MACS_PS3

Neta Grossfeld

10/24/2018

Perspectives of Computational Analysis - Fall 2018

```
library(tidyverse)
library(ggplot2)
set.seed(1234)
```

2.a

```
income_neta <- function(base_inc, rho, g, sigma, n_years, start_year = 2020){

  errors_neta <- rnorm(n_years, mean = 0, sd = sigma)

  income_log_neta <- vector("numeric", n_years)

  loan <- 95000

  for(year in seq_len(n_years)){
    if(year == 1){
      income_log_neta[[year]] <- log(base_inc) + errors_neta[[year]]
    } else {
      income_log_neta[[year]] <- (1-rho)*(log(base_inc)) + g*(year) + rho*(income_log_neta[[year-1
]]) + errors_neta[[year]]
    }
  }

  data_frame(inc = exp(income_log_neta),
             year = 2020 + seq_len(n_years) - 1)
}
```

```
n_sims <- 10000

n_years <- 40

simulated_income_neta <- n_sims %>%
  rerun(income_neta(base_inc = 80000, sigma = .13, rho = .4, g = .025,
                    n_years = n_years, start_year = 2020)) %>%
  bind_rows(.id = "id") %>%
  select(id, year, inc)

#View(simulated_income_neta)
head(simulated_income_neta)
```

MACS 30000
Problem Set 3
Neta Grossfeld
10/24/18

In her article, *Simulation in Sociology,* Moretti makes the argument that the simulations are useful for exploring theories based on the real world, regardless if the simulations' outcomes themselves can be observed in the real world. If these observations were available in the real-world, there would be no need for conducting a computer simulation to explore the boundaries of the theory.

Computer simulations allow us to change certain variables in a theory or model that plausibly represents an observed reality. However, simulations are perhaps most useful in their ability to give us insight into what types of variables, and their different possible values, may affect outcomes based on the theory. Basically, simulations are used as experiments on the theory, which is meant to be representative of the real-world. Even if we cannot tie back certain variables' values to the real-world, there are still benefits to simulations regardless of their validity. For example, in the cellular automata model, it is more useful to compare models and see the effects of differences in certain variables. In the same way, a multi-agent system or model can also be explored. Since we can't observe all the different, possible, and counter-factual scenarios in the real-world, simulations make it possible to explore the theory that is based on observations in the real-world. Rather than defining validity as the degree to which a simulation is representative of the real-world, it is much more interesting to verify the validity of the effect certain variables in the model may have on the outcome, and whether the theory is robust enough to withstand differences in the configuration of parameters.

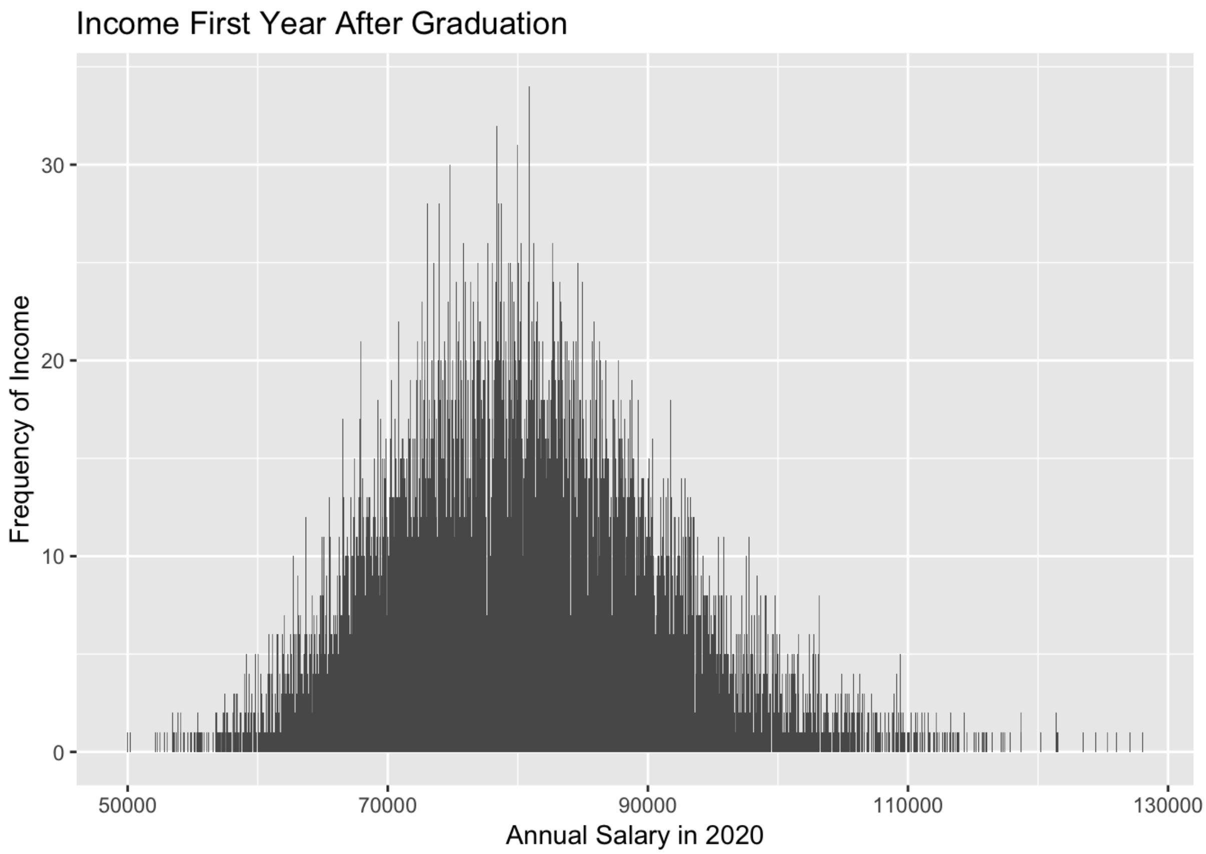
One example of a model that Moretti cites as an example that exhibits the characteristic of "dynamic feedback" is Forrester's 1971 *World Dynamics,* which models population growth. In this model, a new technological innovation could have on relationship between population growth and natural resources. For example, if a new invention has the capability to rapidly increase the production of natural resources, there will likely be a related change in the rate of population growth. An example of a political science system that exhibits dynamic feedback is the presidential election or elections in general in the United States. We can already observe that Trump's election in 2016 has mobilized a huge part of the United States into protesting his policies and actively voting for opponents. A research question could be: What would the differences be in the dynamic feedback of the presidential election if either party had won?

```
## # A tibble: 6 x 3
##   id     year     inc
##   <chr> <dbl>   <dbl>
## 1 1      2020  68382.
## 2 1      2021  81886.
## 3 1      2022 100216.
## 4 1      2023  71322.
## 5 1      2024  91550.
## 6 1      2025 104769.
```

```
simulated_income_neta %>%
  filter(id == 1) %>%
  ggplot(aes(year, inc)) +
  geom_line() +
  labs(title = "Simulated income increase over time (one simulation)",
       x = "Year",
       y = "Annual Income") +
  scale_y_continuous(labels = scales::dollar)
```

2b. Histogram and Percentages Over/Under

```
simulated_income_neta %>%
  filter(year == 2020) %>%
  ggplot(aes(inc)) +
  geom_histogram(binwidth = 50) +
  labs(title = "Income First Year After Graduation",
       x = "Annual Salary in 2020",
       y = "Frequency of Income")
```



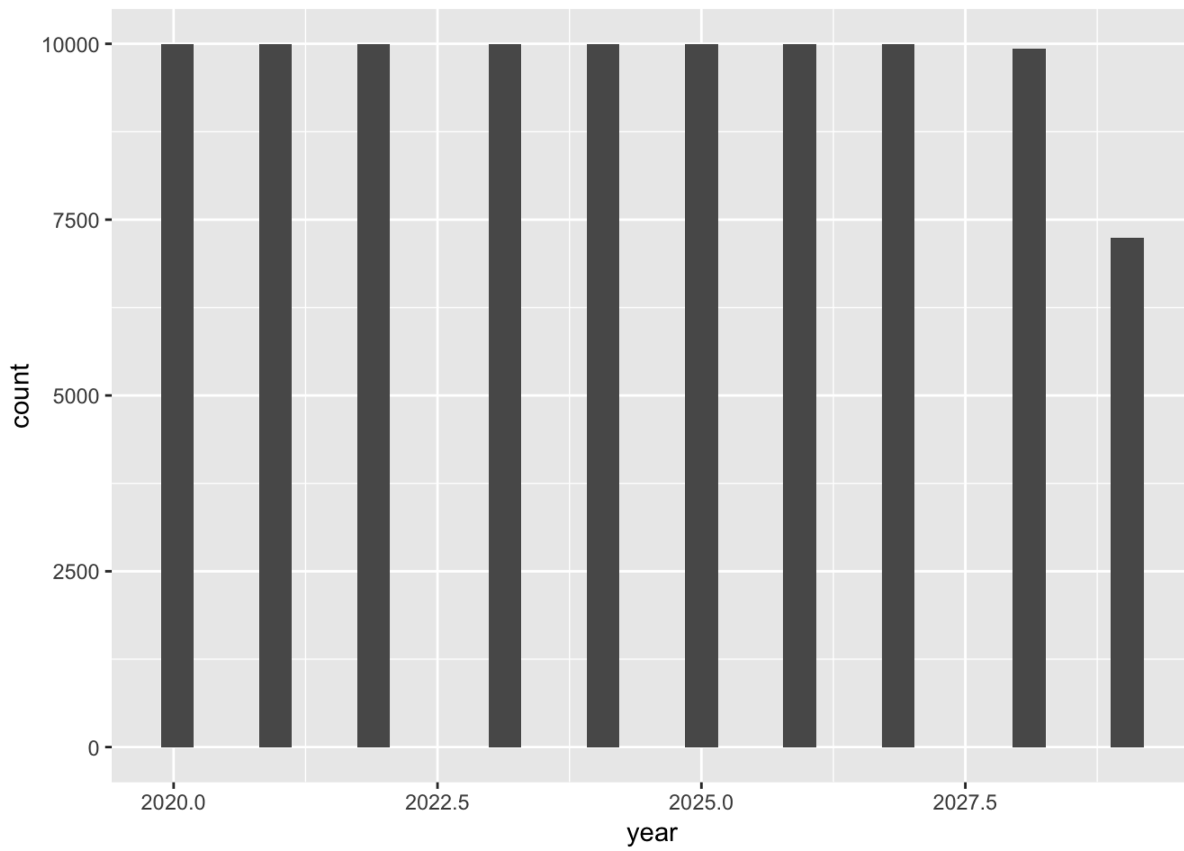

```
percentage_over_under <-
  simulated_income_neta %>%
  filter(year == 2020) %>%
  mutate(over_100k = ifelse(inc > 100000, 1, 0), under_70k = ifelse(inc < 70000, 1, 0))

percent_over = (sum(percentage_over_under$over_100k)/n_sims)*100
# The percentage of students making more than 100k in their first year is 4.32%
percent_under = (sum(percentage_over_under$under_70k)/n_sims)*100
# The percentage of students making less than 70k in their first year is 15.54%

# The distribution is relatively normally distributed, slightly skewed to the left, or the lower en
d of annual income.
```

2c. Loan Payment

```r
loan_neta <- function(base_inc, rho, g, sigma, n_years, start_year = 2020, debt = 95000, debt_pct =
 .1){

  errors_neta <- rnorm(n_years, mean = 0, sd = sigma)

  income_log_neta <- vector("numeric", n_years)

  for(year in seq_len(n_years)){
    if(year == 1){
      income_log_neta[[year]] <- log(base_inc) + errors_neta[[year]]
    } else {
      income_log_neta[[year]] <- (1-rho)*(log(base_inc)) + g*(year) + rho*(income_log_neta[[year-1
]]) + errors_neta[[year]]
    }
  }

  debt_neta <- vector("numeric", n_years)

  data_frame(inc = exp(income_log_neta),
             year = 2020 + seq_len(n_years))

  for(year in seq_len(n_years)){
    if (year == 1){
      debt_neta[[year]]<- debt
    }
    else {
     if(debt_neta[[year - 1]] > 0){
     debt_neta[[year]]<- debt_neta[[year - 1]] - (.1 * exp(income_log_neta[[year]]))
      }else{
        debt_neta[[year]] <- 0
      }
    }
  }

  data_frame(inc = exp(income_log_neta),
             year = 2020 + seq_len(n_years) - 1, debt = debt_neta)

}



simulated_loan_neta <- n_sims %>%
  rerun(loan_neta(base_inc = 80000, sigma = .13, rho = .4, g = .025,
                      n_years = n_years, start_year = 2020, debt = 95000, debt_pct = .1)) %>%
  bind_rows(.id = "id") %>%
  select(id, year, debt) %>%
  filter(debt > 0.0, year <= 2029)

head(simulated_loan_neta)
```

```
## # A tibble: 6 x 3
##   id     year   debt
##   <chr> <dbl>  <dbl>
## 1 1      2020 95000
## 2 1      2021 87151.
## 3 1      2022 77204.
## 4 1      2023 68207.
## 5 1      2024 60550.
## 6 1      2025 50165.
```

```
# histogram
simulated_loan_neta %>%
  ggplot(aes(year)) +
  geom_histogram()
```



2d. Increase Base Salary, Pay Off Loans

```
simulated_loan_neta <- n_sims %>%
  rerun(loan_neta(base_inc = 90000, sigma = .17, rho = .4, g = .025,
                      n_years = n_years, start_year = 2020, debt = 95000, debt_pct = .1)) %>%
  bind_rows(.id = "id") %>%
  select(id, year, debt) %>%
  group_by(id) %>%
  filter(debt > 0.0)

head(simulated_loan_neta)
```

```
## # A tibble: 6 x 3
## # Groups:   id [1]
##   id     year   debt
##   <chr> <dbl>  <dbl>
## 1 1      2020 95000
## 2 1      2021 87690.
## 3 1      2022 79540.
## 4 1      2023 68396.
## 5 1      2024 55569.
## 6 1      2025 42144.
```

```
# histogram
simulated_loan_neta %>%
  ggplot(aes(year)) +
  geom_histogram()
```