Prediction Model

Machine Learning

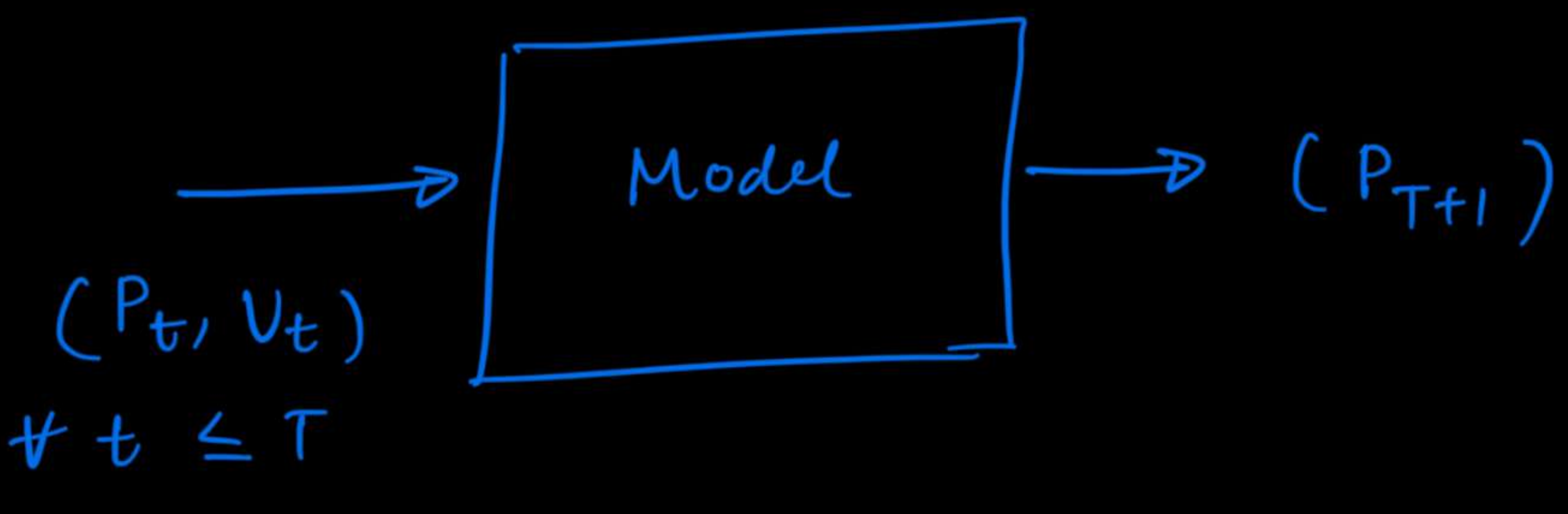
Overview :-

We consider the cryptocurrencies :-

Ethereum, Bitcoin, Cardano, XRP, Solana, Binance coin, Dogecoin, USDC, Tron, Avalanche, Litecoin.

We will build a model which will forecast the prices of these coins, one month from now. For this purpose, we only take the monthly data, since daily data contributes a lot of noise

We only take the closing price and the volume traded. So initially we have a Bivariate time series (P_t, V_t).



Data :-

Let's Analyse the data of one of the coins and see what we can infer.

This is the price chart of Etherium coin. This is a time series data.

We see that the process looks like a Brownian Motion but we don't know yet the parameters. To get the initial estimates of where the price can move in future, we consider the returns calculated as :-

$$r_i = \frac{P_i - P_{i-1}}{P_{i-1}}$$

We consider the model :-

$$① - \boxed{dr(t) = K(\theta - r(t))\,dt + \sigma\,dW(t)}$$ with initial condition $r(0) = r_0$ and $W(t)$ being the standard Brownian motion.

To solve ①, we try $f(r(t), t) = r(t)e^{Kt}$

Using Ito lemma,

$$df(r(t), t) = K r(t) e^{Kt}\,dt + e^{Kt}\,dr(t)$$

$$= K r(t) e^{Kt}\,dt + e^{Kt}\left[K(\theta - r(t)) + \sigma\,dW(t)\right]$$

$$= K\theta e^{Kt}\,dt + \sigma e^{Kt}\,dW(t)$$

$$\therefore f(r(t), t) - f(r(0), 0) = r(t)e^{Kt} - r_0$$

$$= \int_0^t df(r(s), s)$$

$$= \int_0^t K\theta e^{Ks}\,ds + \int_0^t \sigma e^{Ks}\,dW(s)$$

$$\therefore r(t) = r_0 e^{-Kt} + e^{-Kt}\int_0^t K\theta e^{Ks}\,ds + e^{-Kt}\int_0^t \sigma e^{Ks}\,dW(s)$$

Now we calculate mean and variance,

$$E[r(t)] = r_0 e^{-Kt} + \theta e^{-Kt}[e^{Kt} - 1]$$

$$=: r_0 e^{-Kt} + \theta K \Lambda(t) \quad \text{where}$$

$$\Lambda(t) = \int_0^t e^{-Ks}\, ds = \frac{1}{K}[1 - e^{-Kt}]$$

$$\therefore \lim_{t \to \infty} E[r(t)] = \theta$$

$$\text{Ily}, \quad var[r(t)] = \sigma^2 e^{-2Kt} \int_0^t e^{2Ks}\, ds = \frac{\sigma^2}{2K}[1 - e^{-2Kt}]$$

$$= \frac{1}{2}\sigma^2[\Lambda(2t)]$$

Now consider a time interval $[t-h, t]$,

$$r(t) = \theta(1 - e^{-Kh}) + e^{-Kh} r(t-h) + e^{-Kt} I$$

$$I = \int_{t-h}^{t} \sigma e^{Ks}\, dW(s)$$

The discretised process follows AR(1) model with intercept $\theta(1 - e^{-Kh})$, auto-regressive co-efficient e^{-Kh} and innovation process $e^{-Kt} \int_{t-h}^{t} \sigma e^{Ks}\, dW(s)$

Thus,

$$\text{var}\left[e^{-kt} \int_{t-h}^{t} \sigma e^{ks}\, dW(s) \right]$$

$$= \sigma^2 h \alpha(-2kh)$$

$$\alpha(x) = \frac{e^x - 1}{x}, \quad \alpha(0) = 0$$

$\therefore$ we can simulate with $r(0) = r_0$ and moving forward acc to the recursion,

$$\boxed{r(t) = \theta k \Lambda(h) + e^{-kh} r(t-h) + u_t^{(h)}},$$

$$u_t^{(h)} \sim N(0, \sigma^2 h \alpha(-2kh))$$

From this, we get the sample paths which denote the 95% confidence intervals of the returns.

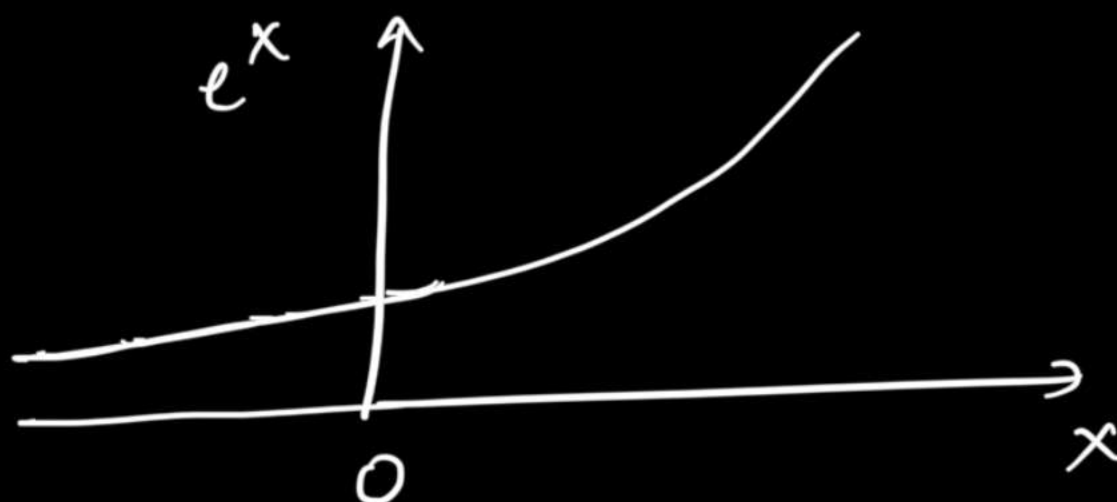
Now, we calculate the mean of these paths at every time instant & thus get a new avg/ expected time series forecast. i.e. let paths be $P_1, P_2, \ldots P_n$ then

$$E[P(t)] = \frac{P_1(t) + P_2(t) + \cdots + P_n(t)}{n}$$

Now to get further insights, we consider the Long-short term memory model.

We create some new features :-

(1) Exponential Moving Average :-



we consider,

$$F_1(t) = r_t + e^{-1} r_{t-1} + e^{-2} r_{t-2} + \dots + e^{-10} r_{t-10}$$

(2) Seasonal component of the returns time series.