Contents

Part I

Семестр 1

1 Линейная регрессия

Линейная регрессия:
Теоретический вид:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

Выборочная регрессия:

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X$$

Остатки регрессии:

$$e_i = Y_i - \hat{Y}_i$$

Задача: Минимизировать остатки регрессии
Критерий: Residual sum of squares

$$RSS = \sum_{i=1}^{n} e_i^2 \to min_{\beta_0, \beta_1}$$

Решение задачи:

$$RSS(\hat{\beta}_0, \hat{\beta}_1) = \sum_{i=1}^{n} (Y_i - \hat{\beta}_0 + \hat{\beta}_1 X_i)^2$$

$$\frac{\partial RSS}{\partial \beta_0} = -2 \sum_{i=1}^{n} (Y_i - \hat{\beta}_0 + \hat{\beta}_1 X_i) = 0 \to \sum_{i=1}^{n} Y_i - n\hat{\beta}_0 - \hat{\beta}_1 \sum_{i=1}^{n} X_i = 0;$$

$$\frac{\partial RSS}{\partial \beta_1} = -2 \sum_{i=1}^{n} (Y_i - \hat{\beta}_0 + \hat{\beta}_1 X_i) X_i = \sum Y_i - \hat{\beta}_0 \sum X_i - \hat{\beta}_1 \sum X_i^2 = 0$$

<u>Система нормальных уравнений</u>

$$\begin{cases} n\hat{\beta}_0 + \beta_1 \sum X_i = \sum Y_i \\ \hat{\beta}_0 \sum X_i + \hat{\beta}_1 \sum X_i^2 = \sum Y_i X_i \end{cases}$$

Решение:

$$\begin{cases} \hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X} \\ \hat{\beta}_1 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} = \frac{\hat{Cov}(X;Y)}{\hat{Var}(X)} \end{cases}$$

1.1 Свойства линейной регрессии

1. $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X} \to$ Линия регрессии проходит через $(\bar{X}, \bar{Y})$
2. Отсутствие систематической ошибки $\to \sum_{i=1}^{n} e_i = 0$
3. $\sum_{i=1}^{n} Y_i = \sum_{i=1}^{n} \hat{Y}_i \to$

4. $\bar{Y} = \hat{Y}_{\text{ср}}$

5. $\sum_{i=1}^{n} X_i e_i = 0 \to$ векторы ортогональны

6. $\sum_{i=1}^{n} \hat{Y}_i e_i = 0 \to$ векторы ортогональны

1.2 Дисперсионный анализ

$$\hat{Var}(Y) = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2 \to \hat{y}_i + e_i$$

$$y_i^2 = \hat{y}_i^2 + 2\hat{y}_i e_i + e_i^2$$

$$\sum y_i^2 = \sum_{i=1}^{n} \hat{y}_i^2 + 2\sum_{i=1}^{n} \hat{y}_i e_i + \sum_{i=1}^{n} e_i^2 \xrightarrow{\sum(\hat{y}_i e_i)=0} \sum_{i=1}^{n} y_i^2 = \sum_{i=1}^{n} \hat{y}_i^2 + \sum_{i=1}^{n} e_i^2$$

$$TSS = \sum_{i=1}^{n} y_i^2$$

$$ESS = \sum_{i=1}^{n} \hat{y}_i^2$$

$$RSS = \sum_{i=1}^{n} e_i^2$$

$$TSS = RSS + ESS$$

Качество подборки регрессии R^2

$$0 < R^2 = \frac{ESS}{TSS} = \frac{\sum(\hat{Y}_i - \bar{Y})^2}{\sum(Y_i - \bar{Y})^2} = \frac{\hat{Var}(\hat{Y})}{\hat{Var}(Y)} < 1$$

Представляет собой долю дисперсии Y, объясняющаяся X

1.3 Интервальные оценки

$$\hat{\beta}_1 = \frac{\hat{Cov}(X,Y)}{\hat{Var}(X)} = \beta_1 + \frac{Cov(X,\varepsilon)}{Var(X)}$$

1.4 Теорема Гаусса-Маркова

1. Модель правильно специфицирована
   (a) Есть все необходимые факторы
   (b) Нет лишних факторов
   (c) Правильно выбрана функциональная форма модели

2. X детерминированы и не равны

3. $E(\varepsilon_i) = 0$

4. $Var(\varepsilon_i) = \sigma_\varepsilon^2$

5. $Cov(\varepsilon_i, \varepsilon_j) = 0$

При выполнении данных условий **оценки модели являются BLUE(best linear unbiased estimator)**

1.5 Стандартные ошибки регрессии

$$\varepsilon_i \sim \mathcal{N}(0, \sigma_\varepsilon^2) \rightarrow \begin{cases} \hat{\beta}_0 \sim \mathcal{N}(\beta_0, \frac{\sum X_i^2 \sigma_\varepsilon^2}{n \sum x_i^2}) \\ \hat{\beta}_1 \sim \mathcal{N}(\beta_1, \frac{\sigma_\varepsilon^2}{\sum x_i^2}) \end{cases}$$

$$\sigma_\varepsilon^2 = \frac{RSS}{n-2}$$

$$\frac{RSS}{\sigma_\varepsilon^2} \sim \chi_{n-2}^2$$

Гипотезы о коэффициентах регрессии

$$\begin{cases} H_0 : \beta_1 = \beta_1^0 \\ H_1 : \beta_1 \neq \beta_1^0 \end{cases}$$

β_1^0 - Математическое ожидание

$$\frac{\hat{\beta}_1 - \beta_1^0}{s.e.(\hat{\beta}_1)} \sim t_{n-2}$$

Гипотеза о значимости коэффициентов

$$\begin{cases} H_0 : \beta_1 = 0 \\ H_1 : \beta_1 \neq 0 \end{cases}$$

$$t = \frac{\hat{\beta}_1}{s.e.(b\hat{e}ta_1)} \sim t_{n-2}$$

$$\boxed{\text{Если P-value} < \alpha \rightarrow \text{Коэффициент является значимым}}$$

1.6 Доверительные интервалы для оценки коэффициентов

$$\hat{\beta}_1 - t_{\frac{\alpha}{2}}^{n-2} \cdot s.e.(\hat{\beta}_1) \leq \beta_1^0 \leq \dots$$

1.7 Предсказания с помощью регрессии

Точечный прогноз:

$$\hat{Y}_{n+1} = \hat{\beta}_0 + \hat{\beta}_1 X_{n+1}$$

Интервальный прогноз:

$$\hat{\beta}_0 + \hat{\beta}_1 X_{n+1} \pm t_{n-2,\alpha/2} \cdot \sqrt{\hat{\sigma}_\varepsilon^2 \left[1 + \frac{1}{n} + \frac{X_{n+1} - \bar{X}}{\sum x_i^2}\right]}$$

Ошибка среднего прогноза

$$\sigma_\varepsilon^2 \cdot \left[(\frac{1}{n} + \frac{\bar{X}^2}{\sum x_i^2}) + \frac{X_{n+2}^2}{\sum x_i^2} - \frac{2X_{n+1}\bar{X}}{\sum x_i^2}\right]$$

1.8 Нормальность распределения остатков

1. Сравнение гистограммы остатков с гистограммой $\mathcal{N}$
2. Q-Q Plot
3. Jarque-Bera test
$$\begin{cases} H_0 : e_i \sim \mathcal{N} \\ H_1 : e_i \not\sim \mathcal{N} \end{cases}$$
$$JB = \frac{n}{6}\left(sk^2 + \frac{1}{4}(k-3)^2\right) \sim \chi^2(2)$$
$$sk = \frac{1}{n}\sum \frac{(X_i - \bar{X})^3}{\hat{\sigma}^3}$$
$$k = \frac{1}{n}\sum \frac{(X_i - \bar{X})^4}{\hat{\sigma}^4}$$
4. Тест Шапиро-Вилка
5. Тест Колмогорова-Смирнова

2 Множественная линейная регрессия

Работаем в пространстве R^n

$$Y_i = \beta_0 + \beta_1 X_{1i} + \ldots + \beta_k X_{ki} + \varepsilon_i, i = 1, \ldots, n$$

Векторный вид:

$$X = \begin{pmatrix} 1 & X_{11} & \ldots & X_{k1} \\ 1 & & & \\ \ldots & \ldots & \ldots & \ldots \\ 1 & X_{1n} & \ldots & X_{kn} \end{pmatrix}_{n \times k+1}$$
$$\beta = \begin{pmatrix} \beta_0 \\ \ldots \\ \beta_k \end{pmatrix}_{k+1 \times 1}$$
$$Y = \beta X + \varepsilon$$

Задача в векторном виде

$$\sum e_i^2 = Y^T Y - 2\hat{\beta}^T X^T Y + \hat{\beta}^T X^T X \hat{\beta} \to min$$
$$\frac{\partial RSS}{\partial \beta} = -2X^T Y + 2X^T X \hat{\beta} = 0$$
$$\hat{\beta} = (X^T X)^{-1} X^T Y$$

2.1 Теорема Гаусса-Маркова для МЛР

Если:

1. Модель правильно специфицирована
2. Не существует линейной связи между регрессорами
3. $E(\varepsilon_i) = 0$
4. $Var(\varepsilon_i) = \sigma_\varepsilon^2$
5. $Cov(\varepsilon_i, \varepsilon_j) = 0$

$$Var[\varepsilon] = \begin{pmatrix} \sigma_\varepsilon^2 & \cdots & 0 \\ & \ddots & \\ 0 & \cdots & \sigma_\varepsilon^2 \end{pmatrix}$$

То:

1. $E(\hat{\beta}) = \beta$ - несмещенная
2. Линейная
3. $Var(X\beta + \varepsilon) = Var(\varepsilon) = \sigma_\varepsilon^2 I_n$

$$\hat{Var}(\hat{\beta}) = \sigma_\varepsilon^2 (X^T X)^{-1}$$
$$s.e.(\hat{\beta}) = \sqrt{\sigma_\varepsilon^2 (X^T X)^{-1}}$$
$$\sigma_\varepsilon^2 = \frac{RSS}{n-k-1}$$

2.2 Коэффициент множественной детерминации

$$R^2 = \sum_{i=1}^{k} \beta_i \frac{\hat{Cov}(X_i, Y)}{\hat{Var}(Y)}$$

Недостаток:
Растет при добавлении любого регрессора, независимо от его качества.
Используем R^2_{adj} - он штрафует за лишние факторы

$$R^2_{adj} = 1 - \frac{RSS/(n-k-1)}{TSS/(n-1)}$$

Если в регрессии нет свободного члена $\rightarrow R^2, R^2_a dj$ не являются показателями качества

1. $\sum e_i \neq 0$
2. $TSS \neq ESS + RSS$
3. $R^2 \neq 1 - \frac{RSS}{TSS}$

2.3 Проверка значимости коэффициентов множественной регрессии

$$\begin{cases} H_0 : \beta_j = 0 \text{ - незначим} \\ H_1 : \beta_j \neq 0 \end{cases}$$

$$t = \frac{\hat{\beta}_j}{s.e.(\hat{\beta}_j)} \sim t_{n-k-1}$$

2.4 Гипотеза об адекватности МЛР

$$\begin{cases} H_0 : \beta_1 = \dots = \beta_k = 0 \\ H_1 : \exists \beta_i \neq 0 \end{cases}$$

$$F(k, n-k-1) = \frac{R^2/k}{(1 - R^2/(n-k-1))} \sim F(k, n-k-1)$$

$$\boxed{Pvalue < \alpha \to \text{Регрессия адекватна}}$$

2.5 Гипотеза о Q линейных ограничений

H_0: Имеют место q линейных ограничений лин. регрессии, эти ограничения независимы

$$\frac{(RSS_R - RSS_U)/q}{RSS_U/(n-k-1)} \sim F(q, n-k-1)$$

3 Фиктивные переменные

Dummy variable: $D_i = \{0, 1\}$

$$Y = \beta_0 + \sigma D + \beta_1 X + \varepsilon$$

Или с учетом углового коэффициента

$$Y = \beta_0 + \sigma D + \beta_1 X + \lambda N + \varepsilon$$

Гипотеза:

$$\begin{cases} H_0 : \sigma = \lambda = 0 \\ H_1 : \sigma^2 + \lambda^2 > 0 \end{cases}$$

$$\frac{(RSS_R - RSS_U)/2}{RSS_U/(n-(k+1))} \sim F(2, n-k-1)$$

Тест Чоу

1. Оцениваем модели по отдельности

$$H_0 : \beta_0' = \beta_0'', \dots, \sigma^2_{\varepsilon'} = \sigma^2_{\varepsilon''}$$

2. Два типа теста Чоу

(a) $F = \frac{(RSS_R - RSS_U)/(k+1)}{RSS_U/(n-2(k+1))}$

(b) $\frac{(RSS_p - [RSS_1 + RSS_2])/(k+1)}{[RSS_1 + RSS_2]/(n-2(k+1))}$

(c) RSS_p - по всем наблюдениям, RSS_1, RSS_2 - по частям данных

Dummy для m-градаций
Разбиваем каждый m на dummy.
При этом нужно брать m - 1 dummy, поскольку иначе сумма дамми переменных будет коллинеарна с β_0 столбцом

Part II
Модуль 3

4 Гетероскедастичность

- Оценки несмещенные
- Несостоятельные
- Неэффективные
- ТГМ не выполняется $\rightarrow$ МНК-оценки не являются BLUE
- Гипотезы не работают

4.1 Тест Голдфелда-Квандта

$$\begin{cases} H_0 : \text{Гомоскедастичность } (\sigma_i^2 = \sigma_\varepsilon^2) \\ H_1 : \text{Гетероскедастичность } (\sigma_i \sim X_{ji}.X_j) \end{cases}$$

1. Упорядочить наблюдения
2. Разделить наблюдения на 3 части
3. Отдельно оценить регрессии и сохранить RSS
4. $F(n_2 - k, n_1 - k) = \frac{RSS_2/(n_2-k)}{RSS_1/(n_1-k)}$

4.2 Тест Глейзера

$$\begin{cases} H_0 : \sigma_i^2 = \sigma_u^2 \\ H_1 : \sigma_i \sim X^\gamma, \gamma = \{\gamma = 1, \gamma = 1/2, \gamma = -1\} \end{cases}$$

1. Сохраняются остатки
2. Если β значима хотя бы для одной из регрессий, то имеем гетероскедастичность:
   (a) $|e_i| = \alpha + \beta X_i + u_i$
   (b) $|e_i| = \alpha + \beta\sqrt{X_i} + u_i$
   (c) $|e_i| = \alpha + \beta\frac{1}{X_i} + u_i$

4.3 Тест Уайта

$$\begin{cases} H_0 : \text{Гомоскедастичность} \\ H_1 : \text{Гетероскедастичность} \end{cases}$$

Вид гетероскедастичности не специфицируется

1. Оценивается регрессия по всем наблюдениям
2. Сохраняются остатки регрессии
3. Оценивается регрессия квадратов остатков на все регрессоры, их квадраты, попарные произведения и константу
4. Находим R^2
5. $\chi^2_{m-1} = nR^2$, где m - число коэффициентов во вспомог регрессии

4.4 Тест Бройша-Пагана

Доп. факторы влияют на σ_i

$$\begin{cases} H_0 : \sigma_i^2 = \sigma_\varepsilon^2 \\ H_1 : \sigma_i^2 \sim f(\alpha_0 + \alpha_1 Z_1 + ... + \alpha_r Z_r) \end{cases}$$

1. Сохраняем остатки e_i, RSS
2. Находится оценка дисперсии возмущений
   $\hat{\sigma_u^2} = \frac{RSS}{n}$
3. Оценивается регрессия e^2 на $Z_1, ..., Z_r \rightarrow$ находим ESS
4. $\frac{ESS}{2\hat{\sigma}^4} \sim \chi^2_r$

4.5 Взвешенный метод обобщенных квадратов

1. Если известны дисперсии для каждого наблюдения
   $\sigma_i \rightarrow \frac{Y_i}{\sigma_i} = \beta_1 \frac{1}{\sigma_i} + \beta_2 \frac{X_i}{\sigma_i} + \frac{u_i}{\sigma_i}$
2. Обычно стандартные отклонения неизвестны $\rightarrow$ Достаточно знать что отклонения пропорциональны некоторой известной переменной Z_i

   $\sigma_i = \lambda Z_i$

   $\frac{Y_i}{Z_i} = \beta_1 \frac{1}{Z_i} + \beta_2 \frac{X_i}{Z_i} + \frac{u_i}{Z_i}$
3. Другой способ борьбы - *Логарифмическое преобразование данных*

4.6 Стандартные ошибки Уайта

Устойчивы к гетероскедастичности

$$\hat{\beta} = \beta + (X'X)^{-1}X'\varepsilon$$

$$nV\hat{ar}(\hat{\beta}) = (\frac{1}{n}X'X)^{-1}(\frac{1}{n}\sum_{s=1}^{n} e_s^2(x_s'x_s))(\frac{1}{n}X'X)^{-1}$$

4.7 Обобщенный метод наименьших квадратов

$$Y = X\beta + \varepsilon$$

Выполнены все условия ТГМ, кроме скалярности ковариационной матрицы ошибок регрессии

$$Var(\varepsilon) = \Omega$$

$$\Omega = C^{-1}\Lambda C, \Lambda \text{ - диагональная матрица (на диагонали - собственные числа)}$$

$$\Omega^{-1/2}Y = \Omega^{-1/2}X\beta + \Omega^{-1/2}\varepsilon$$

$$Y^* = X^*\beta + \varepsilon^*$$

$$\hat{\beta}_{GLS} = (X^{*\prime}X^*)^{-1}(X^{*\prime}Y^*) = (X'\Omega^{-1}X)^{-1}(X'\Omega^{-1}Y)$$

5 Метод максимального правдоподобия

5.1 Регрессия

$$L(\varepsilon|\beta, \sigma^2) = \frac{1}{(2\pi)^{n/2}}(\sigma^2)^{n/2} \cdot exp\left(-\frac{1}{2\sigma^2}\varepsilon'\varepsilon\right) =$$

$$\frac{1}{(2\pi)^{n/2}}(\sigma^2)^{n/2} \cdot exp\left(-\frac{1}{2\sigma^2}(Y - X\beta)'(Y - X\beta)\right)$$

$$l(\varepsilon_1, ..., \varepsilon_n|\beta_0, \beta_1, \sigma^2) = -\frac{n}{2}ln2\pi - \frac{n}{2}ln\sigma^2 - \frac{1}{2\sigma^2}(Y - \beta X)'(Y - \beta X)$$

$$\hat{\beta} = (X'X)^{-1}X'Y, \hat{\sigma}^2 = \frac{e'e}{n}$$

<u>Общие свойства оценков МП</u>

- Инвариантность
- Состоятельность
- Асимптотическая нормальность
- Асимптотическая эффективность

5.2 Тест Вальда

$$Var(\varepsilon) = \sigma_\varepsilon^2 I$$

$$\begin{cases} H_0 : Q\beta = q, rangQ = r \\ H_1 : Q\beta \neq q \end{cases}$$

$$Q\beta_{ML} \sim N(Q\beta, QVar(\hat{\beta_{ML}})Q')$$

$$W = (Q\hat{\beta} - q)'(QVar(\hat{\beta_{ML}}Q')^{-1}(Q\hat{\beta} - q) \sim \chi_r^2$$

Для функций

$$\begin{cases} H_0 : g_j(\beta) = 0 \\ H_1 : \exists j : g_j(\beta) \neq 0 \end{cases}$$

$$r = 1, g(\beta) \approx g(\hat{\beta}) + \frac{\partial g}{\partial \beta}(\hat{\beta})(\beta - \hat{\beta})$$

$$W = g'(\hat{\beta}) \left(\frac{\partial g}{\partial \beta}(\hat{\beta}) Var[\hat{\beta_{ML}}] \frac{\partial g'}{\partial \beta}(\hat{\beta}) \right) g(\hat{\beta}) \sim \chi_r^2$$

$$r > 1 \to \text{То же самое, но в матрицах}$$

Недостаток: Не инвариантен к способу параметризации

5.3 Тест отношения правдоподобия

$$\begin{cases} g_j(\beta) = 0, j = 1, ..., r \\ \exists j : g_j(\beta) \neq 0 \end{cases}$$

$$LR = -2(lnL(\hat{\beta_R}) - lnL(\hat{\beta_{UR}})) \sim \chi_r^2$$

5.4 Тест множителей Лагранжа

$$\begin{cases} H_0 : g(\beta) = \begin{pmatrix} g_1(\beta) \\ ... \\ g_r(\beta) \end{pmatrix} = 0, \\ H_1 : \exists j : g_j(\beta) \neq 0 \end{cases}$$

$$H(\beta, \lambda) = l(\beta) - \lambda' g(\beta) \to max$$

$$\frac{\partial l(\hat{\beta_R})}{\partial \beta_k} - \lambda' \frac{\partial g(\hat{\beta_R})}{\partial \beta_k} = 0$$

$$LM = \left(\frac{\partial l}{\partial \beta}(\hat{\beta_R}) \right)' I^{-1}(\hat{\beta_R}) \left(\frac{\partial l}{\partial \beta}(\hat{\beta_R}) \right) \sim \chi_r^2$$

5.5 Критерий Акаике

Расстояние Кульбака-Лейблера

$$I(f,g) = \int f(x)log\left(\frac{f(x)}{g(x|\theta)}\right)dx$$

$$I(f,g) = \int f(x)log(f(x))dx - \int f(x)log(g(x|\theta))dx$$
$$I(f,g) = E_f[log(f(x))] - E_f[log(g(x|\theta))]$$
$$I(f,g) = C - E_f[log(g(x|\theta))] \to C = \int f(x)log(f(x))dx$$

Критерий Акаике

$$E_yE_x[(log(g(x|\theta(y)))]$$
$$log(L(\hat{\theta}|data)) - K = C - \hat{E}_{\hat{\theta}}[I(f,\hat{g})]$$
$$AIC = -2log(L(\hat{\theta}|data) + 2K \to min$$

5.6 Критерий Шварца

$$BIC = -2ln(L) + Klog(n)$$

6 Модели бинарного выбора

6.1 Модель линейной вероятности

$$p(Y_i = 1) = \beta_0 + \beta_1 X_{1i} + ... + \beta_k X_{ki}$$

Недостатки

- Оцененные значения не всегда $\in$ [0, 1]
- $\varepsilon \nsim N(...)$
- Гетероскедастичность

6.2 Логит-модель

$$p = F(Z) = \frac{1}{1+e^{-Z}}$$
$$\frac{dp}{dZ} = \frac{e^{-Z}}{(1+e^{-Z})^2}$$

Для оценки параметров используется ММП

$$L(\beta) = \prod_{Y_i=1} F(\beta X) \prod_{Y_i=0}(1 - F(\beta X))$$
$$L(\beta) = \prod_{i=1}[F(\beta X)]^{Y_i}[1 - F(\beta X)]^{1-Y_i}$$
$$l(\beta) = \sum_{i=1}^{n}[Y_i lnF(\beta X) + (1 - Y_i)lnF(1 - \beta X)]$$

Условие первого порядка:

$$\sum_{i=1}^{n}[Y_i - \Lambda(\beta X_i)]X_{ji} = 0, j = 0, ..., k$$

Предельный эффект фактора:

$$\frac{\partial p}{\partial X_i} = \frac{dp}{dZ}\frac{\partial Z}{\partial X_i} = f(z)\beta_i = \frac{\varepsilon^{-Z}}{(1+e^{-Z})^2}\beta_i$$

6.3 Пробит-модель

$$f(Z) = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}Z^2}$$

$$\frac{\partial p}{\partial X_i} = \frac{dp}{dZ}\frac{\partial Z}{\partial X_i} = f(z)\beta_i = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}Z^2}\beta_i$$

6.4 Оценка качества бинарных моделей

6.4.1 Odd Ratio

$$OR = \frac{Pr(Y=1)}{Pr(Y=0)}$$

Для логит-модели $X_j \uparrow \rightarrow ln(OR) \uparrow$ на $\beta_j, OR \uparrow$ на e^{β_j}

6.4.2 R^2-МакФаддена

$$R^2_{MF} = 1 - \frac{\hat{l}}{l_0}$$

- $\hat{l}$ - Лог. функция правдоподобия в максимуме
- l_0 - Лог. функция для модели, в которую включена только константа

6.4.3 PseudoR^2

$$\text{Pseudo}R^2 = 1 - \frac{1}{1+\frac{2}{n}(\hat{l}-l_0)}$$

6.4.4 Качество подгонки модели

$$wr_1 = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$$

$$R^2_p = 1 - \frac{wr_1}{wr_0}$$

6.4.5 Выбор порога отсечения

- Sensitivity - Доля правильно идентифицированных 1
- Specificity - Доля правильно идентифицированных 0
- ROC-кривая $= \frac{Sensitivity}{1-Specificity}$

7 Стохастические регрессоры

7.1 Эндогенность

В случае стохастических регрессоров ТГМ выполняется если:

- при любой реализации матрица имеет ранг k
- $\exists \operatorname{plim}_{n\to\infty} \frac{1}{n}(X^T X)$
- $\boxed{\operatorname{plim}_{n\to\infty} \frac{1}{n} X^T \varepsilon = 0}$
  - Если это условие не выполняется $\to$ проблема эндогенности
  - Оценки не являются состоятельными и асимптотически несмещенными

7.2 Инструментальные переменные

Для переменной X переменные $Z_1, ..., Z_l$ *инструментальные*:

- Z сильно коррелируют с X
- Z не коррелируют с ошибками
  - Можно заменить более слабым условием $\operatorname{plim}_{n\to\infty} \frac{1}{n} Z_i \varepsilon = 0$
- $\hat{\beta}_1^{\text{ИП}} = \frac{\hat{\text{Cov}}(Z,Y)}{\hat{\text{Cov}}(Z,X)}$
- m = k $\to \hat{\beta}^{\text{ИП}} = (Z^T X)^{-1} Z^T Y$
- m < k $\to$ Двухшаговый МНК

7.2.1 Двухшаговый МНК

1. Оцениваем $X_j = \alpha_0 + \alpha_1 Z_1 + ...$
   (a) Проекция каждого вектора X в пространство Z
   (b) $\boxed{\hat{X}_j = Z(Z^T Z)^{-1} Z^T X}$
2. Оцениваем $Y = \beta_0 + \beta_1 \hat{X}_1 + ...$
   (a) Каждый вектор X заменяется на свой инструмент $\hat{X}$
   (b) $\boxed{\hat{\beta} = (X^T Z(Z^T Z)^{-1} Z^T X)^{-1} (X^T Z(Z^T Z)^{-1} Z^T Y)}$

7.3 Тест Хаусманна

Определяет проблему эндогенности

- H_0 : Все регрессоры экзогенны
- H_1 : Имеет место проблема экзогенности

$$H = (\hat{\beta}^{\text{ИП}} - \hat{\beta}^{\text{МНК}})^T (\hat{\mathrm{V}}(\hat{\beta}^{\text{ИП}}) - \hat{\mathrm{V}}(\hat{\beta}^{\text{МНК}}))^{-1})(\hat{\beta}^{\text{ИП}} - \hat{\beta}^{\text{МНК}}) \sim \chi^2_{k+1}$$

Тест Ву-Хаусманна

- H_0 : X_1и ε не коррелируют
- H_1 : X_1и ε коррелируют

1. Регрессия всех переменных из X_1 на Z, сохраняем остатки v_j
2. Оцениваем регрессию с учетом остатков $Y = \beta_0 + \beta_1 X_1 + ... + \gamma_1 \hat{v}_1 + ...$
   (a) $H_0 : \gamma_1 = ... = \gamma_k = 0$
   (b) $H_1 : \exists \gamma_1^2 + ... > 0$

Оба теста *асимптотически* дают одинаковые результаты

8 Обобщенный метод моментов

Моментных тождеств берется больше по сравнению с обычным методом моментов

- Берем не равенство моментов, а разницу между выборочным и теоретическим g_i
- Минимизируем разности $\sum_{i=1}^n w_j g_j^2, w_j \propto \frac{1}{var(g_j)}$
- *В общем случае*: $g^T W g \to min$
  - Лучшая матрица W: $W_{opt} = (Var(g(\hat{\theta}_{GMM})))^{-1}$
  - Но θ_{GMM} мы не знаем
- Итерационная процедура
  1. $\sum_{i=1}^n g_j^2 \to min$
     (a) $Var^{-1}(q) = W$
  2. $g^T W g \to min$
     (a) С помощью параметров находим новую W
  3. Повторяем до сходимости
- Стандартный метод инструментальных переменных является частным случаем ОММ
- Если у нас есть L инструментов: $g_i(\beta) = Z_i \varepsilon_i$
- Если инструменты экзогенны, то $E(g_i(\beta)) = 0$ - *Теоретическое тождество* (условие ортогональности)

- *Эмпирическое тождество*: $\bar{g}(\beta) = \frac{1}{N} Z^T \varepsilon$
- Решаем уравнение $\bar{g}(\beta) = 0$
- Если количество инструментов = количество регрессоров → Оценки однозначны и *совпадают с методом инструментальных переменных*
- Если инструментов >, чем регрессоров → в рамках ОММ оптимизируют квардатичную форму $J(\beta) = N(\bar{g}(\beta))^T W \bar{g}(\beta) \to min$
  - Из условия первого порядка $\frac{\partial J(\beta)}{\partial \beta} \to \hat{\beta}_{OMM} = (X^T ZWZ^T X)^{-1} X^T ZWZ^T Y$
  - В зависимости от весовой матрицы W может быть множество оценок

$$(Z^T\varepsilon)^T W(Z^T\varepsilon) \to (Z^T(Y - X\beta))^T W(Z^T(Y - X\beta)) \to$$
$$(Y^T - \beta^T X^T)ZWZ^T(Y - X\beta) =$$
$$Y^T ZWZ^T Y - \beta^T X^T ZWZY^T - Y^T ZWZ^T X\beta + \beta^T X^T ZWZ^T X\beta =$$
$$-2\beta^T X^T ZWZ^T Y + \beta^T X^T ZWZ^T X\beta = 0 \to$$
$$\hat{\beta}_{GMM} = (X^T ZWZ^T X)^{-1} X^T ZWZ^T Y$$

  - $A = (X^T ZWZ^T X)^{-1} X^T ZWZ^T \to Var(\hat{\beta}_{OMM}) = AVar(Y)A^T$
- Выбор оптимальной весовой матрицы
  - Пусть $Var(\varepsilon) = \Omega$
  - $S = \frac{1}{N} E(Z^T \varepsilon\varepsilon^T Z) = \frac{1}{N} E(Z^T \Omega Z)$
  - $W_{opt} = S^{-1}$ → наиболее эффективные оценки ОММ
  - Оценивание матрицы Ω
    - Гомоскедастичность $\Omega = \sigma^2 I$ $S = \frac{\sigma^2}{N} E(Z^T Z)$
    - Гетероскедастичность $\Omega \neq \sigma^2 I$
      - Оцениваем исходное уравнение методом инструментальных переменных
      - На основании остатков $\hat{\varepsilon} = Y - X\hat{\beta}_{IV} \to \hat{\Omega}$
      - $\hat{\Omega}$ - матрица квадратов остатков
      - Можно итерационно подбирать β
- Достоинства и недостатки ОММ
  - Достоинства
    - В отсутствие гетероскедастичности асимптотически не хуже, чем метод инструментальных переменных
    - В случае гетероскедастичности лучше, чем метод инструментальных переменных
  - Недостатки
    - Неэффективно на маленьких выборках

8.1 Тестирование качества инструментов

- Проверка коррелированности эндогенных регрессоров и инструментов
- Если $Y = X_1\beta_1 + X_2\beta_2 + \varepsilon, X_1 и \varepsilon$ коррелируют
  - Z - инструменты
  - Строим регрессию X_1 на Z и посмотреть на R^2 и F-stat > 10, иначе инструменты слабые
  - Проверка экзогенности инструментов
    - *Тест Хансена* $J(\hat{\beta}) = N(\bar{g}(\hat{\beta})^T)\hat{S}^{-1}\bar{g}(\hat{\beta}) \sim \chi^2_{L-K}$
    - При гетероскедастичности $J(\hat{\beta}) = \hat{\varepsilon}^T Z(Z^T\hat{\Omega}Z)^{-1}Z^T\hat{\varepsilon} \sim \chi^2_{L-K}$
    - При гомоскедастичности (*Тест Саргана*) $J(\hat{\beta}) = \frac{1}{\hat{\sigma^2}_\varepsilon}\hat{\varepsilon}^T Z(Z^T Z)^{-1}Z^T\hat{\varepsilon} \sim \chi^2_{L-K}$

9 Системы одновременных уравнений

Пример - модель спроса и предложения

$$\begin{cases} q_t^S = \alpha P_t + \varepsilon_t \\ q_t^D = \beta P_t + \gamma In_t + u_t \end{cases}$$

Подбираем параметры α, β, γ
Если оценить по отдельности, то получим смещенные оценки коэффициентов

$$q_t^S = q_t^D \to \alpha P_T + \varepsilon_t = \beta P_t + \gamma In_t + u_t \to P_t = \frac{\gamma In_t + u_t + \varepsilon_t}{\alpha - \beta}$$

Цена связана с ошибками в обоих уравнениях $\to$ эндогенность
Подставляем в (1)

$$\begin{cases} q_t = \frac{\alpha\gamma}{\alpha-\beta}In_t + \frac{\alpha(u_t-\varepsilon_t)+(\alpha-\beta)\varepsilon_t}{\alpha-\beta} \\ P_t = \frac{\gamma In_t+u_t+\varepsilon_t}{\alpha-\beta} \end{cases}$$

В этих уравнениях нет проблем эндогенности:

$$\pi_1 = \frac{\gamma}{\alpha - \beta}, \pi_2 = \frac{\alpha\gamma}{\alpha - \beta}$$

$$\hat{\pi_1}, \hat{\pi_2} \to \hat{\alpha} = \frac{\hat{\pi_1}}{\hat{\pi_2}}$$

Альтернатива:
Использовать инструментальную переменную дохода вместо цен в (1)

$$\hat{\alpha}_{IV} = \frac{q^T In}{p^T In}$$

Можем однозначно найти только α

9.1 Общий случай СОУ

- Разделяем все переменные на эндогенные $(Y_1, ..., Y_m)$, экзогенные $(X_1, ..., X_k)$
- У каждого Y свое уравнение, у каждого X свой параметр
- *Структурная форма СОУ*

$$\begin{cases} \beta_{11}Y_{1t} + ... + \beta_{1m}Y_{mt} + \gamma_{11}X_{1t} + ...\gamma_{1k}X_{kt} = \varepsilon_{1t} \\ \beta_{21}Y_{1t} + ... + \beta_{2m}Y_{mt} + \gamma_{21}X_{1t} + ...\gamma_{2k}X_{kt} = \varepsilon_{2t} \\ ... \\ \beta_{m1}Y_{1t} + ... + \beta_{mm}Y_{mt} + \gamma_{m1}X_{1t} + ...\gamma_{mk}X_{kt} = \varepsilon_{mt} \end{cases}$$

- В матричной форме

$$Y = \begin{pmatrix} Y_1 \\ ... \\ Y_m \end{pmatrix}, X = \begin{pmatrix} X_1 \\ ... \\ X_m \end{pmatrix}, B = \begin{pmatrix} \beta_{11} & ... & \beta_{1m} \\ ... & & \\ \beta m1 & ... & \beta_{mm} \end{pmatrix}, \Gamma = \begin{pmatrix} \gamma_{11} & ... & \gamma_{1k} \\ ... & & \\ \gamma k1 & ... & \gamma_{kk} \end{pmatrix}$$

$$BY_t + \Gamma X_t = \varepsilon_t$$

$$Y_t = -B^{-1}\Gamma X_t + B^{-1}\varepsilon_t \text{- приведенная форма} \to \Pi = -B^{-1}\Gamma$$

$$Y = \Pi X_t + v_t$$

В структурной форме $m^2 - m + mk \to$ в общем случае не решаемо, некоторые коэффициенты могут быть нулевыми и тогда получится

Будем считать, что $\exists$ q Y и p X с ненулевыми коэффициентами

Y_* - ненулевые, Y_{**} - нулевые

X_x - ненулевые, X_{xx} - нулевые

$$\beta_*^T Y_{*t} + \gamma_x^T X_{xt} = \varepsilon_{1t}$$

$$\begin{pmatrix} Y_{*t} \\ Y_{**t} \end{pmatrix} = \begin{pmatrix} \Pi_{*x} & \Pi_{*xx} \\ \Pi_{**x} & \Pi_{**xx} \end{pmatrix} \begin{pmatrix} X_{xt} \\ X_{xxt} \end{pmatrix} + v_t$$

$$B\Pi = -\Gamma \to \begin{pmatrix} \beta_*' & 0 \end{pmatrix} \begin{pmatrix} \Pi_{*x} & \Pi_{*xx} \\ \Pi_{**x} & \Pi_{**xx} \end{pmatrix} = -\begin{pmatrix} \gamma_x' & 0 \end{pmatrix}$$

$$\beta_*' \Pi_{*xx} = 0$$

Левая часть уравнения размером (k - p), правая - (q - 1)

<u>Необходимое условие идентификации</u>

$$k - p \geq (q - 1) \to \text{можем выразить } \beta$$

$$(k - p) + (m - q) \geq m - 1$$

Число нулевых коэффициентов в уравнении $\geq$ число уравнений - 1

<u>Необходимое и достаточное условие</u>

$$rank\Pi_{*xx} = q - 1$$

- *Виды уравнений*
  - k - p = q - 1 → точно идентифицируемое
    - Косвенный метод наименьших квадратов
    - Оцениваем уравнения приведенной формы и из них выражаем уравнения структурной формы
  - k - p > q - 1 → сверх идентифицируемое
    - Применяется двухшаговый МНК
    - Каждый Y (кроме Y с коэффициентом 1) заменяется на оценку Y из уравнения регрессии Y на все X

9.2 Трехшаговый МНК

Общий вид системы уравнений

$$\begin{pmatrix} y_1 \\ ... \\ y_M \end{pmatrix} = \begin{pmatrix} Z_1 & 0 & ... & 0 \\ ... & ... & ... & ... \\ 0 & 0 & ... & Z_M \end{pmatrix} \begin{pmatrix} \beta_1 \\ ... \\ \beta_M \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ ... \\ \varepsilon_M \end{pmatrix}$$

$Z_1, ..., Z_M$ - включают экзогенные и эндогенные переменные
$E(\varepsilon) = 0$
$E(\varepsilon\varepsilon') = \Sigma$
В трехшаговом МНК находим оценку Σ и потом применяем обобщенный МНК

1. Инструментирование всех эндогенных переменных всеми экзогенными

   $\hat{z}_i = X(X'X)^{-1}X'z_i$

2. Каждый Y в уравнении, кроме Y с коэффициентом 1, заменяется на оценку Y из уравнения регрессии на все X и оценивается каждое уравнение регрессии

3. (a) Сохраняем остатки
   (b) Составляем из них матрицу E
   (c) $\hat{\Sigma} = \frac{E'E}{n}$
   (d) $\hat{B} = \left[\hat{Z}'(\Sigma^{-1} \otimes I)\right]^{-1} \hat{Z}'(\Sigma^{-1} \otimes I)y$
   (e) $V_{\hat{B}} = (\hat{Z}'(\Sigma^{-1} \otimes I)\hat{Z})^{-1}$
   (f) Если известно $var(\varepsilon) = \Omega$, то обобщенный метод наименьших квадратов более эффективен
   $\hat{\beta}_{FGLS} = (X'\hat{\Omega}^{-1}X)^{-1}X'\hat{\Omega}Y$
   $\hat{\Omega} = \hat{\Sigma} \otimes I$
   $Var(\hat{\beta}_{FGLS} = (X'\hat{\Omega}^{-1}X)^{-1}$

Кронекерово произведение (тык)

- $\Sigma \otimes I_N = \begin{pmatrix} \Sigma & ... & ... \\ ... & \Sigma & ... \\ ... & ... & \Sigma \end{pmatrix}$

9.3 SUR. Внешне не связанные уравнения

Справа только X $\rightarrow$ Применить OLS?

Считаем, что эпсилоны в разных уравнениях могут быть связаны (внешние шоки для внутренних Y) $\rightarrow$ Трехшаговый МНК без первого шага

Формулы те же, заменяем Z на X

10 Модели множественного выбора

- $OR = \frac{P(Y=1)}{P(Y=0)} \rightarrow ln(OR) = \beta_0 + \beta_1 X_1 + ...$

10.1 Модели упорядоченного множественного выбора

- Y упорядочен по какому-то критерию (согласен, скорее согласен, ...)
- $Y_i = \{1, 2, ..., m\}$
- $Y_i^* = x_i'\beta + \varepsilon_i$
- $Y_i = j, \text{ if } c_{j-1} < Y_i^* < c_j, j = 1, ..., m$
- $c_0 = -\infty, ..., c_m = \infty$
- С помощью оценки метода правдоподобия оцениваем β, c
- $L = \prod_{j=1}^{m} \prod_{i:Y_i=j} (F(c_j - x_i'\beta) - F(c_{j-1} - x_i'\beta)) \rightarrow max_{\beta,c}$

10.1.1 Гипотеза о параллельности

- $P(Y_i = j) = F(c_j - x_i'\beta) - F(c_{j-1} - x_i'\beta)$
- Проверить, что Y_i принимает значение не больше k
- Просуммировать все вероятности $Y_i \leq k$ по k
- $P(Y_i \leq k|X) = F(c_k - x_i'\beta), k = 1, ..., m$
- Тест Бранта

10.1.2 Отношение шансов

- $\frac{P(Y_i \leq k|X)}{P(Y_i > k|X)} = exp(c_k - x_i'\beta) \rightarrow \frac{P(Y_i > k|X)(X, x_j+1)}{P(Y_i \leq k)|X)(X, x_j+1)} = exp(\beta_j)$

10.1.3 Предельные эффекты

- $\frac{\partial P(Y=j)}{\partial X_k} = -\beta_k f(c_1 - (X\beta))$
- $\frac{\partial P(Y=j)}{\partial X_k} = \beta_k f(c_{m-1} - (X\beta))$

10.2 Мультиномиальные модели

- Ответы не упорядочены
- U_{ij} - полезность j-ой альтернативы для i-го индивида
- $P\{Y_i = j\} = P\{U_{ij} = max\{U_{i1}, ..., U_{iM}\}\}$
- $U_{ij} = x_i'\beta_j + \varepsilon_{ij}$
- Для нормального распределения аналитическое решение не находится
- Задача допускает аналитическое решение, если e_{ij} независимы и имеют функцию распределения $F(x) = exp\{-exp(-x)\}$
- $P(Y_i = 1) = \frac{1}{1+exp(x_i\beta 2)+...+exp(x_i\beta m)}$
- $P(Y_i = j) = \frac{exp(x_i\beta_j}{1+exp(x_i\beta 2)+...+exp(x_i\beta m)}$
- $\frac{P(Y_i=j}{P(Y_i=k)} = \frac{exp(x_i\beta_j}{exp(x_i\beta_k} = exp(x_i'(\beta_j - \beta_k))$
- Сильно предположение о независимости альтернатив

11 Тобит, Sample selection models

11.1 Тобит

- $Y_i = \begin{cases} Y_i^*, \text{ if } Y_i^* > 0 \\ 0, \text{ if } Y_i^* \leq 0 \end{cases}$
- $Y_i^* = x_i^*\beta + u_i$
- $P(Y_i = 0) = P(Y_i^* \leq 0) = P(u_i \leq -x_i'\beta) = P(\frac{u_i}{\sigma_u} \leq \frac{-x_i'\beta}{\sigma_u}) = 1 - \Phi(\frac{-x_i'\beta}{\sigma_u})$
- $f(y|Y \geq c) = \frac{f(y)}{P(Y \geq c)}$, if y $\geq$ c and 0 otherwise
- $E(Y_i|Y_i > 0) = x_i'\beta + \sigma\frac{\phi(x_i'\beta/\sigma)}{\Phi(x_i'\beta/\sigma)}$
- $\frac{\partial E(Y_i}{\partial X_j} = \Phi(x_i'\beta/\sigma)\beta_j$
- $L(\beta, \sigma^2) = \prod_{Y_i=0} P(Y_i = 0) \prod_{Y_i>0} \phi(Y_i)$
- Оценивается градиентным спуском

11.2 Модель Хекмана

- $Y_i^* = x_i'\beta + \varepsilon_i$
- $g_i^* = z_i\gamma + u_i$ (Модель участия)
- $g_i = \begin{cases} 1, g_i^* \geq 0 \\ 0, g_i^* < 0 \end{cases}$

- General model
$$\begin{cases} Y_i = Y_i^*, g = 1 \text{ if } g_i^* \geq 0 \\ Y_i \text{ is not observed, OTW} \end{cases}$$

- $E(Y_i|g_i = 1) = x_i'\beta + E(\varepsilon_i|g_i = 1) = x_i'\beta + E(\varepsilon_i|u_i \geq -z_i'\gamma) = E(Y_i|g_i = 1) = x_i'\beta + \sigma_{\varepsilon u}\lambda(z_i'\gamma)$

- Лямбда Хекмана: $\lambda(z_i'\gamma) = \frac{\phi(z_i'\gamma)}{\Phi(z_i'\gamma)}$

- Можем использовать разные данные для функций $\rightarrow$ более гибкая модель, чем модель Тобита

11.2.1 Оценка

1. Метод правдоподобия - не всегда сходится

2. Двухшаговая процедура (сначала g, потом Y)

12 Ядерные методы

- Ищем оценку E(Y|X = x) = m(x)

- $\int_{-\infty}^{\infty} y\hat{f}(y|x)dy$

- Составляем гистограмму X

- $$\hat{f}(x) = \frac{1}{nh}\sum_{i=1}^{n}\frac{1}{2}I(\left|\frac{x - X_i}{h}\right| \leq 1)$$

- *Свойства ядра*
  - $K(z) \geq 0$
  - $\int_{-\infty}^{\infty} K(z)dz = 1$
  - $\int_{-\infty}^{\infty} zK(z)dz = 0$
  - $\int_{-\infty}^{infty} z^2K(z)dz < \infty$

- Обычно ядра имеют выпуклую форму на промежутке $[-1, 1]$

- $$\hat{f}(x) = \frac{1}{nh}\sum_{i=1}^{n}\frac{1}{2}I(\left|\frac{x - X_i}{h}\right| \leq 1) = \frac{1}{n}\sum_{i=1}^{n} K_h(x - X_i), K_h(\cdot) = \frac{K(\cdot/h)}{h}$$

- Расчет оптимального окна производится через интегрирование MSE по h

- $h_{opt} = \left(\frac{||K||_2^2}{||f^*||_2^2(\mu_2(K))^2 n}\right)^{\frac{1}{5}} \sim n^{-\frac{1}{5}}$

- *Rule of Thumb:*
$$\hat{h}_{rot} = 1.06\hat{\sigma}n^{-\frac{1}{5}}$$

12.1 Ядерная оценка регрессии

Nadaraya-Watson Estimator

$$\hat{m}_k(x) = \frac{n^{-1}\sum_{i=1}^{n} K_h(x - X_i)Y_i}{n^{-1}\sum_{i=j}^{n} K_h(x - X_j)}$$

Среднее по Y в выбранном окне

Part III
Модуль 4

13 Временные ряды

- Данные упорядочены
- Измерения должны быть в каждый момент времени
- Данные могут быть различной частотности
- Временной ряд $\{X_t\}_{t=-\infty}^{\infty}$ - это последовательность случайных величин, заданных на одном вероятностном пространстве
  - Реализация части этой последовательности тоже называется временным рядом

Компоненты временного ряда

- *Тренд* – Временной ряд с трендом $\rightarrow$ Y монотонно изменяется со временем
- *Сезонная компонента* – Повторяющиеся паттерны
- *Циклическая компонента* – Циклы повторяются не через равные промежутки времени
- *Случайная компонента* – Колебания вне циклов

Виды временных рядов

- *Стационарный ряд* – $F(X_{t_1}, ..., X_{t_m}) = F(X_{t_{1+k}}, ..., X_{t_{m+k}})$ для любых моментов m и для любого сдвига k
  - Слишком жесткое требование
- *Стационарный ряд (в широком смысле)*
  - $E(X_t) = \mu, \forall t$
  - $Var(X_t) = \sigma^2, \forall t$
  - $cov(X_t, X_{t+s})$ зависит только от s
  - Пример - *белый шум*: $X_t = \varepsilon_t, \varepsilon_t \sim iid(0, \sigma_\varepsilon^2)$
  - *Линейный временной ряд*

    - $X_t = \mu + \sum_{i=0}^{\infty} \alpha_i \varepsilon_{t-i}$
    - $\alpha_0 = 1, \varepsilon_t \sim iid(0, \sigma_\varepsilon^2)$
  - *AR модель*
    - $X_t = \beta_1 X_{t-1} + \varepsilon_t$
    - $X_t = \beta_0 + \beta_1 X_{t-1} + \varepsilon_t$
    - <u>Проверка стационарности</u>
      - $E(X_t) = \beta_1^t X_0 \to 0$
      - $\sigma_{X_t}^2 = \frac{1-\beta_1^{2t}}{1-\beta_1^2}\sigma_\varepsilon^2 \to \frac{1}{1-\beta_1^2}\sigma_\varepsilon^2, |\beta_1| \leq 1$
      - $cov(X_t, X_{t+s}) = \frac{\beta_1^s}{1-\beta_1^2}\sigma_\varepsilon^2$
  - *Нестанционарный временной ряд*
    - $E(\varepsilon_t) = 0, \forall t$
    - $var(\varepsilon_t) = \sigma_\varepsilon^2, \forall t$
    - $cov(\varepsilon_t, \varepsilon_s) = 0, t \neq s$
    - <u>Random Walk</u>
      - $X_t = X_{t-1} + \varepsilon_t$
      - $X_t = X_{t-2} + \varepsilon_{t-1} + \varepsilon_t$
      - $X_t = X_0 + \varepsilon_1 + ... + \varepsilon_t$
      - $var(X_t) = t\sigma_\varepsilon^2$
    - <u>Случайное блуждание с дрейфом</u>
      - $X_t = \mu + X_{t-1} + \varepsilon_t$
    - <u>Проверка станционарности</u>
      - $\beta_1 > 1 \to$ большая дисперсия, взрывной процесс
  - Нестанционарные временные ряды типа TS
    - *TS* - Этот ряд становится станционарным после выделения тренда
      - $X_t = \beta_0 + \beta_1 t + \beta_2 X_{t-1} + \varepsilon_t, -1 < \beta_2 < 1$
  - Нестанционарные временные ряды типа DS
    - Ряд становится станционарным только в разностях
    - $\Delta X_t = X_t - X_{t+1} = \beta_0 + \varepsilon_t \to E(\Delta X_t) = \beta_0$

Тесты на станционарность рядов

- Автокорреляционная функция (ACF)

  $$\rho_k = \frac{E((X_t - \mu_X)(X_{t+k} - \mu_X))}{\sqrt{E((X_t - \mu_X)^2 E((X_{t|k})^2))}}$$

  $$\rho_k = \frac{\sum((X_t - \bar{X})(X_{t+k} - \bar{X})}{\sqrt{(\sum(X_t - \bar{X})^2 \sum(X_{t+k} - \bar{X})^2)}}$$

  График корреляций по k называется *Коррелограмма* $\to$ Для станционарного ряда быстро убывает

- Частная автокорреляционная функция (PACF)

  PACF(k) вычисляется как МНК оценка коэффициента β_k в регрессии $X_t = \beta_0 + ...\beta_k X_{t-k} + \varepsilon_t$

  Частные автокорреляционные функция для станционарных процессов тоже быстро убывают

- Unit root test

$$\begin{cases} H_0 : \beta_1 = 1 \\ H_1 : \beta_1 < 1 \end{cases}$$

H_0 - нестанционарность

$$\Delta X_t = \delta X_{t-1} + \varepsilon_t$$

$$\begin{cases} H_0 : \delta = 0 \\ H_1 : \delta < 0 \end{cases}$$

$$t = \frac{\delta}{s.e.(\delta)}$$

Распределение статистики Дики-Фуллера не совпадает со Стьюдентом - нужно тау-распределение

H_0 отвергается, если $t < \tau_0^{cr}$, без β_0

H_0 отвергается, если $t < \tau_\mu^{cr}$, с β_0

H_0 отвергается, если $t < \tau_\tau^{cr}$, с детрендированием

- Расширенный тест Дики-Фуллера

$$\Delta X_t = \beta_0 + \delta X_{t-1} + \beta_2 t + \sum_{i=1}^{m} \alpha_i \Delta X_{t-i} + \varepsilon_t$$

$$\begin{cases} H_0 : \delta = 0 \\ H_1 : \delta < 0 \end{cases}$$

Тестовые статистики не изменяются

- KPSS тест

$$y_t = \beta t + r_t + \varepsilon_t$$

$$r_t = r_{t-1} + u_t$$

$$\begin{cases} H_0 : \sigma_u^2 = 0 \text{ (Станционарность)} \\ H_1 : \sigma_u^2 \neq 0 \end{cases}$$

$$S_t = \sum_{s=1}^{t} e_s$$

$$KPSS = \sum_{s=1}^{T} S_t^2 / \hat{\sigma}^2$$

$$\hat{\sigma}^2 = \lim_{T \to \infty} T^{-1} E(S_T^2)$$

13.1 Процессы

Теорема Вольда

Если X_t - стационарный ряд, то его можно представить в виде:

$$X_t = d_t + \sum_{\tau=0}^{\infty} \alpha_\tau \varepsilon_{t-\tau}$$

- d_t - предсказуемый случайный процесс
- ε_t - белый шум
- $\sum_{\tau=0}^{\infty} \alpha_\tau$ - слагаемые не коррелируют

Процессы AR

$$y_t = Y_t - \mu$$

$$AR(p) : y_t = \theta_1 y_{t-1} + \theta_2 y_{t-2} + ... + \theta_p y_{t-p}$$

Лаговый оператор:

$$L^S(Y_t) = Y_{t-S}$$

$$AR(p) : \theta(L) = 1 - \theta_1 L - \theta_2 L^2 - ... - \theta_p L^p$$

MA процессы

$$MA(q) : y_t = \varepsilon_t + \alpha_1 \varepsilon_{t-1} + ... + \alpha_q \varepsilon_{t-q}$$

Процесс всегда стационарный:

$$(E(\varepsilon_t) = 0)$$

<u>Стационарность процесса AR(1)</u>

$$AR(1) : y_t = \theta y_{t-1} + \varepsilon_t \rightarrow (1 - \theta L)^{-1}(1 - \theta L) y_t = (1 - \theta L)^{-1} \varepsilon_t$$

$$(1 - \theta L)^{-1} = \sum_{j=0}^{\infty} \theta^j L^j, |\theta < 1|$$

$|\theta| < 1$ - условие стационарности процесса AR(1)

$$y_t = (1 - \theta L)^{-1} \varepsilon_t$$

$$y_t = \sum_{j=0}^{\infty} \theta^j \varepsilon t - j \rightarrow AR(1) \leftrightarrow MA(\infty), |\theta| < 1$$

Станционарность процесса AR(2)

$$AR(2): y_t = \theta_1 y_{t-1} + \theta_2 y_{t-2} + \varepsilon_t$$

$$(1 - \theta L - \theta_2 L^2) y_t = \varepsilon_t \to (1 - \phi_1 L)(1 - \phi_2 L) y_t = \varepsilon_t$$

$$(1 - \phi_1 L), (1 - \phi_2 L) \text{ - должны быть обратимы}$$

Условие станционарности:

$$|\phi_1| > 1, |\phi_2| < 1$$

Обратное характеристическое уравнение:

$$(1 - \phi_1 z)(1 - \phi_2 z) = 0 \to z_1 = \frac{1}{\phi_1}, z_2 = \frac{1}{\phi_2}$$

При обратимости процесса AR(2):

$$|z_i| > 1$$

$$y_t = \theta_1 y_{t-1} + \theta_2 y_{t-2} + \varepsilon_t$$

Прямое уравнение:

$$\lambda^2 - \theta_1 \lambda - \theta_2 = 0$$

При станционарности:

$$|\lambda_i| < 1$$

ARMA процессы

ARMA(p, q):

$$y_t = \theta_1 y_{t-1} + ... + \theta_p y_{t-p} + \varepsilon_t + \alpha_1 \varepsilon_{t-1} + ... + \alpha_q \varepsilon_{t-q}$$

$$\theta(L) y_t = \alpha(L) \varepsilon_t$$

Если корни обратного характеристического уравнения $\theta(z) = 0$ удовлетворяют условию $|z_j| > 1 \forall j = 1, ..., p \leftrightarrow$ Корни прямого характеристического уравнения удовлетворяют условию $|\lambda_i| < 1, i = 1, ..., p$

Обратимость MA процесса

$$MA(1): y_t = \varepsilon_t + \alpha \varepsilon_{t-1}$$

$$y_t = (1 - \alpha L) \varepsilon_t$$

$$(1 + \alpha L)^{-1} = \sum_{j=0}^{\infty} (-\alpha)^j L^j$$

$$y_t = \alpha \sum_{j=0}^{\infty}(-\alpha)^j y_{t-j-1} + \varepsilon_t$$

$$MA(q) : y_t = \alpha(L)\varepsilon_t$$

Необходимое условие $AR(\infty)$ представления:
Обратимость $\alpha(L)$. Корни прямого характеристического уравнение для MA части должны быть меньше 1 по модулю

13.2 Диагностика моделей

13.2.1 ACF, PACF

ACF:

$$\rho_k = \frac{cov\{Y_t, Y_{t-k}\}}{var\{Y_t\}}$$

PACF:
PACF(k) - чистая корреляция между Y_t и Y_{t-k}. Вычисляется как оценка МНК параметра

<u>AR(1)</u>

ACF:

$$AR(1) : y_t = \theta y_{t-1} + \varepsilon_t$$

$$y_t = \sum_{j=0}^{\infty} \theta^j \varepsilon_{t-j}$$

$$\rho_k = \theta^k$$

PACF:

$$AR(1) : y_t = \theta y_{t-1} + ... + 0 y_{t-p} + \varepsilon_t$$

$$PACF = \begin{cases} \theta, k = 1 \\ 0, k > 1 \end{cases}$$

<u>AR(2)</u>

Для станционарного процесса:

$$\mu = \delta/(1 - \theta_1 - \theta_2)$$

$$y_t = Y_t - \mu$$

$$y_t y_t = \theta_1 y_t y_{t-1} + \theta y_t y_{t-2} + y_t \varepsilon_t$$

$$E(y_t y_t) = \theta_1 E(y_t y_{t-1}) + \theta E(y_t y_{t-2}) + E(y_t \varepsilon_t)$$

$$\gamma_0 = \theta_1 \gamma_1 + \theta_2 \gamma_2 + \sigma_\varepsilon^2$$

Аналогично: домножаем на t - 1, t - 2, t - 3

$$\gamma_1 = \theta_1\gamma_0 + \theta_2\gamma_1$$
$$\gamma_2 = \theta_1\gamma_1 + \theta_2\gamma_0$$
$$\gamma_3 = \theta_1\gamma_2 + \theta_2\gamma_1$$

Решаем систему для гамм:

Условия станционарности

$$\gamma_1 + \gamma_2 < 1$$
$$\gamma_2 - \gamma_1 < 1$$
$$|\gamma_2| < 1$$

Делим на дисперсию (γ_0)

$$\rho_1 = \frac{\theta_1}{1-\theta_2}$$
$$\rho_2 = \frac{\theta_1^2}{1-\theta_2} + \theta_2$$

Для остальных порядков необходимо решить разностное уравнение:

$$\rho_k = \theta_1\rho_{k-1} + \theta_2\rho_{k-2}$$

Решение:

$$\rho_k = C_1\lambda_1^k + C_2\lambda_2^k$$
$$\lambda_1, \lambda_2 : \lambda^2 - \theta_1\lambda - \theta_2 = 0$$
$$\text{Для станционарного процесса} |\lambda_i| < 1$$

ACF для AR(p) exp убывающая

PACF:

$$PACF = \begin{cases} \theta_1, k = 1 \\ \theta_2, k = 2 \\ 0, k > 2 \end{cases}$$

Для AR(p) аналогично

<u>MA(p)</u>

ACF:

$$Y_t = \mu + \varepsilon_t + \alpha\varepsilon_{t-1}$$
$$\gamma_0 = var(Y_t) = (1-\alpha^2)\sigma_\varepsilon^2$$
$$\gamma_1 = cov(Y_t, Y_{t-1}) = cov(\mu + \varepsilon + \alpha\varepsilon_{t-1}, \mu + \varepsilon_{t-1} + \alpha\varepsilon_{t-2}) = \alpha\sigma_\varepsilon^2$$
$$\gamma_2 = \gamma_3 = ... = 0$$
$$\rho_1 = \frac{\alpha}{1+\alpha^2}$$

$$\rho_j = 0, j > 1$$

Аналогично для MA(q):

$$p_j = 0, j > q$$

PACF:

$$MA(1) \leftrightarrow AR(\infty)$$

$$y_t = \alpha \sum_{j=0}^{\infty} (-\alpha)^j y_{t-j-1} + \varepsilon_t$$

Если $|\alpha| < 1$: PACF является exp убывающей, аналогично для MA(q)

Процесс	**ACF**	**PACF**
AR(p)	Exp убывает	= 0 при p > k
MA(q)	= 0 при p > k	Exp убывает
ARMA(p, q)	Exp убывает	Exp убывает

Table 1: Коррелограмма процессов

Если элементы PACF, ACF не превышают $2/\sqrt{T}$ статистически неотличимы от 0

13.3 Способы оценки параметров

1. AR(q) оценивается с помощью МНК
2. MA(q), ARMA(p, q) оцениваются с помощью ММП

13.4 Критерии выбора p и q

1. Проверка, что ошибки в модели являются белым шумом
2. Информационные критерии выбора количества лагов
   (a) Критерий Акаике

   $$AIC = -2lnL + 2k \to min$$

   (b) Критерий Шварца

   $$BIC = -2lnL + (lnT)k \to min$$

   Более сильно штрафует за включение лишних лагов

13.5 ARIMA

Процесс, который становится станционарным в разностях

Y_t - нестанционарный процесс

$\Delta^d Y_t$ - станционарный процесс ARIMA

13.6 Подход Бокса-Дженкинса

1. Проверка ряда на станционарность
2. Если ряд не станционарный - находим разность при которой он является станционарным
3. Для станционарного ряда необходимо выбрать p и q с помощью ACF, PACF
4. Оценка параметров
5. Проверка остатков
6. Использование модели для прогнозирования

13.7 Современный подход

1. Изучение графика ряда - тренд, сезонность
2. Выделение тренда
3. При наличии сезонности - включение дамми переменных
4. Моделирование оставшейся модели

13.8 Автокорреляция случайной составляющей

1. Коррелированность с предыдущими значениями
2. Чаще всего встречается для временных рядов
3. Приводит к нарушению ТГМ
4. Отрицательная автокорреляция - corr < 0

$$AR(1) : \varepsilon_t = \rho\varepsilon_{t-1} + u_t$$

$$u_t \sim i.i.d$$

$$|\rho| < 1 \rightarrow$$

возмущения удовлетворяют марковской схеме первого порядка

Причины автокорреляции

1. Инертность экономических показателей
2. Ошибки спецификации модели, невключение существенных переменных
3. Сглаживание данных

Последствия автокорреляции

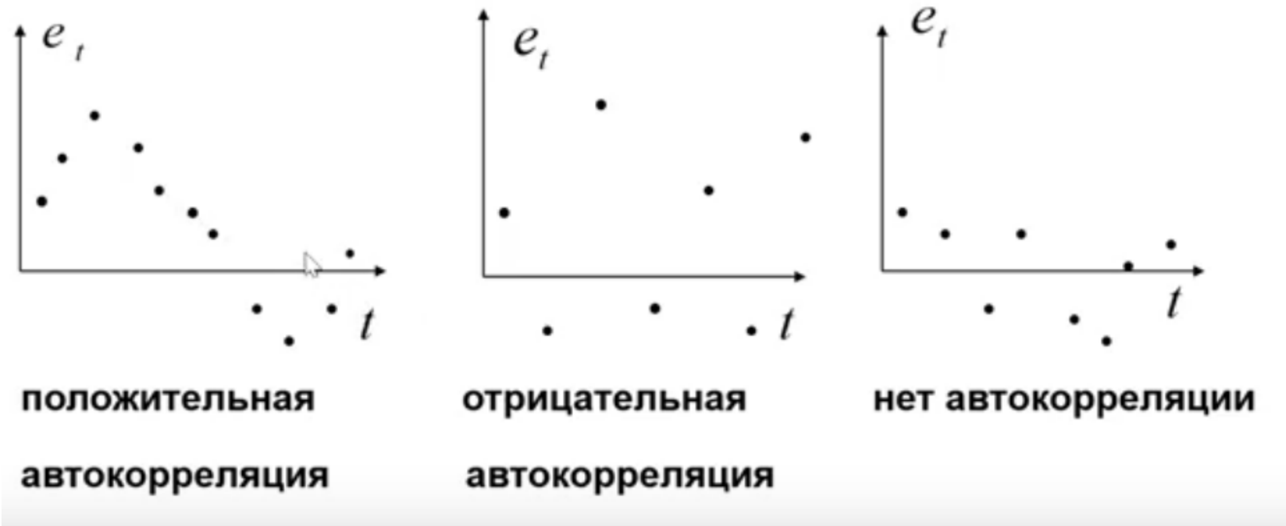
1. Оценки МНК останутся несмещенными, но не будут эффективными

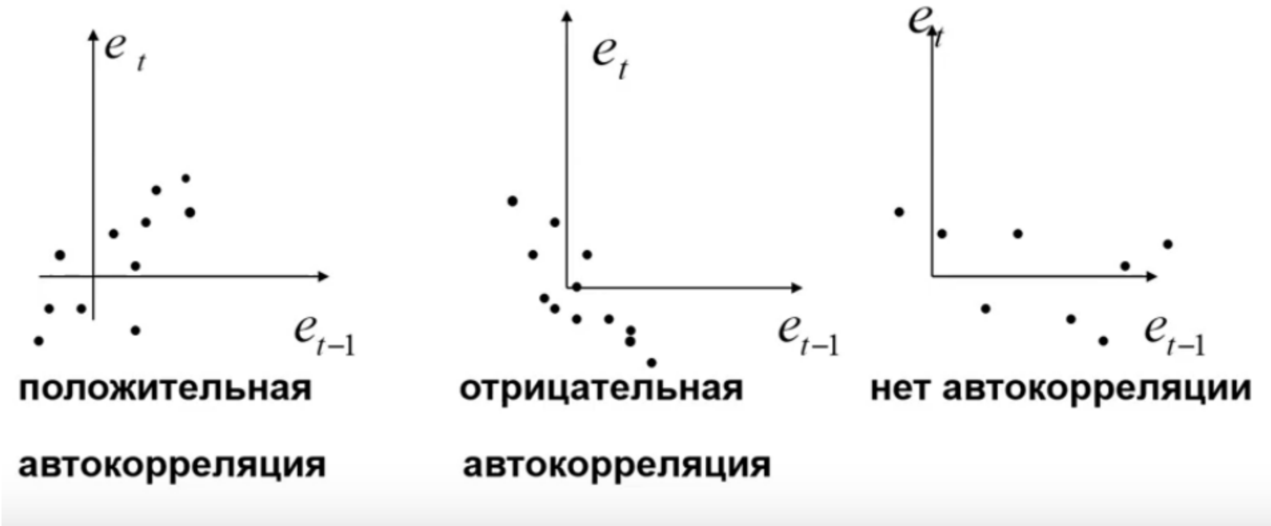
2. Оценки для стандартных ошибок коэффициентов будут заниженными

3. Статистики будут завышенными

Выявление автокорреляции

Визуальный способ выявления автокорреляции





13.8.1 Тест серий

Серия остатков - набор последовательных остатков одного знака
Если есть автокорреляция, то таких серий должно быть немного, но они достаточно длинные

Формальное описание

$$\Big\{ H_0 : \rho = 0 H_1 : \text{ Имеет место автокорреляция первого порядка}$$

1. Оцениваются параметры уравнения регрессии

2. Отмечаем знаки остатков

3. (a) n - число всех наблюдений
   (b) N_1 - число знаков +
   (c) N_2 - число знаков -
   (d) K - число серий
4. Если $K \leq K_{min}$, то имеет место положительная автокорреляция
5. Если $K \geq K_{max}$, то имеет место отрицательная автокорреляция
6. $k \sim N(\frac{2N_1N_2}{N_1+N_2} + 1; \frac{2N_1N_2(2N_1N_2-N_1-N_2)}{(N_1+N_2)^2(N_1+N_2-1)})$

13.8.2 Статистика Дарбина-Уотсона

$$d = \frac{\sum_{t=2}^{T}(e_t - e_{t-1})^2}{\sum_{t=1}^{T} e_t^2}$$

$$\begin{cases} d \to 2 \Rightarrow \rho \approx 0 \\ d \to 0 \Rightarrow \rho \approx 1 \\ d \to 4 \Rightarrow \rho \approx -1 \end{cases}$$

Надо проверять с учетом доверительных интервалов - статистики d_l и d_u мажорируют статистику d сверху независимо от параметров
Если d $< d_l \Rightarrow$ положительная автокорреляция
Если d $> d_u \Rightarrow$ нет положительной автокорреляции
Если между - неопределенность

13.8.3 Устранение автокорреляции

1. Преобразовать исходные данные
2. Использовать стандартные ошибки в форме Ньюи-Веста
3. Использовать ММП

Если мы знаем ρ - можно произвести сдвиг

$$Y_t - \rho Y_{t-1} = \beta_0(1-\rho) + \beta_1(X_t - \rho X_{t-1}) + u_t, u_t - \text{некоррелированные ошибки}$$

$$Y^* = \beta_0 + \beta_1 X_t^* + u_t$$

u_t удовлетворяют ТГМ
Теряется первое наблюдение
Поправка Прайса-Уинстона
Выражается из обобщенного метода МНК

$$Y_1^* = \sqrt{1-\rho^2}Y_1$$

$$Y_t^* = Y_t - \rho Y_{t-1}, t = 2, ..., T$$

13.8.4 Оценка параметра автокорреляции

1. Выражение из Дарбина-Уотсона
2. Процедура Кокрена-Уоркута
   (a) Оцениваем уравнение регрессии и находим остатки
   (b) Оцениваем регрессию остатков на предыдущие, получаем ρ_1
   (c) Преобразуем исходные данные
   (d) Повторяем 1 и 2, получаем ρ_2
   (e) $|\rho_1 - \rho_2| < \varepsilon \to \rho = \rho_2$
   (f) Если процедура не сходится - могли не угадать порядок автокорреляции
3. Двухшаговая процедура Дарбина
   (a)
   $$Y_t = \rho Y_{t-1} + \beta_0(1-\rho) + ...$$
   (b) Оцениваем ρ
4. Метод поиска Хилдрет-Лю на сетке
   (a) Подбираем ρ для которого RSS минимальна
5. Среди регрессоров встречается стохастический Y_{t-1}
   (a) ε удовлетворяют Марковской схеме 1-го уровня
   (b) Тогда статистика Дарбина-Уотсона не применима из-за возникновения проблемы эндогенности
   (c) **Используется h-статистика Дарбина**
   (d)
   $$\begin{cases} H_0 : \rho = 0 \\ H_1 : \rho \neq 0 \end{cases}$$
   (e) $h = \hat{\rho}\sqrt{\frac{n}{1-ns^2_{b_{Y(-1)}}}}$
6. Тест Бройша-Годфри
   (a) H_0 : нет автокорреляции возмущений
   (b) H_1 : имеет место автокорреляция порядка p
   (c)
   $$\begin{cases} H_0 : \rho_1 = \rho_2 = \ldots = \rho_p = 0 \\ H_1 : \rho_1^2 + \rho_2^2 + \ldots + \rho_P^2 \neq 0 \end{cases}$$
   (d) Проверяется гипотеза о том, что ошибки являются белым шумом
   (e) *Алгоритм*
      i. Оцениваются параметры регрессии
      $$Y_t = \beta_0 + \beta_1 X_t + \varepsilon_t$$

ii. Сохраняются остатки регрессии

iii. Оцениваются параметры регрессии

$$e_t = \beta_0 + \beta_1 X_t + r_1 e_{t-1} + \ldots r_P e_{t-P} + \varepsilon_t$$

iv. Сохраняется R^2

v. Тестовая статистика

$$\chi^2 = TR^2$$

vi. Если $\chi^2 > \chi^2_{cr}$, то гипотеза отвергается

7. Стандартные ошибки в форме Ньюи-Веста

$$var[\varepsilon] \sim \Omega = (w_{ij}), w_{ij} = 0, |i-j| > L$$

$$v\hat{a}r[\hat{\beta}] = n(X'X)^{-1}\frac{1}{T}(\sum_{s=1}^{T} e_s^2 x_s x_s' + \sum_{j=1}^{L}\sum_{t=j+1}^{T} w_j e_t e_{t-j}(x_t x_{t-j}' + x_{t-j}x_t'))(X'X)^{-1}$$

8. Q-статистика для проверки белошумности остатков

   (a) Модель ARMA(p, q)

   (b)
   $$\begin{cases} H_0 : \rho_1 = \ldots = \rho_m = 0 \\ H_1 : \rho_1^2 + \ldots + \rho_m^2 > 0 \end{cases}$$

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

   (d) r_k - выборочный коэффициент корреляции остатков ε_t и ε_{t-k}

   (e) m - гиперпараметр

   (f)
   $$Q \sim \chi^2(m-p-q)$$

   (g) Статистика Q используется и для проверки белошумности исходного ряда, тогда тестовая статистика имеет распределение $\chi^2(m)$

9. Статистика Бокса-Льюнга (для малых выборок)

   (a)
   $$Q_k = T(T+2)\sum_{k=1}^{m}\frac{1}{T-k}r_k^2$$

10. *Статистики хорошо работают, когда справа стоят экзогенные факторы, для временных рядов достаточно слабы*

13.9 Моделирование сезонности во временных рядах

13.9.1 Модели ARIMA с сезонностью

1. Включение набора дамми-переменных для каждого месяца кроме одного, чтобы избежать dummy trap

2. Использование Y(-12)

13.9.2 SARIMA

1. Мультипликативная

   (a)
   $$(1-\rho_1 L)\{\Delta ln(wpi_t)-\beta_0\}=(1+\theta_1 L)(1+\theta_{4,1}L^4)\varepsilon_t$$

   $$\Delta ln(wpi_t)=\beta_0+\rho_1\{\Delta ln(wpi_{t-1})-\beta_0\}+\theta_1\varepsilon_{t-1}+\theta_{4,1}\varepsilon_{t-4}+\theta_1\theta_{4,1}\varepsilon_{t-5}+\varepsilon_t$$

   (b) В общем виде SARIMA $(p,d,q)\times(P,D,Q)_s$:

   $$\rho(L^p)\rho_s(L^P)\Delta^d\Delta_s^D z_t=\theta(L^q)\theta_s(L^Q)\varepsilon_t$$

   $$\rho_s(L^P)=(1-\rho_{s,1}L^s-\rho_{s,2}L^{2s}-...\rho_{s,P}L^{Ps})$$
   $$\theta_s(L^Q)=(1-\theta_{s,1}L^s-\theta_{s,2}L^{2s}-...\theta_{s,P}L^{Qs})$$

   (c) Можно подбирать параметры P и Q (сезонные лаги), а не только p и q $\rightarrow$ AIC, BIC

   (d) Оценивается с помощью ММП

2. Аддитивная

   (a)
   $$\Delta ln(wpi_t)=\beta_0+\rho_1\{\Delta ln(wpi_{t-1})-\beta_0\}+\theta_1\varepsilon_{t-1}+\theta_4\varepsilon_{t-4}+\varepsilon_t$$

   $$(1-\rho_1 L)\{\Delta ln(wpi_t)-\beta_0\}=(1+\theta_1 L+\theta_4 L^4)\varepsilon_t$$

   (b) Выше модель для квартальных данных (Добавляется четвертый лаг)

3. В уравнение модели добавляются сезоннные лаги

13.9.3 Процедуры сглаживания ряда

1. STL Decomposition

2. Экспоненциальное сглаживание

   (a)
   $$\hat{Y}_{t+1|t}=\alpha Y_t+(1-\alpha)\hat{Y}_{t|t-1}$$

   (b) ETS
   $$b_t=b_{t-1}+\beta u_t$$
   $$s_t=s_{t-12}+\gamma u_t$$
   $$l_t=\ell_{t-1}+b_{t-1}+\alpha u_t$$
   $$y_t=\ell_{t-1}+b_{t-1}+s_{t-12}+u_t$$

3. Hodrick-Prescott filter

   (a) Ряд разбивается на тренд τ, циклическую компоненту c_t, ошибка ε_t

(b) Подбирается тренд компонента из

$$min_{\tau}(\sum_{t=1}^{T}(y_t - \tau_t)^2 + \lambda \sum_{t=2}^{T-1}[(\tau_{t+1} - \tau_t) - (\tau_t - \tau_{t-1})]^2)$$

$$\lambda = \begin{cases} 100, \text{ for annual data} \\ 1600, \text{ for quarterly data} \\ 14400, \text{ for montly data} \end{cases}$$

(c) Критика: Возникают смещения на концах оцениваемых интервалов

13.10 Прогнозирование с помощью временных рядов

13.10.1 Прогнозирование по модели ARMA(p, q)

По критерию MSE наилучший прогноз на момент T+1:

$$E(Y_{T+1} \mid \Omega_T)$$

Ω_T- вся информация, известная на момент времени T

$$E(\varepsilon_{T+1} \mid \Omega_T) = 0$$
$$E(\varepsilon_T \mid \Omega_T) = \varepsilon_T$$

<u>AR(1)</u>

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \varepsilon_t$$
$$\hat{Y}_{T+1} = E(Y_{T+1} \mid \Omega_T) = \beta_0 + \beta_1 Y_T$$
$$e_{T+1} = Y_{T+1} - \hat{Y}_{T+1} = Y_{T+1} - E(Y_{T+1} \mid \Omega_T) =$$
$$= \beta_0 + \beta_1 Y_T + \varepsilon_{T+1} - \beta_0 - \beta_1 Y_T = \varepsilon_T$$
$$Var(e_{T+1}) = \sigma_\varepsilon^2$$
$$Y_{T+2} = \beta_0 + \beta_1 \hat{Y}_{T+1} = \beta_0 + \beta_1(\beta_0 + \beta_1 Y_T)$$
$$e_{T+2} = \varepsilon_{T+2} + \beta_1 \varepsilon_{T+1}$$
$$Var(e_{T+2}) = \sigma_\varepsilon^2(1 + \beta_1^2)$$
$$e_{T+3} = \varepsilon_{T+3} + \beta_1 \varepsilon_{T+2} + \beta_1^2 \varepsilon_{T+1}$$
$$Var(e_{T+3}) = \sigma_\varepsilon^2(1 + \beta_1^2 + \beta_1^4)$$
$$Var(e_{T+s}) \to \frac{\sigma_\varepsilon^2}{1 - \beta_1^2}$$

<u>MA(1)</u>

$$Y_t = \beta_0 + \varepsilon_t + \alpha_1 \varepsilon_{t-1}$$
$$\hat{Y}_{T+1} = E(Y_{T+1} \mid \Omega_T) = \beta_0 + \alpha_1 \varepsilon_T$$
$$\hat{Y}_{T+s} = \beta_0, s \geq 2$$
$$Var(e_{T+1}) = \sigma_\varepsilon^2$$
$$Var(e_{T+s}) = \sigma_\varepsilon^2(1 + \alpha_1^2), s \geq 2$$

<u>ARMA(1, 1)</u>

$$e_{T+1} = \varepsilon_{T+1}$$
$$Var(e_{T+1}) = \sigma_\varepsilon^2$$
$$e_{T+2} = \varepsilon_{T+2} + \alpha_1 \varepsilon_{T+1} + \beta_1 \varepsilon_{T+1}$$
$$Var(e_{T+2}) = \sigma_\varepsilon^2(1 + (\alpha_1 + \beta_1)^2)$$

13.10.2 Коинтеграция временных рядов

Стационарный временной ряд: $X_t \sim I(0)$
Если только разность порядка d является стационарной, то этот ряд называется интегрированным порядка d: $X_t \sim I(d)$
Свойства интегрированных временных рядов

1. $X_t \sim I(0), Y_t \sim I(1) \to Z_t = X_t + Y_t \sim I(1)$
2. $X_t \sim I(d) \to Z_t = a + bX_t \sim I(d)$
3. $X_t \sim I(d_1), Y_t \sim I(d_2) \to Z_t = aX_t + bY_t \sim I(d_{max(d_1,d_2)})$
4. $X_t \sim I(d), Y_t \sim I(d) \to Z_t = aX_t + bY_t \sim I(d^* \leq d)$

Определение: Коинтегрированные временные ряды

Если $X_t \sim I(d), Y_t \sim I(d)$, а $Z_t = aX_t + bY_t \sim I(0)$, то ряды X_t, Y_t называются коинтегрированными

Коинтеграция
Если:

$$Y_t = \beta_0 + \beta_1 X_t + \varepsilon_t$$
$$\varepsilon_t = Y_t - \beta_0 - \beta_1 X_t$$

Y_t, X_t - временные ряды одного порядка, коэффициент β_1 является значимым, а ошибки нестанционарны, то имеет место *мнимая регрессия*
Если ε_t станционарны, то имеет место *коинтеграция*
Проверка на практике

1. Оцениваем регрессию одного ряда на другой
2. Сохраняем остатки
3. Смотрим на их ACF, PACF
4. Применять тесты DF, ADF нельзя
5. Надо использовать таблицу Маккинона

13.11 Модели с распредленными лагами

1. Модели с распределенными лагами - лаги у X
2. Регрессионные динамические модели - лаги у Y
3. ADL - лаги у X и Y

Проблемы при оценке

1. Модель с лаговыми переменными - вместо текущих значений включаются лаговые значения
2. Если учитывать несколько лагов - возникает проблема мультиколлинеарности

3. Модель геометрических лагов Койка

$$Y_t = \alpha + \beta X_t + \beta\lambda X_{t-1} + \beta\lambda^2 X_{t-2} + \ldots + \varepsilon_t$$
$$0 < \lambda < 1$$

Зависимость уже не является линейной - убирает проблему мультиколлинеарности

Сведение к динамической модели:

$$Y_{t-1} = \alpha + \beta X_{t-1} + \ldots$$
$$Y_t - \lambda Y_{t-1} = \alpha(1-\lambda) + \beta X_t + \varepsilon_t - \lambda\varepsilon_{t-1}$$
$$Y_t = \alpha(1-\lambda) + \beta X_t + \lambda Y_{t-1} + \varepsilon_t - \lambda\varepsilon_{t-1}$$

4. Сильное предположение о экспоненциальном убывании влияния
5. В модели присутствует эндогенность - ее нельзя оценить МНК
6. **Методы оценивания**
   (a) Нелинейный метод оценивания
      i. Вводим переменную
      $$Z_t(\lambda) = X_t + \lambda X_{t-1} + \lambda^2 X_{t-2} + \ldots$$
      ii. Для каждого значения $\lambda = 0, 0.1, \ldots, 1$ оцениваем
      $$Y_t = \alpha + \beta Z_t + \varepsilon_t$$
      iii. Выбираем оценки параметров с минимальным RSS
   (b) ММП
      i. $\varepsilon \sim \mathcal{N}(0, \Omega)$
   (c) Метод инструментальных переменных
      i. Ищем инструменты для Y_{t-1}
      ii. Исходя из предположения, что X экзогенные
7. Модель Ширли-Алмон

$$Y_t = \alpha + \beta_0 X_t + \beta_1 X_{t-1} + \ldots + \varepsilon_t$$
$$\beta_i = c_0 + c_1 i + c_2 i^2 + \ldots + c_p i^p$$
$$Y_t = \alpha + (c_0 X_t + c_0 + c_1 + c_2 + \ldots + c_p) X_{t-1} + \ldots$$
$$Y_t = \alpha + c_0(X_t + X_{t-1} + \ldots) + c_1(X_{t-1} + 2X_{t-2} + \ldots)$$

8. Модель адаптивных ожиданий

$$Y_t = \beta_0 + \beta_1 X^e_{t+1} + \varepsilon_t$$
$$X^e_{t+1} - X^e_t = \lambda(X_t - X^e_t)$$
$$X^e_{t+1} = \lambda X_t + (1-\lambda) X^e_t$$
$$Y_t = \beta_0 + \beta_1(\lambda X_t + (1-\lambda) X^e_t) + \varepsilon_t$$
$$Y_t = \beta_0 + \beta_1 \lambda X_t + \beta_1 \lambda(1-\lambda) X_{t-1} + \beta_1 (1-\lambda)^2 X^e_{t-1} + \varepsilon_t$$
$$Y_t = \ldots$$

(a) Получается модель геометрических лагов Койка

(b) Сводим к динамической модели

$$Y_t = \beta_0 + \beta_1 \lambda X_t + (1-\lambda)(Y_{t-1} - \beta_0 - \varepsilon_{t-1}) + \varepsilon_t$$

(c) Интерпретация оценок параметров

(d) Если X увеличится на 1 единицу, то Y увеличится на β_1 единиц

(e) Станционарный уровень: $Y^* = \beta_0 + \beta_1 X^*$

9. Модель частичной корректировки

$$Y^* = \beta_0 + \beta_1 X_t + \varepsilon_t$$

$$Y_t - Y_{t-1} = \lambda(Y_t^* - Y_{t-1})$$
$$Y_t = \lambda Y_t^* + (1-\lambda)Y_{t-1}$$

(a) Y^* - желаемое значение

(b) Сводим к динамической модели

$$Y_t = \lambda(\beta_0 - \beta_1 X_t + \varepsilon_t) + (1-\lambda)Y_{t-1}$$

(c) Станционарный уровень: $Y^* = \beta_0 + \beta_1 X^*$

10. Модель коррекции ошибками

(a) Если между двумя рядами обнаружена коинтеграция, то эту информацию необходимо использовать, чтобы построить более полную и корректную модель

$$\Delta y_t = \sum_{i=1}^{p} b_i \Delta y_{t-i} + \sum_{i=1}^{q} c_i \Delta x_{t-i} - \gamma(y_{t-1} - \alpha_0 + \beta_1 x_{t-1} + \varepsilon_t)$$

(b) Коэффициенты b, c означают подстройку текущих значений под долгосрочное равновесие

14 Панельные данные

Панельные данные:

$$Y_{it} = \alpha_i + X_{it}'\beta + u_{it}$$

$i = 1, \ldots, N$ - Номер индивида $t = 1, \ldots, T$ - Момент времени

14.1 Способ представления данных

14.1.1 Pooled regression

Используем данные об одном и том же индивиде в разные моменты времени, как разные наблюдения и оцениваем по выборке из $N \cdot T$ наблюдений.
Проблема такого метода:
Индивидуальные регрессии по углам наклона могут сильно отличатся от общей регрессии - можем получить смещенные результаты.
Обычно t имеет более низкую размерность по сравнению с i

14.1.2 Модели с фиксированными переменными

$$Y_{it} = \alpha_i + X_{it}'\beta + u_{it}$$

β одинаковы для всех наблюдений
α разные для индивидов, одинаковы для временных промежутков

$$E(\alpha_i) = const, var(\alpha_i) = 0$$

Оценка моделей:

1. Вводим дамми переменные для всех индивидов
2. Число степеней свободы: NT - N - k
3. Если N велико, а T мало - число степеней свободы мало $\rightarrow$ снижает эффективность + возникает проблема мультиколлинеарности

Сравнение моделей:

$$\begin{cases} H_0 : \alpha_1 = \ldots = \alpha_N \\ H_1 : \exists \alpha_i \neq \alpha_j \end{cases}$$

$$F = \frac{(RSS_{pooled} - RSS_{FE})/(N-1)}{RSS_{FE}/(NT - N - k)}$$

14.1.3 Between Regression

Усредняем значения для всех моментов времени

$$Y_i = \frac{1}{T}\sum_{t=1}^{T} Y_{it}$$

$$Y_i = \alpha_i + X_i'\beta + u_i$$

14.1.4 Within Regression

$$Y_{it} - \bar{Y_i} = (X_{it}' - \bar{X_i}')\beta + \varepsilon_{it} - \varepsilon_i$$

$$\hat{\beta_W} = \hat{\beta_{FE}}, RSS_W$$

При этом, нет проблем со степенями свободы.

14.1.5 Модели со случайными эффектами

$$Y_{it} = X_{it}'\beta + const + \alpha_i + \varepsilon_{it}$$

$$E(\alpha_i) = 0, var(\alpha_i) = \sigma_\alpha^2$$

const позволяет свести мат. ожидание α к 0
Теперь надо оценить только константу и дисперсию alpha

$$\Sigma = \begin{pmatrix} \sigma_\alpha^2 + \sigma_\varepsilon^2 & \sigma_\alpha^2 & \ldots & \sigma_\alpha^2 \\ \sigma_\alpha^2 & \sigma_\alpha^2 + \sigma_\varepsilon^2 & \ldots & \sigma_\alpha^2 \\ \ldots & \ldots & \ldots & \ldots \\ \sigma_\alpha^2 & \ldots & \sigma_\alpha^2 + \sigma_\varepsilon^2 & \sigma_\alpha^2 \\ \sigma_\alpha^2 & \ldots & \ldots & \sigma_\alpha^2 + \sigma_\varepsilon^2 \end{pmatrix}$$

$$\Omega = \begin{pmatrix} \Sigma & & \\ & \ddots & \\ & & \Sigma \end{pmatrix}$$

$$\hat{\beta}_{RE} = (X'\Omega^{-1}X)^{-1}X'\Omega Y$$

Эти модели более эффективны по сравнению с FE. При этом, альфы могут коррелировать с X.

Сравнение моделей:

Сквозная модель vs Модель со случайными эффектами

$$\begin{cases} H_0 : \sigma_\alpha^2 = 0 \\ H_1 : \sigma_\alpha^2 \neq 0 \end{cases}$$

Оценивается с помощью теста Бройша-Пагана или LM

$$F = \frac{\hat{\sigma}_B^2}{\hat{\sigma}_W^2} \sim F(N-k, NT-N-k)$$

$$LM = \frac{NT}{2(T-1)} \left[\frac{T^2 \sum_{i=1}^{N} \varepsilon_i}{\sum_{i=1}^{N} \sum_{t=1}^{T} \varepsilon_{it}^2} - 1 \right]^2$$

FE vs Случайные эффекты

$$\begin{cases} H_0 : RE \leftrightarrow corr(\alpha_i, X_{it}) = 0 \rightarrow \text{Оценки состоятельны, Разница мала} \\ H_1 : FE \leftrightarrow corr(\alpha_i, X_{it}) \neq 0 \rightarrow \text{Только FE состоятельна} \end{cases}$$

Тест Хаусмана:

$$m = \hat{q}' var^{-1}(\hat{q})\hat{q} \sim \chi_k^2$$

$$\hat{q} = \hat{\beta}_{FE} - \hat{\beta}_{RE}$$

14.1.6 Модель с фиксированными индивидуальными эффектами и временными эффектами

$$Y_{it} = \alpha_i + X_{it}'\beta + c_t + u_i$$

Включаются дамми-переменные для каждого момента времени и каждого индивида