Assignment 3

Exercise 1

Load the provided data on terrain *ruggedness* and *GDP* per capita.

1. Reproduce the main result, given in Table 1, Column 5 (they use HC1 standard errors), of Nunn and Puga (2012) by fitting a model with the following (interacted) variables:

$$\log \text{rgdppc}_{2000} \approx (\text{rugged} + \text{dist-coast}) \times \text{africa}.$$

2. Plot posterior samples of the ruggedness effect in Africa against the effect in the rest of the world in a scatterplot. What can you say about the effect?
3. Estimate three additional models — one without the distance to coast, one that uses population in 1400 (use $\log 1 + pop$) instead, and one with both controls.
4. Discuss (conceptually different) approaches to selecting one of these models for inference. *Hint: Consider the difference between causal inference and other inference tasks.*

The marginal likelihood of a model $\mathcal{M}_j$ is a natural measure of *model fit*, and is given by

$$p(\mathcal{D} \mid \mathcal{M}_j) = \iint p(\mathcal{D} \mid \beta_j, \sigma_j) p(\beta_j, \sigma_j) d\beta_j d\sigma_j.$$

When using the *conjugate Normal Inverse-Gamma prior* for linear regression, it is available in closed form.

1. Investigate the sensitivity of the estimates of your model of choice to different prior parameters.
2. Compare the ML using different prior parameters, including prior variances: $\mathbf{\Sigma}_0 = v_p \mathbf{I}$ with $v_p \in \{0.0001, 0.01, 1, 100\}$.
3. Compare the three models using Bayes factors, and explain how they depend on the *model prior* $p(\mathcal{M}_j)$.
4. The *posterior predictive density*, $p(\mathbf{y}^* \mid \mathbf{y}) = \iint p(\mathbf{y}^* \mid \mathbf{y}, \beta_j, \sigma_j) p(\beta_j, \sigma_j) d\beta_j d\sigma_j$, allows us to quantify uncertainty around predictions. To implement this, obtain posterior draws from the model of your choice and use them to *simulate predictions*. Visualise the predictive uncertainty around a subset of the model.

Hint: Consider using the `brms` *package for fitting the models. For model selection, you could use* `bridge_sampler()` *to compute marginal likelihoods,* `waic()` *for an information criterion, or* `loo()` *to compute a leave-one-out cross-validation information criterion.*

Exercise 2

Download the time series on US unemployment rates (mnemonic: UNRATE) from the FRED data base. This is a monthly (seasonally adjusted) time series of unemployment rates running from 1948M1 until today.

- Demean the time series prior to analysis.
- Fit an AR(1) process to this time series using the following priors on the autoregressive coefficient ϕ:
  1. Normal with mean zero and one, and unit variance — $\mathrm{N}(0, 1), \mathrm{N}(1, 1)$.
  2. Beta with shape and rate equal to 0.1, 1, and 10 — $\mathrm{B}(0.1, 0.1), \mathrm{B}(1, 1), \mathrm{B}(10, 10)$,
- Plot the prior distributions and give an intuition for the implied prior assumption for ϕ.
- Plot and interpret the posterior distributions of ϕ -— are they sensitive to the prior?