Stat 565

Forecasting

Feb 4 2016

Charlotte Wickham

stat565.cwick.co.nz

Midterm

Next Thursday in class.

Closed book, no cheat sheet.

Last year’s midterm posted.

- Define the concept of stationarity and describe its importance in time series analysis.
- Define basic stationary time series models: white noise, AR(1) and MA(1).
- Define the autocovariance and autocorrelation functions and derive the autocorrelation function for basic time series models.
- Apply the Box-Jenkins modelling approach to identify, fit, check and forecast SARIMA models for time series data.

library(forecast)

Has some "automatic" ways to select arima models (and seasonal ARIMA models).

```
auto.arima(log(oil$price))
```

Finds the model (up to a certain order) with the lowest AIC. You should still check if a simpler model has almost the same AIC.

Forecasting

Basic Idea: Given an ARMA model, and some past data, we want to predict the future.

Let x_n^{n+k} denote the predictor of x_{n+k} from the values up to x_n.

Technically: We will find a linear function of past values to predict future values that minimizes the prediction mean squared error (one definition of a good predictor).

Use ARIMA model directly

Plug in zero for future Z_t

Plug in conditional expectation for future X_t.

Plug in observed values for past X_t and Z_t .

Example: predict a SARIMA(1,0,0) x $(0,1,1)_{12}$ one step ahead

Your turn

What is the one step ahead prediction for an AR(1) process?

Derive predictor

$$x_n^{n+k} = \sigma^2 \sum_{j=k}^{\infty} \psi_j w_{n+k-j}$$

Skip, see Shumway & Stoffer
section 3.5 if interested

BASIC IDEA: use phi form, best
guess for future white noise is zero.

Show error in prediction is

$$\mathrm{Var}(x_n^{n+k}) = \sigma^2 \sum_{j=0}^{k-1} \psi_j^2$$

Skip, see Shumway & Stoffer
section 3.5 if interested

But we don't know ϕ and θ?

Plug in our estimates and get **approximate** predictions.

These do not take into account the uncertainty in our estimates.

`predict` in R on an `arima` fit.

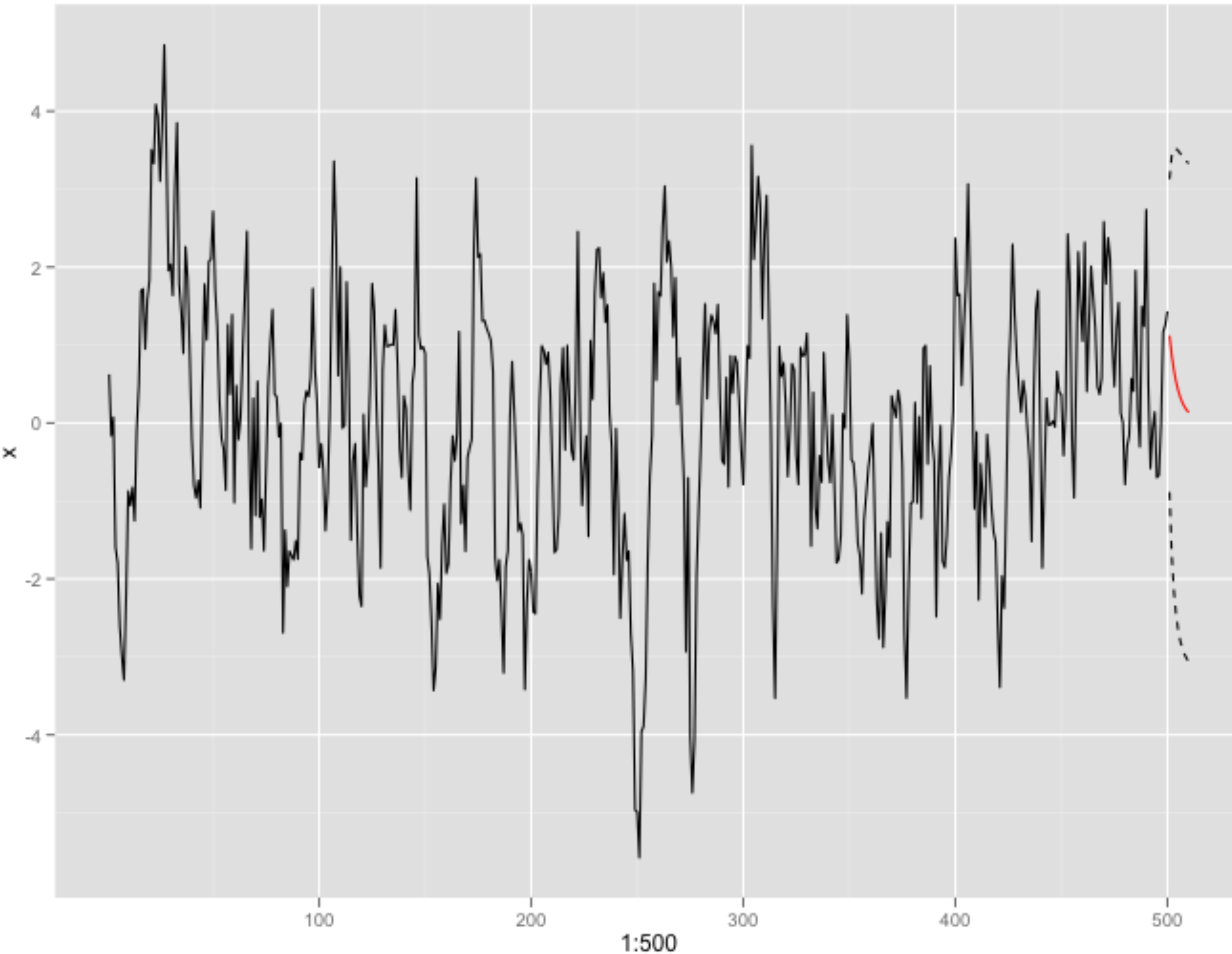
```r
x <- arima.sim(model = list(ar = 0.8), 500)
fit_ar1 <- arima(x, order = c(1, 0, 0 ))
predict(fit_ar1, n.ahead = 10)

predict(fit_ar1, n.ahead = 10)
pred.df <- as.data.frame(predict(fit_ar1,
              n.ahead = 10))
```

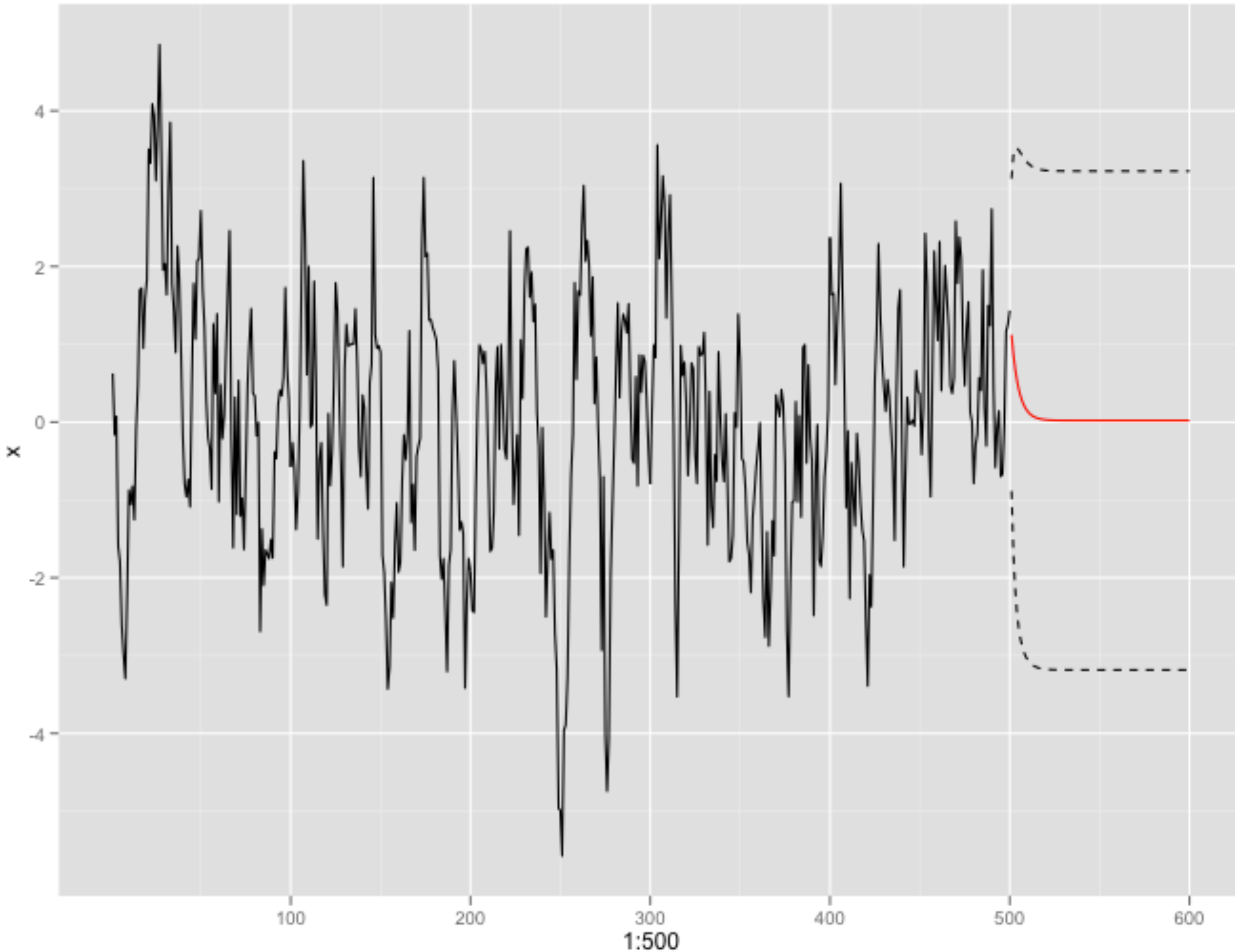
```r
qplot(1:500, x, geom = "line") +
     geom_line(aes(x = 501:510, pred - 2*se), data = pred.df,
         linetype = "dashed") +
    geom_line(aes(x = 501:510, pred + 2*se), data = pred.df,
         linetype = "dashed") +
    geom_line(aes(x = 501:510, pred), data = pred.df, colour = "red")

pred.df.100 <- as.data.frame(predict(fit_ar1, n.ahead = 100))

qplot(1:500, x, geom = "line") +
     geom_line(aes(x = 501:600, pred - 2*se), data = pred.df.100,
         linetype = "dashed") +
    geom_line(aes(x = 501:600, pred + 2*se), data = pred.df.100,
         linetype = "dashed") +
    geom_line(aes(x = 501:600, pred), data = pred.df.100, colour = "red")
```



4
2
0
-2
-4
x
100
200
300
400
500
1:500



4
2
0
-2
-4
x
100
200
300
400
500
600
1:500

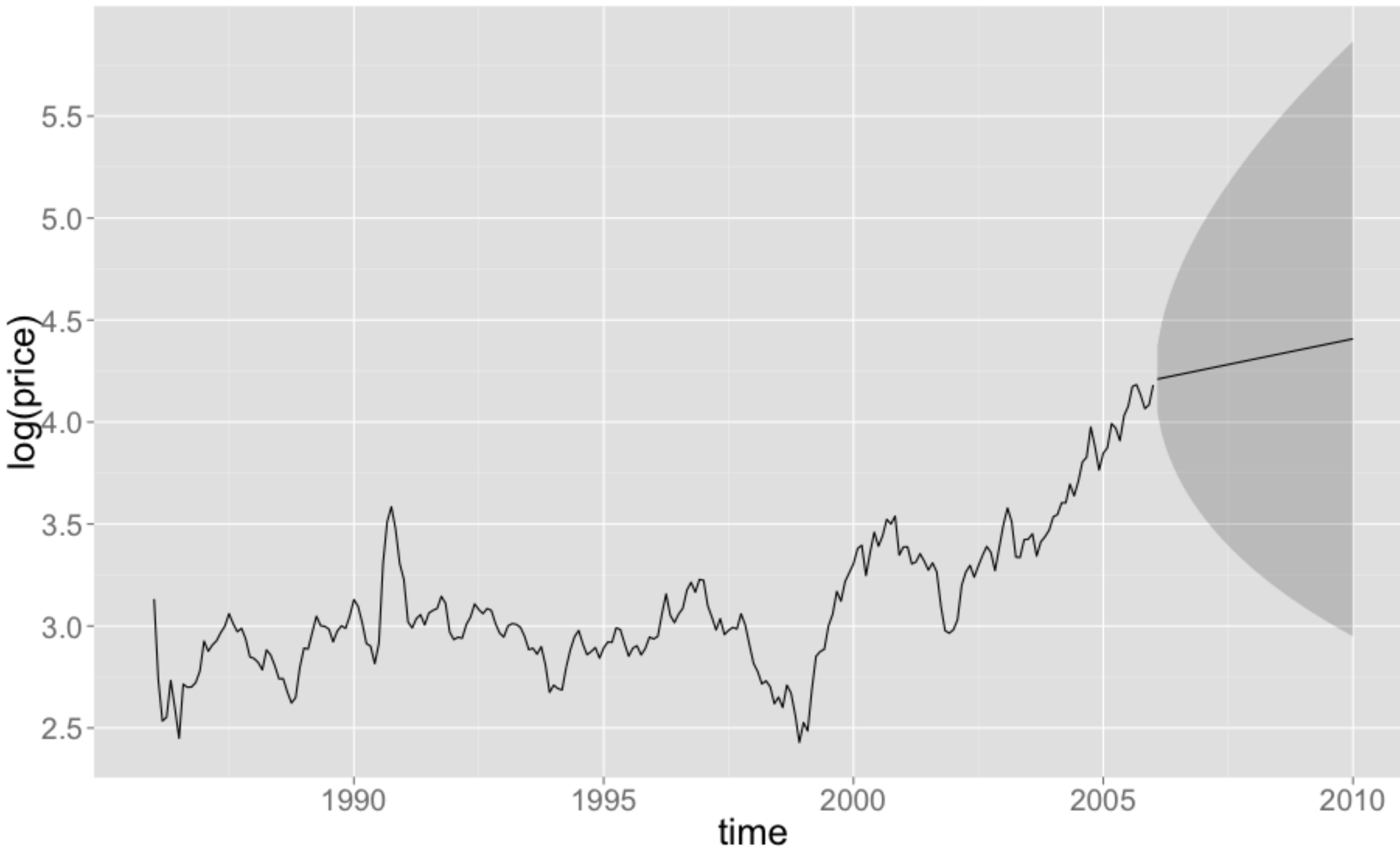
Careful with differencing

To use `predict` to forecast, you need to provide any `xreg` terms you used to fit the model to `newxreg`.

```
(fit_ma1 <- arima(log(oil$price), order = c(0, 1, 1),
 xreg = 1:n))

pred.df <- as.data.frame(predict(fit_ma1, n.ahead = 48,
           newxreg = (n + 1):(n+48)))
pred.df$time <- max(oil$time) + (1:48)/12
```

```
qplot(time, log(price), data = oil, geom = "line") +
  geom_ribbon(aes(ymin = pred- 2*se,
                  ymax = pred + 2*se, y = NULL),
             data = pred.df, alpha = 0.2) +
  geom_line(aes(y = pred), data = pred.df)
```



Some other forecasting approaches

Model a deterministic trend

Exponential smoothing

Holt-Winters

Your turn

$x_t = \beta t + w_t$

$x_t = x_{t-1} + w_t$

Are these series stationary?

What about their first difference?

Trend stationary & difference stationary

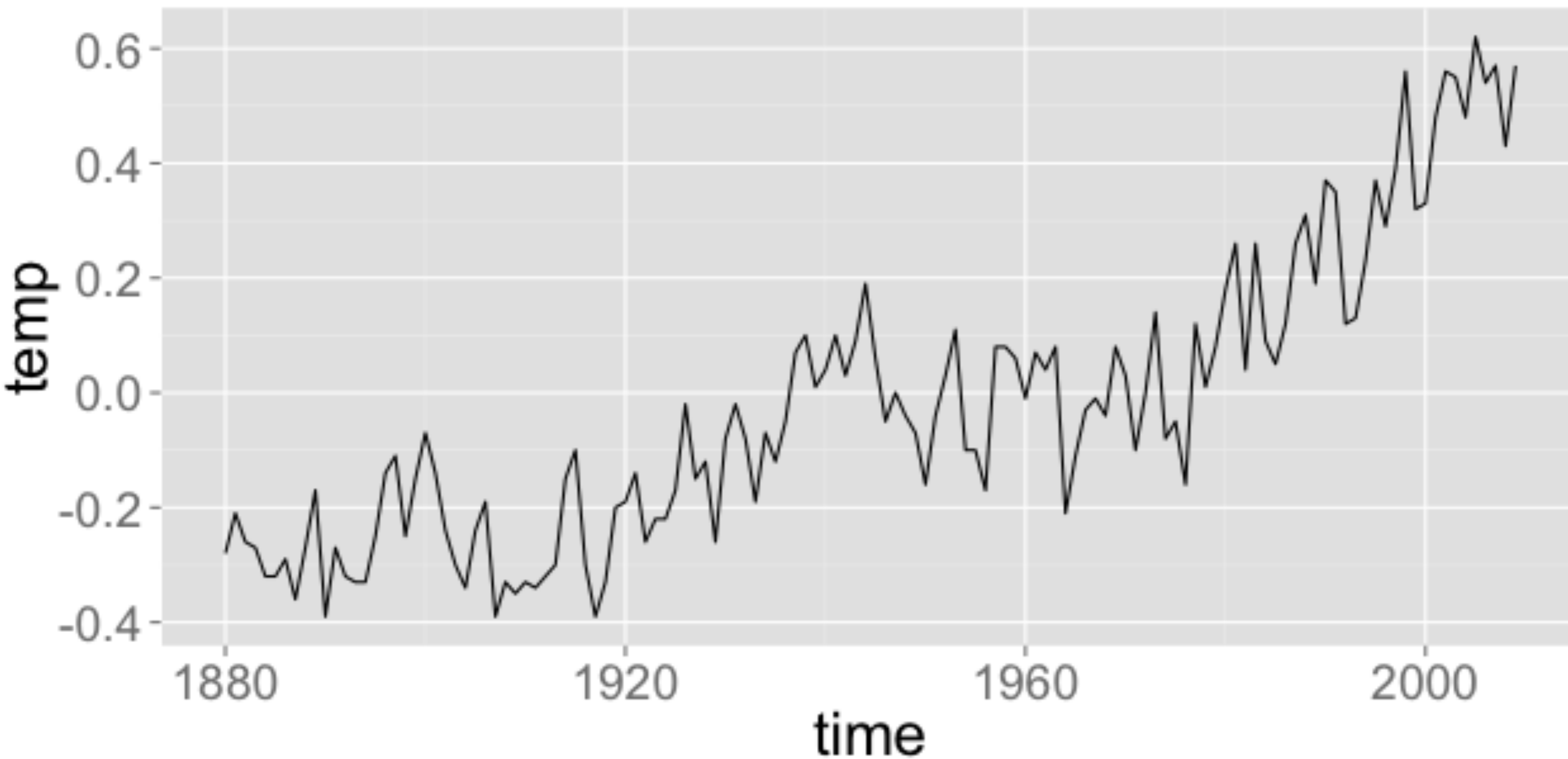
We can make a distinction between processes that are stationary up to a deterministic trend, and processes that are stationary after differencing.

Called *trend stationary* or *difference stationary*.

Or *deterministic trend* and *stochastic trend*.

Unfortunately, you can find both a trend stationary model, and difference stationary model that fits a given series equally well.

Global temperatures



Let's try two approaches

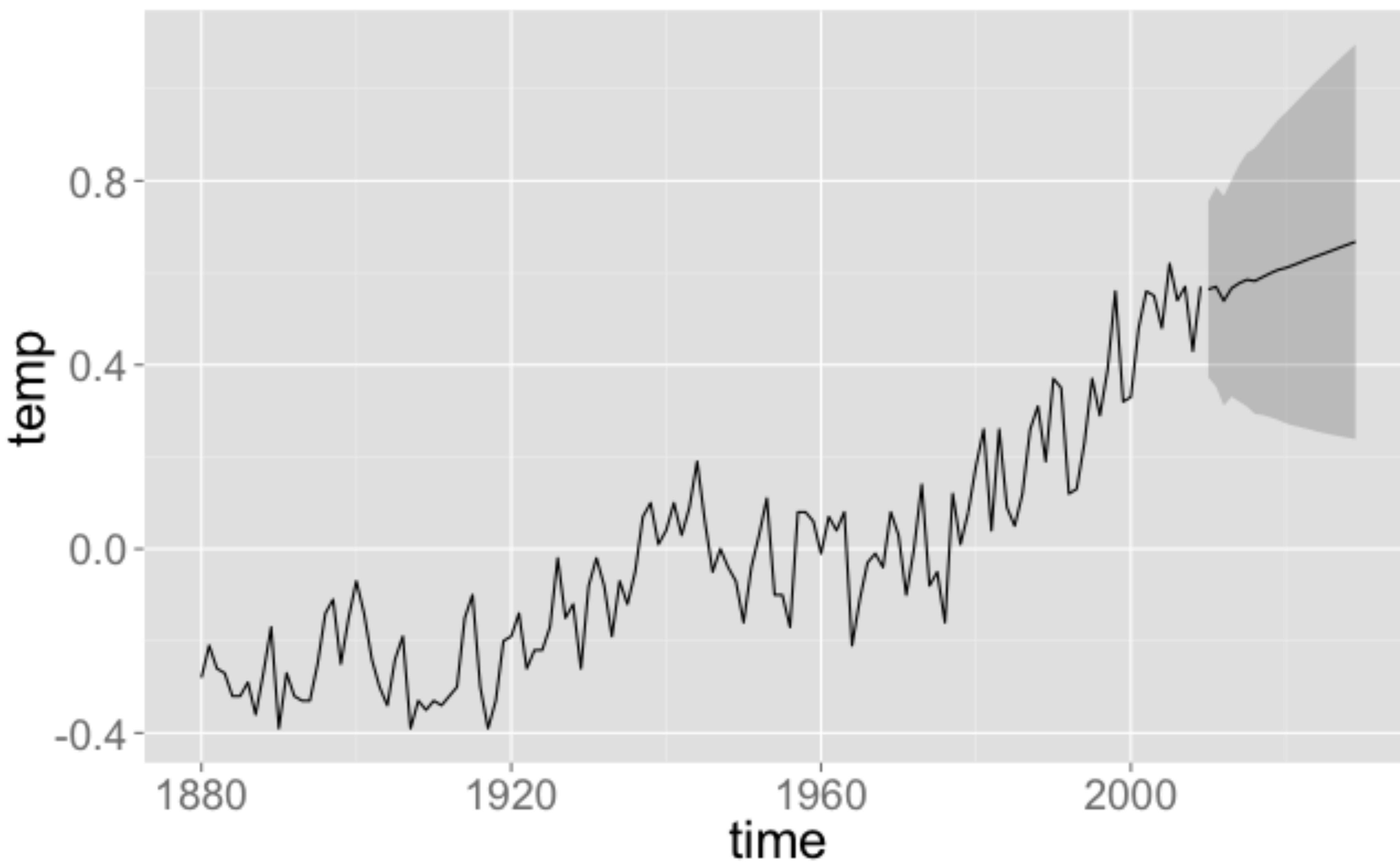
1. Difference for stationarity, then model differenced series as ARMA.

Difference once, then ARMA(3, 0) stochastic trend

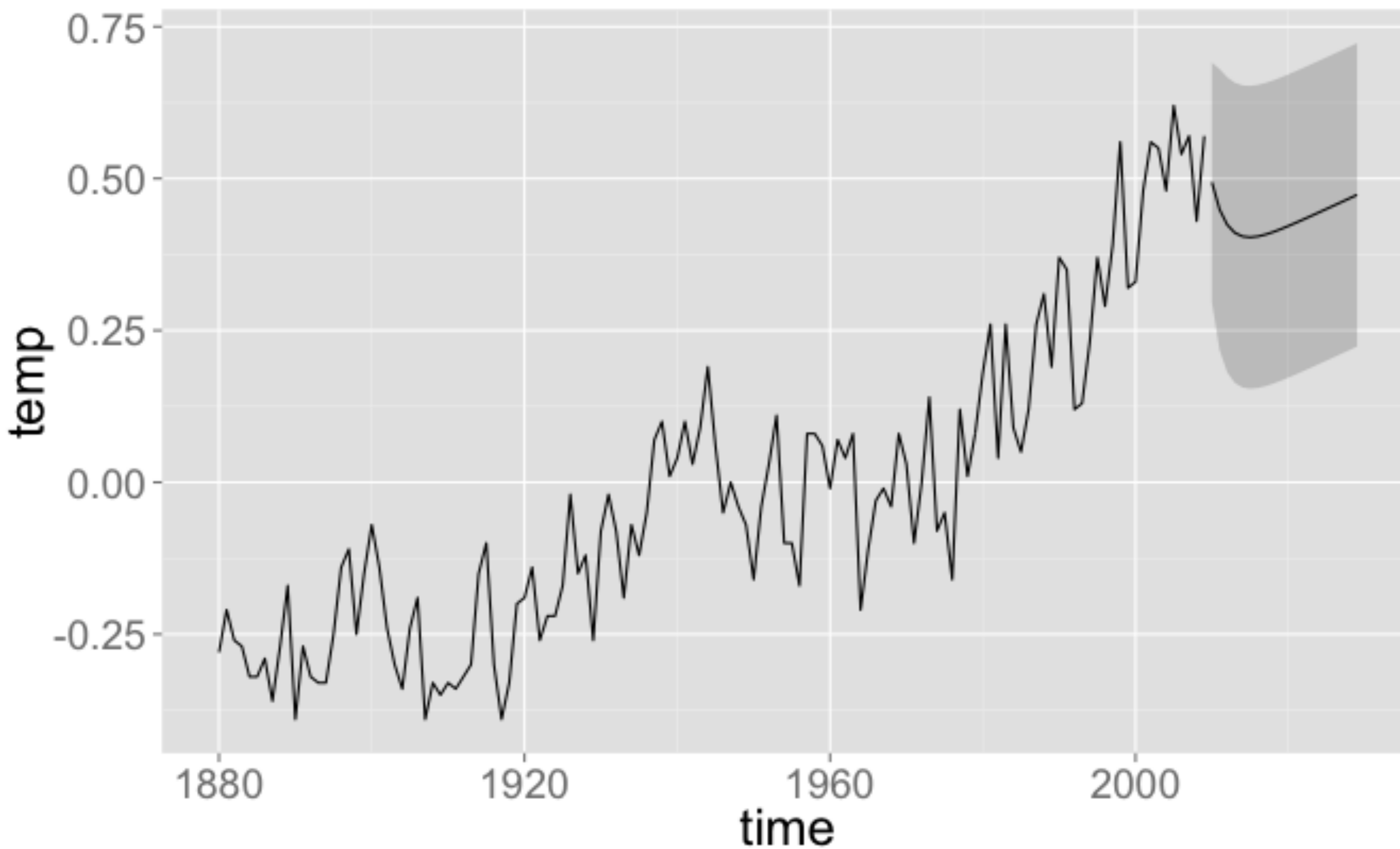
might be something else.

2. Remove linear trend, then model residuals as ARMA. deterministic trend

Linear trend + AR(1)



0.8
0.4
0.0
-0.4
temp
1880
1920
1960
2000
time


0.75
0.50
0.25
0.00
-0.25
temp
1880
1920
1960
2000
time