Задача 3 – Регрессия

Дебит: $q_i(t) = q_{0i} f(t) \qquad \left(Q = q_{0i} \int\limits_{t_1}^{t_2} f(t)\,dt \right)$

Модель: $\ln q_i(t_j) = \ln q_{0i} + \ln\left(f(t_j)\right) + \varepsilon_{ij}$

i – индекс скважины, j – индекс времени.

$\exists$ $A_{ij} = \ln q_i(t_j) = \ln \frac{Q_{ij}}{T_j}$ ~ текущий дебит i-й скв.

$B_i = \ln q_{0i} = \ln \frac{Q_{i0}}{T_0}$ ~ начальный дебит i-й скв.

$\Rightarrow \quad A_{ij} = B_i + \ln f(t_j) + \varepsilon_{ij}$

1. $f(t) = e^{-Dt}, \quad D>0$

$A_{ij} = B_i - D t_j + \varepsilon_{ij}$

Невязка: $\sum\limits_{ij} \varepsilon_{ij}^2 \xrightarrow[D]{} \min$

$$D = \arg\min_{D>0} \sum_{ij} \left(A_{ij} - B_i + D t_j\right)^2 = \arg\min_{D>0} \Phi$$

$$\frac{\partial \Phi}{\partial D} = 0 \Rightarrow \sum_{ij}\left(D t_j^2 + A_{ij} t_j - B_i t_j\right) = 0$$

$$D = \frac{1}{\sum\limits_{ij} t_j^2} \left(\sum_{ij} B_i t_j - \sum_{ij} A_{ij} t_j \right)$$

2. $f(t) = (1+t)^{-a}, \quad a>0$

$A_{ij} = B_i - a\ln(1+t_j) + \varepsilon_{ij}$

$$a = \arg\min \sum_{ij} \left(A_{ij} - B_i + a\ln(1+t_j)\right)^2 = \arg\min \Phi$$

$$a = \frac{1}{\sum_{ij} \ln^2(1+t_j)} \left(\sum_{ij} B_i \ln(1+t_j) - \sum_{ij} A_{ij} \ln(1+t_j) \right)$$

3. $f(t) = (1 + bDt)^{-1/b}$, $\{b, D\} > 0$

$$A_{ij} = B_i - \frac{1}{b} \ln(1 + bDt_j) + \varepsilon_{ij}$$

$$\{b, D\} = \arg\min_{\substack{b>0 \\ D>0}} \Phi = \arg\min \sum_{ij} \left(A_{ij} - B_i + \frac{1}{b} \ln(1+bDt_j) \right)^2$$

$$\left\{ \frac{\partial \Phi}{\partial b} = 0 \quad \frac{\partial \Phi}{\partial D} = 0 \right\}$$

Получается система нелинейных уравнений.

Если <u>линеаризуется</u>, то стандартная регрессия, матрица плана сст.-д.

Если <u>не линеаризуется</u>, то численными методами (метод итераций, построение нормального уравнения, методы поиска, например м-д Хука-Дживса)

4. $f(t) = \begin{cases} (1+t)^{-a}, & t \le \tau \\ (1 + bD(t-\tau))^{-1/b}, & t > \tau \end{cases}$; $\{a, b, \tau\} > 0$

Непрерывность при $t = \tau$: $f'_t\big|_{\tau-0} = f'_t\big|_{\tau+0} \Rightarrow$

$$\Rightarrow D = \frac{-a(1+\tau)^{-a}}{(1+\tau)}$$

$$f(t) = \begin{cases} (1+t)^{-a}, & t \le \tau \\ \left(1 - b(1+\tau)^{-a-1} a (t-\tau)\right)^{-\frac{1}{b}}, & t > \tau \end{cases}$$

$$\Theta(t-\tau) = \begin{cases} 1, & t>\tau \\ 0, & t\leqslant\tau \end{cases}$$

$$\ln f(t) = -a\left(1-\Theta(t-\tau)\right)\ln(1+t) - \frac{1}{b}\,\Theta(t-\tau)\ln\left(1-b(1+\tau)^{-a-1}a(t-\tau)\right)$$

$$\{a, b, \tau\} = \underset{\substack{a>0\\ b>0\\ \tau>0}}{\operatorname{argmin}} \sum_{ij}\left(A_{ij}-B_i-\ln f_{a,b,\tau}(t_j)\right)^2$$

Анализ остатков

1. Коэффициент детерминации (доля объяснённой дисперсии)

$$R^2 = \left(1 - D_1/D_2\right)$$

$$D_1 = \frac{1}{N}\sum_i (q_i-\tilde{q}_i)^2$$

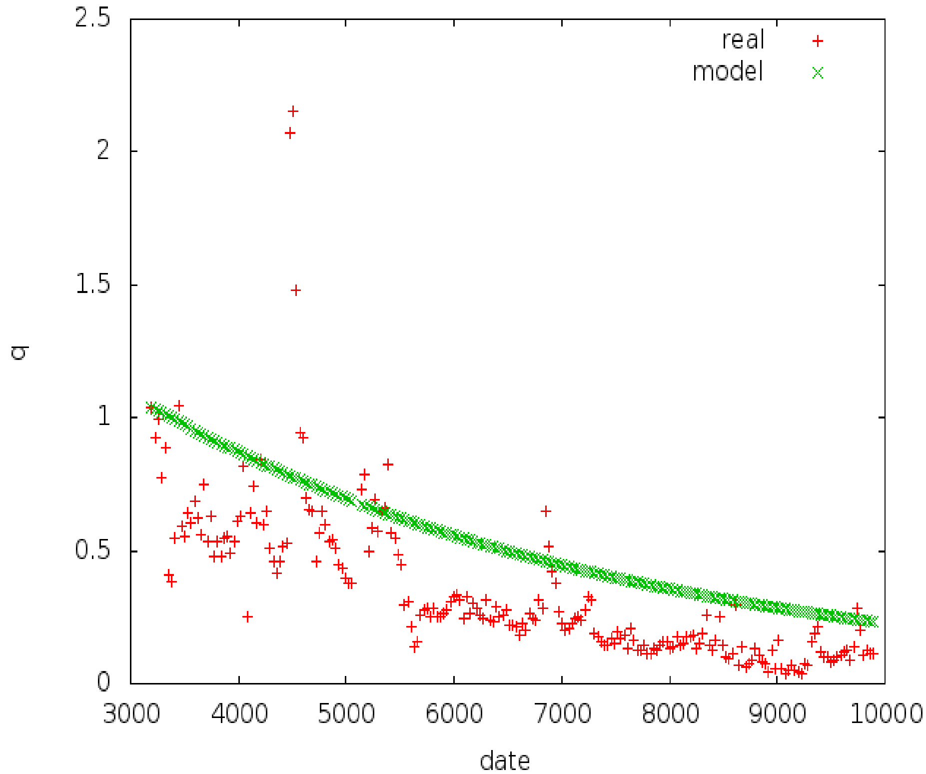
$$D_2 = \frac{1}{N}\sum_i (q_i-\langle q\rangle)^2$$

$$\langle q\rangle = \frac{1}{N}\sum_i q_i$$

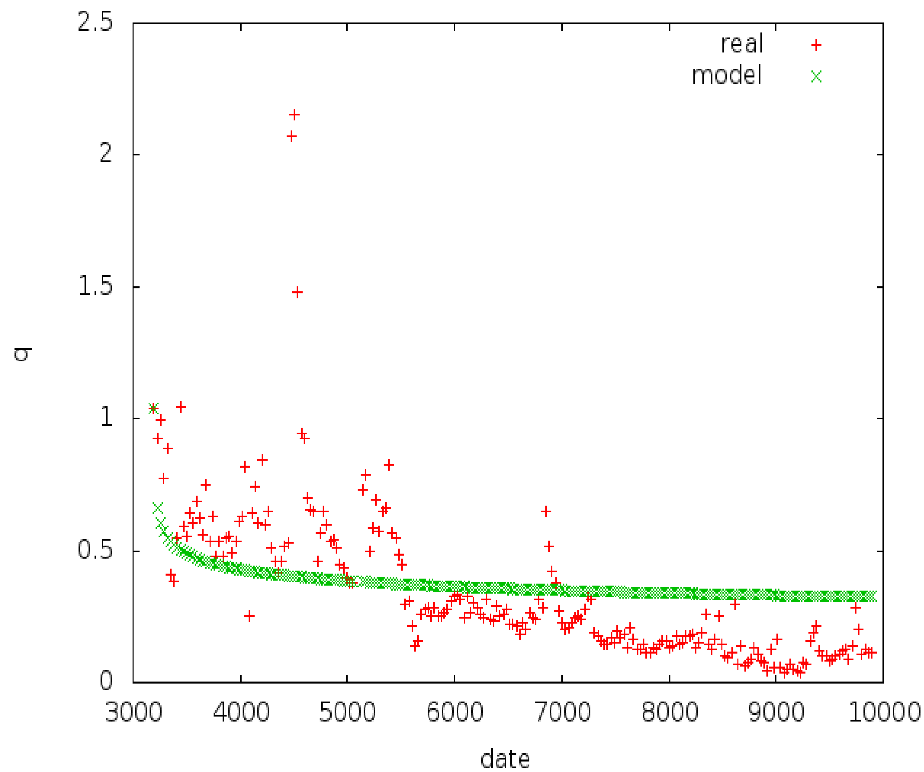
q_i – реальные данные

$\tilde{q}_i$ – модельные данные

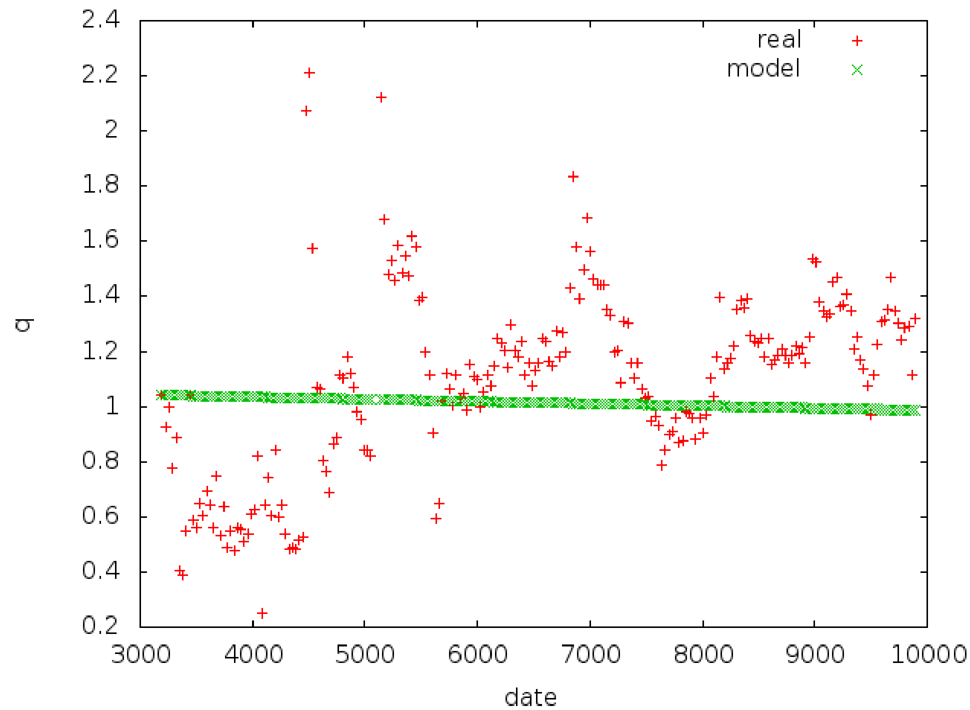
2. Гетероскедастичность (непостоянство дисперсии остатков)



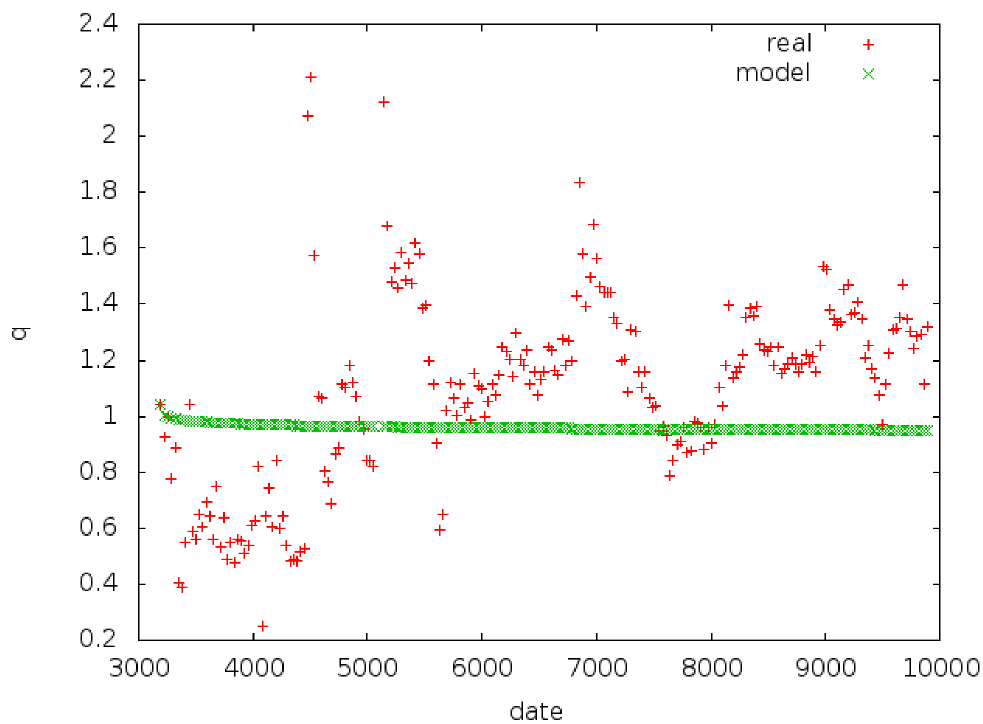
Реальный и модельный дебит нефти в скв. 2138, функция падения $f(t) = e^{-Dt}$, $R^2 = 0.077$ (по всему месторождению).



Реальный и модельный дебит нефти в скв. 2138, функция падения $f(t) = (1+t)^{-a}$, $R^2 = 0.047$ (по всему месторождению).



Реальный и модельный дебит воды в скв. 2138, функция падения $f(t) = e^{-Dt}$, $R^2 = 0.159$ (по всему месторождению).



Реальный и модельный дебит воды в скв. 2138, функция падения $f(t) = (1+t)^{-a}$, $R^2 = 0.149$ (по всему месторождению).