Perspectives on computational analysis HW3

Angela Zorro Medina

10/24/2018

Question 1

Simulation in Sociology, Moretti (2002). (3 points). Read the paper, Moretti (2002), and write a one-to-two-page paper responding to the following questions.

a)The author discusses the role of simulation as a tool for exploring theory. She also highlights the importance of establishing validity of the simulative model of the theory. That is, validity is the degree to which the theoretical constructs and their computational implementation are representative of the real-world. What are some potential weaknesses in validity that the author describes with regard to multi-agent systems and cellular automata? **b)**The author highlights dynamic feedback as a key characteristic that computer simulation is good at modeling. Dynamic feedback is where some initial stimulus changes behavior, and then that change in behavior creates new stimuli which in turn cause further behavioral change. Give an example of a model that the author cites from Sociology that exhibits this characteristic. Come up with an example of a research question on a political science topic where the underlying system exhibits dynamic feedback.

In her paper "Computer Simulation in Sociology," Sabrina Moretti explains the contribution of computer simulation to sociology, specifically dynamic systems. Dynamics systems simulations are a computational tool used to analyze complex systems that cannot be studied using traditional mathematical models. In social sciences, computer simulations have been used to analyze the variables of a global system. Despite the benefits of using computational simulations, it has not been widely implemented in all areas of social sciences. On the contrary, while they have been successfully used in economics, they have been less successful in other fields, like global politics and international relationships.

The author explains two main concerns regarding the system dynamics that have caused that they have not been equally successful in the different fields of social sciences. The first problem refers to the fact that system dynamic models cannot be used to explain changes in the structure, these models can only explain changes inside a fixed dynamic structure. In social sciences like economics when you are studying the effect of a change in a variable into another these models become extremely useful, however in other subfields like sociology, in the cases where you are trying to explain the change in the structure itself these models cannot be very useful.

The second concern is related to the quantification of the model. The principal limitation to social sciences to be studied as natural sciences is that complex social phenomenon are difficult to quantify. Not all social variables have the same measurement problems, some variables can be easily translated into numeric values, especially in economics, but there are others that cannot be universally quantified. When we approach a variable that cannot be universally quantified, there is a large space for interpretation, and when we face interpretation subjectivity arise. As a consequence, the studies on those variables are not comparable, if one author A is defining x in a different way than author B, then the results of A and B cannot be compared. Moreover, even x is defined in the same way by author A and B; they could be measuring it differently. This second problem states the concern that computational simulations cannot evaluate if the model represents reality and the model cannot tell us the connection between theory and empirical data.

The validity problem highlighted by the author can be observed closely by looking at the cellular automata and multi-agent models. The cellular automata tool allows making simulative models of biological evolution and self-reproduction. This model is based on homogenous agents that have determined positions in a specific framework. By assuming homogeneity in the behavior of agents and their modality of interactions, the cellular automata model ignores that individuals modified their attitudes, opinions, and behaviors at different moments. Moreover, the cellular automata cannot overcome the validity problem by simply applying

transitional rules at different moments because the construction of the groups that are going to make the transition at the same time still imposes important restrictions that do not reflect reality.

Multi-agent models also have a validity problem. These models aim to simulate artificial intelligence that is capable of adapting itself to the external world. Multi-agent models must have a detailed list of rules of agents' behavior, like protocols of communication and decision-making procedures that consider environmental changes and the interactions with the other agents of the system. The definition of the set of rules is the main reason why these models have validity weaknesses. If multi-agent systems use a model of rationality that is not realistic or understandable, then it cannot reflect reality. In light of this, the success of the multi-agent system relies on our capacity to develop models of rationality that are realistic. On another aspect, the multi-agent model presents a second challenge related to the standardization of the variables that come from the psychological theory. To overcome this problem, we need to find a way to standardized emotions, motivations, desires, among others, to have comparable data and comparable conclusions.

To address the validity problem, the author suggests using an extension of the "Turing Test," proposed by Taber and Timpone (1996). According to this test, if the researcher cannot distinguish between the model behavior and the real behavior, then the model can be considered valid. By looking at the simulated data and comparing it with the correspondent real data, we can verify how far or how close is our model from reality. Ideally, we would not be able to identify which is the model and which is the real data because they are both identical. Although the "Turing Test" provides some guidelines to evaluate how our model is close to reality, it does not provide any useful information regarding the measurement problems. These problems cannot be addressed with a computational tool; a computer cannot explain us exactly what does in mean to be 3 in a sociability scale from 1 to 10 without a theoretical framework defining what 3 means. What we can achieve is a consensus that allows us to have the same metric scale when evaluating sociability, and for that computational tools can help us to include different dimensions that satisfy the different approaches.

Despite the problematics highlighted above, the author explains that dynamic systems provide new opportunities to analyze the dynamics of social phenomena. Moretti explains the dynamic feedback is one of the most useful features of a dynamic system to understand changes in social life. One of the examples cited by Moretti to emphasize her point is the work of Reynolds in 1994 about cultural transmission. Moretti explains that Reynolds applies evolutionary theory to study the emergence of cooperation in a population of Peruvians herders. Reynolds uses genetic algorithms to see how the process of cooperation derived from different rules of crossover and mutation.

In conclusion, while Morelli highlights the problems and reasons why computational simulations have not been widely implemented across all fields in social sciences, it is clear that by using this method social research can analyze different social phenomenon in a way that was not possible before. The problematics presented in the paper are not minor. However, that does not mean that it is not a useful tool that can be used understanding the limitations it has. An interesting research question that can be addressed using this simulation method is, for example, the potential consequences of killing Jamal Khashoggi in the international relations between Saudi Arabia and the United States. The general underlying question is how "illegal" acts done by a member of a state can affect the international relations between different states when large amounts of capital investment are involved. This question could be addressed as the simulation done by Guetzkow (1959), and its results can help us to contribute to the theory of development in international relationships while giving us some guidelines on how members o the state should be trained to address these situations.

References Guetzkow, H. (1959). A use of simulation in the study of inter-nation relations. Behavioral Science, 4(3), 183-191. Moretti, S. (2002). Computer simulation in sociology: What contribution?. Social Science Computer Review, 20(1), 43-57. Reynolds, R. G. (1994). Learning to co-operate using cultural algorithms. In J. Doran&G. N. Gilbert (Eds.), Simulating societies: The computer simulation of social phenomena (pp. 223-244). London: UCL Press. Taber, C. S., Timpone, R. J., & Timpone, R. J. (1996). Computational modeling (Vol. 113). Sage.

Question 2

Simulating your income (7 points) Assume that all of you will graduate from MACSS program at the University of Chicago in June 2020. Your annual income from the time you graduate to the end of your life is generated by the following process,

$$ln(inc_{2020}) = \ln(inc_0) + \ln(\epsilon_{2020})$$

$$ln(inc_t) = (1 - \rho)[\ln(inc_0) + g(t - 2020)] + \rho \ln(inc_{t-1}) + ln(\epsilon_t)$$

for
$$2021 <= t <= 2059$$

where the variable $inco_t$ is your annual income in year t>=2020, $inco_0$ is the average starting income ($t = 2018$) for a MACSS student, ρ???$[0, 1)$ reflects some positive dependence of today's income on last period's income, g is a long-run annual growth rate for your annual salary, and ϵ_t is an error term that is distributed lognormal $LN(0, \sigma)$ where σ is the standard deviation of the log of the error term. That is, $ln(\epsilon_t)$ is distributed normal $N(0, \sigma)$.

a)(3 points) Let the standard deviation of your income process be $\sigma = 0.13$, let the persistence be $\rho = 0.4$, let the long-run growth rate of income be $g = 0.025$, and let average initial income be $inc_0 = 80,000$. Assume you will work for 40 years after you graduate (2020 to 2059). Simulate 10,000 different realizations of your lifetime income. Do this by first drawing 10,000 sets of 40 normally distributed errors with mean 0 and standard deviation $\sigma = 0.13$. Then plug those into the income process (1) to simulate your lifetime income. Plot one of the lifetime income paths. Make sure your axes are correctly labeled and your plot has a title.

```
## -- Attaching packages -----------------------------------------------------------------

## v ggplot2 3.0.0      v purrr   0.2.5
## v tibble  1.4.2      v dplyr   0.7.7
## v tidyr   0.8.1      v stringr 1.3.1
## v readr   1.1.1      v forcats 0.3.0

## -- Conflicts -------------------------------------------------------------- t
## x dplyr::arrange()   masks plyr::arrange()
## x purrr::compact()   masks plyr::compact()
## x dplyr::count()     masks plyr::count()
## x tidyr::extract()   masks magrittr::extract()
## x dplyr::failwith()  masks plyr::failwith()
## x dplyr::filter()    masks stats::filter()
## x dplyr::id()        masks plyr::id()
## x dplyr::lag()       masks stats::lag()
## x dplyr::mutate()    masks plyr::mutate()
## x dplyr::rename()    masks plyr::rename()
## x purrr::set_names() masks magrittr::set_names()
## x dplyr::summarise() masks plyr::summarise()
## x dplyr::summarize() masks plyr::summarize()

##   id year      inc
## 1  1 2020 91013.33
## 2  1 2021 72883.03
## 3  1 2022 58793.88
## 4  1 2023 71664.11
## 5  1 2024 80968.35
## 6  1 2025 72087.85
```

Simulated income increase over time (one simulation)
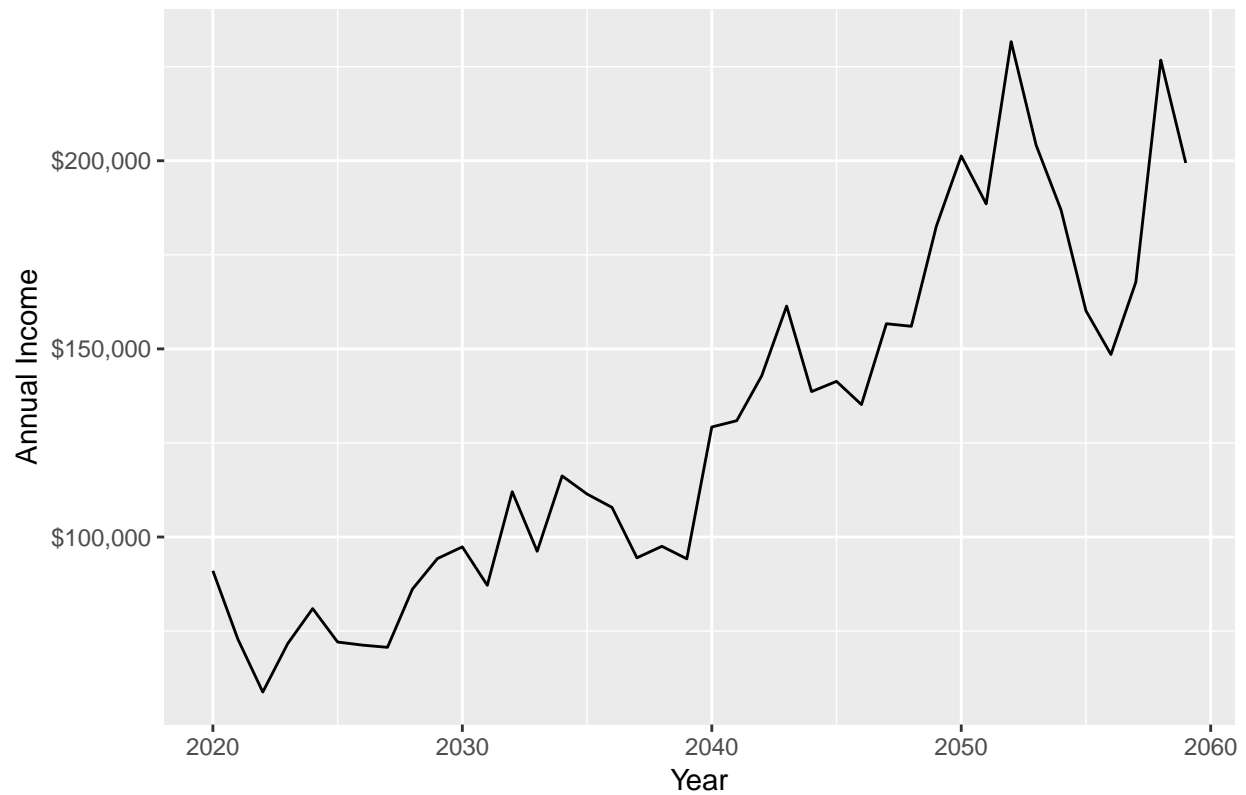

b)**(1 points)** Plot a histogram with 50 bins of year t=2020 initial income for each of the 10,000 simulations. What percent of your class will earn more than 100,000 in the first year out of the program? What percent of the class will earn less than 70,000? Is the distribution normally distributed (i.e., symmetric and bell curved)?

```
##          max_inc   min_inc
## 2020 129691.2   47020.37
## 2021 147542.5   50645.34
## 2022 138547.3   49698.88
## 2023 152333.1   53831.36
## 2024 149055.9   51503.23
## 2025 154666.7   49227.17
## 2026 157628.1   54618.25
## 2027 170235.6   56467.03
## 2028 163300.0   56366.66
## 2029 179523.7   58206.00
## 2030 176661.7   57604.54
## 2031 181598.1   62160.98
## 2032 174620.4   65048.39
## 2033 207697.6   64240.89
## 2034 203456.6   69396.72
## 2035 247746.3   64907.49
## 2036 211659.0   72933.22
## 2037 212954.9   71950.28
## 2038 226160.4   79386.42
## 2039 224000.6   74818.02
## 2040 224954.1   80845.97
## 2041 232070.6   78965.96
```

```
## 2042 245117.2  85317.26
## 2043 250299.5  82537.05
## 2044 279121.0  86513.51
## 2045 288251.8  90894.13
## 2046 298078.8  94868.71
## 2047 296989.7  91193.89
## 2048 279952.5  94842.28
## 2049 284256.8  99321.36
## 2050 308052.6  94915.87
## 2051 310391.3  97659.43
## 2052 322249.4 107657.79
## 2053 313259.6 101863.98
## 2054 318497.9 105186.19
## 2055 349441.5 110921.89
## 2056 319608.5 106033.90
## 2057 338644.2 115108.00
## 2058 349196.4 119663.50
## 2059 401333.0 124834.65

##      perct_below_70000 perct_above_100000
## [1,]             0.207             0.0354
```

Income 50 breaks 2020

income
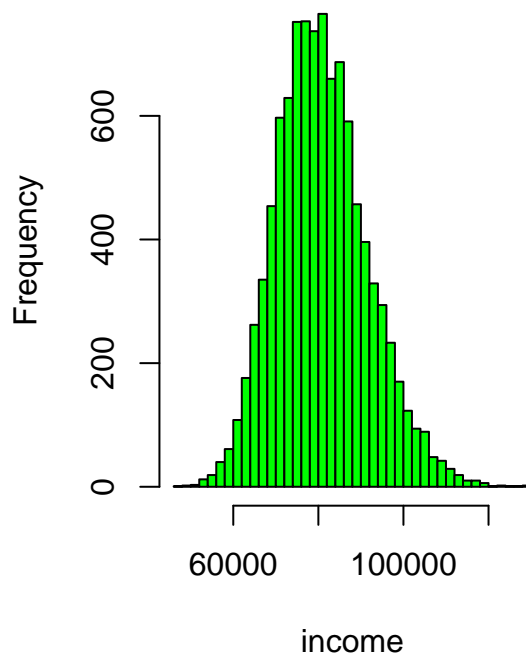
From the histrogram we can observe that the distribution follows a symmetric an curve bell, which means it follows a normal distribution.

c)(**2 points**)Suppose you graduate from the MACSS program with \$95,000 of zero-interest debt. You will use 10% of your annual salary after you graduate to pay of this loan. Plot the histogram of how many years it takes to pay of the loan in each of your 10,000 simulations. This histogram will only have as many bins as

you have unique years in which people pay of their debt. In what percent of the simulations are you able to pay of the loan in 10 years (on or before t = 2029)?

```
##
## Attaching package: 'data.table'

## The following objects are masked from 'package:dplyr':
##
##     between, first, last

## The following object is masked from 'package:purrr':
##
##     transpose

##      payment_prior_2029 payment_post_2029
## [1,]             0.2667            0.7333
```

Year in which Debt is pay
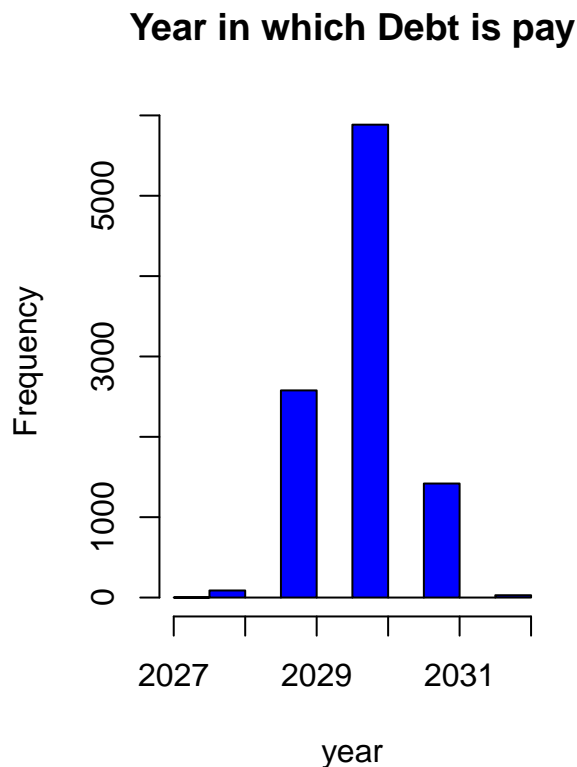
(d) **(1 points)** Now suppose that the Chicago MACSS program becomes very well known in the next two years, and the skills you are learning are demanded more by employers. This increases the average starting salary to inc0 = 90000,but the standard deviation in incomes increases also to 0,17. Plot the new histogram of how many years it takes to pay of your loan of 95,000 in your new 10,000 simulations with the new standard deviation and the new average initial salary. In what percent of the simulations are you able to pay off the loan in 10 years (on or before t = 2029)?

```
##   id year        inc
## 1  1 2020 105029.91
## 2  1 2021 128675.56
## 3  1 2022 129169.92
## 4  1 2023 111480.31
```

```
## 5  1 2024 106818.04
## 6  1 2025  93052.12

##       payment_prior_20291 payment_post_20291
## [1,]               0.8343             0.1657
```

Year in which Debt is pay