Data Science.
Lectures. Weeks 3-4.
Обобщённый метод наименьших квадратов

Мороз Екатерина

Contents

1 Классическая ЛММР (Линейная модель множественной регрессии)

Матричная форма записи: $Y = X\beta + \varepsilon$
Y – n-мерный вектор наблюдений за переменной y, снятых с объектов, попаших в выборку;
X – матрица (n x k), где k - это количество регрессоров (объясняющих переменных), включая константу,
n – объем выборки.

X – детерминированная (неслучайная) матрица:
Объекты, попавшие в выборку, имеют неслучайные значения;
При изменении выборки матрица X не меняется, меняется только ε - за счет этого будет меняться Y.

$E(\varepsilon) = 0$ – случайная ошибка в среднем отсутствует (мат. ожидание равно 0)
Все дисперсии случайных ошибок одинаковы: σ^2 (неизвестна)
$V(\varepsilon) = E(\varepsilon\varepsilon^T) = \sigma^2 I_n$ – случайные ошибки не коррелируют.

Коварицонная матрица вектора случайных ошибок

$$E\varepsilon\varepsilon^T = E\begin{pmatrix}\varepsilon_1\\ ...\\ \varepsilon_n\end{pmatrix}\begin{pmatrix}\varepsilon_1 & ... & \varepsilon_n\end{pmatrix} = \begin{pmatrix}E\varepsilon_i\varepsilon_j\end{pmatrix}_{i,j=1,...,n} = \begin{pmatrix}\sigma^2 & ... & 0\\ ... & ... & ...\\ 0 & ... & \varepsilon_n\end{pmatrix} = \sigma^2 I_n$$

$$cov(\varepsilon_i\varepsilon_j) = E\varepsilon_i\varepsilon_j = \begin{cases}\sigma^2, & i = j\\ 0, & i \neq j\end{cases}$$

1.1 Обычный метод наименьших квадратов

(МНК; OLS - ordinary least squares)

$\hat{\beta}_{\text{мнк}} = (X^TX)^{-1}X^TY$

Теорема:
Полученные оценки являются несмещенными
Вектор средних значений для $\hat{\beta}_{\text{мнк}}$: $E(\hat{\beta}_{\text{мнк}}) = \beta$

Доказательство:
$\hat{\beta}_{\text{мнк}} = (X^TX)^{-1}X^T(X\beta + \epsilon) = (X^TX)^{-1}X^T\beta + (X^TX)^{-1}X^T\varepsilon = \beta + (X^TX)^{-1}X^T\varepsilon$
$E(A\xi) = AE(\xi)$

Ковариационная матрица вектора $\hat{\beta}_{\text{мнк}}$
$V(\hat{\beta}_{\text{мнк}}) = \sigma^2(X^TX)^{-1}$

Доказательство:
$V(\hat{\beta}_{\text{мнк}}) = E(\hat{\beta}_{\text{мнк}} - \beta)(\hat{\beta}_{\text{мнк}} - \beta)^T = E((X^TX)^{-1}X^T\varepsilon[(X^TX)^{-1}X^T\varepsilon]^T) =$
$= (X^TX)^{-1}X^TE(\varepsilon\varepsilon^T)X(X^TX)^{-1} = \sigma^2(X^TX)^{-1}$
$(AB)^T = B^TA^T; (A^{-1})^T = (A^T)^{-1}$

1.2 Теорема Гаусса-Маркова для КЛММР

$\hat{\beta}_{\text{мнк}}$ обладает свойством оптимальности:
Если рассмотреть K - класс всех оценок β_j, несмещенных и линейных по $y_1...y_n$, то $D\hat{\beta}_j = \min D\tilde{\beta}$

Доказательство:
1) $\hat{\beta}_{\text{мнк},j} \in K_j$ $(E\hat{\beta}_{\text{мнк}} = \beta; \hat{\beta} = (X^TX)^{-1}XY$ – линейная функция от $Y)$
2) $\varphi_c(\beta) = c^T\beta = c_1\beta_1 + ... + c_k\beta_k$ – параметр для оценивания
$c = (c_1...c_k)^T, c_j = const$ (по Y)
$K_c = \{b^TY\}$ – класс оценок для $c^T\beta$
$b \in R^n$ – может зависеть от X
Докажем, что $D(c^T\hat{\beta}_{\text{мнк}}) \leq D(b^TY) \forall b^TY \in K_c$
2.1) $c^T\hat{\beta}_{\text{мнк}} \in K_c$:
– $Ec^T\hat{\beta} = c^T\beta$ - несмещенность
– $c^T\hat{\beta} = c^T(X^TX)^{-1}X^TY = b_0^TY$ - линейность
2.2) $D(c^T\hat{\beta}) = E(c^T\hat{\beta} - c^T\beta)^2 = E[c^T(\hat{\beta} - \beta)][...]^T = E[c^T(\hat{\beta} - \beta)(\hat{\beta} - \beta)^Tc] =$
$= c^T(E(\hat{\beta} - \beta)(\hat{\beta} - \beta)^T)c = \sigma^2c^T(X^TX)^{-1}c$
$E(A\xi B) = A(E\xi B)$
3) Условие несмещенности:
$E(b^TY) = E(b^T(X\beta + \varepsilon)) = b^TX\beta + b^TE\varepsilon = b^TX\beta \equiv c^T\beta \Longrightarrow b^TX = c^T$
$Ec^T\hat{\beta} = c^T\beta = B^TX\beta$
4) $D(b^TY) = E(b^TY - c^T\beta)^2 = E(b^T(X\beta + \varepsilon) - c^T\beta)^2 = E(b^TX\beta - c^T\beta + b^T\varepsilon)^2 =$
$= E(b^T\varepsilon)(b^T\varepsilon)^T = E(b^T\varepsilon\varepsilon^Tb) = b^T(E\varepsilon\varepsilon^T)b = \sigma^2b^Tb \geq 0$

2 Мультиколлинеарность

ε_i
$y_i = \beta_1 + \beta_2x_i^{(2)} + ... + \beta_kx_i^{(k)} + \varepsilon_i$, $i = 1, ..., n$
$Y = X\beta + \varepsilon$, $X - (n \text{ x } k)$
$E\varepsilon = O_n$, $V(\varepsilon) = \sigma^2I_n$

Ковариационная матрица случайных ошибок является классической
МНК-оценки коэффициентов: $\hat{\beta} = (X^TX)^{-1}X^TY$
Ковариационная матрица МНК-оценок: $V(\hat{\beta}) = \sigma^2(X^TX)^{-1}$

Мультиколлинеарность может быть полная и частичная. На практике встречается только частичная.

2.1 Полная мультиколлинеарность

$rankX < k \Longleftrightarrow det(X^TX) = 0$, то есть столбцы матрицы связаны линейной зависимостью:
$\exists c = (c_1, ..., c_k)^T$, $c \neq O_k : c_1X^{(1)} + ... + c_kX^{(k)}$ или $c^TX = O^T$

Встречается редко, когда на этапе спецификации мы не замечаем связи между регрессорами. В таком случаем МНК не может быть реализоаван. С ней легко бороться: необходимо удалить лишние регрессоры.

2.2 Частичная (практическая) мультиколлинеарность

$det(X^TX) \approx 0$
Какая-то из переменных очень близка к линейной комбинации остальных.

2.3 Признаки мультиколлинеарности

1. Малое значение $det(X^T X)$

2. Матрица парных коэффициентов корреляций $\hat{R} = (\hat{r}_{ij})_i, j = 1, ..., k$ содержит числа близкие к 1. Один регрессор сильно связан с другим

3. Среди коэффициентов детерминации $R^2(x^{(j)}/X(j)), j = 1, ..., k$ одного регрессора по остальным есть близкие к 1 ($X(j)$ - набор объясняющих переменных, всех, кроме $x^{(j)}, (dimX(j) = k - 1)$)

Пояснение к п. 3:
В матрице $V(\hat{\beta}) = \sigma^2(X^T X)^{-1}$ на главной диагонали стоят дисперсии
$D\hat{\beta}_j = \dfrac{\sigma^2}{n(1 - R^2(x(j)/X(j))}, [s.e.(\hat{\beta}_j)]^2 = S^2_{\hat{\beta}_j} = \dfrac{\hat{\sigma^2}}{n(1 - R^2...)}$, где $\hat{\sigma^2} = S^2 = \dfrac{ESS}{n - k}$
s.e. - большие $\Longrightarrow t_{\text{стат}}$ - маленькие $\Longrightarrow$ оценки коэффициентов незначимы
(много незначащих факторов в модели)

2.4 Последствия мультиколлинеарности

– Неустойчивость оценок коэффициентов к составу исходных данных (если в выборке изменить одну-две пары x и y, то значимые оценки могут стать незначимыми и наоборот)

Пример:$y_i = x_i + 2z_i + \varepsilon_i$
Пусть $x \approx 2z$ (если $r(x, z) \approx 1$)
Тогда могут быть эквивалентные представления модели:
$y_i = 4z_i + \varepsilon_i'$, $y_i = -x_i + 6z_i + \varepsilon_i''$, $y_i = 5x_i + 6z_i + \varepsilon_i'''$
Разные исследователи будут получать существенно разные выводы

– Знаки оценок меняются

– Большие значения стандартных ошибок (дисперсий)

2.5 Как бороться с мультиколлинеарностью?

Три подхода:

– Гребневая регрессия (смещенное оценивание коэффициентов)

– Метод главных компонент (переход к некоррелированным регрессорам)

– Снижение размерности пространства регрессоров (отбор наиболее информативных обясняющих переменных) - **самый распространенный метод**

2.5.1 Гребневая регрессия

Отказ от МНК-оценивания, переход к смещенному оцениванию коэффициентов:
Вместо $\hat{\beta}_{\text{мнк}} = (X^T X)^{-1} X^T Y$ вводим класс оценок $\hat{\beta}(\alpha) = (X^T X + \alpha I_k)^{-1} X^T Y$
В частности, $\hat{\beta}(0) = \hat{\beta}_{\text{мнк}}$ - несмещенная оценка

$E\hat{\beta}(\alpha) = \beta + \delta(\alpha), \delta(\alpha) \neq 0$ при $\alpha \neq 0$
Утв. $\exists \alpha_0 \neq 0 : E(\hat{\beta}(\alpha_0) - \beta)^2 < E(\hat{\beta}_{\text{мнк}} - \beta)^2$ - найдется α, у которой средний квадрат отклонения меньше, чем у МНК-оценки.

На практике α находят методом подбора. Чаще всего при $0,2 < \alpha < 0,4$ достигается минимум средне-квадратичной оценки.

2.5.2 Метод главных компонент

Временный переход к некоррелированным данным:
1) Центрирование исходных данных
$\{(y_i, x_i^1, ..., x_i^k, k = 1, ..., n\} \Longrightarrow \{(\tilde{y_i}, \tilde{x_i^1}, ..., \tilde{x_i^k}, k = 1, ..., n\}$
$\tilde{y_i} = y_i - \bar{y}, \tilde{x_i}^{(j)} = x_i^{(j)} - \bar{x}^{(j)}$
2) Построение главных компонент для центрированных регрессоров
$\tilde{x_i^1}, ..., \tilde{x_i^k} \Longrightarrow f^{(1)}, ..., f^{(k)}$
Каждая главная компонента является линейной комбинацией переменных $\tilde{x_i}^1, ..., \tilde{x_i}^k$
- Все главные компоненты попарно некоррелированны
- По главным комнонентам можно однозначно восстановить исходные регрессоры (и наоборот)
3) Построение регрессии $\tilde{y}$ по $f^{(1)}, ..., f^{(k)}$
$\tilde{y_i} = c_1 f_i^{(1)} + ... + c_k f_i^{(k)} + \tilde{\sum}_i$
4) Возврат к $\tilde{x_i}^1, ..., \tilde{x_i}^k$
Результат: регрессия y по $x_i^1, ..., x_i^k$

2.5.3 Отбор наиболее информативных регрессоров (снижение размерности)

- Метод пошаговой регрессии
- Метод полного перебора регрессий

$p = k - 1$ – количество структурных регрессоров (априорный набор регрессоров $x^{(2)}, ..., x^{(k)}$)
Перебор моделей по числу выбранных регрессоров: $p = 1, 2, ..., p$

Два варианта перебора: с увеличением и уменьшением числа объясняющих переменных

Полный перебор с добавлением регрессоров
$p = 1 : R^2(y/x^{(j)}) \longrightarrow max_{j=1,...,p} = R^2(1)$ - построение всех возможных регрессий, когда y зависит только от одного x
$p = 2 : R^2(y/x^{(j_1)}, x^{(j_2)}) \longrightarrow max_{j_1,j_2=1,...,p,j_1=J-2} = R^2(2)$ - повторение для двух x и так далее

Последовательность $R^2(p)$ не убывает

Правило остановки:
(i) Прирост $R^2(p+1) - R^2(p)$ стал статистически незначимым
(ii) R^2_{adj} досиг своего максимума
(iii) Нижняя доверительная граница для R^2_{adj} достигла максимума. **Самый правильный вариант**

Частичный перебор
Пошаговая регрессия.
Отличие от полного перебора: при переходе к следующему шагу $(p+1)$ один из иксов выбирается таким, какой был на предыдущем шаге, это снижает количество переборов.

На практике отбор регрессоров происходит методом исключения незначимых переменных

3 Обобщенная ЛММР, обобщенный МНК

$Y = X\beta + \varepsilon$
X – детерминированная (n x k), $rankX = k, k < n$
$E\varepsilon = O_n, V(\varepsilon) = \Omega$
Техническое условие: $\Omega > 0$ – положительная определенность

Свойства:
1) $\Omega^T = \Omega$
2) $\exists \Omega^{-1}, \Omega^{-1} > 0$
3) $\exists P$ невырожденная (n x n) : $\Omega^{-1} = P^T P$

Два метода оценивания: МНК, ОМНК

3.1 МНК для КЛММР и ОЛММР

Критерий: $(Y - Xb)^T(Y - Xb) \longrightarrow min_b$
Оценка: $\hat{\beta}_{\text{мнк}} = (X^TX)^{-1}X^TY$

Свойства:
– Несмещенность $E\hat{\beta}_{\text{мнк}} = \beta$ (используется только $E\varepsilon = O_n$)
– Состоятельнойсть $\frac{1}{n}X^TX \longrightarrow Q_{x^Tx} > 0$
– Линейность по Y
– Оптимальность в случае КЛММР (теорема Гаусса-Маркова)
– Неоптимальность в случае ОЛММР

Что делать?
Если $V(\varepsilon) = \sigma^2 I_n$, то $V(\hat{\beta}_{\text{МНК}} = \sigma^2(X^TX)^{-1}$
В этом случае $\hat{\sigma}^2_{\text{МНК}} = \dfrac{1}{n-k}(Y - X\hat{\beta}_{\text{мнк}})^T(Y - X\hat{\beta}_{\text{мнк}})$ – несмещенная оценка для σ^2

Если $V(\varepsilon) = \Omega(\Omega > 0)$, то $V(\hat{\beta}_{\text{МНК}} = (X^TX)^{-1}X^T\Omega X(X^TX)^{-1}$

Что в случае ОЛММР оценивает формула $\hat{\sigma}^2_{\text{МНК}}$? Сигмы нет, формула ничего не оценивает

Надо уметь состоятельно оценивать матрицу Ω

3.2 ОМНК для ОЛММР

(1) $Y = X\beta + \varepsilon$ – ОЛММР (2) $PY = PX\beta + P\varepsilon$ – КЛММР
ОМНК для модели (1) это МНК, примененный к преобразованным данным, то есть к модели (2)

Критерий:
$(PY - PXb)^T(PY - PXb) \longrightarrow min_b$
$(Y - Xb)^TP^TP(Y - Xb) = (Y - Xb)^T\Omega^{-1}(Y - Xb) \longrightarrow min_b$

Оценка:
$\hat{\beta}_{\text{омнк}} = [(PX)^T(PX)]^{-1}(PX)^T(PX) = (X^T\Omega^{-1}X)^{-1}X^T\Omega^{-1}Y$

Свойства ОМНК оценок:
- Несмещенность $E\hat{\beta}_{\text{омнк}} = \beta$
- Состоятельность
- Линейность по Y
- Оптимальность (теорема Айткена)

$V(\hat{\beta}_{\text{омнк}}) = (X^T\Omega^{-1}X)^{-1}$

Есть ковариационная матрица $V(\varepsilon) = \Omega$ неизвестна,
то ОМНК-оценка $\hat{\beta}_{\text{омнк}} = (X^T\Omega^{-1}X)^{-1}X^T\Omega^{-1}Y$ не реализуема

Частный случай: $\Omega = \sigma^2\Omega_0$
σ^2 – неизвестный параметр
$\Omega_0 > 0$ – известная матрица $\exists P_0 : \Omega_0^{-1} = P_0^TP_0$
$\hat{\beta}_{\text{омнк}} = (X^T\Omega_0^{-1}X)^{-1}X\Omega_0^{-1}Y$ – реализуемая оценка

Критерий: $(Y - Xb)^T\Omega_0^{-1}(Y - Xb) \longrightarrow min_b$
Ковариационная матрица: $V(\hat{\beta}_{\text{омнк}}) = \sigma^2(X^T\Omega_0^{-1}X)^{-1}$
Оценивание σ^2 в рамках ОЛММР с $V(\varepsilon) = \sigma^2\Omega_0$
Модель (2): $P_0Y = P_0X\beta + P_0\varepsilon, E(P_0\varepsilon) = O$
$V(P_0\varepsilon) = E(P_0\varepsilon)(P_0\varepsilon)^T = P_0E(\varepsilon\varepsilon^T)P_0^T = P_0\sigma^2\Omega_0P_0^T = \sigma^2P_0(P_0^TP_0)^{-1}P_0^T = \sigma^2I_n$

Критерий:$(P_0Y - P_0Xb)^T(P_0Y - P_0Xb) \longrightarrow min_b$

Оценка σ^2
$\hat{\sigma}^2_{\text{омнк}} = \dfrac{1}{n-k}(P_0Y - P_0X\hat{\beta}_{\text{омнк}})^T(P_0Y - P_0X\hat{\beta}_{\text{омнк}}) = \dfrac{1}{n-k}(Y - X\hat{\beta}_{\text{омнк}})^T\Omega_0^{-1}(Y - X\hat{\beta}_{\text{омнк}})$

Коэффициент детерминации R^2 в обобщенной ЛММР

$R^2 = 1 - \dfrac{ESS}{TSS}$

$R^2 = 1 - \dfrac{(Y - X\hat{\beta}_{\text{омнк}})^T \Omega_0^{-1} (Y - X\hat{\beta}_{\text{омнк}})}{(Y - \bar{Y}_i)^T \Omega_0^{-1}} (Y - \bar{Y}_i)$

R^2 скорее всего не имеет смысла, так как в преобразованной матрице вида PX столбец единиц отсутствует.

3.3 Выводы:

Выбор метода оценивания зависит от предпосылок модели.
Надо максимально учитывать специфику модели, в частности, учесть характер зависимости регрессионных ошибок.
Если мы знаем Ω, то можно применить ОМНК, он даст нам лучшие несмещенные, линейные оценки.
МНК тоже даст несмещенные, состоятельные, линейные, но не наилучшие оценки.

3.4 Частный случай: ЛММР с гетероскедастическими (и некоррелированными) ошибками

Гетероскедастичность: дисперсии случайных ошибок значительно различаются

(1) $Y = X\beta + \varepsilon, V(\varepsilon) = \Omega$ $\Omega = \begin{pmatrix} \sigma_1^2 & & 0 \\ & ... & \\ 0 & & \sigma_n^2 \end{pmatrix}, D\varepsilon_i = \sigma_i^2 \neq const$

$\Omega^{-1} = \begin{pmatrix} \frac{1}{\sigma_1^2} & & 0 \\ & ... & \\ 0 & & \frac{1}{\sigma_n^2} \end{pmatrix}, P = \begin{pmatrix} \frac{1}{\sigma_1} & & 0 \\ & ... & \\ 0 & & \frac{1}{\sigma_n} \end{pmatrix}$

(1) $Y = X\beta + \varepsilon \Longrightarrow$ (2) $PY = PX\beta + P\varepsilon$

(1) $y_i = \beta_1 x_i^{(1)} + ... + \beta_k x_i^{(k)} + \varepsilon_i$ (2) $\dfrac{y_i}{\sigma_i} = \dfrac{\beta_1 x_i^{(1)}}{\sigma_i} + ... + \dfrac{\beta_k x_i^{(k)}}{\sigma_i} + \dfrac{\varepsilon_i}{\sigma_i}$

ОМНК $\equiv$ ВМНК (взвешенный МНК, weight least squares)

Критерий: $\sum_{i=1}^n \dfrac{1}{\sigma^2}(y_i - \beta_1 x_i^{(1)} - ... - \beta_k x_i^{(k)} \longrightarrow min_b$

Суть: если σ_i большое, то i-ое уравнение включается в сумму с маленьким весом

Имеем $n + k$ параметров модели:
$\beta_1, ..., \beta_k$ и $\sigma_1^2, ..., \sigma_n^2$, но n наблюдений $\Longrightarrow$ состоятельно оценить все (независимые) параметры нельзя
Но если $\sigma_1^2, ..., \sigma_n^2$ выражаются через некоторое (небольшое) количество параметров, то состоятельно оценить можно

Что делать? $\Omega = diag(\sigma_1^2, ..., \sigma_n^2)$
Ω известна $\longrightarrow$ ВМНК

Ω неизвестна $\longrightarrow$ Форма гетероскедастичности $\longrightarrow$ Известна $\longrightarrow$ Двухшаговая процедура

Неизвестна $\longrightarrow$ МНК-оценки + поправки Уайта

Функциональная форма гетероскедастичности:
$\sigma_i^2 = f(z_i^{(1)}, ..., z_i^{(p)})$, где z - некоторые из иксов или функции от них

3.4.1 Поправки Уайта

(ОЛММР + МНК)
$\hat{\beta}_{\text{мнк}}$ - несмещенные, состоятельные (но не оптимальные оценки)
$s.e.(\hat{\beta}_{j,\text{мнк}}) = \sqrt{\hat{\sigma}^2_{\text{мнк}}[(X^TX)^{-1}}$ - несостоятельные оценки
$h.c.s.e.(\hat{\beta}_{j,\text{мнк}})$ стандартные ошибки, поправленные с учетом гетероскедастичности

Они являются состоятельными оценками стандартных отклонений

Оценивание ковариационной матрицы мектора МНК-оценок β по Уайту:
$\hat{V}(\hat{\beta}_{\text{мнк}}) = (X^TX)^{-1}XT\hat{\Omega}X(X^TX)^{-1}$
где $\hat{\Omega} = diag(e_1^2, ..., e_n^2), e_i$ – МНК-остаток
$X^Tdiag(e_1^2, ..., e_n^2)X = \sum_{s=1}^n e_s^2 X_s X_s^T - -(k$ x $k)$ матрица
где $X_s = (X_s^{(1)}, ..., X_s^{(k)})^T$ - s-я строка матрицы X, поставленная в столбец

3.4.2 Двухшаговая процедура оценивания параметров гетероскедастической ОЛММР

(1) $Y = X\beta + \epsilon, V(\varepsilon) = \Omega = diag(\sigma_1^2, ..., \sigma_n^2)$
Постулируем линейную форму гетероскедастичности:
$\sigma_i^2 = \gamma_0 + \gamma_1 z_i^{(1)} + ... + \gamma_p z_i^{(p)} = \gamma_0 + Z_i\gamma$ (пусть это истинная, правильно угаданная форма)

Шаг 1. Применение классического МНК
1) МНК (1), получаем остатки и оценку: $\hat{\beta}, e_1, ..., e_n \Longrightarrow e_i^2$
2) Оцениваем вспомогательную регрессию
$e_i^2 = \gamma_0 + Z_i^T\gamma +$ ошибка
МНК $\Longrightarrow \hat{\gamma}_0, \hat{\gamma}$
Доказанный факт: $\hat{\gamma}_0, \hat{\gamma}$ оценки состоятельны
3) получаем окончательные оценки для σ^2: $\hat{\sigma}_i^2 = \hat{\gamma}_0 + Z_i^T\hat{\gamma}$

Шаг 2. Оценка β в модели (1) доступным (реализуемым) ОМНК, то есть применяем взвешенный МНК, в котором вместо весов $1/\sigma_i^2$ берем $1/\hat{\sigma}_i^2$
Оценки $\hat{\beta}$ не являются оптимальными, но они асимптотически оптимальны

> **Замечание:**
> Взвешенный МНК это обычный МНК, примененный к преобразованным данным
> $(y_i, x_i^{(1)}, ..., x_i^{(k)}) \longrightarrow (\frac{y_i}{\hat{\sigma}_i}, \frac{x_i^{(1)}}{\hat{\sigma}_i}, ..., \frac{x_i^{(k)}}{\hat{\sigma}_i})$

Проблема: $\hat{\sigma}_i = \sqrt{\hat{\sigma_i}^2}$
Не $\hat{\sigma_i}^2$ все могут оказаться положительными
Что делать?
- выкидывать те i, для которых $\hat{\sigma_i}^2 \leq 0$
Некорректный способ, так как выбор выкидываемых наблюдений не случаен

На практике пытаются использовать не аддитивную, а мультимликативную форму гетероскедастичности вида: $\hat{\sigma_i}^2 = exp\{\gamma_0 + \gamma_1 z_i^{(1)} + ... + \gamma_p z_i^{(p)}\} = e^{\gamma_0 + Z_i^T\gamma}$

3.4.3 Двухшаговая процедура доступного ОМНК-оценивания параметров обобщенной модели (1) с экспоненциальной формой гетероскедастичности

Шаг 1.

1) МНК (1) $\hat{\beta}_{\text{мнк}} \Longrightarrow e$ - МНК-остатки

2) Постулируем форму гетероскедастичности
$\sigma_i^2 = exp\{\gamma_0 + Z_i^T\gamma\}$, т.е. $ln(\sigma_i^2) = \gamma_0 + Z_i^T\gamma$
Оцениваем вспомогательную регрессию

$ln(\sigma_i^2) = \gamma_0 + Z_i^T \gamma + \text{ошибка}_i$
Получаем состоятельные МНК-оценки $\hat{\gamma}_0, \hat{\gamma}$, параметров γ_0, γ
Получаем состоятельные оценки дисперсий регрессионных ошибок $\hat{\sigma_i}^2 =$, параметров γ_0, γ

3) Получаем окончательные оценки для $\sigma^2 = exp\{\hat{\gamma}_0 + Z_i^T \hat{\gamma}\} > 0$

Шаг 2.

Оцениваем β в модели (1) доступным ОМНК (проблемы отрицательных оценок дисперсий σ^2 здесь нет)

1. До двухшаговой процедуры: есть ли гетероскедастичность в ошибках?
2. После двухшаговой процедуры: избавились ли мы от гетероскедастичности?

Если форма гетероскедастичности ошибок $\varepsilon_1, ..., \varepsilon_n$ исходной модели $y_i = \beta_1 x_i^{(1)} + ... + \beta_k x_i^{(k)}$ оценена правильно, то в модели для преобразованных данных $\frac{y_i}{\sigma_i} = \frac{\beta_1 x_i^{(1)}}{\sigma_i} + ... + \frac{\beta_k x_i^{(k)}}{\sigma_i} + \frac{\varepsilon_i}{\sigma_i}$ ошибки станут гомоскедастичными.