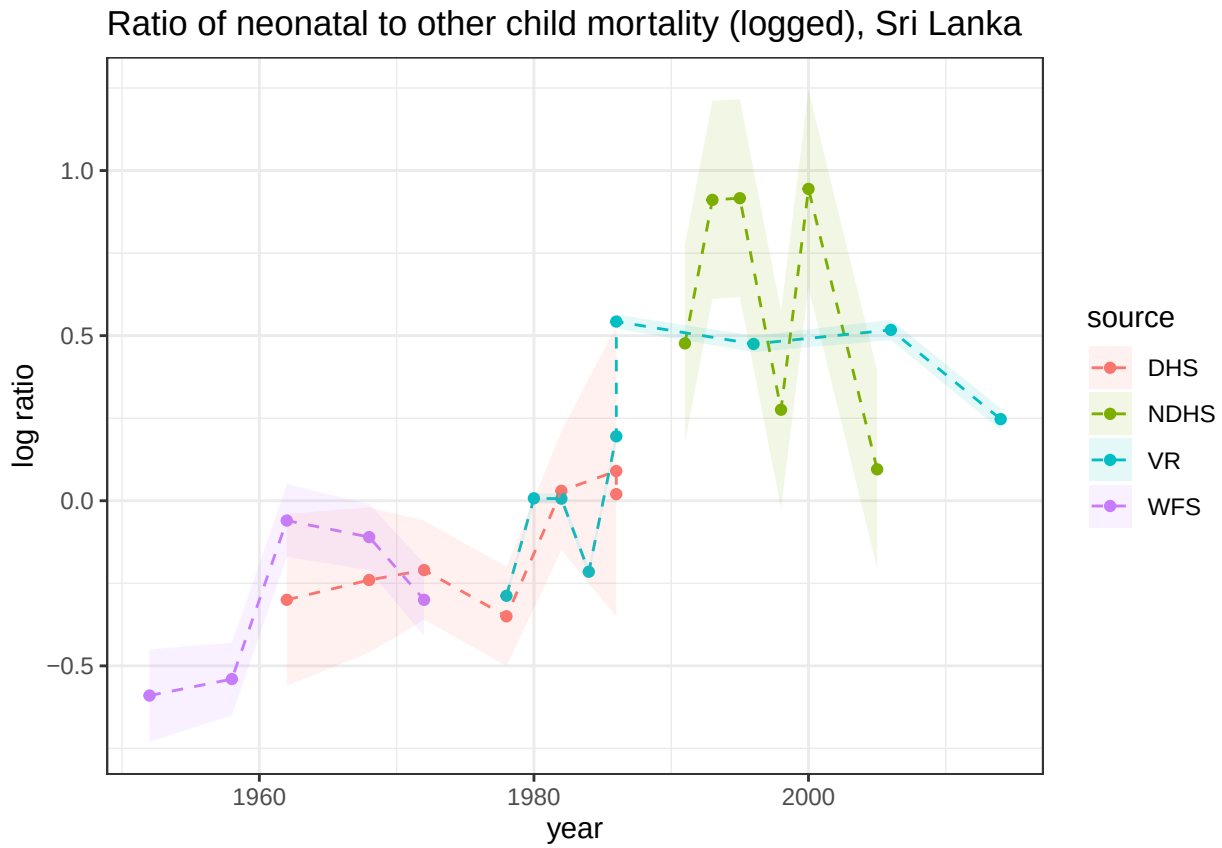
STA2201 Lab #10

Timothy Jordan Regis

22/03/2023

Child mortality in Sri Lanka

In this lab you will be fitting a couple of different models to the data about child mortality in Sri Lanka, which was used in the lecture. Here's the data and the plot from the lecture:



Fitting a linear model

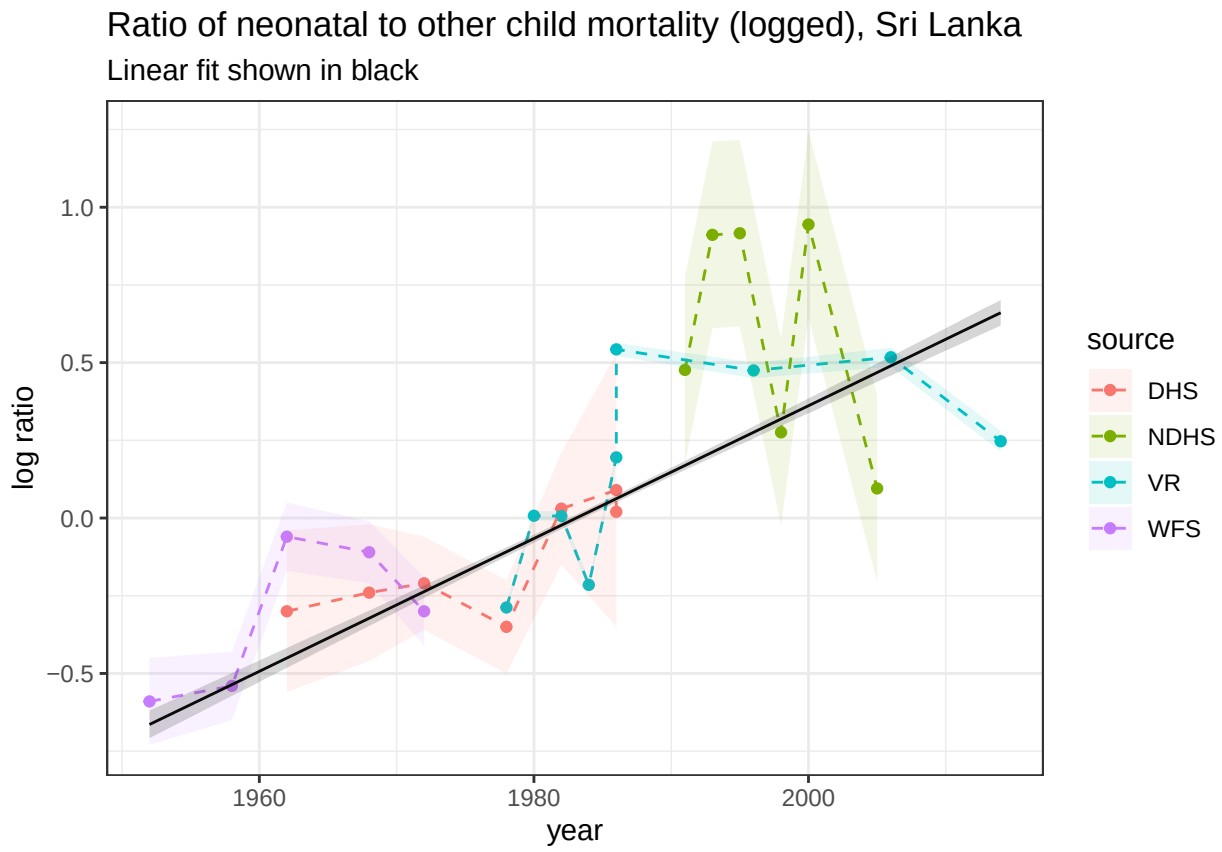
Let's firstly fit a linear model in time to these data. Here's the code to do this:

Extract the results:

```
## # A tibble: 6 x 9
```

```
##       t .variable .value .lower .upper .width .point .interval  year
##   <int> <chr>      <dbl>  <dbl>  <dbl>  <dbl> <chr>  <chr>     <int>
## 1     1 mu        -0.664 -0.708 -0.619   0.95 median qi         1952
## 2     2 mu        -0.643 -0.686 -0.599   0.95 median qi         1953
## 3     3 mu        -0.621 -0.663 -0.579   0.95 median qi         1954
## 4     4 mu        -0.600 -0.640 -0.559   0.95 median qi         1955
## 5     5 mu        -0.579 -0.618 -0.539   0.95 median qi         1956
## 6     6 mu        -0.557 -0.595 -0.519   0.95 median qi         1957
```

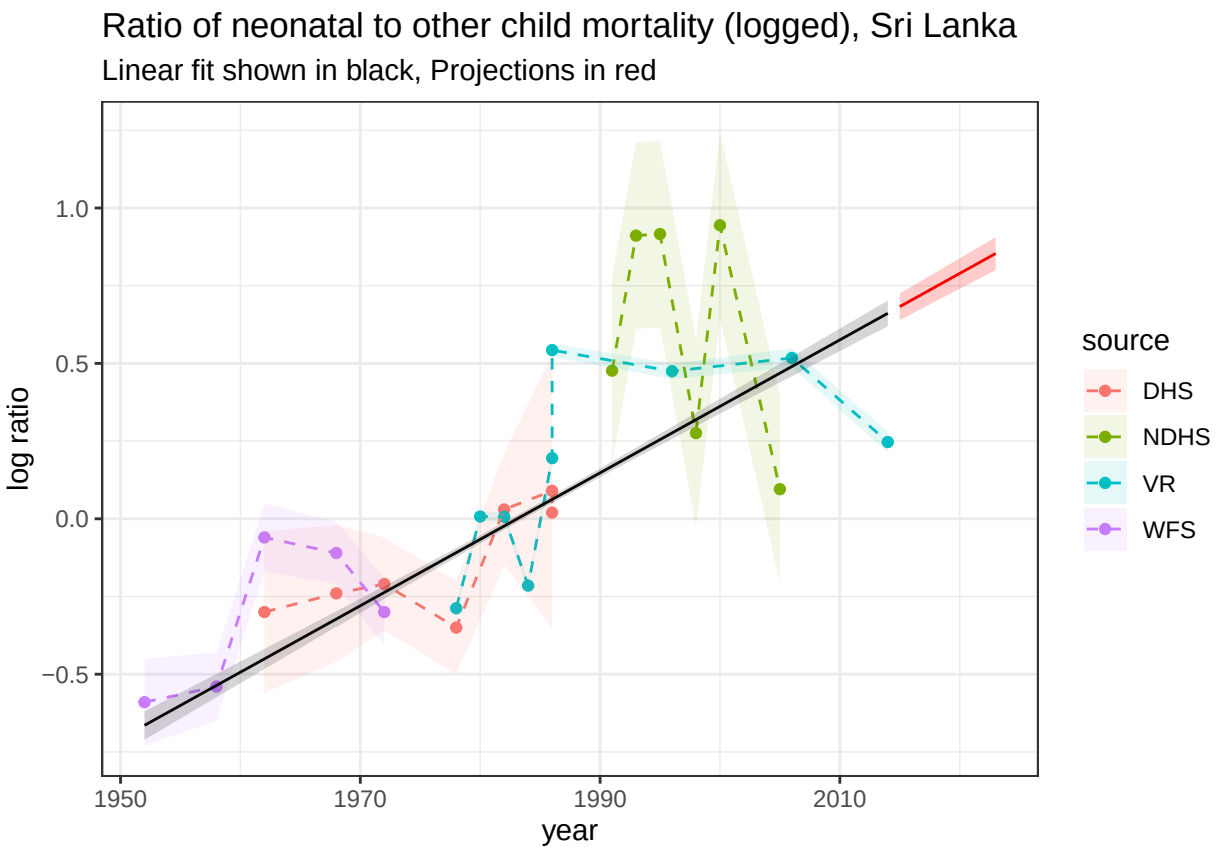
Plot the results:



Question 1

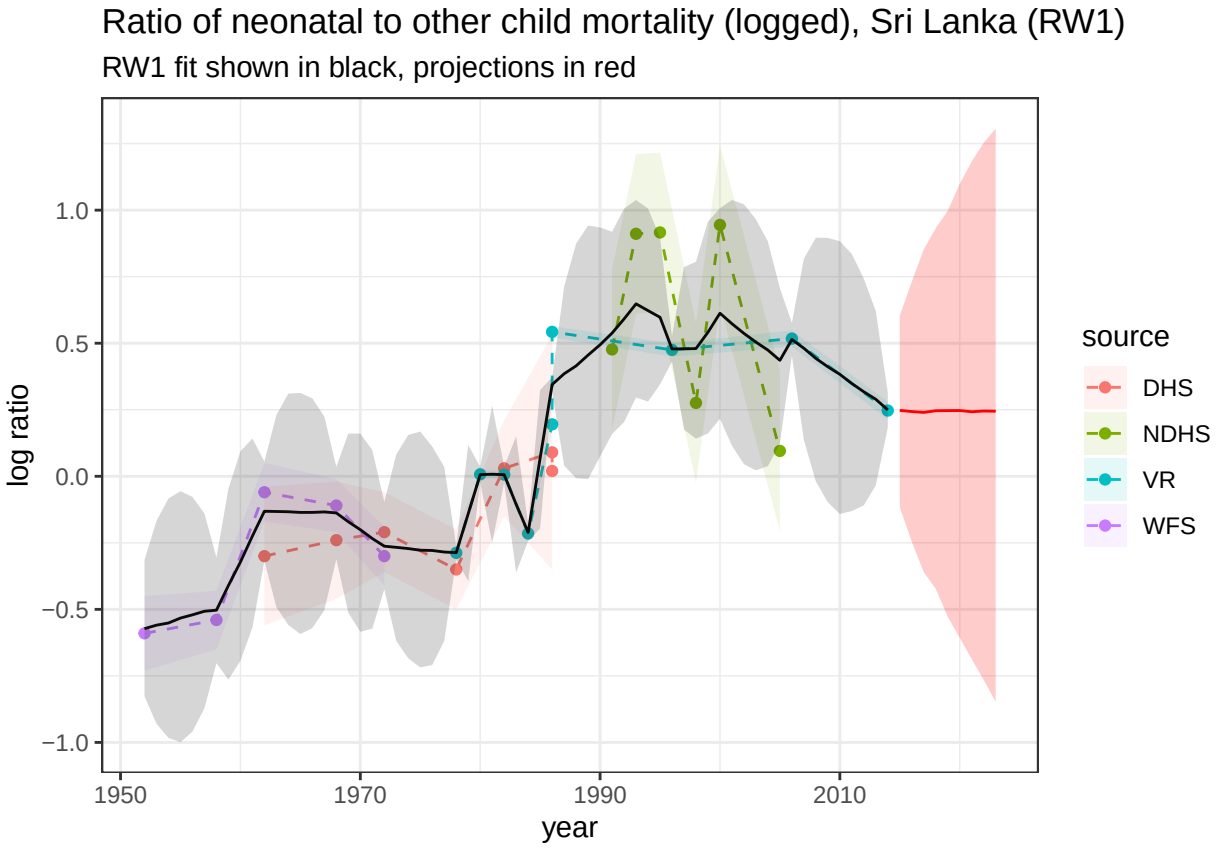
Project the linear model above out to 2023 by adding a `generated quantities` block in Stan (do the projections based on the expected value μ). Plot the resulting projections on a graph similar to that above.

Random walks



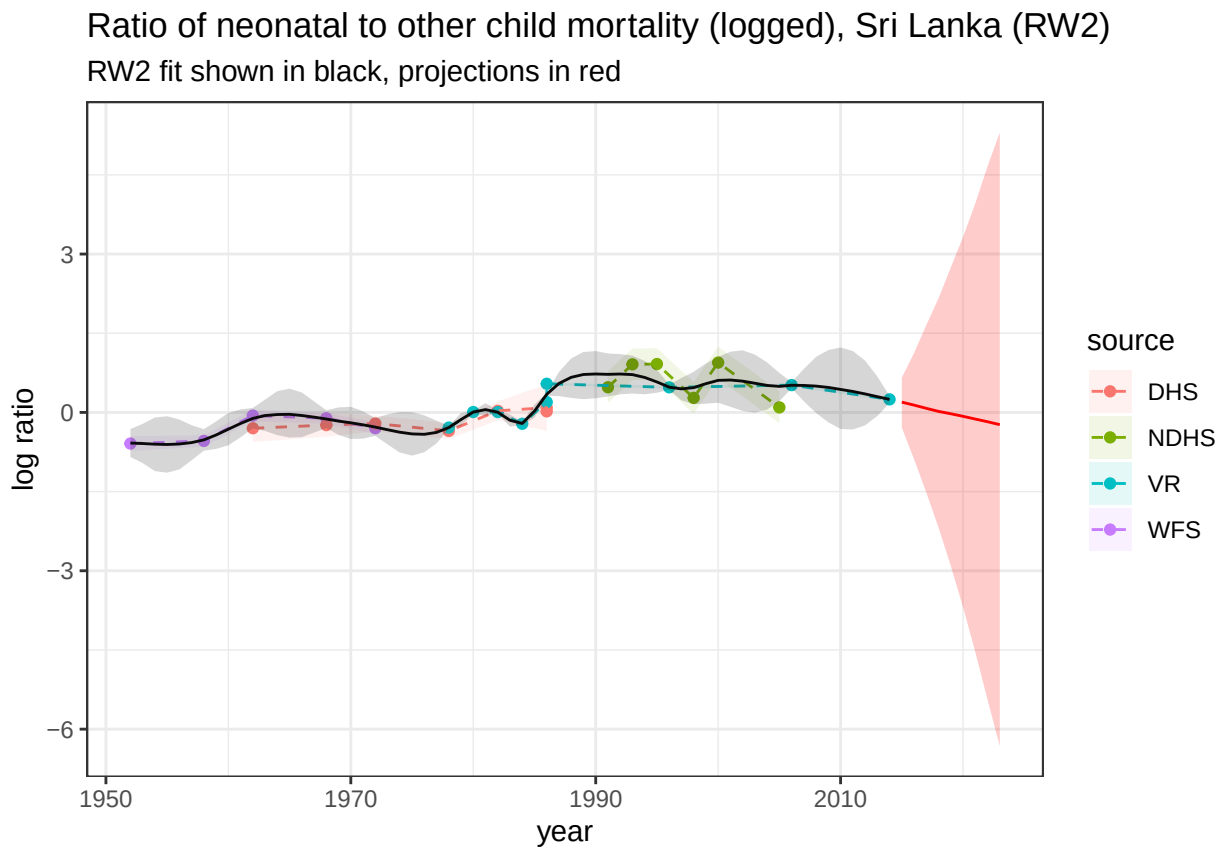
Question 2

Code up and estimate a first order random walk model to fit to the Sri Lankan data, taking into account measurement error, and project out to 2023.



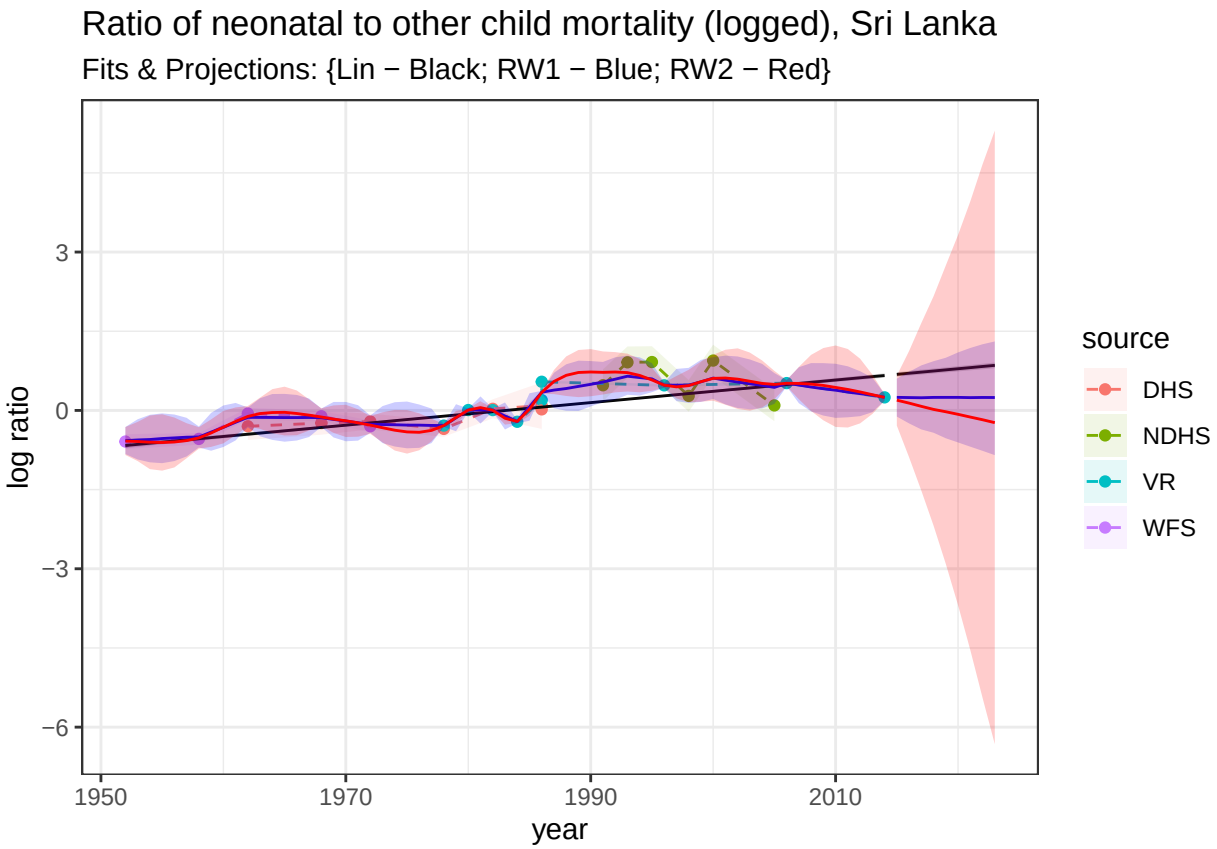
Question 3

Now alter your model above to estimate and project a second-order random walk model (RW2).



Question 4

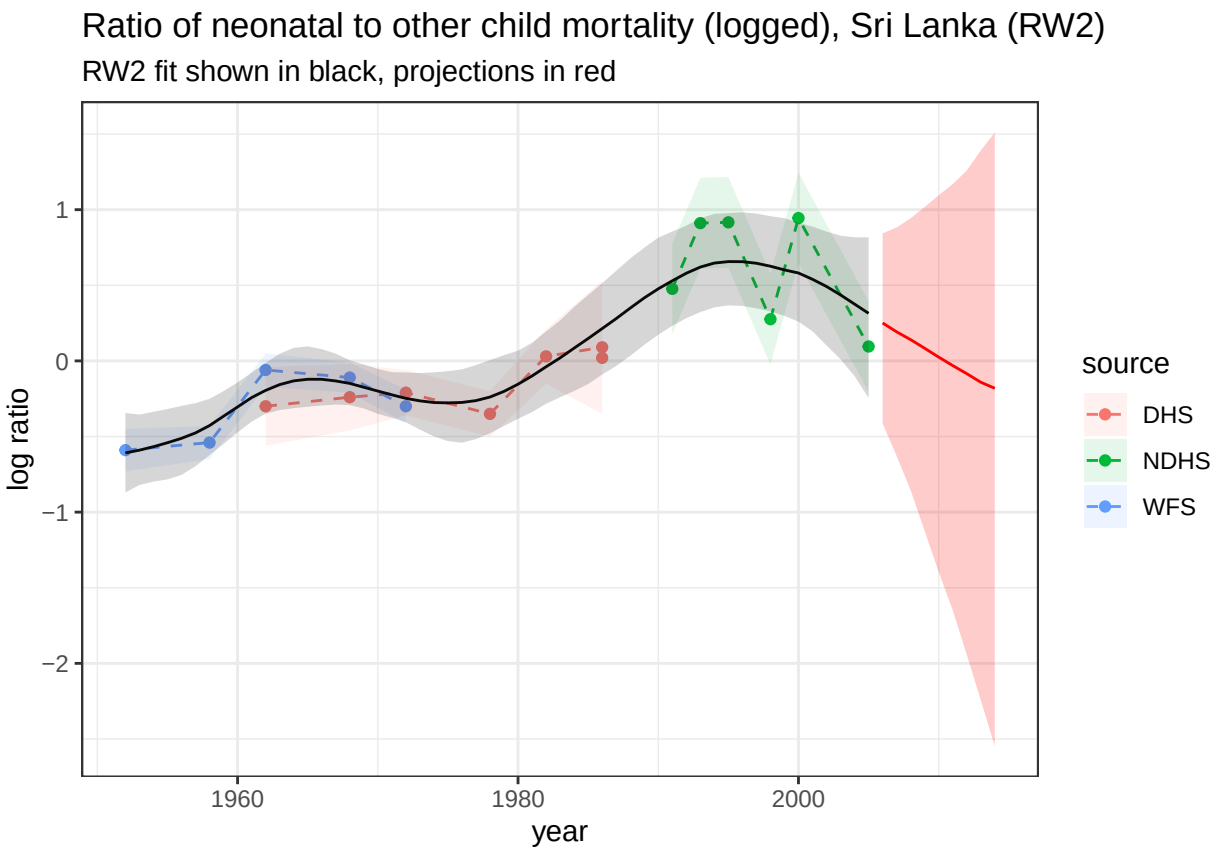
Run the first order and second order random walk models, including projections out to 2023. Compare these estimates with the linear fit by plotting everything on the same graph.



Comparing all 3 on the same graph, we have the linear fit in black, with a slightly higher trend prediction, and very low variation, we have the RW1 fit in blue, with a shallow trend prediction and a moderate amount of variation, and finally we have the RW2 fit in red, with a low trend prediction and an extremely high amount of variation.

Question 5

Rerun the RW2 model excluding the VR data. Briefly comment on the differences between the two data situations.



In the event of the no VR data, we see that we have lost many years with observations as the VR source contained the most recent information. We also see now that our curve as well as our standard errors along our curves, are much smoother than in our previous plots, which had rough curves with a strong degree of variation in the standard errors along it.

Question 6

Briefly comment on which model you think is most appropriate, or an alternative model that would be more appropriate in this context.

Overall, I think the RW1 model does the best out of our options here. Its estimates sit in between those from the linear and RW2 models, and has a standard error which captures both reasonably well. In comparison, the linear model has a higher prediction, which we do think does a better job of identifying the preexisting trend in the data, but it has a very small variation around these estimates, which we worry may be overly confident. On the other hand, the RW2 model has a lower prediction, and an extremely high variation around its estimates, which may be under confident, and is somewhat unhelpful for being conclusive about our results, thus leading us to favor the RW1 model.

The best test would likely be to predict new data and compare it to the real data using a train-test split.