Computational Statistics & Probability

Problem Set 3 - Information Criteria and Interactions
Due: 23:59:59 30.nov.2022

Fall 2022

Instructions

Assignments must be submitted through Canvas. See the course Canvas page for policies covering collaboration, acceptable file formats (`.Rmd` & `.pdf`), and late submissions. Completed assignments must include executable code (`.Rmd`) and a corresponding knitted markdown file (`pdf`). An R Markdown cheat sheet is available.

1. Collider Bias and Information Criteria

Return to the textbook example in §6.3.1, which explores the relationship between age, marriage and happiness.

```
library(rethinking)
d <- sim_happiness( seed=1515 , N_years=1000)
d2 <- d[ d$age>17 , ]  # only adults
d2$A <- (d2$age - 18) / (65 - 18)
d2$mid <- d2$married + 1
precis(d)
```

Compare the two models, `m6.9` and `m6.10`, using both PSIS and WAIC.

```
# Set up. Here are the two models from Chapter 6:
m6.9 <- quap(
  alist(
    happiness ~ dnorm( mu , sigma ),
    mu <- a[mid] + bA*A,
    a[mid] ~ dnorm( 0 , 1),
    bA ~ dnorm( 0 , 2 ),
    sigma ~ dexp( 1 )
  ) , data = d2 )

m6.10 <- quap(
  alist(
    happiness ~ dnorm( mu , sigma ),
    mu <- a + bA*A,
    a ~ dnorm( 0 , 1 ),
    bA ~ dnorm( 0 , 2 ),
    sigma ~ dexp( 1 )
  ) , data = d2)
```

a) Which model is expected to make better predictions according to these information criteria?

```
# Compare by PSIS:
compare( m6.9, m6.10 , func=PSIS)
```

```
##           PSIS       SE    dPSIS     dSE    pPSIS       weight
## m6.9  2771.068 37.13856   0.0000      NA 3.752925 1.000000e+00
## m6.10 3102.028 27.71311 330.9597 34.1829 2.415411 1.358367e-72
```

```r
# Compare by WAIC
compare( m6.9, m6.10 , func=WAIC)
```

```
##           WAIC       SE    dWAIC      dSE    pWAIC       weight
## m6.9  2771.393 36.93196   0.0000       NA 3.960565 1.000000e+00
## m6.10 3101.906 27.68192 330.5132 34.02321 2.330043 1.698146e-72
```

```r
# Model m6.9 is expected to make better predictions than m6.10 by both information
# criteria. However, m6.9 is the incorrect causal model whereas m6.10 is the
# correct causal model. These two models illustrate that we should NOT use
# PSIS or WAIC to choose among models unless we have, ex ante, a clear understanding
# of the generative causal model.
```

b) On the basis of the causal model, how should you interpret the parameter estimates from the model preferred by PSIS and WAIC?

```r
# Because we know exactly how the data is simulated, we know that there is
# no association between age and happiness.  And this is what we see in m6.10.
# However, with the introduction of marital status in m6.9, we see that age is
# negatively associated with happiness.  This is a collider bias: unconditionally
# there is no association between happiness and age, but conditioning on the common
# effect, marriage status, induces a negative association between age and happiness.

# Consider the coefficients of m6.9:
precis( m6.9, depth=2 )
```

```
##             mean         sd       5.5%      94.5%
## a[1]  -0.1947491 0.06521259 -0.2989714 -0.0905268
## a[2]   1.2161249 0.08876809  1.0742563  1.3579934
## bA    -0.7325968 0.11708179 -0.9197161 -0.5454775
## sigma  1.0199308 0.02325838  0.9827594  1.0571022
```

```r
# We can only interpret these paramter estimates wrt the causal model, which is:
#
#    H -> M <- A
#
# where H is happiness, A age, and M marriage.

# The parameter bA is a collider bias: there is only a conditional association,
# not an actual causal effect.  The parameters a[1] and a[2] are intercepts for
# unmarried and married, respectively.  However, marriage does not influence
# happiness but instead is a consequence of happiness.  So, marriage does not
# accurately estimate the effect of marriage on happiness. Instead, these
# parameters measure the association between marriage and happiness. But the
# estimate includes a bias because the model also includes age.  To see this
# consider the following model, which stratifies happiness by marriage status
# but ignores age:


m6.9.1 <- quap(
  alist(
    happiness ~ dnorm( mu , sigma ),
    mu <- a[mid],
```

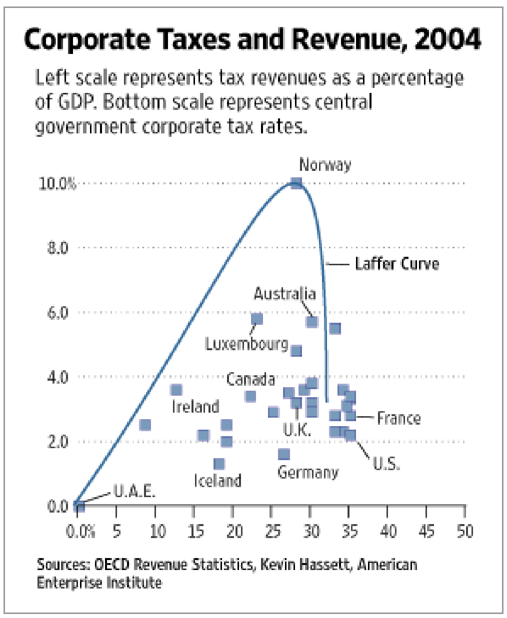
```
    a[mid] ~ dnorm( 0 , 1),
    sigma ~ dexp( 1 )
  ) , data = d2 )
precis( m6.9.1, depth=2 )
```

```
##              mean         sd       5.5%      94.5%
## a[1]  -0.5050091 0.04321821 -0.5740802 -0.4359381
## a[2]   0.7667104 0.05325169  0.6816039  0.8518168
## sigma  1.0409067 0.02373597  1.0029720  1.0788413
```

```
# The estimates for a[1] and a[2] are different without age.  Thus, all parameters
# of m6.9 are non-causal associations.
```

2 Laffer Curve

In 2007 *The Wall Street Journal* published an editorial arguing that raising corporate tax rates increases government revenues only to a point, after which higher tax rates produce less revenue for governments. The editorial included the following graph of corporate tax rates in 29 countries plotted against tax revenue, over which a Laffer curve was drawn.



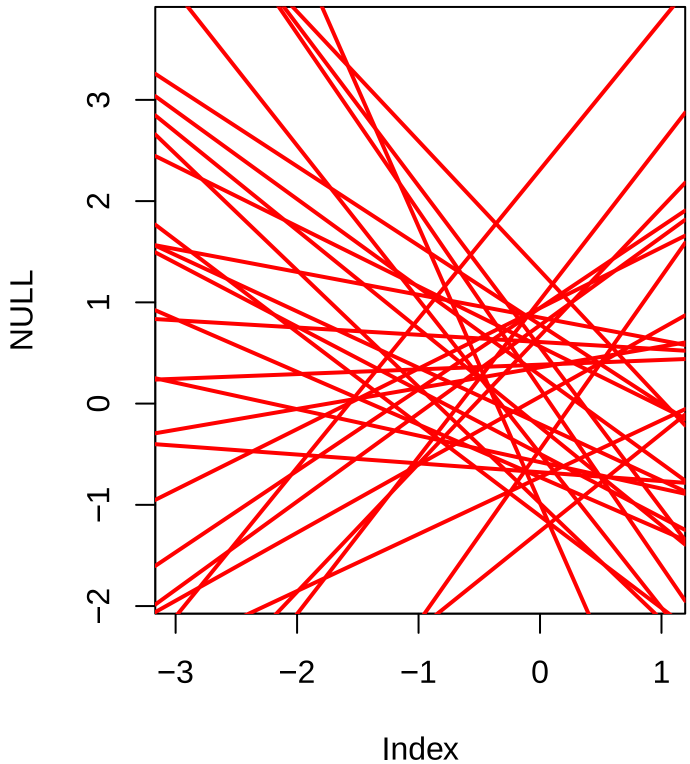
The data used in this plot are available in the `rethinking` package.

```
library(rethinking)
data(Laffer)
d <- Laffer
precis( d )
```

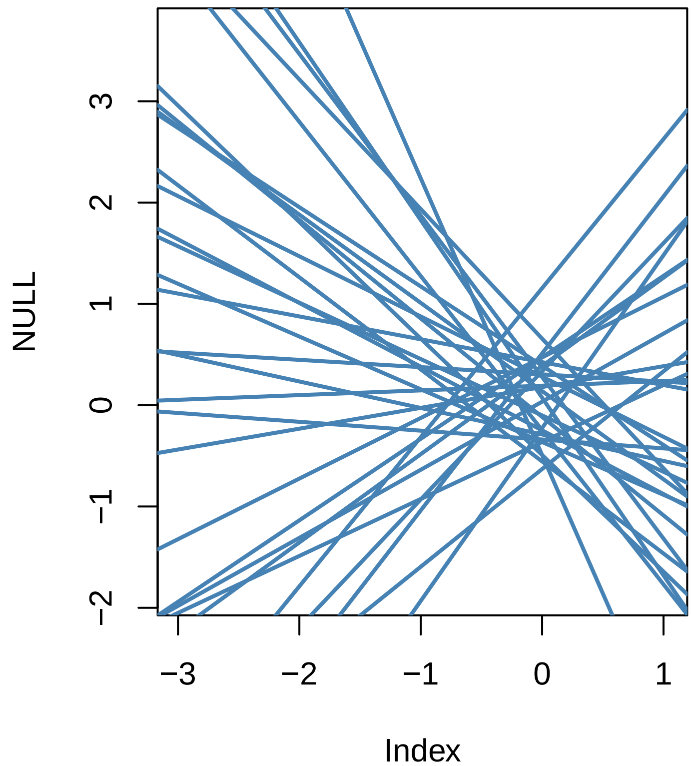
a) Using this data, fit a basic regression that uses tax rate to predict tax revenue. Simulate and justify your priors.

```
# First, we standardize our variables
d$tax <- standardize( d$tax_rate )
d$rev <- standardize( d$tax_revenue )

# Next, we simulate priors
set.seed(303)
```

```r
n <- 30
a <- rnorm(n,0,1)
b <- rnorm(n,0,1)
# blank(bty="n")
plot( NULL , xlim=range(d$tax) , ylim=range(d$rev) )
for ( i in 1:n ) abline( a[i] , b[i] , lwd=2 , col="red" )
```




```r
# It is critical that you NOT pick priors to fit the data per se; rather, you
# should pick priors that encode your prior knowledge before seeing data.

# The relationship between tax rates and revenue is very complicated. The notion
# that there is some point after which raising taxes yields diminishing returns is
# an old idea.  (That is, the curve is not Laffer's contribution; his contribution
# was his analysis of why.)  It is unclear whether such a phenomenon appears in the
# range (0 to 50%) that modern economies tax corporations, however. Thus, we shall
# stick with the weak N(0,1) prior on the slope intercept, `b`, affording the
# the possibility of a positive or negative slope

# A deviation of N(0,1) around the mean appears excessive, however.  So, I introduce
# some mild regularization with a N(0,0.5) prior on `a`:
#
set.seed(303)
n <- 30
a <- rnorm(n,0,0.5)
b <- rnorm(n,0,1)
# blank(bty="n")
plot( NULL , xlim=range(d$tax) , ylim=range(d$rev) )
for ( i in 1:n ) abline( a[i] , b[i] , lwd=2 , col="steelblue" )
```




```
#
# This is still a very weak set of priors.  It only encodes the background
# assumption that country tax revenue rates (as percentage of GDP) do not vary
# one standard deviation from the average. Hence, the intercept term `a` has a
# slightly tighter prior.  Otherwise, the slope prior `b` allows for positive and
# negative relationships between rate and revenue.

# Note that the reasons given for a prior can be interrogated and priors changed
# in response to criticism.

# Next, compute the average tax rate (is zero w/ standardized variables):
xbar <- mean(d$tax)

# a simple linear regression model:
m2 <- quap(
  alist(
    revenue ~ dnorm( mu , sigma ),
    mu <- a + b*(tax - xbar),
    a ~ dnorm(0,0.5),
    b ~ dlnorm(0,1),
    sigma ~ dunif( 0 , 2 )
  ), data=list(revenue=d$rev,tax=d$tax) )
precis( m2 )
```

```
##                mean        sd        5.5%     94.5%
## a     0.0000363016 0.1633871 -0.26108782 0.2611604
## b     0.3299587104 0.1534237  0.08475806 0.5751594
## sigma 0.9309746159 0.1222327  0.73562319 1.1263260
```

```r
# Define a sequence for computing MAP estimate (mu) and PI intervals:
rate.seq <- seq(from=-3.5, to=3.5, length.out=50)

# use link to compute mu for each sample from the posterior
# and for each weight in `weight.seq`
mu <- link( m2 , data=data.frame(tax=rate.seq) )

# summarize the distribution of mu
mu.mean <-apply( mu , 2, mean )  # mean of each column (axis '2') of the
# matrix mu

# compute the 89% PI credibility interval
mu.PI <- apply( mu, 2, PI, prob=0.89 )
```

b) Now construct and fit *any* curved model you wish to the data. Plot your straight-line model and your new curved model. Each plot should include 89% PI intervals.

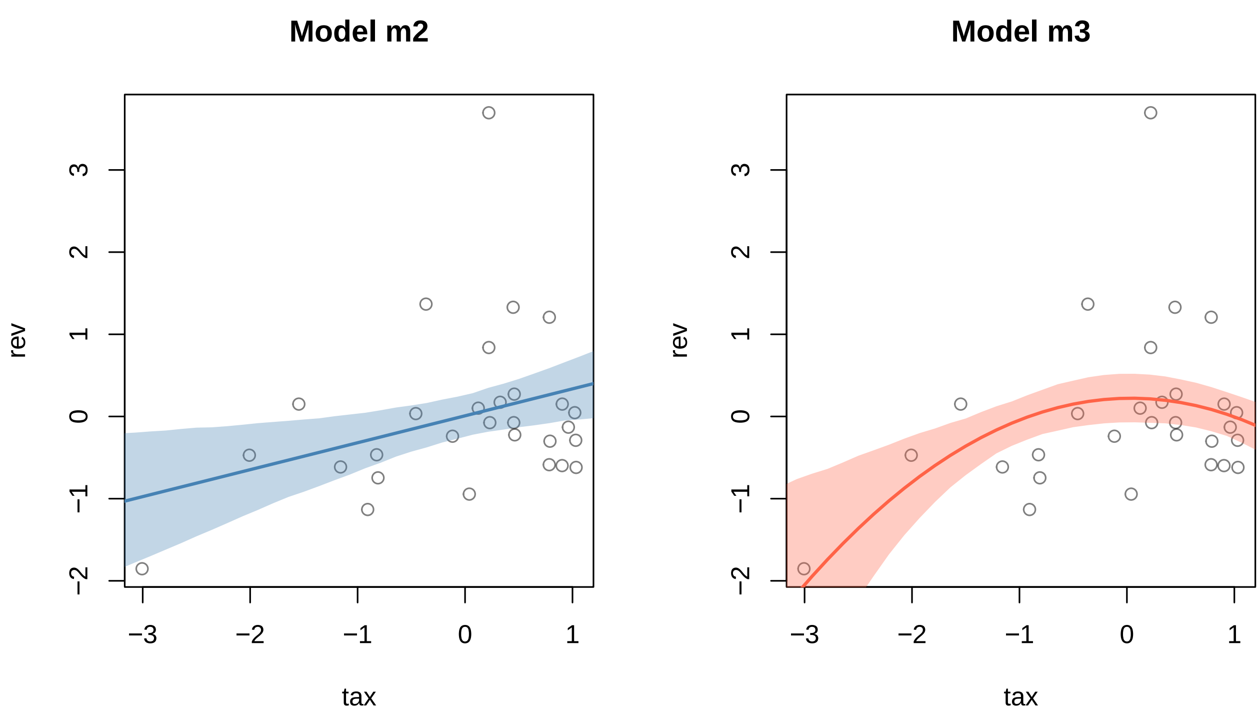
```r
# The Laffer curve in the editorial appears to be a hand-drawn Quadratic Bézier
# curve -- at least to my eye. As a first step towards this curve, we have model
# m3, a degree 2 polynomial, which includes a `squared tax` variable `tax_2`:

# quadtratic `tax squared` variable:
d$tax_2 <- d$tax^2

# Next, we weaken the priors to allow the model to produce a higher-peaked
# posterior distribution.  Note that attempts to induce a more peaked posterior
# distribution in the manner of the editorial's graph generates errors, as the
# initial parameter values are too far away from the MAP estimates.

m3 <- quap(
  alist(
    revenue ~ dnorm( mu , sigma ),
    mu <- a + b1*tax + b2*tax_2,
    a ~ dnorm(0,1),
    b1 ~ dlnorm(0,2),
    b2 ~ dnorm(0, 2),
    sigma ~ dunif( 0 , 8 )
  ), data=list(revenue=d$rev,tax=d$tax,tax_2=d$tax_2) )
precis( m3 )
```

```
##              mean         sd        5.5%       94.5%
## a      0.22188803 0.18550360 -0.07458255  0.51835862
## b1     0.01866559 0.03702467 -0.04050698  0.07783817
## b2    -0.23602774 0.09564724 -0.38889050 -0.08316497
## sigma  0.88542614 0.11626824  0.69960703  1.07124524
```

```r
# use link to compute mu for each sample from the posterior
# and for each weight in `weight.seq`
mu2 <- link( m3 , data=data.frame(tax=rate.seq, tax_2=rate.seq^2) )

# summarize the distribution of mu
mu.mean2 <-apply( mu2 , 2, mean )  # mean of each column (axis '2') of the
                                   # matrix mu2

# compute the 89% PI credibility interval
```

```r
mu.PI2 <- apply( mu2, 2, PI, prob=0.89 )
```

```r
#
# plot setting for side-by-side output: 1 row, 2 columns
par(mfrow=c(1,2))

# --plot m2 ---
plot( rev ~ tax , data=d, col=col.alpha("black",0.5), main="Model m2" )
#plot MAP line, aka the mean mu for each weight
lines( rate.seq, mu.mean, col="steelblue", lwd=2)
#plot a shaded region for 89% PI
shade( mu.PI , rate.seq, col=col.alpha("steelblue",0.33) )

# --plot m3 ---
plot( rev ~ tax , data=d, col=col.alpha("black",0.5), main="Model m3")
#plot MAP line, aka the mean mu for each weight
lines( rate.seq, mu.mean2, col="tomato", lwd=2)
#plot a shaded region for 89% PI
shade( mu.PI2 , rate.seq, col=col.alpha("tomato",0.33) )
```



c) Using WAIC or PSIS, compare a straight-line model to your curved model. What conclusions would you draw from comparing your two models?

```r
# Compare  m2 & m3  by PSIS:
set.seed(2002)
compare( m2, m3 , func=PSIS)
```

```
## Some Pareto k values are very high (>1). Set pointwise=TRUE to inspect individual points.
## Some Pareto k values are very high (>1). Set pointwise=TRUE to inspect individual points.
```

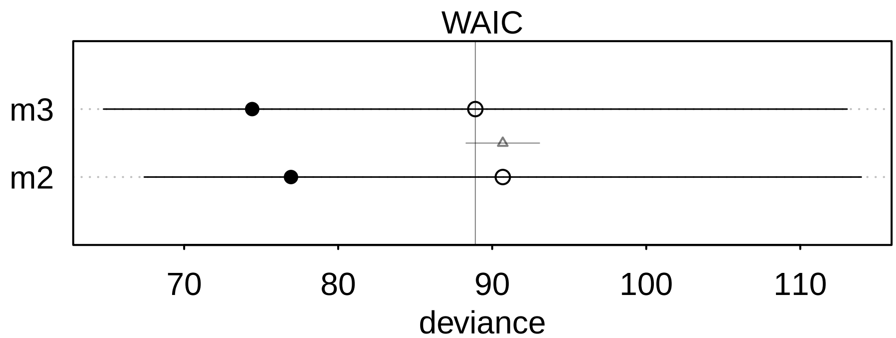
```
##        PSIS       SE    dPSIS      dSE    pPSIS    weight
## m3 101.1592 35.97211 0.00000       NA 13.46013 0.5443531
```

```
## m2 101.5149 34.45184 0.35576 2.751581 12.33191 0.4556469
```

```r
# First, note that PSIS raises an alert that there is at least one outlier where
# the smoothing approximation that PSIS uses is unreliable.  We will return to
# analyze this point for question (d).

# Compare m2 & m3 by WAIC
set.seed(2003)
compare( m2, m3 , func=WAIC)
```

```
##          WAIC       SE    dWAIC      dSE    pWAIC    weight
## m3 87.80086 23.20469 0.000000       NA 6.696022 0.7244322
## m2 89.73396 22.30202 1.933108 2.371508 6.348237 0.2755678
```

```r
# In comparing the two models we may ask, are m2 and m3 easily distinguished by
# their expected out-of-sample accuracy?  To answer, we need to consider the
# error in each estimate.  Let's consider the WAIC score, plotted below
#set.seed(2003)
plot(compare( m2, m3) )
```




```r
# The filled points are in-sample deviance values, the open points are WAIC values,
# and the line segments show standard error of each WAIC (SE). The triangle is
# difference in WAIC (dWAIC) and the corresponding line segment standard error of
# this difference (dSE)

# Although the polynomial model m333 has a lower WAIC score than the straight-line
# model m2, notice that WAIC regards these two models as effectively indistinguishable.
# There is less than 2 units of deviance (dWAIC = 1.9331), and the standard error
# of this difference is even greater (dSE = 2.3715).  Put differently, the standard
# error of the difference between m2 and m3 is even larger than the difference
# itself. Thus, we cannot distinguish between m2 and m3 on the basis of WAIC.
```

d) There is one country with a high tax revenue which is an outlier. Use PSIS and WAIC to measure the importance of this outlier in the two models you fit.

```r
# Recall that PSIS returned an error that the shape parameter k was above 1, which
# corresponds to a warning that the fitted Pareto distribution has infinite variance.

# The warning message appears when values of k exceed 0.5, indicating one or more
# data points that have outsized influence on the out-of-sample accuracy.  In short,
# an outlier simultaneously has outsized influence on the fitting a distribution
# but is, by the nature of being an outlier, unlikely to recur out of sample.

# The editorial identifies the extreme outlier as Norway. Let's see the influence
# of Norway on our model.
```

```r
set.seed(2002)
compare( m2, m3 , func=PSIS)
```

```
## Some Pareto k values are very high (>1). Set pointwise=TRUE to inspect individual points.
## Some Pareto k values are very high (>1). Set pointwise=TRUE to inspect individual points.
```

```
##        PSIS       SE   dPSIS      dSE    pPSIS    weight
## m3 101.1592 35.97211 0.00000       NA 13.46013 0.5443531
## m2 101.5149 34.45184 0.35576 2.751581 12.33191 0.4556469
```

```r
set.seed(2002)
PSIS_m2 <- PSIS(m2, pointwise=TRUE)
```

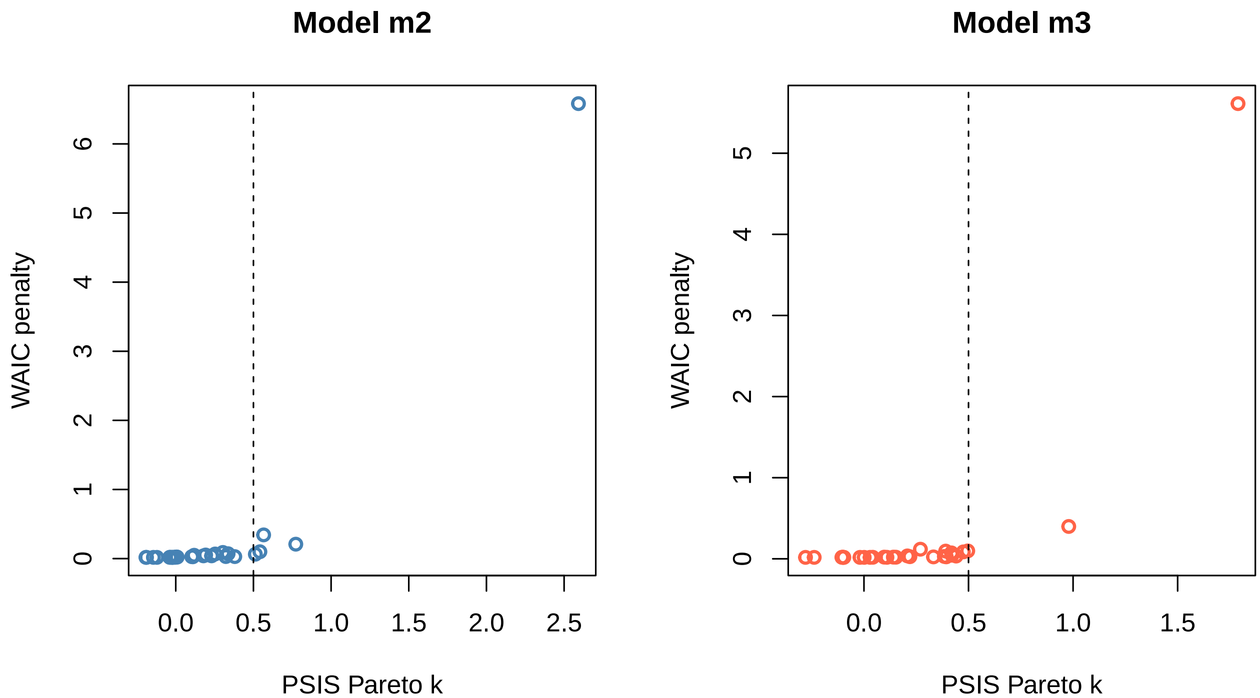
```
## Some Pareto k values are very high (>1). Set pointwise=TRUE to inspect individual points.
```

```r
set.seed(2002)
WAIC_m2 <- WAIC(m2, pointwise=TRUE)


set.seed(2002)
PSIS_m3 <- PSIS(m3, pointwise=TRUE)
```

```
## Some Pareto k values are very high (>1). Set pointwise=TRUE to inspect individual points.
```

```r
set.seed(2002)
WAIC_m3 <- WAIC(m3, pointwise=TRUE)
```

```r
#
# plot setting for side-by-side output: 1 row, 2 columns
par(mfrow=c(1,2))

# -- plot m2
plot( PSIS_m2$k, WAIC_m2$penalty , xlab="PSIS Pareto k",
      ylab="WAIC penalty", col="steelblue", lwd=2,
      main="Model m2")
abline(v = 0.5, lty=2)

# -- plot m3
plot( PSIS_m3$k, WAIC_m3$penalty , xlab="PSIS Pareto k",
      ylab="WAIC penalty", col="tomato", lwd=2,
      main="Model m3")
abline(v = 0.5, lty=2)
```




```
# "Norway" has an extremely high Pareto k (>1) on the x-axis and an extremely
# high WAIC penalty (>2) for both models. What these plots illustrate is that
# Norway has a much higher influence on the posterior distribution than
# any other country, causing a risk to overfitting.

# We could refit our pair of models with a Student-t distribution to reduce the
# leverage of Norway and therefore reduce the risk of overfitting, but this is
# exactly the opposite effect than the original Laffer curve in the editorial
# calls for.  To increase the curature of our model to more closely resemble
# the Laffer curve would require (among other things) that Norway exhibit even
# more influence on the fit of the model -- enough influence, a priori, to
# resist the influence of the other 28 countries.
```

e) Given your analysis, what conclusions do you draw about the relationship between tax rate and tax revenue? Do your conclusions support the original Laffer curve plot used in the editorial?

```
# What we see from our analysis is a curve in search of suitable data rather than
# data in search of a suitable curve.  Our analysis compared a base-line simple
# linear regression to a second-degree polynomial, and we found limitations to
# the curvature of the model that we could fit to the data. The first arose in
# setting priors: if we tried to introduce priors that admitted the mere possibility
# of a more curved posterior, an error is raised that the starting values are too
# far from the maximum a posteriori (MAP) values.  Second, the point -- Norway -- that
# the original Laffer curve fit at its apex is flagged by PSIS as an outlier and
# an overfitting risk.  The standard way for dealing with this dataset would be to
# fit the data with a Student-t distribution which would *decrease* the importance
# of Norway, rather than increase it, leading to an m3 model that was less influenced
# rather than more influenced.

# In short, the Laffer curve in the editorial could only arise from an overwhelmingly
# strong prior, one that essentially outweighed the evidence encoded in the data.
# Thus, rather than fitting a model to data, this editorial overlays a preconceived
```

```
# model of the relationship between corporate tax rates and government tax revenues
# that is inconsistent with the actual relationship in the data.
```