Feb 17, Interaction Effects

Stephen R. Proulx

2/1/2021

Today's objectives:

- Construct models with interaction terms
- Develop an understanding of how interaction models differ from conducting separate analyses

National GDP and ruggedness of terrain

Load the data and do a little processing.

```
data(rugged)
d <- as_tibble(rugged)


# make log version of outcome and  extract countries with GDP data
d <- d %>%
  mutate(log_gdp = log(rgdppc_2000))%>%
  drop_na(rgdppc_2000) %>%
  mutate(log_gdp_std=log_gdp/mean(log_gdp), rugged_std = rugged/max(rugged)) %>%
  mutate(rugged_std=rugged_std - mean(rugged_std))%>%
  select(log_gdp,rugged,log_gdp_std,rugged_std,country, cont_africa)
```

Separate models

Run two separate models, one for non-African countries, the other for African countries.

The model is:

$$\text{log\_gdp} \sim \text{Normal}(\mu, \sigma)\mu = a + b * \text{rugged\_std}a \sim \text{Normal}(?, 1)b \sim \text{Normal}(?, 1)\sigma \sim \text{Exponential}(1)$$

```
m7.1 <-
  quap(alist(
    log_gdp ~ dnorm( mu , sigma ),
    mu <- a+b*rugged_std,
    a  ~ dnorm(1,1),
    b  ~ dnorm(0,1),
    sigma ~ dexp(1)
  ), data=filter(d,cont_africa==1) )
```

```r
m7.2 <-
  quap(alist(
    log_gdp  ~ dnorm( mu , sigma ),
    mu <- a+b*rugged_std,
    a  ~ dnorm(1,1),
    b  ~ dnorm(0,1),
    sigma ~ dexp(1)
  ), data=filter(d,cont_africa==0) )


precis(m7.1)
```

```
##            mean         sd        5.5%    94.5%
## a     7.4100944 0.13025276  7.20192538 7.618263
## b     0.7838193 0.53376418 -0.06923893 1.636878
## sigma 0.8949621 0.09092172  0.74965168 1.040273
```

```r
precis(m7.2)
```

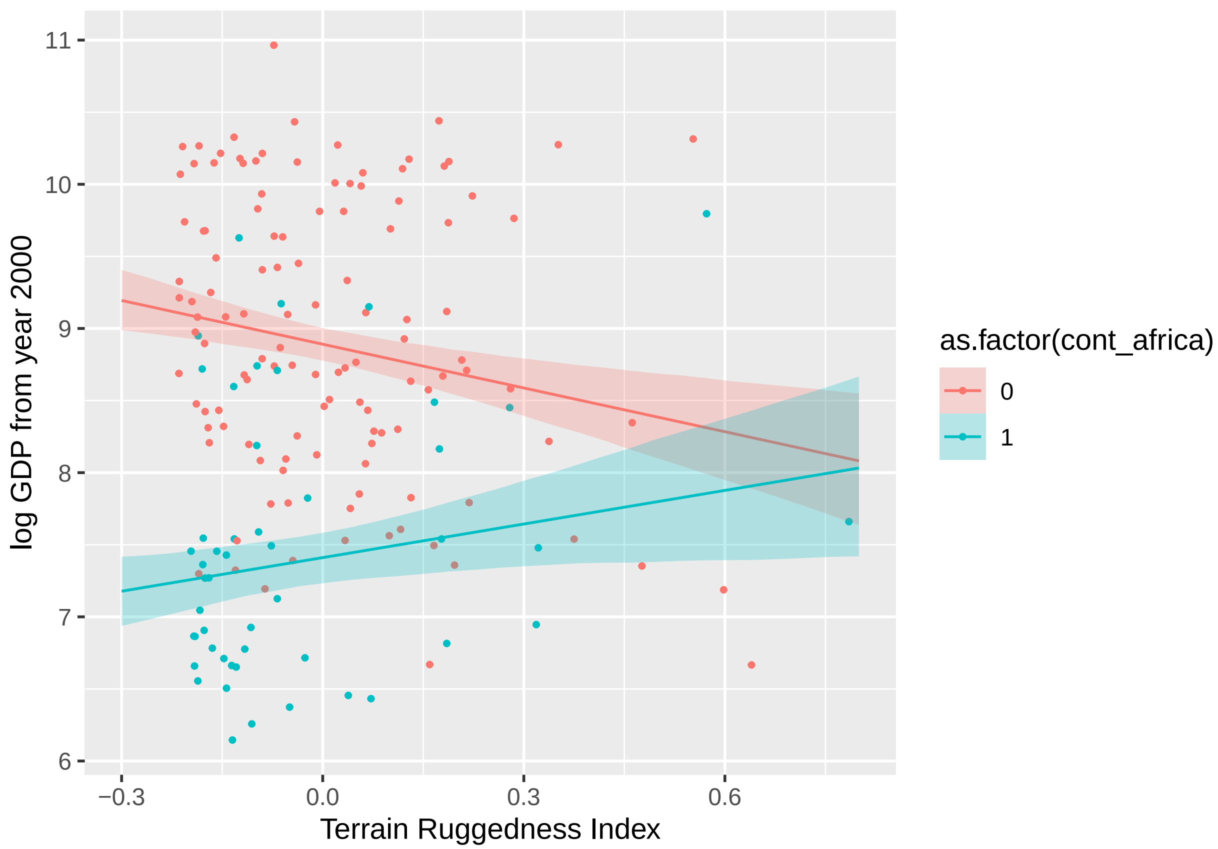
```
##             mean         sd       5.5%      94.5%
## a      8.8904883 0.08639771  8.7524081  9.0285685
## b     -0.9994984 0.43487930 -1.6945195 -0.3044773
## sigma  0.9474013 0.06089293  0.8500827  1.0447200
```

Here's some code to visualize this. It looks long and complicated, but you've used parts of this before. We just need to do some extra work to be able to plot the Africa/non-Africa countries on the same graph.

```r
nd <-
  tibble(rugged_std = seq(from = -0.3, to = 0.8, length.out = 30))

f_m7.1 <-
  link_df(m7.1, data = nd) %>%
  as_tibble() %>%
  group_by(rugged_std) %>%
  summarise(mean_mu=mean(mu),
            lower_mu=quantile(mu,0.1),
            upper_mu=quantile(mu,0.9))%>%
  ungroup()%>%
  mutate(cont_africa = 1)
```

```
## Warning: The `x` argument of `as_tibble.matrix()` must have unique column names if `.name_repair` is
## Using compatibility `.name_repair`.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_warnings()` to see where this warning was generated.

## Note: Using an external vector in selections is ambiguous.
## i Use `all_of(i)` instead of `i` to silence this message.
## i See <https://tidyselect.r-lib.org/reference/faq-external-vector.html>.
## This message is displayed once per session.
```

```r
f_m7.2 <-
  link_df(m7.2, data = nd) %>%
  as_tibble() %>%
  group_by(rugged_std) %>%
  summarise(mean_mu=mean(mu),
            lower_mu=quantile(mu,0.1),
            upper_mu=quantile(mu,0.9))%>%
  ungroup()%>%
  mutate(cont_africa =0)

#put them back into a single data-frame for plotting purposes
f <-
  bind_rows(f_m7.1, f_m7.2)


ggplot(d,aes(x = rugged_std)) +
  geom_point(aes(y = log_gdp, color = as.factor(cont_africa)),
           size = 2/3) +
  geom_smooth(data = f,
            aes(y = mean_mu, ymin = lower_mu, ymax = upper_mu,
                fill = as.factor(cont_africa), color = as.factor(cont_africa)),
            stat = "identity",
            alpha = 1/4, size = 1/2)+
  xlab("Terrain Ruggedness Index" ) +
  ylab("log GDP from year 2000")
```



Models with interaction terms

In this section we will use all the data at once.

We will build up to an interaction model, starting with less complex models and adding terms. For model 7.3 we want:

$$\text{log\_gdp} \sim \text{Normal}(\mu, \sigma)\mu = a + b * \text{rugged\_std} a \sim \text{Normal}(?, 1) b \sim \text{Normal}(?, 1) \sigma \sim \text{Exponential}(1)$$

For model 7.4 we want:

$$\text{log\_gdp} \sim \text{Normal}(\mu, \sigma)\mu = a + br * \text{rugged\_std} + bc * \text{cont\_Africa} a \sim \text{Normal}(?, 1) br \sim \text{Normal}(?, 1) bc \sim \text{Normal}(?, 1) \sigma \sim \text{Exp}$$

I'm going to specify m7.5 slightly differently than in the book, and explain later:

$$\text{log\_gdp} \sim \text{Normal}(\mu, \sigma)\mu = anA * (1 - \text{cont\_Africa}) + aA * (\text{cont\_Africa}) + brnA * \text{rugged\_std} * (1 - \text{cont\_Africa}) + brA * \text{rugged\_s}$$

```
m7.3 <-
  quap(alist(
    log_gdp ~ dnorm( mu , sigma ),
    mu <- a+br*rugged_std,
    a  ~ dnorm(1,1),
    br ~ dnorm(0,1),
    sigma ~ dexp(1)
  ), data=d)



m7.4 <-
  quap(alist(
    log_gdp ~ dnorm( mu , sigma ),
    mu <- a+br*rugged_std + bc * cont_africa,
    a  ~ dnorm(1,1),
    br  ~ dnorm(0,1),
    bc  ~ dnorm(0,1),
    sigma ~ dexp(1)
  ), data=d)

m7.5 <-
  quap(alist(
    log_gdp  ~ dnorm( mu , sigma ),
    mu <- anA*(1-cont_africa) + aA* cont_africa  + brnA*rugged_std*(1-cont_africa) + brA *rugged_std* co
   anA  ~ dnorm(1,1),
    aA  ~ dnorm(1,1),
    brnA  ~ dnorm(0,1),
    brA  ~ dnorm(0,1),
    sigma ~ dexp(1)
  ), data=d )
```

Now use compare to see how they stack up in terms of WAIC.

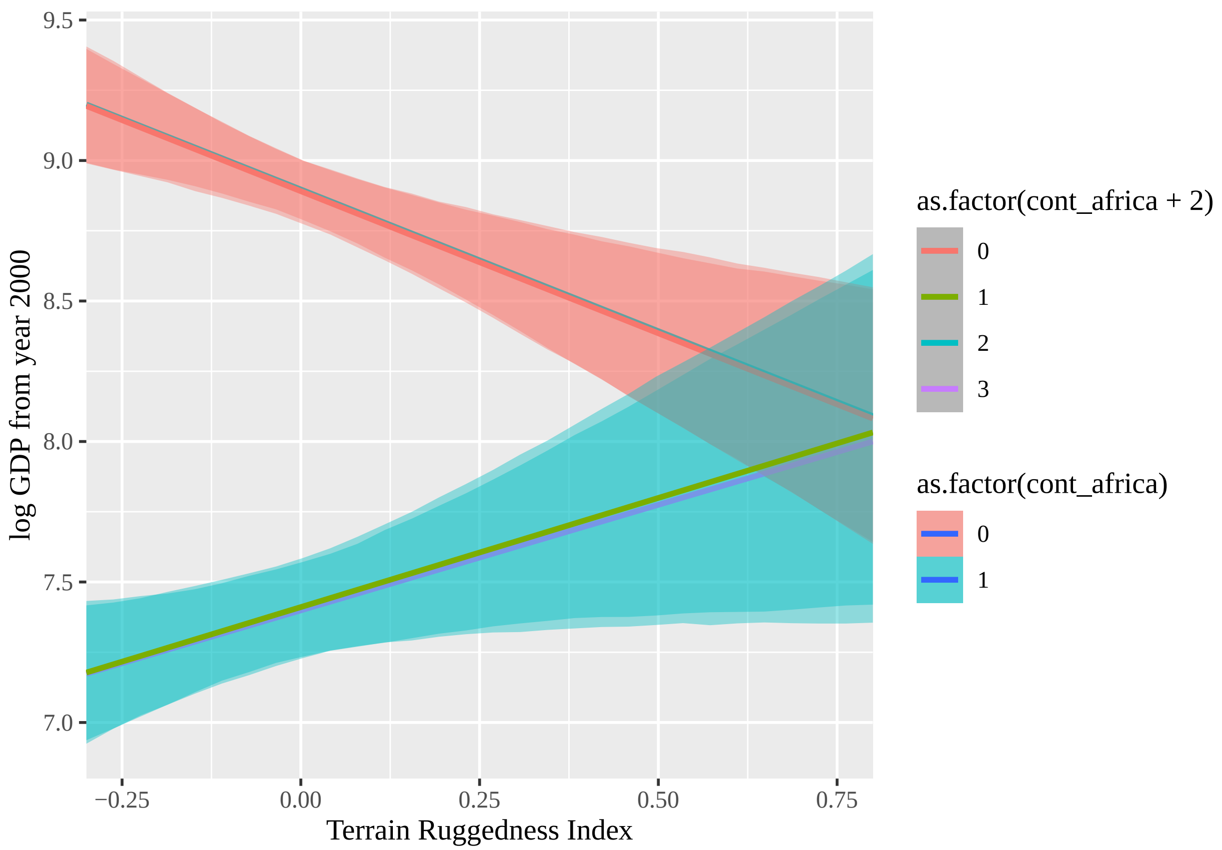
```
compare(m7.3,m7.4,m7.5)
```

```
##          WAIC       SE     dWAIC       dSE    pWAIC       weight
## m7.5 470.6011 15.98758  0.000000        NA 4.963194 9.583407e-01
## m7.4 476.8725 15.30645  6.271359  5.424311 4.334327 4.165927e-02
## m7.3 539.8909 13.26018 69.289805 16.113827 2.585007 8.618465e-16
```

And now let's see how the predicitons differ between the interaction model with the "run them separately model." What are these models doing the same or differently? The only difference is that sigma is estimated separately for the groups of countries.

```
nd1 <-
  tibble(rugged_std = seq(from = -0.3, to = 0.8, length.out = 30),cont_africa=0)
nd2 <-
  tibble(rugged_std = seq(from = -0.3, to = 0.8, length.out = 30),cont_africa=1)


f_m7.5_A <-
  link_df(m7.5,  data = nd1) %>%
  as_tibble() %>%
  group_by(rugged_std) %>%
  summarise(mean_mu=mean(mu),
            lower_mu=quantile(mu,0.1),
            upper_mu=quantile(mu,0.9))%>%
  ungroup()%>%
  mutate(cont_africa =0,
         model="regular")


f_m7.5_N <-
  link_df(m7.5,  data = nd2) %>%
  as_tibble() %>%
  group_by(rugged_std) %>%
  summarise(mean_mu=mean(mu),
            lower_mu=quantile(mu,0.1),
            upper_mu=quantile(mu,0.9))%>%
  ungroup()%>%
  mutate(cont_africa =1,
         model="regular")

#put them back into a single data-frame for plotting purposes
f_7.5 <-
  bind_rows(f_m7.5_A, f_m7.5_N)





ggplot(d,aes(x = rugged_std)) +
  geom_smooth(data = f_7.5,
              aes(y = mean_mu, ymin = lower_mu, ymax = upper_mu,
                 fill = as.factor(cont_africa), color = as.factor(cont_africa+2)),
              stat = "identity") +
    geom_smooth(data = f,
              aes(y = mean_mu, ymin = lower_mu, ymax = upper_mu,
                  fill = as.factor(cont_africa), color = as.factor(cont_africa)),
```

```r
              stat = "identity")+ scale_x_continuous("Terrain Ruggedness Index", expand = c(0, 0)) +
  ylab("log GDP from year 2000") +
  theme(text = element_text(family = "Times"))
```



Can we compare the "run them separately" model with the "run them together" model?

We can do a bit better to see how these models compare with each other by writing a model that includes all the data, but allows for sigma to vary by continent.

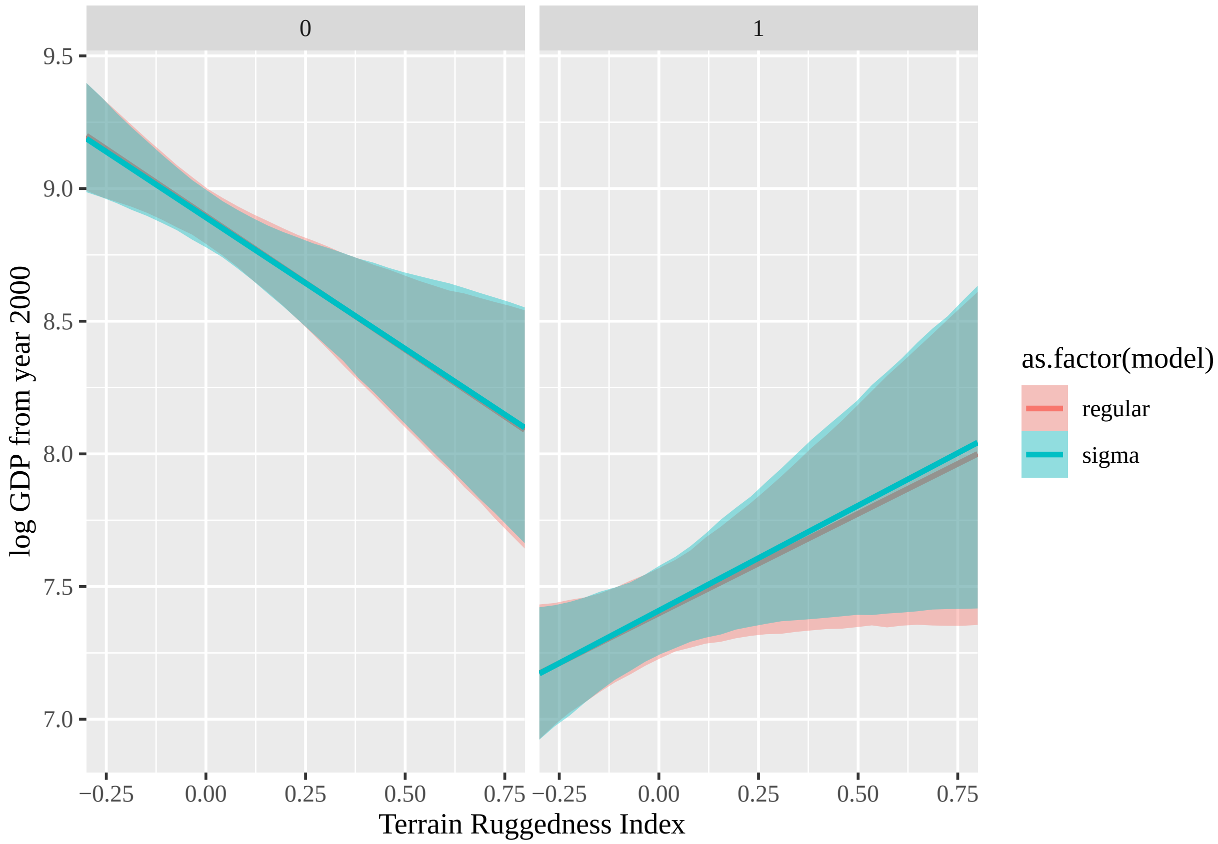
```r
m7.5.sigma <-
  quap(alist(
    log_gdp  ~ dnorm( mu , sigma ),
    mu <- anA*(1-cont_africa) + aA* cont_africa + brnA*rugged_std*(1-cont_africa) + brA *rugged_std* co
    sigma <- sigmanA*(1-cont_africa) + sigmaA* cont_africa,
    anA  ~ dnorm(1,1),
    aA  ~ dnorm(1,1),
    brnA  ~ dnorm(0,1),
    brA  ~ dnorm(0,1),
    sigmanA ~ dexp(1),
    sigmaA ~ dexp(1)
  ), data=d)
```

```r
compare(m7.5, m7.5.sigma)
```

```
##                 WAIC       SE    dWAIC     dSE    pWAIC    weight
## m7.5       470.4614 15.83204 0.000000      NA 4.933004 0.7544126
## m7.5.sigma 472.7060 16.52879 2.244573 1.46604 6.066878 0.2455874
```

How do their predictions compare?

```
nd1 <-
  tibble(rugged_std = seq(from = -0.3, to = 0.8, length.out = 30),cont_africa=0)
nd2 <-
  tibble(rugged_std = seq(from = -0.3, to = 0.8, length.out = 30),cont_africa=1)


f_m7.5_sigma_A <-
  link_df(m7.5.sigma,  data = nd1) %>%
  as_tibble() %>%
  group_by(rugged_std) %>%
  summarise(mean_mu=mean(mu),
            lower_mu=quantile(mu,0.1),
            upper_mu=quantile(mu,0.9))%>%
  ungroup()%>%
  mutate(cont_africa =0,
         model="sigma")


f_m7.5_sigma_N <-
  link_df(m7.5.sigma,  data = nd2) %>%
  as_tibble() %>%
  group_by(rugged_std) %>%
  summarise(mean_mu=mean(mu),
            lower_mu=quantile(mu,0.1),
            upper_mu=quantile(mu,0.9))%>%
  ungroup()%>%
  mutate(cont_africa =1,
         model="sigma")

#put them back into a single data-frame for plotting purposes
f_7.5_sigma <-
  bind_rows(f_m7.5_sigma_A, f_m7.5_sigma_N)

#join the two model predictions

f_7.5_joined <- bind_rows(f_7.5,f_7.5_sigma)




ggplot(filter(f_7.5_joined),aes(x = rugged_std)) +
  geom_smooth(aes(y = mean_mu, ymin = lower_mu, ymax = upper_mu,
                  fill = as.factor(model), color = as.factor(model)),
              stat = "identity") +
  scale_x_continuous("Terrain Ruggedness Index", expand = c(0, 0)) +
  ylab("log GDP from year 2000") +
  theme(text = element_text(family = "Times"))  +
  facet_wrap(~cont_africa)
```



###Compare with the alternative formulation In the video lecture, McElreath talks about the standard way to formulate interaction models, and I recreate this here. Later we will see in more detail how to formulate models that consider differences rather as a way to have more meaningful priors.

```
m7.5 <-
  quap(alist(
    log_gdp  ~ dnorm( mu , sigma ),
    mu <- anA*(1-cont_africa) + aA* cont_africa  + brnA*rugged_std*(1-cont_africa) + brA *rugged_std* c
   anA  ~ dnorm(1,1),
    aA  ~ dnorm(1,1),
    brnA  ~ dnorm(0,1),
    brA  ~ dnorm(0,1),
    sigma ~ dexp(1)
  ), data=d )

m7.5.alt <-
  quap(alist(
    log_gdp  ~ dnorm( mu , sigma ),
    mu <- aglobal + bA* cont_africa  +  br*rugged_std  + brXA *rugged_std* cont_africa,
    aglobal  ~ dnorm(1,1),
    bA  ~ dnorm(0,1),
    br  ~ dnorm(0,1),
    brXA  ~ dnorm(0,1),
    sigma ~ dexp(1)
  ), data=d )
```

They do come out a little bit different from each other, but this is all the priors as we'll see.

```
precis(m7.5.alt)
```

```
##                mean         sd       5.5%       94.5%
## aglobal  8.8783943 0.08512947  8.7423410  9.01444767
## bA      -1.3454595 0.15831074 -1.5984707 -1.09244836
## br      -0.7478931 0.40612953 -1.3969666 -0.09881971
## brXA     1.3640350 0.62059502  0.3722043  2.35586575
## sigma    0.9371093 0.05112846  0.8553961  1.01882244
```

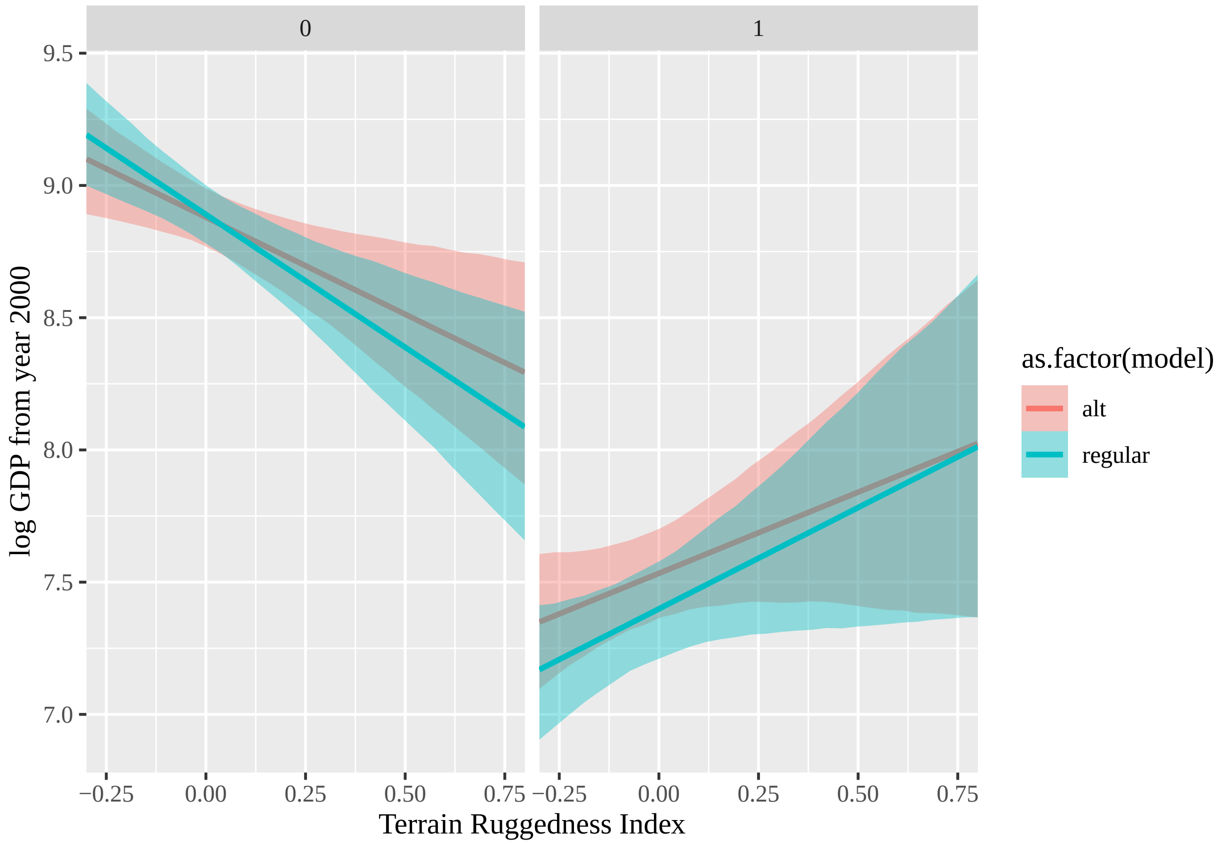
```
precis(m7.5)
```

```
##              mean         sd       5.5%      94.5%
## anA    8.8919764 0.08519572  8.7558172  9.0281357
## aA     7.3997206 0.13450998  7.1847477  7.6146936
## brnA  -1.0048677 0.43010340 -1.6922560 -0.3174794
## brA    0.7587887 0.54839884 -0.1176585  1.6352360
## sigma  0.9352643 0.05101595  0.8537309  1.0167976
```

And now plot the estimates of mu.

```
nd1 <-
  tibble(rugged_std = seq(from = -0.3, to = 0.8, length.out = 30),cont_africa=0)
nd2 <-
  tibble(rugged_std = seq(from = -0.3, to = 0.8, length.out = 30),cont_africa=1)


f_m7.5_A <-
  link_df(m7.5,  data = nd1) %>%
  as_tibble() %>%
  group_by(rugged_std) %>%
  summarise(mean_mu=mean(mu),
            lower_mu=quantile(mu,0.1),
            upper_mu=quantile(mu,0.9))%>%
  ungroup()%>%
  mutate(cont_africa =0,
         model="regular")


f_m7.5_N <-
  link_df(m7.5,  data = nd2) %>%
  as_tibble() %>%
  group_by(rugged_std) %>%
  summarise(mean_mu=mean(mu),
            lower_mu=quantile(mu,0.1),
            upper_mu=quantile(mu,0.9))%>%
  ungroup()%>%
  mutate(cont_africa =1,
         model="regular")

#put them back into a single data-frame for plotting purposes
```

```r
f_7.5  <-
  bind_rows(f_m7.5_A, f_m7.5_N)

f_m7.5_alt_A <-
  link_df(m7.5.alt,  data = nd1) %>%
  as_tibble() %>%
  group_by(rugged_std) %>%
  summarise(mean_mu=mean(mu),
            lower_mu=quantile(mu,0.1),
            upper_mu=quantile(mu,0.9))%>%
  ungroup()%>%
  mutate(cont_africa =0,
         model="alt")


f_m7.5_alt_N <-
  link_df(m7.5.alt,  data = nd2) %>%
  as_tibble() %>%
  group_by(rugged_std) %>%
  summarise(mean_mu=mean(mu),
            lower_mu=quantile(mu,0.1),
            upper_mu=quantile(mu,0.9))%>%
  ungroup()%>%
  mutate(cont_africa =1,
         model="alt")

#put them back into a single data-frame for plotting purposes
f_7.5_alt <-
  bind_rows(f_m7.5_alt_A, f_m7.5_alt_N)

#join the two model predictions

f_7.5_joined <- bind_rows(f_7.5,f_7.5_alt)




ggplot(filter(f_7.5_joined),aes(x = rugged_std)) +
  geom_smooth(aes(y = mean_mu, ymin = lower_mu, ymax = upper_mu,
               fill = as.factor(model), color = as.factor(model)),
            stat = "identity") +
  scale_x_continuous("Terrain Ruggedness Index", expand = c(0, 0)) +
  ylab("log GDP from year 2000") +
  theme(text = element_text(family = "Times"))  +
  facet_wrap(~cont_africa)
```



So the alternative model gets more of a slope out of Africa and more of an intercept in Africa. That said, overall they really do perform quite similarly, especially over the range where there is lots of data.

We'll loosen up the priors and see how they change

```
m7.5 <-
  quap(alist(
    log_gdp  ~ dnorm( mu , sigma ),
    mu <- anA*(1-cont_africa) + aA* cont_africa  + brnA*rugged_std*(1-cont_africa) + brA *rugged_std* co
   anA  ~ dnorm(1,10),
    aA  ~ dnorm(1,10),
    brnA  ~ dnorm(0,10),
    brA  ~ dnorm(0,10),
    sigma ~ dexp(10)
  ), data=d )

m7.5.alt <-
  quap(alist(
    log_gdp  ~ dnorm( mu , sigma ),
    mu <- aglobal + bA* cont_africa  +  br*rugged_std  + brXA *rugged_std* cont_africa,
    aglobal  ~ dnorm(1,10),
    bA  ~ dnorm(0,10),
    br  ~ dnorm(0,10),
    brXA  ~ dnorm(0,10),
    sigma ~ dexp(10)
  ), data=d )
```

Now they are more similar (you need to do some subtraction to completely see this)

```r
precis(m7.5.alt)
```

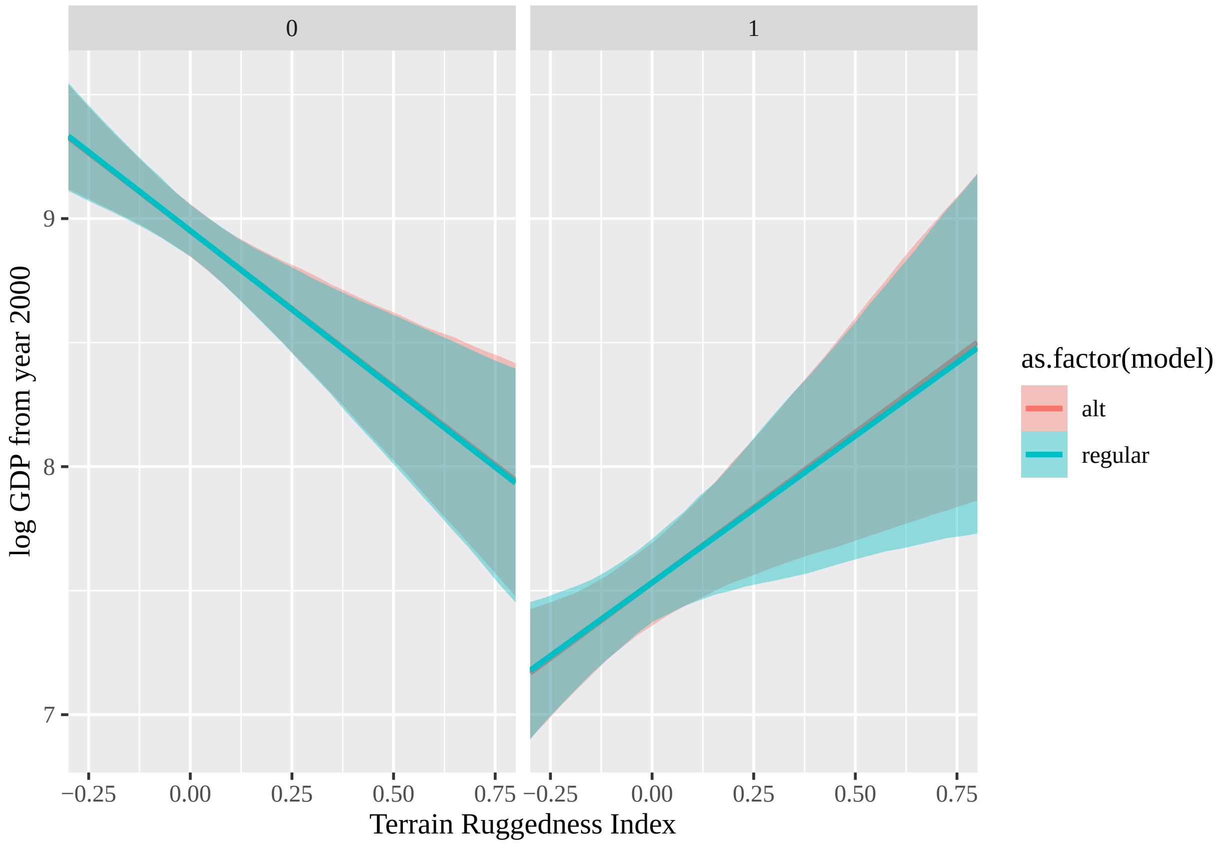
```
##               mean         sd       5.5%      94.5%
## aglobal  8.9520229 0.08287799  8.8195678  9.0844779
## bA      -1.4229232 0.15568890 -1.6717441 -1.1741023
## br      -1.2499856 0.46114677 -1.9869872 -0.5129840
## brXA     2.4221116 0.78330047  1.1702461  3.6739770
## sigma    0.9087113 0.04741743  0.8329291  0.9844935
```

```r
precis(m7.5)
```

```
##             mean         sd       5.5%      94.5%
## anA    8.9523023 0.08287853  8.8198464  9.0847582
## aA     7.5276776 0.13179749  7.3170398  7.7383155
## brnA  -1.2560652 0.46161935 -1.9938221 -0.5183083
## brA    1.1786028 0.63449269  0.1645609  2.1926446
## sigma  0.9086799 0.04741337  0.8329042  0.9844556
```

```r
nd1 <-
  tibble(rugged_std = seq(from = -0.3, to = 0.8, length.out = 30),cont_africa=0)
nd2 <-
  tibble(rugged_std = seq(from = -0.3, to = 0.8, length.out = 30),cont_africa=1)


f_m7.5_A <-
  link_df(m7.5,  data = nd1) %>%
  as_tibble() %>%
  group_by(rugged_std) %>%
  summarise(mean_mu=mean(mu),
            lower_mu=quantile(mu,0.1),
            upper_mu=quantile(mu,0.9))%>%
  ungroup()%>%
  mutate(cont_africa =0,
         model="regular")


f_m7.5_N <-
  link_df(m7.5,  data = nd2) %>%
  as_tibble() %>%
  group_by(rugged_std) %>%
  summarise(mean_mu=mean(mu),
            lower_mu=quantile(mu,0.1),
            upper_mu=quantile(mu,0.9))%>%
  ungroup()%>%
  mutate(cont_africa =1,
         model="regular")

#put them back into a single data-frame for plotting purposes
f_7.5  <-
  bind_rows(f_m7.5_A, f_m7.5_N)
```

```r
f_m7.5_alt_A <-
  link_df(m7.5.alt,  data = nd1) %>%
  as_tibble() %>%
  group_by(rugged_std) %>%
  summarise(mean_mu=mean(mu),
            lower_mu=quantile(mu,0.1),
            upper_mu=quantile(mu,0.9))%>%
  ungroup()%>%
  mutate(cont_africa =0,
         model="alt")


f_m7.5_alt_N <-
  link_df(m7.5.alt,  data = nd2) %>%
  as_tibble() %>%
  group_by(rugged_std) %>%
  summarise(mean_mu=mean(mu),
            lower_mu=quantile(mu,0.1),
            upper_mu=quantile(mu,0.9))%>%
  ungroup()%>%
  mutate(cont_africa =1,
         model="alt")

#put them back into a single data-frame for plotting purposes
f_7.5_alt <-
  bind_rows(f_m7.5_alt_A, f_m7.5_alt_N)

#join the two model predictions

f_7.5_joined <- bind_rows(f_7.5,f_7.5_alt)




ggplot(filter(f_7.5_joined),aes(x = rugged_std)) +
  geom_smooth(aes(y = mean_mu, ymin = lower_mu, ymax = upper_mu,
               fill = as.factor(model), color = as.factor(model)),
            stat = "identity") +
  scale_x_continuous("Terrain Ruggedness Index", expand = c(0, 0)) +
  ylab("log GDP from year 2000") +
  theme(text = element_text(family = "Times"))  +
  facet_wrap(~cont_africa)
```


0
1
log GDP from year 2000
9
8
7
−0.25
0.00
0.25
0.50
0.75
Terrain Ruggedness Index
as.factor(model)
alt
regular