быть приведены к стационарному виду с помощью взятия последовательных разностей.

1 тип. Процесс с детерминированным полиномиальным трендом

$x_t = P_k(t)$, где $P_k(t)$ - полином степени k от t , а ϵ_t - стационарный процесс, не обязательно белый шум.

2 тип. Процесс случайного блуждания

$$x_t = \mu + x_{t-1} + \epsilon_t$$

В этом случае $\Delta x_t = \mu + \epsilon_t$, и процесс x_t называется случайным блужданием с дрейфом.

В чем-то они схожи: у обоих есть линейный тренд, но они отличаются случайной частью. В первом случайная часть - это текущий шок, текущие возмущения, а во втором - это накопленные возмущения от всех предыдущих шоков.

Ряд второго типа можно привести к стационарному виду только взятием первой разности, а ряд первого типа можно привести к стационарному виду как взятием разности, так и выделением линейного тренда.

1 тип. Процесс, приводимый к стационарному путем выделения линейного тренда - TSP (trend stationary process).

Это процесс вида $x_t = \alpha + \beta t + \epsilon_t$, он приводится к стационарному процессу путем включения в регрессию линейного тренда. Это, в принципе, процесс, у которого есть детерминированный тренд.

2 тип. Процесс, приводимый к стационарному путем взятия первой разности - DSP (diferencing stationary process).

Это процесс вида $x_t = \mu + x_{t-1} + \epsilon_t$.

Билет 19: Различия в поведении процессов TSP и DSP. Гипотеза единичного корня. Проблема принадлежности временного ряда к классу DS/TS (неправильная идентификация). Методика Дики-Фуллера. Критерий Дики-Фуллера

1 Различия в поведении процессов TSP и DSP

1 тип. *Процесс, приводимый к стационарному путем выделения линейного тренда - TSP* (trend stationary process).

Это процесс вида $x_t = \alpha + \beta t + \epsilon_t$, он приводится к стационарному процессу путем включения в регрессию линейного тренда. Это, в принципе, процесс, у которого есть детерминированный тренд.

2 тип. *Процесс, приводимый к стационарному путем взятия первой разности - DSP* (diferencing stationary process).

Это процесс вида $x_t = \mu + x_{t-1} + \epsilon_t$.

Свойства процессов TSP и DSP

Процесс TSP	Процесс DSP
Не стационарен из-за непостоянного тренда.	Не стационарен из-за непостоянной дисперсии.
Конечная память о шоках, он забывает об ошибке на предыдущем шаге сразу. Если вместо белого шума будет стоять более общий процесс $ARMA(p, q)$, то, конечно, шоки сказываются некоторое время, но их влияние со временем ослабевает, т.е. процесс с конечной памятью в случайных воздействиях.	Так как в явном решении стоит сумма всех предыдущих ϵ , то шоки помнятся все время. Это процесс с бесконечной памятью. Экономически это не очень понятно, шоки не должны сказываться постоянно.

Для решения вопроса об отнесении X_t к классу TS (стационарных или стационарных относительно тренда) или DS (разностно стационарных) процессов имеется целый ряд различных процедур. Однако все эти процедуры имеют различные недостатки. Процедуры, оформленные в виде формальных статистических критериев, как правило, имеют достаточно низкую мощность, что ведет к весьма частому неопровержению исходной (нулевой гипотезы), когда она в действительности не выполняется. В то же время, невыполнение теоретических предпосылок, на которых основывается критерий, при применении его к реальным данным приводит к отличию реально наблюдаемого размера критерия от заявленного уровня значимости. Вследствие последнего обстоятельства теряется контроль над вероятностью ошибки первого рода, и это может приводить к слишком частому отвержению нулевой гипотезы, когда она в дей-

ствительности верна. В связи с таким положением вещей исследователи обычно используют при анализе рядов на принадлежность их к классу TS или DS не один, а несколько критериев и подкрепляют выводы, полученные с использованием формальных критериев (с установленными уровнями значимости) графическими процедурами.

2 Гипотеза единичного корня

Если ряд X_t имеет тип $ARIMA(p, d, q)$, то в результате его (d -кратного дифференцирования) мы получаем стационарный ряд $\Delta^d x_t$, типа $ARMA(p, q)$:

$$\sum_{j=0}^p \alpha_j \Delta^d x_{t-1} = \sum_{j=0}^q \beta_j \epsilon_{t-1}$$

или

$$\alpha_p(L) \Delta^d x_t = \beta_q(L) \epsilon_t,$$

где $\alpha_p(L)$ и $\beta_q(L)$ - полиномы от оператора обратного сдвига L , имеющие степени p и q , соответственно.

Заметим, что $\Delta x_t = (1 - L)x_t$, так что $\Delta^d x_t = (1 - L)^d x_t$ и $\alpha_p^*(L)(1 - L)^d x_t = \beta_q(L) \epsilon_t$ или $\alpha_p(L)x_t = \beta_q(L) \epsilon_t$

где $\alpha_p(L) = \alpha_p^*(L)(1 - L)^d$ - полином степени $(p + d)$.

Поскольку ряд $\Delta^d x_t$ стационарный, все p корней полинома $\alpha_p^*(z)$ находятся за пределами единичного круга, так что полином $\alpha_p^*(z)$ имеет p корней за пределами единичного круга и d корней на границе этого круга, а точнее, корень $z = 1$ кратности d .

Таким образом, ряд X_t представляется нестационарной моделью $ARMA(p+d, q)$, в которой авторегрессионный полином $a(L)$ имеет ровно d корней, равных 1, а все остальные корни по модулю больше 1. Поэтому проверка нулевой гипотезы H_0 о том, что некоторый ARMA ряд X_t является DS-рядом (т.е. нестационарным рядом), может быть сведена к проверке гипотезы о том, что авторегрессионный полином $a(L)$ имеет хотя бы один корень, равный 1. Это оправданно, если исходить из предположения, что (z) не имеет корней внутри единичного круга, т.е. исключить из рассмотрения “взрывные” модели. При этом о гипотезе H_0 кратко говорят как о **гипотезе единичного корня** (*UR - unit root hypothesis*), хотя точнее было бы говорить о гипотезе авторегрессионного единичного корня. В качестве альтернативной тогда выступает **TS гипотеза** о том, что рассматриваемый ARMA ряд - стационарный.

Критерии, в которых за исходную (нулевую) гипотезу берется гипотеза TS, служат скорее для подтверждения результатов проверки DS-гипотезы. В этом случае вместо проверки гипотезы единичного корня у полинома $a(z)$ проверяется гипотеза о наличии единичного корня $z = 1$ у уравнения $b^*(z) = 0$, где $b^*(L)$ - полином от оператора обратного сдвига L в представ-

лении в виде процесса скользящего среднего. $X_t = b * (z) \in t$ ряда разностей. $X_t = X_t - X_{t-1}$ исходного процесса X_t .

$$\begin{aligned} H_0 : X &\sim ARMA(p + d, q) \sim \text{DS-ряд} \Rightarrow \\ \Rightarrow H_0 : \exists \text{ хотя бы 1 корень } z = 1 \text{ у полинома } \alpha_p(L) : \alpha_p(z) < 0 \\ H_1 : X_t &\sim ARMA(p + d, q) - \text{стационарный ряд} \end{aligned}$$

$$\begin{aligned} H_0 : X_t &\sim ARMA(p + d, q) - \text{TS-ряд} \Rightarrow \\ H_1 : \exists z = 1 : \beta_q(z) &= 0 \text{ полином} \\ \beta_1^*(z) &= 0 \end{aligned}$$

3 Методика Дики-Фуллера. Критерий Дики-Фуллера

Рассмотрим модель: $x_t = \alpha + \rho x_{t-1} + \beta t + \epsilon_t$.

Эта модель вобрала в себя черты обеих моделей: TS и DS, и гипотезы о характере ряда можно записать в виде простых гипотез о ее параметрах:

$$H_0: \text{ряд является DS} \Rightarrow \begin{cases} \rho = 1 \\ \beta = 0 \end{cases}$$

H_1 : ряд является TS $\Rightarrow \rho < 1$, (ϵ - не просто белый шум, а некоторый стационарный ряд)

Нулевая гипотеза относится к классу общих линейных гипотез.

При традиционном подходе для ее проверки нужно оценить 2 регрессии: $x_t = \alpha + \rho x_{t-1} + \beta t + \epsilon_t$, и $x_t = \alpha + x_{t-1} + \epsilon_t$. И затем проверить значимость разности остаточных сумм квадратов, используя F -статистику.

Рассмотрим другой подход для более простой версии этой модели: $x_t = \alpha + \rho x_{t-1} + \epsilon_t$, т.е. без включения линейного тренда.

В этом случае гипотезы принимают вид:

$H_0: \rho = 1 \Rightarrow$ ряд является DS.

$H_1: \rho < 1 \Rightarrow$ ряд является TS.

Проверим гипотезу о том, что $\rho = 1$ при помощи t -статистики.

После вычитания из обеих частей уравнения x_{t-1} его можно переписать в другом виде:

$$\Delta x_t = \alpha + (\rho - 1)x_{t-1} + \epsilon_t.$$

Пусть $\rho - 1 = \gamma$, тогда проверяемые гипотезы примут вид:

$H_0 : \gamma = 0 \Rightarrow$ ряд является DS,

$H_1 : \gamma < 0 \Rightarrow$ ряд является TS.

В классической линейной регрессии для проверки такой гипотезы применяется односторонняя t -статистика. Но в случае выполнения нулевой гипотезы, ряд x_t является случайным блужданием, его дисперсия стремится к бесконечности при увеличении t , и $\frac{\hat{\gamma}}{s.e.(\hat{\gamma})}$, где $s.e.(\hat{\gamma})$ - оценка дисперсии $\hat{\gamma}$, не является распределением Стьюдента. Следовательно, и асимптотическое распределение этой статистики не является нормальным. Причиной является невы-

полнение условий Центральной предельной теоремы в этом случае. Критические точки этого распределения приходится рассчитывать численно, используя имитационное моделирование Монте-Карло. Впервые это распределение было выведено и затабулировано в работе Дикки и Фуллера и носит их имя. Тест, использующий для проверки типа нестационарности распределение статистики $\frac{\hat{\gamma}}{s.e.(\hat{\gamma})}$, при условии $\gamma = 0$, т.е. когда процесс принадлежит типу DS, называется **тестом Дикки-Фуллера**, и обозначается как **DF-тест**. При условии, что нулевая гипотеза о том, что $\gamma = 0$, выполнена, мы имеем процесс типа случайного блуждания (DSP). Именно для этого случая не работает t -статистика.