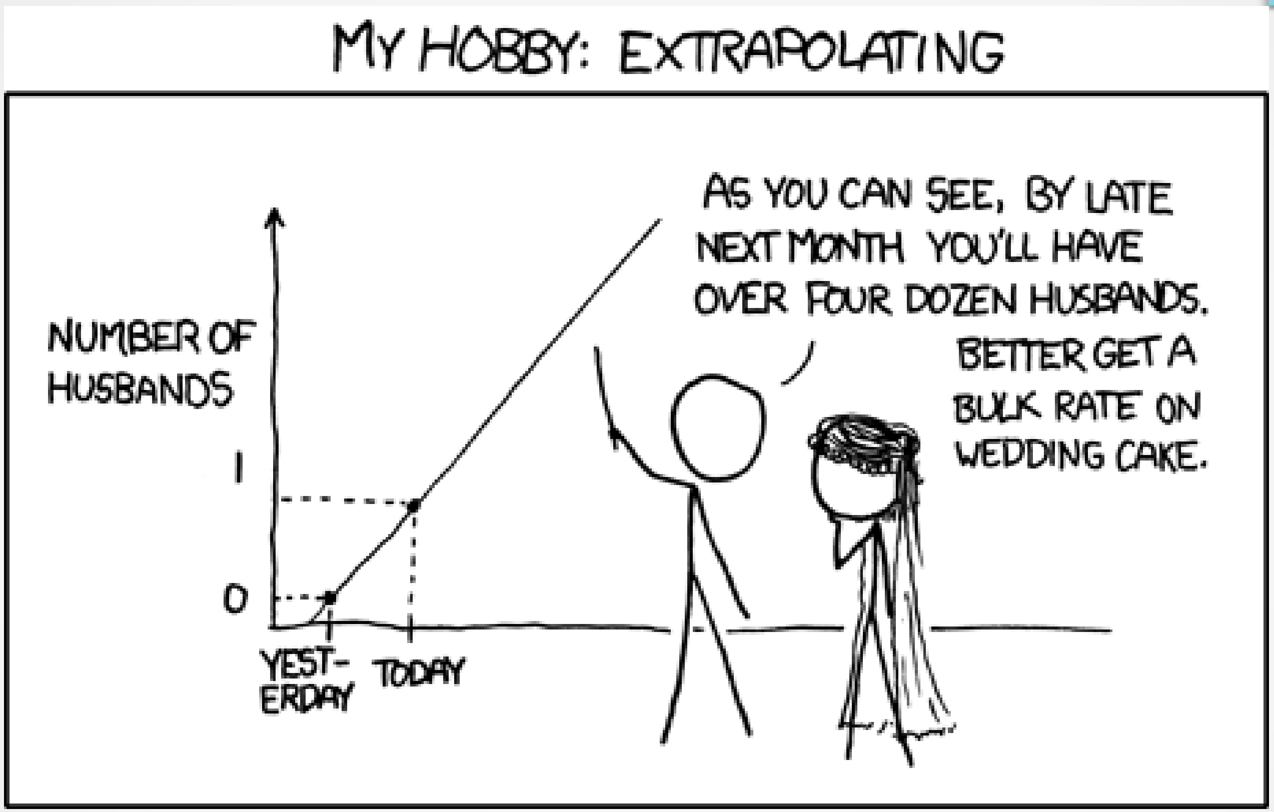
STATS 331

xkcd.com



Introduction to Bayesian Statistics

Semester 2, 2016

SET – Student Evaluations

- You should have received an email from the university about completing evaluations online
- Please do these – we *do* read them and consider what you say!

Today's Lecture

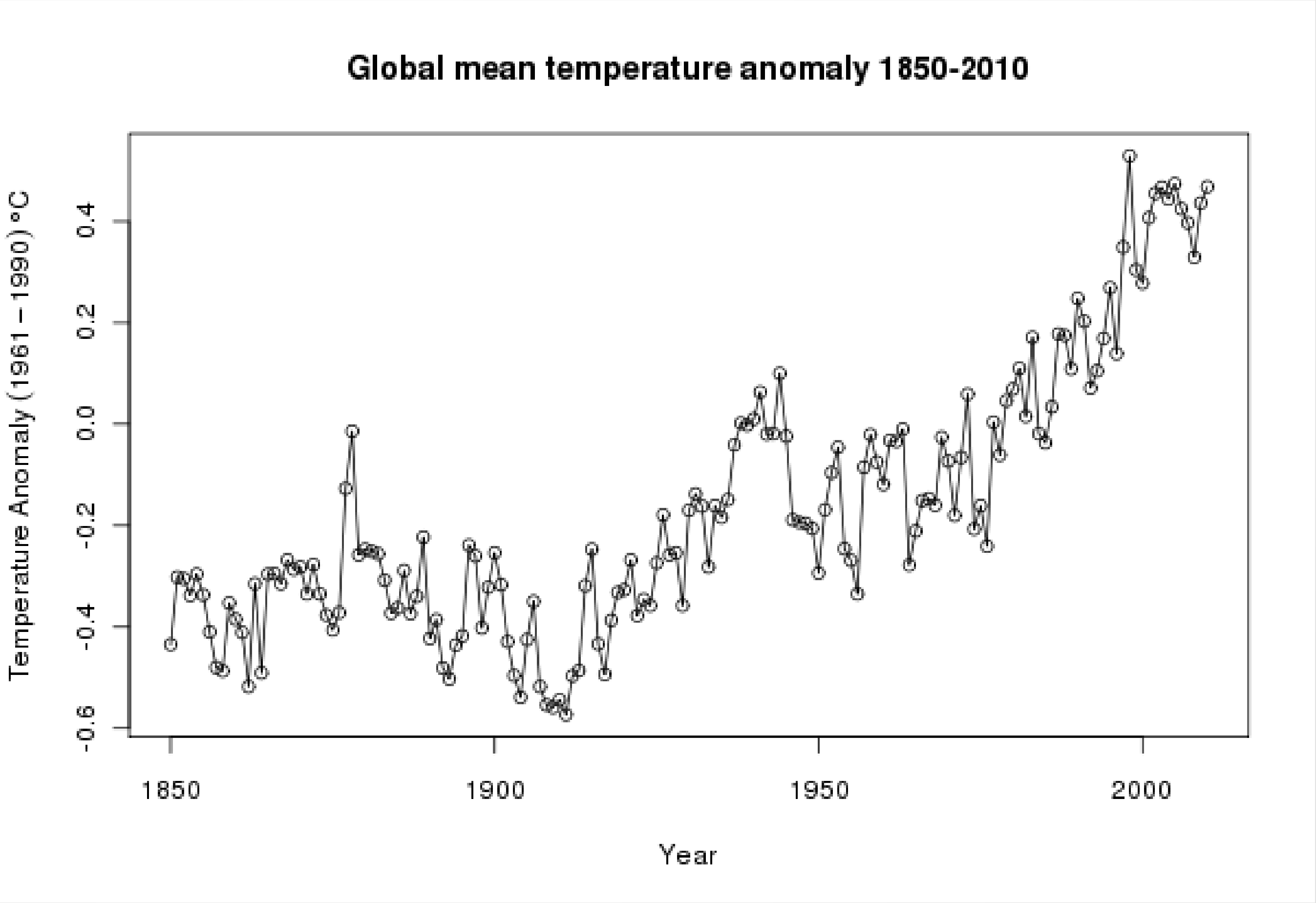
- Time series
- Why? Time Series models are really useful, and quite fun

Question

- How many of you have studied a course containing “time series models” such as the AR(1) model?
- It's ok if you haven't seen this!

What is a Time Series?

- Double meaning
- Meaning 1: any quantity varying over time



The Other Meaning

"A **probability distribution** for a quantity plotted over time"

- i.e. not any single curve, but a *probability distribution over the set of possible curves*

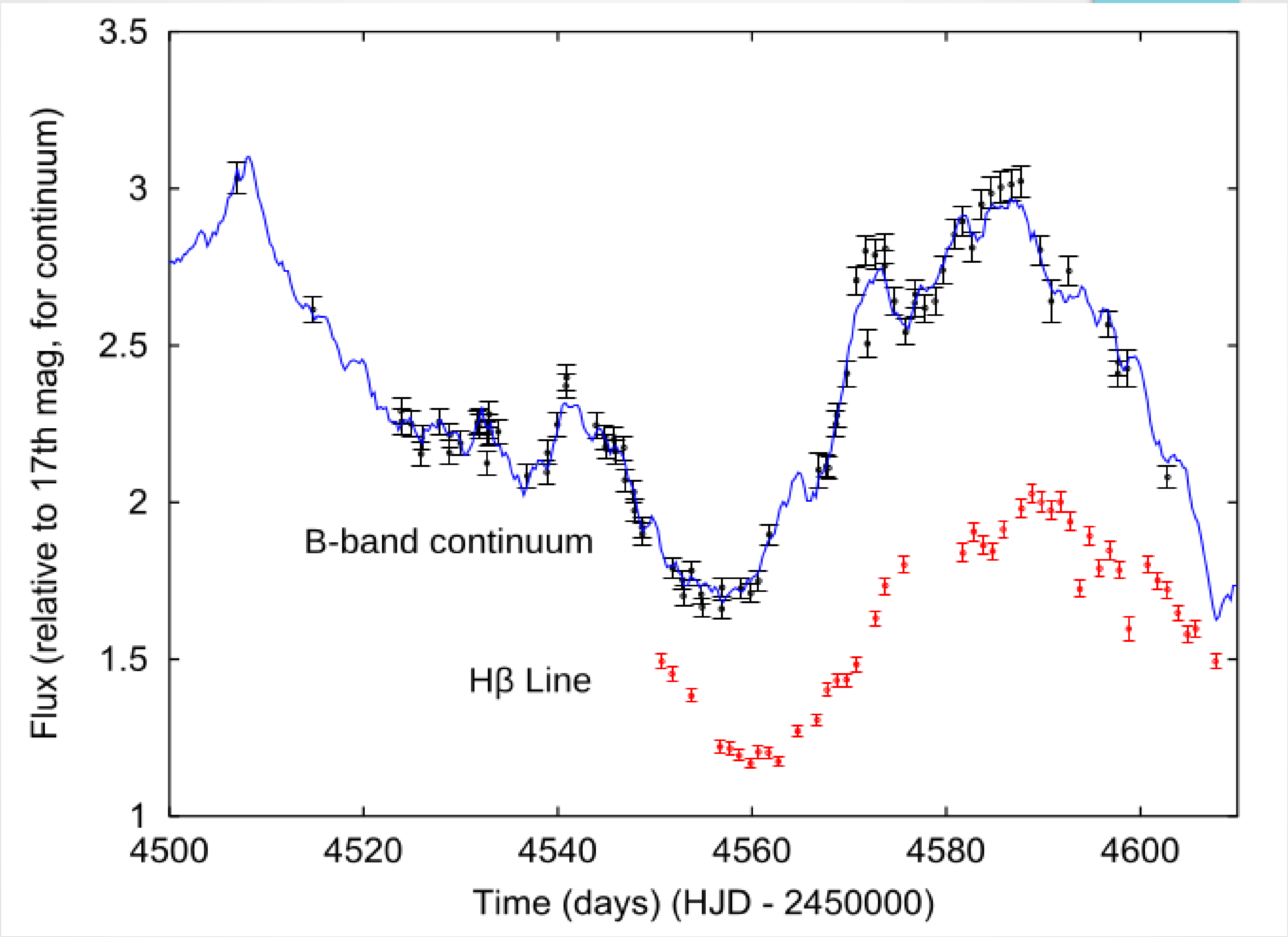
Applications

- The applications of time series models are **immense**

Astronomy

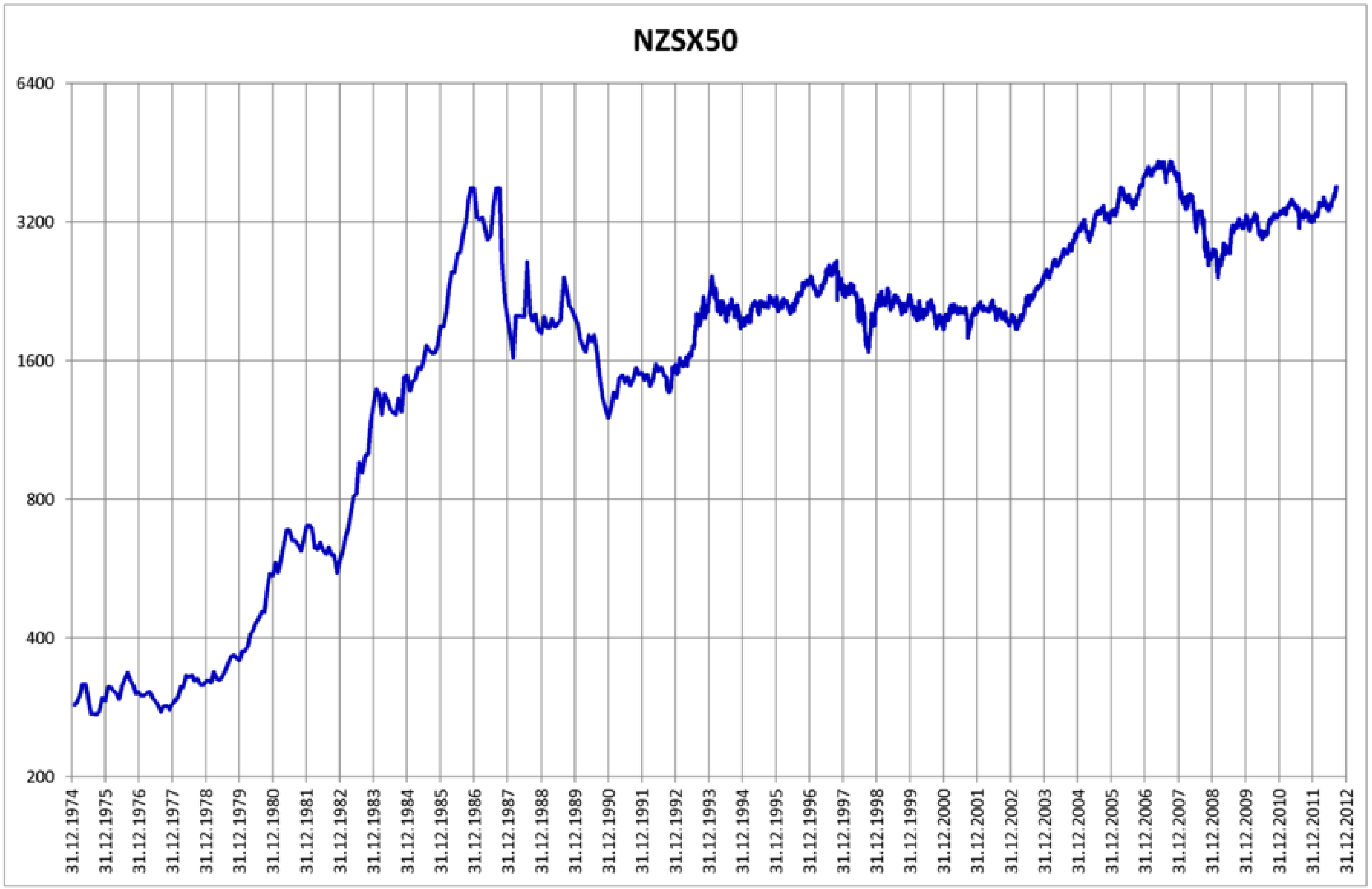
Artist's impression of a quasar. Credit: ESO/M. Kornmesser

Astronomy



3.5
3
2.5
2
1.5
1
4500
4520
4540
4560
4580
4600
Flux (relative to 17th mag, for continuum)
Time (days) (HJD - 2450000)
B-band continuum
Hβ Line

Finance (Obviously)



The AR(1) Model

- AR stands for auto-regressive
- Good simple-ish model for a quantity that fluctuates a bit over time
- Time is **discrete**. e.g. {0, 1, 2, …}
- The AR(1) is a **probabilistic model**, it is a probability distribution over the set of possible trajectories/curves

The Basic Idea of AR(1)

- The quantity at current time is given by the quantity at the last time plus an “innovation”
- We specify a distribution for the innovations.

$$y_i = \mu + \alpha(y_{i-1} - \mu) + \epsilon_i$$

$$\epsilon_i \sim \mathcal{N}(0, \sigma^2)$$

Thinking about the AR(1)

- If the errors were zero, it would be an exponential decay to the mean value

The innovations “keep the fluctuations alive”

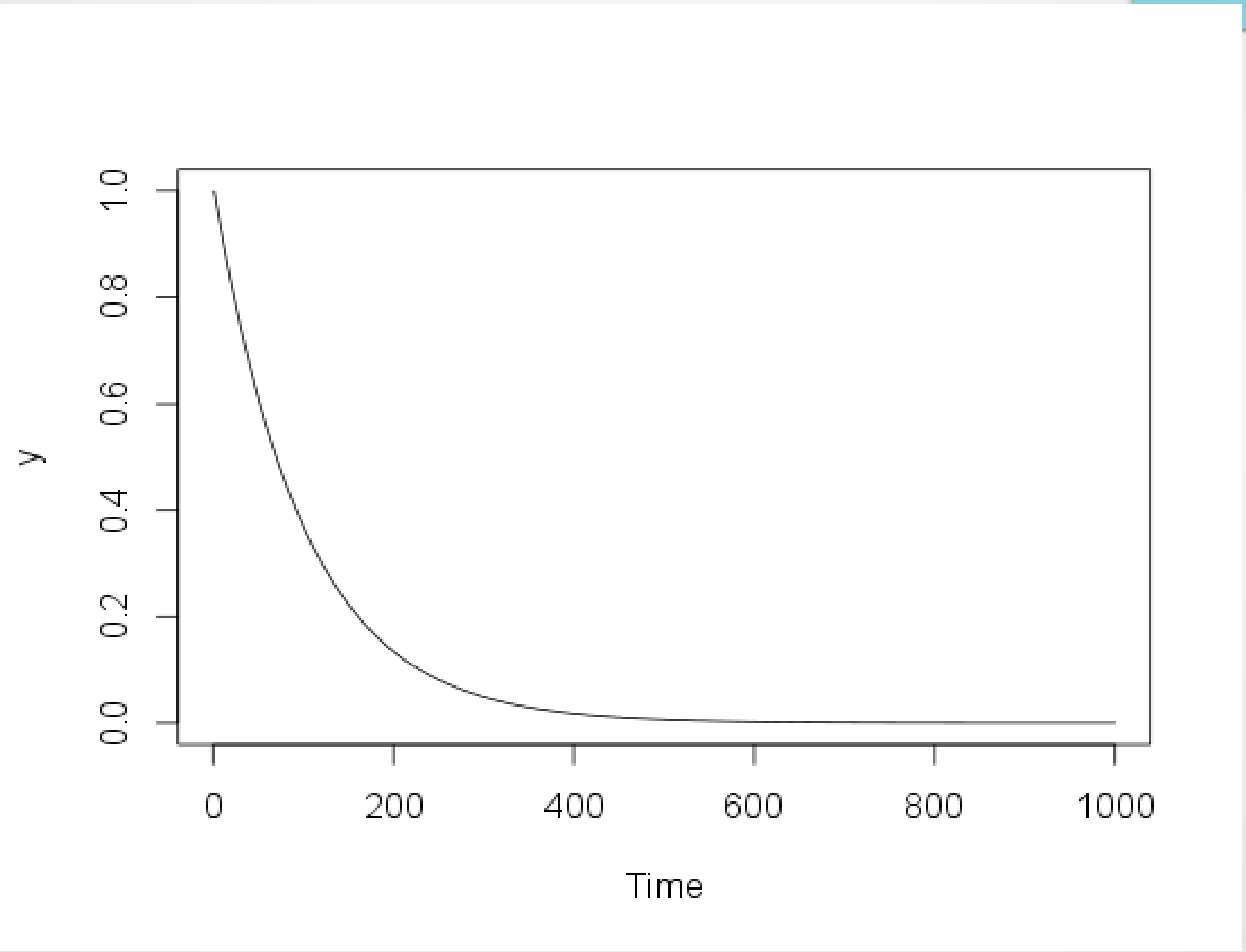
$$y_i = \mu + \alpha(y_{i-1} - \mu) + \epsilon_i$$

$$\epsilon_i \sim \mathcal{N}(0, \sigma^2)$$

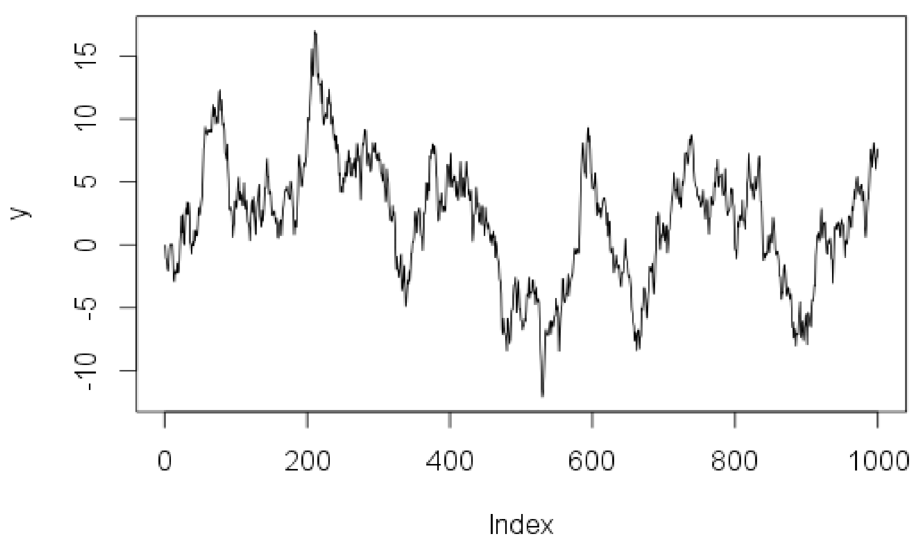
AR(1) Simulation in R

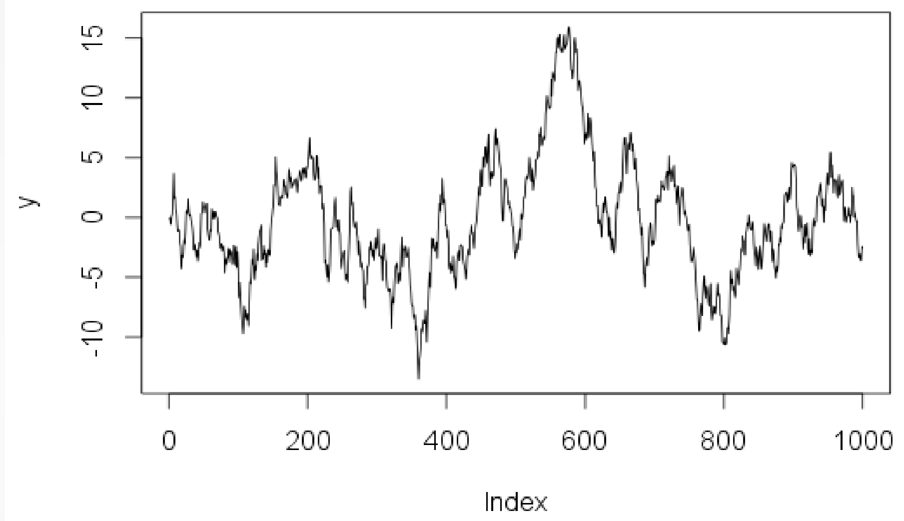
```
N = 1000 # Length of simulation
y = rep(1, N) # Storage
mu = 0    # Mean value
alpha = 0.99  # How much of old value to “remember”
sigma = 1 # How much to “kick” each time
for(i in 2:N)
{
   y[i] = mu + alpha*(y[i-1] – mu) + sigma*rnorm(1)
}
```

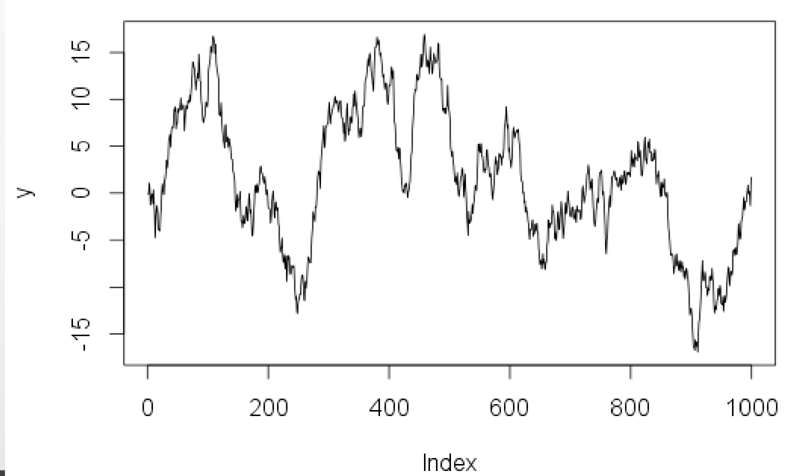
With No Innovations



With innovations – Typical time series from AR(1)







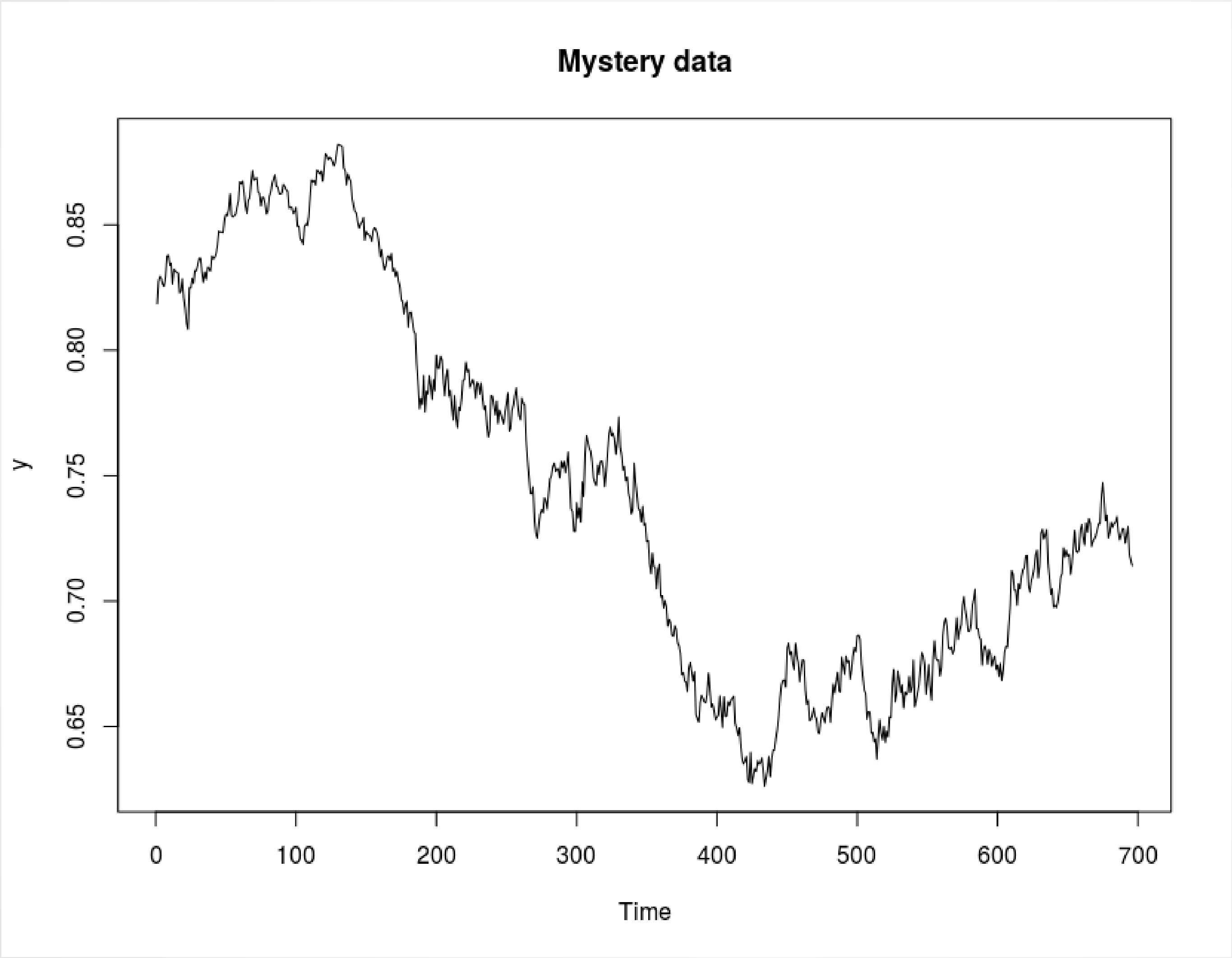
An AR(1) is a Probability Distribution

- The AR(1) is a probability distribution for a sequence of quantities
- It can therefore be used either as a *prior* or as a *sampling distribution* (likelihood) in a Bayesian model.
- In today’s example it will be the latter

An AR(1) is a Probability Distribution

- The sample space is the set of possible curves
- Some curves more probable than others
- A single curve is **one point** in the sample space of the AR(1).

An Application

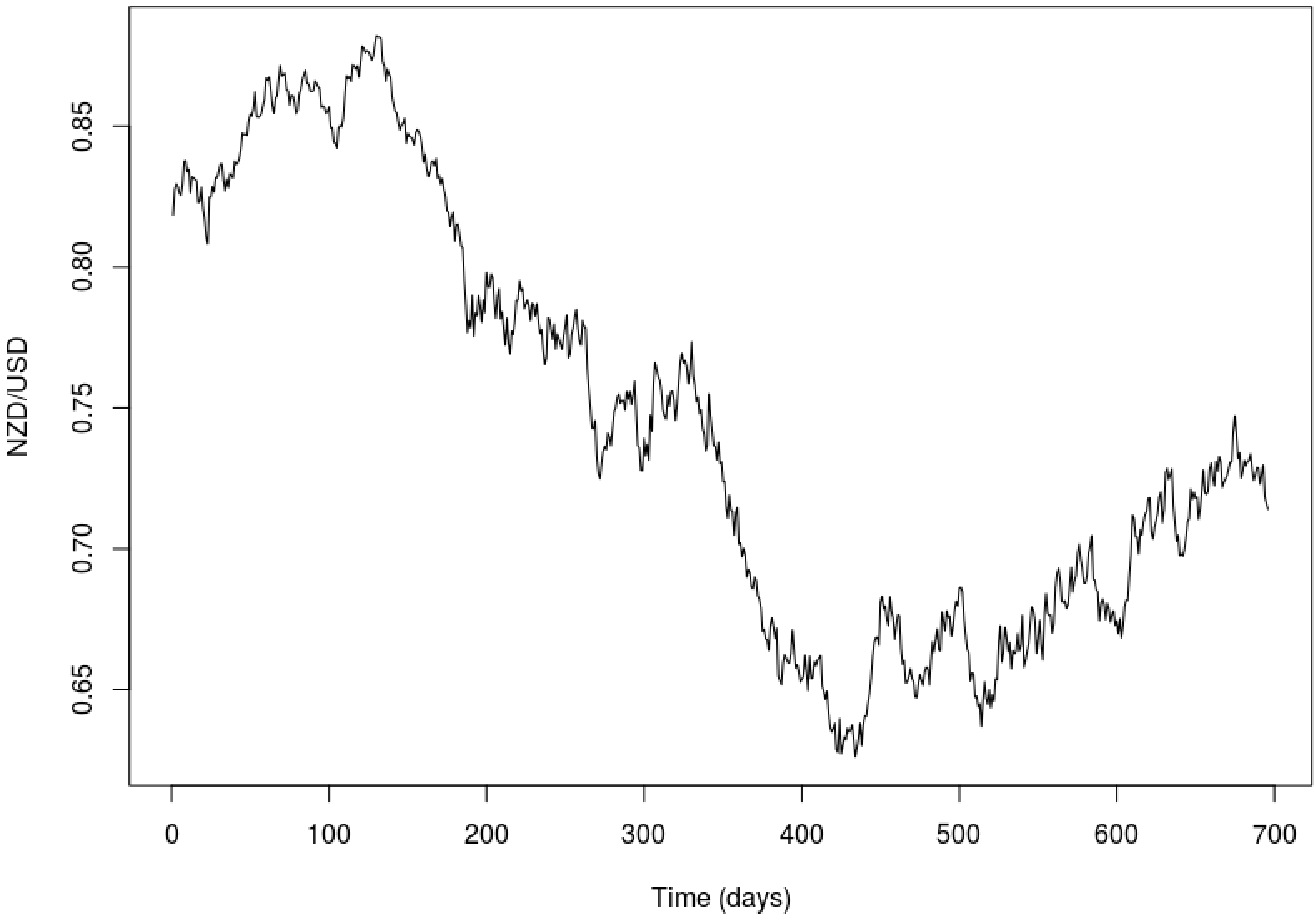


- You can probably guess what it is


ONE DOLLAR


FEDERAL RESERVE NOTE
THE UNITED STATES OF AMERICA
L 11180916 G
ONE DOLLAR

Non-mystery data



Bayesian

- The AR(1) will give us our likelihood, because it's a probabilistic model for data, that depends on parameters

- Let's go Bayesian and estimate the parameters (by calculating the posterior)

Unknown Parameters

mu, alpha, sig

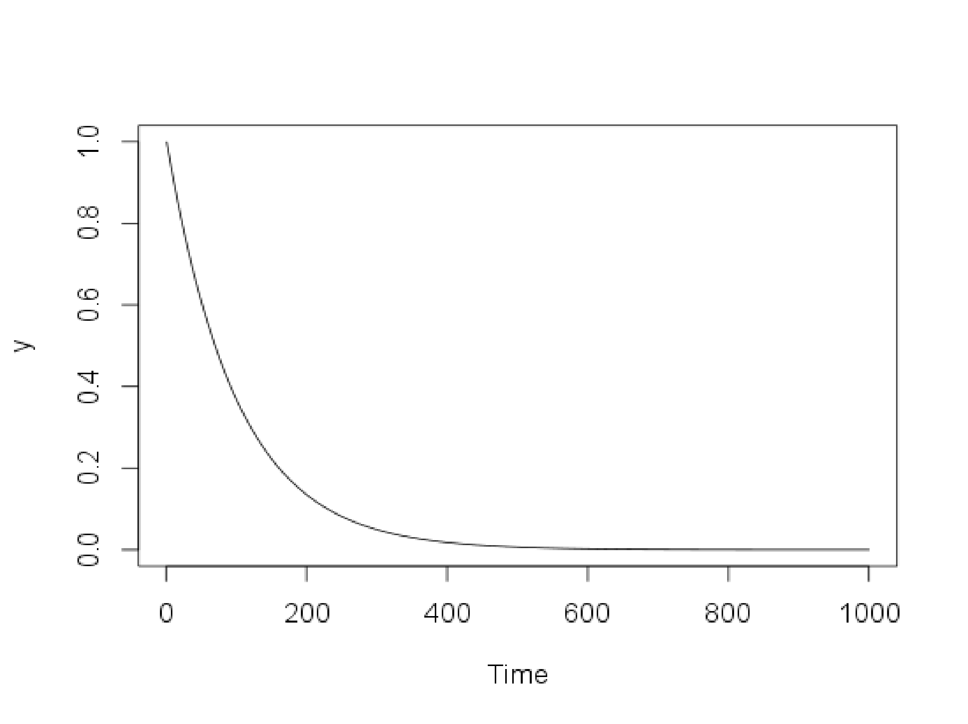
$$y_i = \mu + \alpha(y_{i-1} - \mu) + \epsilon_i$$

$$\epsilon_i \sim \mathcal{N}(0, \sigma^2)$$

A Transformed Parameter

- alpha is a bit of an inconvenient parameter
- Instead define timescale L...like “half life”

$$L = -\frac{1}{\log(\alpha)}$$



JAGS Code – 3 parameters

```
# All fairly vague priors
mu ~ dnorm(0, 1/1000^2)
log_L ~ dunif(-10, 10)
L <- exp(log_L)
log_sigma ~ dunif(-10, 10)
sigma <- exp(log_sigma)

# Define alpha using deterministic node
alpha <- exp(-1/L)
```

JAGS Code – Likelihood

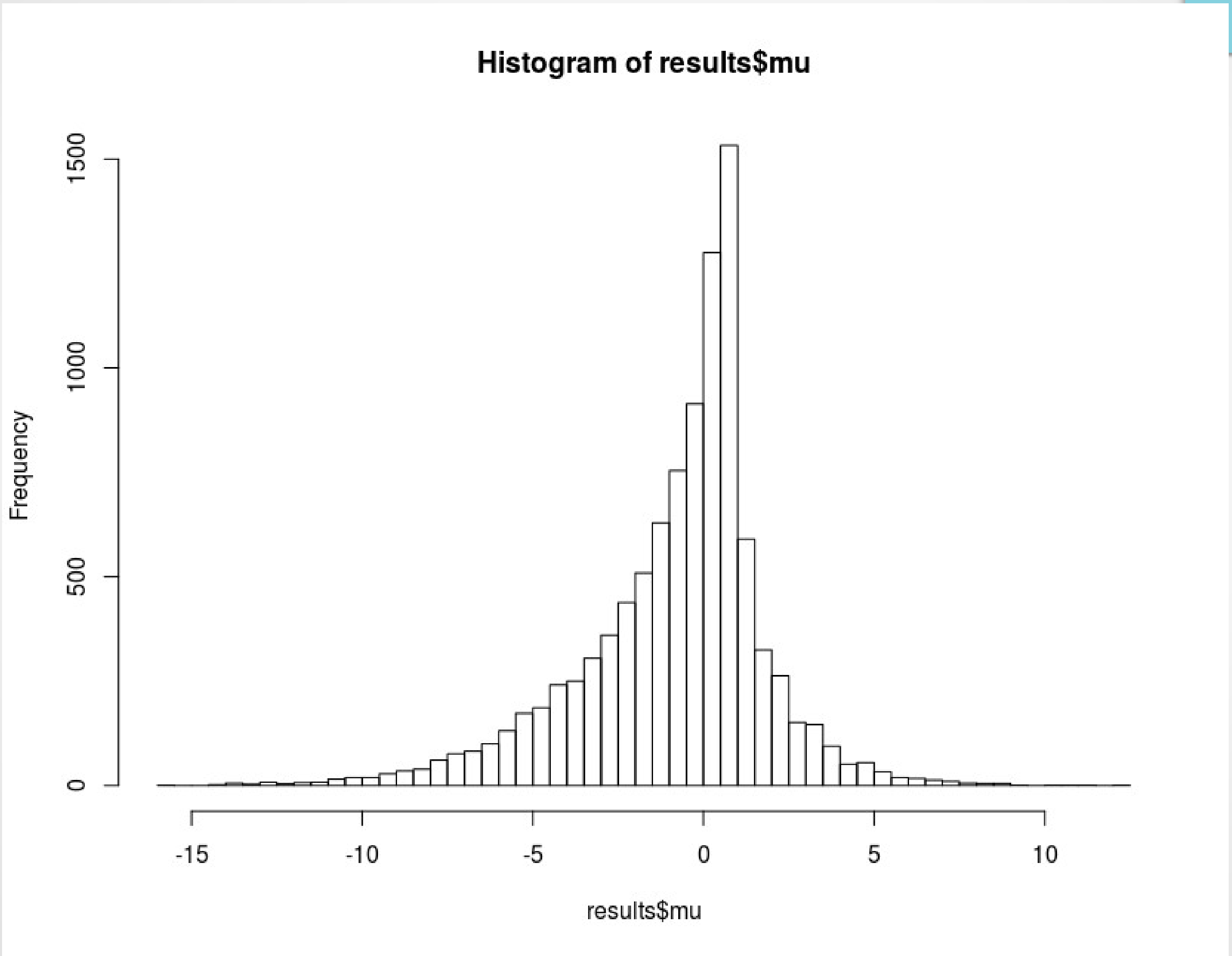
- Looks almost like the R simulation code!
- We are assuming `y[1]` is uninformative

```
for(i in 2:N)
{
   y[i] ~ dnorm(mu + alpha*(y[i-1] - mu), 1/sigma^2)
}
```

Results!

- The posterior distributions (of course!) for the three parameters

Huh?


Histogram of results$mu
Frequency
0
500
1000
1500
-15
-10
-5
0
5
10
results$mu

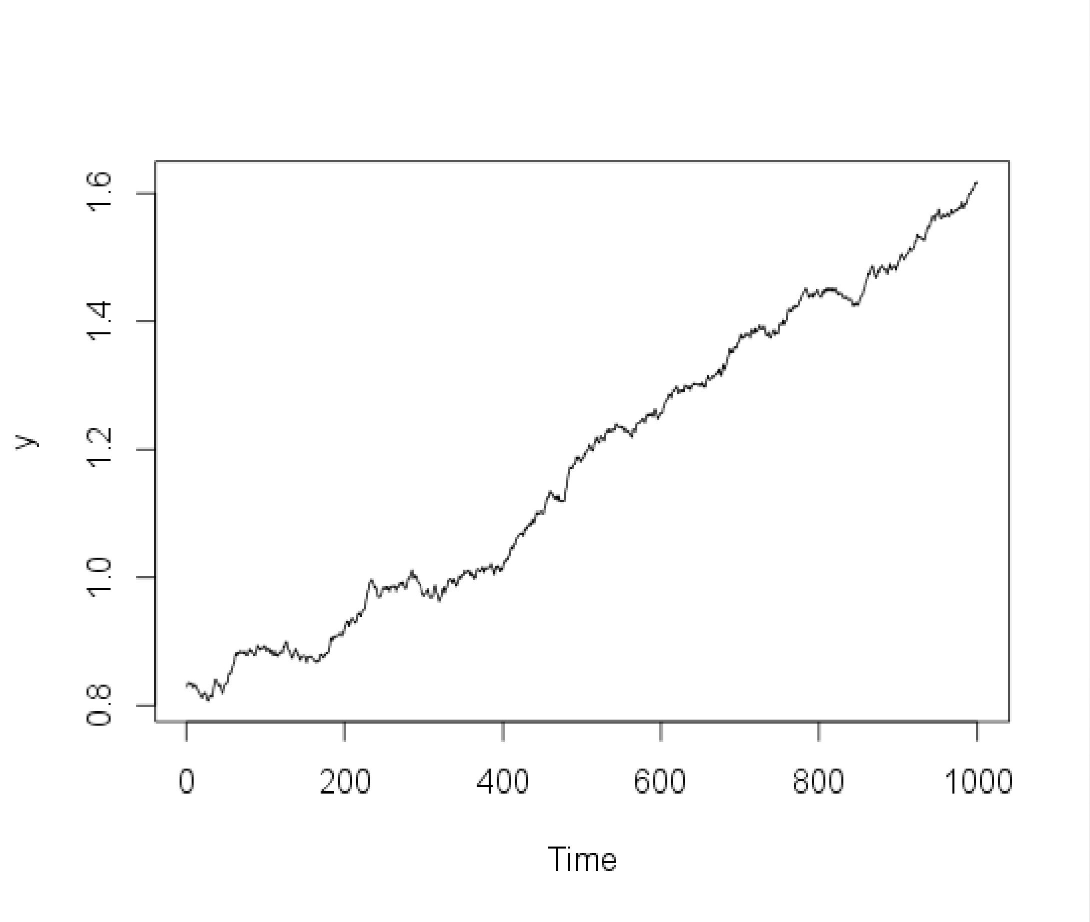
Remember mu

- `mu` is the “mean level” that the NZD “fluctuates around”

- Why the big uncertainty?

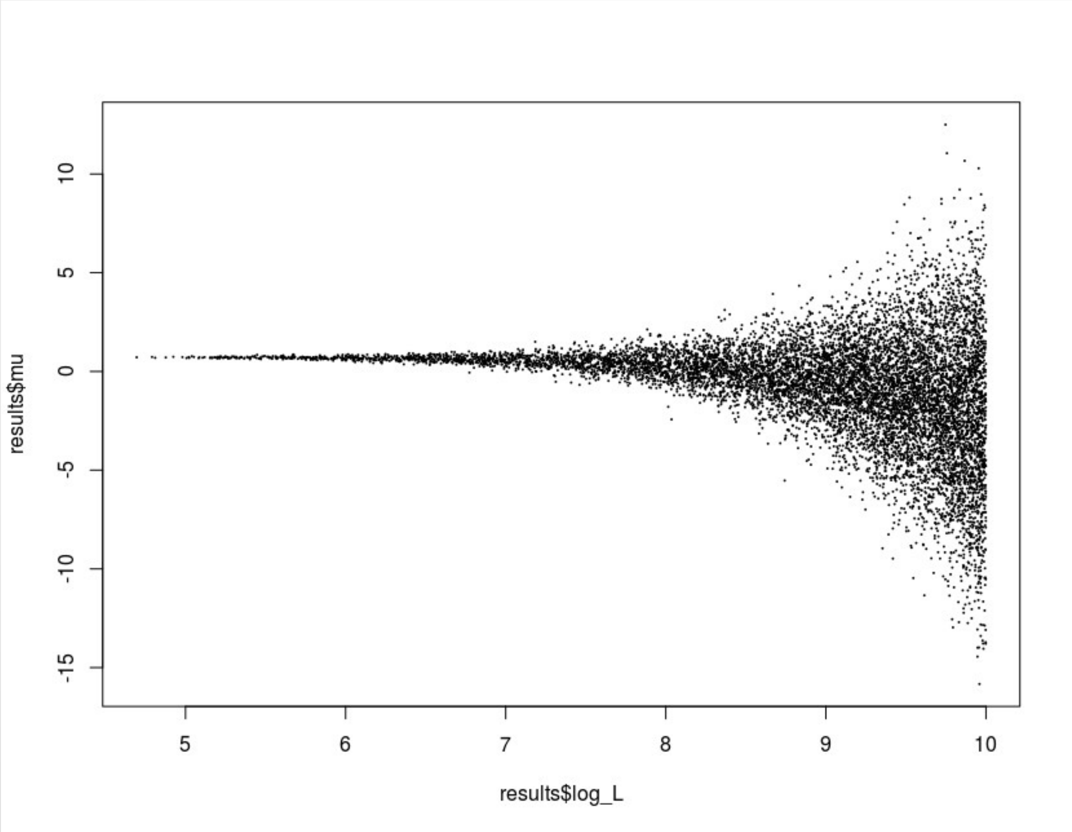
It Makes Sense

- Nothing **in this data set** says that we're not looking at a really small part something that fluctuates a lot!



Joint Posterior, `mu` and `log_L`

```
plot(results$log_L, results$mu, cex=0.1)
```



If the timescale is short, then we have a good measurement of `mu`.

If the timescale is long, then we have basically no idea about `mu`.

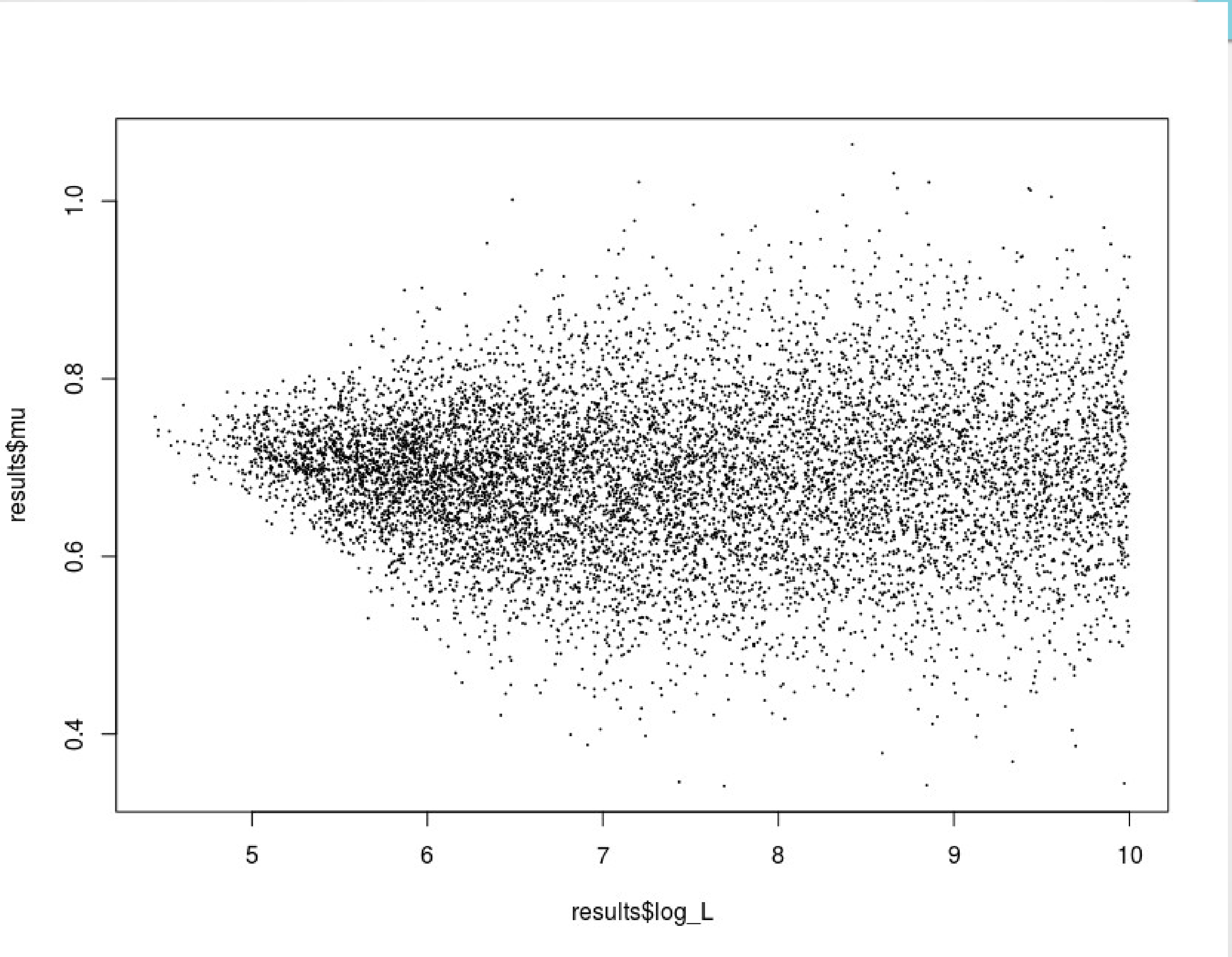
And we can't tell which is true based on a small dataset.

Informative Prior

```
mu ~ dnorm(0.7, 1/0.1^2)
```

- Seems reasonable to me. Could, if you wanted to, take into account external data about the NZD/USD ratio or about other currencies.

That's More Like It



Predicting the Future

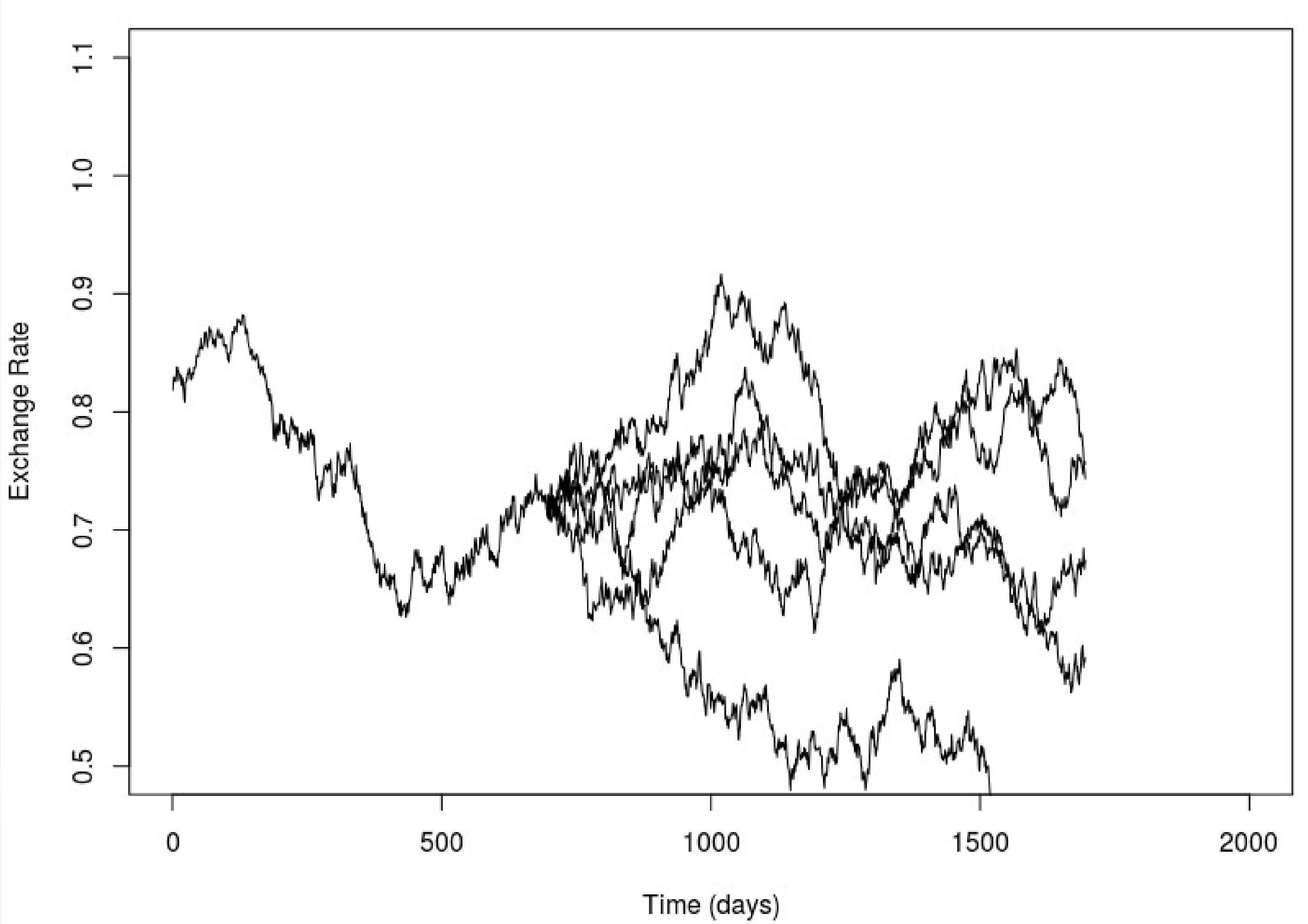
- Same procedure as in previous models.

- Make extra variables, similar to those for the data, but with a different name.

Predicting the Future

```
y_future[1] <- y[N]
for(i in 2:1000)
{
   y_future[i] ~ dnorm(mu +
     alpha*(y_future[i-1] - mu), 1/sigma^2)
}
```

The Money Plot



R code for plot

```
plot(data$y, type='l', xlim=c(0, 2000),
ylim=c(0.5, 1.1), xlab='Time (days)',
ylab='Exchange Rate')

t = seq(data$N, data$N+999)

lines(t, results$y_future[5,])

lines(t, results$y_future[15,])

lines(t, results$y_future[25,])

lines(t, results$y_future[35,])

lines(t, results$y_future[45,])
```

One Year Forecast

- Let's look at the posterior samples for `y_future[366]`

```
results$y_future[, 366]
```

Disclaimer

- I like this example, but simplistic models of financial things can cause **trouble in the real world!**
- The AR(1) model makes quite strong assumptions about what is likely to happen long term, and doesn't know anything about the real causes of variability

Today

- A time series model – the AR(1)
- It's just another kind of probability distribution
- Bayesian works the same always...*different situations just use different probability distributions for the prior and the likelihood*