Lecture 3: Kalman filter

April 5, 2016

Likelihood statistics

Problem:

We often can't write the probability of data given parameters

Examples:

1. Tag-recapture
   - What's the probability of tagging an animal in 2008, seeing it again in 2010 and 2011, and then never seeing it again?
2. Time-series
   - What's the probability distribution for escapement of chinook salmon in the snake river in 2011, given that you've sampled escapement from 1980-2010?
3. Occupancy
   - Three volunteers look for an endangered butterfly at a site, and only two find it. These volunteers sample at a new site, and none see the butterfly. What is the probability that is present but wasn't detected?

Likelihood statistics

Solution:

- Introduce "latent" variables

$$\Pr(y, \varepsilon|\theta) = \Pr(y|\theta, \varepsilon)\Pr(\varepsilon)$$

  – where ε is a unobserved random variable
  – $\Pr(\varepsilon)$ is a "prior" or "hyper-distribution" for latent variables
  – ε is sometimes called "augmented data"
    - Left side of the joint-likelihood
- Calculate the marginal likelihood of parameters when integrating across random effects

$$\Pr(y|\theta) = \int \Pr(y|\theta_1, \varepsilon)\Pr(\varepsilon|\theta_2)\,\mathrm{d}\varepsilon$$

  – *Marginalize – take a weighted average of likelihoods, where weights are given according to the probability of random effects,* $\Pr(\varepsilon|\theta_2)$

Definitions

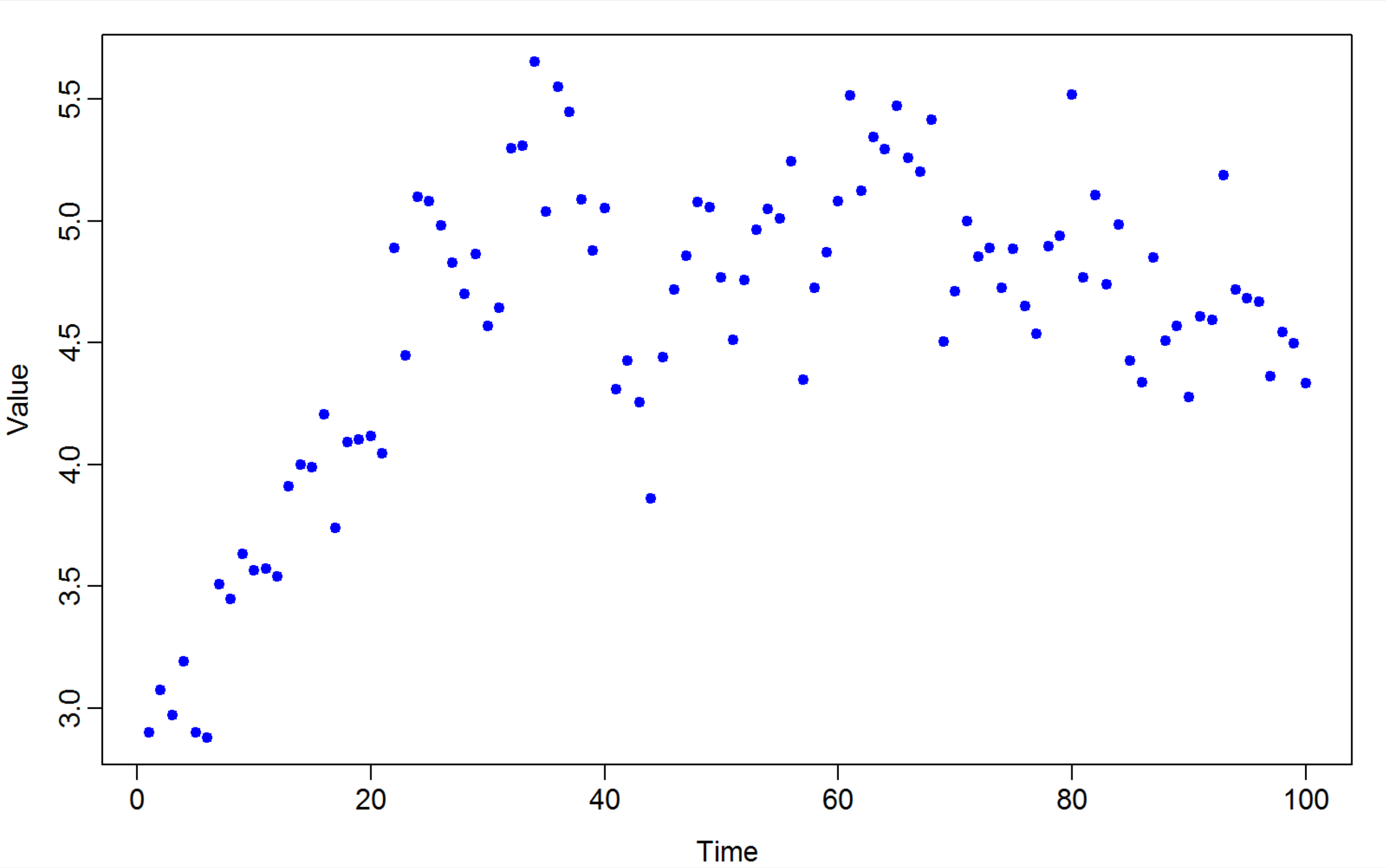
Term	Definition
Random effect	Coefficient that is “exchangeable” with one or more other coefficients
Hyperdistribution	Distribution for “exchangeable” random effects
Exchangeable	No information is available to distinguish between residual variability in random effects
Fixed effect	Coefficient that is not exchangeable with others, and which hence is estimated without a hyperdistribution
Mixed-effect model	Model with both fixed and random effects

Directed random walk

$$x_{t+1} = x_t + \varepsilon_t$$
$$\varepsilon_t \sim Normal(\alpha, \sigma_x^2)$$
$$\log(y_t) \sim Normal(x_t, \sigma_y^2)$$

– Time-series follows a random-walk with a trend α

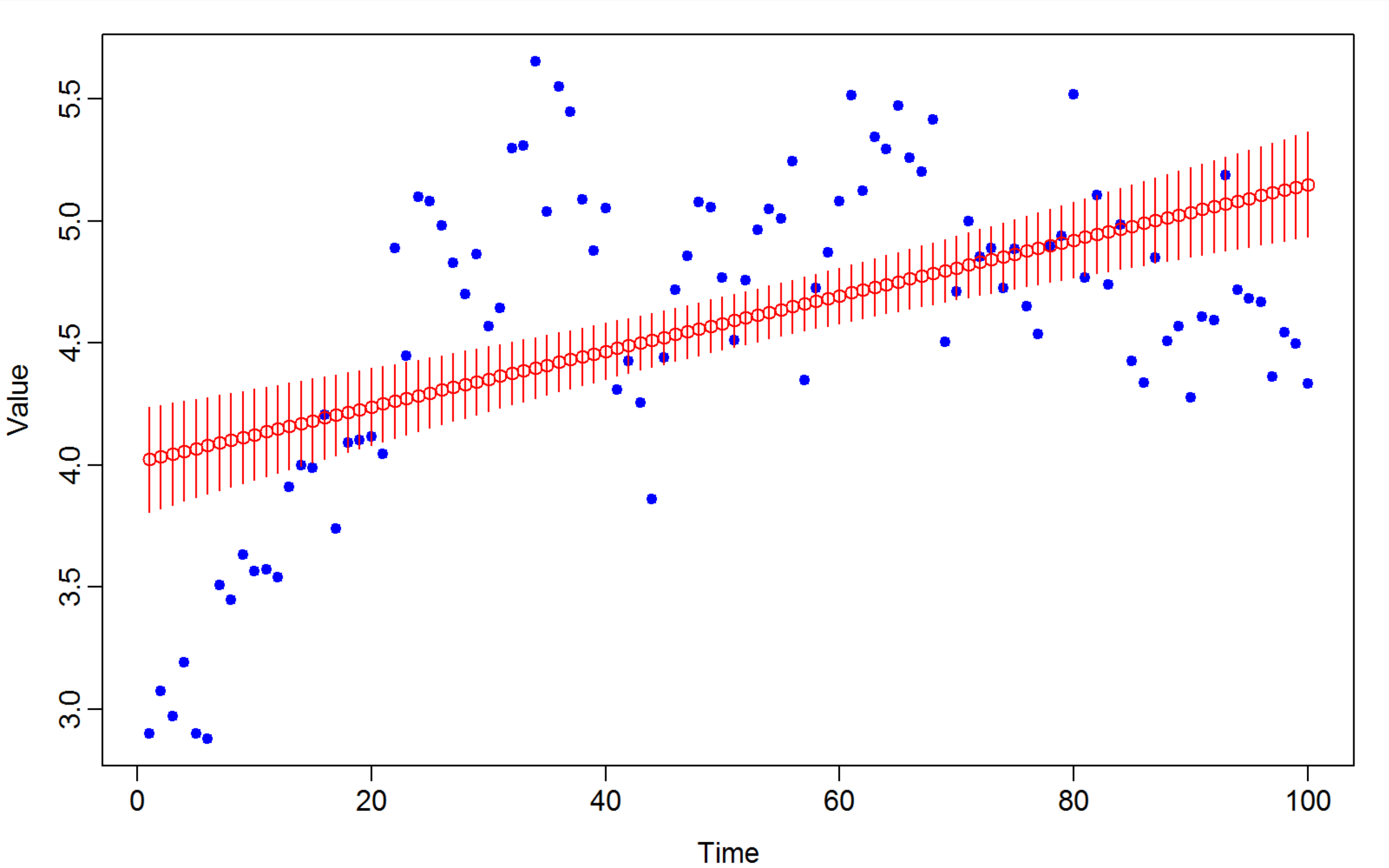
Data set



Approaches

1. Linear model
2. Loess smoother
3. Generalized additive model
4. Kalman filter

Linear model

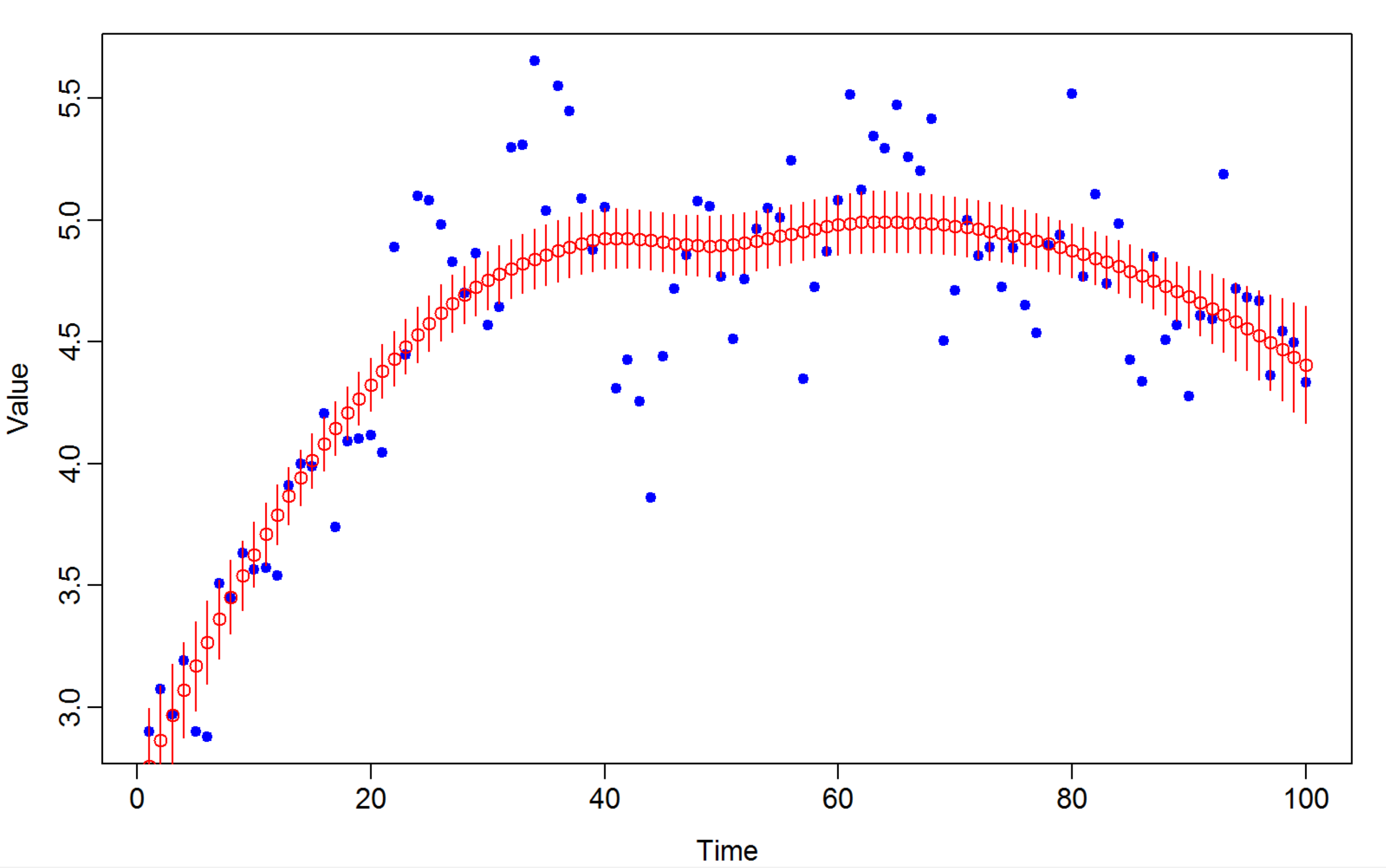


Linear model

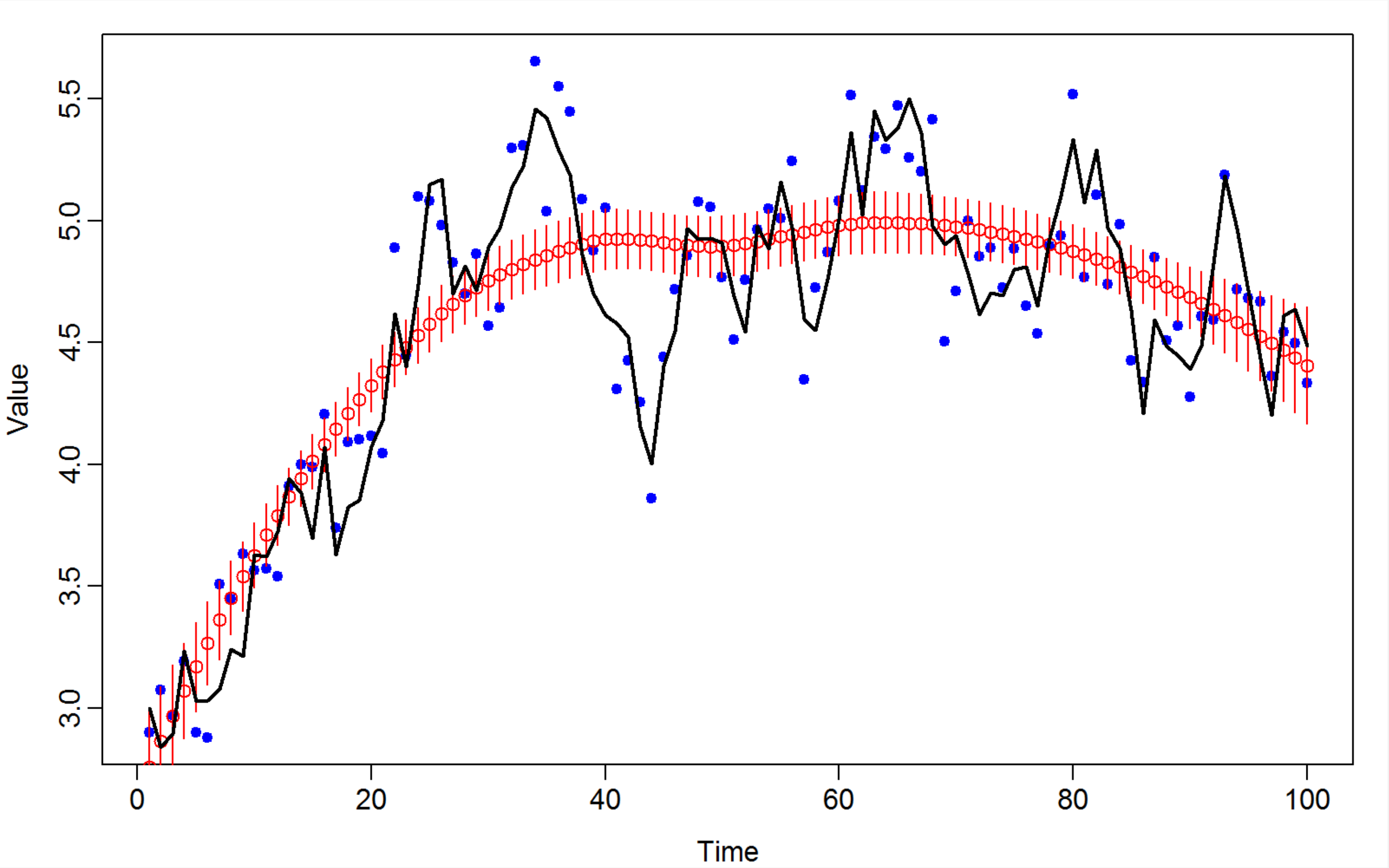
Problems

- Huge residuals at beginning and ending
- Predictive variance is larger at beginning and end of series
- Doesn’t contain the true value very often
- Not sufficiently flexible

Loess smoother



Loess smoother

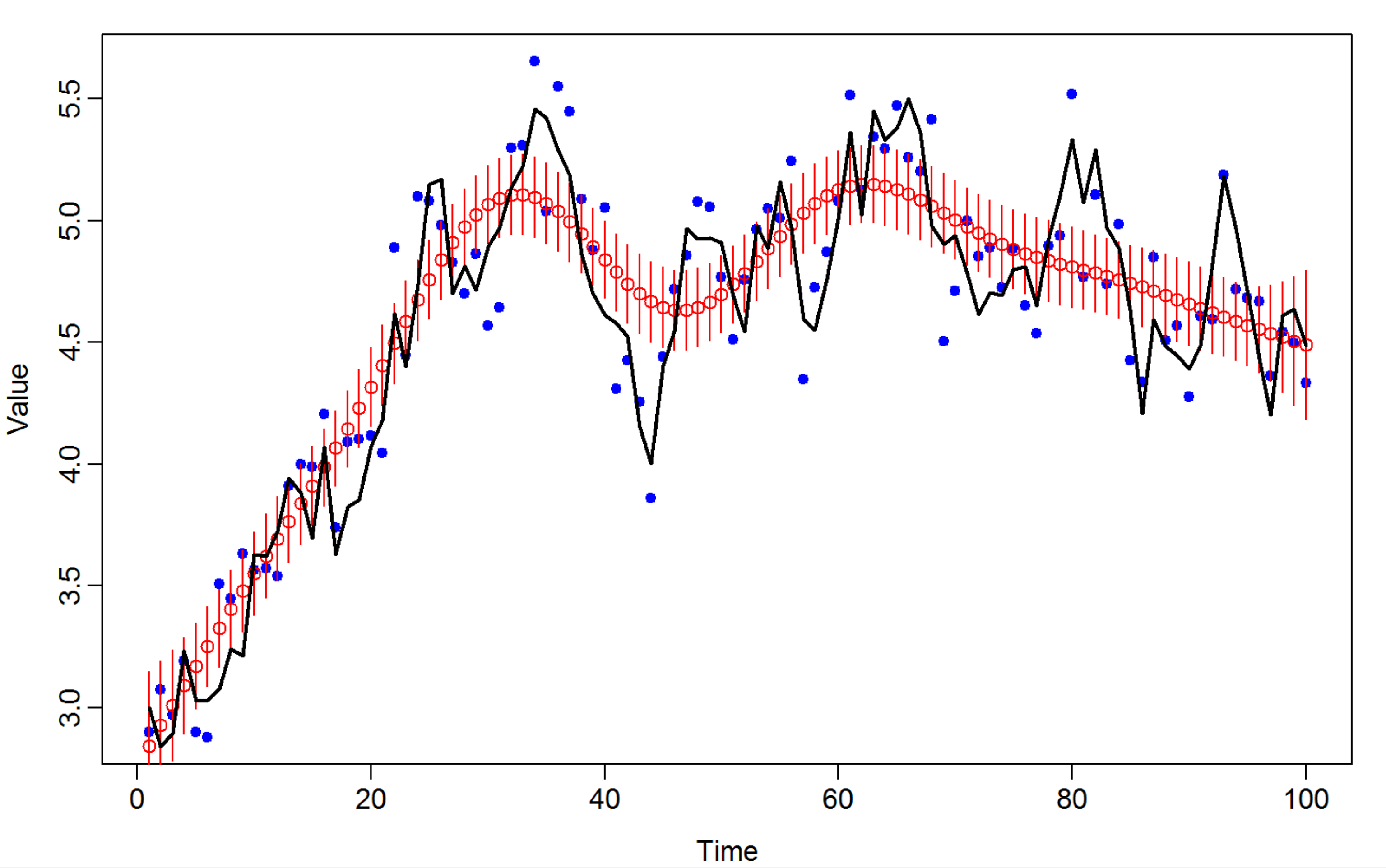


Loess smoother

Problems

- No model to specify
  - No interpretation of parameters as data-generating process
- Confidence intervals don’t include the true values
- Seems to “oversmooth”
  - i.e., strings of years are over/under-estimated

Generalized additive model



Generalized additive model

Problems

- Confidence intervals don’t include the true values
- Misses some fine-scale variation

Kalman filter

$$x_{t+1} = x_t + \varepsilon_t$$

$$\varepsilon_t \sim Normal(\alpha, \sigma_x^2)$$

$$\log(y_t) \sim Normal(x_t, \sigma_y^2)$$

Steps:

- Code up data-generating process
- Treat ε_t as random effect
- Estimate parameters using maximum likelihood

Kalman filter

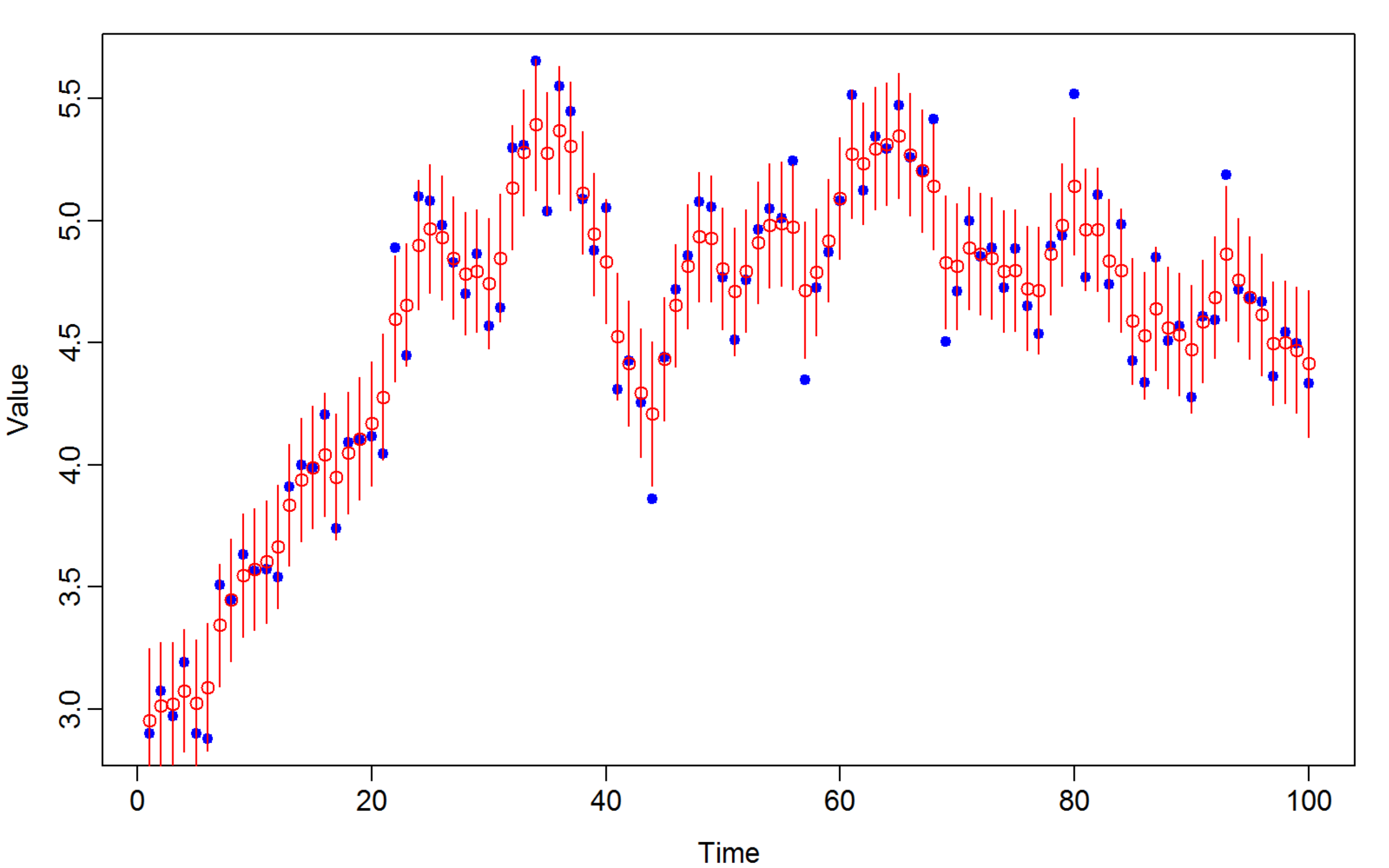
$$\boldsymbol{\theta} = (x_0, \alpha, \sigma_x^2, \sigma_y^2)$$

$$\hat{\boldsymbol{\theta}} = \mathrm{argmax}_{\boldsymbol{\theta}} \left(\int \Pr\left(\mathbf{y} \middle| \mathbf{x}, \sigma_y^2\right) \Pr(\mathbf{x}|\alpha, x_0, \sigma_x^2)\, d\mathbf{x} \right)$$

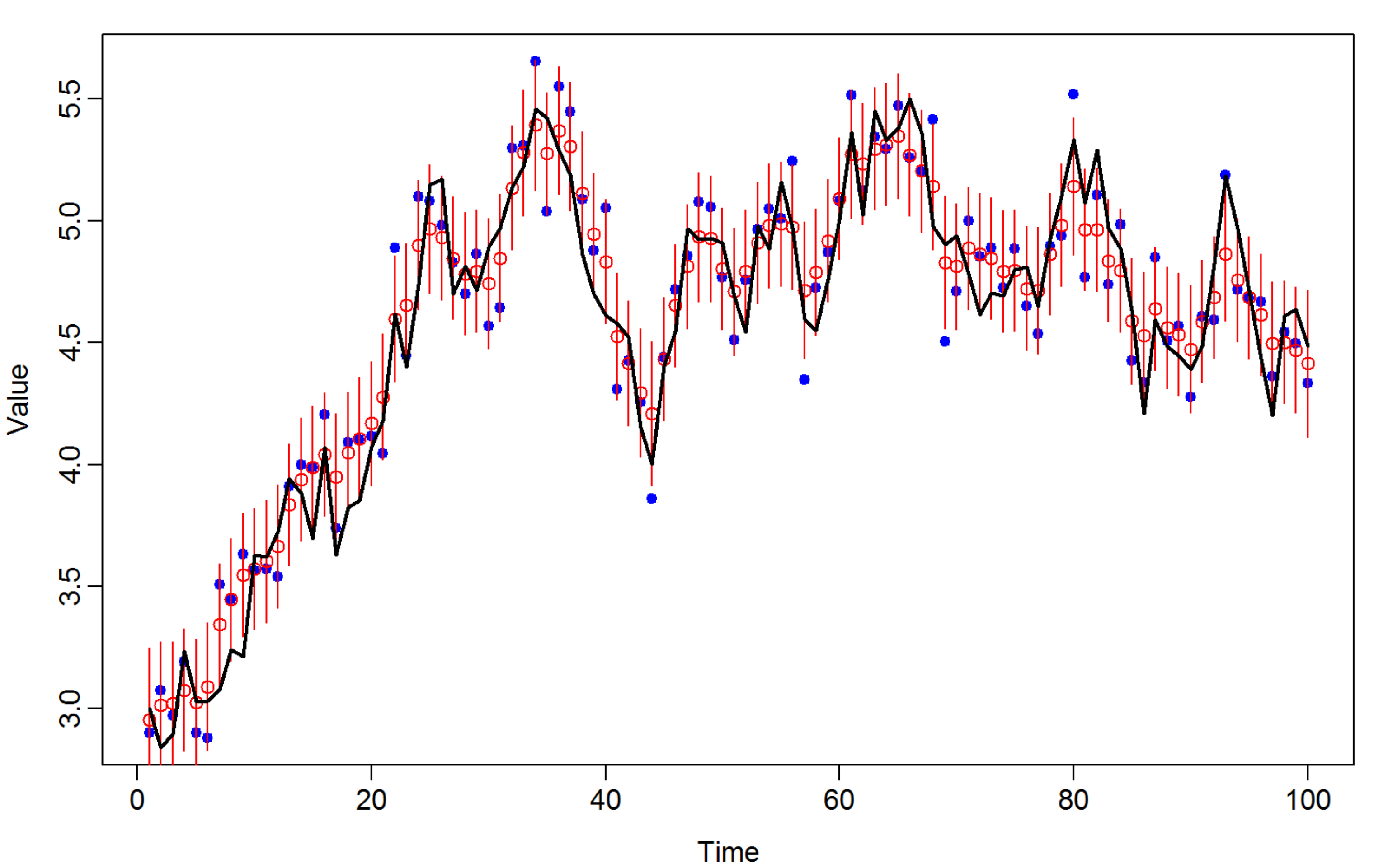
… where we can split this into smaller integrals …

$$= \mathrm{argmax}_{\theta} \left(\prod_{t=2}^{n_t} \int \Pr\left(y_t \middle| x_t, \sigma_y^2\right) \Pr(x_t|\alpha, x_{t-1}, \sigma_x^2)\, dx_t dx_{t-1} \right)$$

Kalman filter



Kalman filter



Kalman filter

Improvements

- Specifies an explicit model
- Confidence intervals include the true values
- Seems to behave intuitively
  - Predictions are shrunk towards data, and neighboring predictions