linear and bilinear regression

athos.fiori

November 2020

1 Introduction

Imagine we would like to infere models as the four one presented in the figure

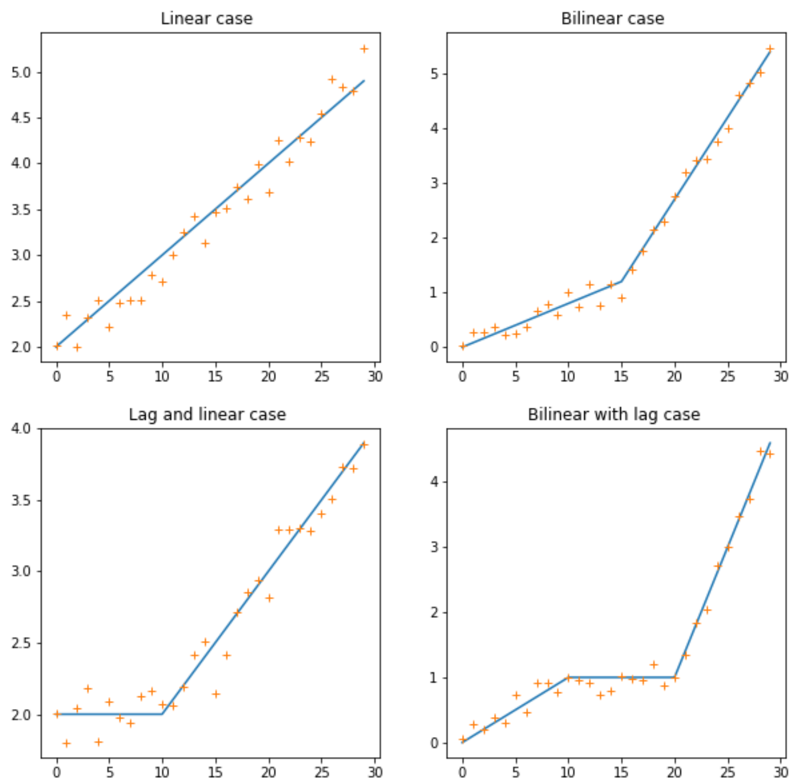

Figure 1: The four models (blu lines) we would like to predict. The linear model is treated in the first section. The Lag plus growth is treated in the second section. The bilinear and growth-lag-growth is treated in the third section.

1.1 Linear regression

Let's start by doing a simple linear regression. We consider the observations $y(t) = \delta t + x_0$ deviating from the linear model only trough gaussian noise i.e.

$$y_i = \delta t_i + x_0 + \epsilon \tag{1}$$

with $\epsilon \sim \mathcal{N}(0, \sigma^2)$. The likelihood of the dataset $D = (y_1, ..., y_N)$ reads

$$p\left(D|\delta, x_0, \sigma^2\right) = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{N}(y_i - \delta t_i - x_0)^2} \tag{2}$$

Define $\vec{Y} = (y_1, \ldots, y_N)^T$, $\vec{T} = (0, 1, \ldots, N)^T$ and $\vec{O} = (1, \ldots, 1)^T$ then we can write 2 as

$$p\left(D|\delta, x_0, \sigma^2\right) = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} e^{-\frac{1}{2\sigma^2}\left(\vec{Y} - \delta\vec{T} - x_0\vec{O}\right)^T\left(\vec{Y} - \delta\vec{T} - x_0\vec{O}\right)} \tag{3}$$

If we use an uniform prior $p(x_0) = dx_0$ and marginalize over x_0 we find

$$p\left(D|\delta, \sigma^2\right) = \frac{1}{\sqrt{N}(2\pi\sigma^2)^{\frac{N-1}{2}}} e^{-\frac{N}{2\sigma^2}Var(y-\delta t)} = \frac{1}{\sqrt{N}(2\pi\sigma^2)^{\frac{N-1}{2}}} e^{-\frac{1}{2\sigma^2}\left(\vec{Y}^c - \delta\vec{T}^c\right)^T\left(\vec{Y}^c - \delta\vec{T}^c\right)} \tag{4}$$

where the superscript c stands for the centered vector i.e. $\vec{Y}^c = \vec{Y} - \langle y\rangle\vec{O}$ and similar for T.
We now marginalize over σ using the prior $p(\sigma) = \frac{1}{\sigma}d\sigma$ and we obtain

$$P(D|\delta) \propto \left(\left(\vec{Y}^c - \delta\vec{T}^c\right)^T\left(\vec{Y}^c - \delta\vec{T}^c\right)\right)^{\frac{1-N}{2}} \tag{5}$$

The maximum likelihood estimator (MLE) is found by solving

$$\frac{d\log P(D|\delta)}{d\delta} = 0 \tag{6}$$

and we find

$$\delta^\star = \frac{\vec{Y}^c \cdot \vec{T}^C}{\vec{T}^C \cdot \vec{T}^C} = \frac{Cov(y,t)}{Var(t)} \tag{7}$$

The solution of

$$\frac{d\log P(D|\sigma^2, \delta^\star)}{d\sigma^2} = 0 \tag{8}$$

imply the MLE to be

$$\sigma^{2\star} = \frac{Var(y - \delta^\star t)}{1 - \frac{1}{N}} = \frac{\left(\vec{Y} - \delta^\star\vec{T}^c\right)^2}{N-1} \tag{9}$$

and finally the MLE of the offset x_0 is given by

$$x_0^\star = -\delta^\star\langle t\rangle + \langle y\rangle \tag{10}$$

1.2 Lag plus linear regression

Define $\vec{Y} = (y_1, \ldots, y_N)^T$, $\vec{T}_\tau = (\underbrace{0, \ldots, 0}_{\tau}, 1, \ldots, N-\tau)^T$ and $\vec{O} = (1, \ldots, 1)^T$ where $\tau \in [0, \ldots, N]$. Then we can write the likelihood of the lag+linear growth model as

$$p\left(D|\delta, x_0, \sigma^2, \tau\right) = \frac{1}{\left(2\pi\sigma^2\right)^{\frac{N}{2}}} e^{-\frac{1}{2\sigma^2}\left(\vec{Y}-\delta\vec{T}_\tau - x_0\vec{O}\right)^T\left(\vec{Y}-\delta\vec{T}_\tau - x_0\vec{O}\right)} \tag{11}$$

Marginalize over x_0 and σ is exactly as before and we find

$$P(D|\delta, \tau) \propto \left(\left(\vec{Y}^c - \delta\vec{T}^c_\tau\right)^T \left(\vec{Y}^c - \delta\vec{T}^c_\tau\right)\right)^{\frac{1-N}{2}} \tag{12}$$

Consider an uniform prior for δ, the marginal likelihood is found through the Laplace method

$$P(D|\tau) = \int d\delta e^{\log P(D|\tau,\delta)} = e^{\log P(D|\tau,\delta^\star)} \sqrt{\frac{2\pi}{\left|\frac{\partial^2 \log P(D|\tau,\delta^\star)}{\partial^2\delta}\right|}} \tag{13}$$

where as before

$$\delta^\star = \frac{\vec{Y}^c \cdot \vec{T}^c_\tau}{\vec{T}^c_\tau \cdot \vec{T}^c_\tau} \tag{14}$$

Let's define

$$f(\delta) = \left(\vec{Y}^c - \delta\vec{T}^c_\tau\right)^T \left(\vec{Y}^c - \delta\vec{T}^c_\tau\right) \tag{15}$$

then

$$\left.\frac{\partial f}{\partial \delta}\right|_{\delta^\star} = 0 \tag{16}$$

and

$$\left.\frac{\partial^2 f}{\partial^2 \delta}\right|_{\delta^\star} = 2\vec{T}^c_\tau \cdot \vec{T}^c_\tau \tag{17}$$

so

$$\left.\frac{\partial^2 \log P(D|\tau, \delta^\star)}{\partial^2\delta}\right|_{\delta^\star} = \frac{1-N}{2}\frac{f''}{f^2} \tag{18}$$

In order to find the $\tau^\star$ which maximises the likelihood 24 we compute this likelihood for $\tau = 1, 2, \ldots, N$ and select $\tau^\star$ as the one with the largest likelihood. Then we build the vector $\vec{T}^c_{\tau^\star}$ and we can find the MLE slope

$$\delta^\star = \frac{\vec{Y}^c \cdot \vec{T}^c_{\tau^\star}}{\vec{T}^c_{\tau^\star} \cdot \vec{T}^c_{\tau^\star}} \tag{19}$$

measurment error

$$\sigma^{2\star} = \frac{\left(\vec{Y} - \delta^\star\vec{T}^c_{\tau^\star}\right)^2}{N-1} \tag{20}$$

and offset

$$x_0^\star = \frac{\left(\vec{Y}^c - \delta^\star \vec{T}^c_{\tau^\star}\right) \cdot \vec{O}}{N} \tag{21}$$

1.3 Growth lag growth or bilinear case

Define $\vec{Y} = (y_1, \ldots, y_N)^T$, $\vec{T}_{\tau_1} = (\underbrace{1, \ldots, \tau_1}_{\tau_1}, \underbrace{\tau_1, \ldots, \tau_1}_{N-\tau_1})^T$,$\vec{T}_{\tau_2} = (\underbrace{0, \ldots, 0}_{\tau_2}, \underbrace{1, \ldots, \tau_1}_{N-\tau_2})^T$ and $\vec{O} = (1, \ldots, 1)^T$ where $\tau_1, \tau_2 \in [0, \ldots, N]$ and $\tau_2 \geq \tau_1$. Then the likelihood can be written as

$$p\left(D|\delta_1, \delta_2, x_0, \sigma^2, \tau_1, \tau_2\right) = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} e^{-\frac{1}{2\sigma^2}\left(\vec{Y} - \delta_1 \vec{T}_{\tau_1} - \delta_2 \vec{T}_{\tau_2} - x_0 \vec{O}\right)^2} \tag{22}$$

again marginalize over x_0 and σ^2 is the same as before and we find

$$P(D|\delta_1, \tau_1, \delta_2, \tau_2) \propto \left(\left(\vec{Y}^c - \delta_1 \vec{T}^c_{\tau_1} - \delta_2 \vec{T}^c_{\tau_2}\right)^T \left(\vec{Y}^c - \delta_1 \vec{T}^c_{\tau_1} - \delta_2 \vec{T}^c_{\tau_2}\right)\right)^{\frac{1-N}{2}} \tag{23}$$

To marginalize over δ_1, δ_2 we again use the Laplace method

$$P(D|\tau_1, \tau_2) = e^{\log P(D|\tau_1, \delta_1^\star, \tau_2, \delta_2^\star)} \sqrt{\frac{2\pi}{\left|\frac{\partial^2 \log P(D|\tau_1, \delta_1^\star, \tau_2, \delta_2^\star)}{\partial \delta_1 \partial \delta_2}\right|}} \tag{24}$$

where

$$\delta_1^\star = \frac{(\vec{T}^c_{\tau_1} \cdot \vec{Y}^c)(\vec{T}^c_{\tau_2} \cdot \vec{T}^c_{\tau_2}) - (\vec{T}^c_{\tau_2} \cdot \vec{Y}^c)(\vec{T}^c_{\tau_1} \cdot \vec{T}^c_{\tau_2})}{\left(\vec{T}^c_{\tau_1} \cdot \vec{T}^c_{\tau_2}\right)^2 - (\vec{T}^c_{\tau_1} \cdot \vec{T}^c_{\tau_1})(\vec{T}^c_{\tau_2} \cdot \vec{T}^c_{\tau_2})} \tag{25}$$

$$\delta_2^\star = \frac{(\vec{T}^c_{\tau_2} \cdot \vec{Y}^c) - \delta_1^\star (\vec{T}^c_{\tau_1} \cdot \vec{T}^c_{\tau_2})}{\vec{T}^c_{\tau_2} \cdot \vec{T}^c_{\tau_2}} \tag{26}$$

The determinant of the Hessian matrix reads

$$\left|\frac{\partial^2 \log P(D|\tau_1, \delta_1^\star, \tau_2, \delta_2^\star)}{\partial \delta_1 \partial \delta_2}\right| = (1-N)^2 \frac{(\vec{T}^c_{\tau_1})^2 (\vec{T}^c_{\tau_2})^2 - (\vec{T}^c_{\tau_1} \cdot \vec{T}^c_{\tau_2})^2}{f^2} \tag{27}$$

where

$$f = \left(\left(\vec{Y}^c - \delta_1 \vec{T}^c_{\tau_1} - \delta_2 \vec{T}^c_{\tau_2}\right)^T \left(\vec{Y}^c - \delta_1 \vec{T}^c_{\tau_1} - \delta_2 \vec{T}^c_{\tau_2}\right)\right) \tag{28}$$

Again we compute numerically all the possible likelihood combinations $P(D|\tau_1, \tau_2)$ for $\tau_1 = 0, \ldots, N$;$\tau_2 = \tau_1, \ldots, N$. We then pick the best MLE $\tau_1^\star$ and $\tau_2^\star$. The MLE $\delta_{1,2}$ have been computed before and not surprisingly the MLE measurment error

$$\sigma^{2\star} = \frac{\left(\vec{Y} - \delta_1^\star \vec{T}^c_{\tau_1^\star} - \delta_2^\star \vec{T}^c_{\tau_2^\star}\right)^2}{N-1} \tag{29}$$

and offset

$$x_0^\star = \frac{\left(\vec{Y}^c - \delta_1^\star \vec{T}^c_{\tau_1^\star} + \delta_2^\star \vec{T}^c_{\tau_2^\star}\right) \cdot \vec{O}}{N} \tag{30}$$