MACS30000 Assignment 3
Dr. Richard Evans
Submitted by: Junho Choi
Due October 24, 2018 (11:30 AM)

Problem 1

Moretti (2002) highlights the potential for simulation methods to be applied to questions in the social sciences. The author elaborates that, often, major questions in social sciences have to do with non-linear, dynamic systems. These are often very complex for mathematical methods to pin down, and computer simulations are described as a better alternative to such methods. In addition, different methods, such as system dynamics, multi-agent systems, and cellular automata, are showcased by the author.

While praising the new frontier of research that could be opened by active use of such simulation methods. the author also exhibits a moderate amount of criticism towards the possible hindrances that the methods could have. The issue at hand with methods is the one of "(external) validity"; that is, one may ask the question of how extendable the results and processes of simulation are to the real world. For instance, consider multi-agent systems; the approach taken by the method is one that uses autonomous individual, micro-level behaviors that may exhibit macro-level "emergent properties" that are not explainable simply by the said specified behaviors. However, there still lies the issue of how the said behaviors are determined. As the author describes, "each verified behavior is a mere deduction of the theory from which the model derives" (Moretti 2002, p. 54). Therefore, if such behaviors are misspecified by some theory or model, validity of the entire set of simulations will be at concern.

Another problem related to validity could be identified by examining the method of cellular automata. The author describes that the method of cellular automata is quite akin to that of multi-agent systems (Moretti 2002, p. 47). Therefore, one may also argue that cellular automata bears the problem of misspecifying the underlying behaviors, leading to questionable validity. In addition to this, however, cellular automata have an additional internal structure that could be problematic in representing the real-world phenomena – that updates (to each cell) are made synchronously. As the author points out, this is a very strong assumption as individuals (represented by "cells" in cellular automata) are very likely to make amends to one's behaviors and opinions at different moments (Moretti 2002, p. 47).

This is not to argue, however, that all approaches using simulation methods are invalid or prone to misrepresent the real world. Exactly because of the potential drawbacks that simulation methods can have, researchers do not just rely on a single simulation. Instead, by employing tactics such as changing the underlying parameters or making adjustments to the structures they attempt to minimize the possibility of causing issues with validity. For both multi-agent systems and cellular automata, different models and theories could be used for tuning the underlying behaviors. For cellular automata's problem of synchronous updating, the author also mentions that one can attempt to avoid this by selectively choosing a part of units to be changed (Moretti 2002, p. 47).

In addition, simulation methods can be extremely useful when "dynamic feedback" is a concern for the specifications to a theory or a model. In order to promote this idea, I provide an example inspired by a series of political situations in South Korea.[1] Consider a case in which political support for a certain party diminishes, and the portion of the said party creates a new party to hopefully attract a fraction of old supporters as well as some new ones. The strategies taken by this new party (and counter-strategies taken by existing parties) will then affect the levels of political support for all existing parties. But this may now urge some other new faction to emerge from the old to start a new party. Therefore, in this hypothetical situation, there is a dynamic feedback between the strategies of political parties (and sects within them) and the levels of support. This relationship would be quite difficult to represent using mathematical formulations, and it may be much efficient and tractable to use computer simulation.

The usefulness of simulation in dynamic feedback is also highlighted by the author, where an example from the pioneer of system dynamics is given. While explaining how the method of system dynamics came to be, the author gives an example by Forrester (1971) in which the global economic growth affects the environment, and how this in turn affects the global economic growth as well (Moretti 2002, p.44).

In summary, while the various computer simulation methods may not be without their drawbacks, they can be extremely useful tools for the social sciences. This would be even more valid in the modern society where there is greater interaction between systems and, therefore, an even greater amount of information to process.

Problem 2

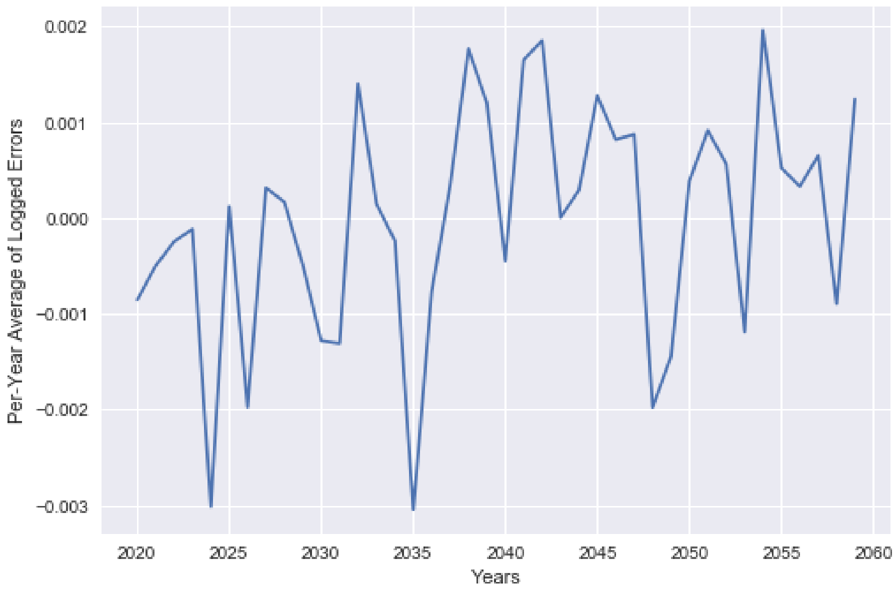
For the coding part of the problem, I refer to `A3_junhoc.ipynb` (and its `.pdf` version, `A3_junhoc_ipynb.pdf`) for further details. If coding appears in any part of the sub-problems, I will write `[CODE]` to notify this. For the reference of a specific sub-problem, see under the heading of respective sub-problem in the `.ipynb` file.

(a) `[CODE]` I first begin by estimating the 40-years-worth of logged errors (i.e. $\ln(\epsilon_{2020})$ to $\ln(\epsilon_{2059})$) which are said to be distributed normally with mean 0 and standard deviation 0.13. This estimation is done 10,000 times as given in the problem. To

[1] To briefly illustrate what happened, dissatisfaction towards the Democratic Party of Korea (DP) led to a sect of the said party forming a new political party, called People's Party (PP). PP's strategy was rather effective that it quickly became one of the major political players in South Korea, earning many seats at the Korean Congress. Almost mimicking this, a sect of the conservative Saenuri Party (SP) broke up and started a new political party (Bareun Party, BP) in the hopes of gaining popularity as well. BP's strategy was not as successful as that of PP; and later, as both parties' popularity waned in comparison to those of DP and SP, the two merged and created a new party.

ensure that the logged errors are well-behaving, I have plotted the per-year average of the logged errors against the years 2020 to 2059; this is shown in Figure 1.

Figure 1: Per-Year Averages of Logged Errors Plotted Against Years



The above figure shows that the range of averaged (logged) errors are within the range of $[-0.004, 0.004]$, which is much smaller than the previously-mentioned standard deviation of 0.13. I will use these (logged) errors throughout the sub-problems.

Codes to simulate the income process (according to the steps mentioned in the problem) are further described in the `.ipynb` files as well. I report the plot of one randomly-chosen income process (plotted against years) in Figure 2. While there certainly are many ups-and-downs in the said figure, the simulated income increases along the years (of one's lifetime) as predicted by the question.

(b) `[CODE]` **Is the distribution normally distributed?** While one may be inclined to say that the histogram for the 10,000 simulated 2020 incomes (shown in Figure 3's Panel A) is normal enough, as the errors are log-normal we expect the distribution of simulated 2020 incomes to be log-normal as well. This we can confirm by actually plotting the logged income distribution (shown in Figure 3's Panel B), which exhibits normality more so than the actual income distribution. Therefore, I conclude that the distribution is not normally distributed; rather, it is slightly skewed to the right.

What percent of the class will earn more than \$100,000 in the first year out of the program? Less than \$70,000? According to the simulation that I have conducted (which can be seen in the `.ipynb` file), one can see that only **4.0%** of the class is expected to earn more than \$100,000 in the first year. Those earning less than \$70,000 in the first year, according to my simulation, is approximately **14.82%**.

Figure 2: Simulated Income (Example)

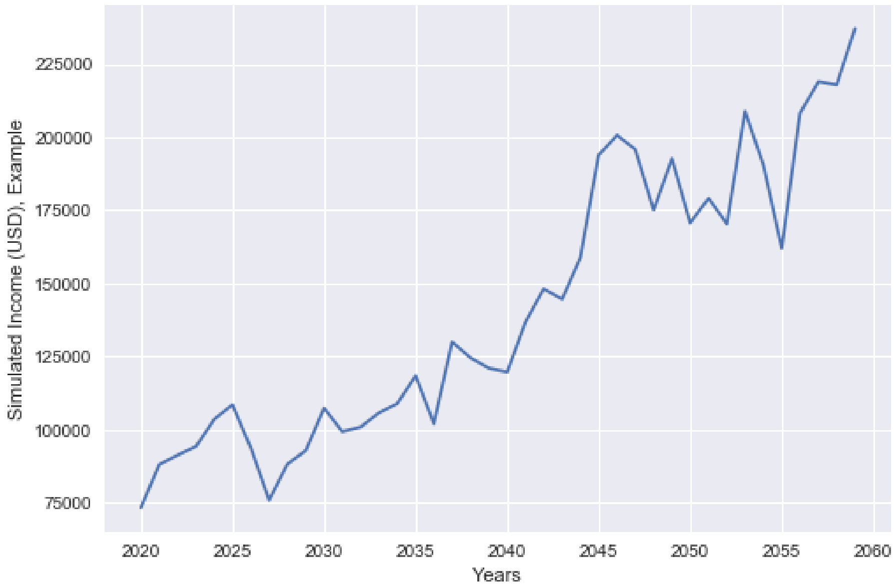
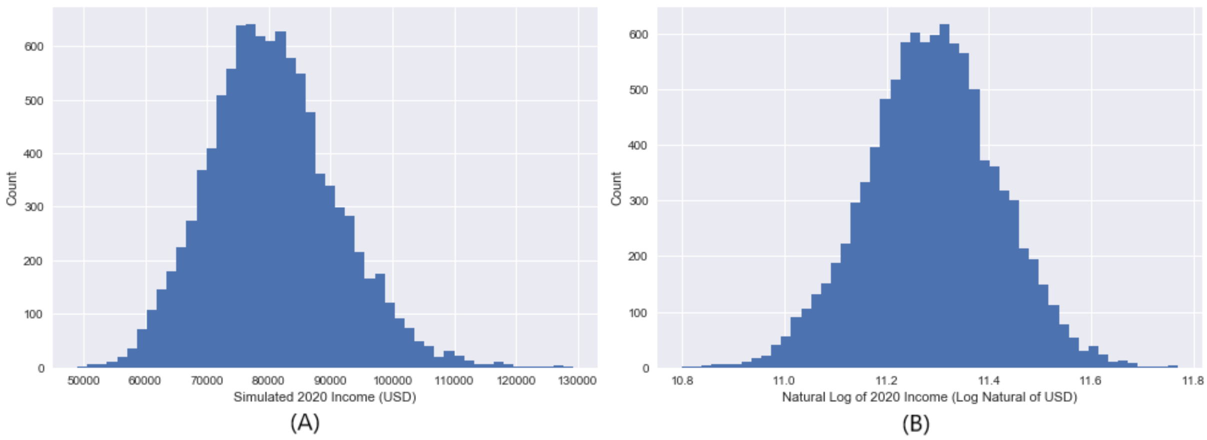


Figure 3: Simulated 2020 Incomes (A) and Logged 2020 Incomes Distributions (B)



(c) `[CODE]` **Plotting the histogram for years in debt repayment.** In the simulations that I have conducted with $\rho = 0.4$, $g = 0.025$, initial income $inc_0 = 80000$ (in USD), and $\sigma = 0.13$, the years to pay off (in full) one's debt of $\$95,000$ are distributed as in Figure 4. The least amount of time taken in full debt repayment is 9 years, whereas the most time taken is 13 years.

What percent of the simulations pay off the loan in 10 years? According to the simulations that I have conducted, approximately **16.08%** of the simulations take 10 years or less at repaying their debt (of $95,000).

(d) `[CODE]` **Plotting the histogram for years in debt repayment, with the new simulations.** In this new set of simulations, the values for ρ and g have not changed, but the initial income is now $inc_0 = 90000$ (in USD) and standard deviation of error is $\sigma = 0.17$. The histogram for years in debt repayment, then, can be shown as in

Figure 5. One can observe that now, the least amount of time taken in full debt repayment is 7 years, whereas the most time taken is 13 years (which is the same as before).

What percent of the simulations pay off the loan in 10 years? According to this new set of simulations, approximately **76.08%** of the simulations take 10 years or less at repaying their debt (of $95,000). This is approximately 4.7 times that of the previous set of simulations.

Figure 4: Years to Debt Repayment, First Set of Simulations (Problem 2-(c))

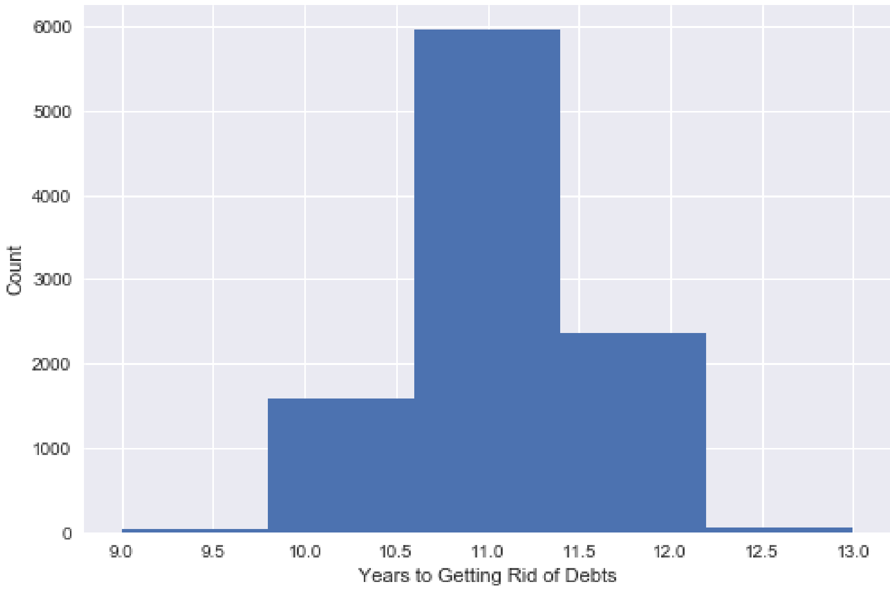


Figure 5: Years to Debt Repayment, Second Set of Simulations (Problem 2-(d))