Wind project initial constants

Khaoula Ben Chaabane

Ecole Polytechnique de Tunisie

25/04/2020

Introduction

In these slides, I have determined the initial parameters needed for the model of the wind power production. These parameters are ε , θ_0 and α.

Model

The wind power production is modeled as follows, where X_t is the normalized real production :

$$\begin{cases} dX_t = \left(\dot{p}_t - \theta_t \left(X_t - p_t\right)\right) dt + \sqrt{2\alpha\theta_0 X_t \left(1 - X_t\right)} dW_t, \quad t \in [0, T] \\ X_0 = x_0 \in [0, 1] \end{cases}$$

We may introduce the following model for the forcecast error of the normalized wind power production where X_t is the real production, p_t the forecast and $V_t = X_t - p_t$ is the error :

$$\begin{cases} dV_t = -\theta_t V_t dt + \sqrt{2\alpha\theta_0 \left(V_t + p_t\right)\left(1 - V_t - p_t\right)} dW_t, \quad t \in [0, T] \\ V_0 = v_0 \in [-1 + \varepsilon, 1 - \varepsilon] \end{cases}$$

Model

To guarantee a unique solution for the process X_t, θ_t needs to be bounded for $t \in [0, T]$. We have that :

$$\theta_t = \max\left(\theta_0, \frac{\alpha\theta_0 + |2\dot{p}_t|}{2\min(1-p_t, p_t)}\right)$$

This is not true for θ_t if $p_t \to 0^+$ or $p_t \to 1^-$. Therefore we need to ensure that $p_t \in [\varepsilon, 1-\varepsilon]$ for some $0 < \varepsilon < \frac{1}{2}$, $\forall\ t \in [0, T]$.

Model

We define then the corrected forecast :

$$p_t^{\varepsilon} = \begin{cases} \varepsilon & \text{if} \quad p_t < \varepsilon \\ p_t & \text{if} \quad \varepsilon \leq p_t < 1 - \varepsilon \\ 1 - \varepsilon & \text{if} \quad p_t \geq 1 - \varepsilon \end{cases}$$

and the corrected (and bounded) drift coefficient is therefore :

$$\theta_t^{\varepsilon} = \max\left(\theta_0, \frac{\alpha\theta_0 + 2\,|\dot{p}_t^{\varepsilon}|}{2\min\left(1 - p_t^{\varepsilon}, p_t^{\varepsilon}\right)}\right)$$

Least Square Minimization : LSM

In order to evaluate the constants of our model we apply the least square method on the forecast error V_t.
We consider the transition $\Delta V_i = V_{i+1} - V_i$ with $\Delta t = t_{i+1} - t_i$.
$(V_{i+1}|V_i)$ is a random variable which conditional mean can be approximated by the solution of the following system :

$$\begin{cases} \mathrm{d}\mathbb{E}[V] = -\theta_t^{\varepsilon}\mathbb{E}[V]\mathrm{d}t \\ \mathbb{E}\left[V\left(t_i\right)\right] = V_i \end{cases}$$

evaluated in t_{i+1} (i.e., $\mathbb{E}\left[V\left(t_{i+1}\right)\right]$).

Then, the random variable $\left(V_{i+1} - \mathbb{E}\left[V\left(t_{i+1}\right)\right]\right)$ has a mean equal to 0 approximately.
If we assume that $\theta_t^{\varepsilon} = c \in \mathbb{R}^{+}$ for all $t \in \left[t_i, t_{i+1}\right]$, then $\mathbb{E}\left[V\left(t_{i+1}\right)\right] = V_i e^{-c\Delta t}$.
If we have a total of n transitions, we can write the regression problem for the conditional mean with L^2 loss function as :

$$
\begin{aligned}
c^* &\approx \arg\min_{c \geq 0} \left[\sum_{i=1}^{n} \left(V_{i+1} - \mathbb{E}\left[V\left(t_{i+1}\right)\right]\right)^2\right] \\
&= \arg\min_{c \geq 0} \left[\sum_{i=1}^{n} \left(V_{i+1} - V_i e^{-c\Delta t}\right)^2\right] \qquad (1)
\end{aligned}
$$

We take the first order approximation of $e^{-c\Delta t}$ w.r.t. c :

$$e^{-c\Delta t} = 1 - c\Delta t + O\left((c\Delta t)^2\right)$$

and introduce it in equation (1). We get

$$c^* \approx \arg\min_{c\geq 0} \underbrace{\left[\sum_{i=1}^{n} \left(V_{i+1} - V_i(1 - c\Delta t)\right)^2\right]}_{=f(c)}$$

As $f(c)$ is convex in c, solving (5) (finding c^*) is equivalent to solving

$$\frac{\partial f}{\partial c}\left(c^{**}\right) = 0$$

and choosing $c^* = \max\left\{0, c^{**}\right\}$

Least Square Minimization : LSM

$$\begin{aligned}
\frac{\partial f}{\partial c} &= \sum_{i=1}^{n} 2\left(-V_i\right)\left(-\Delta t\right)\left(V_{i+1} - V_i\left(1 - \theta_0 \Delta t\right)\right) \\
&= \sum_{i=1}^{n} 2 V_i \Delta t\left(V_{i+1} - V_i(1 - c\Delta t)\right) \\
&= \sum_{i=1}^{n} 2 V_{i+1} V_i \Delta t - 2 V_i^2 \Delta t + 2 V_i^2 \Delta t^2 c
\end{aligned}$$

Then, c^{**} satisfies the following :

$$c^{**} \approx \frac{\sum_{i=1}^{n} V_i\left(V_i - V_{i+1}\right)}{\Delta t \cdot \sum_{i=1}^{n} \left(V_i\right)^2}$$

Quadratic variation

We approximate the SDE by its E-M scheme. In particular, we approximate the Itô quadratic variation with the discrete one :

- Itô process quadratic variation : $[V]_t = \int_0^t \sigma_s^2 \mathrm{d}s$
- Discrete process quadratic variation : $[V]_t = \Sigma_{0<s\leq t} (\Delta V_s)^2$

Then, considering Δt the time between the measurements, we approximate :

$$\theta_0^* \alpha^* \approx \frac{\sum_{i=1}^n (\Delta V_i)^2}{2\Delta t \sum_{i=1}^n (V_i + p_i)(1 - V_i - p_i)}$$

Estimation of $(\theta_0, \alpha, \varepsilon)$

In this section, we will use the approximation made previously to estimate the parameters $(\theta_0, \alpha, \varepsilon)$ of the SDE. Let us define $(\theta_0^*, \alpha^*, \varepsilon^*)$ as their estimators.

If we fix ε, we define the forecast error $\forall i \in 1...n$ $V_i = X_i - p_i^\varepsilon$.
If we also fix θ_0 and α, we can define the set of indexes :

$$\mathrm{I} = \{i \in \{1, \ldots, n\} : \text{ the LSM estimation will estimate } \theta_0\}$$
$$\mathrm{J} = \left\{j \in \{1, \ldots, n\} : \text{ the } LSM \text{ estimation will estimate } \frac{\theta_0 \alpha}{\varepsilon}\right\}$$

We will proceed then to approximate these sets in order to estimate our parameters.

Estimation of $(\theta_0, \alpha, \varepsilon)$

To use the LSM estimation, we assumed that $\theta_t^\varepsilon = c \in \mathbb{R}^+$, and we defined θ_t^ε :

$$\theta_t^\varepsilon = \max\left(\theta_0, \frac{\alpha\theta_0 + 2\left|\dot{p}_t^\varepsilon\right|}{2\min\left(1 - p_t^\varepsilon, p_t^\varepsilon\right)}\right)$$

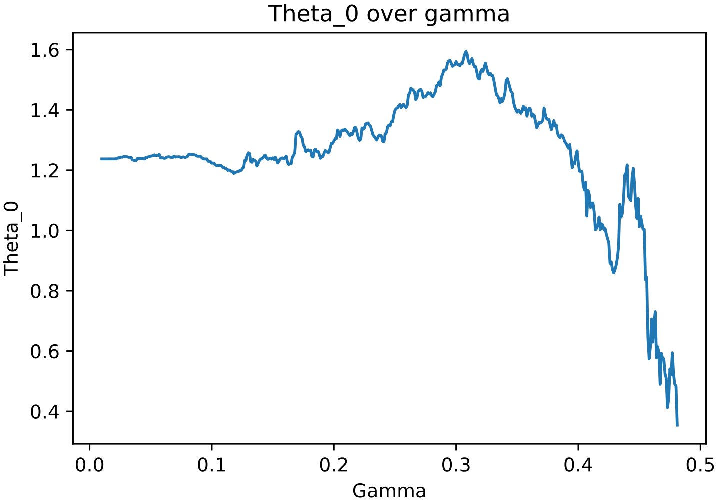
From the definition of θ_t^ε : We have that for $\varepsilon << 1$, and $p_t = \varepsilon$ or $p_t = 1 - \varepsilon$, the approximation $\theta_t^\varepsilon \approx \frac{\theta_0 \alpha}{\varepsilon}$ holds. Then, for ε small enough, J can be approximated by the following :

$$\mathrm{J} \approx \mathrm{J} = \left\{j \in \{1, \dots, n\} : p_j^\varepsilon \in \{\varepsilon, 1 - \varepsilon\}\right\}$$

and θ_t^ε, we have that it is more likely that $\theta_t^\varepsilon = \theta_0$ if $p_t^\varepsilon \approx \frac{1}{2}$. Then, we can approximate I by

$$\mathrm{I} \approx \tilde{\mathrm{I}} = \{i \in \{1, \dots, n\} : p_i \in (\gamma, 1 - \gamma)\}, \gamma \approx \frac{1}{2}, \gamma < \frac{1}{2}$$

Estimation of θ_0^*



We estimate $\theta_0 * = 1.25$

Estimation of α^*

With the previous approximation made of the quadratic variation we can estimate $\theta_0 * \alpha* = 0.094$ therefore, with our given estimation of θ_0* we find that : $\alpha* = 0.08$

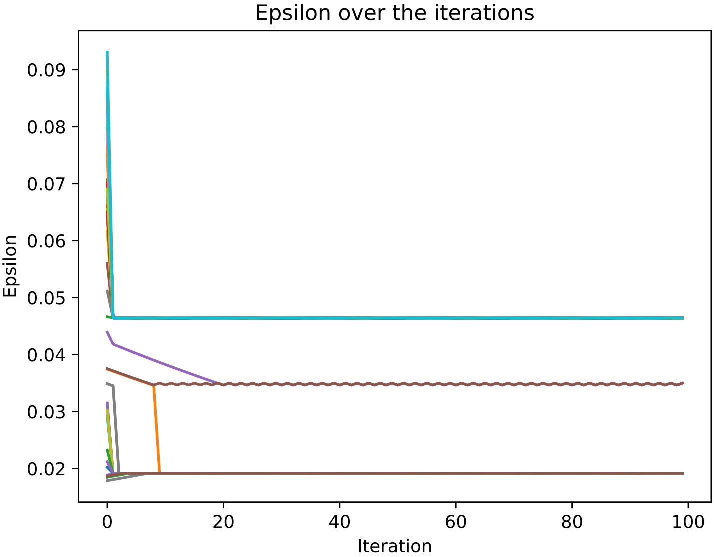
Estimation of ε^*

Now that we have an approximated value of $\theta_0\alpha$, if we can estimate $\frac{\theta_0\alpha}{\varepsilon}$, then we can estimate ε. We showed previously that for $\varepsilon << 1$, the LSM estimation using indexes from J is an estimator for $\frac{\theta_0\alpha}{\varepsilon} =: k$
The goal is to find values for ε that satisfy $\varepsilon << 1$. For that we start by randomly choosing a small initial value for ε (that we will call ε_0), and iterating we aim to converge to some local minimum. We proceed with the following steps :

- We sample ε_0 from $U[0.01, 0.1]$ and load $\varepsilon \leftarrow \varepsilon_0$
- We create $\tilde{J}$ and use the LSM estimation to find k.
  - If $k < \theta_0^*$, then the assumption $\theta_t^\varepsilon = c \in \mathbb{R}^+$ is wrong and we reduce the value of ε, i.e., $\varepsilon \leftarrow \varepsilon * 0.999$.
  - If $k \geq \theta_0^*$, we load $\varepsilon \leftarrow \frac{\theta_0^*\alpha^*}{k}$(we allow a maximum relative change of 1%).

  We repeat this step 100 times.
- We repeat steps 1 and 2, 50 times.

Estimation of ε^*



We can see from the plots that $\varepsilon * = 0.018$ is a good approximation.

Conclusion

To conclude, the estimations of the SDE parameters that we found are : $(\theta_0^*, \alpha^*, \varepsilon^*) = (1.25, 0.08, 0.018)$.
The code computing this process can be found in the file 'Wind$_p$*roject*$_i$*ntial*$_g$*uess.ipynb*'.