Gaussian Processes Part 2

Max Kochurov

MSU

February 1, 2023

Agenda

❶ Introduction

❷ GP approach
- Introduction
- Non-periodic part
- Periodic part
- The Model

❸ Modelling
- Priors
- Parametrization
- Seasonality

❹ Stochastic Volatility

Time Series, Classical Approach

If data has seasonality, you usually use STL decomposition. However,

- Parameters are not interpretable, only decomposition is available
- No uncertainty estimates
- Quite strict on input values
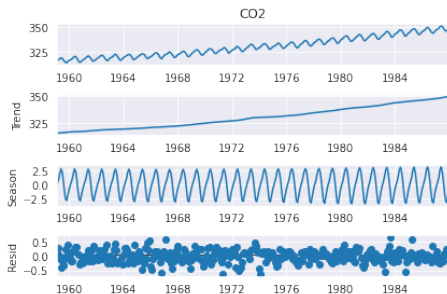- Significantly less flexible in modelling

Figure: STL decomposition for CO2 data, Statsmodels

GP decomposition

A Gaussian process can handle a complicated set of assumptions in addition to what STL provides

- Granular seasonality (year + quarter + month + week)
- Changepoint models
- Flexible likelihood Function
- Panel regression models
- Missing values

Typical Model

Typical model is additive

$$x_t \sim \underbrace{g(t)}_{non-periodic} + \underbrace{s(t)}_{periodic} + \underbrace{h(t)}_{holidays}$$

Reference

See more in Prophet preprint [1]. Every time series model is unique

Reminder

$x \in \mathbb{R}^n$, $y \in \mathbb{R}$

$$Y \sim \mathcal{GP}(m(x), k(x, x'))$$

Reminder

$x \in \mathbb{R}^n$, $y \in \mathbb{R}$

$$Y \sim \mathcal{GP}(m(x), k(x, x'))$$

$$\begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} m(x_1) \\ \vdots \\ m(x_N) \end{bmatrix}, \begin{bmatrix} k(x_1, x_1) & \dots & k(x_1, x_N) \\ \vdots & \ddots & \vdots \\ k(x_N, x_1) & \dots & k(x_N, x_N) \end{bmatrix} \right)$$

1. $\mathcal{GP}$ Gaussian Process - simply, a normal distribution with special mean $m(x)$ and covariance $k(x, x')$

Reminder

$x \in \mathbb{R}^n$, $y \in \mathbb{R}$

$$Y \sim \mathcal{GP}(m(x), k(x, x'))$$

$$\begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix} \sim \mathcal{N} \left(\begin{bmatrix} m(x_1) \\ \vdots \\ m(x_N) \end{bmatrix}, \begin{bmatrix} k(x_1, x_1) & \dots & k(x_1, x_N) \\ \vdots & \ddots & \vdots \\ k(x_N, x_1) & \dots & k(x_N, x_N) \end{bmatrix} \right)$$

1. $\mathcal{GP}$ Gaussian Process - simply, a normal distribution with special mean $m(x)$ and covariance $k(x, x')$
2. $m(x)$ - mean function, e.g.
   - Linear regression $m(x) = x^\top \beta$
   - Simply Constant or Zero $m(x) = c$
   - Other custom functions $m(x) = \sin(x)$

Reminder

$x \in \mathbb{R}^n$, $y \in \mathbb{R}$

$$Y \sim \mathcal{GP}(m(x), k(x, x'))$$

$$\begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix} \sim \mathcal{N} \left(\begin{bmatrix} m(x_1) \\ \vdots \\ m(x_N) \end{bmatrix}, \begin{bmatrix} k(x_1, x_1) & \dots & k(x_1, x_N) \\ \vdots & \ddots & \vdots \\ k(x_N, x_1) & \dots & k(x_N, x_N) \end{bmatrix} \right)$$

1. $\mathcal{GP}$ Gaussian Process - simply, a normal distribution with special mean $m(x)$ and covariance $k(x, x')$
2. $m(x)$ - mean function, e.g.
   - Linear regression $m(x) = x^\top \beta$
   - Simply Constant or Zero $m(x) = c$
   - Other custom functions $m(x) = \sin(x)$
3. $k(x, x')$ - kernel function, simply - measure of similarity for x and x'
   - $[K]_{ij} = k(x_i, x_j)$ is an SPD matrix

Non-periodic Part (mean function)

- Growth models
- Linear trend models
- Changepoint models

Non-periodic Part (mean function)

- Growth models
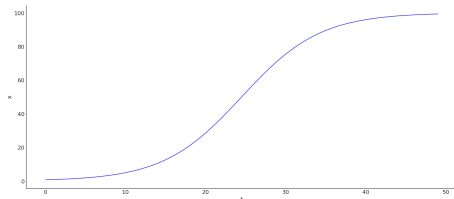- Linear trend models
- Changepoint models

Figure: Growth Model

$$x = \frac{c}{1 + \exp(-k(t - m))}$$

Non-periodic Part (mean function)

- Growth models
- Linear trend models
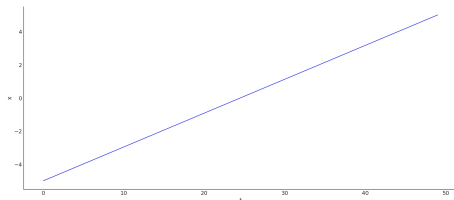- Changepoint models

Figure: Linear Trend Model

$$x = \frac{c}{1 + \exp(-k(t - m))}$$

Non-periodic Part (mean function)

- Growth models
- Linear trend models
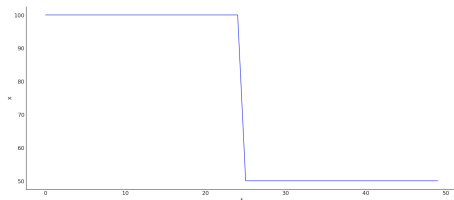- Changepoint models

Figure: Changepoint Model

$$x = \begin{cases} c_1, & t < m \\ c_2, & t \geq m \end{cases}$$

Non-periodic Part (mean function)

- Growth models
- Linear trend models
- Changepoint models

Extentions

Extensions are possible, e.g. time dependent saturation in the growth model. See in [1]
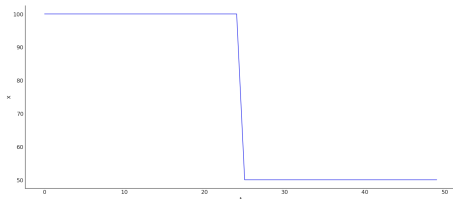

Figure: Changepoint Model

$$x = \begin{cases} c_1, & t < m \\ c_2, & t \geq m \end{cases}$$

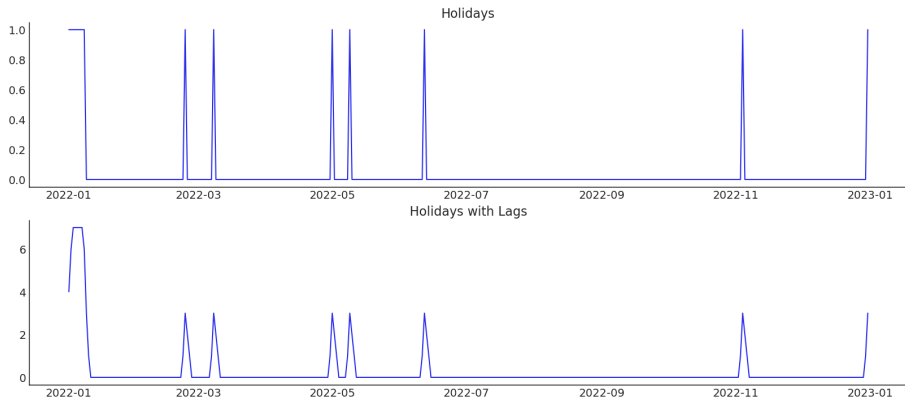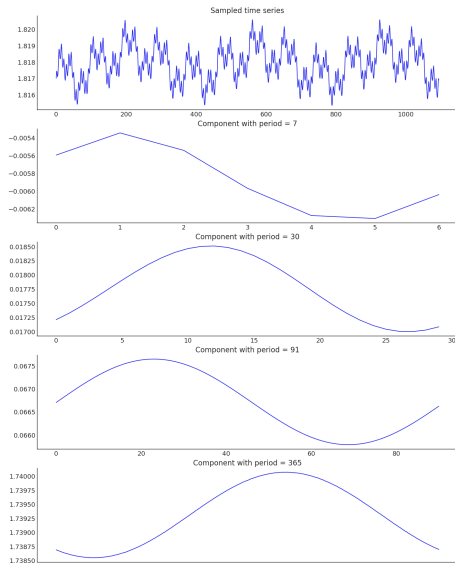# Holidays

$$h(t) = \text{is-holiday}(t)$$

Figure: Holiday features

Periodic part (cov function)

Granularities are important here. Multiple Periodic kernels can be used.

- Yearly
- Quarterly
- Monthly
- Weekly

Lengthscales for Periodic Part

Hyperparameters

Common sense driven lenthscale choice

- Week - couple of days make a change ($ls \approx 3$)
- Month - week makes sense for a dramatic change ($ls \approx 7$)
- Quarter - month makes sense for a dramatic change ($ls \approx 30$)
- Year - quarter makes sense for a dramatic change ($ls \approx 90$)

Lengthscales for Periodic Part

Hyperparameters

Common sense driven lenthscale choice

- Week - couple of days make a change ($ls \approx 3$)
- Month - week makes sense for a dramatic change ($ls \approx 7$)
- Quarter - month makes sense for a dramatic change ($ls \approx 30$)
- Year - quarter makes sense for a dramatic change ($ls \approx 90$)

In practice

Hard to infer lengthscales, better just fix them if informed

Putting All Together

$$m(t) = \underbrace{g(t)}_{non-periodic} + \underbrace{h(t)}_{holidays}$$

$$
\begin{aligned}
k(*, *) = \quad & \text{Periodic}(\mathsf{p} = 365, \mathsf{l} = 90) \\
+ \quad & \text{Periodic}(\mathsf{p} = 90, \mathsf{l} = 30) \\
+ \quad & \text{Periodic}(\mathsf{p} = 30, \mathsf{l} = 7) \\
+ \quad & \text{Periodic}(\mathsf{p} = 7, \mathsf{l} = 3)
\end{aligned}
$$

Putting All Together

$$m(t) = \underbrace{g(t)}_{non-periodic} + \underbrace{h(t)}_{holidays}$$

$$k(*, *) = \alpha_{365}\, \mathrm{Periodic}(\mathsf{p} = 365, \mathsf{l} = 90)$$
$$+ \alpha_{90}\, \mathrm{Periodic}(\mathsf{p} = 90, \mathsf{l} = 30)$$
$$+ \alpha_{30}\, \mathrm{Periodic}(\mathsf{p} = 30, \mathsf{l} = 7)$$
$$+ \alpha_{7}\, \mathrm{Periodic}(\mathsf{p} = 7, \mathsf{l} = 3)$$

Missing Parts

1. Weights for periodic components

Putting All Together

$$m(t) = \underbrace{g(t)}_{non-periodic} + \underbrace{h(t)}_{holidays}$$

$$k(*, *) = \alpha_{365}\, \text{Periodic}(p = 365, l = 90)$$
$$+ \alpha_{90}\, \text{Periodic}(p = 90, l = 30)$$
$$+ \alpha_{30}\, \text{Periodic}(p = 30, l = 7)$$
$$+ \alpha_{7}\, \text{Periodic}(p = 7, l = 3)$$
$$+ \beta\, \text{ExpQuad}(l = 90) + \text{WhiteNoise}(\gamma)$$

Missing Parts

1. Weights for periodic components
2. Trent violations from $g(t)$ that can't be explained by periodic component and observation noise

Example

CO2 data (Manua Loa, Hawaii)

- Monthly observations
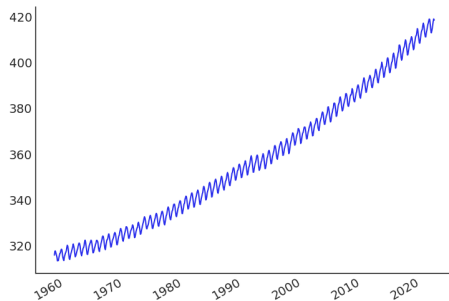- Clean data
- Unclear trend function

Figure: CO2 level

Example

CO2 data (Manua Loa, Hawaii)

- Monthly observations
- Clean data
- Unclear trend function

Where to start?

❶ Decide on a trend model
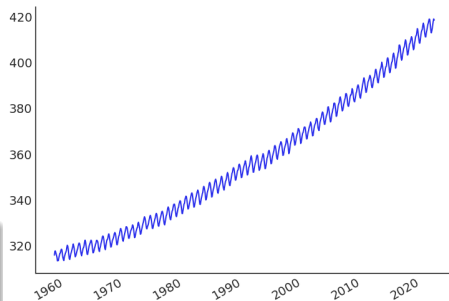
❷ Add periodic component

Figure: CO2 level

Priors

The exponential Growth Model

$$y_t = y_0 \cdot g^t$$

1. The baseline level for CO2 (data tells it is ≈ 310)
2. Growth rate for CO2 (some plausible value)
   - unlikely decay
   - no astronomic values

Prior Predictive Checks (Naive Attempt)

$$g \sim \text{Normal}(1.1, 0.1)$$
$$\epsilon \sim \text{LogNormal}(0, 1)$$
$$y_0 \sim \text{Normal}(310, 10)$$
$$y_t \sim \text{Normal}(y_0 \cdot g^t, \epsilon)$$

Figure: Prior Predictive Check

Prior Predictive Checks (Naive Attempt)

$$g \sim \text{Normal}(1.1, 0.1)$$
$$\epsilon \sim \text{LogNormal}(0, 1)$$
$$y_0 \sim \text{Normal}(310, 10)$$
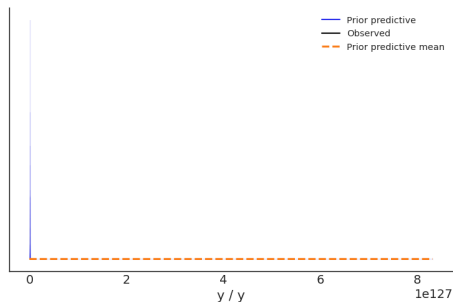$$y_t \sim \text{Normal}(y_0 \cdot g^t, \epsilon)$$

PPC Observation

Model is unlikely to be correctly specified. Prior predictive ranges to 10^{127} **parts per million** of CO2. Air has $2.5 \cdot 10^{25}$ molecules in m^3
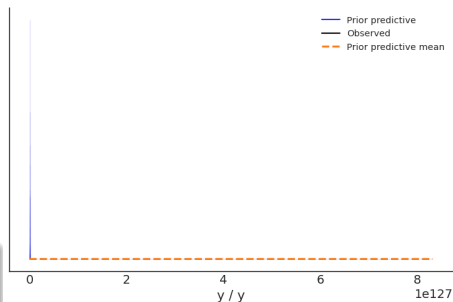

Figure: Prior Predictive Check

Prior Predictive (Problem Driven)

Thinking loudly

I do not believe in a century we can exceed 400% growth (we would have died already) and I suppose it should be at least 20% (it would not be a problem otherwise)

Prior Predictive (Problem Driven)

Thinking loudly

I do not believe in a century we can exceed 400% growth (we would have died already) and I suppose it should be at least 20% (it would not be a problem otherwise)

1. 400% in a century is 0.115% per month
2. 20% in a century is 0.015% per month

Prior Predictive (Problem Driven)

> **Thinking loudly**
>
> I do not believe in a century we can exceed 400% growth (we would have died already) and I suppose it should be at least 20% (it would not be a problem otherwise)

1. 400% in a century is 0.115% per month
2. 20% in a century is 0.015% per month

$$g \sim \text{Normal}(1.00065, 0.0005)$$

Prior Predictive (Problem Driven)

$g \sim \text{Normal}(1.00065, 0.0005)$

$\epsilon \sim \text{LogNormal}(0, 1)$

$y_0 \sim \text{Normal}(310, 10)$

$y_t \sim \text{Normal}(y_0 \cdot g^t, \epsilon)$

Figure: Prior Predictive Check

Prior Predictive (Problem Driven)

$$g \sim \text{Normal}(1.00065, 0.0005)$$
$$\epsilon \sim \text{LogNormal}(0, 1)$$
$$y_0 \sim \text{Normal}(310, 10)$$
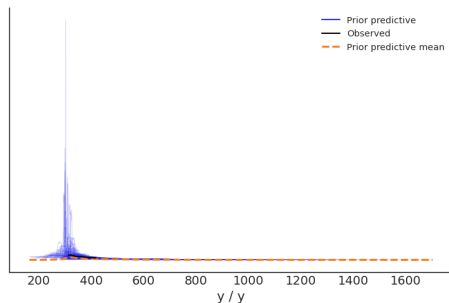$$y_t \sim \text{Normal}(y_0 \cdot g^t, \epsilon)$$

PPC Observation

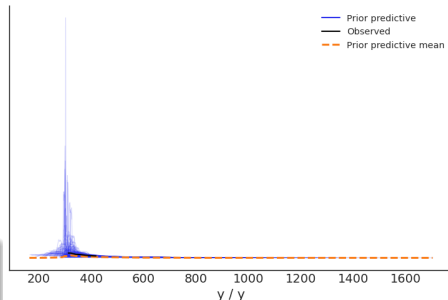Model acts as expected ranging CO2 levels in plausible regions. It seems to be good enough to go to the next step.

Figure: Prior Predictive Check

Inference

1. In sampling any value could appear

2. Here we got a very large value for some parameter

3. We can try to reparametrize

```
:   with model:
        trace = pm.sample()

    Auto-assigning NUTS sampler...
    Initializing NUTS using jitter+adapt_diag...
    /home/ferres/.miniconda3/envs/bayes-econ/lib/python3.10/site-p
    ric.py:57: RuntimeWarning: overflow encountered in multiply
      return bound(*args, **kwds)
    Multiprocess sampling (4 chains in 4 jobs)
    NUTS: [initial, g, eps]
```

Figure: Overflow in sampling

Log Model

$$\log y_t = \log y_0 + g \cdot t$$

$$g \sim \text{Normal}(1.00065, 0.0005)$$
$$\epsilon \sim \text{LogNormal}(-2, 0.5)$$
$$\log y_0 \sim \text{Normal}(\log 310, 10/310)$$
$$y_t \sim \text{LogNormal}(\log y_0 + g \cdot t, \epsilon)$$

Important

ϵ has a different meaning in the model. It is now a relative error.

Prior Predictive (Reparametrized)

$$g \sim \text{Normal}(1.00065, 0.0005)$$
$$\epsilon \sim \text{LogNormal}(-2, 0.5)$$
$$\log y_0 \sim \text{Normal}(\log 310, 10/310)$$
$$y_t \sim \text{LogNormal}(\log y_0 + g \cdot t, \epsilon)$$

Figure: Prior Predictive Check

Prior Predictive (Reparametrized)

$$g \sim \text{Normal}(1.00065, 0.0005)$$
$$\epsilon \sim \text{LogNormal}(-2, 0.5)$$
$$\log y_0 \sim \text{Normal}(\log 310, 10/310)$$
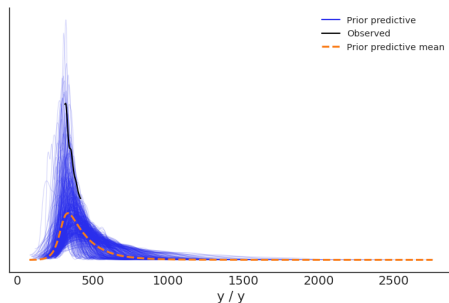$$y_t \sim \text{LogNormal}(\log y_0 + g \cdot t, \epsilon)$$

PPC Observation

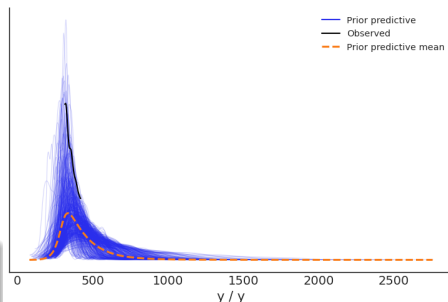Model acts as expected ranging CO2 levels in plausible regions. It seems to be good enough to go to proceed with sampling.

Figure: Prior Predictive Check

Inference

1. Everything is good
2. We can visualize

```
with model:
    trace = pm.sample()

Auto-assigning NUTS sampler...
Initializing NUTS using jitter+adapt_diag...
Multiprocess sampling (4 chains in 4 jobs)
NUTS: [initial, g, eps]
```
100.00% [8000/8000
Sampling 4 chains, 0 divergences]
```
Sampling 4 chains for 1_000 tune and 1_000 draw iterations (4
s total) took 6 seconds.
```

Figure: Good sampling

Inference

1. Everything is good
2. We can visualize
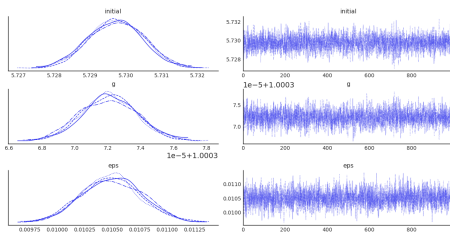3. Trace plot is a good visual inspection if sampling went well

Figure: Good Trace Plot

```
with model:
    trace = pm.sample()
```

```
Auto-assigning NUTS sampler...
Initializing NUTS using jitter+adapt_diag...
Multiprocess sampling (4 chains in 4 jobs)
NUTS: [initial, g, eps]
```

100.00% [8000/8000

Sampling 4 chains, 0 divergences]

```
Sampling 4 chains for 1_000 tune and 1_000 draw iterations (4
s total) took 6 seconds.
```

Figure: Good sampling

Posterior Predictive

Another way to validate your model
is posterior predictive.

- How model performs vs
  observed data?
- Is there any missing effect to
  include?

Let's see

Posterior Predictive

Another way to validate your model
is posterior predictive.

- How model performs vs
  observed data?
- Is there any missing effect to
  include?

Let's see

1. The fit is good
2. Model does not reflect
   seasonalities
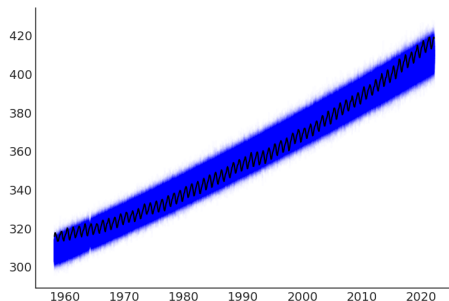3. We can fix it with a better
   model

Figure: Posterior Predictive Plot

Conclusions so far

- The reparametrized model is decent
- It does not account for seasonality
- We have the mean function for the next modelling step

$$g \sim \text{Normal}(1.00065, 0.0005)$$
$$\epsilon \sim \text{LogNormal}(-2, 0.5)$$
$$\log y_0 \sim \text{Normal}(\log 310, 10/310)$$
$$y_t \sim \text{LogNormal}(\log y_0 + g \cdot t, \epsilon)$$

Adding Seasonality

CO_2 data (Manua Loa, Hawaii)

- Monthly observations
- Yearly seasonality makes sense
- Trend function is clear from a simpler model

Figure: CO2 level

Adding Seasonality

CO2 data (Manua Loa, Hawaii)

- Monthly observations
- Yearly seasonality makes sense
- Trend function is clear from a simpler model
- Seasonality can be added with a periodic kernel

Figure: CO2 level

$$k(*, *) = \quad \text{Periodic}(p = \quad , l = \quad)$$

Adding Seasonality

CO2 data (Manua Loa, Hawaii)

- Monthly observations
- Yearly seasonality makes sense
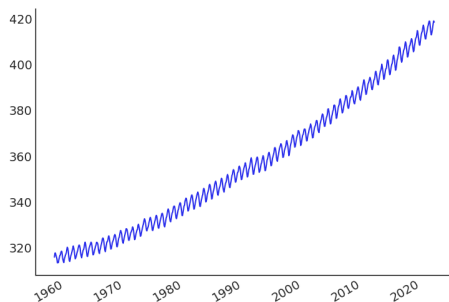- Trend function is clear from a simpler model
- Seasonality can be added with a periodic kernel
- Period is 12 months

Figure: CO2 level

$$k(*, *) = \quad \text{Periodic}(\mathsf{p} = 12, \mathsf{l} = \quad)$$

Adding Seasonality

CO2 data (Manua Loa, Hawaii)

- Monthly observations
- Yearly seasonality makes sense
- Trend function is clear from a simpler model
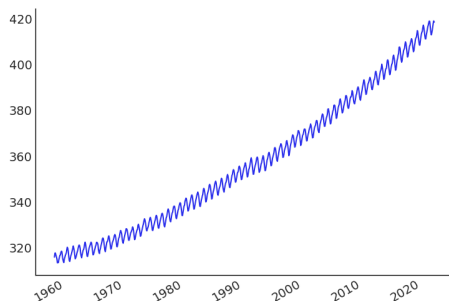- Seasonality can be added with a periodic kernel
- Period is 12 months
- Appropriate lengthscale is ≈ 3 months

Figure: CO2 level

$$k(*, *) = \quad \text{Periodic}(\mathsf{p} = 12, \mathsf{l} = 3)$$

Adding Seasonality

CO2 data (Manua Loa, Hawaii)

- Monthly observations
- Yearly seasonality makes sense
- Trend function is clear from a simpler model
- Seasonality can be added with a periodic kernel
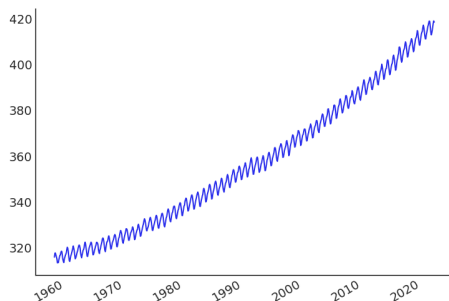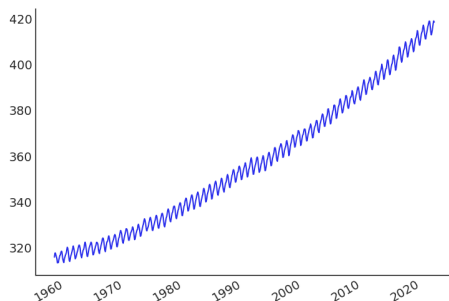- Period is 12 months
- Appropriate lengthscale is ≈ 3 months
- Do not forget to scale the covariance!
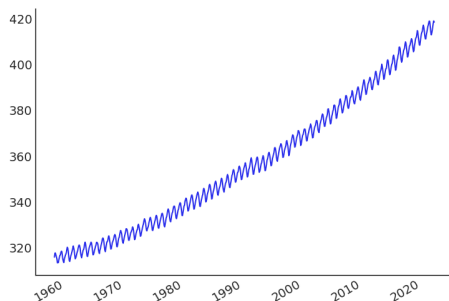
Figure: CO2 level

$$k(*, *) = \alpha_k \, \text{Periodic}(\mathsf{p} = 12, \mathsf{l} = 3)$$

Seasonality Change

$$K(*, *) = \alpha_k \, \text{Periodic}(\mathsf{p} = 12, \mathsf{l} = 3)$$

- Periodic kernel is a strong prior.
- Seasonality pattern is assumed constant (what if ...)

Seasonality Change

$$K(*, *) = \alpha_k \, \text{Periodic}(\mathsf{p} = 12, \mathsf{l} = 3)$$

- Periodic kernel is a strong prior.
- Seasonality pattern is assumed constant (what if ...)
- We can make seasonality patterns change (kernel math!)

Seasonality Change

$$K(*, *) = \alpha_k \, \text{Periodic}(\mathsf{p} = 12, \mathsf{l} = 3)$$

- Periodic kernel is a strong prior.
- Seasonality pattern is assumed constant (what if ...)
- We can make seasonality patterns change (kernel math!)
- We need a decay in correlation between distant year patterns

Seasonality Change

$$K(*, *) = \alpha_k \, \text{Periodic}(\mathsf{p} = 12, \mathsf{l} = 3) \cdot \text{Exponential}(\mathsf{l} = 1200)$$

- Periodic kernel is a strong prior.
- Seasonality pattern is assumed constant (what if ...)
- We can make seasonality patterns change (kernel math!)
- We need a decay in correlation between distant year patterns
- Multiplication with the Exponential kernel should work just fine

Seasonality Change

$$K(*, *) = \alpha_k \, \text{Periodic}(\text{p} = 12, \text{l} = 3) \cdot \text{Exponential}(\text{l} = 1200)$$

- Periodic kernel is a strong prior.
- Seasonality pattern is assumed constant (what if ...)
- We can make seasonality patterns change (kernel math!)
- We need a decay in correlation between distant year patterns
- Multiplication with the Exponential kernel should work just fine
- Lengthscale in the Exponential kernel is how strong patterns correlate across months
- I choose 1200, century long patterns
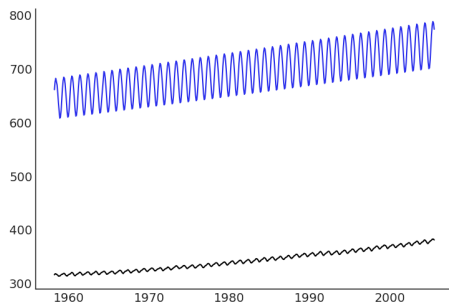
Decay Visualized
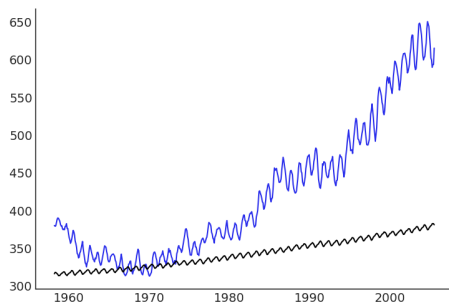

Figure: Mean without decay

Figure: Mean with decay

Decay model

No decay model is the limit case of decay model with lengthscale $\to \infty$

The Model

$$g \sim \text{Normal}(1.00065, 0.0005)$$
$$\epsilon \sim \text{LogNormal}(-2, 0.5)$$
$$\log y_0 \sim \text{Normal}(\log 310, 10/310)$$
$$\alpha_k \sim \text{LogNormal}(-2, 0.5)$$
$$K(*, *) = \alpha_k \text{ Periodic}(\text{p} = 12, \text{l} = 3) \cdot \text{Exponential}(\text{l} = 1200)$$
$$\log \bar{y}_t = \mathcal{GP}(\log y_0 + g \cdot t, K)$$
$$y_t \sim \text{LogNormal}(\log \bar{y}_t, \epsilon)$$

Improve the model

1. Figure out a better ϵ prior
2. Infer better lengthscales (hyperpriors)
3. Infer structural medium term shift (additive kernel)
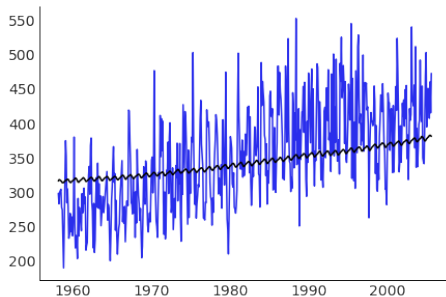
Prior Predictive Check

Figure: $\epsilon \sim \text{LogNormal}(-2, 0.5)$
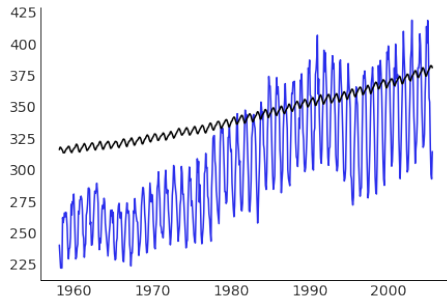
Figure: $\epsilon \sim \text{LogNormal}(-4, 0.5)$

Inference

- `pm.sample()` is slow
- GPU sampling is also slow
- `pm.find_MAP()` makes sense and fast
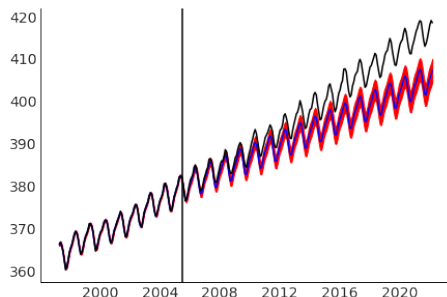
Figure: Predictions with MAP

Key takeaways

- Look at prior predictive to validate initial model
- Interpretable building blocks create interpretable model
- Sasonality model is done with Periodic kernel
- Seasonality might be flexible

Stochastic Volatility Model
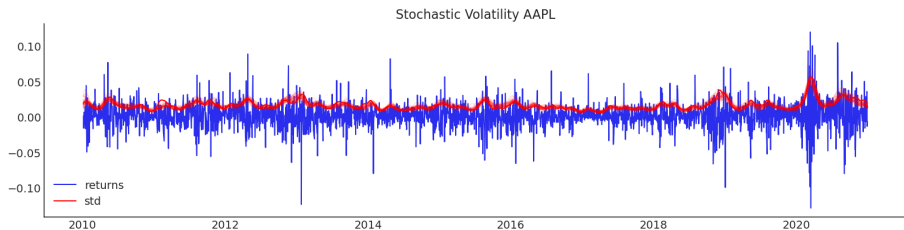
Figure: Stochastic Volatility Estimation for AAPL

Read More

https://github.com/quantopian/bayesalpha

Key Ideas

- Returns are constant
- Volatility is not constant
- Model volatility as a Gaussian process
- Use approximations to speed up the model
- Use GPU to do fast inference

Priors

- Returns

- Volatility

- Time Component

Priors

- Returns
  - Expect year return at orders $\pm 100\%$

- Volatility

- Time Component

Priors

- Returns
  - Expect year return at orders $\pm 100\%$
  - Daily return $\pm(2^{\frac{1}{250}} - 1)$

- Volatility

- Time Component

Priors

- Returns
  - Expect year return at orders $\pm 100\%$
  - Daily return $\pm(2^{\frac{1}{250}} - 1)$
  - The prior is insane but yet motivated, name your prior!
- Volatility

- Time Component

Priors

- Returns
  - Expect year return at orders $\pm 100\%$
  - Daily return $\pm(2^{\frac{1}{250}} - 1)$
  - The prior is insane but yet motivated, name your prior!
- Volatility
  - Expect signal to noise $\frac{\text{mean}}{\text{std}} \approx 1$

- Time Component

Priors

- Returns
  - Expect year return at orders $\pm 100\%$
  - Daily return $\pm(2^{\frac{1}{250}} - 1)$
  - The prior is insane but yet motivated, name your prior!
- Volatility
  - Expect signal to noise $\frac{\text{mean}}{\text{std}} \approx 1$
  - $\log \text{std} = \log(2^{\frac{1}{250}} - 1)$
- Time Component

Priors

- Returns
  - Expect year return at orders $\pm 100\%$
  - Daily return $\pm(2^{\frac{1}{250}} - 1)$
  - The prior is insane but yet motivated, name your prior!
- Volatility
  - Expect signal to noise $\frac{\text{mean}}{\text{std}} \approx 1$
  - $\log \text{std} = \log(2^{\frac{1}{250}} - 1)$
- Time Component
  - After the model is framed

The Model

I did the modelling choice to make it as simple as possible

$$\text{return} \sim \text{Normal}(0, 2^{\frac{1}{250}} - 1)$$

$$\text{log std} \sim \text{Normal}(\log(2^{\frac{1}{250}} - 1), 0.05)$$

$$\text{ls} \sim \text{Gamma}(30, 5)$$

$$\alpha_{\text{vol}} \sim \text{Exponential}(100)$$

$$K(*, *) = \alpha_{\text{vol}} \, \text{Matern32}(\text{ls})$$

$$\Delta_t^{\text{log std}} \sim \mathcal{GP}(0, K)$$

$$\text{obs}_t \sim \text{Normal}(\text{return}, \exp(\text{log std} + \Delta_t^{\text{log std}}))$$

The Model

I did the modelling choice to make it as simple as possible

$$\text{return} \sim \text{Normal}(0, 2^{\frac{1}{250}} - 1)$$

$$\text{log std} \sim \text{Normal}(\log(2^{\frac{1}{250}} - 1), 0.05)$$

$$\text{ls} \sim \text{Gamma}(30, 5)$$

$$\alpha_{\text{vol}} \sim \text{Exponential}(100)$$

$$K(*, *) = \alpha_{\text{vol}} \, \text{Matern32}(\text{ls})$$

$$\Delta_t^{\text{log std}} \sim \mathcal{GP}(0, K)$$

$$\text{obs}_t \sim \text{Normal}(\text{return}, \exp(\text{log std} + \Delta_t^{\text{log std}}))$$

The Model

I did the modelling choice to make it as simple as possible

$$\text{return} \sim \text{Normal}(0, 2^{\frac{1}{250}} - 1)$$

$$\text{log std} \sim \text{Normal}(\log(2^{\frac{1}{250}} - 1), 0.05)$$

$$\text{ls} \sim \text{Gamma}(30, 5)$$

$$\alpha_{\text{vol}} \sim \text{Exponential}(100)$$

$$K(*, *) = \alpha_{\text{vol}} \text{Matern32(ls)}$$

$$\Delta_t^{\text{log std}} \sim \mathcal{GP}(0, K)$$

$$\text{obs}_t \sim \text{Normal}(\text{return}, \exp(\text{log std} + \Delta_t^{\text{log std}}))$$

The Model

I did the modelling choice to make it as simple as possible

$$\text{return} \sim \text{Normal}(0, 2^{\frac{1}{250}} - 1)$$

$$\log \text{std} \sim \text{Normal}(\log(2^{\frac{1}{250}} - 1), 0.05)$$

$$\text{ls} \sim \text{Gamma}(30, 5)$$

$$\alpha_{\text{vol}} \sim \text{Exponential}(100)$$

$$K(*, *) = \alpha_{\text{vol}} \, \text{Matern32}(\text{ls})$$

$$\Delta_t^{\log \text{std}} \sim \mathcal{GP}(0, K)$$

$$\text{obs}_t \sim \text{Normal}(\text{return}, \exp(\log \text{std} + \Delta_t^{\log \text{std}}))$$

The Model

I did the modelling choice to make it as simple as possible

$$\text{return} \sim \text{Normal}(0, 2^{\frac{1}{250}} - 1)$$

$$\log \text{std} \sim \text{Normal}(\log(2^{\frac{1}{250}} - 1), 0.05)$$

$$\text{ls} \sim \text{Gamma}(30, 5)$$

$$\alpha_{\text{vol}} \sim \text{Exponential}(100)$$

$$K(*, *) = \alpha_{\text{vol}} \, \text{Matern32}(\text{ls})$$

$$\Delta_t^{\log \text{std}} \sim \mathcal{GP}(0, K)$$

$$\text{obs}_t \sim \text{Normal}(\text{return}, \exp(\log \text{std} + \Delta_t^{\log \text{std}}))$$

The Model

I did the modelling choice to make it as simple as possible

$$\text{return} \sim \text{Normal}(0, 2^{\frac{1}{250}} - 1)$$

$$\log\text{std} \sim \text{Normal}(\log(2^{\frac{1}{250}} - 1), 0.05)$$

$$\text{ls} \sim \text{Gamma}(30, 5)$$

$$\alpha_{\text{vol}} \sim \text{Exponential}(100)$$

$$K(*, *) = \alpha_{\text{vol}} \, \text{Matern32}(\text{ls})$$

$$\Delta_t^{\log\text{std}} \sim \mathcal{GP}(0, K)$$

$$\text{obs}_t \sim \text{Normal}(\text{return}, \exp(\log\text{std} + \Delta_t^{\log\text{std}}))$$

The Model

I did the modelling choice to make it as simple as possible

$$\text{return} \sim \text{Normal}(0, 2^{\frac{1}{250}} - 1)$$

$$\log \text{std} \sim \text{Normal}(\log(2^{\frac{1}{250}} - 1), 0.05)$$

$$\text{ls} \sim \text{Gamma}(30, 5)$$

$$\alpha_{\text{vol}} \sim \text{Exponential}(100)$$

$$K(*, *) = \alpha_{\text{vol}} \, \text{Matern32}(\text{ls})$$

$$\color{red}{\Delta_t^{\log \text{std}} \sim \mathcal{GP}(0, K)}$$

$$\text{obs}_t \sim \text{Normal}(\text{return}, \exp(\log \text{std} + \Delta_t^{\log \text{std}}))$$

The Model

I did the modelling choice to make it as simple as possible

$$\text{return} \sim \text{Normal}(0, 2^{\frac{1}{250}} - 1)$$

$$\log \text{std} \sim \text{Normal}(\log(2^{\frac{1}{250}} - 1), 0.05)$$

$$\text{ls} \sim \text{Gamma}(30, 5)$$

$$\alpha_{\text{vol}} \sim \text{Exponential}(100)$$

$$K(*, *) = \alpha_{\text{vol}} \, \text{Matern32}(\text{ls})$$

$$\Delta_t^{\log \text{std}} \sim \mathcal{GP}(0, K)$$

$$\text{obs}_t \sim \text{Normal}(\text{return}, \exp(\log \text{std} + \Delta_t^{\log \text{std}}))$$

Splines

Make the model faster!

Spline interpolation
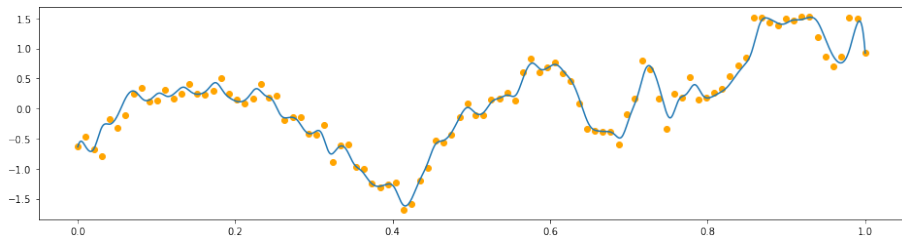
`pmx.utils.spline.bspline_interpolation`

Figure: B-Spline interpolation

$$y' = B(x, x', d)y,$$

where $\{x, y\}$ original function, d - degree,
x' - new domain, y' interpolated function

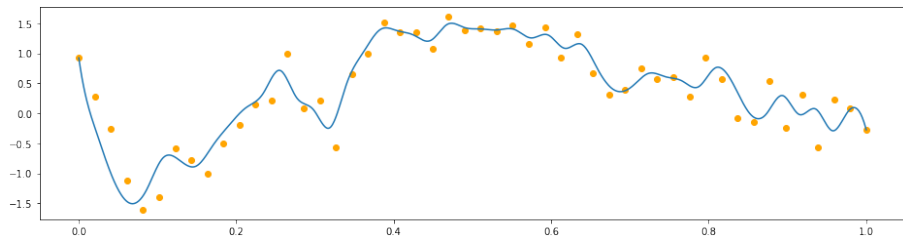# Spline interpolation

`pmx.utils.spline.bspline_interpolation`

Figure: B-Spline interpolation (fewer points)

$$y' = B(x, x', d)y,$$

where $\{x, y\}$ original function, d - degree,
x' - new domain, y' interpolated function

Why B-Spline

$$y' = B(x, x', d)y,$$

1. x, x' are usually given
2. y - is random (Gaussian Process)
3. y' - is a smooth interpolation
4. $B(x, x', d)$ is static
5. $B(x, x', d)$ is sparse

To remember
1. Bad extrapolation when degree $d > 2$

Spline Interpolation

- $d = 1$ - linear interpolation
- $d = 2$ - quadratic interpolation
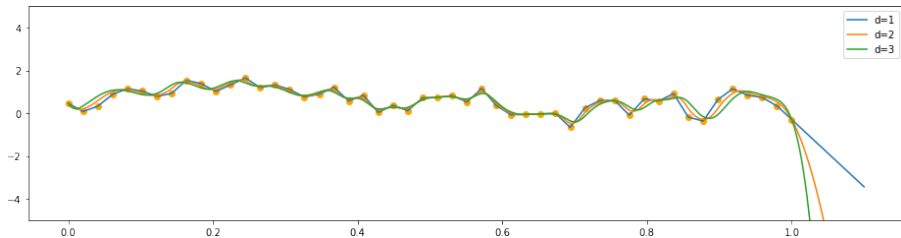- $d = 3$ - cubic interpolation

Figure: Spline interpolation

Splines for Stochastic volatility

$$\text{return} \sim \text{Normal}(0, 2^{\frac{1}{250}} - 1)$$

$$\text{log std} \sim \text{Normal}(2^{\frac{1}{250}} - 1, 0.05)$$

$$\text{ls} \sim \text{Gamma}(30, 5)$$

$$\alpha_{\text{vol}} \sim \text{Exponential}(100)$$

$$K(*, *) = \alpha_{\text{vol}} \text{ Matern32(ls)}$$

$$\bar{\Delta}_{t'}^{\text{log std}} \sim \mathcal{GP}(0, K)$$

$$\Delta_t^{\text{log std}} = \text{B-Spline}(\bar{\Delta}_{t'}^{\text{log std}}, \text{d} = 3)$$

$$\text{obs}_t \sim \text{Normal}(\text{return}, \exp(\text{log std} + \Delta_t^{\text{log std}}))$$

Splines for Stochastic volatility

$$\text{return} \sim \text{Normal}(0, 2^{\frac{1}{250}} - 1)$$

$$\text{log std} \sim \text{Normal}(2^{\frac{1}{250}} - 1, 0.05)$$

$$\text{ls} \sim \text{Gamma}(30, 5)$$

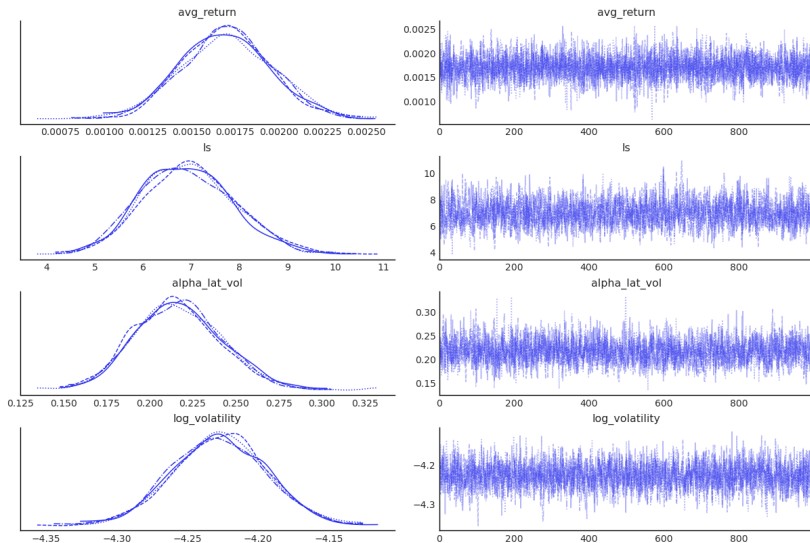$$\alpha_{\text{vol}} \sim \text{Exponential}(100)$$

$$K(*, *) = \alpha_{\text{vol}} \, \text{Matern32}(\text{ls})$$

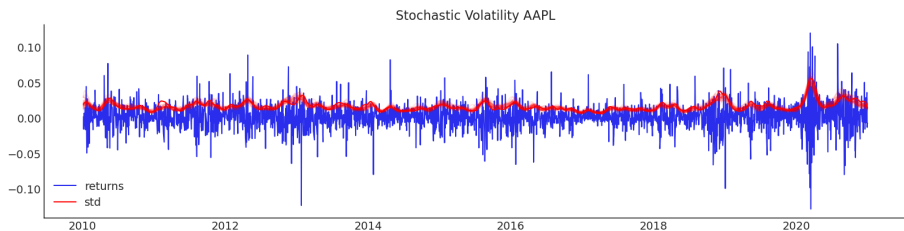$$\bar{\Delta}_{t'}^{\text{log std}} \sim \mathcal{GP}(0, K)$$

$$\Delta_t^{\text{log std}} = \text{B-Spline}(\bar{\Delta}_{t'}^{\text{log std}}, \text{d} = 3)$$

$$\text{obs}_t \sim \text{Normal}(\text{return}, \exp(\text{log std} + \Delta_t^{\text{log std}}))$$

Results

Takeaways

Stochastic Volatility AAPL

- Interpretable parameters
- Splines make model much faster
- Uncertainties for
  - average volatility
  - average returns
  - stochastic variation (α_{vol})

References I

📄 T. SJ and L. B.
Forecasting at scale.
2017.