GR5065 Homework 4

Ben Goodrich

Due March 28, 2022 at 4PM

```
set.seed(20220328)
```

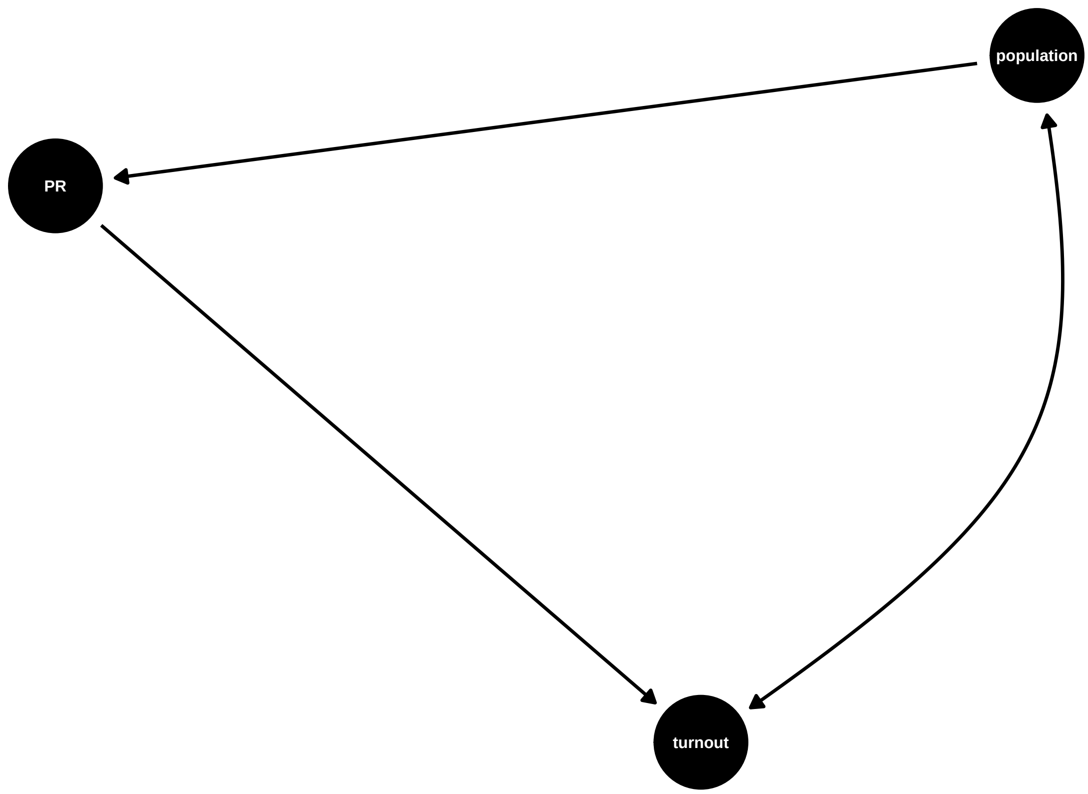
1 Voter Turnout in France

```
library(dplyr)
Eggers <- readRDS("Eggers.rds") %>%
  mutate(PR = as.integer(rrv >= 0)) # has a PR system
```

1.1 Regression Discontuity Designs and Directed Acylic Graphs

Explain, using a DAG, how regression discontinuity designs identify the average causal effect under the Adjustment Criterion. You can use the regression discontinuity design in Eggers' paper if that helps you explain, but the logic of it would be true for any "forcing variable" (which is population in this case), treatment (system of government in this case), and outcome (voter turnout in this case).

```
library(ggdag)
dagify(turnout ~ PR, PR ~ population, turnout ~~ population) %>%
  ggdag(text_size = 2) + theme_dag_blank()
```



In a regression discontinuity design, the treatment variable (PR system in this case) has exactly one parent, known as the "forcing variable" (population in this case). The Adjustment Criterion can be satisfied by conditioning on all parents of treatment. However, conditioning on a continuous forcing variable is less straightforward than conditioning on a binary variable. The choice made in this paper is to filter to municipalities whose populations are sufficiently close to the threshold required for a PR system. But, you also need to include the forcing variable in the regression model, although it is not necessary to assume it has a linear relationship to the outcome. Often researchers will try to use a more flexible functional form, in which case they should definitely be using Bayesian methods to propagate the uncertainty about the functional form through to the predictions.

1.2 Drawing from the Prior Predictive Distribution

As explained on page 144, Eggers estimates a basic model (before adding additional covariates, which you do not need to worry about on this homework) where the percentage of voter turnout is conditionally normal, given the population of the municipality. There are five unknown parameters to estimate:

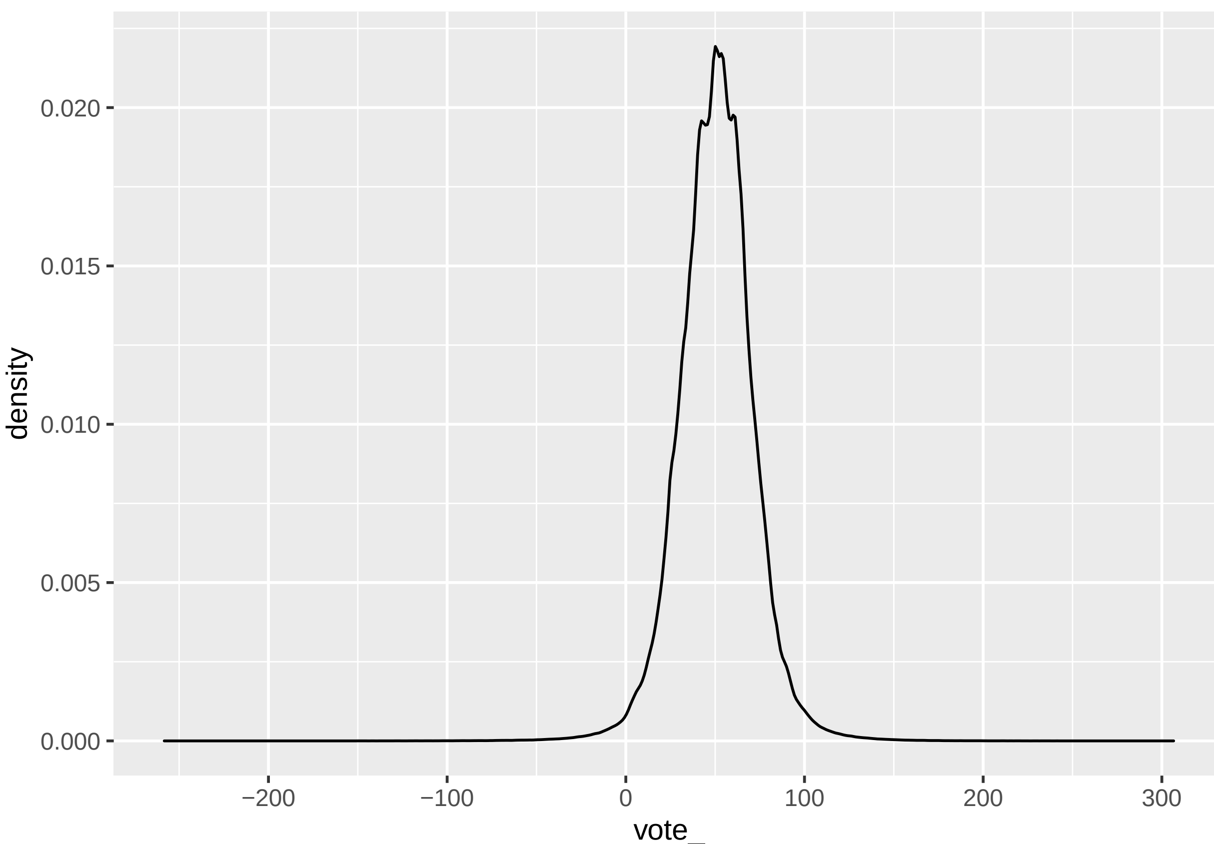
1. β_0, which is the intercept
2. τ, which is the coefficient on the dummy variable for a PR system (due to population being at least $3,500$)
3. β_1, which is the coefficient on the logarithm of the ratio of population to $3,500$ so that `rrv` would be zero in the data if the population were exactly $3,500$
4. β_2, which is the coefficient on the interaction between `PR` and `rrv`, which can be interpreted as how much more sensitive voter turnout is to log population in PR systems than in plurality-rule systems
5. σ, which is the standard deviation of the errors when predicting the percentage of voter turnout with only the previous four variables only

Choose priors for the unknowns and draw 1000 realizations from the prior predictive distribution of voter turnout for each of the $N = 35891$ municipalities in the dataset using your own R code (as opposed to calling `stan_glm` with `prior_PD = TRUE`). In this case, you should *not* center the predictors because `Eggers$rrv` is already constructed so that it is zero when the population is exactly $3,500$. Thus, the intercept, β_0, can essentially be interpreted as the expected percentage of voter turnout for a municipality that has $3,499$ people and thus a plurality-rule system.

```
PR  <- Eggers$PR
rrv <- Eggers$rrv
vote_ <- t(replicate(1000, {
  beta_0_ <- rnorm(1, mean = 50, sd = 15)
  tau_ <- rnorm(1, mean = 0, sd = 1)
  beta_1_ <- rnorm(1, mean = 0, sd = 5)
  beta_2_ <- rnorm(1, mean = 0, sd = 2.5)
  mu_ <- beta_0_ + tau_ * PR + beta_1_  * ifelse(PR == 0, rrv, 0) +
                    (beta_2_ + beta_1_) * ifelse(PR == 1, rrv, 0)
  sigma_ <- rexp(1, rate = 0.1)
  epsilon_ <- rnorm(n = length(mu_), mean = 0, sd = sigma_)
  y_ <- mu_ + epsilon_
  y_
}))
colnames(vote_) <- Eggers$com.name
```

1.3 Checking the Prior Predictive Distribution

```
library(ggplot2)
tibble(vote_ = c(vote_[ , Eggers$PSDC99 >= 1750 & Eggers$PSDC99 <= 5250])) %>%
  ggplot + geom_density(aes(vote_))
```



These are mostly between 0 and 100, as they should be when the outcome is a percentage, and span the entire range, which is appropriate when we are not very sure what the percentage turnout will be in municipal elections for a country we are not too familiar with.

```
quantile(vote_[ , Eggers$PSDC99 == 390350], probs = c(0.0625, 0.25, 0.5, 0.75, 0.925))
```

```
##      6.25%        25%        50%        75%      92.5%
##   1.039369  29.803668  52.486776  75.947362 102.050422
```

This is putting about a 6% chance on turnout being negative and a similar chance on turnout being above 100%, both of which are impossible. While this would not be horrible, clearly there is room for improvement if we were to widen the bandwidth window to include the largest municipality. The reason why there is a greater chance under the prior of Toulouse having inadmissible values is because we are leaning more heavily on the assumption that the effect of having larger population is linear among municipalities with PR systems. When the outcome variable is a percentage, the relationships cannot be linear over the entire range of unbounded predictors. Researchers often make linearity assumptions over subsets of that range, which is what Eggers is doing here with the small bandwidth window(s).

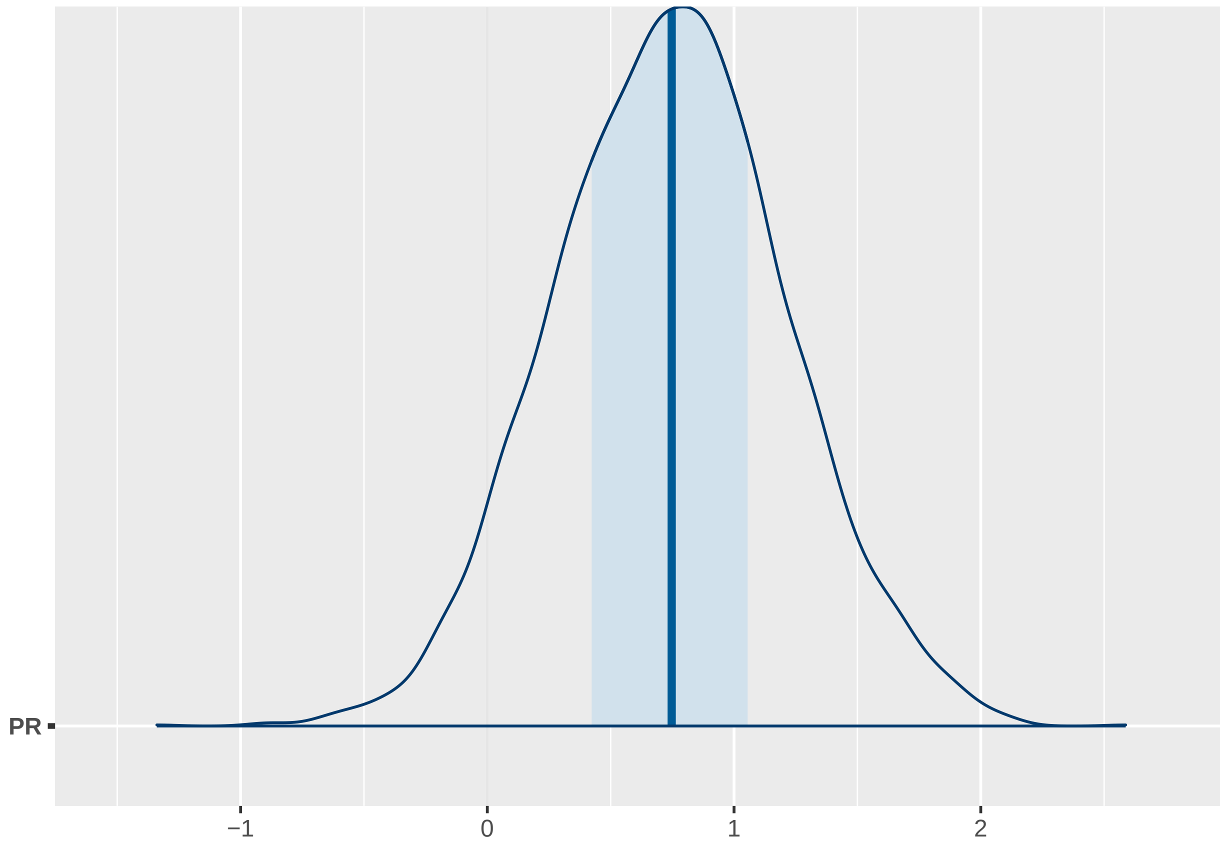
1.4 Posterior Distribution

```
library(rstanarm)
options(mc.cores = parallel::detectCores())
```

```
post_RDD <- stan_glm(to.2008 ~ rrv * PR, data = Eggers, family = gaussian,
                     subset = PSDC99 >= 1750 & PSDC99 <= 5250,
                     prior_intercept = normal(location = 50, scale = 15),
                     prior = normal(location = 0, scale = c(5, 1, 2.5)),
                     prior_aux = exponential(rate = 0.1), seed = 20220328)
```

1.5 Interpretation

```
plot(post_RDD, plotfun = "areas", pars = "PR")
```



Our beliefs about the effect of introducing a PR system is fairly Gaussian, with a center a bit less than 1 and a standard deviation of about $\frac{2}{3}$. Thus, the probability that the effect is positive is considerably greater than the chance it is negative; specifically the posterior probability of being negative is

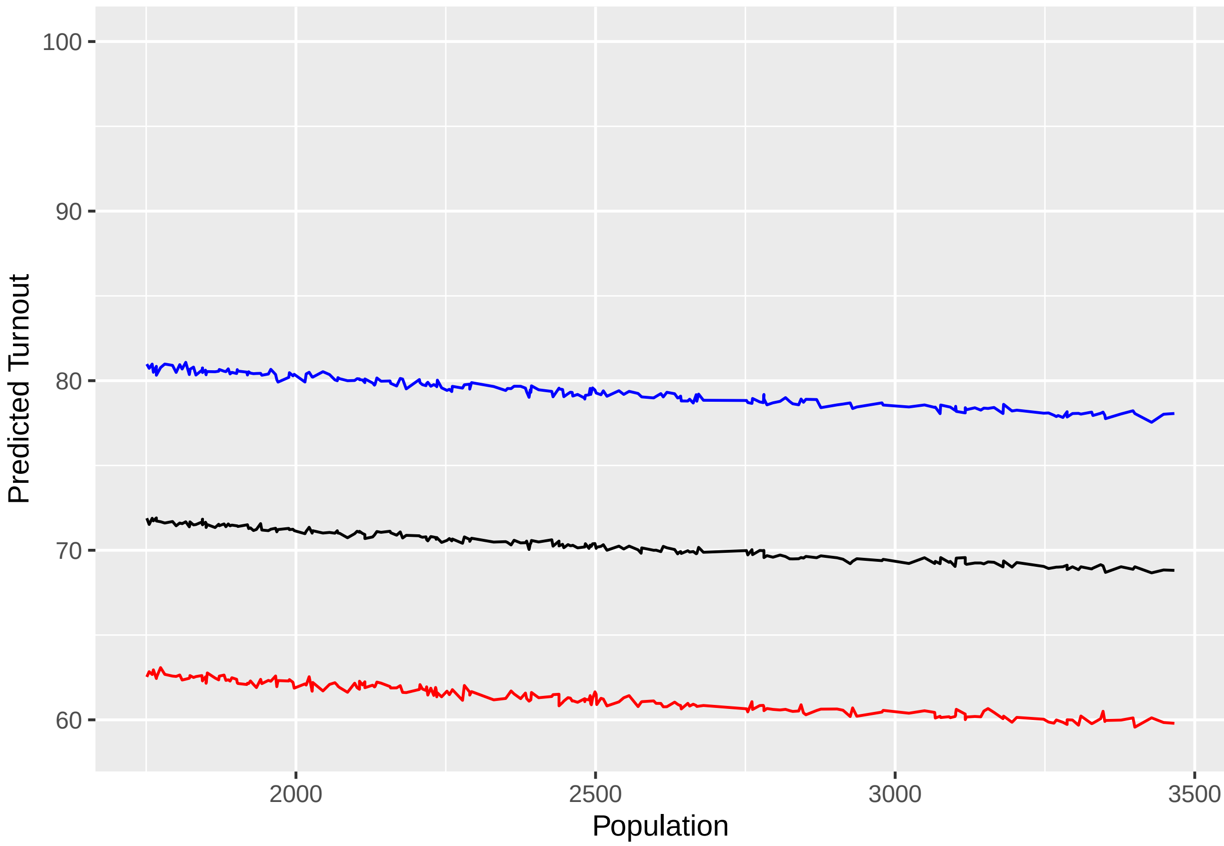
```
mean(as.data.frame(post_RDD)$PR < 0)
```

```
## [1] 0.05325
```

Fortunately, this is not a Frequentist p-value otherwise, but a Frequentist p-value is often mistaken for the posterior probability that a causal effect "has the wrong sign".

1.6 Prediction

```
Eggers_missing <- filter(Eggers, PSDC99 >= 1750, PSDC99 <= 5250, is.na(to.2008)) %>%
  select(-to.2008)
PPD <- posterior_predict(post_RDD, newdata = Eggers_missing)
quants <- t(apply(PPD, MARGIN = 2, FUN = quantile, probs = c(0.1, 0.5, 0.9)))
colnames(quants) <- c("low", "median", "high")
bind_cols(Eggers_missing, as.data.frame(quants)) %>%
  ggplot + geom_line(aes(x = PSDC99, y = low), color = "red") +
  geom_line(aes(x = PSDC99, y = median), color = "black") +
  geom_line(aes(x = PSDC99, y = high), color = "blue") +
  xlab("Population") + ylab("Predicted Turnout") + ylim(59, 100)
```



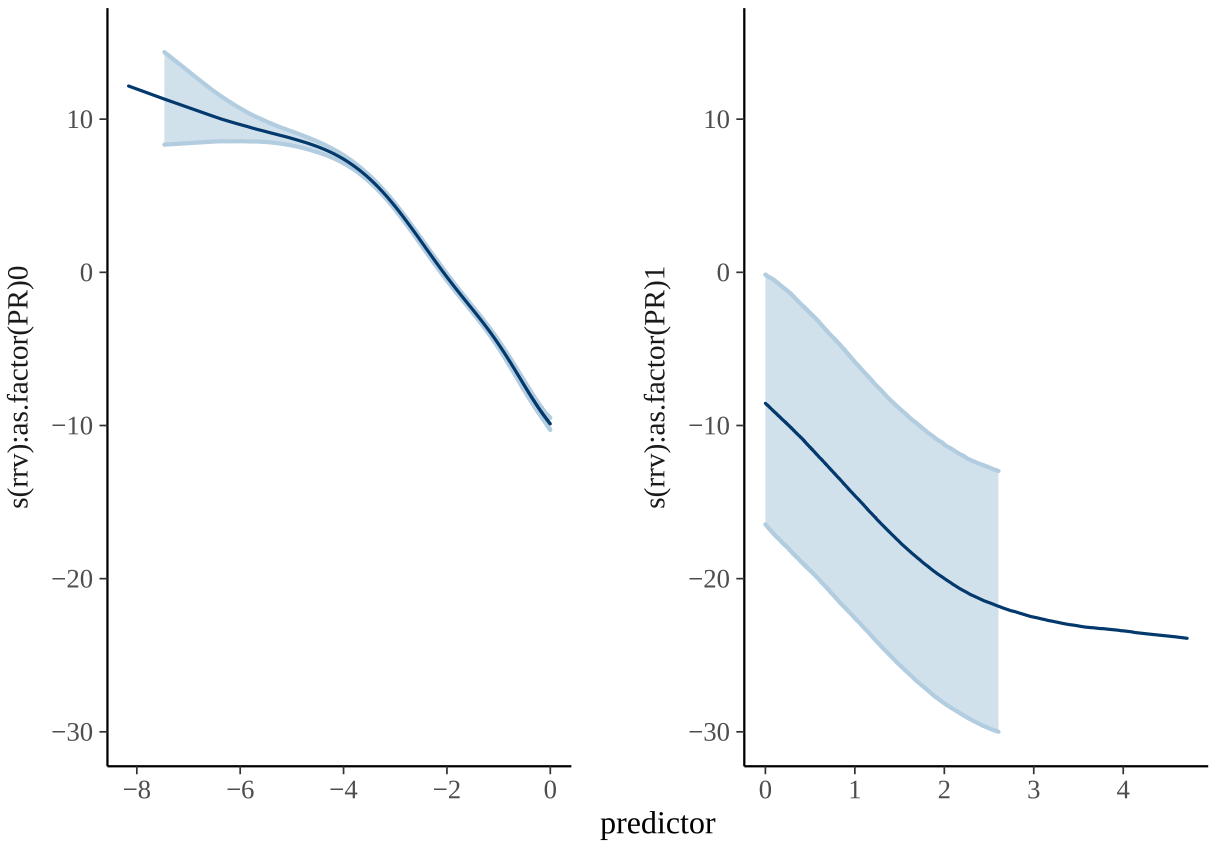
All of these predictions are admissible, with a slowly decreasing median of around 70% and bounds on an 80% predictive interval that are about 10% in either direction.

1.7 Addendum

It is more common to see people use more flexible functional forms for the forcing variable in regression discontinuity designs. However, these can be a disaster if you only obtain point estimates and do not propagate the uncertainty in the functional form all the way through to the final conclusions. Doing so also allows you to expand (or eliminate) the bandwidth window. It would look like

```
post_GAM <- stan_gamm4(to.2008 ~ PR + s(rrv, by = as.factor(PR)),
                       data = Eggers, family = gaussian,
                       prior_intercept = normal(location = 50, scale = 15),
                       prior_aux = exponential(rate = 0.1), seed = 20220328)
```

```
plot_nonlinear(post_GAM) + ylim(-30, 15)
```



As it turns out, it appears as if the relationship is fairly linear over a wide range before and after the threshold for establishing a PR system. We can then estimate the causal effect of crossing the threshold with

```
counterfactual <- data.frame(PR = 0:1, rrv = 0)
mu <- posterior_epred(post_GAM, newdata = counterfactual) # 4000 x 2
summary(mu[ , 2] - mu[ , 1])
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -0.5899  0.9125  1.2129  1.2084  1.5157  2.7115
```

2 Minimum Wage Increases

```
library(haven)
Manning <- as_factor(read_dta("ManningElusiveEmployment.dta")) %>%
  mutate(agecat = as.factor(agecat),
         quarterly_date = as.factor(quarterly_date - 75))
glimpse(Manning)
```

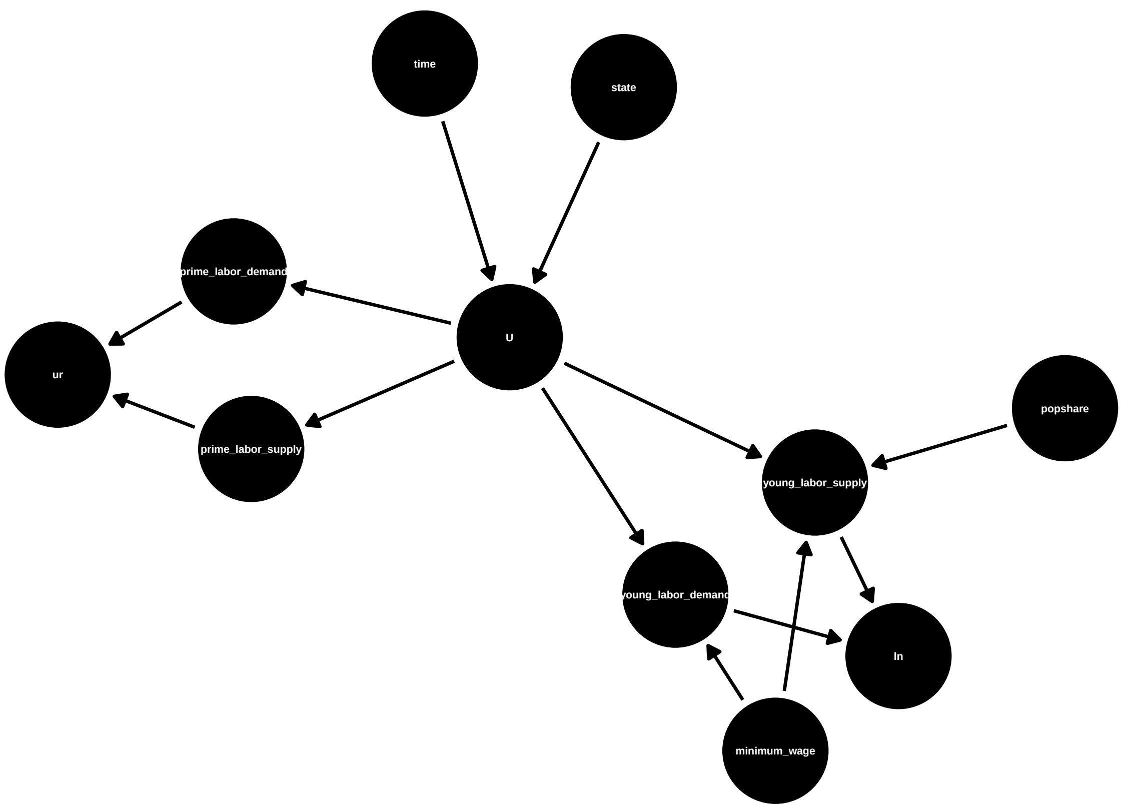
```
## Rows: 41,818
## Columns: 15
## $ year           <dbl> 1979, 1979, 1979, 1979, 1979, 1979, 1979, 1979, 1979, 1~
## $ region         <fct> east south central division, east south central divisio~
## $ statefips      <fct> alabama, alabama, alabama, alabama, alabama, alaska, al~
## $ agecat         <fct> 1, 2, 3, 4, 5, 1, 2, 3, 4, 5, 1, 2, 3, 4, 5, 1, 2, 3, 4~
## $ qtr            <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1~
## $ popshare       <dbl> 0.1194241, 0.1388525, 0.1032453, 0.3838268, 0.2546513, ~
## $ ln             <dbl> -1.0879076, -0.4460275, -0.3935704, -0.3444984, -0.5263~
```

```
## $ lemp          <dbl> 12.53646, 13.32907, 13.08522, 14.44738, 13.85520, 10.28~
## $ lpop          <dbl> 13.62437, 13.77510, 13.47879, 14.79188, 14.38158, 11.23~
## $ ne            <dbl> 826013.62, 960392.75, 714110.19, 2654790.75, 1761330.62~
## $ ur            <dbl> 0.9393107, 0.9393107, 0.9393107, 0.9393107, 0.9393107, ~
## $ statecensus   <fct> "alabama", "alabama", "alabama", "alabama", "alabama", ~
## $ lw            <dbl> 6.087777, 6.294876, 6.277453, 6.401202, 6.350931, 6.306~
## $ quarterly_date <fct> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1~
## $ lmin          <dbl> 5.669881, 5.669881, 5.669881, 5.669881, 5.669881, 5.828~
```

2.1 Directed Acylic Graph

Start with the "stylized representation of the impact of the minimum wage on employment" in Figure 5 on page 23 and then draw a DAG that corresponds to what you think Manning thinks the data-generating process is for "Specification 1" in Table 2. Is the total causal effect of changing the minimum wage identified using the Adjustment Criterion in that DAG?

```r
dagify(young_labor_supply ~ minimum_wage + U + popshare,
       young_labor_demand ~ minimum_wage + U,
       prime_labor_supply ~ U,
       prime_labor_demand ~ U,
       U ~ state + time,
       ln ~ young_labor_supply + young_labor_demand,
       ur ~ prime_labor_supply + prime_labor_demand,
       latent = "U") %>%
  ggdag(text_size = 1.5, node_size = 18) + theme_dag_blank()
```



This DAG makes it more explicit that Manning is assuming that changes in the minimum wage affect the labor supply of and labor demand for young people, but do not affect the supply and demand for prime-age workers who typically earn more than the minimum wage. There is an unobservable variable U that consists of many factors that affect the economy, such as monetary policy. However, young labor supply and demand are colliders along the path from minimum wage to U and naturally block them, while leaving the causal paths from minimum wage through young labor supply and demand to the young employment rate unblocked.

Manning also conditions on several other variables that are theoretically irrelevant to the identification of the causal effect (if infinite data were available), but are perhaps practically relevant with finite data to reduce the conditional variance in the outcome(s) of interest and thereby obtain more precise estimates. In particular, the prime-age unemployment rate is a function of U, so conditioning on it reduces the unexplained variation in the young employment rate. But provided there is no causal path from the minimum wage to the prime-age unemployment rate, conditioning on `ur` does not alter the fact that the Adjustment Criterion is satisfied to identify the Average Causal Effect of the minimum wage on the young employment rate. Similarly, U may be a consequence of things that vary by state over time or are similar at a moment in time across states, which are crudely adjusted for by introducing many dummy variables.

It is less clear what `popshare`, which is the share of young people in the state at a point in time, is doing in Manning's model. Manning says that it is "to account for the fact that labor market outcomes for teens may be affected by how many of them there are", but then does not include an interaction term between `popshare` and `lmin`. However, because the minimum wage and the young employment rate are put into the linear regression model in logarithms, conditioning on any variable affects the nonlinear relationship between the levels of these variables.

2.2 Frequentist Inference

Manning only uses Frequentist estimators in this paper, but does not test any null hypotheses. There is nothing in the paper to suggest that Manning is interested in making any Frequentist inference, nor is it clear what Frequentist inference the data might speak to, since the data are not a random sample from any population (i.e. Manning has data on all states over the period of interest).

Rather, Manning seems to use the confidence intervals in Figures 1 through 4 as if they were Bayesian credible intervals. The entire article hinges around the fact that almost all economists historically believed that increasing the minimum wage above the market rate would have a substantial negative effect on employment because employers would be unwilling to pay that much for entry-level workers. However, many economists changed their minds in recent decades as evidence, such as that in Manning's paper, accumulated that the employment effect was small and perhaps might be positive. Thus, Figure 3 is intended to establish what Manning believes about the effect of minimum wage on the young employment rate, and thus it would have made more sense had Minning used Bayesian estimation techniques.

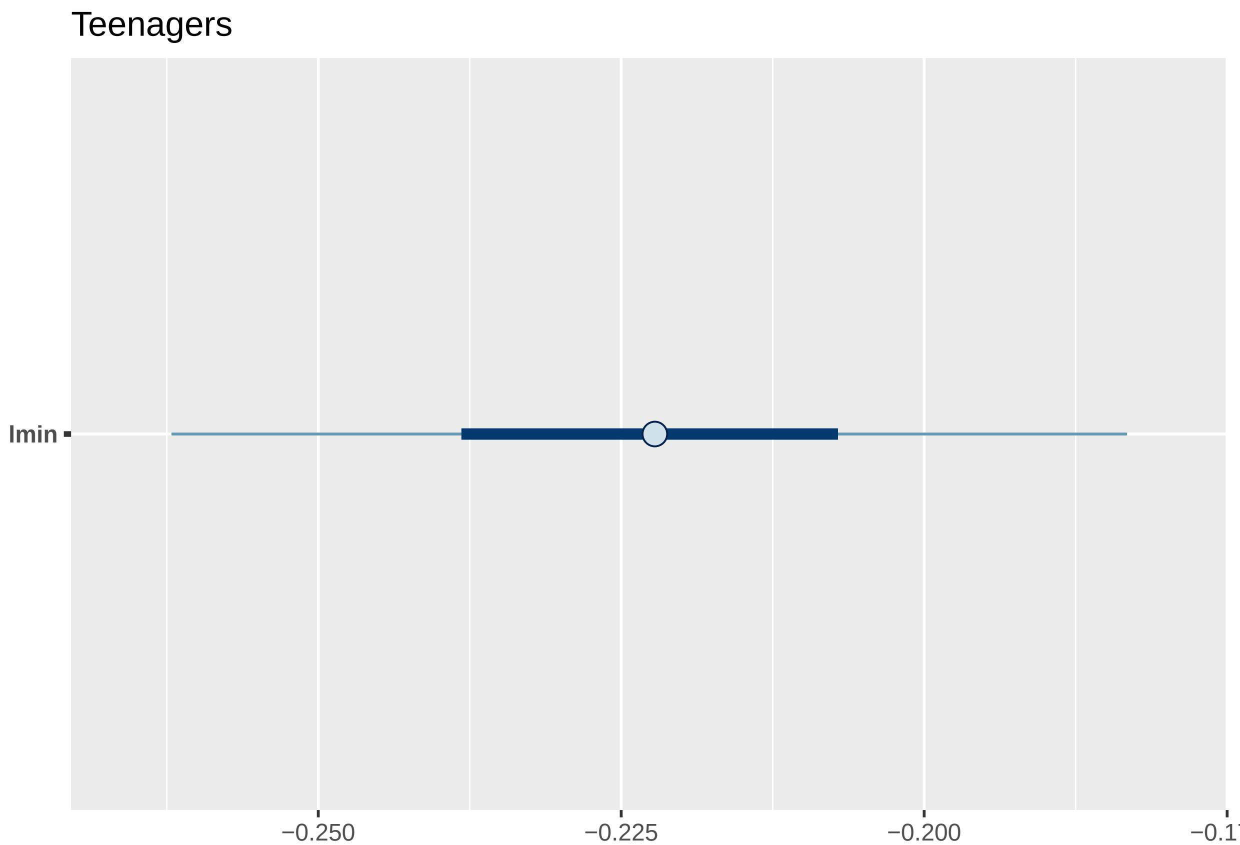
In addition, an important question is which model specification should be used because the conclusions differ somewhat. Since models are not random variables, Frequentist estimation techniques cannot answer questions like "What is the probability that specification 1 is best?" We have not gotten into Bayesian answers to such questions yet, but answering them is considerably more natural under the subjective view of probability.

2.3 Bayesian Inference

```r
post_teen <- stan_lm(ln ~ lmin + popshare + ur + statefips + quarterly_date,
                     data = Manning, subset = agecat == 1,
                     prior_intercept = normal(location = log(0.5), scale = 1),
                     prior = R2(0.2, what = "mode"), seed = 20220328)
post_20s <- update(post_teen, subset = agecat == 2)
```

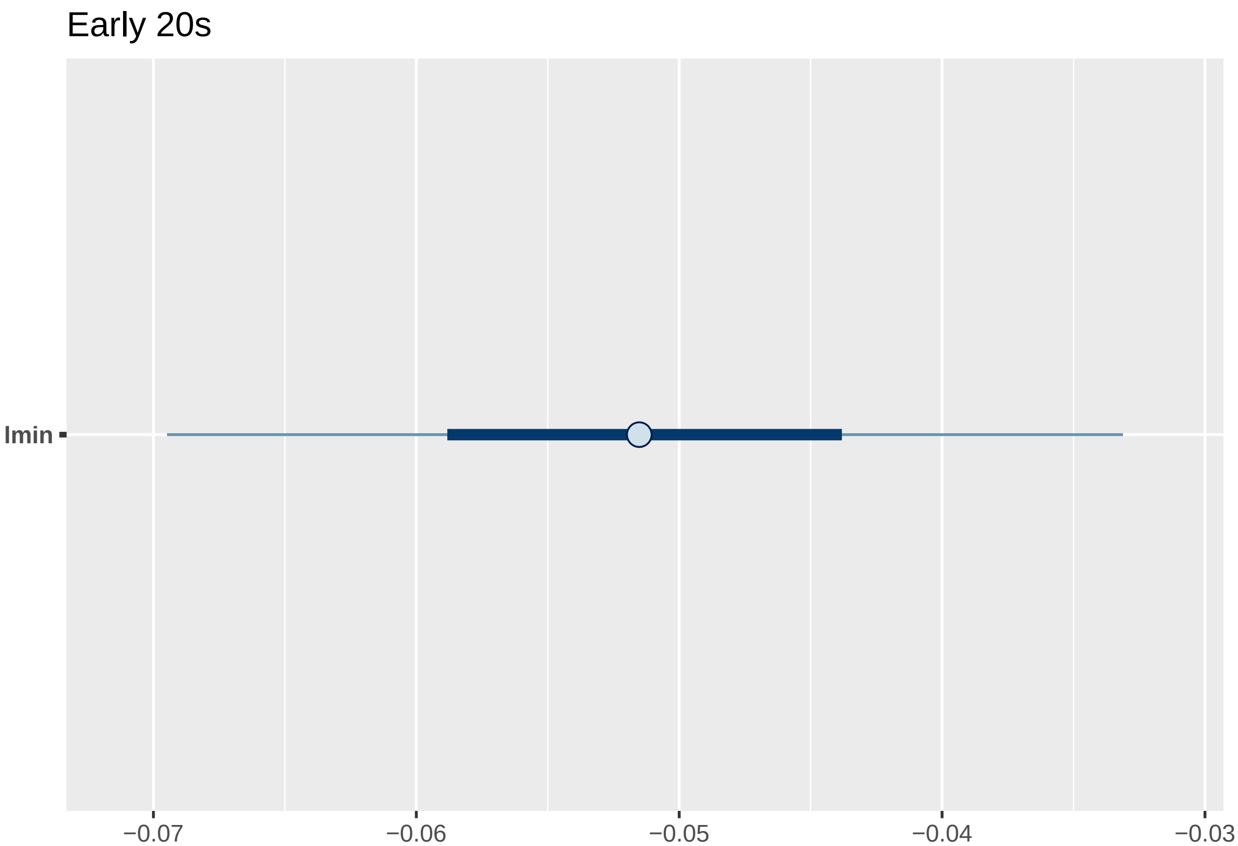
2.4 Interpretation

```r
plot(post_teen, pars = "lmin") + ggtitle("Teenagers")
```



This is similar to the corresponding 95% confidence interval in Figure 3 of Manning's paper in the sense that it is to the left of zero, but ours is much narrower and can be interpreted in a more straightfoward manner as saying the effect is almost certainly between -0.275 and -0.15.

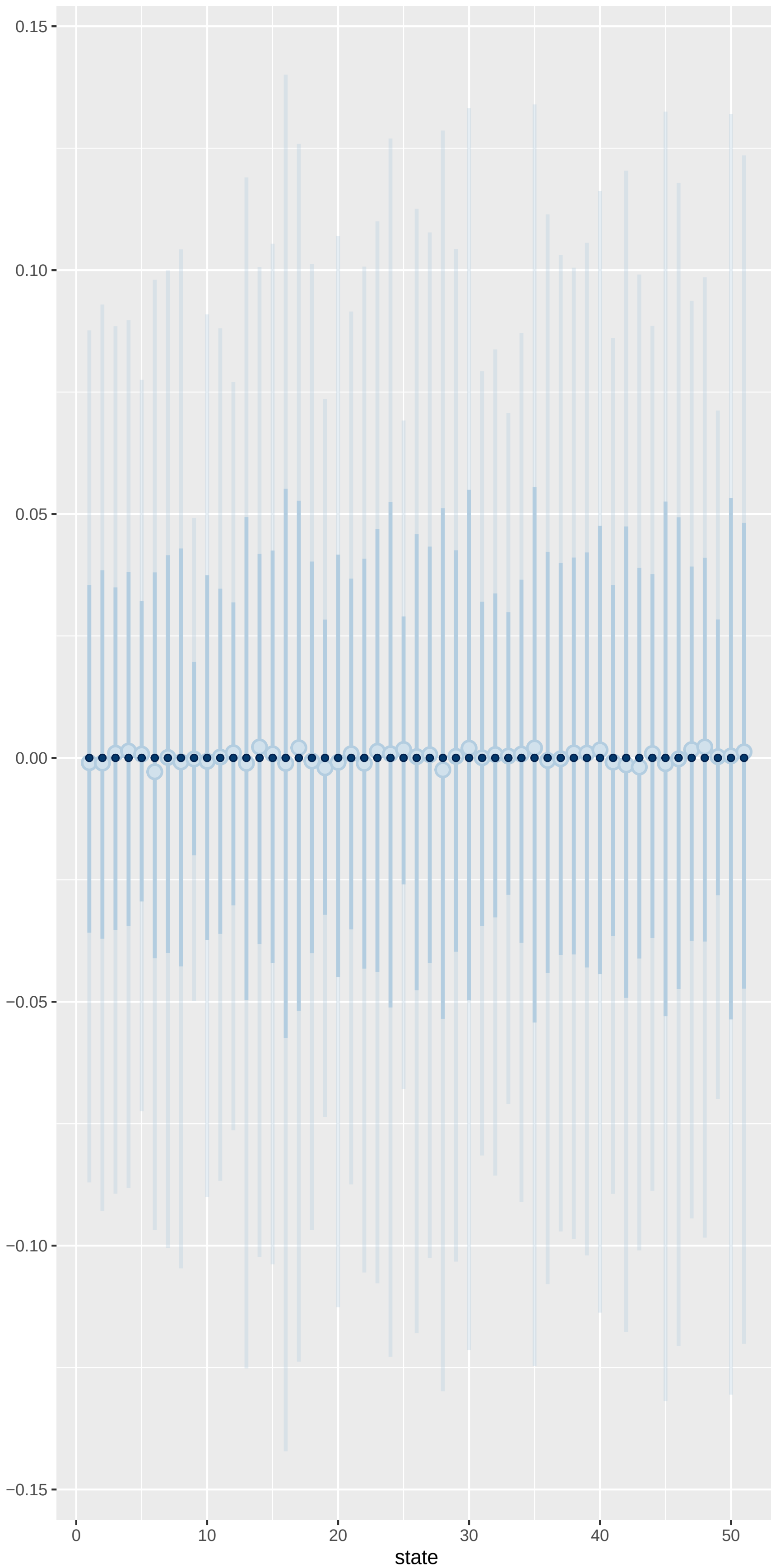
```r
plot(post_20s, pars = "lmin") + ggtitle("Early 20s")
```



This is similar to the corresponding 95% confidence interval in Figure 4 of Manning's paper but the differences are essentially the same as in the previous plot for teenagers.

2.5 Prediction

```
recent <- filter(Manning, agecat == 1, year == 2019, qtr == 4)
factual <- exp(posterior_predict(post_teen, newdata = recent))
recent_ <- mutate(recent, lmin = log(pmax(exp(lmin), 15 * 40)))
counterfactual <- exp(posterior_predict(post_teen, newdata = recent_))
difference <- counterfactual - factual
bayesplot::ppc_intervals(y = rep(0, ncol(difference)), yrep = difference) +
  theme(legend.position = "none") + xlab("state")
```


0.15
0.10
0.05
0.00
−0.05
−0.10
−0.15
0
10
20
30
40
50
state

Based on this plot, we can say that any effect of changing the federal minimum wage to $15 per hour would have such a small effect on employment that it is completely obscured by the noise in the predictions.

I suspect that the data are not documented correctly and the column called `lmin` is not actually the logarithm of anything but the level of the minimum wage in 2019 dollars per hour. But that does not matter for the purposes of this homework.

2.6 Prior Predictive Distribution

First, we set up all the variables and center them

```
Manning <- mutate(Manning, statefips = factor(statefips))
teen <- model.matrix(~ lmin + popshare + ur +
                     statefips + quarterly_date,
                     data = filter(Manning, agecat == 1))
# drop the column of 1s and center everything else
teen <- apply(teen[ , -1], MARGIN = 2, FUN = function(x) x - mean(x))
```

Then, it is much easier to draw from the predictive distribution

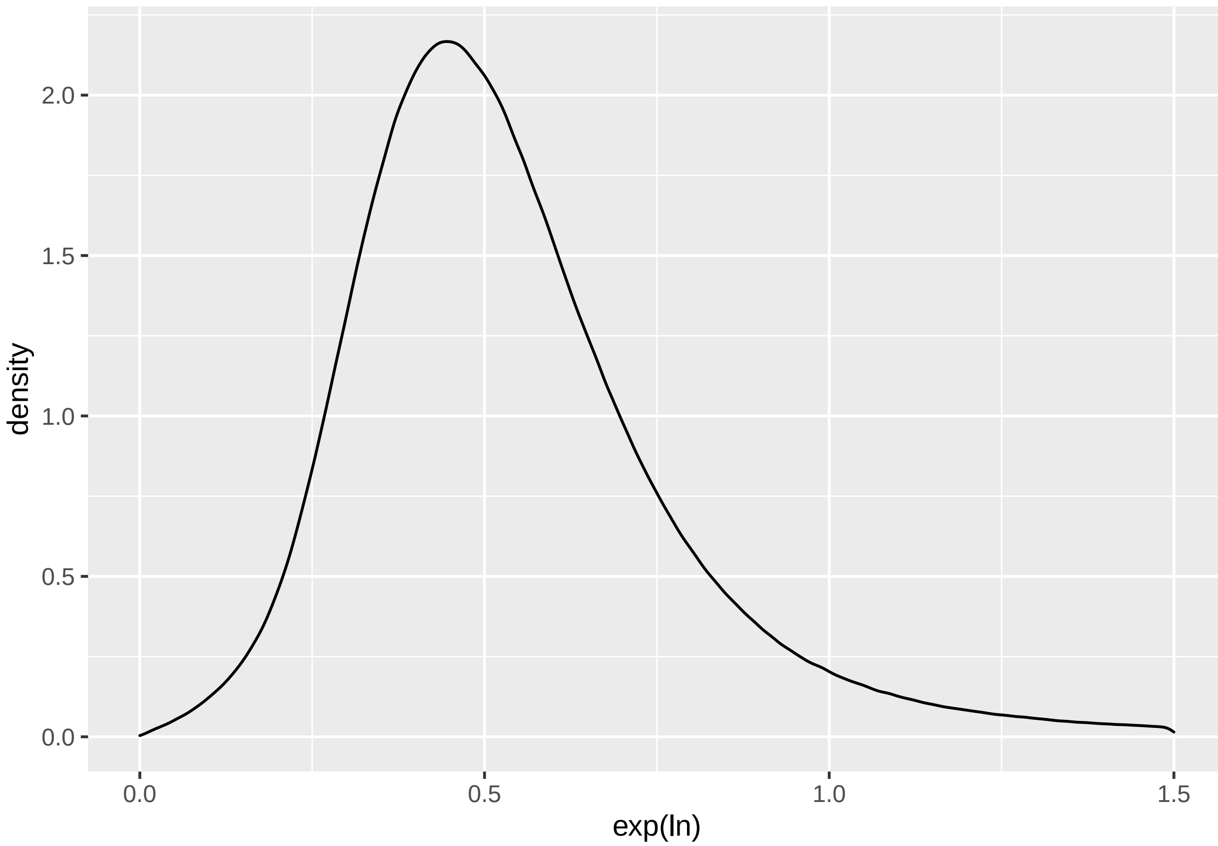
```
prior_predictive <- t(replicate(1000, {
  # first stage
  alpha_1 <- rnorm(1, mean = log(20), sd = 1)
  beta_1  <- rnorm(1, mean = -0.1, sd = 0.2) # in percent
  delta_1 <- rnorm(ncol(teen) - 3, mean = 0, sd = 1)

  # intermediate stage
  pred <- alpha_1 + c(teen[ , -(1:2)] %*% c(beta_1, delta_1))

  # second stage
  w_0 <- median(pred - teen[, "lmin"])
  alpha_2 <- rnorm(1, mean = log(0.5), sd = 0.25)
  beta_2 <- rnorm(1, mean = -0.1, sd = 0.1)
  gamma_2 <- rnorm(1, mean = 0, sd = 0.05)
  lambda_2 <- rnorm(1, mean = 0, sd = 0.1)
  theta_2 <- rnorm(1, mean = -0.1, sd = 0.1)
  delta_2 <- rnorm(ncol(teen) - 3, mean = 0, sd = 0.1)
  mu_2 <- alpha_2 + beta_2 * teen[ , "lmin"] +
    gamma_2 * (pred - teen[ , "lmin"] - w_0)^2 +
    c(teen[ , -1] %*% c(lambda_2, theta_2, delta_2))
  sigma_2 <- rexp(1, rate = 4)
  epsilon_2 <- rnorm(length(mu_2), mean = 0, sd = sigma_2)
  y_2 <- mu_2 + epsilon_2
  return(y_2)
}))
```

This still puts some prior probability on the even that more than 100% of teenagers have jobs, but its center and most of its mass is quite reasonable. This illustrates that getting halfway decent draws from a prior predictive distribution is not too difficult if you think about the sample space of the outcome, center the predictors so that the intercept is the expected outcome for a unit with average predictors, and make the parameters have the right units.

```
ggplot(tibble(ln = c(prior_predictive))) +
  geom_density(aes(x = exp(ln))) + xlim(0, 1.5)
```


density
2.0
1.5
1.0
0.5
0.0
0.0
0.5
1.0
1.5
exp(ln)