MODELING TIME SERIES DATA

1. Introduction. In these notes, I will provide my own perspective on mathematical modeling time series data. For the time being, I will consider as a first example (of many to come):

- Retail Sales

The data presented here can be accessed through: https://fred.stlouisfed.org/ and will also be provided in my Github page at https://github.com/Ricard0000/Financial-Models. The file is a CSV file named "*real_sales_per_day*". The purpose of this lecture note is to give the reader an insight into multiscale modeling: Modeling systems which change in time according to different time scales. For example, daily changes might be modeled with $\Delta t_1 \sim \mathcal{O}(hours)$, monthly changes might be modeled by $\Delta t_2 \sim \mathcal{O}(days)$, and yearly changes $\Delta t_3 \sim \mathcal{O}(months)$. In other words, the exact dynamics of financial data (Price or number of units sold or number of consumer etc...) denoted by $S(t)$, is given by

$$\frac{S(t)}{dt} = \mathcal{F}(s) \qquad \Longleftrightarrow \qquad S(t) = S_0 + \int_0^t \mathcal{F}(s)\, ds \tag{1.1}$$

where $\mathcal{F}(s)$ models the time rate of change of $S(t)$. It can more specifically be decomposed as:

$$S(t + \Delta t) = S(t) + \Delta t_1 \mathcal{F}_1(t) + \Delta t_2 \mathcal{F}_2(t) + \Delta t_3 \mathcal{F}_3(t) \cdots . \tag{1.2}$$

where we used the left-end-point approximation (Forward-Euler approximation) to evaluate the integral in equation 1.1 starting at t and ending at $t + \Delta t$. Equation (1.2) is what I claim is an appropriate model for time-series data involving multiple scales (At least in the Forward-Euler approximation).

The purpose:
1) To show multiscale modeling is essential and more accurately models physical systems (for which financial data is most certainly a part of) which change depending on the magnitude of time observations Δt_i.
2) To give more accurate alternatives to conventional methods such as ARIMA ([1]) and other time series models that do not give a precise multiscale model as to how $S(t)$ is changing.
3) To come up with a Differential Equation (DE) or a Stochastic Differential Equation (SDE) modeling $S(t)$ (which is otherwise not provided using ARIMA or other standard fitting algorithms including physics informed neural networks [2]).

2. Modeling Data. A plot of the data we wish to model is shown in figure 1 where I scaled the data by multiplying the actual data by $\dfrac{1}{\max(Data) - \min(Data)}$. There are 312 data points the starting point corresponds to the year 1992. At around t=200 (around 2008) was the year of the financial crash. Every 12 grid points in this data corresponds to 1-year.

Modeling using linear regression. It its a standard procedure to use linear regression as a model for predictions. In terms of an ODE:

$$\frac{dS}{dt} = m \qquad \text{where } m \text{ is a constant.} \tag{2.1}$$

However, this data is clearly non linear due to the 2008 crash, and the clear yearly periodic behavior. Lets deal with modeling of the crash first. We will use linear regression for for the left side of the crash $t \leq 200$ and the right side of the crash ($t > 192$). In terms of an ODE,

$$\frac{dS}{dt} = m - b\delta(t - 192) \tag{2.2}$$

represents a linear solution with an abrupt change of b at $t = 192$. Why? Take the limit of $\dfrac{dS}{dt}$ from the left

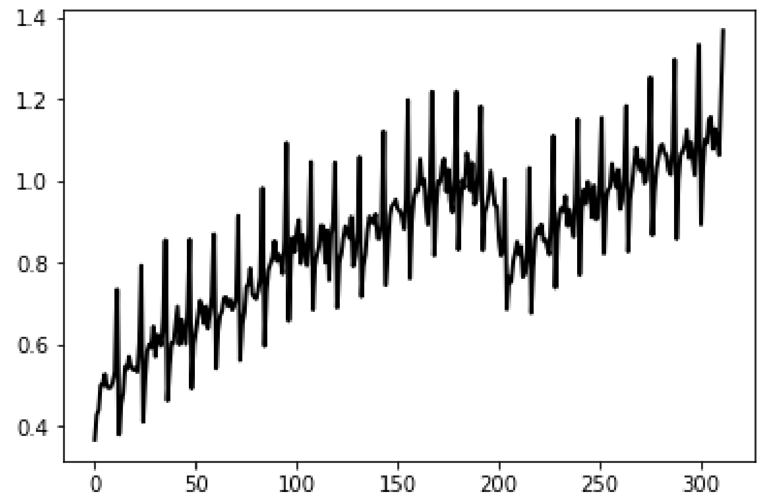

FIG. 1. *Retail Sales data in the United States*

and right:

$$\begin{aligned}
\lim_{t\to 192^+} S(t) - \lim_{t\to 192^-} S(t) &= \lim_{t\to 192^+} \int_0^t m - b\delta(s-192)\,ds - \lim_{t\to 192^-} \int_0^t m - b\delta(s-192)\,ds \\
&= (m-b) - m \qquad \text{since the left integral contains 192 in} \\
&\qquad\qquad\qquad\quad \text{its domain and the right does not,} \\
&= b.
\end{aligned} \tag{2.3}$$

Clearly this abrupt change is not physically realistic. There should be a smooth transition from the left of $t = 192$ to the right of $t = 192$. It is known that

$$\lim_{\sigma\to 0} \frac{1}{\sqrt{2\pi\sigma}} \exp\left(\frac{-x^2}{2\sigma}\right) \to \delta(t) \tag{2.4}$$

Thus a smooth transition from the left to the right region is possible using a Gaussian term as part of dS/dt where small values of σ represents a very sharp change/crash. For this reason, we can model the data (with the exception of the periodic trends) using the ODE:

$$\frac{dS}{dt} = m - \frac{b}{\sqrt{2\pi\sigma}} \exp\left(\frac{-(t-192)^2}{2\sigma}\right) \tag{2.5}$$

where now m, b, and σ are learned parameters from data. This model of course assumes that the slope of the regression line to the left and right of the crash are the same. Since this may not be the case we introduce a piece-wise model for this data:

$$\begin{aligned}
\frac{dS}{dt} &= m_1 - \frac{b}{\sqrt{2\pi\sigma}} \exp\left(\frac{-(x-192)^2}{2\sigma}\right), \qquad t < 192 \\
\frac{dS}{dt} &= m_2 - \frac{b}{\sqrt{2\pi\sigma}} \exp\left(\frac{-(x-192)^2}{2\sigma}\right), \qquad t > 192
\end{aligned} \tag{2.6}$$

This can now be implemented on a computer (see the attached code on Github). The result is shown in figure 2 where the red curve is the smooth piece-wise linear model.

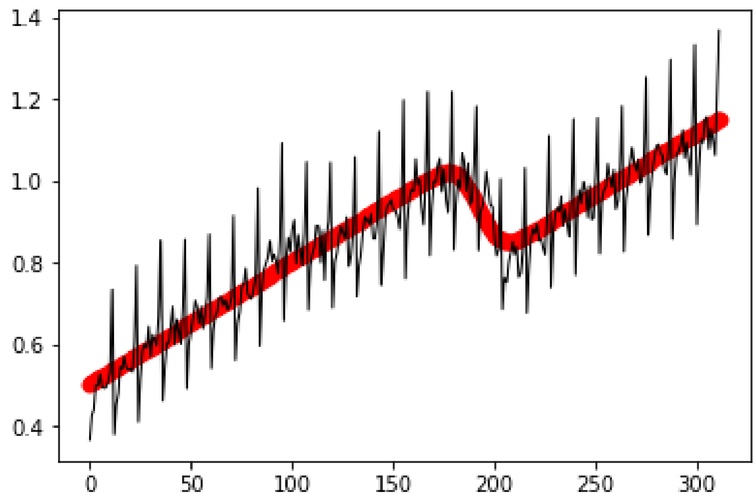

Fig. 2. *Retail Sales data, Red: smooth piece-wise linear model.*

Now one needs to look at seasonal trends, this will be explained later and done with the help of a simple Neural network. We also need to analyze the residuals of our model. For the time being, I will leave the reader in suspense and a plot in 3 with the finished model shown in red. This will be explained at a later date.

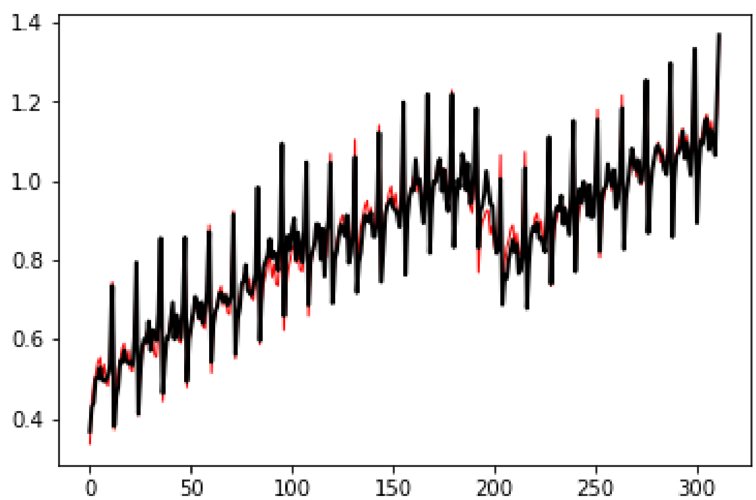

Fig. 3. *Retail Sales data, Red: smooth piece-wise linear model with yearly trends.*

REFERENCES

[1] D. Asteriou and S. Hall, *ARIMA Models and the Box-Jenkins Methodology*, 01 2016, pp. 275–296, https://doi.org/10.1057/978-1-137-41547-9_13.

[2] M. Raissi, P. Perdikaris, and G. E. Karniadakis, *Physics informed deep learning (part i): Data-driven solutions of nonlinear partial differential equations*, 2017, https://arxiv.org/abs/1711.10561.