GR5065 HW2

1 Economic Growth

> **It is important that you complete 1.1 through 1.3 before Thursday, February 23rd and then you can quickly do 1.4 sometime during the day on February 23rd.**

The concept of the total output of an economy is used in almost every social science model where the units of observation are countries (or country-periods). However, no country "counts" up all the transactions that occur within its jurisdiction, as if everything were recorded in a blockchain. Rather, the "data" on Gross Domestic Product (GDP) or GDP per capita is the output of some model, and different models produce different predictions of the same theoretical concept.

In addition, there are (at least) two different approaches to modeling it, either as total expenditures or as total income received. To better understand the two methods that are used by the U.S. government to estimate its *annualized growth rate*, read this paper.

To try reduce confusion, let's use μ_t to refer to the *concept* of real economic growth at time t, and let GDP growth refer to an *estimate* of μ_t via the expenditure approach. Gross Domestic Income (GDI) growth is also an *estimate* of μ_t via the income approach. You can download illustrative data on GDI and GDP growth from the start of 1970 through the third quarter of 2022, along with the level of the unemployment rate (UR), via

```r
library(dplyr)
FRED <- "https://fred.stlouisfed.org/graph/fredgraph.csv?id="
SERIES <- c(GDI = "A261RL1Q225SBEA",
            GDP = "A191RL1Q225SBEA",
            UR  = "LRUN64TTUSQ156S")
dataset <- readr::read_csv(paste0(FRED, paste(SERIES, collapse = ",")),
                           progress = FALSE, show_col_types = FALSE,
                           na = ".") %>%
  rename(quarter_startdate = DATE,
```

```
        GDI = A261RL1Q225SBEA,
        GDP = A191RL1Q225SBEA,
        UR  = LRUN64TTUSQ156S) %>%
  na.omit %>%
  arrange(desc(quarter_startdate))
```

If you look at `dataset`, you will see that in the last few quarters estimated GDI and GDP have been quite different from each other, although you do not need to actually use `dataset` to complete this question.

Suppose we have a measurement model for the United States where both GDI and GDP at time t are independent of each and normally distributed with expectation μ_t and time-invariant precision τ (recall that the precision is the reciprocal of the variance).

1.1 Prior for τ

Suppose that your prior distribution for τ is a Gamma distribution with shape $a > 0$ and rate $b > 0$ where $\mathbb{E}\tau = \frac{a}{b} = \frac{4}{9}$. Choose values of a and b that are consistent with this information and define `a` and `b` in a R chunk.

1.2 Prior for $\mu_t \mid \tau$

The article above mentions Okun's Law, which there are several versions of. For our purposes, suppose the picture at the top right of that Wikipedia link represents Okun's Law as

$$\mu_t\left(x_t\right) \equiv 3.2 + (-1.8)\, x_t$$

where x_t is the *change* in the unemployment rate between quarter t and quarter $t-1$. We are only concerned with the situation where t is the fourth quarter of 2022. The unemployment rate in October of 2022 was 3.62 percent and the unemployment rate in December of 2022 was 3.43 percent for a change of $x_t = -0.19$.

Suppose your prior for $\mu_t \mid \tau$ is normal with expectation, m, provided by Okun's law and precision equal to $v\tau$, where $v > 0$ reflects how well you think Okun's law predicts quarterly economic growth. You might note that the R^2 in the Wikipedia link above is 0.463 . Define `m` and `v` in a R chunk.

1.3 Prior Predictive Distribution

Create a tibble called `draws` with $R = 10^7$ rows where you

1. Create a column called `tau` where τ is drawn from a Gamma distribution with `shape` a and `rate` b
2. Create a column called `mu` where μ is drawn from a normal distribution with `mean` m and `sd` $\frac{1}{\sqrt{v\tau}}$.
3. Create a column called `GDP` that is drawn from a normal distribution with `mean` `mu` and `sd` $\frac{1}{\sqrt{\tau}}$
4. Create a column called `GDI` that is also drawn from a normal distribution `mean` `mu` and `sd`$\frac{1}{\sqrt{\tau}}$

Note that τ influences both the prior distribution for μ_t and the distribution of the future data on GDP and GDI. Use the `geom_density` function twice in the ggplot2 package to plot the density of `GDI` and `GDP` (in different colors). Does this plot represent a reasonable set of beliefs about what GDI and GDP will be *reported* for the fourth quarter of 2022 by the U.S. government? If not, go back and revise your choice of `a`, `b`, and / or `v`.

1.4 Posterior Distribution

> **You will have to complete this subproblem after the U.S. government releases its report for the fourth quarter of 2022 on Thursday February 23, 2023 at 8:30AM from the top of** https://www.bea.gov/ .

Once you have two data points (GDP and GDI) for the fourth quarter of 2022, define them as `GDP_t` and `GDI_t` in a R chunk. These will both be reported to one decimal place. Subset `draws` to keep only those rows that round to the observed values with something like

```
posterior_draws <- filter(draws,
                          round(GDP, digits = 1) == GDP_t,
                          round(GDI, digits = 1) == GDI_t)
```

Then, use the `geom_hex` function in the ggplot2 package to visualize the posterior distribution of μ_t and τ from `posterior_draws`. How would you describe your updated beliefs about μ_t and τ?

2 Climate Change

Read this paper by Stevens et al. (2016) including its online appendices, especially the fifth appendix. This paper is about the Equilibrium Climate Sensitivity (ECS), which is an unknown value that indicates how much the average temperature of the Earth will change in the distant future as a result of the "forcing" event of doubling the amount of carbon dioxide in the atmosphere since the start of the industrial revolution. As with the poker problem on HW1, you do not need to know much about climate modeling to answer the questions below because you can ask on Ed Discussion about anything you do not understand. You do, however, need to apply what we have learned about probability with continuous random variables.

2.1 Prior

In equation S3 on page 13 of the appendix, Stevens et al. (2016) states

> The uncertainty in the forcing, F, and feedback, λ, used to generate the prior [for the Equilibrium Climate Sensitivity] are large, resulting in a prior having the form
>
> $$\text{ECS} = -\frac{F + \sigma_F}{\lambda + \sigma_\lambda}$$
>
> where σ_F is a gaussian distributed random variable with [expectation zero and] standard deviation $0.2F$. Likewise, σ_λ is a gaussian distributed random variable with [expectation zero and] standard deviation 0.5λ. For F and λ we adopt 3.7 and -1.6 ... respectively.

This implies that the numerator is a Gaussian random variable with expectation F and standard deviation $0.2F$, while the denominator is a Gaussian random variable with expectation λ and standard deviation $|0.5\lambda|$.

Create a tibble called `draws` with $R = 10^7$ rows with a column for the numerator, a column for the denominator, and a column called `ECS` for the ratio (note the negative sign in front).

2.2 Truncation

In the next sentence, Stevens et al. (2016) states — apparently after the first version of the article was originally published — that their actual prior on the ECS was only over $\Theta = [0, 10]$ since negative values for ECS are basically impossible and very large values are implausible.

- Redefine `draws` as the realizations in the previous subproblem that satisfy this condition.
- What proportion of the original R realizations satisfy this condition?
- Use the `geom_density` function in the ggplot2 package to visualize the prior on the ECS of Stevens et al. (2016) over the $[0, 10]$ interval

2.3 Ratio of Normal Random Variates

Let the continuous random variable, X, be normally distributed with expectation μ_X and standard deviation σ_X. Similarly, let the continuous random variable, Y, be normally distributed with expectation μ_Y and standard deviation σ_Y . Then, if X and Y are independent, the Probability Density Function (PDF) of the random variable $Z = \frac{X}{Y}$ over $\Theta = \mathbb{R}$ is helpfully given on Wikipedia, where $\Phi(t)$ is the standard normal Cumulative Distribution Function (CDF) and is implemented in R by the `pnorm` function, whose default arguments are `mean` of zero and `sd` of one.

This PDF for Z is applicable to the untruncated distribution of

$$\text{ECS} = \frac{F + \sigma_F}{-(\lambda + \sigma_\lambda)} = \frac{F + \sigma_F}{-\lambda - \sigma_\lambda} = Z$$

so the denominator is a normal random variable with expectation $-\lambda$ and standard deviation $-0.5\lambda > 0$ when $\lambda = -1.6$.

Write a R function called `dratio` that evaluates the PDF of Z. It should start like

```
dratio <- function(z, mu_X = 3.7, sigma_X = 0.74, mu_Y = 1.6, sigma_Y = 0.8) {
  # fill in the rest by defining a, b, c, d, etc. as in the Wikipedia article
}
```

2.4 Posterior

The function that Stevens et al. (2016) uses for the likelihood is given in equation S4 on page 13 of the appendix as

$$P(e_j \mid \chi) = \frac{(1 - 2e_j)\,\text{erf}(2\chi - 2\chi_j) + 1}{2}$$

with $e_j \in [0, \frac{1}{2}]$ but this is very confusing. Pretend that they equivalently wrote that as

$$L(\chi; e_j, c_j) = \frac{(1 - 2e_j)\, 2\Phi\left(\sqrt{2}\,(2\chi - 2c_j) - 1\right) + 1}{2}$$

which in a R function would be

```
L <- function(chi, e, c) {
  ( (1 - 2 * e) * (2 * pnorm(sqrt(2) * (2 * chi - 2 * c)) - 1) + 1 ) / 2
}
```

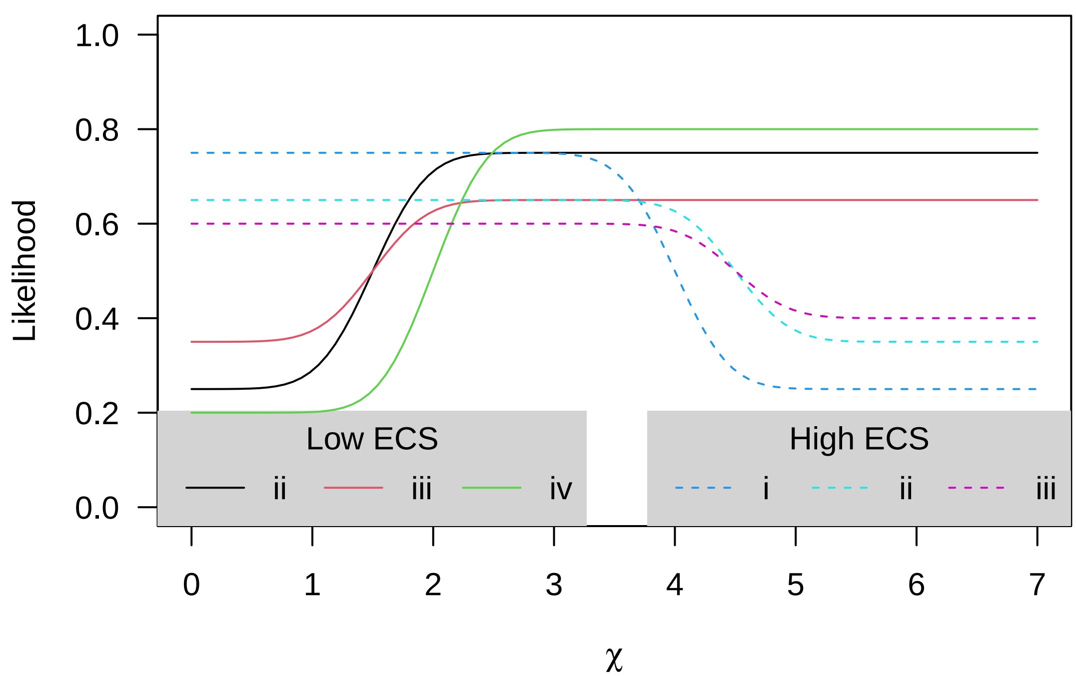
Here, χ is the unknown Equilibrium Climate Sensitivity (ECS), and

```
e <- c(Lii = .25, Liii = .35, Liv = .2, Hi = .75, Hii = .65, Hiii = .6)
c <- c(Lii = 1.5, Liii = 1.5, Liv = 2, Hi = 4, Hii = 4.5, Hiii = 4.5)
```

are "data" but really the authors' assessment of data given in table S2 of the appendix on page 28. Here `c` is a vector of hypothesized values of the ECS and `e` is a vector of probabilities of observing evidence more extreme than the hypothesized value.

Figure 2a in Stevens et al. (2016) can then be recreated with

```
curve(L(chi, e[1], c[1]), from = 0, to = 7, ylim = 0:1, xname = "chi",
      xlab = expression(chi), ylab = "Likelihood", las = 1)
legend("bottomleft", legend = c("ii", "iii", "iv"), lty = 1, col = 1:3,
       title = "Low ECS", ncol = 3, box.lwd = NA, bg = "lightgrey")
legend("bottomright", legend = c("i", "ii", "iii"), lty = 2, col = 4:6,
       title = "High ECS", ncol = 3, box.lwd = NA, bg = "lightgrey")
for (j in 2:3) curve(L(chi, e[j], c[j]), add = TRUE, col = j, xname = "chi")
for (j in 4:6) curve(L(chi, e[j], c[j]), add = TRUE, col = j, xname = "chi", lty = 2)
```



Use your `dratio` and `L` functions along with `integrate` to reproduce the black and blue curves in Figure 2b of Stevens et al. Note that "the" likelihood used is actually the product of the six contributions to it.