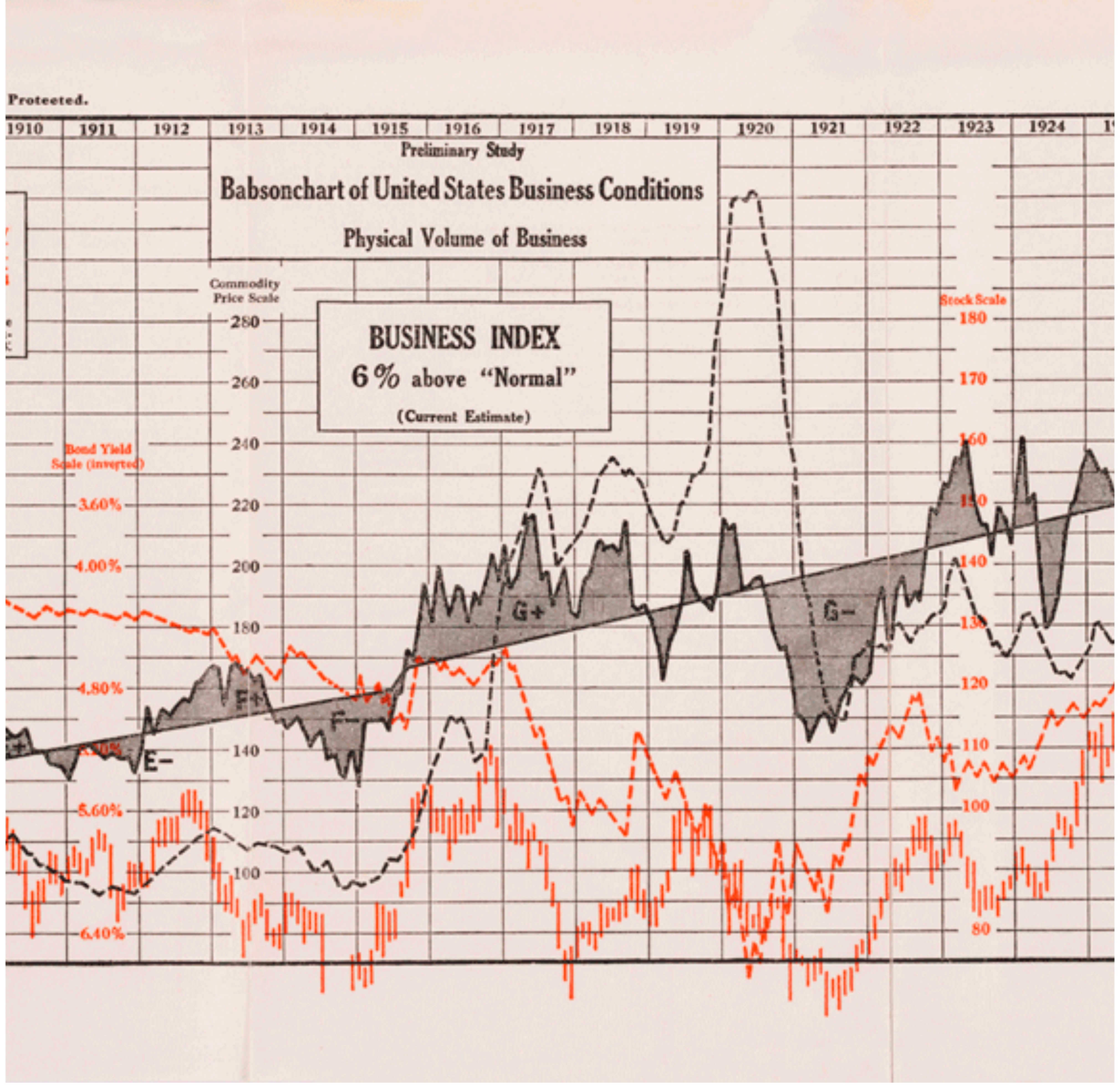
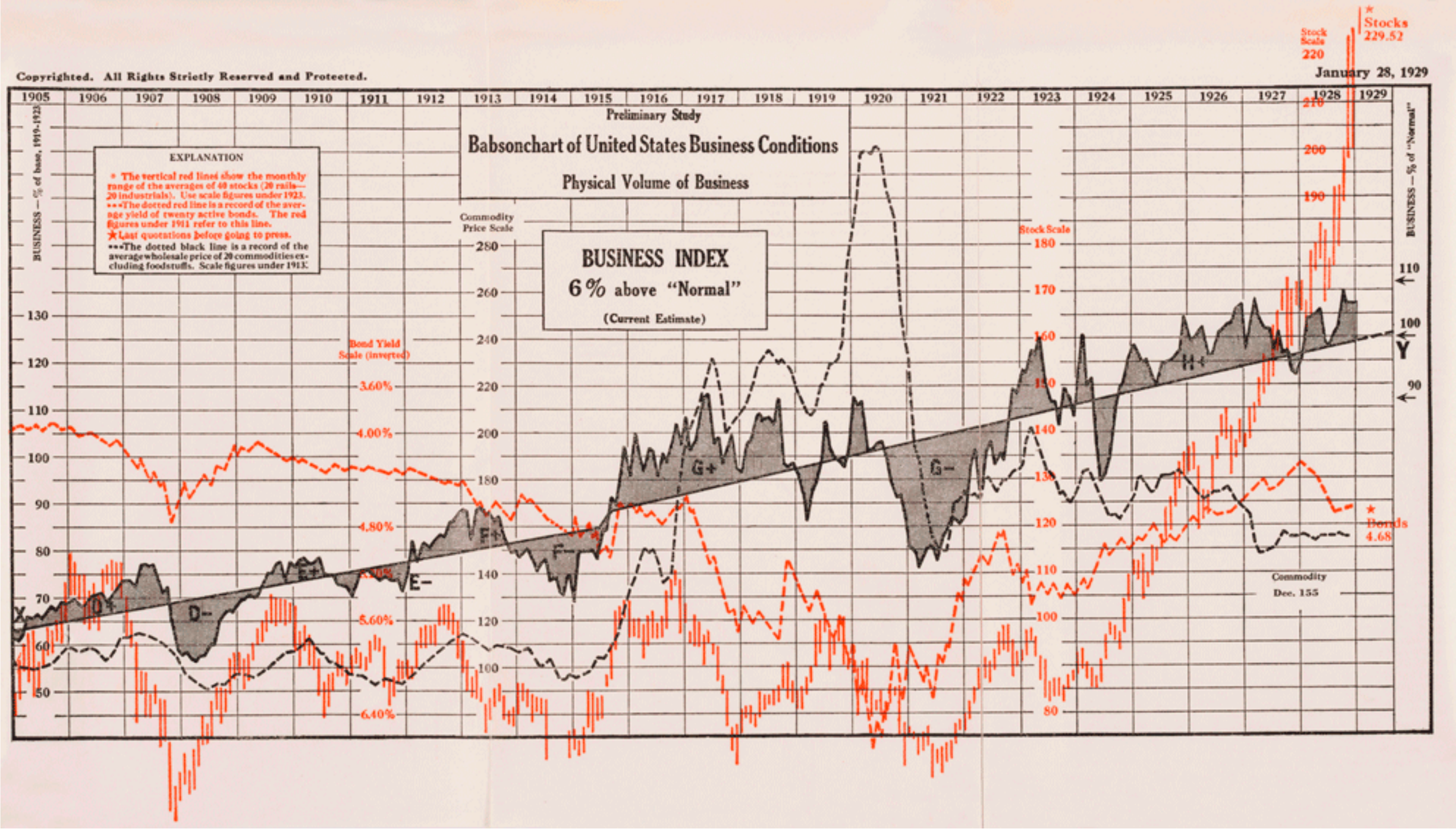
Prophet
Forecasting at Scale

Sean J. Taylor (Lyft)
Ben Letham (Facebook)

Background

- We have many applications that require forecasts.
- Often even a single metric must be forecast numerous times (e.g. for each country)
- Not many people have forecasting training or experience.
- Not many existing solutions or tools.




Copyrighted. All Rights Strictly Reserved and Protected.
January 28, 1929
Preliminary Study
Babsonchart of United States Business Conditions
Physical Volume of Business
BUSINESS INDEX
6 % above "Normal"
(Current Estimate)
EXPLANATION
* The vertical red lines show the monthly range of the averages of 40 stocks (20 rails—20 industrials). Use scale figures under 1923.
The dotted red line is a record of the average yield of twenty active bonds. The red figures under 1911 refer to this line.
★ Last quotations before going to press.
The dotted black line is a record of the average wholesale price of 20 commodities excluding foodstuffs. Scale figures under 1913.
1905 1906 1907 1908 1909 1910 1911 1912 1913 1914 1915 1916 1917 1918 1919 1920 1921 1922 1923 1924 1925 1926 1927 1928 1929
Commodity Price Scale
Bond Yield Scale (inverted)
Stock Scale
★ Stocks 229.52
★ Bonds 4.68
Commodity Dec. 155
BUSINESS — % of "Normal"

Many applications

Capacity planning

- How many servers, employees, meals, parking spaces, etc., are we going to need?

Goal setting

- How much would a metric grow by next year if we did nothing at all?

Anomaly detection

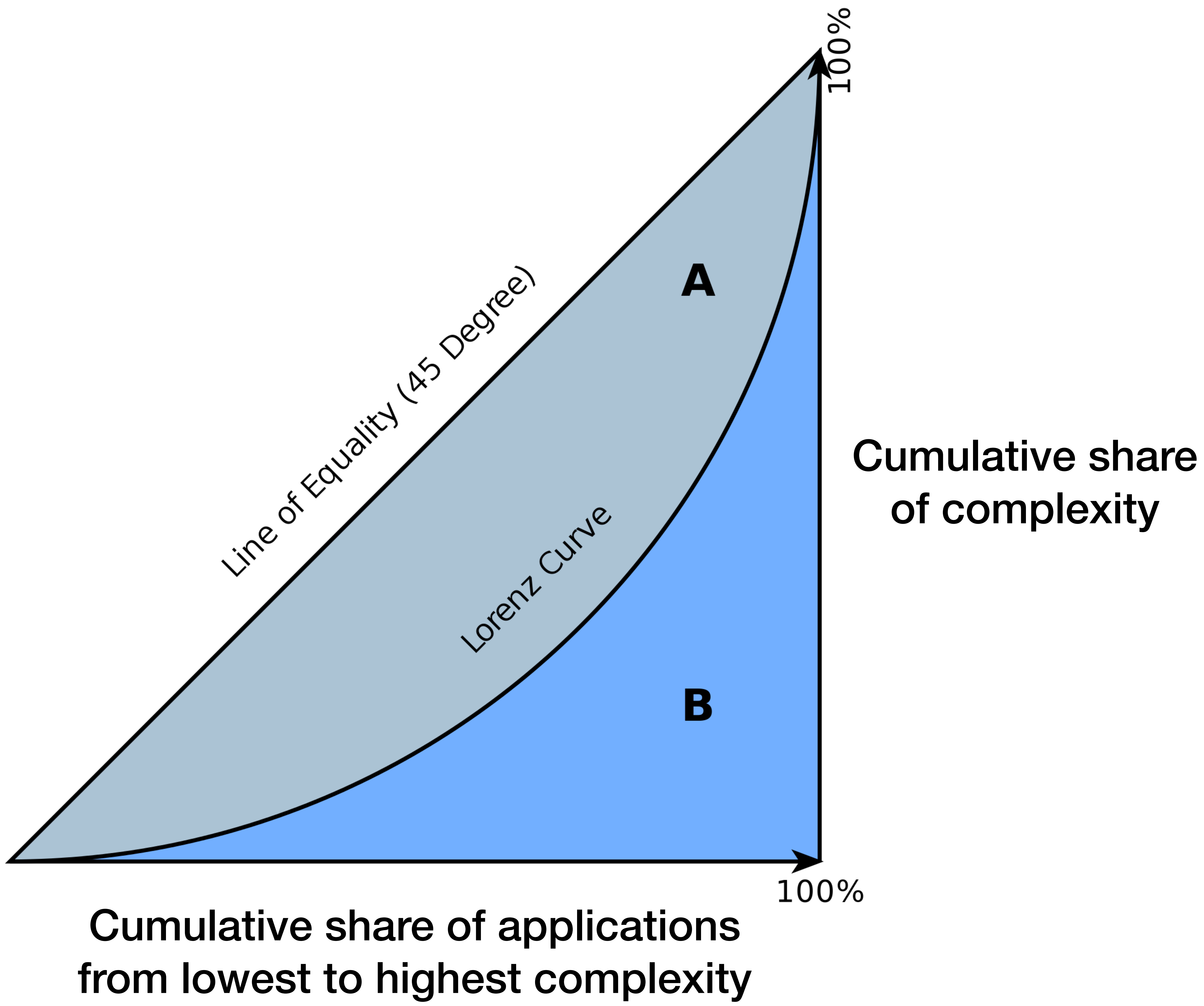
- Is this spike in bug reports due to some actual problem or because it's a holiday in Brazil?

Stuff we haven't thought of yet

- Forecasts can become components in complex data pipelines.

Pareto principle for forecasting

- Many business applications can be well handled by a relatively small class of curves.
- No need to cover complex forecasting problems which can benefit from most advanced approaches (e.g. LSTMs).
- **Scale to more applications** by making forecasting quick, simple, and repeatable for human analysts.
- **Scale to more users** by making the tool easy to use for beginners with a path to improve models for experts.



Prophet

semi automate forecasting

- find similarities across forecasting problems
- build a tool that can solve *most* of them
- make it easy to use + teach everyone to use it
- give a path forward to improving forecasts

Implementation

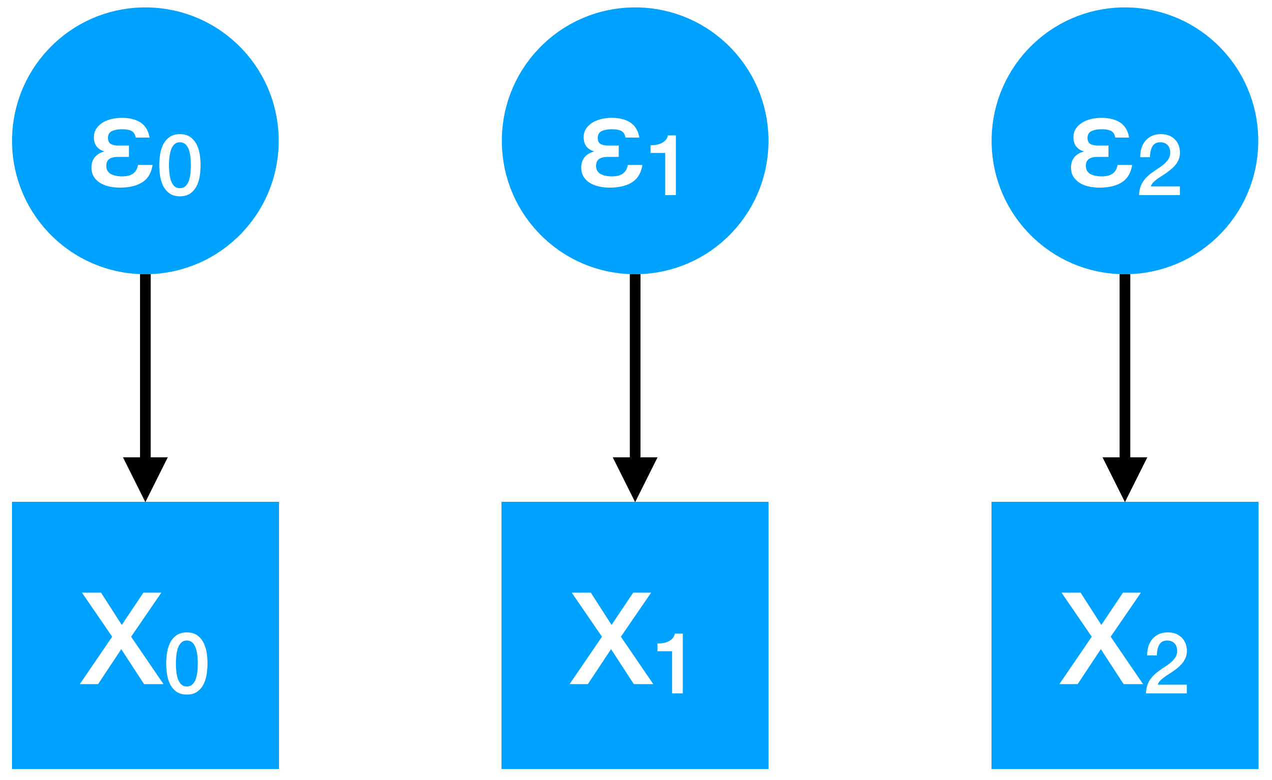
- Python and R packages
  - CRAN: prophet
  - PyPI: fbprophet
- Core procedure implemented in Stan (a probabilistic programming language).
- Version 0.1 released Feb 2017
- Version 0.5 released May 2019
- >8000 Github stars

Python API

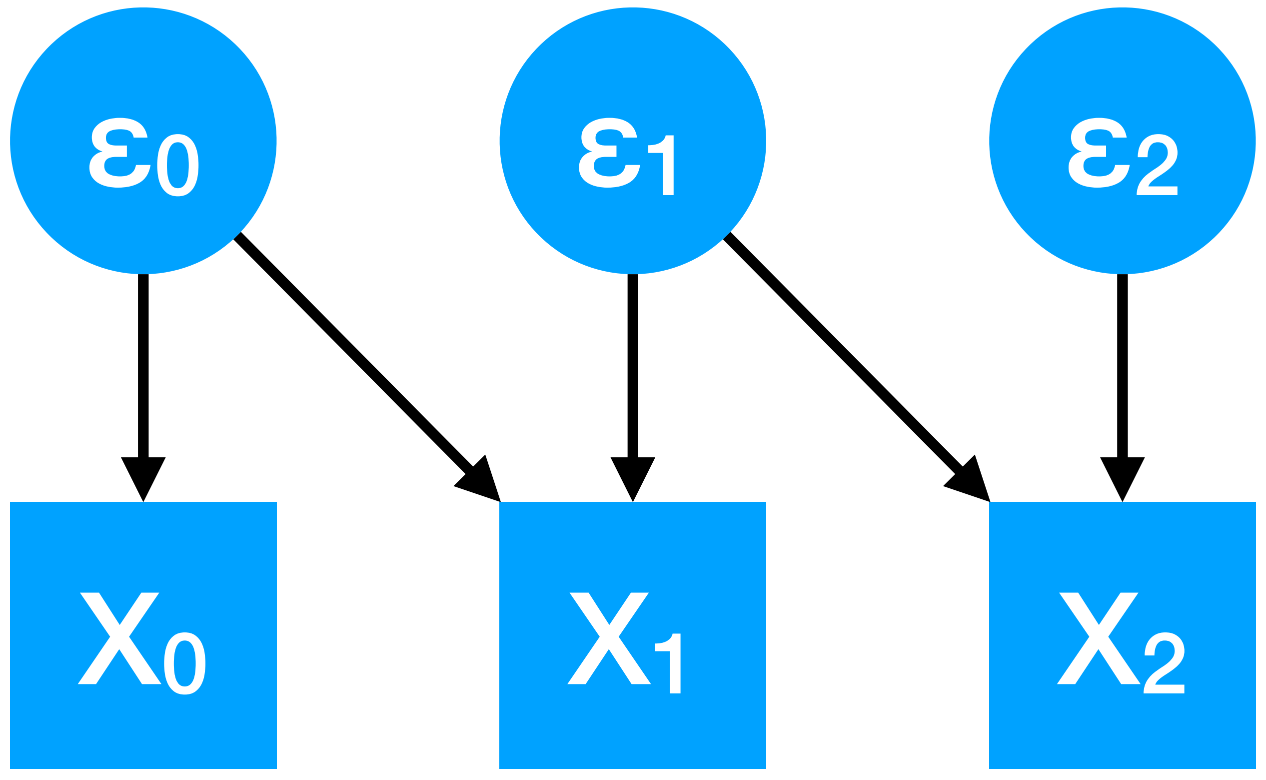
```
>>> from fbprophet import Prophet

>>> m = Prophet()

>>> m.fit(data)

>>> future =
m.make_future_dataframe(periods=365)

>>> forecast = m.predict(future)
```

Review of
time series methods

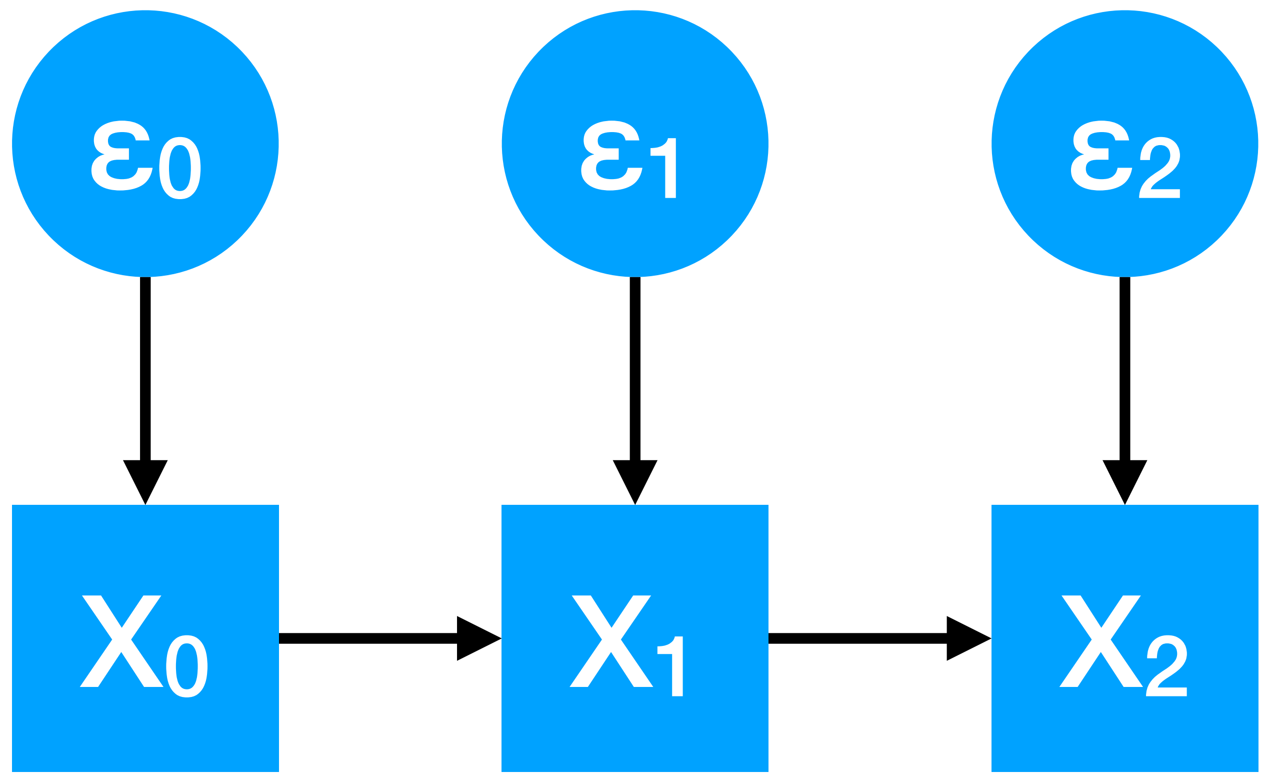
AR and MA models



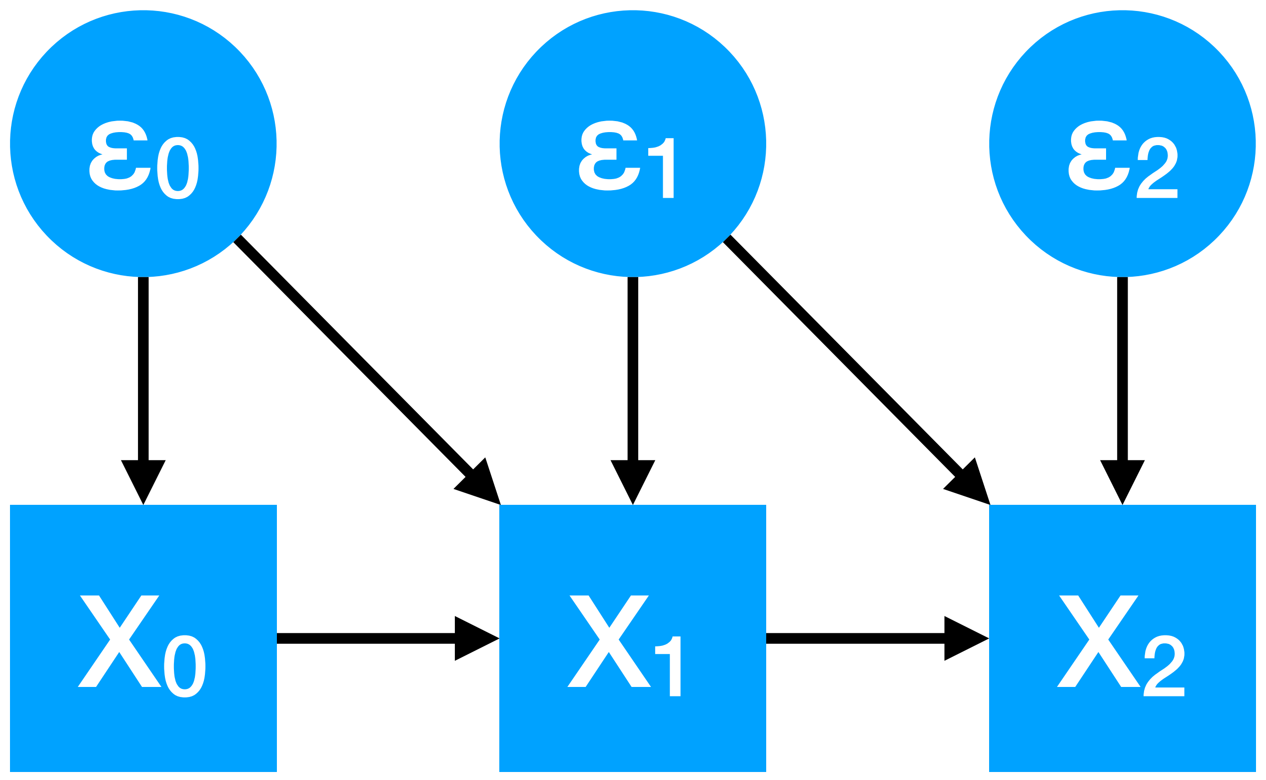
White noise



MA(1)

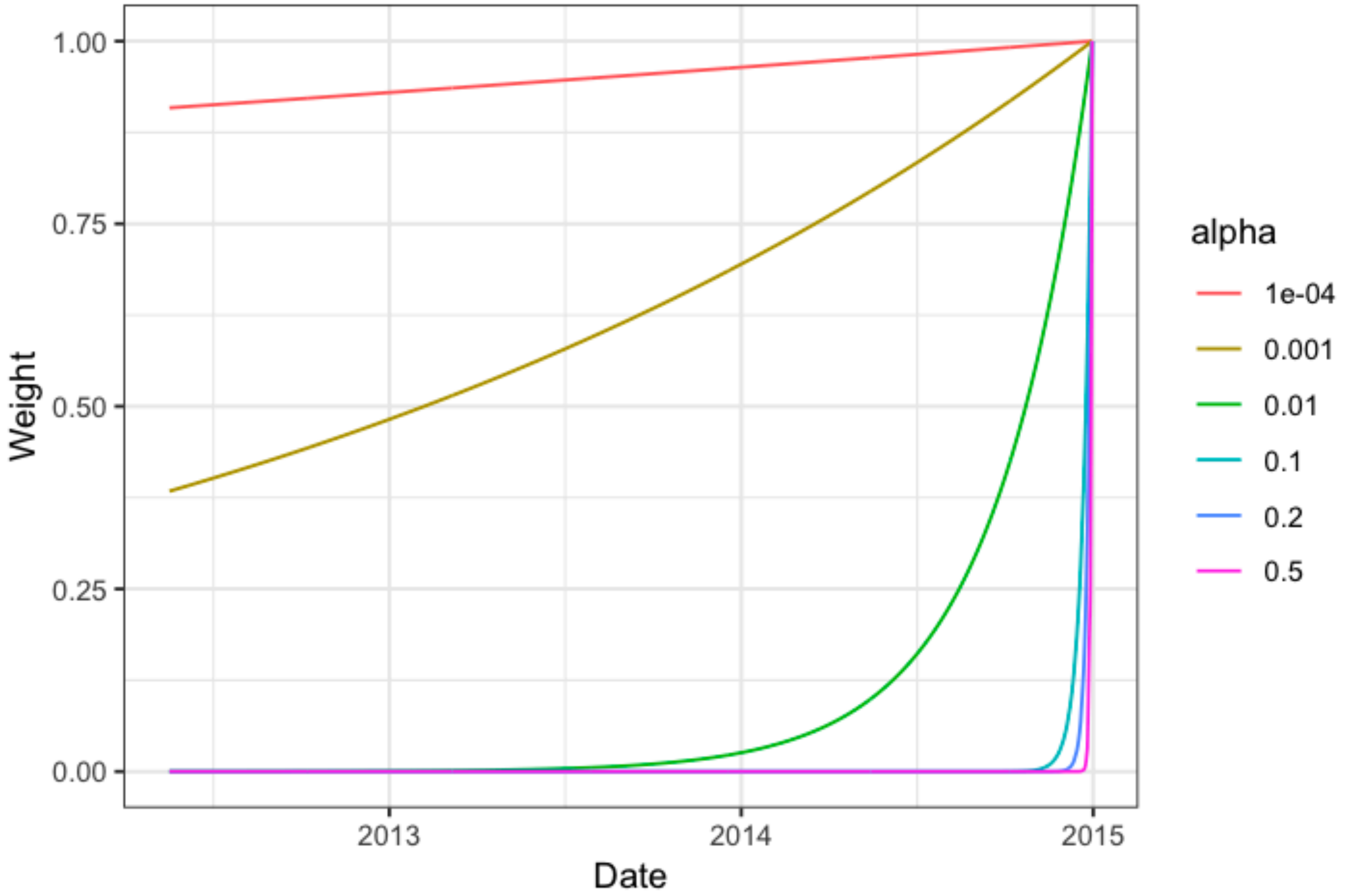


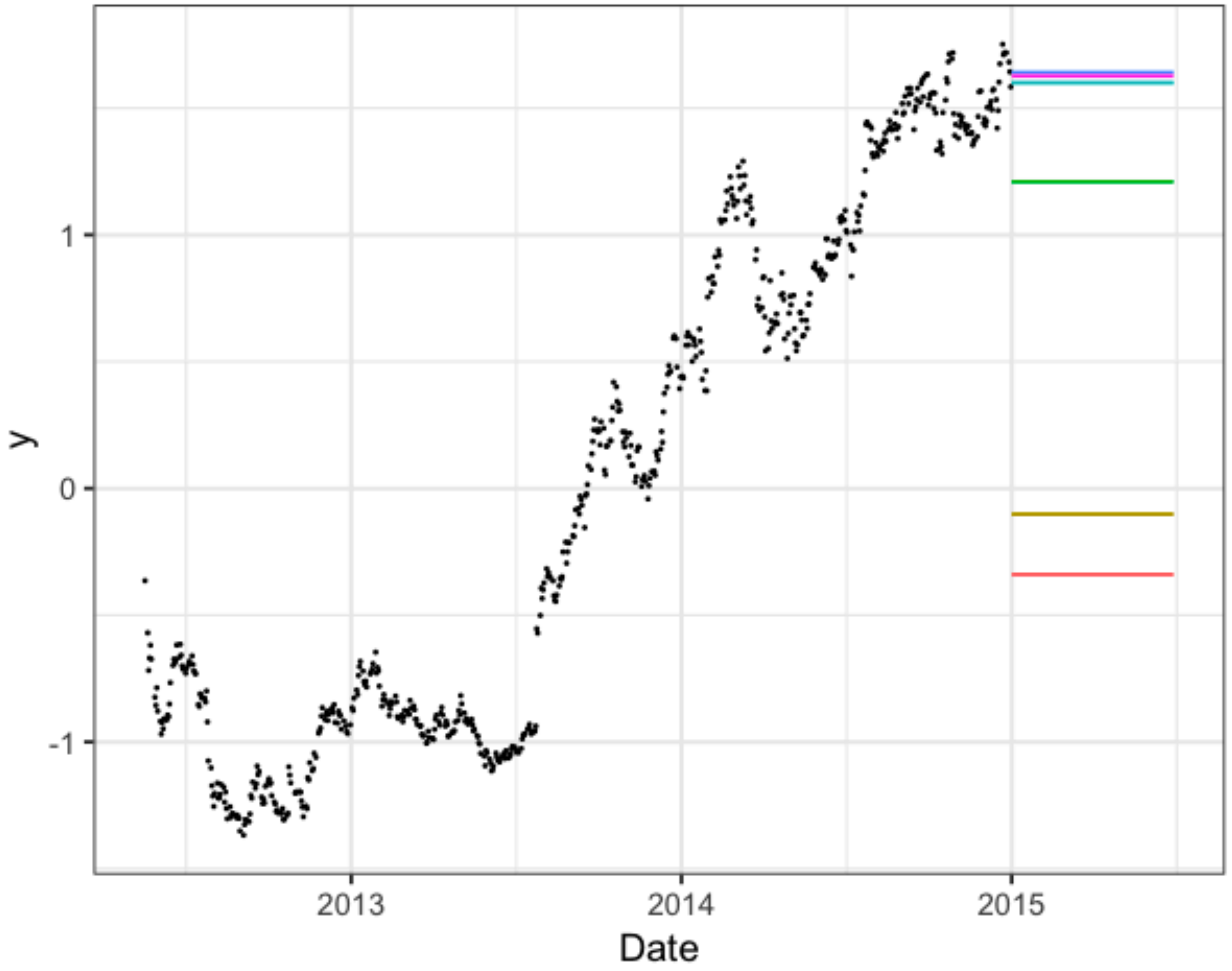
AR(1)



ARMA(1,1)

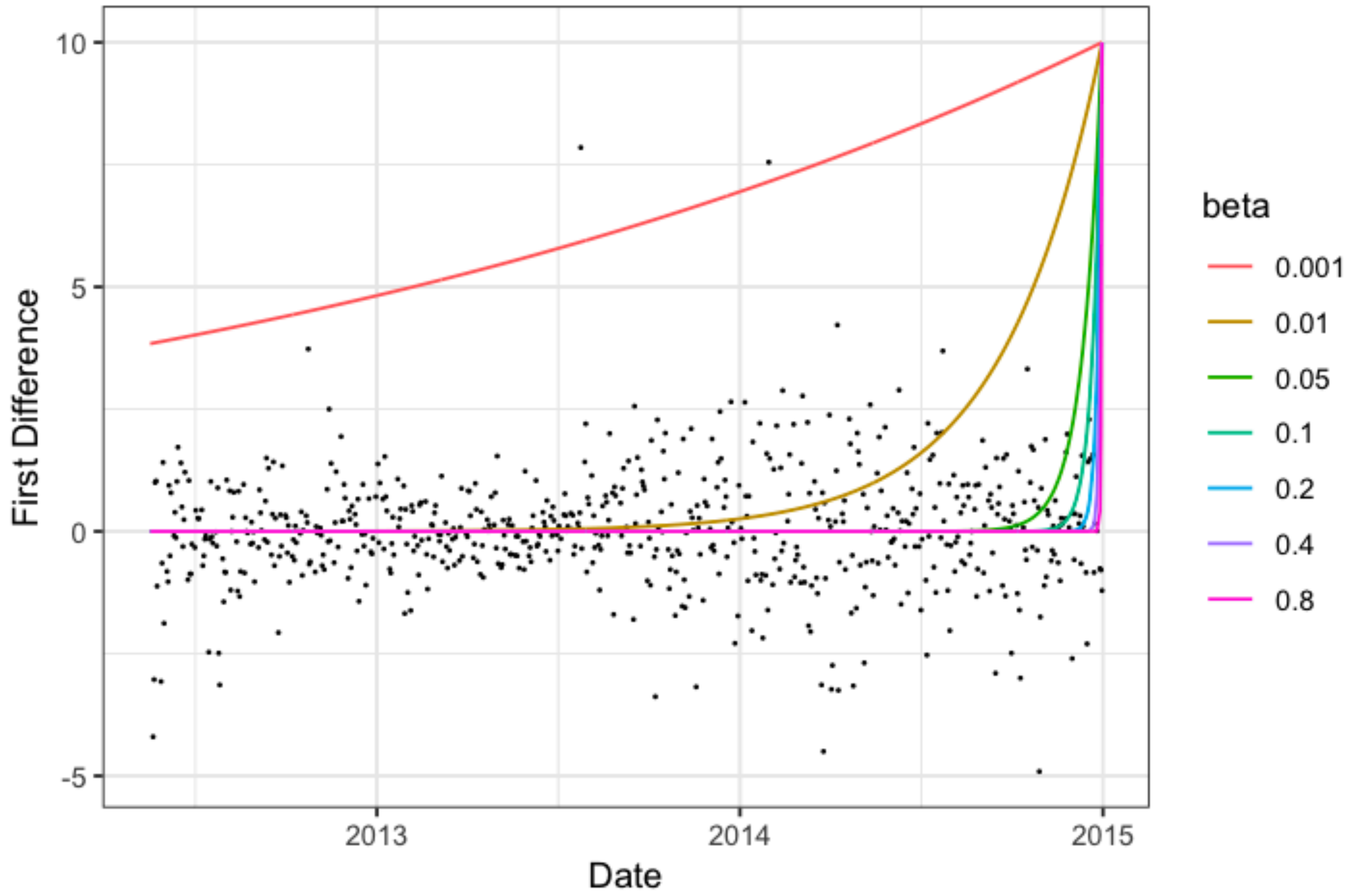
Exponential smoothing

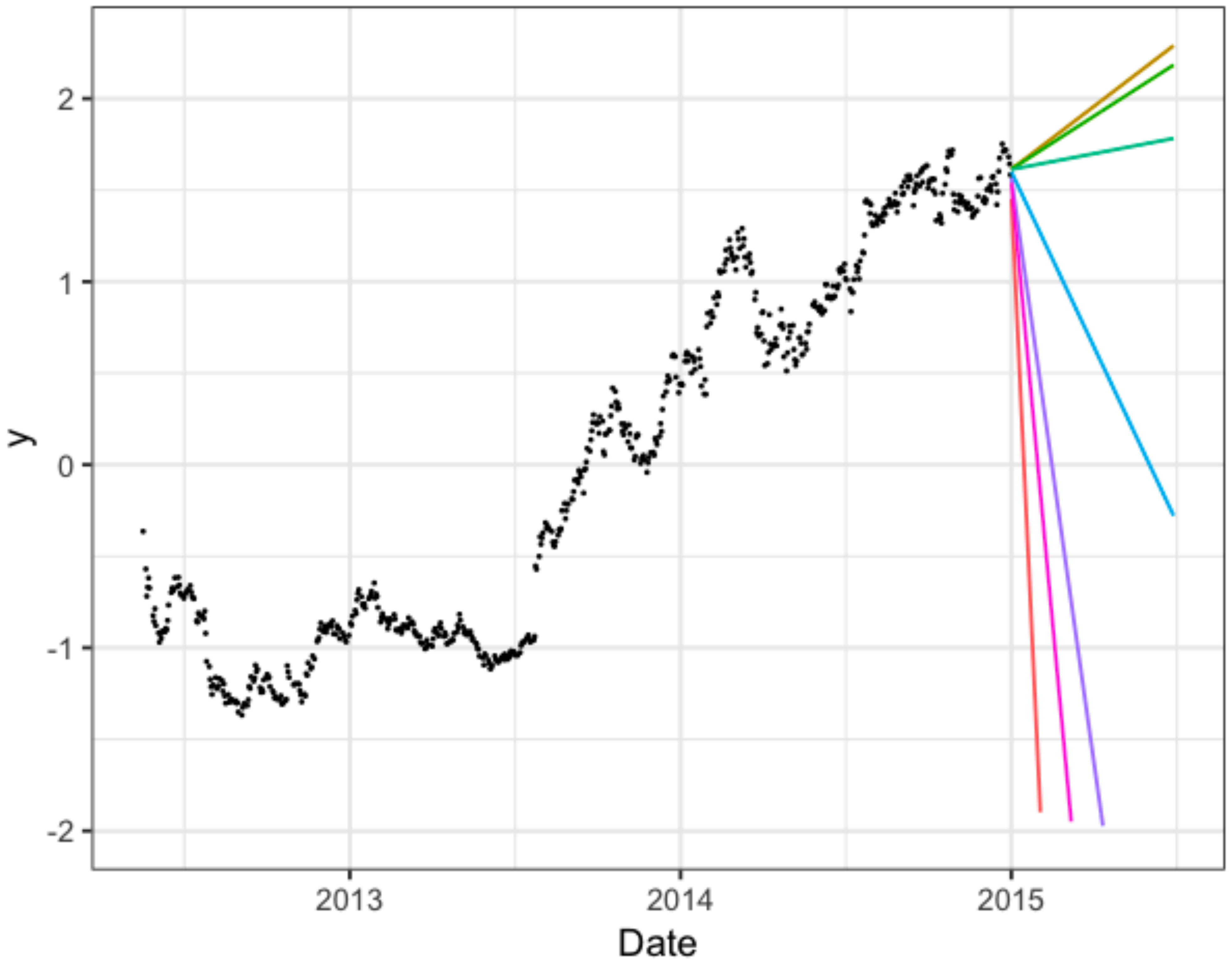




$$S_t = \alpha X_t + (1 - \alpha) S_{t-1}$$

Double exponential smoothing





$$S_t = \alpha X_t + (1-\alpha)(S_{t-1} + B_{t-1})$$

$$B_t = \beta(S_t - S_{t-1}) + (1-\beta)B_{t-1}$$

Business time series features

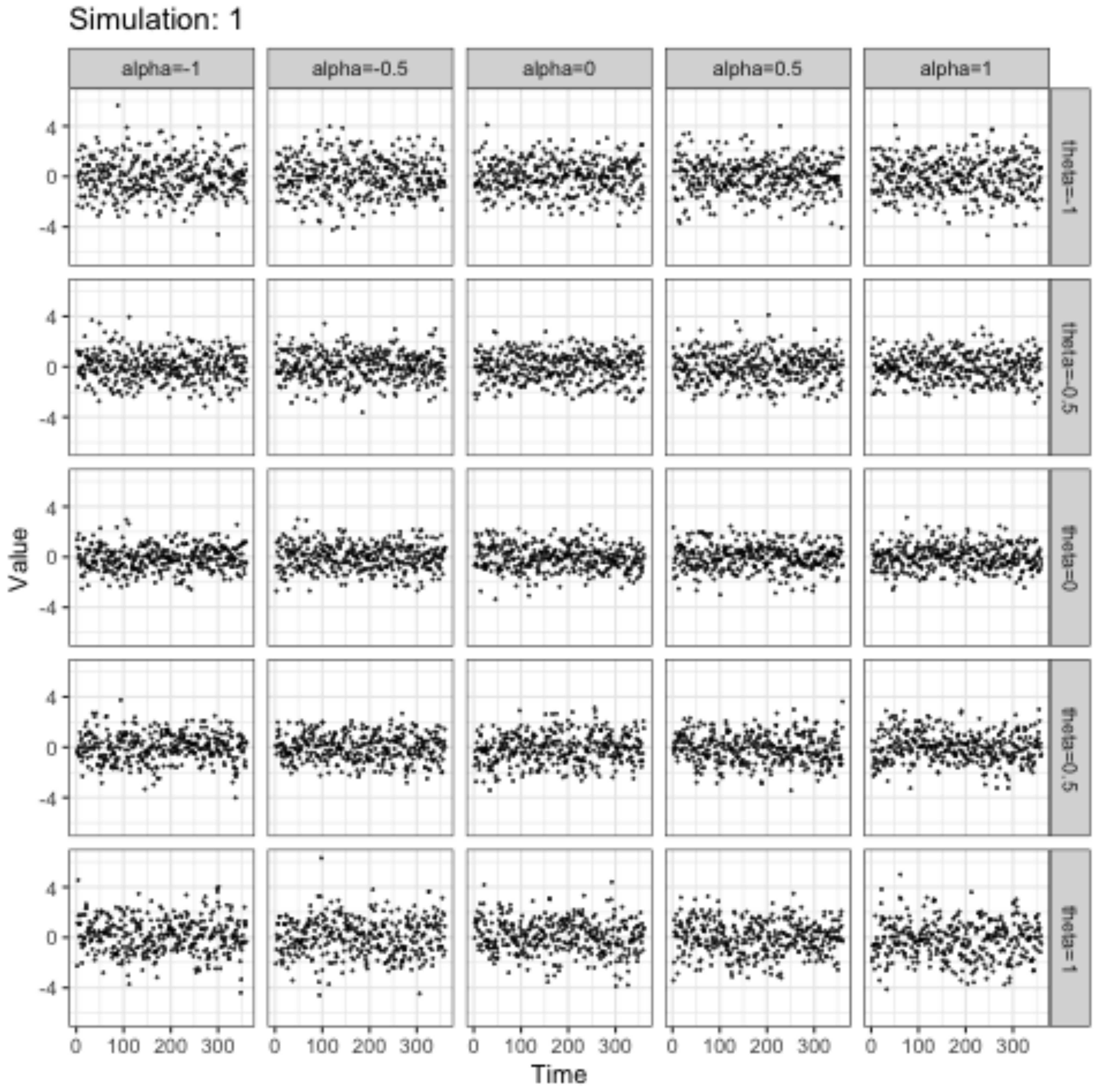
- outliers
- multiple seasonalities
- changes in trends
- abrupt changes

Sequence models

ARMA(p,q)

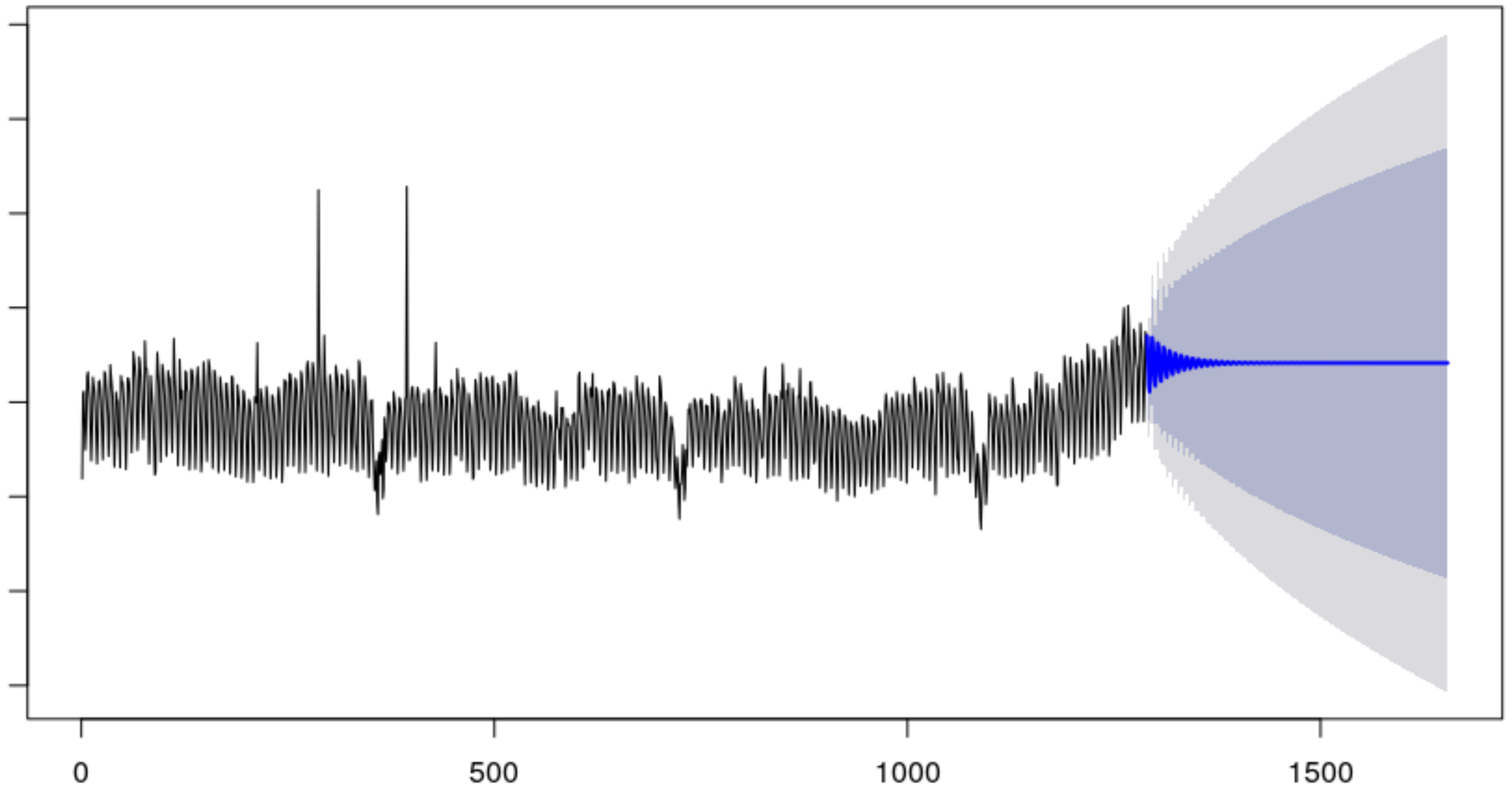
$$X_t = \sum_{i=1}^{p} \alpha_i X_{t-i} + \sum_{i=1}^{q} \theta_q \epsilon_{t-q} + \epsilon_t$$

- **Problem**: parameters don't correspond to any human-interpretable properties of the time series.

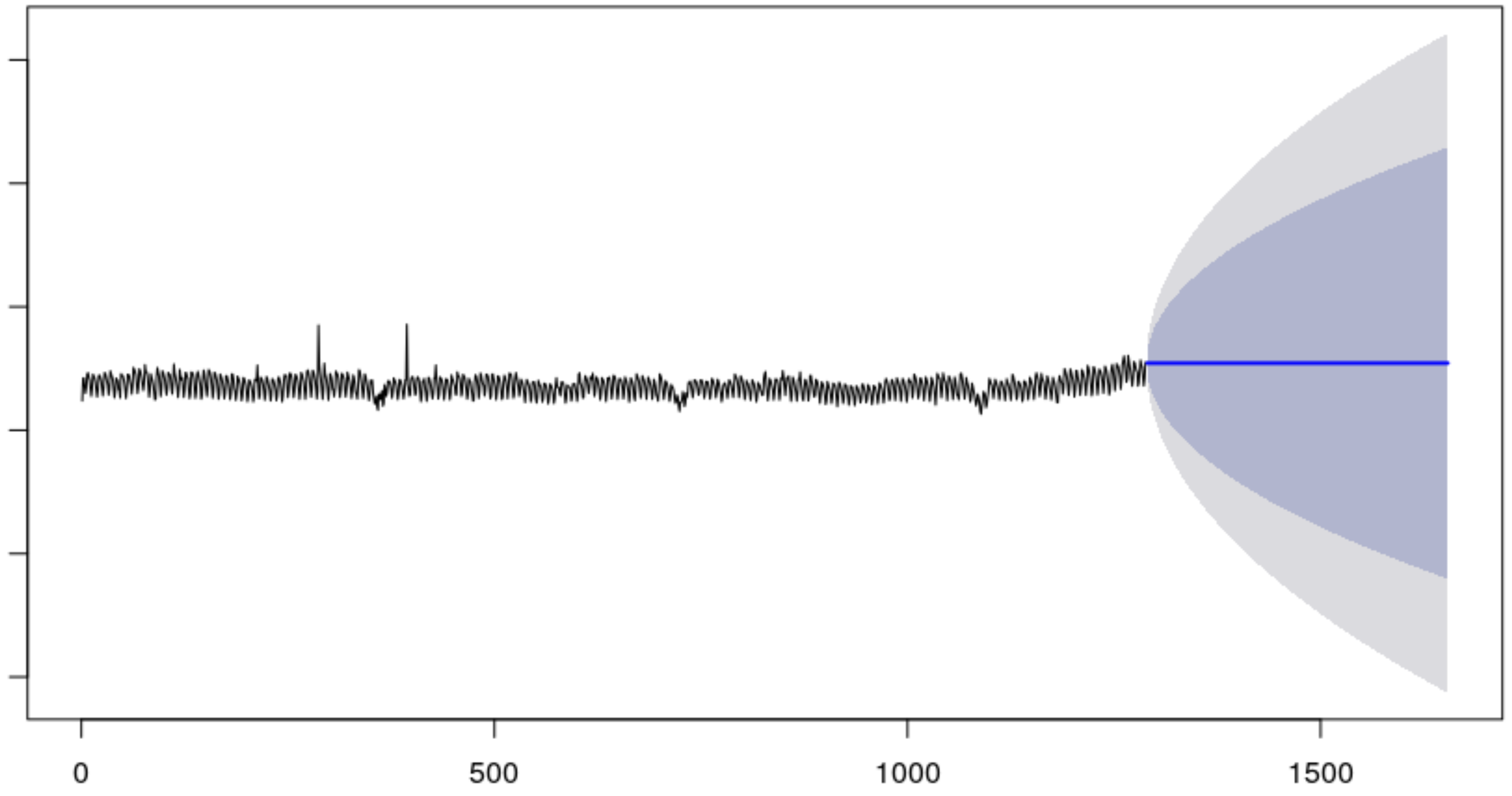


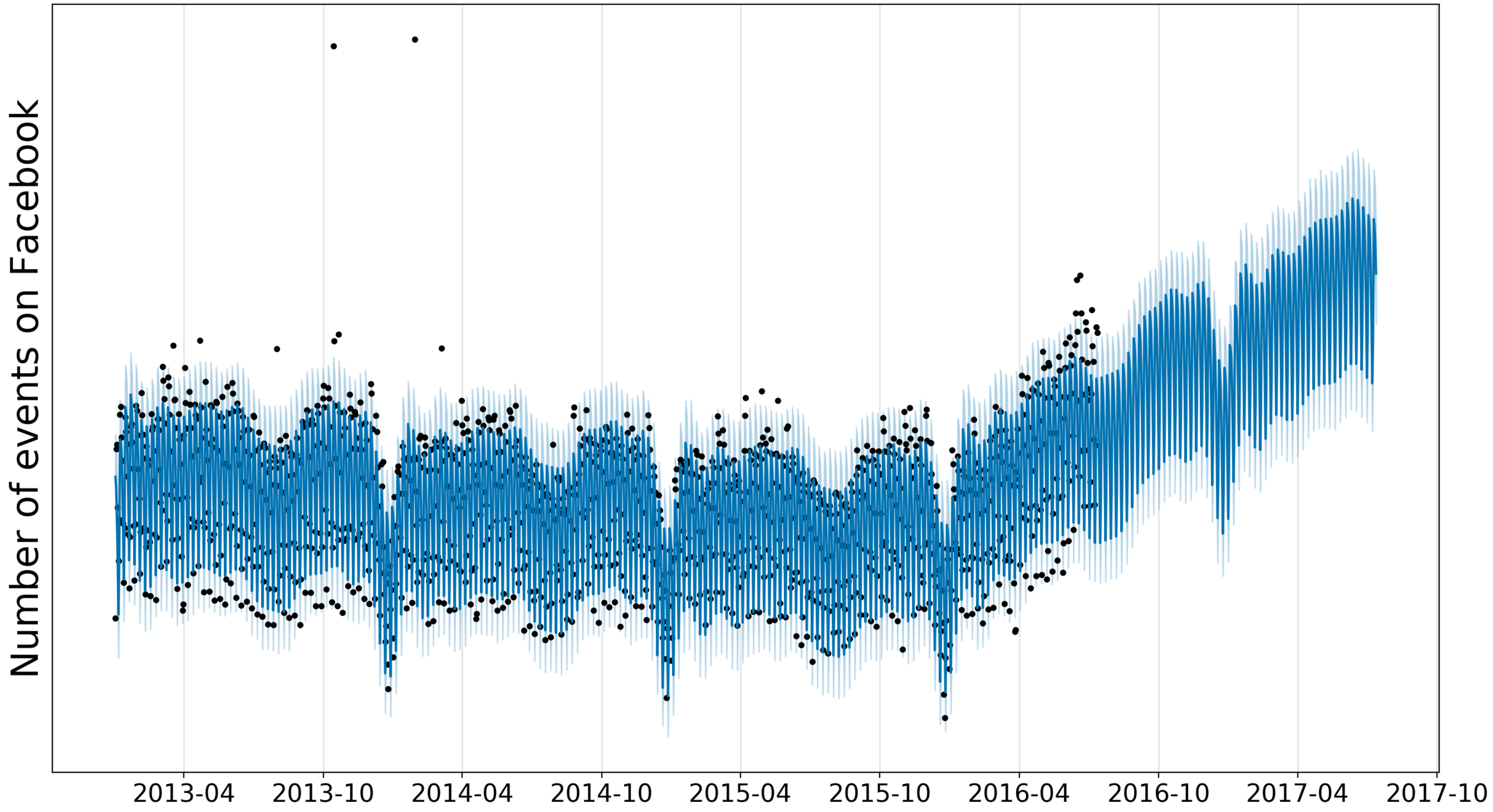
Parameters should capture structure

Forecasts from ARIMA(3,1,2)



Forecasts from ETS(A,N,N)





Additive model for curve fitting

$$\begin{aligned} y(t) = {} & \text{piecewise\_trend}(t) + \\ & \text{seasonality}(t) + \\ & \text{holiday\_effects}(t) + \\ & \text{i.i.d. noise} \end{aligned}$$

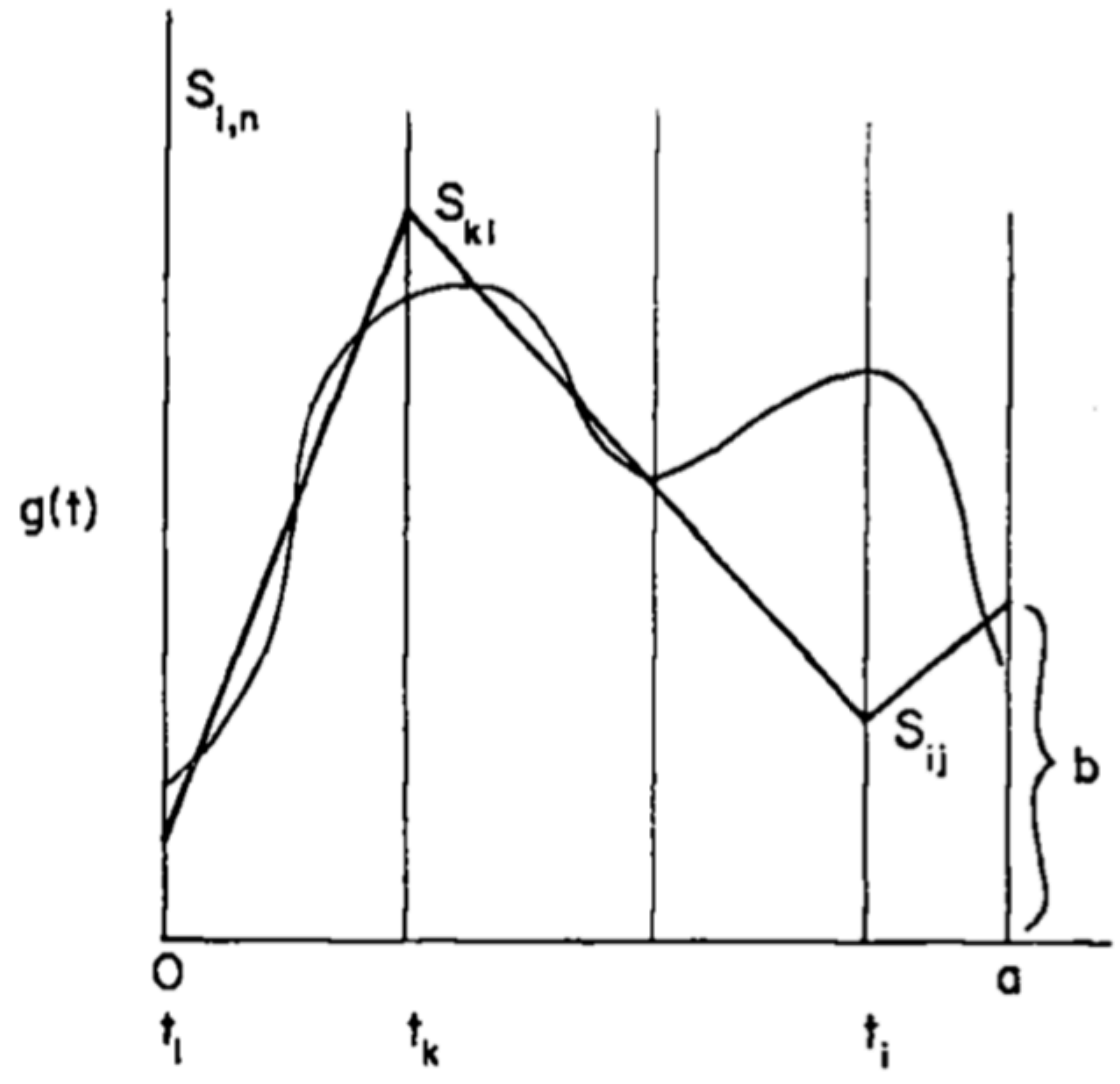
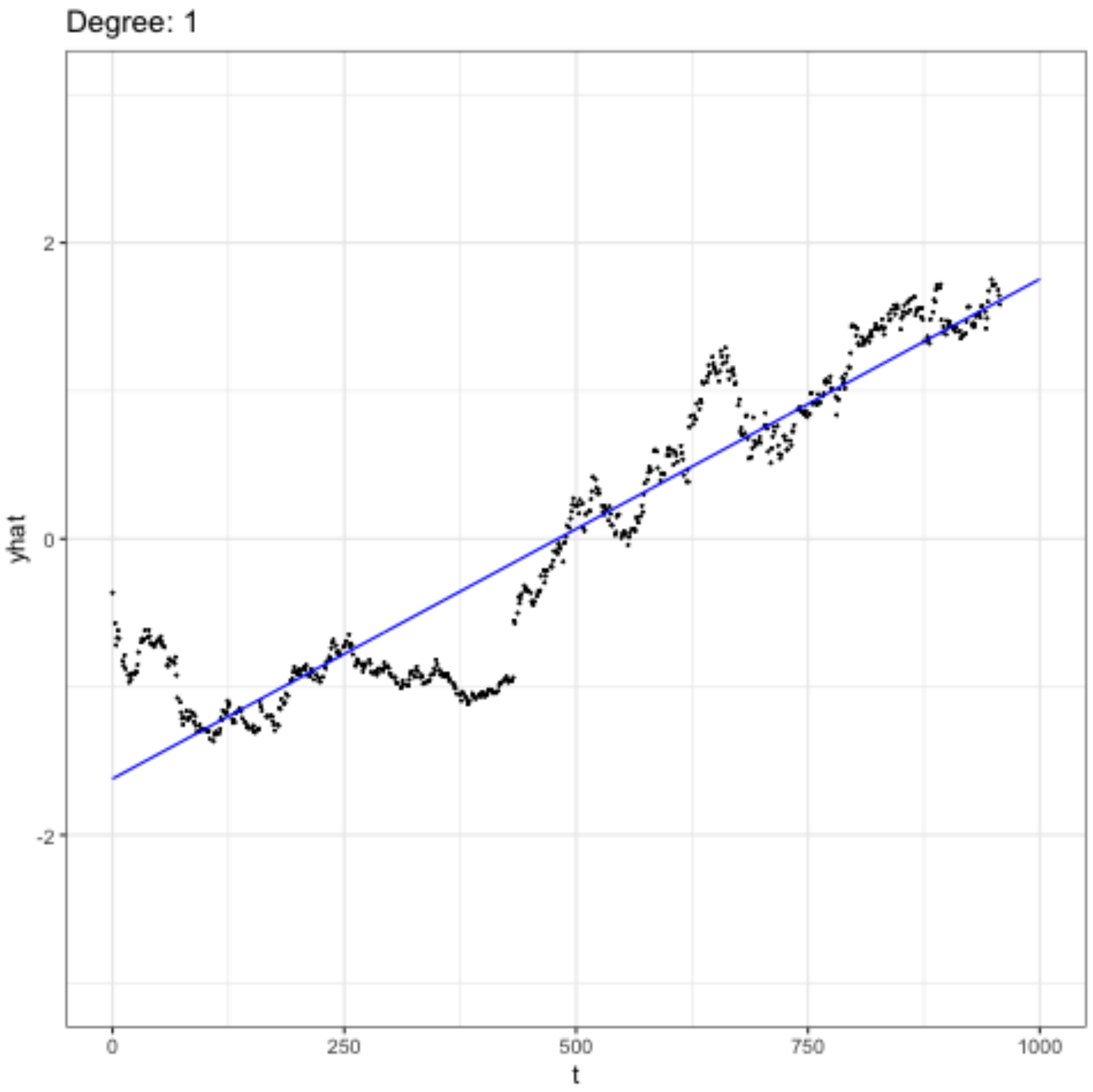


FIG. 1. Curve fitting by segmented straight lines.

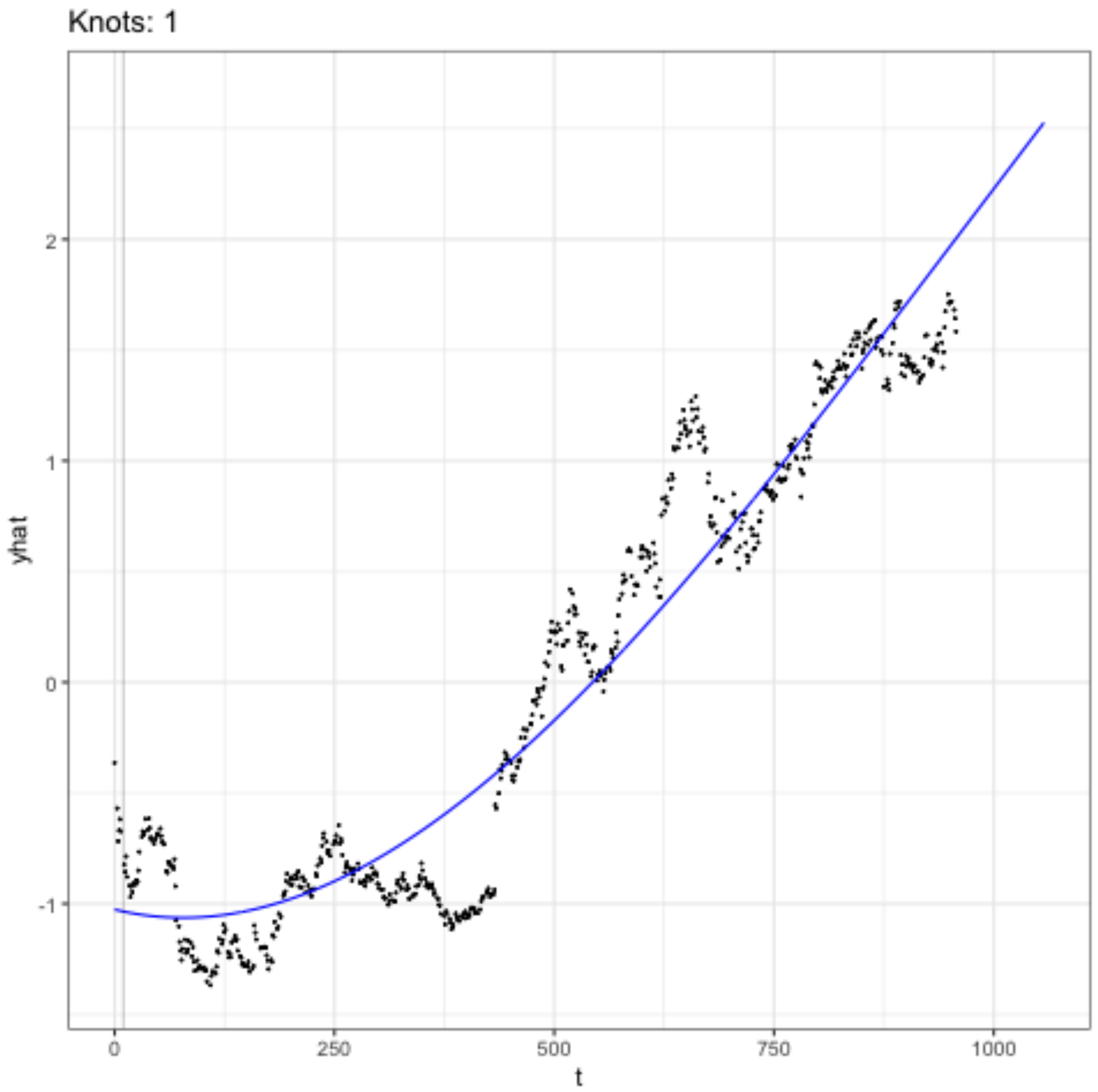
Polynomials

- Polynomials are a natural choice for fitting curves.
- We can control the complexity of the fit using the degree of the polynomial.
- But polynomials are terrible at extrapolation.



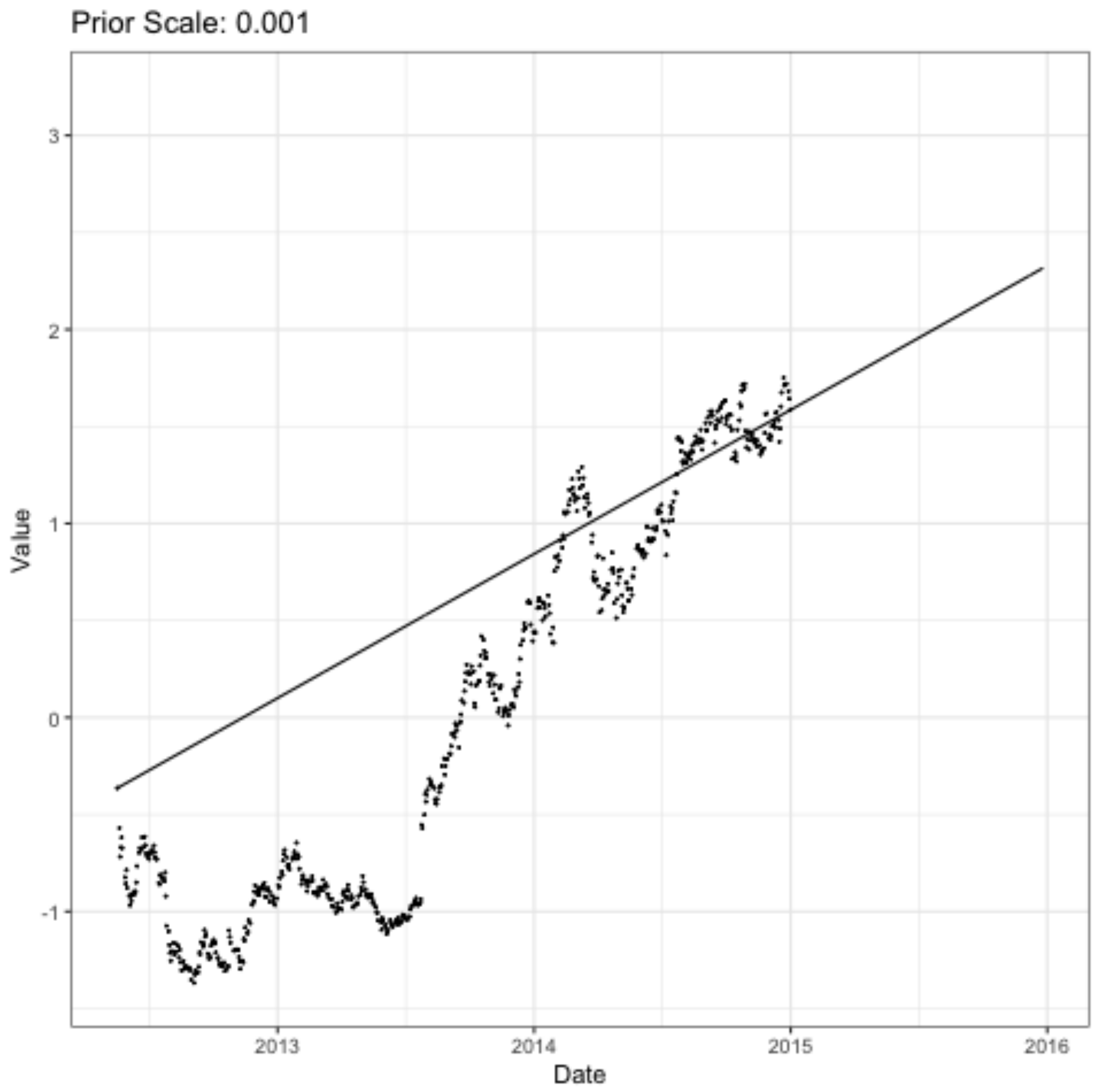
Splines

- Splines are piecewise polynomial curves.
- They can have lower interpolation error than polynomials with fewer terms.



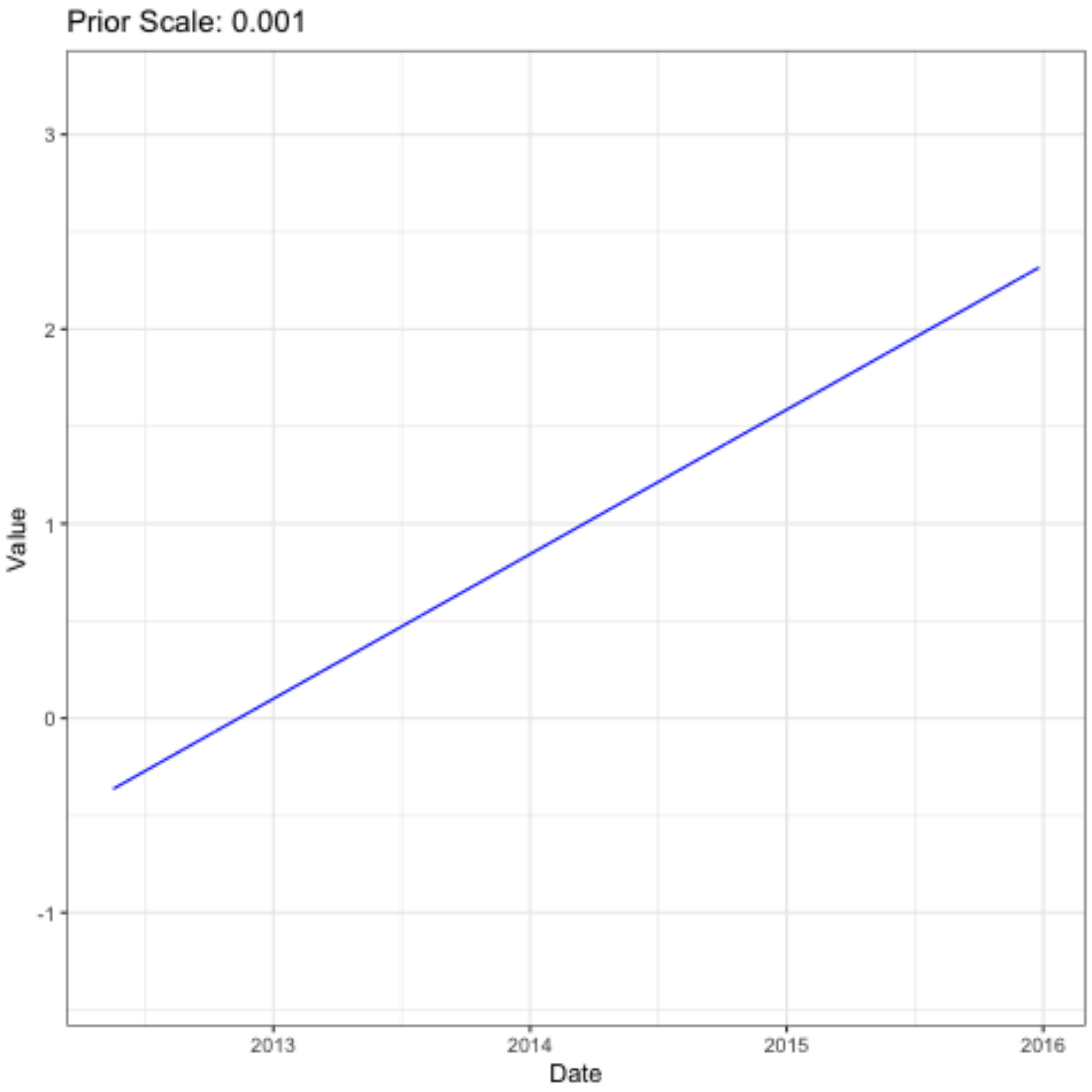
Piecewise linear

- The main curve that Prophet uses is piecewise linear.
- These curves are simple to fit and tend to extrapolate well.
- The hard part is deciding which “knots” or changepoints to use.



Changepoint selection in action

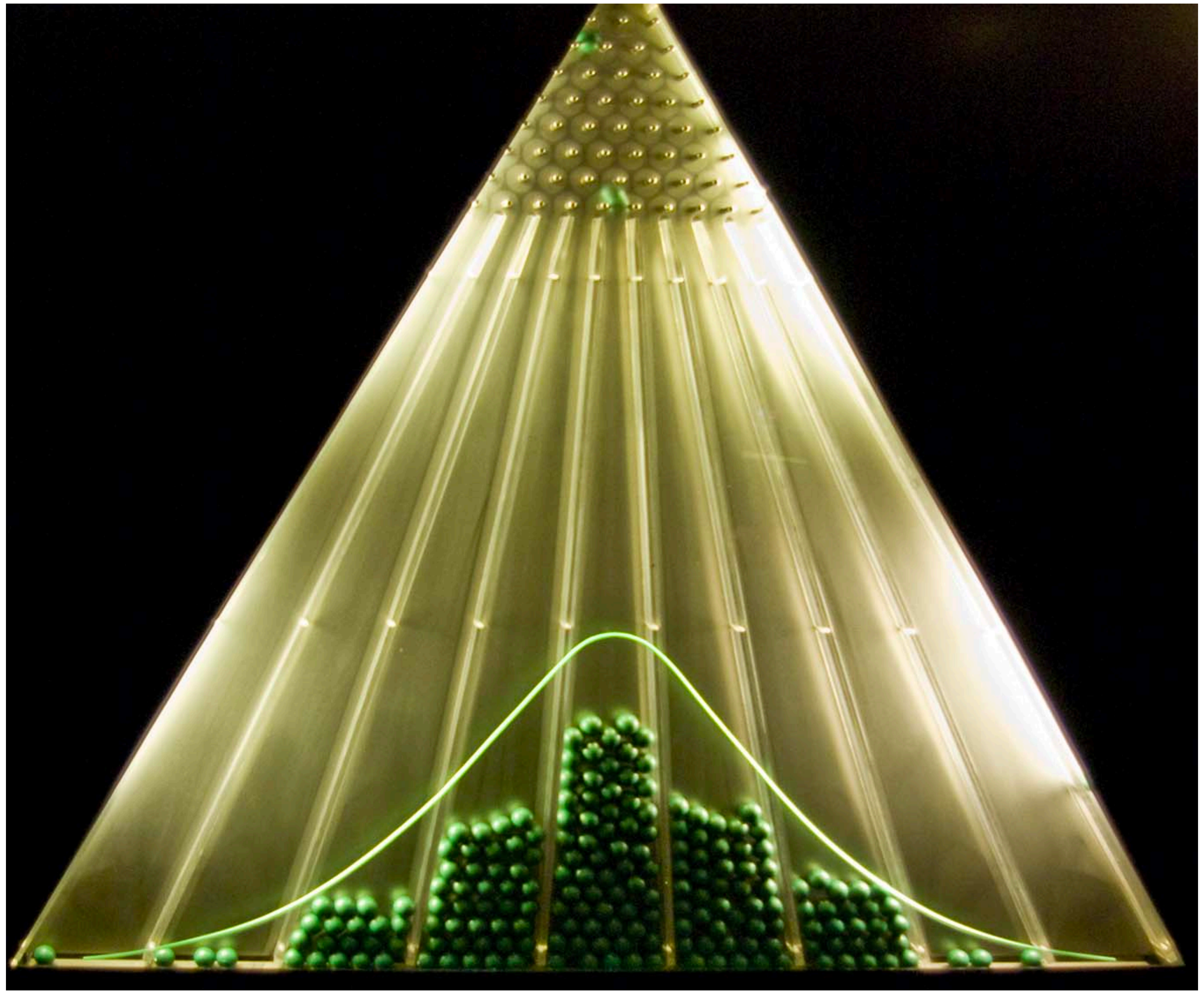
- We generate a grid of potential changepoints.
- Each changepoint is an opportunity for the underlying curve to change its slope.
- Apply a Laplace prior (equivalent to L1-penalty) to changes to select simpler curves.
- Smaller prior scales result in fewer changepoints and less flexible curves.



Estimating uncertainty

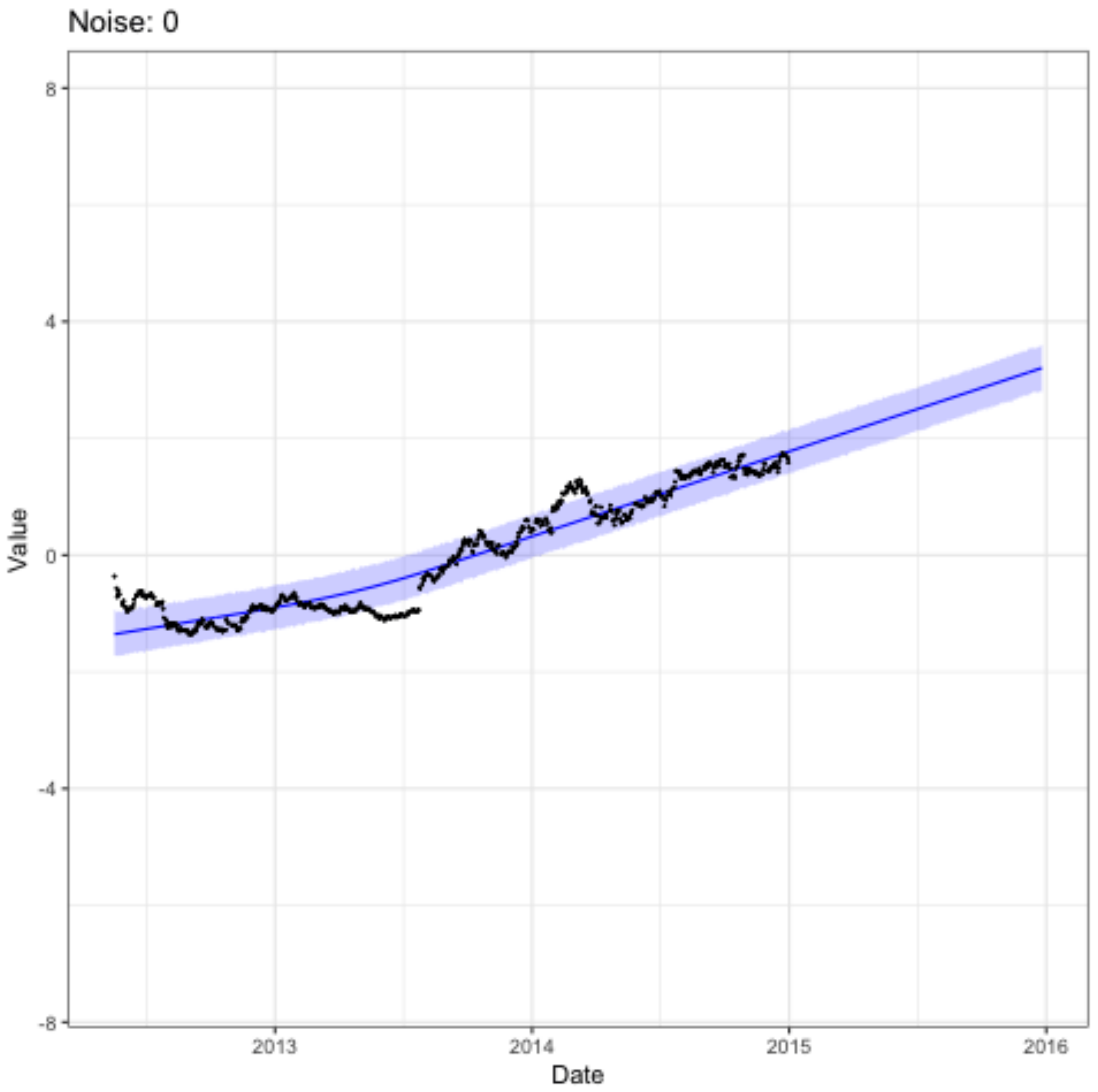
Three sources of uncertainty:

- irreducible noise (🤷‍♂️)
- parameter uncertainty (HMC)
- trend forecast uncertainty (simulation)



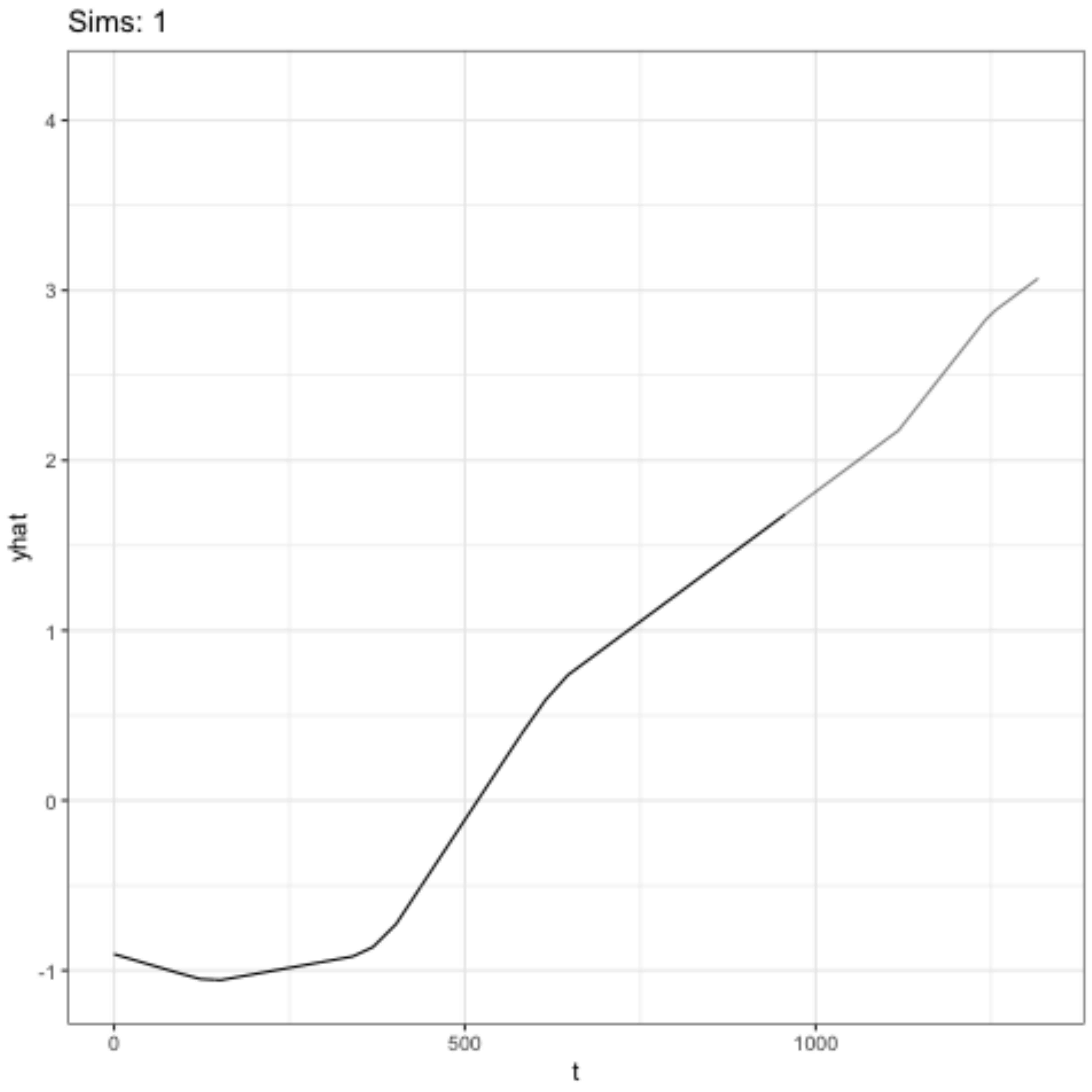
Irreducible uncertainty

- Anything Prophet cannot fit is modeled as mean-zero i.i.d. random noise.
- This creates tube-shaped uncertainty in the forecast.
- Large uncertainty indicates the model has fit the historical data poorly.



Trend change simulation

- At each date in the forecast we allow the trend to change.
- The **rate** of change is estimated based on how many changepoints were selected.
- The **distribution** of changes is selected based on their magnitudes.



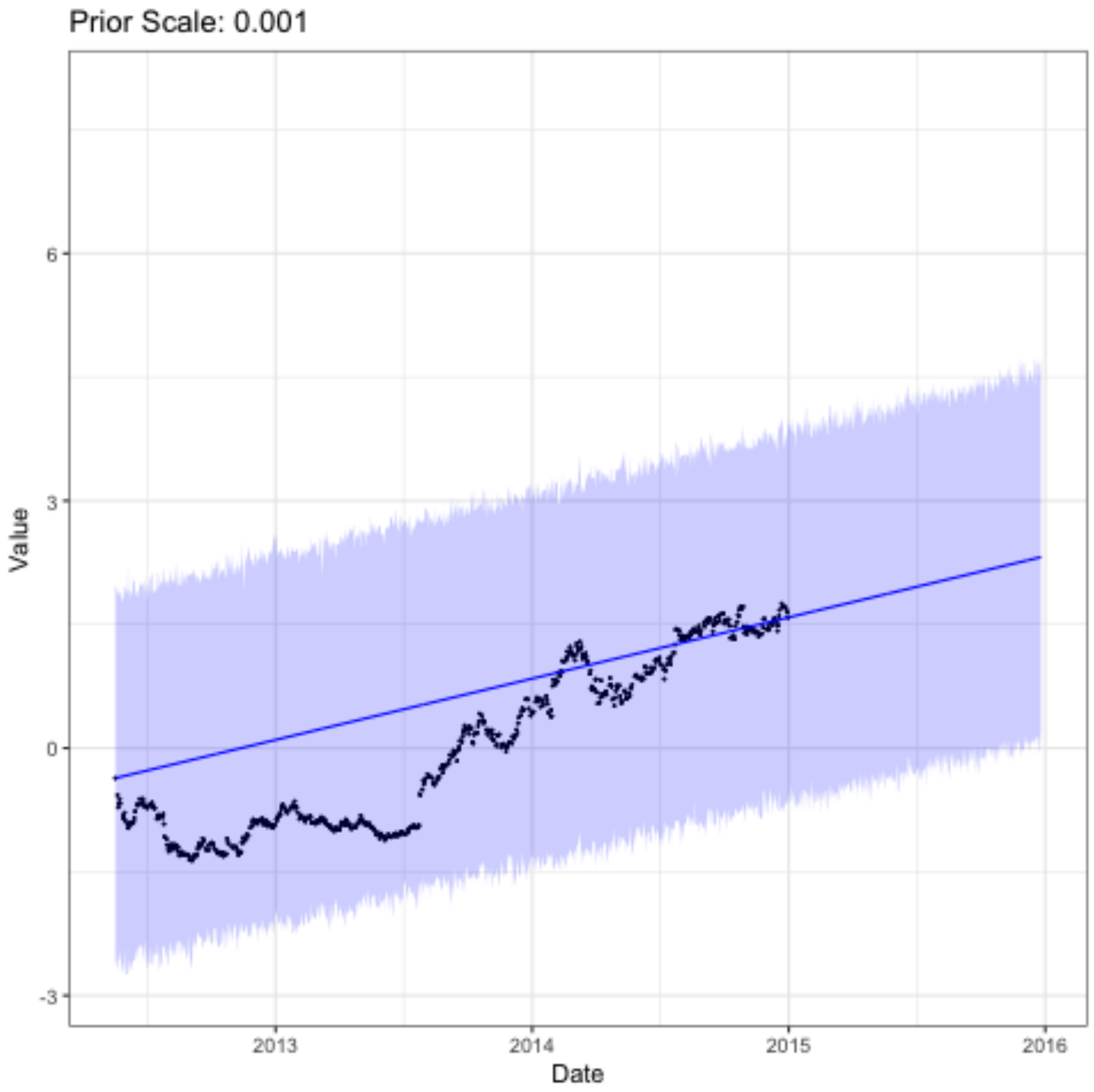
Tuning

If you run a forecasting procedure and you don’t like the forecast what can you?

- Adjust the input data you supply.
- Manually edit the results in a spreadsheet.
- Change the parameters you used for your model.

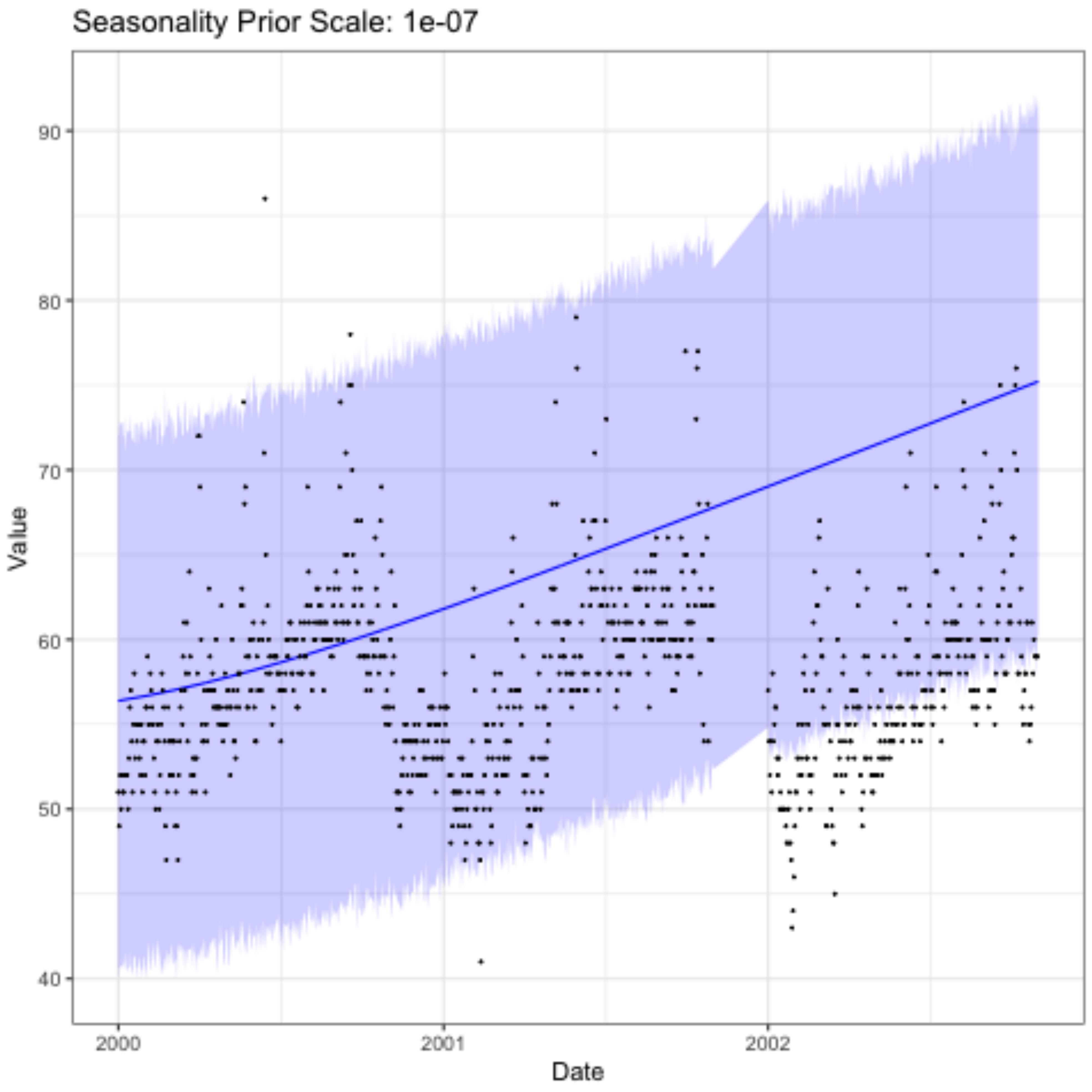
Changepoint prior scale

- How likely we are to include changepoints in the model.
- Controls flexibility of the curve.
- <u>Rigid curves</u>: large i.i.d. errors (tube shaped)
- <u>Flexible curves</u>: large trend uncertainty (cone shaped)



Seasonality prior scale

- Regularizes the parameters on the Fourier expansion.
- Overfitting seasonality can also be controlled by turning off various types of seasonal patterns or using fewer Fourier terms.

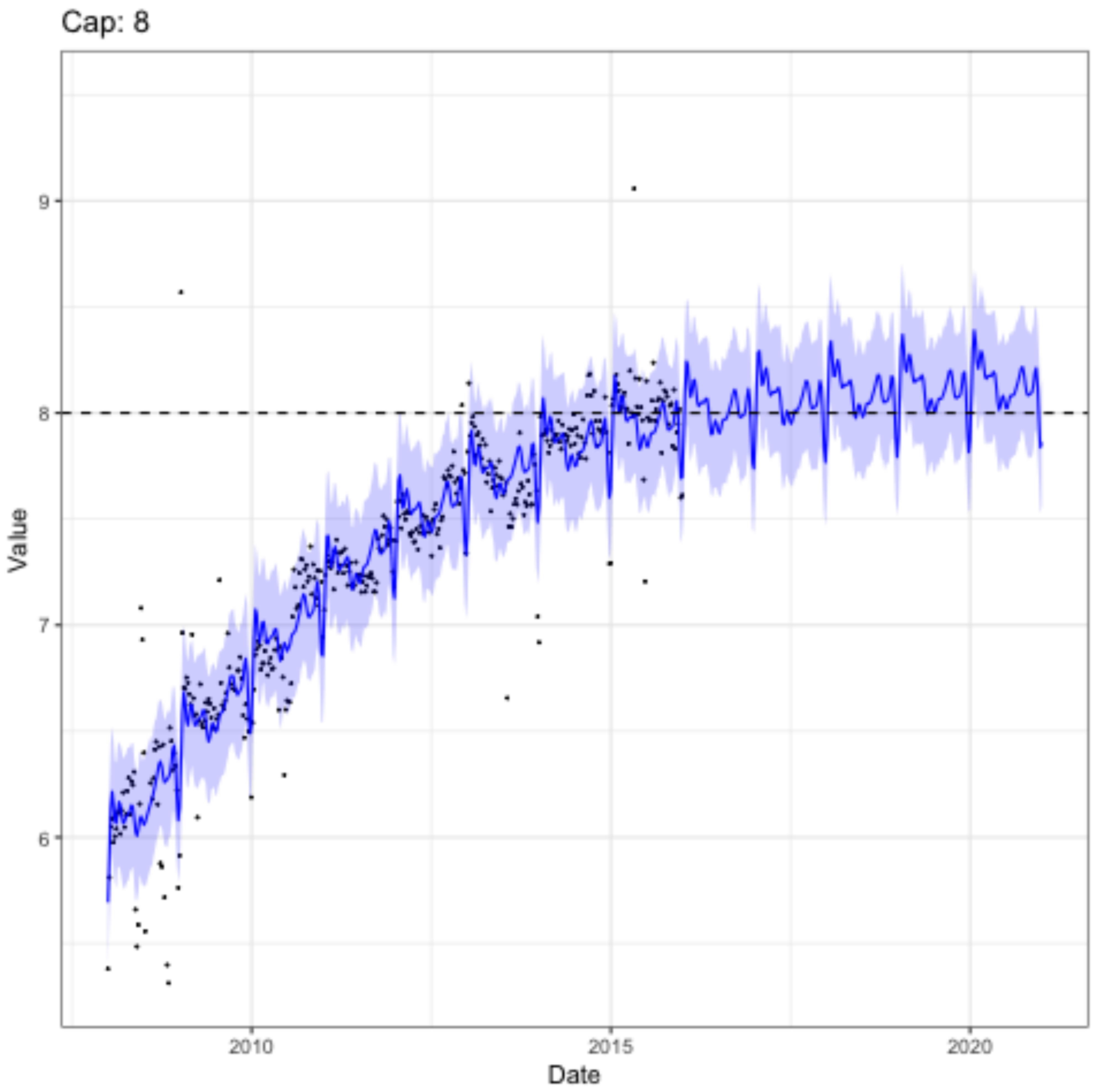


Capacities

- Piecewise logistic growth curves have a capacity parameter that we do not fit from data.

- Often we can use obvious constraints as upper and lower bounds on forecasts.

- The user can specify the capacity as a constant or as a time series.



Takeaways

- Forecasting “at scale” is 25% technology problem 75% people problem.
- Prophet is a simple model (with some tricks) but covers many important use-cases at Facebook and elsewhere.
- Simple is good! Prophet works robustly and fails in understandable ways.
- Using curve-fitting with interpretable parameters allows users to input their domain knowledge into forecasts.