Prediction of Power Generation

Waleed Alhaddad

April 22, 2019

Reader's Guide

List of changes in this iteration:

- Matched data with every forecast
- Normalized according to installed power of each period
- Interpolated the forecast using cubic splines.

Next steps:

- upgrade the numerical integrator and use the interpolated forecast
- Check performace using the interpolated forecasts and data

Note :

- Green slides: possible future extensions
- Red slides: notes to be removed

Toy model with Synthetic Data

We generate paths from an SDE with given parameters using Monte Carlo simulation and then fit the model to check the retrievability of the parameters. We consider the following cases:

1: Constant parameter Wiener process: finite and infinite horizon.

2: Controlled Wiener process within an interval.

3: Controlled Wiener process centered around a given function.

Stochastic Differential Equation

Given the SDE

$$\begin{aligned} dX_t &= a(X_t;\boldsymbol{\theta})dt + b(X_t;\boldsymbol{\theta})dW_t \quad t > 0 \\ X_0 &= x_0 \end{aligned} \tag{1}$$

Consider a set of N observations, $X = \{X_{t_0^N}, X_{t_1^N}, \ldots, X_{t_N^N}\}$ observed within intervals Δ_N.

- Δ_N: is any function of the number of samples, N. For example, $\Delta_N = \frac{1}{N}$.
- $a(X_t;\boldsymbol{\theta})$ is the drift function (add technical conditions here)
- $b(X_t;\boldsymbol{\theta})$ is the diffusion function (add technical conditions here)
- $\boldsymbol{\theta}$ is a vector of constant parameters

likelihood function

We have the Likelihood function of the transition density

$$\mathscr{L}(\theta; X) = \prod_{i=1}^{n} \rho(X_{i+1}|X_i, \boldsymbol{\theta})\rho(X_0) \tag{2}$$

Consider a Gaussian transition density,

$$\mathscr{L}(\boldsymbol{\theta}; X) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi b(X_i, \boldsymbol{\theta})^2 dt}} \exp\Big(\frac{-((X_{i+1} - X_i) - a(X_i, \boldsymbol{\theta})dt)^2}{2b(X_i, \boldsymbol{\theta})^2 dt}\Big)\rho(X_0)$$

We take the improper prior $\rho(X_0) = 1$.

Case 1: Constant Parameters

For a fixed $b(X_t; \boldsymbol{\theta}) = \sigma$ and a fixed $a(X_t; \boldsymbol{\theta}) = \mu$. We observe that varying the sampling and discretization parameters affects the convergence behavior.
We consider the following three situations:

1.1 Infinite Horizon expansion caused by increasing samples N
(increasing N and Δ_N is fixed).

1.2 Infinite Horizon expansion caused by increasing intervals Δ_N
(fixed N and increasing decreasing Δ_N).

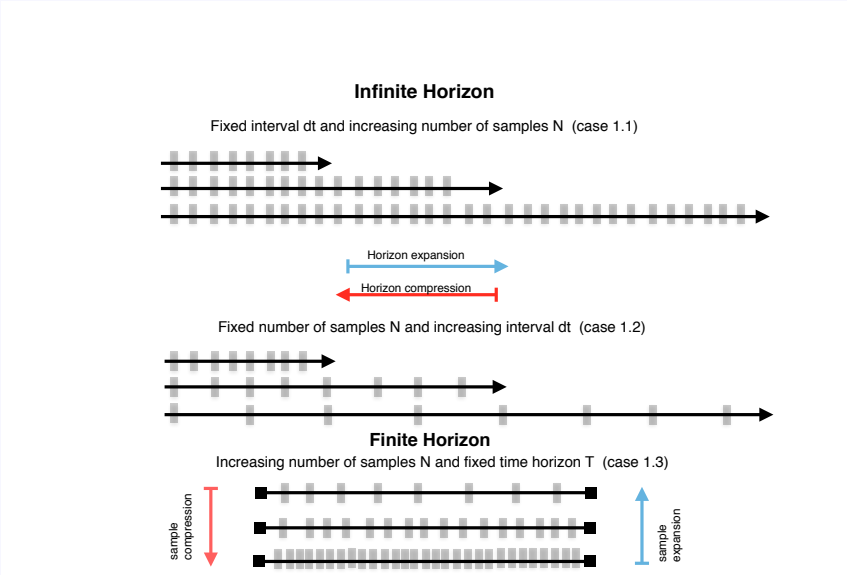
1.3 Finite Horizon expansion with sample compression, i.e where $\Delta_N = \frac{T}{N}$
(increasing N, fixed T).

Case 1: Implementation

In the following, we use L-BFGS-B algorithm to optimize the likelihood with bounds on σ being positive. And we use $\mu = 100$ and $\sigma = 10$ for the path generation. We denote the estimators of μ and σ by $\hat{\mu}$ and $\hat{\sigma}$ where,

$$\hat{\mu} = \arg\max_{\mu \in \mathbb{R}} \mathscr{L}(\boldsymbol{\theta}; X) \qquad \hat{\sigma} = \arg\max_{\sigma \in \mathbb{R}^+} \mathscr{L}(\boldsymbol{\theta}; X) \tag{3}$$

The above cases can be seen clearly in the illustration below.



Case 1.1: Constant Parameters

Infinite Horizon expansion caused by increasing samples N

An increase of the number of samples N in a growing time interval T. In this case, we obtain convergence of the order $\frac{1}{\sqrt{N}}$ as expected.

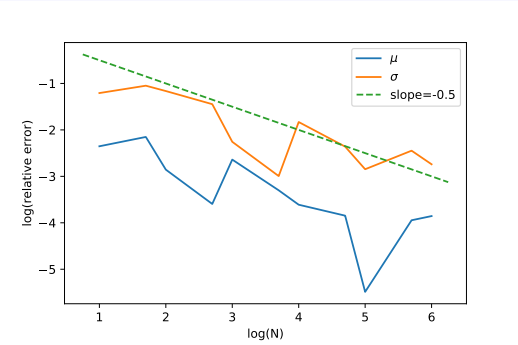


Figure 1: convergence in the case of infinite horizon expansion caused by increasing samples N

Case 1.2: Constant Parameters

Infinite Horizon expansion caused by increasing intervals dt

In this case, the estimator $\hat{\sigma}$ of the diffusion coefficient is inconsistent. This is because the time interval between the samples is increasing and so we do not have as much information about the variation of the process. However, the drift estimator $\hat{\mu}$ is still consistent. We have not seen any theoretical proof of this situation.

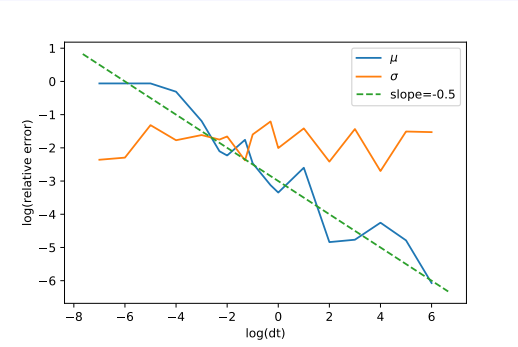


Figure 2: convergence rate in the case of infinite horizon expansion caused by increasing intervals Δ_N

Case 1.3: Constant Parameters

Finite Horizon expansion with sample compression

In this case, we observe increasing number of samples in a fixed time interval (sample compression).In this case the drift estimator μ is inconsistent while the diffusion estimator σ is consistent. This situation matches the theoretical discussion in [1].

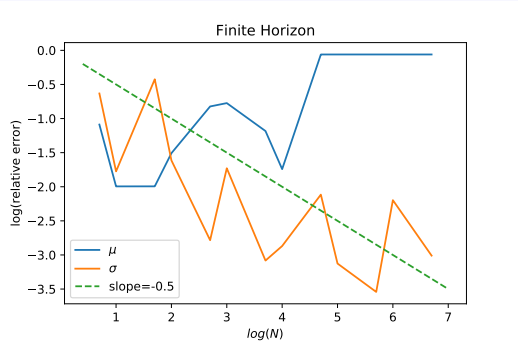


Figure 3: Convergence rate in the case of finite horizon expansion with sample compression

Remarks

1. It is easier to obtain convergence in the diffusion coefficient. Because the estimator of the diffusion has access to information from the infinite variation of the wiener process while the drift does not. We can show this theoretically from the efficiency condition of the estimators.
2. The convergence generally follows Monte Carlo's convergence rate in terms of sampling as expected. This statement is wrong in a high frequency setting, as we have different convergence rate for estimator of the diffusion and than for the estimator of the drift as we will see next. The estimator of the drift converges slower than the Monte Carlo rate given that both are consistent.

Motivation of high frequency SDE data

High frequency situations are not uncommon and extreme but may happen for data that looks reasonable. For instance, in some cases weekly data is treated as high frequency data . A situation is considered high frequency with respect to the amount of variation in the drift term [1] (or in regards to the speed at which it reverts to the mean).
This also important for building confidence intervals as the limiting distribution of our estimator may not be Gaussian. For example, in the setting of high frequency data and fixed time T (case 1.3 above) the limiting distribution of the estimator is path dependent and is given by a normal variance-mixture distribution which can be heavy tailed with no finite moments [1]. However, in this case, a modification of the estimator can help in achieving a Gaussian limiting distribution of the estimator.
Please note that in the setting of high frequency SDE data, we have two possibilities:

- Finite Horizon (fixed time T): this corresponds to case 1.3 above.
- Infinite Horizon: this corresponds to other two cases.

Convergence of the drift estimator for high frequency SDE (Case 1.3 / Finite Horizon)

In a high frequency setting with finite horizon (case 1.3), there are **no** consistent estimators of the drift term $a(X_t, \boldsymbol{\theta})$ (see [1]) as we break the following condition :
Given the following SDE,

$$\begin{aligned} dX_t &= a(X_t;\alpha)dt + b(X_t;\beta)dW_t \quad t > 0 \\ X(0) &= x_0 \end{aligned} \tag{4}$$

the drift estimator $\hat{\alpha}$ of the parameter set α of the system is consistent if and only if

$$n\Delta_n \to \infty \tag{5}$$

And this condition contradicts having a finite horizon as we trivially have $T = N\Delta_N \to \infty$. If the condition (5) is not satisfied, we cannot have a consistent estimator $\hat{\alpha}$ of the drift. However, the estimator $\hat{\beta}$ of the diffusion term $b(X_t;\beta)$ is still consistent regardless.

Convergence of the drift estimator for high frequency SDE (Case 1.3 / Finite Horizon)

To show this result numerically, we pick any Δ_n such that $T = n\Delta_n \to \infty$. In this example, we modify the non-consistent drift parameter estimator (case 1.3) such that

$$dt = \frac{T}{N} \quad \to \quad dt = \frac{T}{\sqrt{N}} \tag{6}$$

Then we recover back the convergence of the other cases as this replacement turns the finite horizon to an infinite horizon.

Numerical Convergence for high frequency SDE

In this numerical example, we take $\Delta_N = \frac{T}{\sqrt{N}}$ as an example.

$$\text{Finite Horizon: } dt = \frac{T}{N} \quad \text{versus} \quad \text{Infinite Horizon: } dt = \frac{T}{\sqrt{N}} \tag{7}$$

We confirm condition (5) numerically. It is clear that in the finite horizon case the estimator of the diffusion $\hat{\sigma}$ converges while the estimator of the drift $\hat{\mu}$ is not consistent. Whereas, in the infinite horizon case, we obtain the expected convergence in both parameters.

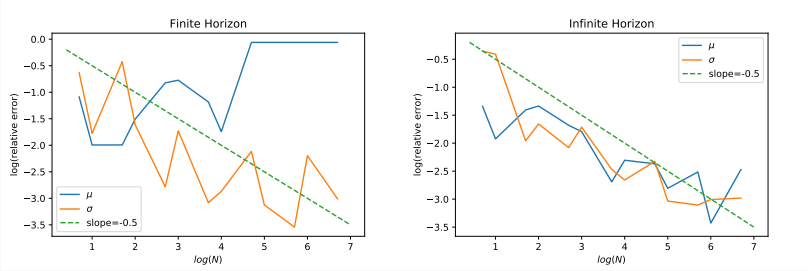


Figure 4: Right side: Finite horizon with $dt = \frac{T}{N}$, Left side: Infinite horizon with $dt = \frac{T}{\sqrt{N}}$

Convergence for high frequency SDE data

We can show that the score function of the likelihood is an efficient martingale estimating function and that the obtained estimator is optimal[1]. Then we have the following convergence rates [3]:
For a finite Horizon (fixed T)

$$\hat{\alpha} \text{ is not consistent} \quad \text{and} \quad \textit{relative error}(\hat{\beta}) \propto \frac{1}{\sqrt{N}} \tag{8}$$

For an infinite Horizon ($T \to \infty$)

$$\textit{relative error}(\hat{\alpha}) \propto \frac{1}{\sqrt{N\Delta_N}} \quad \text{and} \quad \textit{relative error}(\hat{\beta}) \propto \frac{1}{\sqrt{N}} \tag{9}$$

Note that the estimator $\hat{\alpha}$ of the drift is affected by the time discretization while the estimator $\hat{\beta}$ of the diffusion is independent of the time discretization.

[1] If we consider an approximate martingale estimator function such as the pseudo likelihood, then we require an extra condition that $n\Delta_n^{2\kappa-1} \to 0$ where κ is the order of Δ_N in RHS of the expectation (11)

High Frequency SDE General Convergence

Consider a set of N observations, $X = \{X_{t_0^N}, X_{t_1^N}, \ldots, X_{t_N^N}\}$ and the following martingale estimator function,

$$G_n(\gamma) = \sum_{i=1}^{N} g(\Delta_N, X_{t_i^N}, X_{t_{i-1}^N}; \gamma) \tag{10}$$

where $t_i^n = i\Delta_n$, $\gamma = (\alpha, \beta)$ and $g = (g_1, g_2)$. In our martingale case[2],we must have,

$$E_\gamma\Big(g(\Delta_N, X_{t_i^N}, X_{t_{i-1}^N}; \gamma)\Big| X_{t_{i-1}^N}\Big) = 0 \tag{11}$$

where the $\hat{\beta}$ estimator is given by solving $G_n(\gamma) = 0$.

[2]The results can also be adapted to a non martingale estimating function such as a pseudo likelihood

High Frequency SDE - Possible Extensions

- Try to understand the limit of when we need to switch to the high frequency from the low frequency regime. That is how it depends on the speed of the mean reversion in our model.
- Get the martingale estimator function of the Gaussian transition density explicitly.
- State the optimality and efficiency of the estimators explicitly.
- Obtain the stated theoretical convergence rate.
- Try to show theoretically why in case 1.2 , the diffusion estimator is not consistent.
- Understand and state the conditions under which we have a unique parameter estimation to our SDE.

References

[1] Jakobsen, N. M., & Sørensen, M. (2017). Efficient estimation for diffusions sampled at high frequency over a fixed time interval. Bernoulli, 23(3), 1874-1910. doi:10.3150/15-bej799

[2] Sørensen, M. (2007). Efficient Estimation for Ergodic Diffusions Sampled at High Frequency. SSRN Electronic Journal. doi:10.2139/ssrn.1150694

[3] Kessler, M., Lindner, A., & Sørensen, M. (2012). Statistical methods for stochastic differential equations. Boca Raton, FL: Chapman & Hall/CRC.

Paths as IID samples

We have the Likelihood function of the transition density

$$\mathscr{L}(\theta; X) = \prod_{j=1}^{M} \prod_{i=1}^{N} \rho(X_{j,i+1} | X_{j,i}, \boldsymbol{\theta}) \rho(X_0) \tag{12}$$

Consider a Gaussian transition density,

$$\mathscr{L}(\boldsymbol{\theta}; X) = \prod_{j=1}^{M} \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi b(X_{j,i}; \boldsymbol{\theta})^2 dt}} \exp\Big(\frac{-((X_{j,i+1} - X_{j,i}) - a(X_{j,i}; \boldsymbol{\theta}) dt)^2}{2 b(X_{j,i}; \boldsymbol{\theta})^2 dt} \Big) \rho(X_0)$$

We take the improper prior $\rho(X_0) = 1$.

Paths as IID samples - Implementation

As before, we use L-BFGS-B algorithm to optimize the likelihood with bounds on σ being positive. We denote the estimators of μ and σ by $\hat{\mu}$ and $\hat{\sigma}$ where,

$$\hat{\mu} = \arg\max_{\mu \in \mathbb{R}} \mathscr{L}(\boldsymbol{\theta}; X) \qquad \hat{\sigma} = \arg\max_{\sigma \in \mathbb{R}^+} \mathscr{L}(\boldsymbol{\theta}; X) \tag{13}$$

Paths as IID samples - Convergence

We have Monte Carlo convergence in terms of samples as expected.

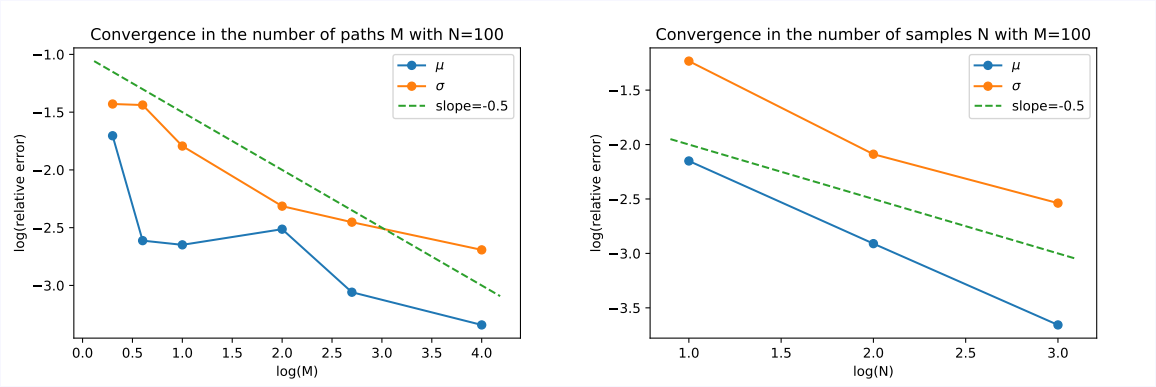


Figure 5: Optimized using L-BFGS-B algorithm. True parameters used are $\mu = 100$ and $\sigma = 10$.M is the number of paths and N is the number of samples.

Case 2: Controlled Wiener process in a unit box

We have the SDE,

$$\begin{aligned} dX_t &= a(X_t;\boldsymbol{\theta})dt + b(X_t;\boldsymbol{\theta})dW_t \quad t > 0 \\ X_0 &= 0 \end{aligned} \tag{14}$$

Consider a set of N observations, $X = \{X_{t_0^N}, X_{t_1^N}, \ldots, X_{t_N^N}\}$ on intervals of Δ_N.
We choose,

$$a_{i+1}(X_i;\boldsymbol{\theta}) = \mu X_i(\frac{1}{2} - X_i) \qquad b_{i+1}(X_i;\boldsymbol{\theta}) = \sigma X_i(1 - X_i) \tag{15}$$

We choose $0 < \sigma << 1$ and $\mu \in (0,1]$. Then, we have contained the Wiener process in the interval $[0,1]$ with high probability.

Case 2: Controlled Wiener process in a unit box

We see that the process is contained within the unit box as we kill the diffusion coefficient if we touch the boundary. Additionally, the drift coefficient is always directed towards the center line $\frac{1}{2}$ of the unit box.

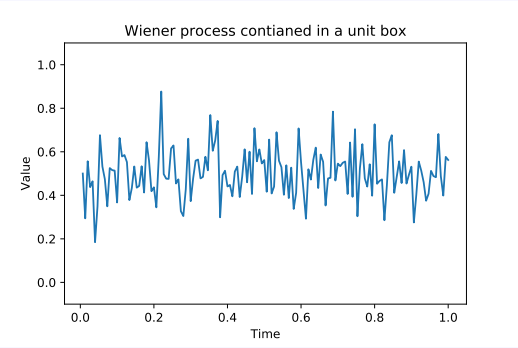


Figure 6: With high probability the wiener process stays inside the box as $\mu \in (0,1]$ and $\sigma << 1$

.

Case 2: Controlled Wiener process in a unit box - Parameter Inference

We obtain the same convergence as before,

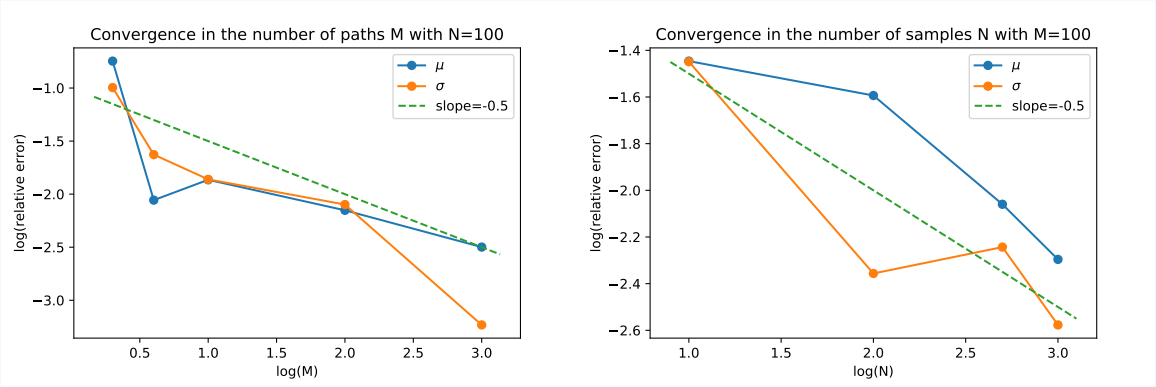


Figure 7: Optimized using L-BFGS-B algorithm. True parameters used are $\mu = 0.5$ and $\sigma = 0.1$. M is the number of paths and N is the number of samples.

Case 3: Controlled Wiener process around a given function

We have the SDE,

$$\begin{aligned} dX_t &= a(X_t;\boldsymbol{\theta})dt + b(X_t;\boldsymbol{\theta})dW_t \quad t>0 \\ X_0 &= 0 \end{aligned} \tag{16}$$

Consider a set of N observations, $X = \{X_{t_0^N}, X_{t_1^N}, \ldots, X_{t_N^N}\}$ on intervals of Δ_N.
We choose,

$$a_{i+1}(X_i;\boldsymbol{\theta}) = \mu X_i(\frac{1}{2}(1+p_i) - X_i) \qquad b_{i+1}(X_i;\boldsymbol{\theta}) = \sigma X_i(1-X_i) \tag{17}$$

We choose $0 < \sigma << 1$ and $\mu \in (0,1]$. Then, we have contained the Wiener process in the interval $[0,1]$ with high probability. Additionally, we add mean reversion to $\frac{1}{2}(1+p_i)$ where $p_i = \sin(\frac{i}{N}2\pi)$.

Case 3: Controlled Wiener process following a given function

We see that the process follows the sine wave closely when the drift parameter $\mu \approx 1$ and that we have a shift artifact when the drift parameter $\mu << 1$. Note that the process does not escape the unit box as the diffusion is controlled and the drift is directed towards the function which is always inside the unit box.

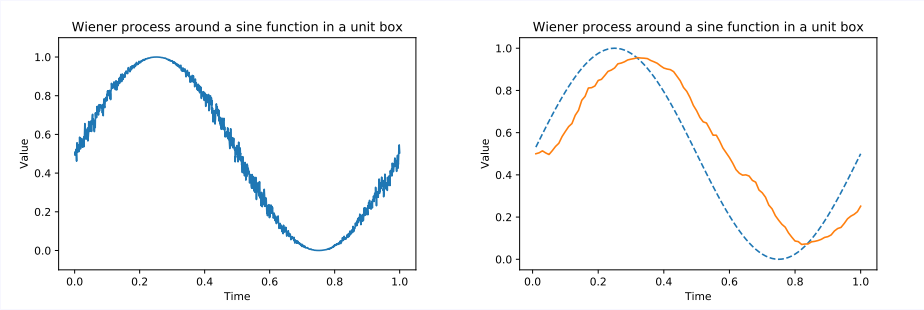


Figure 8: Illustrative examples: On the right, we see the controlled wiener process responding to the boundary by killing the diffusion at the peaks of the sine wave. On the right we observe an unintended shift artifact when the drift parameter $\mu << 1$.

Case 3: Controlled Wiener process around a function - Paramter Inference

We obtain the same convergence as before,

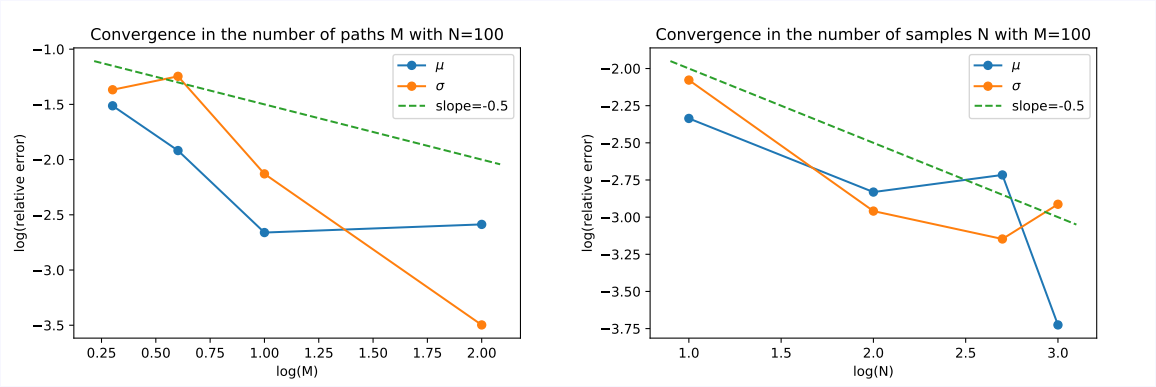


Figure 9: Optimized using L-BFGS-B algorithm. True parameters used are $\mu = 100$ and $\sigma = 10$.M is the number of paths and N is the number of samples.

Case 4: Following deterministic functions

As we construct processes that follow or revert to deterministic function it is natural to state the SDE in terms of deviations from the deterministic function. Then, we have for $V_t = X_t - p_t$

$$\begin{aligned} dV_t &= a(V_t; \boldsymbol{\theta}_t)dt + b(V_t; \boldsymbol{\theta}_t)dW_t \quad t > 0 \\ V_0 &= v_0 \end{aligned} \tag{18}$$

where we choose $a(V_t; \boldsymbol{\theta}_t) = -\theta_t V_t$ to have mean reversion with exponential speed in time controlled by some strictly positive parameter θ_t. To keep the process within a unit box, we choose $b(V_t; \boldsymbol{\theta}) = \sqrt{2\alpha_t \theta_t p_t (1-p_t) X_t (1-X_t)}$

Case 4: Implementation

In the previous sections we relied on the Euler discretization of SDEs in order to obtain our estimators through the Maximum Likelihood approach. However, this introduces errors in the likelihood similar to the observed error in the wiener process following the sine wave in (9). To prevent such artifacts, we solve for the moments of the process defined by the SDE in (18). We have,

$$\begin{aligned} \mathbb{E}[V_t^k] &= \mathbb{E}\Big[\Big(\int a(V_t;\boldsymbol{\theta}_t)dt + \int b(V_t;\boldsymbol{\theta}_t)dW_t\Big)^k\Big] \quad t>0 \\ V_0 &= v_0 \end{aligned} \tag{19}$$

For a Gaussian approximation, we are interested in the first two moments.

Case 4: Implementation - first moment

The first moment ($k = 1$) is given by,

$$\begin{aligned} m_1(t) = \mathbb{E}V_t &= \mathbb{E}\Big[\int a(V_t; \boldsymbol{\theta}_t)dt\Big] + \mathbb{E}\Big[\int b(V_t; \boldsymbol{\theta}_t)dW_t\Big] \\ &= \mathbb{E}\Big[\int a(V_t; \boldsymbol{\theta}_t)dt\Big] \end{aligned} \tag{20}$$

In our case, the drift is linear and is given by $a(V_t; \boldsymbol{\theta}_t) = -\theta_t V_t$. then, we have

$$m_1(t) = \mathbb{E}V_t = \Big[\int -\theta_t \mathbb{E}V_t dt\Big] \iff \frac{dm_1(t)}{dt} = -\theta_t m_1(t) \quad ; \quad m_1(0) = \mathbb{E}[V_0] = v_0 \tag{21}$$

This is an ODE which has solutions of the form,

$$m_1(t) = v_0 \exp(-\int_0^t \theta_s \, ds) \tag{22}$$

Case 4: Implementation - second moment

The second moment ($k = 2$) is given by,

$$= \mathbb{E}V_t^2 = \mathbb{E}\Big[\Big(\int a(V_t;\boldsymbol{\theta}_t)dt + \int b(V_t;\boldsymbol{\theta}_t)dW_t\Big)^2\Big] \tag{23}$$

Let $Y_t = g(V_t) = V_t^2$ then $g_t = 0$, $g^{'}(V_t) = 2V_t$ and $g^{''}(V_t) = 2$. Using *Itô* and Taylor,

$$\begin{aligned} dY_t = \frac{\partial g}{\partial t}dt + \frac{\partial g}{\partial v}\Big(a(V_t;\boldsymbol{\theta}_t)\ dt + b(V_t;\boldsymbol{\theta}_t)dW_t\Big) \\ + \frac{1}{2}\frac{\partial^2 g}{\partial v^2}\Big(a(V_t;\boldsymbol{\theta}_t)\ dt + b(V_t;\boldsymbol{\theta}_t)dW_t\Big)^2 + \mathscr{O}(dV_t^3) \end{aligned} \tag{24}$$

As $dt \to 0$, we have

$$\begin{aligned} dY_t &= \frac{\partial g}{\partial t} + \Big(a(V_t;\boldsymbol{\theta}_t)\frac{\partial g}{\partial v} + \frac{b^2(V_t;\boldsymbol{\theta}_t)}{2}\frac{\partial^2 g}{\partial v^2}\Big)dt + b(V_t;\boldsymbol{\theta}_t)\frac{\partial g}{\partial v}dW_t \\ &= 2\Big(a(V_t;\boldsymbol{\theta}_t)V_t + \frac{b^2(V_t;\boldsymbol{\theta}_t)}{2}\Big)dt + 2V_t b(V_t;\boldsymbol{\theta}_t)dW_t \end{aligned} \tag{25}$$

Case 4: Implementation - second moment

The second moment ($k = 2$) is given by,

$$\begin{aligned}\mathbb{E}[V_t^2] = \mathbb{E}\,Y_t &= 2\mathbb{E}\Big[\int \Big(a(V_t;\boldsymbol{\theta}_t)V_t + \frac{b^2(V_t;\boldsymbol{\theta}_t)}{2}\Big)dt\Big] + 2\mathbb{E}\Big[\int V_t b(V_t;\boldsymbol{\theta}_t)dW_t\Big] \\ &= 2\mathbb{E}\Big[\int \Big(a(V_t;\boldsymbol{\theta}_t)V_t + \frac{b^2(V_t;\boldsymbol{\theta}_t)}{2}\Big)dt\Big]\end{aligned} \tag{26}$$

In our case, the drift is $a(V_t;\boldsymbol{\theta}_t) = -\theta_t V_t$ and the diffusion is given by $b(V_t;\boldsymbol{\theta}) = \sqrt{2\alpha_t\theta_t p_t(1-p_t)X_t(1-X_t)} = \sqrt{2\alpha_t\theta_t p_t(1-p_t)(V_t+p_t)(1-V_t-p_t)}$. Then,

$$\begin{aligned}m_2(t) = \mathbb{E}[V_t^2] &= 2\int \Big(-\theta_t\mathbb{E}[V_t^2] + \alpha_t\theta_t p_t(1-p_t)\mathbb{E}[(V_t+p_t)(1-V_t-p_t)]\Big)dt \\ &= 2\int \Big(-\theta_t\mathbb{E}[V_t^2] + \alpha_t\theta_t p_t(1-p_t)(p_t(1-p_t) + (1-2p_t)\mathbb{E}[V_t] - \mathbb{E}[V_t^2])\Big)dt \\ &= 2\int \Big(-\mathbb{E}[V_t^2][\theta_t + \alpha_t\theta_t p_t(1-p_t)] + \mathbb{E}[V_t][\alpha_t\theta_t p_t(1-p_t)(1-2p_t)] + \alpha_t\theta_t p_t^2(1-p_t)^2\Big)dt\end{aligned}$$

Case 4: Implementation - second moment

The second moment $m_2(t)$ is given by solving the ODE,

$$\frac{dm_2(t)}{dt} = 2\Big(- m_2(t)[\theta_t + \alpha_t\theta_t p_t(1-p_t)] + m_1(t)[\alpha_t\theta_t p_t(1-p_t)(1-2p_t)] + \alpha_t\theta_t p_t^2(1-p_t)^2\Big)$$

where $m_2(0) = \mathbb{E}[V_0^2] = v_0^2$ and has a solution of the form,

$$m_2(t) = m_2(0)\exp\Big(-2\int_0^t [\theta_\tau + \alpha_\tau\theta_\tau p_\tau(1-p_\tau)]\ d\tau\Big) +$$

$$2\int_0^t \exp\Big(-2\int_0^s [\theta_\tau + \alpha_\tau\theta_\tau p_\tau(1-p_\tau)]\ d\tau\Big)\Big(m_1(t-s)[\alpha_{t-s}\theta_{t-s}p_{t-s}(1-p_{t-s})(1-2p_{t-s})]$$

$$+ \alpha_{t-s}\theta_{t-s}p_{t-s}^2(1-p_{t-s})^2\Big)\ ds$$

Following deterministic functions - Cases

We consider the following cases:

4.1 Constant parameters $\boldsymbol{\theta}_t = \boldsymbol{\theta}$

- 4.1.1 reverting to a constant function
- 4.1.2 reverting to a deterministic function

4.2 Time dependent parameters $\boldsymbol{\theta}_t$

Case 4.1: Following deterministic functions - constant parameters

In this case, the first moment $m_1(t)$ reduces to,

$$m_1(t) = v_0 \exp(-t\theta) \tag{27}$$

And the second moment $m_2(t)$ reduces to,

$$m_2(t) = m_2(0)\exp\Big(-2[t\theta + \alpha\theta \int_0^t p_\tau(1-p_\tau)\ d\tau]\Big) +$$
$$2\int_0^t \exp\Big(-2[s\theta + \alpha\theta \int_0^s p_\tau(1-p_\tau)\ d\tau]\Big)\Big(m_1(t-s)[\alpha\theta p_{t-s}(1-p_{t-s})(1-2p_{t-s})]$$
$$+ \alpha\theta p_{t-s}^2(1-p_{t-s})^2\Big)\ ds$$

Case 4.1.1: Following a constant function - constant parameters

For constant forecast $p_t = \frac{1}{2}$ the second moment reduce further,

$$m_2(t) = m_2(0)\exp\Big(-2t\theta(1+\frac{\alpha}{4})\Big) + \frac{\alpha\theta}{8}\int_0^t \exp\Big(-2s\theta(1+\frac{\alpha}{4})\Big)\, ds$$
$$= m_2(0)\exp\Big(-2t\theta(1+\frac{\alpha}{4})\Big) + \frac{\alpha}{16(1+\frac{\alpha}{4})}\Big(1-\exp(-2t\theta(1+\frac{\alpha}{4})\Big)$$

Then the variance is given by,

$$\mathbb{V}[V_t] = m_2(0)\exp\Big(-2t\theta(1+\frac{\alpha}{4})\Big) + \frac{\alpha}{16(1+\frac{\alpha}{4})}\Big(1-\exp(-2t\theta(1+\frac{\alpha}{4})\Big) - v_0^2\exp(-2t\theta) \tag{28}$$

Note that the second moment $m_2(0) = v_0^2$ as the previous point $V_{t_{i-1}} = v_0$ deterministic.

Case 4.1.1: Following a constant function - constant parameters

To conclude, we have the following mean and variance functions,

$$\begin{aligned}\mathbb{E}[V_{t_i}|V_{t_{i-1}} = v_0] &= v_0 \exp(-(t_i - t_{i-1})\theta) \\ \mathbb{V}[V_{t_i}|V_{t_{i-1}} = v_0] &= \frac{\alpha}{16(1+\frac{\alpha}{4})}\Big(1 - \exp(-2(t_i - t_{i-1})\theta(1+\frac{\alpha}{4})\Big)\end{aligned} \tag{29}$$

Next, we discretize and obtain the likelihood.

Case 4.1.1: Following a constant function - constant parameters

To obtain the likelihood, consider a set of N observations, $X = \{X_{t_0^N}, X_{t_1^N}, \ldots, X_{t_N^N}\}$ on intervals of Δ_N. We have $V = X - \frac{1}{2}$. Then,

$$\begin{aligned}\mathbb{E}[V_i|V_{i-1} = v_0] &= v_0 \exp(-\Delta_N \theta) \\ \mathbb{V}[V_i|V_{i-1} = v_0] &= \frac{\alpha}{16(1+\frac{\alpha}{4})}\Big(1 - \exp(-2\Delta_N\theta(1+\frac{\alpha}{4})\Big)\end{aligned} \tag{30}$$

The Likelihood of the transition density is given by,

$$\mathscr{L}(\theta; X) = \prod_{j=1}^{M}\prod_{i=1}^{N} \rho(V_{j,i+1}|V_{j,i}, \boldsymbol{\theta}) \tag{31}$$

And the Gaussian transition density,

$$\mathscr{L}(\boldsymbol{\theta}; V) = \prod_{j=1}^{M}\prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\mathbb{V}[V_{j,t_i}|V_{j,t_{i-1}} = v_0]}} \exp\Big(\frac{-(V_{j,i+1} - \mathbb{E}[V_{j,t_i}|V_{j,t_{i-1}} = v_0])^2}{2\mathbb{V}[V_{t_i}|V_{j,t_{i-1}} = v_0]}\Big)$$

Case 4.1.1: Following a constant function - constant parameters

We obtain the following convergence using the analytical moments obtained from the V_t process and enforcing that the parameters θ and α are non-negative.

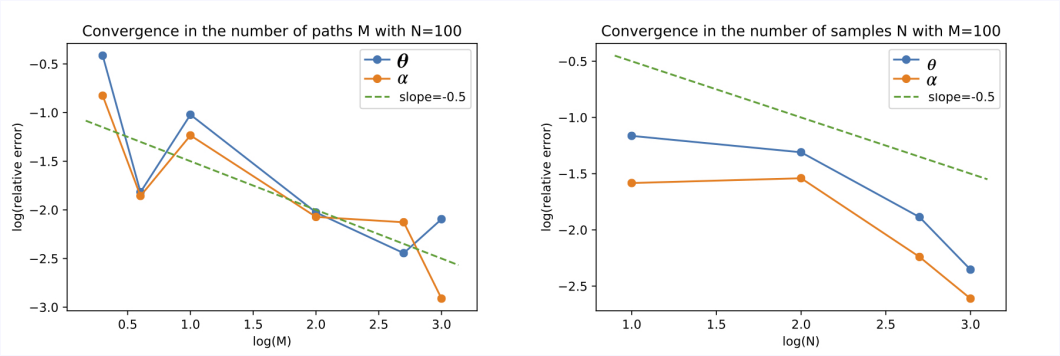


Figure 10: This is done for true parameters $\theta = 0.5$ and $\alpha = 1$ using L-BFGS-B optimization algorithm

Case 4.1.1: Following a constant function - constant parameters

Further, we note that the inference on the drift needs special care in situations of high speed reversion to the mean. Such situations seem to occur in the following cases:

- $\theta >> 1$ and big enough α
- $\Delta_N >> 1$

In both cases, we see a flattening in the likelihood in the direction of the drift parameter θ which causes the optimizer to return inconsistent estimates of the drift parameter θ. These two cases are applicable if the function or the process oscillates at a high frequency with respect to the other.

Case 4.1.1: Following a constant function - constant parameters

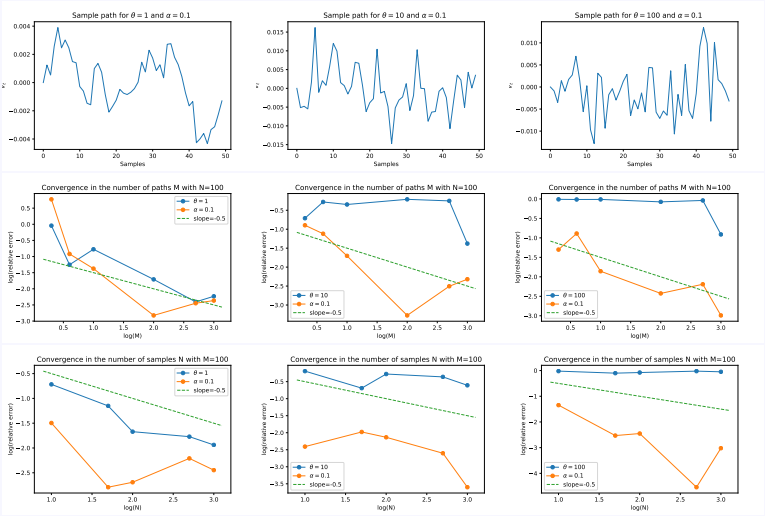


Figure 11: We can see how an increase of $\theta >> 1$ with big enough α causes sharp oscillations that make the inference of θ unfeasible.

Case 4.1.1: Following a constant function - constant parameters

In the second case of increasing Δ_N, we see the following flattening of the likelihood in the direction of θ.

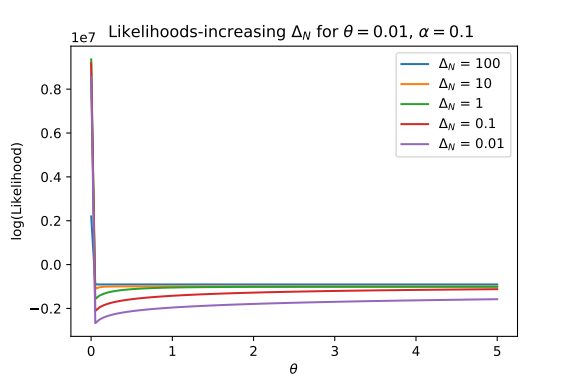


Figure 12: We can see how an increase of time intervals $\Delta_N >> 1$ causes a flattening of the likelihood making the valley of the optimal θ shallower until the inference of θ becomes unfeasible.

Case 4.1.1: Following a constant function - constant parameters

Question: How can we quantify the mean reversion of an SDE ?
The reversion to the mean is not only determined by the parameter θ but also by the diffusion term. We can see this in the following example,

$$dX_t = (1 + X_t)dt + (1 + X_t^2)dW_t \tag{32}$$

which at first glance does not seem mean reverting, but in fact, it is mean reverting. The reversion is driven solely by diffusion term in this case. This is a case of volatility induced stationary, as the non-linearity in the diffusion term are responsible for the mean reversion. A similar non-linearity is present in our model and quantifying the effective mean reversion will help determine if we are in the convergent regime where our estimator of the drift is consistent or not.
Regarding, the specific application to wind forecasts this should not be an issue as the forecast and the recorded data are unlikely to oscillate around each other at a "high" frequency.

Case 4.1.1: Following a constant function - constant parameters

Answer: Yes, it is true that such SDE's are stationary even though they are not "mean reverting". The "mean reversion" or the "pull" is induced solely by the diffusion term. We can show that

$$dX_t = (1 + X_t)dt + (1 + X_t^2)dW_t \tag{33}$$

is stationary. Even though its associated process with constant diffusion σ

$$dX_t = (1 + X_t)dt + \sigma dW_t \tag{34}$$

is clearly non-stationary. We can do that using the tools we will introduce next to check and find the stationary distributions of our SDEs. Namely, using the scale and speed densities.

Case 5.1.1: Theoretical Background

Consider an SDE,

$$\begin{aligned} dX_t &= a(X_t;\boldsymbol{\theta})dt + b(X_t;\boldsymbol{\theta})dW_t \quad t > 0 \\ X_0 &= x_0 \end{aligned} \tag{35}$$

We define the scale density $s(x)$,

$$s(x) = exp\Big[\int^x -\frac{2a(y;\boldsymbol{\theta})}{b^2(y;\boldsymbol{\theta})}\ dy\Big] \tag{36}$$

And the speed density $m(x)$,

$$m(x) = \frac{1}{b^2(y;\boldsymbol{\theta})\ s(x)} \tag{37}$$

Case 5.1.1: Theoretical Background

A unique stationary distribution of the process X_t exists if and only if

$$\int_{x_0}^{r} s(x)\ dx = \infty \quad \int_{l}^{x_0} s(x)\ dx = \infty \quad \int_{l}^{r} m(x)\ dx < \infty \tag{38}$$

where $x_0 \in (l, r)$ and (l, r) is an interval where $b^2(y; \boldsymbol{\theta}) > 0$

Case 5.1.1: Theoretical Background

Considering the SDE of a process following a constant forecast p,

$$\begin{aligned} dX_t &= -\theta(X_t - p)dt + \sqrt{2\theta\alpha p(1-p)X_t(1-X_t)}dW_t \quad t > 0 \\ X_0 &= x_0 \end{aligned} \tag{39}$$

We obtain the scale density $s(x)$ with integrability condition $\frac{p}{a} \leq -1$ and $\frac{1-p}{a} \leq -1$

$$s(x) = exp\Big[\int^x -\frac{2a(y;\boldsymbol{\theta})}{b^2(y;\boldsymbol{\theta})}\ dy\Big] = exp\Big[\int_{x_0}^x \frac{y-p}{ay(y-1)}\ dy\Big] \propto (1-x)^{\frac{1-p}{a}} x^{p/a} \tag{40}$$

where $a = -\alpha p(1-p) < 0$.Then the speed density $m(x)$ becomes,

$$m(x) = \frac{1}{b^2(y;\boldsymbol{\theta})\ s(x)} \propto (1-x)^{-\frac{1-p}{a}-1} x^{-\frac{p}{a}-1} \tag{41}$$

We note that $(l,r) = (-\infty,\infty)$ is where $b^2(y;\boldsymbol{\theta}) > 0$. Let $x_0 \in (l,r)$. Taking $x_0 = 1$ we see that

$$\int_{x_0}^r s(x)\ dx = \infty \quad \int_l^{x_0} s(x)\ dx = \infty \quad \int_l^r m(x)\ dx < \infty \tag{42}$$

Case 5.1.1: Theoretical Background

Under the condition,

$$\begin{cases} \frac{p}{a} \leq -1 \\ \frac{1-p}{a} \leq -1 \end{cases} \iff \min(p, 1-p) \geq -a \iff \min(p, 1-p) \geq \alpha p(1-p) \tag{43}$$

For example, if forecast is constant $p = \frac{1}{2}$, the condition becomes $\alpha \leq 2$.
we conclude that there exists a unique stationary distribution of X_t and its proportional to the speed density,

$$m(x) \propto (1-x)^{-\frac{1-p}{a}-1} x^{-\frac{p}{a}-1} \tag{44}$$

which is the kernal of a Beta distribution with shape parameters

$$Beta\left(\frac{p}{\alpha(1-p)}, \frac{1}{\alpha p}\right) \tag{45}$$

Case 5.1.1: Theoretical Background

Now we consider the centered process $V_t = X_t - p$,

$$\begin{aligned} dX_t &= -\theta V_t dt + \sqrt{2\theta\alpha p(1-p)(V_t+p)(1-V_t-p)}dW_t \quad t > 0 \\ X_0 &= x_0 \end{aligned} \tag{46}$$

We obtain the scale density $s(x)$,

$$s(x) = exp\Big[\int^x -\frac{2a(y;\boldsymbol{\theta})}{b^2(y;\boldsymbol{\theta})}\, dy\Big] = exp\Big[\int_{x_0}^x \frac{y-p}{ay^2+by+c}\, dy\Big] \tag{47}$$

where,

$$\begin{cases} a &= \alpha p(1-p) \\ b &= \alpha p(1-p)(2p-1) \\ c &= \alpha p^2(1-p)^2 \end{cases} \tag{48}$$

Case 5.1.1: Theoretical Background

To simplfy, we choose the constant forecast to be $p = \frac{1}{2}$. Then we have,

$$\begin{cases} a &= \frac{\alpha}{4} \\ b &= 0 \\ c &= \frac{\alpha}{16} \end{cases} \tag{49}$$

and the scale density becomes,

$$\begin{aligned} s(x) &= exp\Big[\int^x -\frac{2a(y;\boldsymbol{\theta})}{b^2(y;\boldsymbol{\theta})}\, dy\Big] = exp\Big[\int_{x_0}^x \frac{y}{ay^2+c}\, dy\Big] = exp\Big[\int_{x_0}^x \frac{y}{\frac{\alpha}{16}+\frac{\alpha y^2}{4}}\, dy\Big] \\ &\propto \left(4v^2+1\right)^{2/\alpha} \qquad (\alpha > 0) \end{aligned} \tag{50}$$

Then we have the speed density $m(x)$,

$$m(x) = \frac{1}{b^2(y;\boldsymbol{\theta})\; s(x)} \propto \left(4v^2+1\right)^{-\frac{\alpha+2}{\alpha}} \tag{51}$$

Case 5.1.1: Theoretical Background

We proceed to check if it is stationary.
We note that $(l, r) = (-\infty, \infty)$ is where $b^2(y; \boldsymbol{\theta}) > 0$. Let $x_0 \in (l, r)$. Taking $x_0 = 0$, we see that

$$\int_{x_0}^{r} s(x)\ dx = \infty \quad \int_{l}^{x_0} s(x)\ dx = \infty \quad \int_{l}^{r} m(x)\ dx < \infty \tag{52}$$

So we conclude that it is a stationary distribution of V_t and its unique. Hence, the kernal of the invarient distribution of the process V_t with constant forecast $p = \frac{1}{2}$ is $\pi(v) = \frac{\sqrt{\pi}\ \Gamma(\frac{1}{2}+\frac{2}{\alpha})}{\Gamma(\frac{2+\alpha}{\alpha})} \left(1 + 4v^2\right)^{-\frac{\alpha+2}{\alpha}}$. Note that it almost resembles a t-student distribution.

Case 5.1.1: Theoretical Background

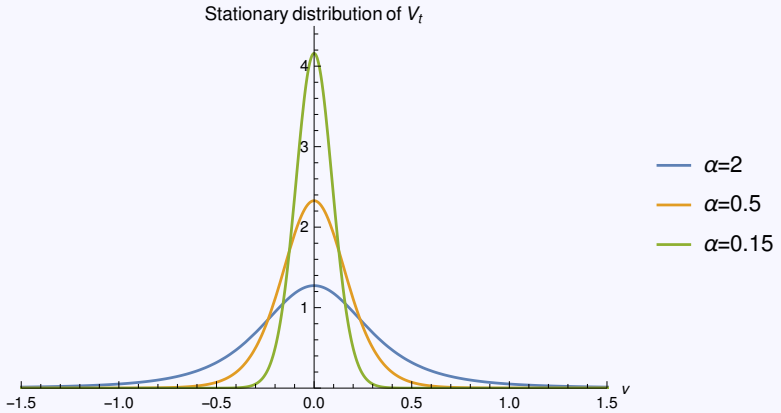


Figure 13: The stationary distribution $\pi(v)$ of the process V_t is symmetric with heavy tails.

Following deterministic functions - Cases

In an improved model, we incorperate Beta density transitions. We repeat the same cases as follows:

5.1 Constant parameters $\boldsymbol{\theta}_t = \boldsymbol{\theta}$

- 5.1.1 reverting to a constant function
- 5.1.2 reverting to a deterministic function

5.2 Time dependent parameters $\boldsymbol{\theta}_t$

However, first lets discuss why and how do we incorperate the Beta transition densities.

Case 5.1.1: Following a constant function - constant parameters

Recall that both X_t and V_t are controlled processes that are constrained within a unit box. The symmetry of the Gaussian transition density of these two processes indicate that we have some probability outside the unit box once the forecast touches the boundry. There are many ways to fix this behaviour:

1. Use a Beta transition density for all the transitions while matching the moments of the Gaussian transition density.
2. Use a Beta transition density only when the transition density of the process touches the boundry of the unit box.This depends on the width and position of the transition density.
   - 2.1 match the parameters of the Beta transition with the moments of the Gaussian transition.
   - 2.2 Not to match the moments of the Gaussian and introdude the beta parameters as two additional hyperparameters (increasing the degrees of freedom).

Case 5.1.1: Following a constant function - constant parameters

If we like to enforce that the Beta transition matches the moments of the Gaussian, then we can solve for the paramters of the Beta distribution in terms of the parameters of the Gaussian transition density.
First, lets consider the non-centered process X_t. In that case, we choose standard Beta distribution which has support on $x \in [0,1]$. Then solving for its parameters in terms of the parameters of the Gaussian transition denisty, to obtain

$$\alpha = \Big(\frac{1-\tilde{\mu}}{\tilde{\sigma}^2} - \frac{1}{\tilde{\mu}}\Big)\tilde{\mu}^2 \qquad \beta = \alpha\Big(\frac{1}{\tilde{\mu}} - 1\Big) \tag{53}$$

where $\tilde{\sigma}^2$ is the variance of the Gaussian transition density of the process and $\tilde{\mu}$ is its mean.

Case 5.1.1: Following a constant function - constant parameters

To ensure the existence of such a Beta distribution with the parameters required

$$\alpha = \Big(\frac{1-\tilde{\mu}}{\tilde{\sigma}^2} - \frac{1}{\tilde{\mu}}\Big)\tilde{\mu}^2 \qquad \beta = \alpha\Big(\frac{1}{\tilde{\mu}} - 1\Big) \tag{54}$$

we have the followind conditions by the definition of the standard Beta distribution:

1. $x \in [0,1]$.
2. Its shape parameters α and β must be positive definite.

Solution (this only applies to Euler realizations due to time discretization):

1. Discard realizations X_t outside the range $(0,1)$.
2. Ensure that $\alpha > 0$ and $\beta > 0$ with high probability.

The first condition is automatically satisfied with high probability. The second condition introduces the following restrictions:

$$\alpha > 0,\ \beta > 0 \iff 0 < \tilde{\mu}_t < 1 \text{ and } \tilde{\sigma}_t^2 < \tilde{\mu}_t - \tilde{\mu}_t^2 \tag{55}$$

while the first restriction $0 < \tilde{\mu}_t < 1$ holds with very high probability. The second restricts the second moment of the process X_t not to exceed its mean, i.e $\mathbb{E}[X_t^2] \leq \tilde{\mu}_t \quad \forall t$

Case 5.1.1: Following a constant function - constant parameters

Next, we consider the centered process V_t. Since the process V_t is inside the range $(-1,1)$ with high probability, the standard Beta distribution cannot model its transitions. Therefore, we translate the distribution to have support on $[a,b]=[-1,1]$. We use the linear transformation $y=g(x)=a+(b-a)x$ then we have

$$g^{-1}(y)=\frac{a-y}{b-a} \quad \text{and} \quad \left|\frac{\partial g^{-1}(y)}{\partial y}\right|=\frac{1}{b-a} \tag{56}$$

And the general Beta distribution becomes,

$$f_Y(y)=f_X(g^{-1}(y))\left|\frac{\partial g^{-1}(y)}{\partial y}\right|=\frac{1}{B(\alpha,\beta)(b-a)}\left(\frac{a-y}{b-a}\right)^{\alpha-1}\left(1-\frac{a-y}{b-a}\right)^{\beta-1} \tag{57}$$

and we have,

$$\mathbb{E}[Y]=a+(b-a)\mathbb{E}[X]=a+(b-a)\frac{\alpha}{\alpha+\beta} \tag{58}$$

$$\mathbb{V}[Y]=(b-a)^2\mathbb{V}[X]=(b-a)^2\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} \tag{59}$$

Case 5.1.1: Following a constant function - constant parameters

Next, we need to solve the system in terms of the mean and variance of the process V_t.

$$\begin{cases} \tilde{\mu}_t = a + (b-a)\frac{\alpha}{\alpha+\beta} \\ \tilde{\sigma}_t^2 = (b-a)^2 \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} \end{cases} \tag{60}$$

to obtain,

$$\alpha = \frac{(\tilde{\mu}_t - a)\left(\tilde{\mu}_t^2 - \tilde{\mu}_t(a+b) + \tilde{\sigma}_t^2 + ab\right)}{\tilde{\sigma}_t^2(a-b)} \tag{61}$$

$$\beta = \frac{(\tilde{\mu}_t - b)\left(\tilde{\mu}_t^2 - \tilde{\mu}_t(a+b) + \tilde{\sigma}_t^2 + ab\right)}{\tilde{\sigma}_t^2(b-a)} \tag{62}$$

then for our interval of interest $[a,b] = [-1,1]$, we have

$$\alpha = -\frac{(1+\tilde{\mu}_t)(\tilde{\mu}_t^2 + \tilde{\sigma}_t^2 - 1)}{2\tilde{\sigma}_t^2} \qquad \beta = \frac{(\tilde{\mu}_t - 1)(\tilde{\mu}_t^2 + \tilde{\sigma}_t^2 - 1)}{2\tilde{\sigma}_t^2} \tag{63}$$

Case 5.1.1: Following a constant function - constant parameters

As before, for the general Beta distribution on the interval $[a, b] = [-1, 1]$, we must have that

1. $x \in [-1, 1]$.
2. Its shape parameters α and β must be positive definite.

Solution (this only applies to Euler realizations due to time discretization):

1. Discard realizations V_t outside the range $(-1, 1)$.
2. Ensure that $\alpha > 0$ and $\beta > 0$ with high probability.

The first condition is automatically satisfied with high probability. The second condition introduces the following restrictions:

$$\alpha > 0,\ \beta > 0 \iff 0 < \tilde{\mu}_t < 1 \text{ and } \tilde{\sigma}_t^2 < 1 - \tilde{\mu}_t^2 \tag{64}$$

while the first restriction $-1 < \tilde{\mu}_t < 1$ holds with very high probability. The second restricts the second moment of the process V_t not to exceed one, i.e. $\mathbb{E}[V_t^2] \leq 1 \quad \forall t$

Case 5.1.1: Following a constant function - constant parameters

If we approximate all transitions by a Beta transition density, we obtain the log-likelihood,

$$\ell(\boldsymbol{\theta}; V) = \sum_{j=1}^{M} \sum_{i=1}^{N} (\alpha - 1) \log(V_{j,i+1}) + (\beta - 1) \log(1 - V_{j,i+1}) - \log(B(\alpha, \beta)) \tag{65}$$

where $B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$ is the Beta function. And enforcing to match the transition moments of the Gaussian, we have

$$\alpha = \Big(\frac{1 - E[V_{j,t_i} | V_{j,t_{i-1}} = v_0]}{\mathbb{V}[V_{t_i} | V_{j,t_{i-1}} = v_0]} - \frac{1}{E[V_{j,t_i} | V_{j,t_{i-1}} = v_0]} \Big) E[V_{j,t_i} | V_{j,t_{i-1}} = v_0]^2 \tag{66}$$

$$\beta = \alpha \Big(\frac{1}{\mathbb{E}[V_{j,t_i} | V_{j,t_{i-1}} = v_0]} - 1 \Big) \tag{67}$$

Case 5.1.1: Following a constant function - constant parameters

Another approach is to only approximate by a Beta distribution if the transition density of X_t is sufficiently crossing the boarder of the unit box. For example, we approximate by a Beta the samples that satisfy either of the following conditions,

$$\begin{cases} \mu_{t_i} + 3\sigma_{t_i} > 1 \\ \mu_{t_i} - 3\sigma_{t_i} < 0 \end{cases} \iff \begin{cases} E[V_{j,t_i} | V_{j,t_{i-1}} = v_0] + p + 3\sqrt{\mathbb{V}[V_{t_i} | V_{j,t_{i-1}} = v_0]} > 1 \\ E[V_{j,t_i} | V_{j,t_{i-1}} = v_0] + p - 3\sqrt{\mathbb{V}[V_{t_i} | V_{j,t_{i-1}} = v_0]} < 0 \end{cases} \tag{68}$$

If a sample $V_{j,t_{i-1}}$ satisfies either of these conditions, we denote it by $\bar{V}_{j,t_{i-1}}$ and the number of such samples by $\bar{N}$.

Case 5.1.1: Following a constant function - constant parameters

We then construct the following likelihood,

$$\mathscr{L}(\theta; X) = \prod_{j=1}^{M} \left[\prod_{i=1}^{N-\bar{N}} \rho(V_{j,i+1} | V_{j,i}, \boldsymbol{\theta}) \prod_{i=1}^{\bar{N}} \bar{\rho}(\bar{V}_{j,i+1} | \bar{V}_{j,i}, \boldsymbol{\theta}) \right] \tag{69}$$

and the log-likelihood becomes,

$$\begin{aligned} \ell(\boldsymbol{\theta}; V) = \sum_{j=1}^{M} \Big(\sum_{i=1}^{\bar{N}} (\alpha - 1) \log(\bar{V}_{j,i+1}) + (\beta - 1) \log(1 - \bar{V}_{j,i+1}) - \log(B(\alpha, \beta)) \\ + \sum_{i=1}^{N-\bar{N}} - \frac{(V_{j,i+1} - \mathbb{E}[V_{j,t_i} | V_{j,t_{i-1}} = v_0])^2}{2\mathbb{V}[V_{t_i} | V_{j,t_{i-1}} = v_0]} - \frac{1}{2} \Big(2\pi \mathbb{V}[V_{j,t_i} | V_{j,t_{i-1}} = v_0] \Big) \Big) \end{aligned} \tag{70}$$

where α and β are given as in (66) and (67).

Case 5.2.1: Time dependent Forecast

We approximate all transitions by a Beta transition density, we obtain the log-likelihood,

$$\ell(\boldsymbol{\theta}; V) = \sum_{j=1}^{M}\sum_{i=1}^{N}(\alpha-1)\log(\frac{V_{j,i+1}+1}{2}) + (\beta-1)\log(1-\frac{V_{j,i+1}+1}{2}) - \log(2B(\alpha,\beta)) \quad (71)$$

where α and β are given as in (72):

$$\alpha = -\frac{(1+\tilde{\mu}_t)(\tilde{\mu}_t^2+\tilde{\sigma}_t^2-1)}{2\tilde{\sigma}_t^2} \qquad \beta = \frac{(\tilde{\mu}_t-1)(\tilde{\mu}_t^2+\tilde{\sigma}_t^2-1)}{2\tilde{\sigma}_t^2} \quad (72)$$

where we compute the moments numerically before each realization using Euler

$$\tilde{\mu} = V_{j,i-1}e^{-\Delta N\theta} \quad (73)$$

$$\begin{aligned}\tilde{\sigma}_t^2 = V_{j,i-1}^2 + 2\Delta N(-V_{j,i-1}^2(\theta+\alpha\theta p_i(1-p_i)) + \\ V_{j,i-1}(\alpha\theta p_i(1-p_i)(1-2p_i)) + \alpha\theta p_i^2(1-p_i)^2) - \tilde{\mu}^2\end{aligned} \quad (74)$$

where p_i is the discretization of the time dependent forecast.

Case 5.2.1: Time dependent Forecast

We generate samples using the previous Euler scheme to numerically compute the moments before each realization.We obtain the following sample paths

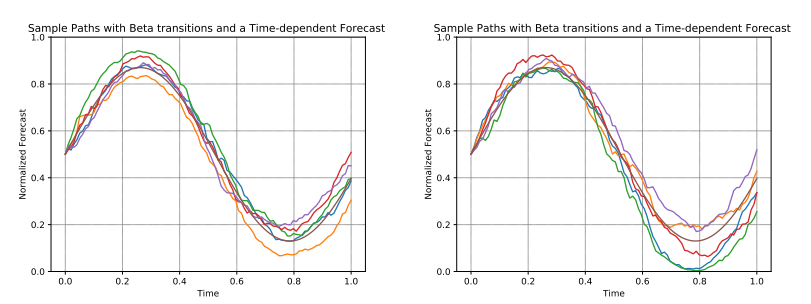


Figure 14: Sample Paths using Beta transitions LHS: $\theta = 0.5$ and time dependent path variablitiy $\alpha(t) = \alpha p_t(1 - p_t)$ where $\alpha = 0.3$. RHS: $\theta = 0.5$, $\alpha = 0.5$

Case 5.2.1: Time dependent Forecast

We perform a convergence test for speed of reversion parameter $\theta = 0.5$ and time dependent path variablitiy $\alpha(t) = \alpha p_t(1 - p_t)$ where $\alpha = 0.3$. We obtain the following convergence result

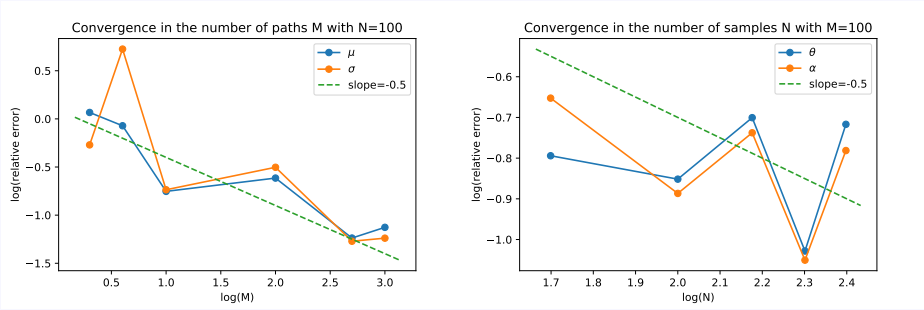


Figure 15: Convergence results for Beta transitions with $\theta = 0.5$ and time dependent path variablitiy $\alpha(t) = \alpha p_t(1 - p_t)$ where $\alpha = 0.3$

Real data and Forecast

- We have compiled 1265 forecasts from two full years (2016-2017) along with the measured real power generation.
- The real power generation data has been matched to the forecast using timestamps (Unix time). We have 72 hourly observations for each forecast.
- Data is normalized according to the installed power at that specific period of the forecast.

Real data and Forecast

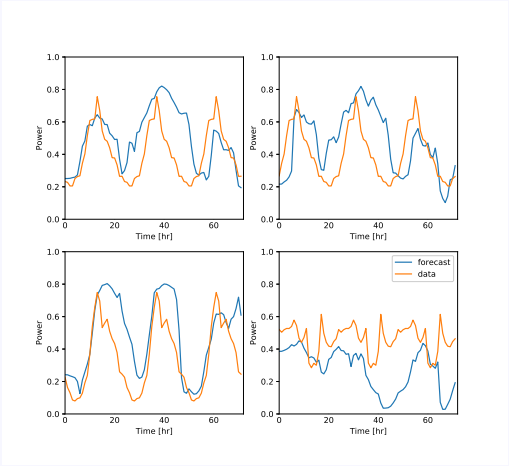


Figure 16: Random selection of forecasts and their associated real power generation data

Forecast cubic splines interpolation

For better numerical integration of the ODEs in order to find the moments of the stochastic process V_t, we need to interpolate between the discrete values of the forecast (as the given forecast is not fine-grained).
In this case, we choose to use cubic splines.

Forecast cubic splines interpolation

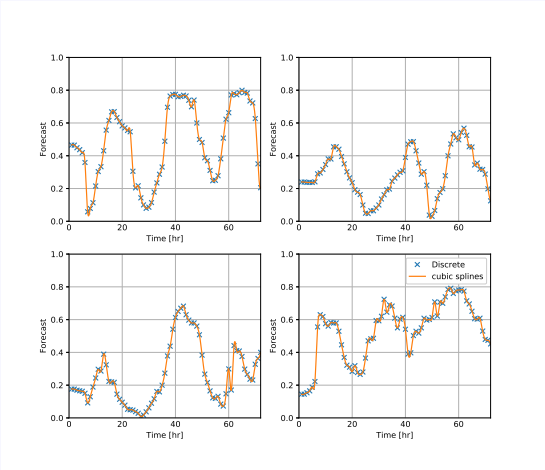


Figure 17: Random selection of forecasts discrete values and the associated cubic splines interpolation

Reader's Guide

List of changes in this iteration:

- Matched data with every forecast
- Normalized according to installed power of each period
- Interpolated the forecast using cubic splines.

Next steps:

- upgrade the numerical integrator and use the interpolated forecast
- Check performace using the interpolated forecasts and data

Note :

- Green slides: possible future extensions
- Red slides: notes to be removed