

Московский государственный университет имени М. В. Ломоносова
Факультет космических исследований

Курсовая работа
студента 402 группы
Шигина Глеба Сергеевича

Трансформация данных в GLM моделях

Научный руководитель:
с.н.с., к.ф.-м.н.
Шкляев Александр Викторович

Москва, 2022

Содержание

| | | |
|----------|---|-----------|
| 1 | Введение | 2 |
| 2 | Трансформация данных | 2 |
| 2.1 | Трансформация только зависимой переменной при помощи inverse response plot | 2 |
| 2.1.1 | Inverse response plot | 3 |
| 2.1.2 | Выбор степенного преобразования | 4 |
| 2.2 | Трансформация только зависимой переменной при помощи преобразования Бокса-Кокса | 5 |
| 2.3 | Трансформация только предиктора методом Бокса-Кокса | 6 |
| 2.4 | Трансформация и предиктора, и зависимой переменной | 6 |
| 3 | Практическое сравнение | 8 |
| 3.1 | Задачи с одним предиктором | 9 |
| 3.2 | Задачи с несколькими предикторами | 11 |
| 4 | Заключение | 11 |
| A | Многомерное преобразование Бокса-Кокса | 11 |
| B | Графики | 12 |

1 Введение

ТВА

2 Трансформация данных

Существуют классы задач, для которых мы знаем, что математическое ожидание $\mathbf{E}(Y|X)$ является линейной функцией от X . Это может быть теория, подкрепленная экспериментальными данными [Weisberg et al., 1978], либо статистическая информация о данных. Например, пусть y_i и x_i – выборки их нормальных распределений со средними μ_X, μ_Y соответственно, дисперсиями σ_X, σ_Y соответственно и корреляцией ρ_{XY} . Тогда можно показать (см. [Berger and Casella, 2001, стр. 550]), что

$$y_i | x_i \sim N \left(\mu_Y - \rho_{XY} \frac{\sigma_Y}{\sigma_X} \mu_X + \rho_{XY} \frac{\sigma_Y}{\sigma_X} x_i, \sigma_Y (1 - \rho_{XY}^2) \right). \quad (1)$$

Это можно переписать как $y_i | x_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2)$, где

$$\beta_0 = \mu_Y - \rho_{XY} \frac{\sigma_Y}{\sigma_X} \mu_X, \beta_1 = \rho_{XY} \frac{\sigma_Y}{\sigma_X}, \sigma^2 = \mathbf{D}(Y|X) = \sigma_Y^2 (1 - \rho_{XY}^2), \quad (2)$$

то есть мы получили линейную регрессию Y по X :

$$\mathbf{E}(Y|X = x_i) = \beta_0 + \beta_1 x_i. \quad (3)$$

Однако, у нас не всегда есть возможность узнать истинную зависимость зависимой переменной от предикторов. Нам необходимо понять, какие преобразования бывают и как выбрать среди них наилучшее.

Для удобства на данном этапе ограничимся одним **предиктором** X и **зависимой переменной** Y .

2.1 Трансформация только зависимой переменной при помощи inverse response plot

Предположим, что истинная регрессионная модель Y по X имеет вид:

$$Y = g(\beta_0 + \beta_1 X + \varepsilon), \quad (4)$$

где g – некоторая функция, вообще говоря, нам неизвестная. Модель (4) может быть приведена к линейному виду путем преобразования Y с помощью обратной функции g^{-1} :

$$g^{-1}(Y) = \beta_0 + \beta_1 X + \varepsilon. \quad (5)$$

Например, если $Y = \log(\beta_0 + \beta_1 X + \varepsilon)$, то $g(x) = \log(x)$, значит $g^{-1}(x) = \exp(x)$, и $\exp(Y) = \beta_0 + \beta_1 X + \varepsilon$.

В данной работе в качестве способа получения оценки g^{-1} предлагается рассмотреть метод обратного отклика (*inverse response plot*) с подбором функции из степенного семейства (*power family*), включающим в себя семейство преобразований Бокса-Кокса.

2.1.1 Inverse response plot

В работе [Cook and Weisberg, 1994] было показано, что если X имеет эллиптически симметричное распределение (что является менее жестким ограничением, чем нормальность), то g^{-1} можно оценить из scatter plot'a, где по горизонтальной оси откладываются истинные значения y , а по вертикальной – значения \hat{y} , полученные из регрессии на исходных данных: $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$. Такой график называют **inverse response plot**.

Пример

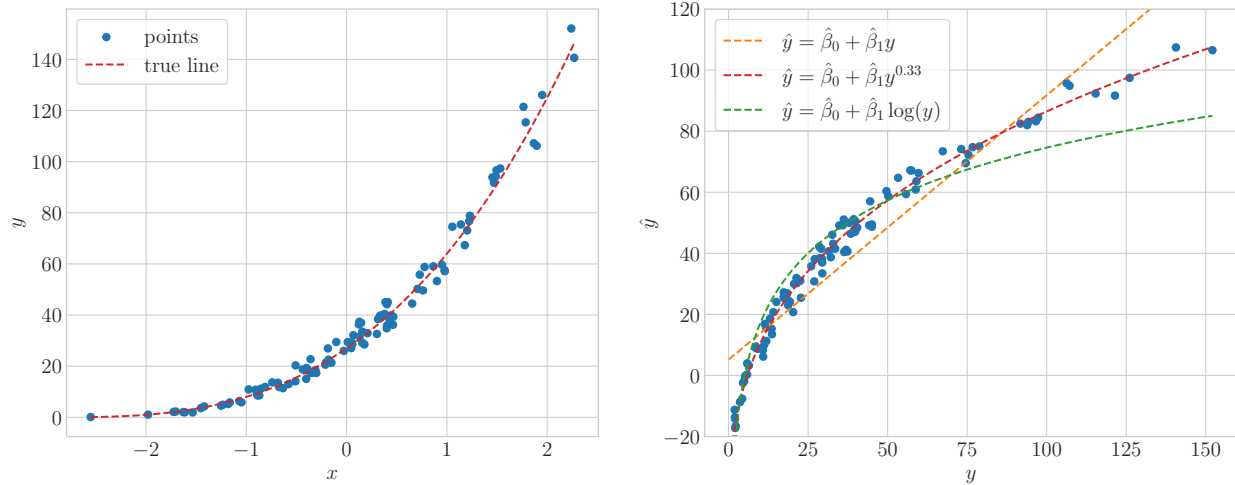


Рис. 1: Scatter plot Y от X (слева) и inverse response plot (справа)

Пусть $X \sim N(0, 1)$, $\varepsilon \sim N(0, 0.1)$, $Y = (3 + X + \varepsilon)^3$. Была построена выборка размера $N = 100$. Получившуюся зависимость можно увидеть на рисунке 1. Построим линейную регрессию y по x без преобразования данных. Получим некоторые оценки \hat{y} . Inverse response plot можно также увидеть на рисунке 1. Помимо точек, на графике присутствуют три пунктирные кривые – они показывают результаты линейных регрессий \hat{y} по y , $y^{0.33}$ и $\log(y)$ соответственно. Видно, что кривая $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 y^{0.33}$ наиболее близка к исходным точкам. Это ожидаемо, так как из построения Y следует, что искомое нами преобразование для Y имеет вид $g^{-1}(Y) = Y^{1/3}$.

2.1.2 Выбор степенного преобразования

Семейство преобразований (*transformation family*) – это параметризованное множество преобразований, где каждому значению параметра (или параметров) отвечает некоторый уникальный представитель семейства.

Одним из таких семейств является **степенное семейство** (*power family*). Оно определено для положительных X и имеет вид:

$$\psi(Y, \lambda) = \begin{cases} Y^\lambda, & \lambda \neq 0 \\ \log(Y), & \lambda = 0 \end{cases}. \quad (6)$$

Семейство параметризовано числом λ , и значение $\lambda = 0$ принимается не за тождественную единицу (ведь $Y^0 \equiv 1$), а за логарифмическое преобразование $\log(Y)$. Здесь возникает проблема, затрудняющая работу с этим семейством – его представители $\psi(Y, \lambda)$ не являются непрерывным по λ . Поэтому удобнее работать с так называемым **нормированным** (или **отмасштабированным**) **степенным семейством** (*scaled power family*):

$$\psi_S(Y, \lambda) = \begin{cases} \frac{Y^\lambda - 1}{\lambda}, & \lambda \neq 0 \\ \log(Y), & \lambda = 0 \end{cases}. \quad (7)$$

Несложно видеть, что $\psi_S(Y, \lambda)$ и $\psi(Y, \lambda)$ отличаются только преобразованием сдвига и масштаба, и линейная регрессия будет давать аналогичные результаты при двух этих преобразованиях, отличаться будут только веса предикторов. При этом функция $\psi_S(Y, \lambda)$ непрерывна по λ , и $\log(Y)$ является естественным представителем семейства, так как $\lim_{\lambda \rightarrow 0} \psi_S(y, \lambda) = \log(y) \forall y > 0$.

Суммируя все вышесказанное, для нахождения оценки g^{-1} мы рассматриваем модели вида

$$\mathbf{E}(\hat{y}|Y = y) = \hat{\beta}_0 + \hat{\beta}_1 \psi_S(y, \lambda). \quad (8)$$

При фиксированном λ модель (8) представляет собой простую линейную регрессию с предиктором $\psi_S(y, \lambda)$ и зависимой переменной \hat{y} . Оптимальным параметром $\hat{\lambda}$ предлагается считать тот, который минимизирует **остаточную сумму квадратов** (*residual sum of squares*):

$$RSS(\lambda) = \sum_{i=1}^n \left(\hat{y}_i - \hat{\beta}_0 - \hat{\beta}_1 \psi_S(y_i, \lambda) \right)^2. \quad (9)$$

2.2 Трансформация только зависимой переменной при помощи преобразования Бокса-Кокса

В своей работе [Box and Cox, 1964] Бокс и Кокс рассматривали модифицированное семейство степенных преобразований:

$$\psi_M(Y, \lambda) = GM(Y)^{1-\lambda} \cdot \psi_S(Y, \lambda) = \begin{cases} GM(Y)^{1-\lambda} \cdot \frac{Y^\lambda - 1}{\lambda}, & \lambda \neq 0, \\ GM(Y) \cdot \log(Y), & \lambda = 0 \end{cases}, \quad (10)$$

где $GM(Y) = \sqrt[n]{y_1 y_2 \dots y_n}$ – среднее геометрическое выборки.

Метод Бокса-Кокса основывается на предположении, что для некоторого неизвестного λ после преобразования зависимая переменная $\psi_M(Y, \lambda)$ такова, что $\psi_M(y_i, \lambda)$ – независимые нормально распределенные сл.в. с постоянной дисперсией σ^2 и математическим ожиданием

$$\mathbf{E}(\psi_M(Y, \lambda) | X = x) = \beta_0 + \beta_1 x. \quad (11)$$

Из этих предположений предлагается брать такое λ , которое максимизировало бы правдоподобие. В нормальной модели логарифм правдоподобия имеет вид:

$$\begin{aligned} \log(L) = \ell = -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i) = \\ = -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log(\sigma^2) - \frac{1}{2\sigma^2} RSS. \end{aligned} \quad (12)$$

Оценка максимума правдоподобия дает нам $\hat{\sigma}^2 = RSS/n$. Отсюда получаем:

$$\ell = -\frac{n}{2} \log(2\pi) - \frac{n}{2} - \frac{n}{2} \log(RSS/n). \quad (13)$$

Можно показать, что якобиан замены преобразования Бокса-Кокса $\psi_M(Y, \lambda)$ равен 1 при любом λ . А значит при фиксированном λ после трансформации зависимой переменной вид функции правдоподобия останется прежним:

$$\ell(\lambda) = -\frac{n}{2} \log(2\pi) - \frac{n}{2} - \frac{n}{2} \log(RSS(\lambda)/n). \quad (14)$$

Так как только последний член выражения (14) зависит от λ , то решение задачи максимизации правдоподобия $L(\lambda)$ (или логарифма правдоподобия $\ell(\lambda)$) по λ эквивалентна задаче минимизации $RSS(\lambda)$ по λ :

$$\ell(\lambda) = -\frac{n}{2} \log(2\pi) - \frac{n}{2} - \frac{n}{2} \log(RSS(\lambda)/n) \rightarrow \max_{\lambda} \Leftrightarrow RSS(\lambda) \rightarrow \min_{\lambda} \quad (15)$$

Здесь и далее для удобства предлагается под преобразованием Бокса-Кокса понимать $\psi_S(X, \lambda)$ для предикторов и $\psi_M(Y, \lambda)$ для зависимой переменной. Обозначим его $\psi_*(U, \lambda)$.

Пример

Снова обратимся к сгенерированному примеру из пункта 2.1.1. Из построения нам известно, что искомое значение λ для преобразования Y равно 0.333. Сравним два метода, которые мы обсудили ранее:

- Inverse response plot выдает оценку $\hat{\lambda} = 0.341$;
- Метод Бокса-Кокса выдает оценку $\hat{\lambda} = 0.273$.

В этом конкретном примере метод Бокса-Кокса оказался ощутимо дальше реального значения параметра λ , чем inverse response plot. Однако, как мы увидим далее, преимущество одного метода над другим не так очевидно.

2.3 Трансформация только предиктора методом Бокса-Кокса

Мы снова рассмотрим преобразование Бокса-Кокса, на этот раз определенных для строго положительного X :

$$\psi_*(X, \lambda) = \begin{cases} \frac{X^\lambda - 1}{\lambda}, & \lambda \neq 0 \\ \log(X), & \lambda = 0 \end{cases}. \quad (16)$$

Аналогично пункту 2.2, мы рассматриваем модели вида

$$\mathbf{E}(Y|X = x) = \alpha_0 + \alpha_1 \psi_*(x, \lambda). \quad (17)$$

При фиксированном λ модель (17) представляет собой простую линейную регрессию с предиктором $\psi_*(x, \lambda)$ и зависимой переменной y . Аналогично задаче преобразования зависимой переменной, для разных значений λ строится регрессия методом наименьших квадратов (МНК-регрессия) и выбирается то $\hat{\lambda}$, которое минимизирует $RSS(\lambda)$.

В качестве альтернативы можно использовать вариацию метода Бокса-Кокса, которая пытается сделать распределение преобразованного предиктора X как можно более нормальным. Заметим, что в этом случае регрессионная модель отсутствует, и метод Бокса-Кокса модифицирован для применения непосредственно к X .

2.4 Трансформация и предиктора, и зависимой переменной

В случае, когда необходимо трансформировать Y и X , существует несколько альтернативных подходов. Ниже эти будут сформулированы эти подходы, где необходимо – предоставлена мотивация подхода, и далее будет проведено практическое сравнение.

Итак, предлагаемые варианты:

1. Попарная минимизация RSS предикторов из \mathbf{X} и преобразование Y с помощью inverse response plot (если необходимо).

Предлагается для каждого предиктора подобрать преобразование Бокса-Кокса, минимизирующее RSS в регрессии Y только по этому предиктору. Также, может быть полезно трансформировать Y не после этих действий, а до, например, когда возникает гетероскедастичность.

2. Приведение всех предикторов к нормальному виду и преобразование Y с помощью inverse response plot (если необходимо).

Этот метод основывается на идее, что для нормально распределенных величин существует линейная зависимость (см. начало раздела 2).

3. Минимизация определителя матрицы выборочных ковариаций $|V(\mathbf{X})|$ и преобразование Y с помощью inverse response plot (если необходимо).

Этот вариант является расширением предыдущего, так как будут учитываться попарные корреляции между предикторами. В работе [Velilla, 1993] метод Бокса-Кокса распространяется на случай многомерных случайных величин. Связь минимизации $|V(\mathbf{X})|$ с решением нашей задачи показана в приложении А.

4. Минимизация RSS одновременно по предикторам и зависимой переменной. В отличие от предыдущего метода, здесь будет учитываться связь зависимой переменной с предикторами.

Для сравнения подходов будет использоваться графический метод, включающий в себя следующие графики:

1. Residuals vs fitted.

Это график остатков регрессии в зависимости от прогноза зависимой переменной. Он позволяет оценить адекватность линейной модели. В адекватной модели ожидается, что линия средних (на графиках – красная кривая) будет горизонтальной прямой, проходящей через 0.

2. Leverage plot.

Рычаг (*leverage*) h_i для i -го наблюдения – это диагональный элемент матрицы $X(X^T X)^{-1} X^T$, соответствующий коэффициенту при y_i в прогнозе \hat{y}_i (который представляет собой линейную функцию от y). У каждого наблюдения по горизонтальной оси откладывается рычаг, а по вертикальной – стандартизированные остатки

$$\frac{r_i}{\hat{\sigma}\sqrt{1-h_i}}.$$

Эта стандартизация основана на том, что $\mathbf{D}r_i = \sigma^2(1-h_i)$. Полученный график позволяет отследить отдельные значения, существенно влияющие на коэффициенты модели.

3. Scale-location plot.

Он позволяет оценить гомоскедастичность. Это график корня из стандартизованных остатков $r_i/(\sqrt{1-h_i}RSS/n)$ от прогноза зависимой переменной.

4. Quantile-Quantile plot. Это график стандартизованных остатков от теоретических квантилей некоторого распределения (в нашем случае – нормального). Он позволяет оценить, насколько стандартизованные остатки соответствуют заданному распределению.

Совокупность этих графиков назовём diagnostic plots. На рисунке 2 представлен пример внешнего вида рассматриваемых графиков.

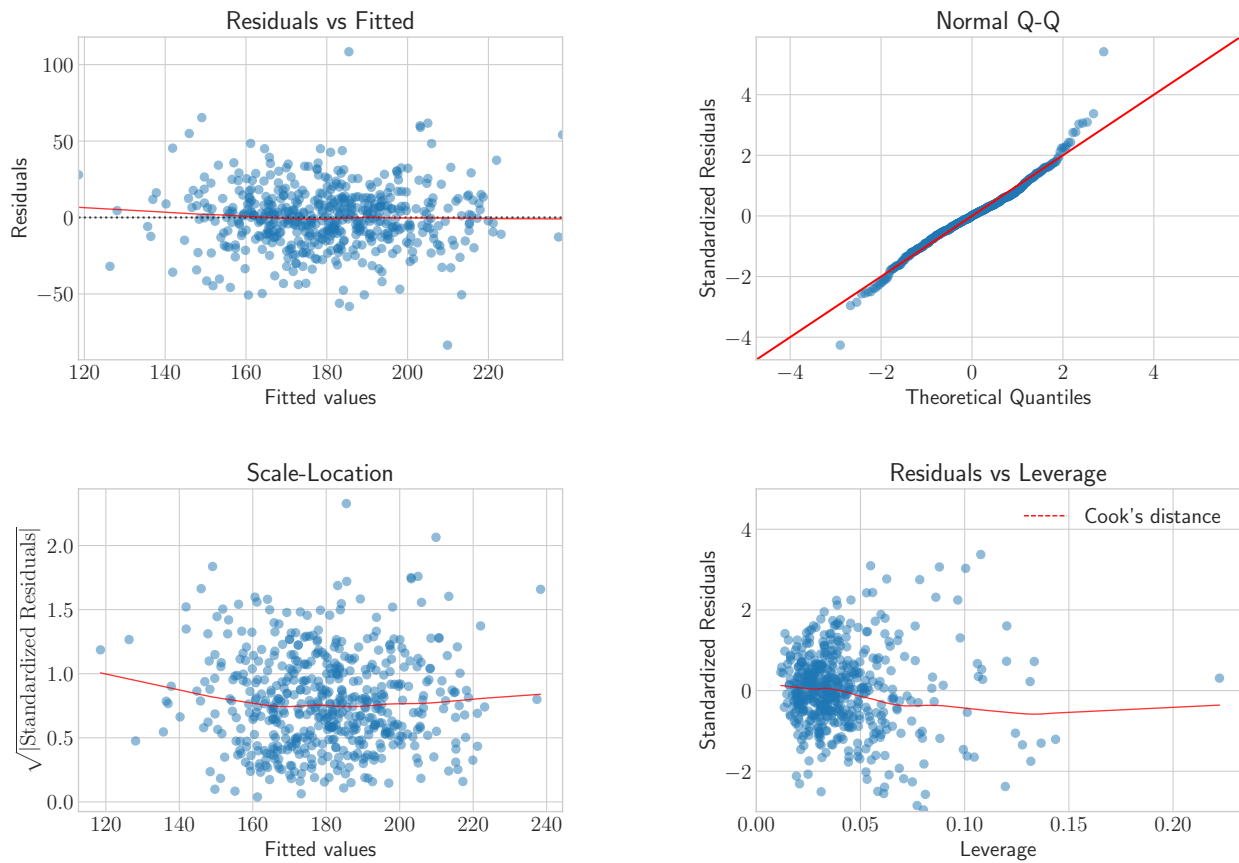


Рис. 2: Пример diagnostic plots, на которых будет происходить сравнение подходов.

3 Практическое сравнение

Здесь мы проведем сравнение методов, сформулированных в разделе 2.4.

3.1 Задачи с одним предиктором

В случае одного предиктора, варианты (2) и (3) трансформации данных представляют собой одно и то же (нормализация многомерной случайной величины сводится к одномерному случаю). Поэтому на данном этапе исключим из рассмотрения вариант (3) и сосредоточимся на трех оставшихся.

Со всеми графиками к этим примерам можно ознакомиться в приложении В.

Пример 3.1. $X \sim \text{Gamma}(2, 1)$, $\varepsilon \sim N(0, 2)$, $Y = 5 + X^2 + \varepsilon$.

В данном примере мы рассмотрим предиктор X из гамма-распределения, нормальными остатками и простой зависимостью Y от X . Ожидается, что для x будет подобрано $\lambda = 2$, а y останется неизменным. Diagnostic plots для трансформаций (1), (2) и (3) показаны на рисунках 4, 5 и 6 соответственно.

Из графиков видно, что нормализация $x + \text{inverse plot}$ для y показали себя плохо, а оставшиеся два подхода дают близкие результаты. И действительно, минимизация RSS для подбора преобразования только для x дает значение $\lambda = 2.04$, а в использовании inverse plot для y не было необходимости; совместная минимизация RSS и по x , и по y дала значения параметров в преобразовании Бокса-Кокса 2.10 и 1.03 соответственно.

Пример 3.2. $X \sim \exp(2)$, $\varepsilon \sim N(0, 1)$, $Y = (2 + 0.8X + \varepsilon)^3$.

В этом примере мы хотим получить преобразование для y с $\lambda = 1/3$, а x оставить без изменений. Графики 7, 8, 9 отвечают преобразованиям (1), (2), (4) соответственно.

Здесь снова нормализация предиктора в отрыве от зависимой переменной не дала приемлемого результата. Подход (1) дает $\lambda = 2.28$ для x и $\mu = 1.06$ на inverse plot для y . В итоге часть точек превращаются в выбросы и теряется нормальность остатков. Это связано с тем, что Y является кубическим многочленом от X и ε :

$$Y = 0.512X^3 + 1.92X^2\varepsilon + 3.84X^2 + 2.4X\varepsilon^2 + 9.6X\varepsilon + 9.6X + \varepsilon^3 + 6\varepsilon^2 + 12\varepsilon + 8.$$

И для получения простой регрессии нам сначала нужно преобразовать y к $y^{1/3}$. Однако подход (1) предлагает действовать в обратном порядке – сначала преобразовать x , а потом y . Ввиду сложной изначальной зависимости y от x мы не можем описать ее единственным предиктором $\psi_*(x, \lambda)$, отсюда и получаются плохие результаты.

Совместный подбор параметров через минимизацию RSS снова дает наиболее близкие к исходным значения 1.04 для x и 0.30 для y . Правильность этого подхода также подтверждается diagnostic plot'ом.

Пример 3.3. $X \sim \chi^2(3)$, $\varepsilon \sim N(0, 0.2)$, $Y = \arctg(X) + \varepsilon$.

Здесь имеется нестепенная зависимость Y от X . Поэтому ни один из методов не сможет дать "правильный" ответ, однако для нас представляет интерес, какой из методов окажется наиболее близким. Diagnostic plot'ы 10, 11, 12 отвечают преобразованиям (1), (2), (4) соответственно.

После нормализации x на графиках страдает Residuals vs Fitted график, что ставит под сомнение адекватность построенной модели. Методы (1) и (4) снова дали похожие результаты, возможно, потому что для y не требуется никакого преобразования, и оба метода концентрируются на x . Имеем:

- $\lambda = 0.01$ для x и $\mu = 0.55$ для y в методе (1);
- $\lambda = -0.01$ для x и $\mu = 0.91$ для y в методе (4).

Существенным отличием является то, что у метода (4) остатки получились существенно более похожими на нормальные, чем в методе (1). Отличия остальных трех графиков не являются настолько значительными.

Пример 3.4. $X \sim R[0, 3]$, $\varepsilon \sim \exp(1)$, $Y = X\varepsilon$.

Теперь рассмотрим гетероскедастичную модель, ошибка ε здесь является мультипликативной. Решением такой задачи будет являться логарифмирование и предиктора, и зависимой переменной. Тогда задача будет иметь вид:

$$\log Y = \log X + \log \varepsilon = \log X + \mathbf{E} \log \varepsilon + \tilde{\varepsilon}, \quad (18)$$

где $\tilde{\varepsilon} \sim N(0, \tilde{\sigma}^2)$.

В этом примере все рассматриваемые методы получили разные результаты, однако качественно полученные модели оказались достаточно похожими.

- $\lambda = 3$ для x (крайнее значение из заданного диапазона) и $\mu = 1$ для y (inverse plot не применялся) в подходе (1);
- $\lambda = 0.78$ для x и $\mu = -0.17$ для y в подходе (2);
- $\lambda = 1.01$ для x и $\mu = -0.07$ для y в подходе (4);

Во всех моделях имеется наклон на графике Scale-Location. Это говорит нам о том, что в результате преобразований мы не смогли достичь гомоскедастичности (наиболее близким к горизонтальной линии оказался Scale-Location plot для подхода (4)). Также мы можем наблюдать наличие "хвостов" на графиках QQ-plot. Наиболее похожие на нормальное распределение остатки получились в подходе (2).

Пример 3.5. $X \sim R[0, 3]$, $\varepsilon \sim N(0, 0.25)$, $Y = \exp(X\varepsilon)$.

Этот пример является некоторой комбинацией примеров 3.3 и 3.4.

Рассматриваемые нами методы не рассчитаны на работу с нестепенными зависимостями, однако не всегда можно с помощью только визуальных методов отличить одну задачу от другой. Например, визуальной зависимости y от x в примерах 3.4 и 3.5 очень похожи (см. рисунок 3d и 3e).

Как можно видеть на diagnostic plot'ах на рисунках 16, 17 и 18 (соответствуют подходам (1), (2) и (4)), у всех методов получились достаточно плохие модели: нормальности остатков нет ни у кого, во всех случаях присутствует гетероскедастичность.

3.2 Задачи с несколькими предикторами

4 Заключение

ТВА

А Многомерное преобразование Бокса-Кокса

Пусть у нас есть многомерная сл.в. $\mathbf{X} = (X_1, X_2, \dots, X_p)$. Введем преобразование:

$$\psi_M(\mathbf{X}, \boldsymbol{\lambda}) = (\psi_M(X_1, \lambda_1), \psi_M(X_2, \lambda_2), \dots, \psi_M(X_p, \lambda_p)). \quad (19)$$

Предположим, что существует такое $\boldsymbol{\lambda}$, что

$$\psi_M(\mathbf{X}, \boldsymbol{\lambda}) \sim N(\boldsymbol{\mu}, \mathbf{V}), \quad (20)$$

где \mathbf{V} – некоторая симметричная положительно определенная матрица, которую мы хотим оценить. Если \mathbf{x} – это наблюдения из распределения \mathbf{X} , то правдоподобие имеет вид

$$L(\boldsymbol{\lambda}) = \prod_{i=1}^n \frac{1}{(2\pi|\mathbf{V}|)^{1/2}} \exp \left(-\frac{1}{2} (\psi_M(\mathbf{x}_i, \boldsymbol{\lambda}) - \boldsymbol{\mu})^T \mathbf{V}^{-1} (\psi_M(\mathbf{x}_i, \boldsymbol{\lambda}) - \boldsymbol{\mu}) \right), \quad (21)$$

где $|\mathbf{V}|$ – определитель матрицы \mathbf{V} . Тогда логарифм правдоподобия имеет вид:

$$\begin{aligned} \log(L(\boldsymbol{\lambda})) = \ell(\boldsymbol{\lambda}) = & -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log(|\mathbf{V}|) - \\ & - \frac{1}{2} \sum_{i=1}^n \left(-\frac{1}{2} (\psi_M(\mathbf{x}_i, \boldsymbol{\lambda}) - \boldsymbol{\mu})^T \mathbf{V}^{-1} (\psi_M(\mathbf{x}_i, \boldsymbol{\lambda}) - \boldsymbol{\mu}) \right). \end{aligned} \quad (22)$$

При фиксированном $\boldsymbol{\lambda}$ это логарифм правдоподобия многомерного нормального распределения. Мы можем найти оценки максимума правдоподобия для $\boldsymbol{\mu}$ и \mathbf{V} :

$$\boldsymbol{\mu}(\boldsymbol{\lambda}) = \frac{1}{n} \sum_{i=1}^n \psi_M(\mathbf{x}_i, \boldsymbol{\lambda});$$

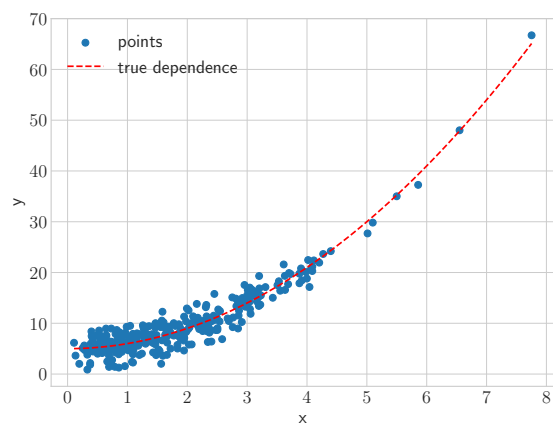
$$\mathbf{V}(\boldsymbol{\lambda}) = \frac{1}{n} \sum_{i=1}^n (\psi_M(\mathbf{x}_i, \boldsymbol{\lambda}) - \boldsymbol{\mu})(\psi_M(\mathbf{x}_i, \boldsymbol{\lambda}) - \boldsymbol{\mu})^T.$$

Подставляя эти оценки в (22), получим:

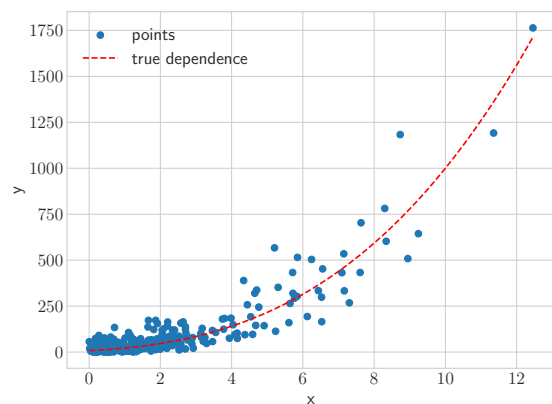
$$\ell(\boldsymbol{\lambda}) = -\frac{n}{2} - \frac{n}{2} \log(2\pi) - \frac{n}{2} \log(|\mathbf{V}(\boldsymbol{\lambda})|) \quad (23)$$

Максимизация (23) по $\boldsymbol{\lambda}$ равносильна минимизации определителя $\mathbf{V}(\boldsymbol{\lambda})$ по $\boldsymbol{\lambda}$.

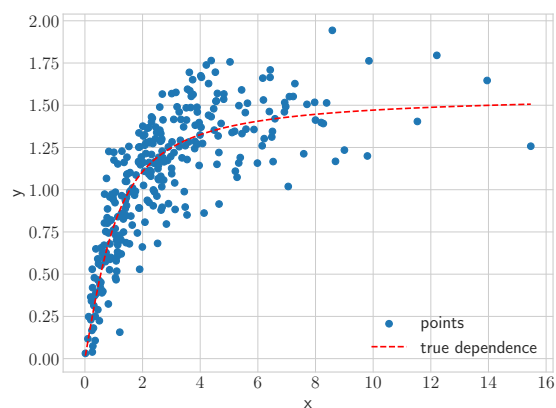
В Графики



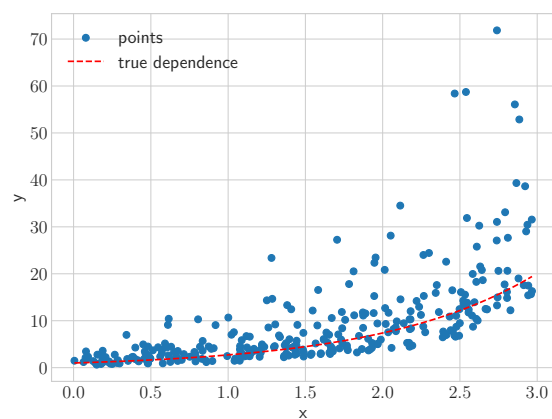
(a) Scatter plot y от x для примера 3.1



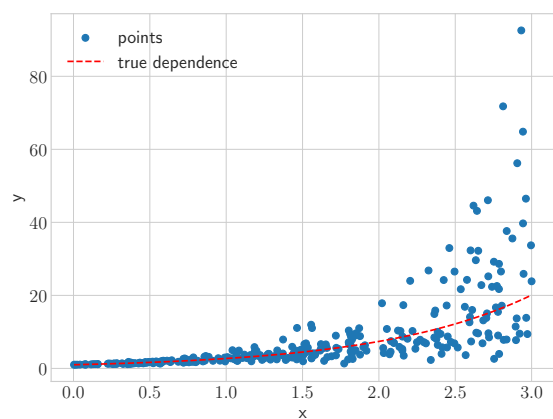
(b) Scatter plot y от x для примера 3.2



(c) Scatter plot y от x для примера 3.3



(d) Scatter plot y от x для примера 3.4



(e) Scatter plot y от x для примера 3.5

Рис. 3: Scatter plot'ы для примеров 3.1-3.5

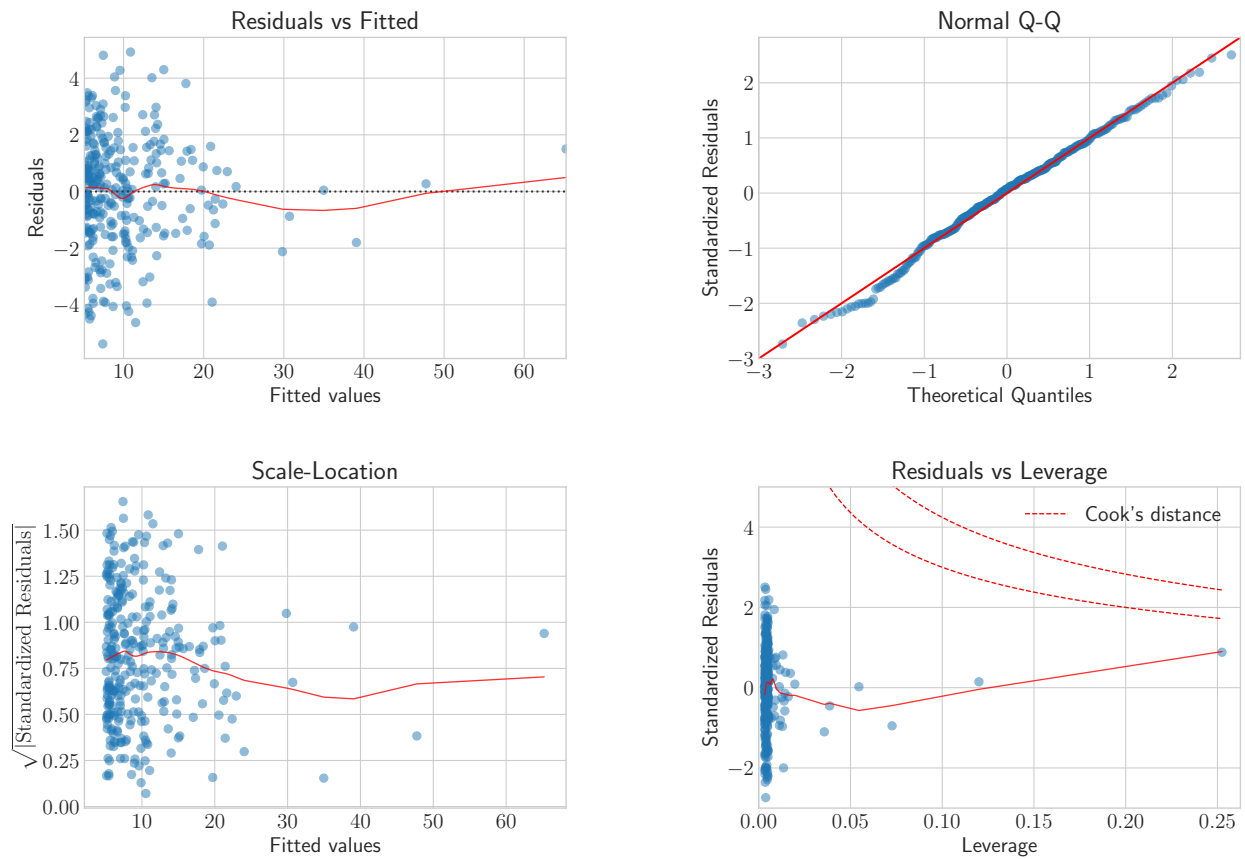


Рис. 4: Diagnostic plot для подхода (1) в примере 3.1

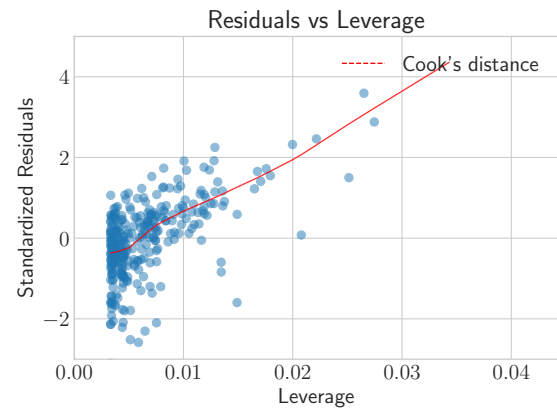
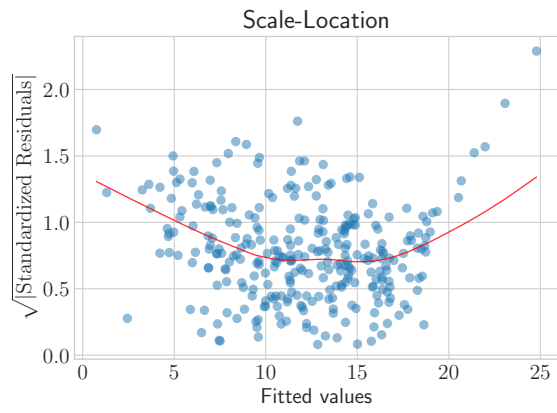
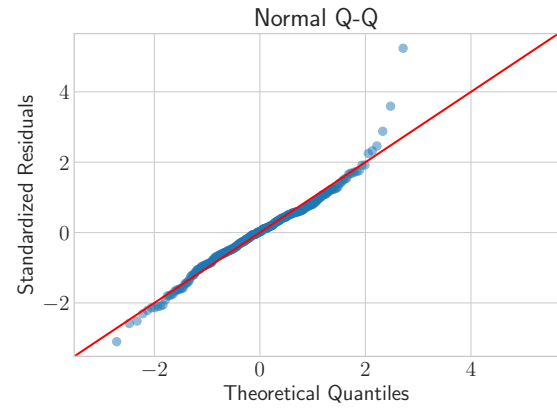
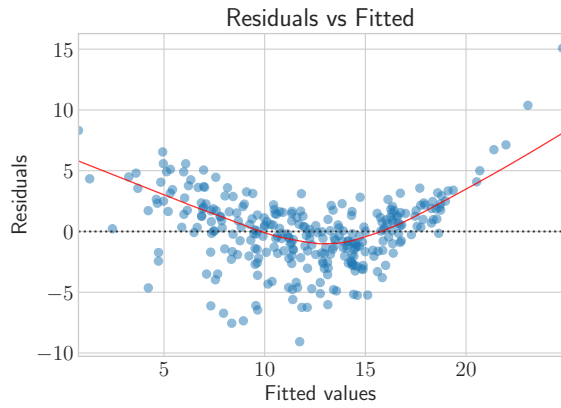


Рис. 5: Diagnostic plot для подхода (2) в примере 3.1

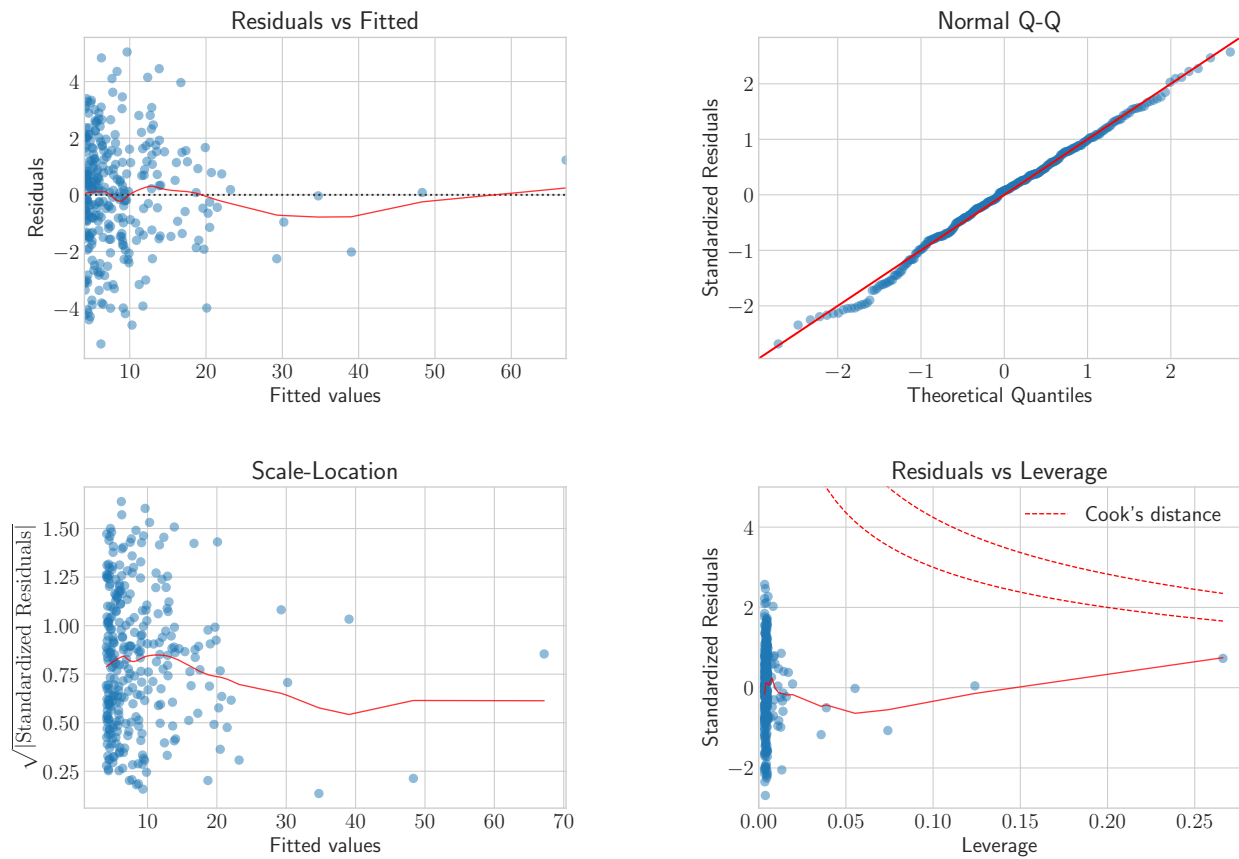


Рис. 6: Diagnostic plot для подхода (4) в примере 3.1

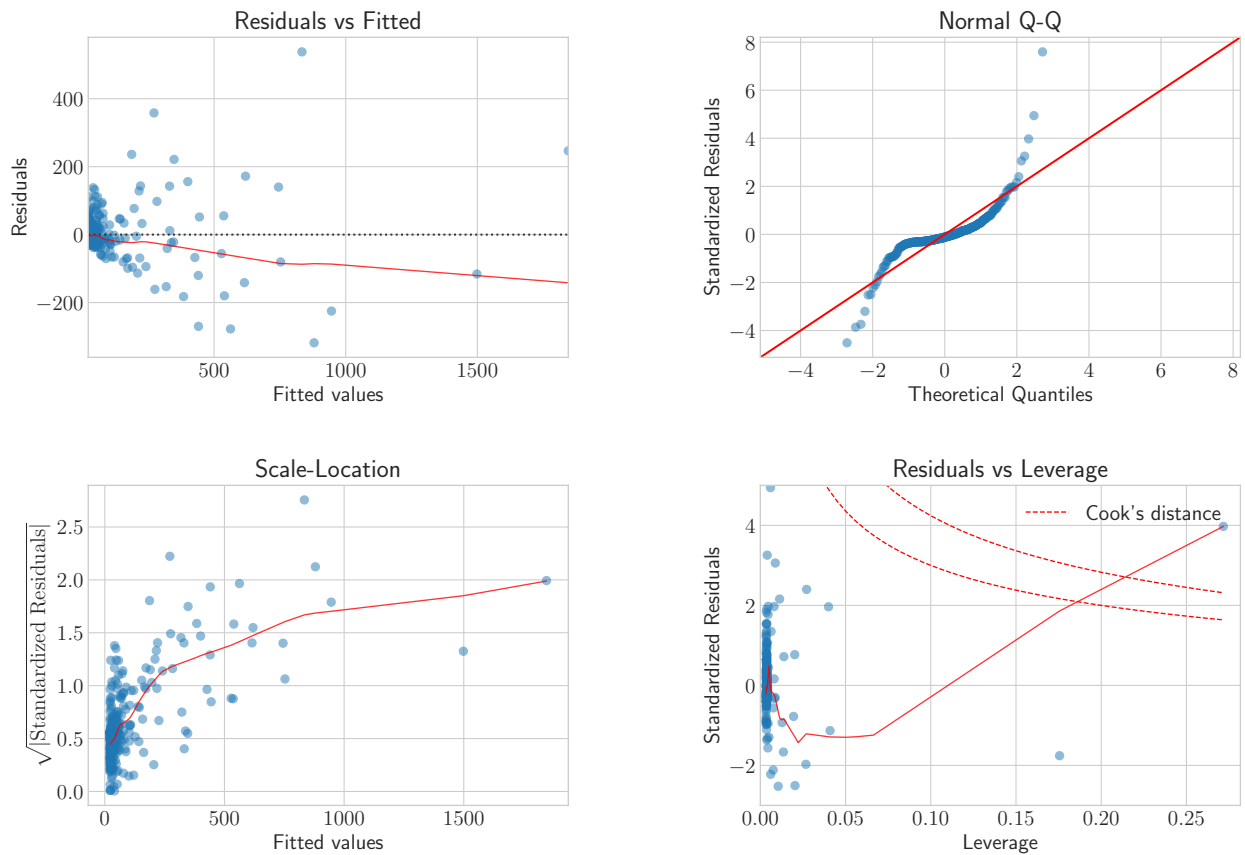


Рис. 7: Diagnostic plot для подхода (1) в примере 3.2

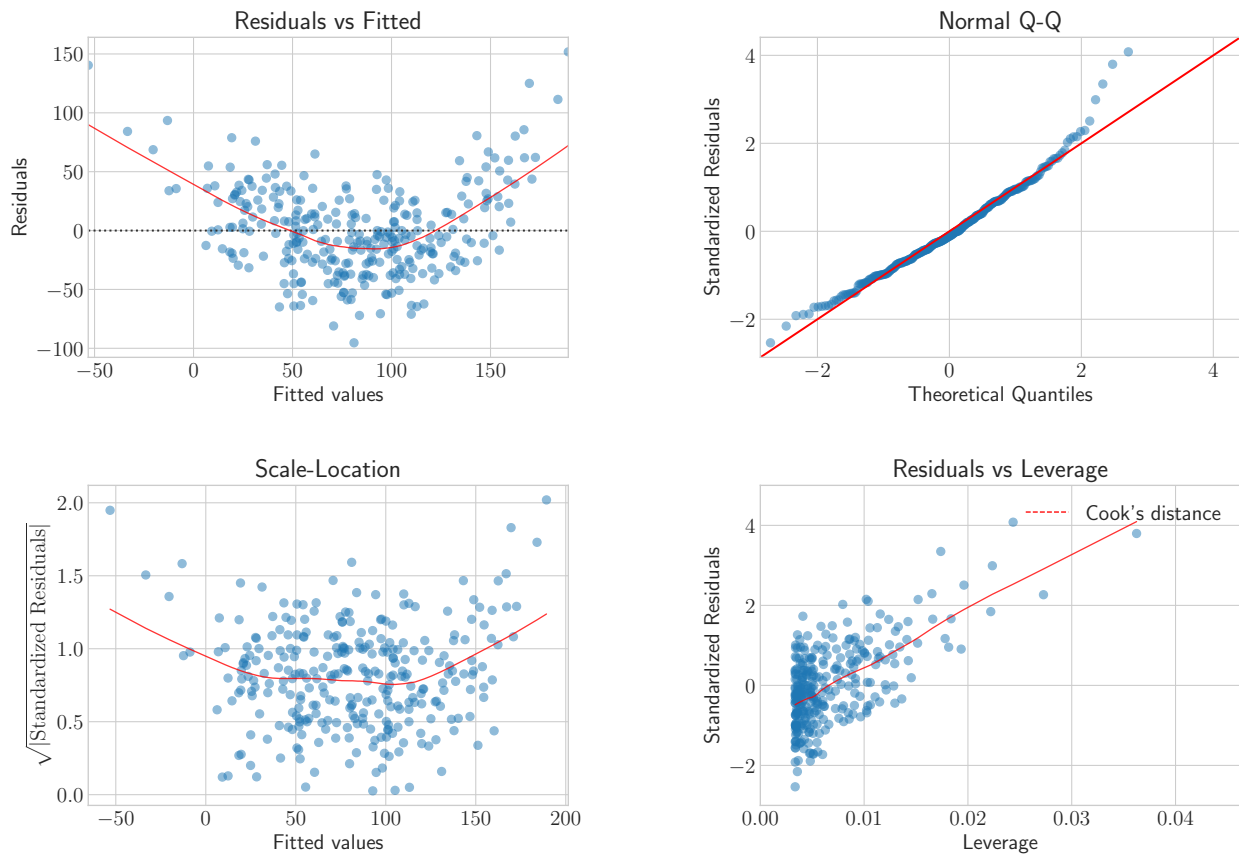


Рис. 8: Diagnostic plot для подхода (2) в примере 3.2

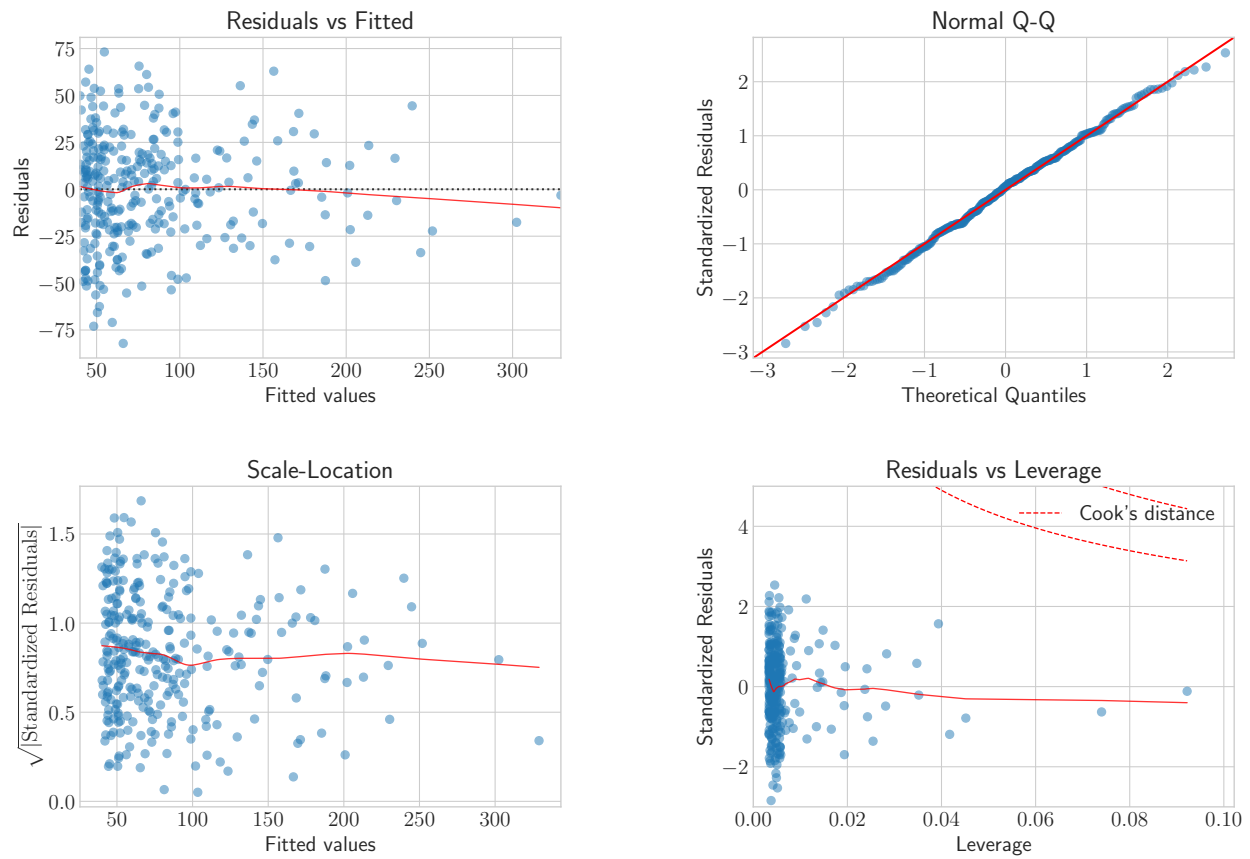


Рис. 9: Diagnostic plot для подхода (4) в примере 3.2

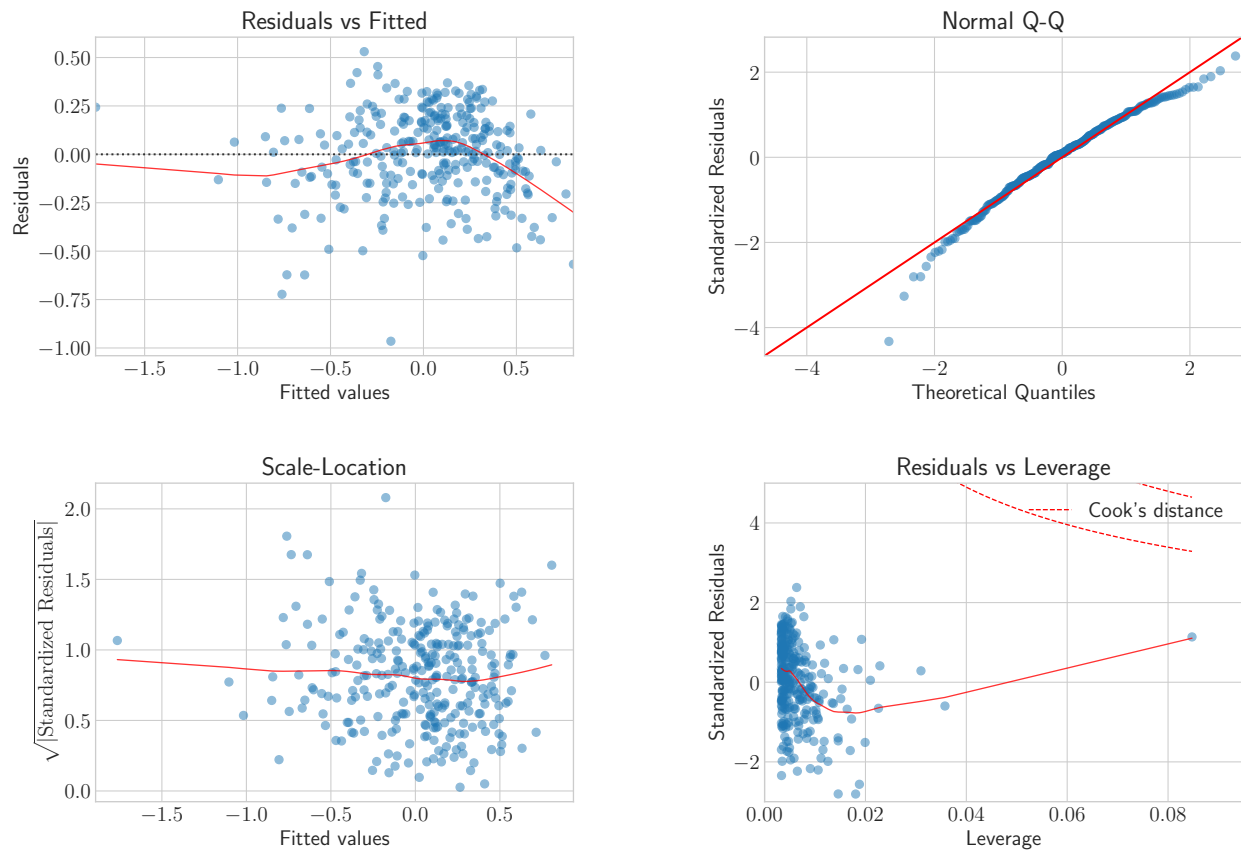


Рис. 10: Diagnostic plot для подхода (1) в примере 3.3

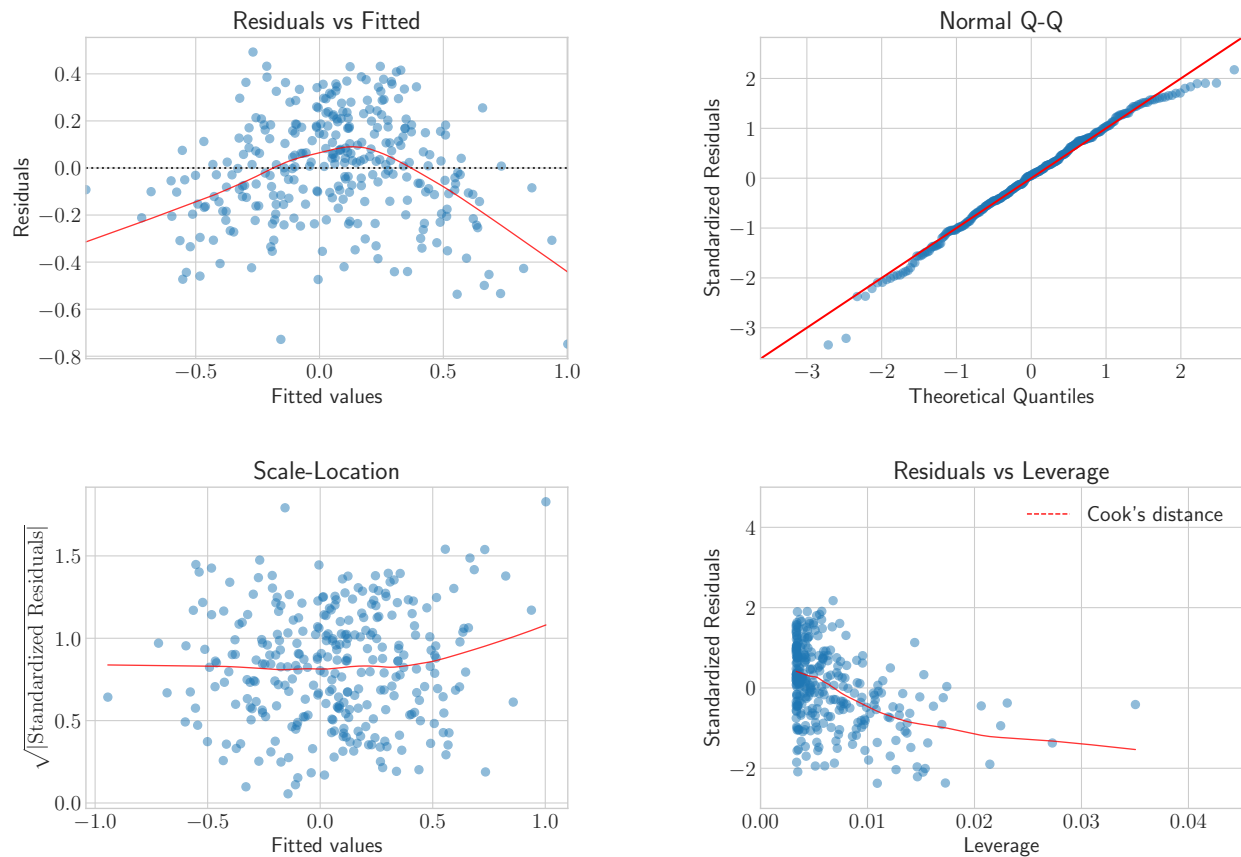


Рис. 11: Diagnostic plot для подхода (2) в примере 3.3

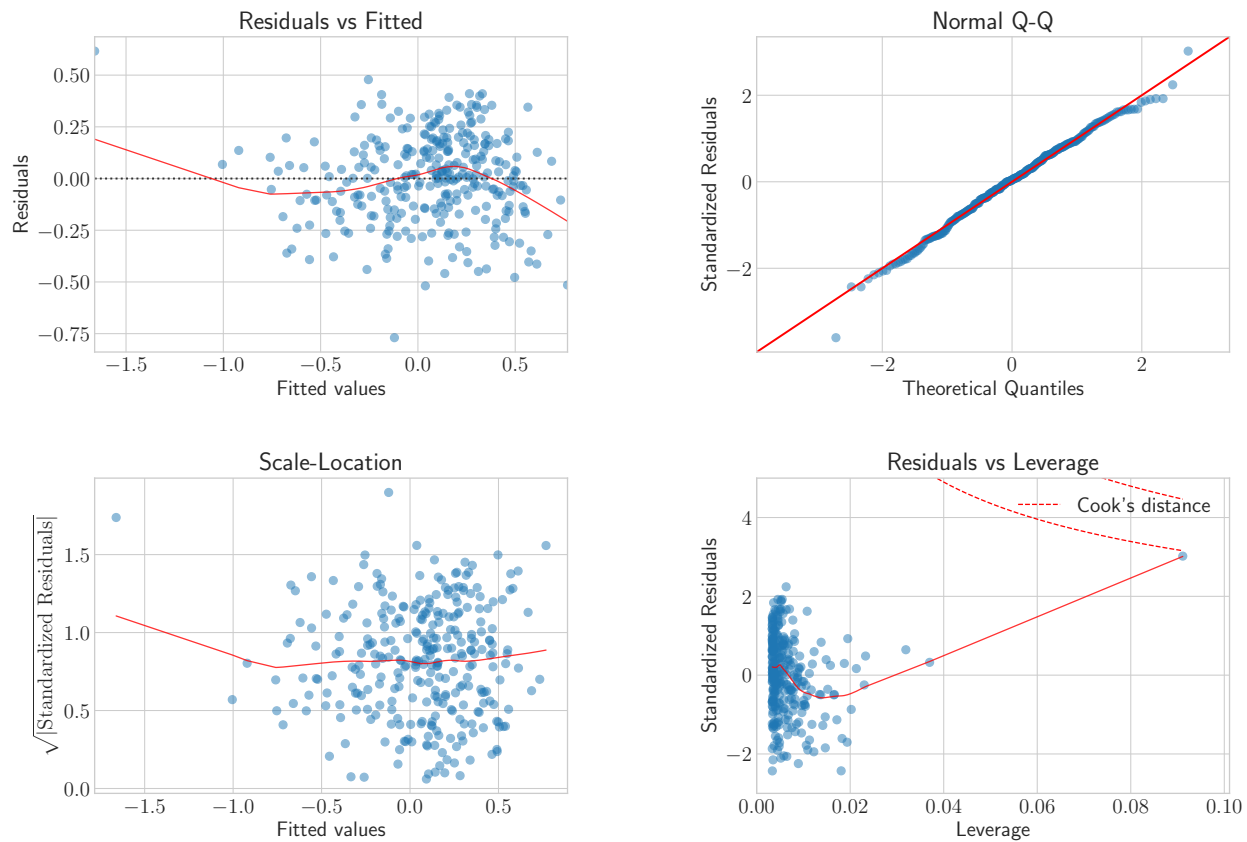


Рис. 12: Diagnostic plot для подхода (4) в примере 3.3

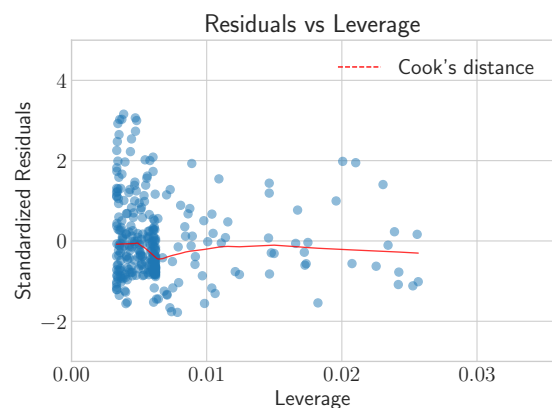
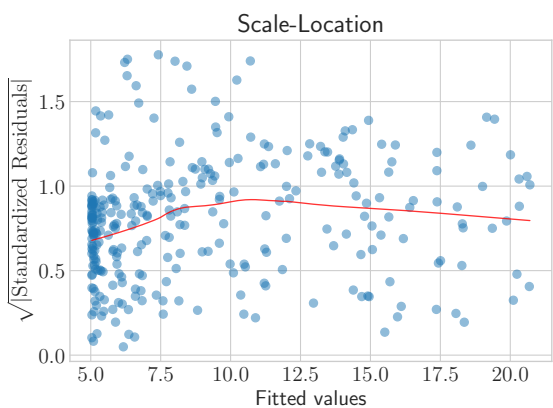
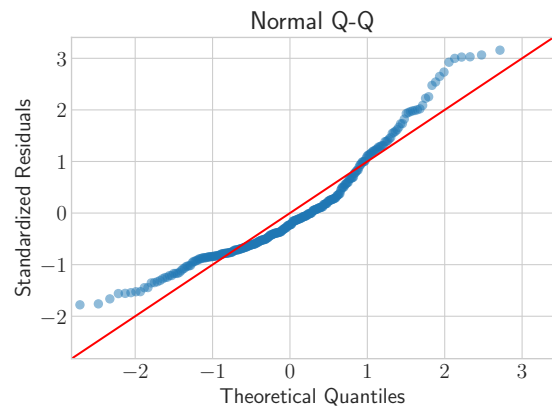
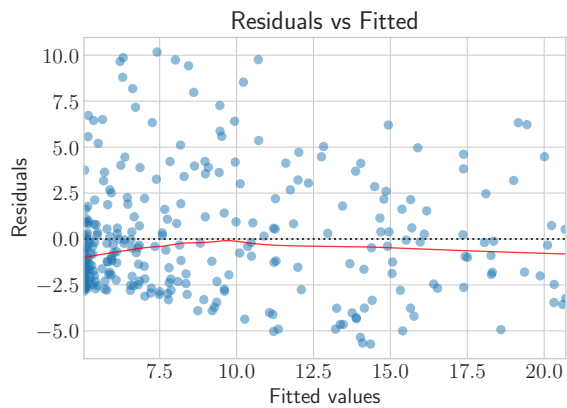


Рис. 13: Diagnostic plot для подхода (1) в примере 3.4

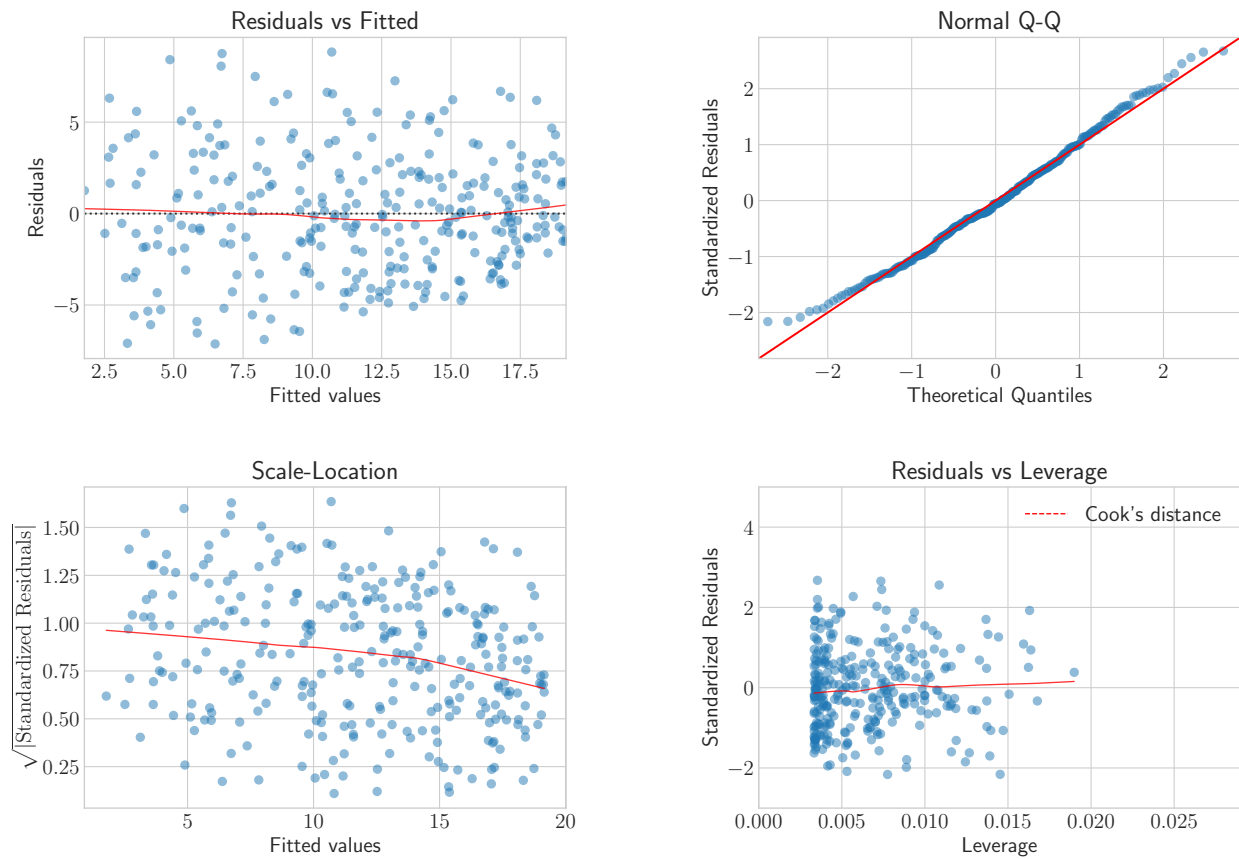


Рис. 14: Diagnostic plot для подхода (2) в примере 3.4

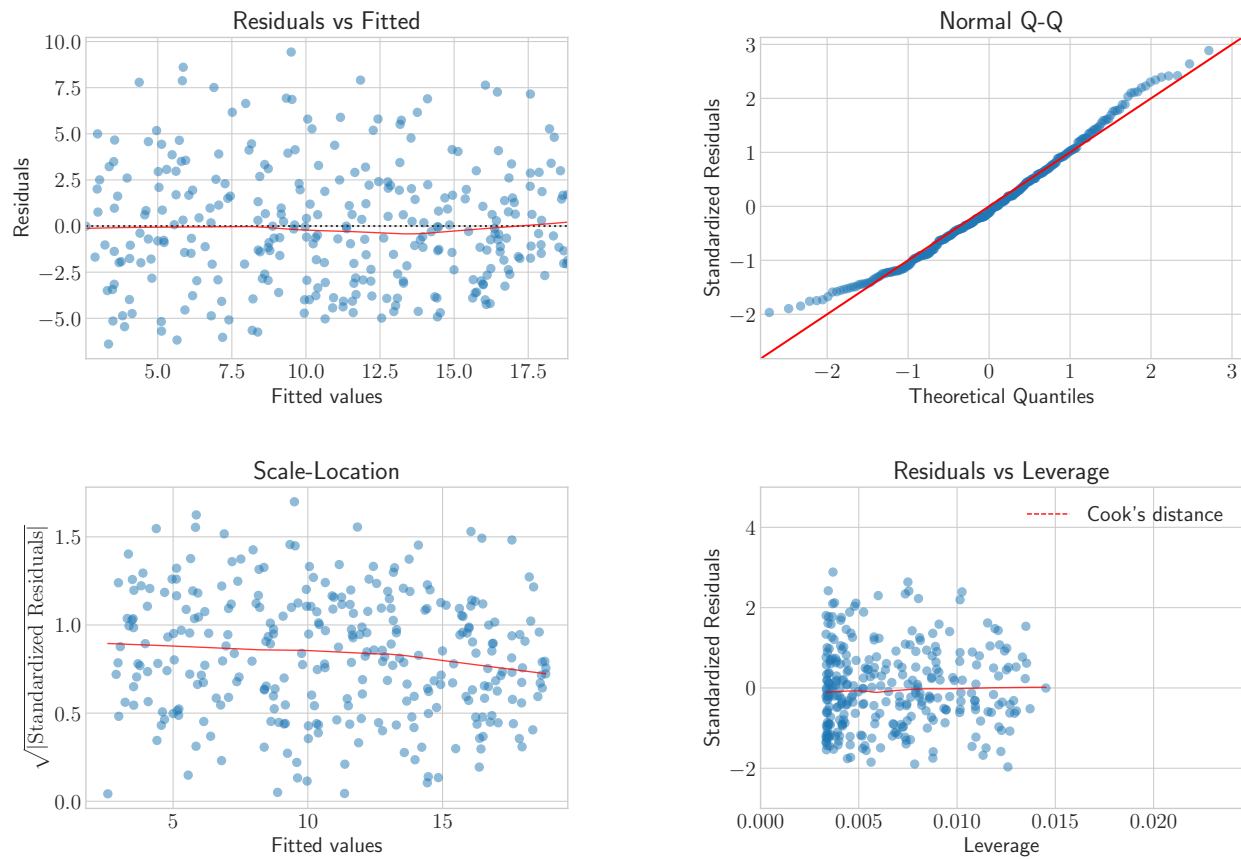


Рис. 15: Diagnostic plot для подхода (4) в примере 3.4

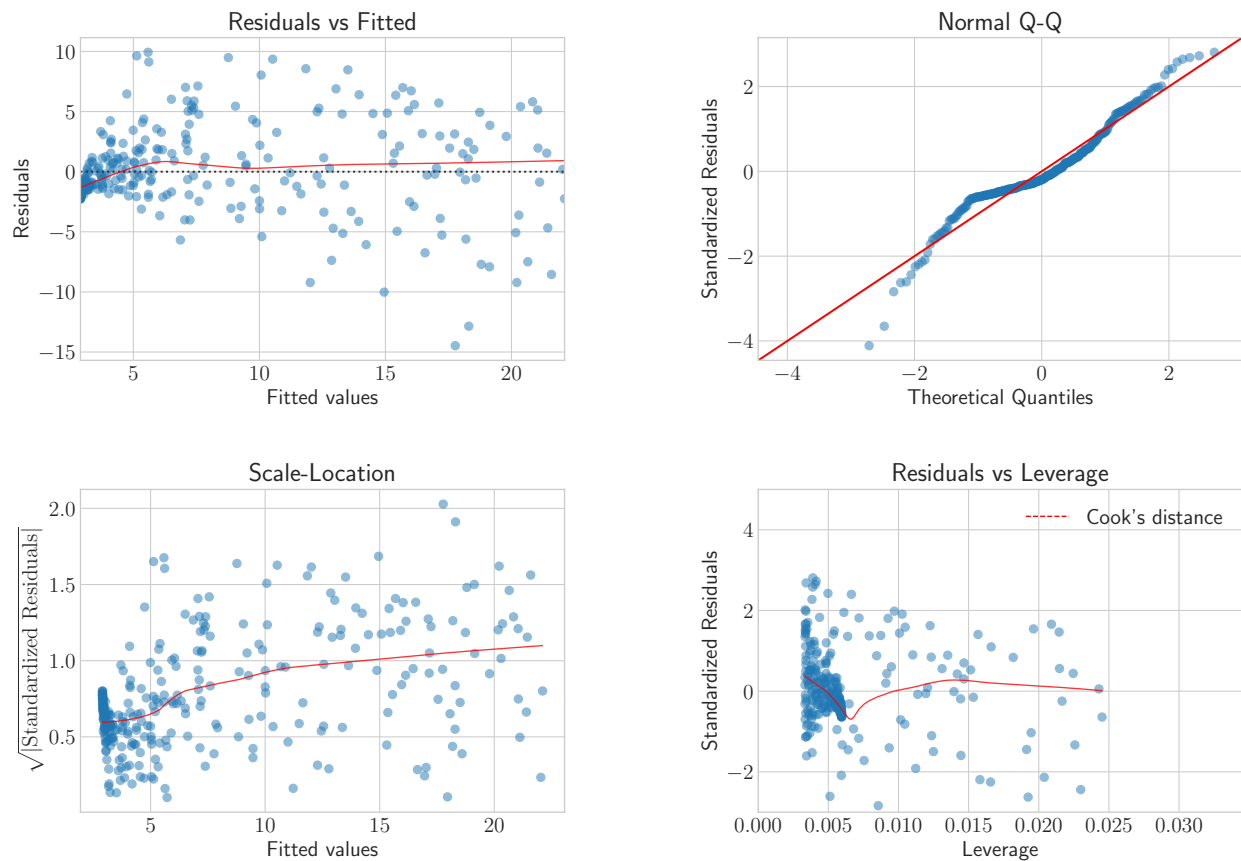


Рис. 16: Diagnostic plot для подхода (1) в примере 3.5

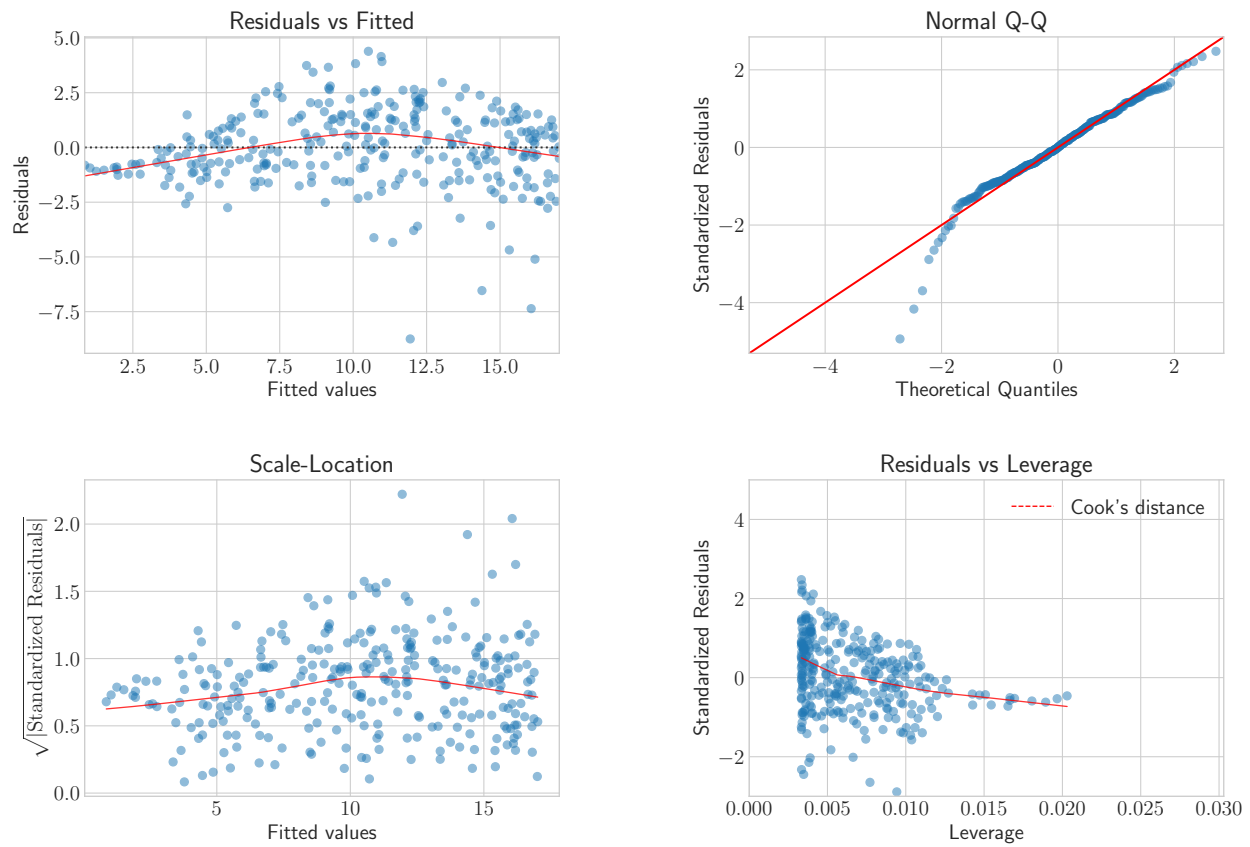


Рис. 17: Diagnostic plot для подхода (2) в примере 3.5

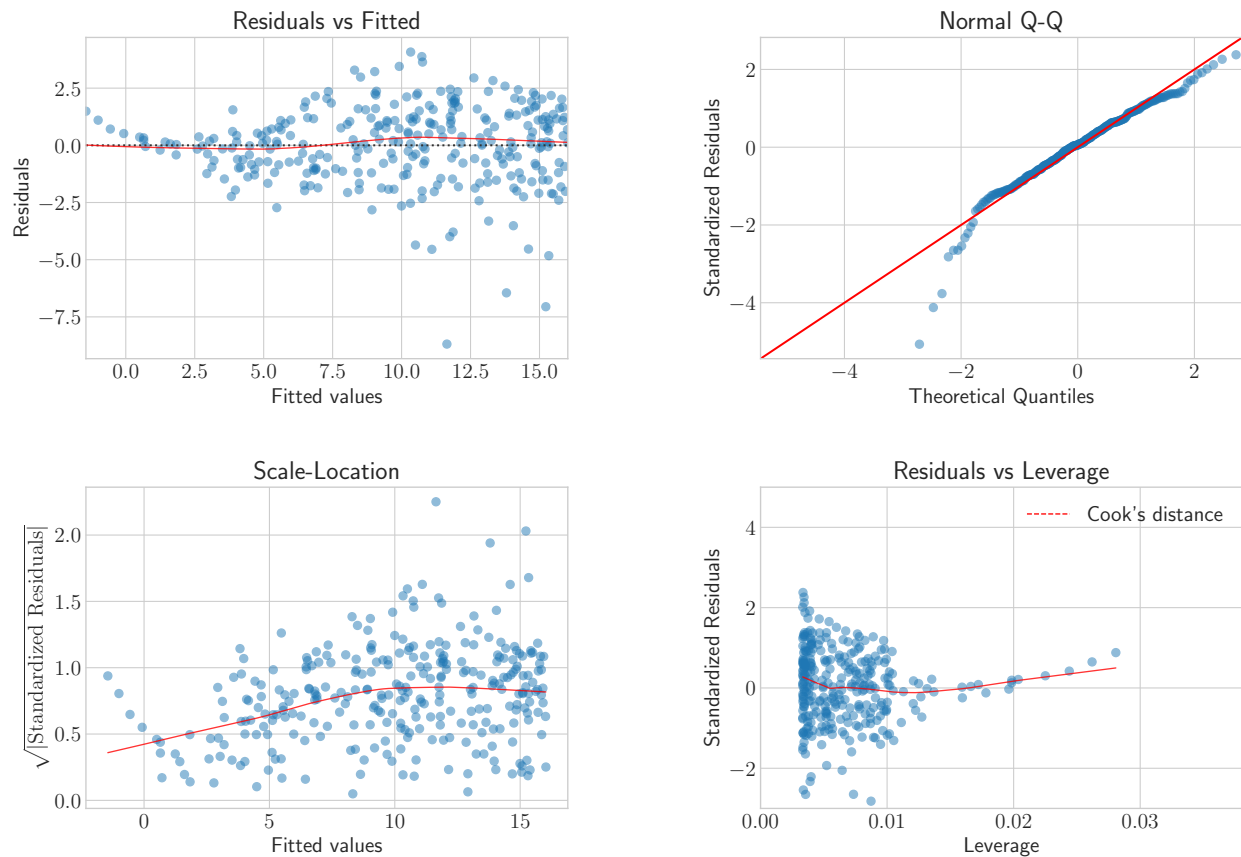


Рис. 18: Diagnostic plot для подхода (4) в примере 3.5

Список литературы

- [Berger and Casella, 2001] Berger, R. and Casella, G. (2001). Statistical Inference. Duxbury Press, Florence, AL, 2 edition.
- [Box and Cox, 1964] Box, G. E. P. and Cox, D. R. (1964). An analysis of transformations. Journal of the Royal Statistical Society. Series B (Methodological), 26(2):211–252.
- [Cook and Weisberg, 1994] Cook, R. D. and Weisberg, S. (1994). Transforming a response variable for linearity. Biometrika, 81(4):731–737.
- [Velilla, 1993] Velilla, S. (1993). A note on the multivariate Box–Cox transformation to normality. Statistics & Probability Letters, 17(4):259–263.
- [Weisberg et al., 1978] Weisberg, H., Beier, E., Brody, H., Patton, R., Raychaudhuri, K., Takeda, H., Thern, R., and Van Berg, R. (1978). s -dependence of proton fragmentation by hadrons. ii. incident laboratory momenta 30-250 gev/c. Phys. Rev. D, 17:2875–2887.