SMU Data Science Review

Volume 6
Number 2 *Summer 2022*

Article 5

2022

Reinforcement Learning for Predicting the US GDP Output Gap

Paul Swenson
Southern Methodist University, paulswenson2@gmail.com

Anish Patel
Southern Methodist University, anish@smu.edu

David Stroud
Southern Methodist University, david@davidstroud.me

Jules Stacy
jules.stacy07@gmail.com

Reinforcement Learning for Predicting the US GDP Output Gap

Paul R Swenson [1], Anish Patel [1], David Stroud , Jules Stacy
[1]Master of Science in Data Science, Southern Methodist University,
Dallas, TX 75275 USA

paulswenson2@gmail.com
anish@smu.edu
david@davidstroud.me
jules.stacy07@gmail.com

Abstract. If a bank can successfully predict the economic fundamentals of the country they operate in, they have a significant advantage when determining interest rate risk against their competitors. This in turn is advantageous when determining investment risk against their competitors. This paper explores using Reinforcement Learning (RL) as a method for predicting The United States' Gross Domestic Product (GDP) on a quarterly basis. Various RL algorithms are compared based on how accurately they predict the GDP output gap for the following quarter. This research was unable to accurately predict the GDP output gap on a quarterly basis, but further research could be done by including additional features and reward functions. Limitations for these findings are likely due to the small quantity of data on economics in the United States.

1 Introduction

Gross Domestic Product (GDP) is one of the most widely used statistics in economics when discussing how well a country's economy is behaving. GDP is calculated by getting the sum of the value of all goods produced by a country. It is used in a variety of ways from influencing political policy, to helping banks and investors determine their practices for the coming months. One such way a bank would choose interest rates is by estimating what the GDP will look like for the upcoming quarter. If a bank can successfully predict the economic position of the country they operate in in advance, they have a significant advantage when determining investment risk against their competitors.

The economic position of a country can be estimated by a metric called the output gap. The output gap uses two values: real GDP (what a country produces), and potential GDP (what a country could produce if it utilized all its available resources). The output gap is derived by taking the difference between real GDP and potential

GDP. If the output gap of a country is positive, that means that the country is producing more than it is estimated to be able to produce. If the output gap is negative, that means that the country is producing less than it is estimated to be able to produce. Both scenarios have significant implications on many facets of economic behavior, too numerous to list in this paper, from unemployment rates to cost of goods.

Not all methods for determining the output gap are substantially complex. Some methods can be as simple as linear regression to predict a continuous, linear trend in its behavior. GDP may not follow a linear process as the economy has ups and downs, so there are some limitations to using linear regression. Time series models are an obvious choice for this problem as they can determine seasonal behavior as well as linear trends. ARIMA models have been used to some success as well as more complex time series models which utilize neural networks in hopes to uncover more complex trends in the data. One more interesting development in the world of artificial intelligence (AI) is the concept of gamification, or the process of turning a data problem into a game in which the models attempt to achieve the best score in a closed environment.

Our research will utilize a subset of AI called reinforcement learning (RL) to understand the current economic position of the United States based on data collected by the World Bank Group. RL utilizes the idea of gamification to turn the process of predicting the output gap into a problem of sequential steps with the objective of minimizing the prediction error. All that to say, our model will take the current economic position and the variables associated with it and attempt to converge on the optimal decision process when predicting the output gap.

There have been many advancements in the world of AI agents trained to solve games over the last decade. Chess was one of the first major board games to be solved in 1996 by the agent (AI game player) Deep Blue. Deep Blue utilized two key methods for its success: alpha-beta minimax search coupled with its heuristic static evaluation function [1]. Other games have been solved since; Go and Starcraft II are the most notable recent examples by Deep Mind, both of which used Reinforcement Learning (RL) [2][3]. RL has also been utilized in industrial engineering to solve cooling problems in server farms, investigate optimal actor behavior in simulated economic spaces, and even to build investment portfolios. These methods can also be utilized in similar ways to determine the economic position of the United States. This paper will compare the efficacy of various RL algorithms in the creation of agents which most accurately predict the output gap.

2 Literature Review

2.1 Gamification

The concept of gamification applies elements of games into various applications. One key element of games is the reward element. In the case of video games, a reward could be the satisfaction of mastering a skill or advancing to the next level in the game.

In a board game like chess, the reward can be the pieces captured from an opponent, and ultimately winning a game. When considering games, it is equally important to consider the pain or displeasure that the player feels from actions that result in negative consequences. For example, in a video game a player will feel displeasure from an action which causes them to lose. There is still value in these painful actions, however, since they allow the player to learn what actions not to take. Reinforcement learning agents use these elements of gamification by creating a game-like environment out of all different kinds of problems and training against those environments.

One common area gamification utilized in non-game spaces is in the Marketing domain. Advertisements (ads) in the past had a hard time understanding how successful their ads were in reaching customers and generating new business. Recent developments in software and AI have changed the way advertising works with the addition of tracking user interactions with advertisements. By recording these interactions, researchers can turn users' ad experiences into a game in which a user clicking on an advertisement indicates a win, and a user not clicking on an advertisement indicates a loss. Any data problem able to be reframed in this way has the potential to be optimized using RL.

2.2 Background for Reinforcement Learning

RL can be described as a framework for predictive models to learn how to interact with an environment through experience. RL starts by defining a scheme consisting of an agent which learns and makes decisions and an environment containing everything that the agent has knowledge of [Fig 1]. For this research, the environment will be the economic data and the agent will be the model predicting the output gap. RL agents learn to make predictions by interacting with the environment through various actions. Each of these actions changes the state of the environment and yields rewards that the agent will try to maximize [4].

Figure 1: Reinforcement Learning Scheme

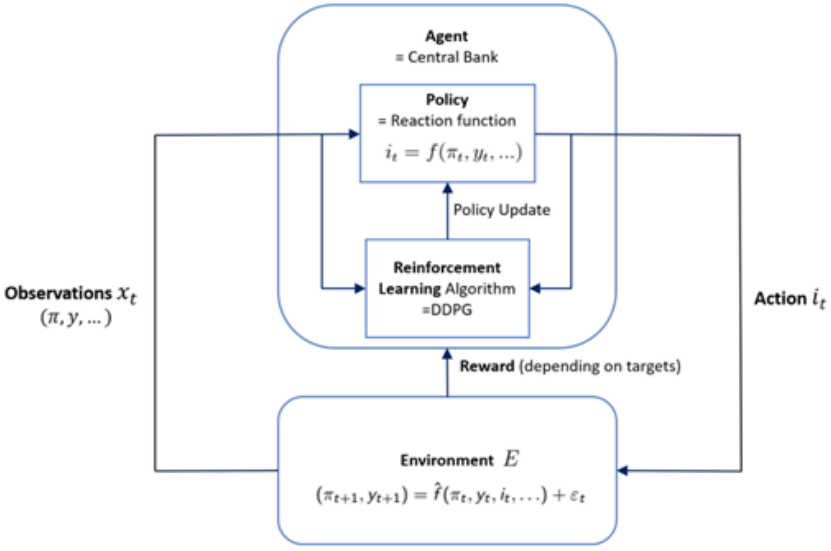

Figure 1: Reinforcement Learning Scheme[5].

One real world analogy to RL can be described through training a dog to sit in a room. The agent is the dog, the environment is the room, and the reward is a dog treat. If the dog takes the action to sit, it gets a treat or reward, which is a form of positive feedback. If the dog takes another action like jumping or running, it will get punishment in the form of negative verbal feedback or by withholding a treat. After many training sessions, the dog will eventually learn which actions to take to maximize the amount of dog treats that it receives.

When an agent takes actions to interact with the environment, it enters a process called the Markov Decision Process (MDP). The agent will enter the MDP at the current state of the environment S_0 and pick an action a to transition to the corresponding state S_n. Some MDPs also include transition probabilities, or the likelihood of changing from one state to another. The total representation of all states and actions available to the agent can be represented by a directed graph such as the example below. States are represented in green, actions are represented in orange, and the state transition probabilities are represented by the values along the edges between each node.

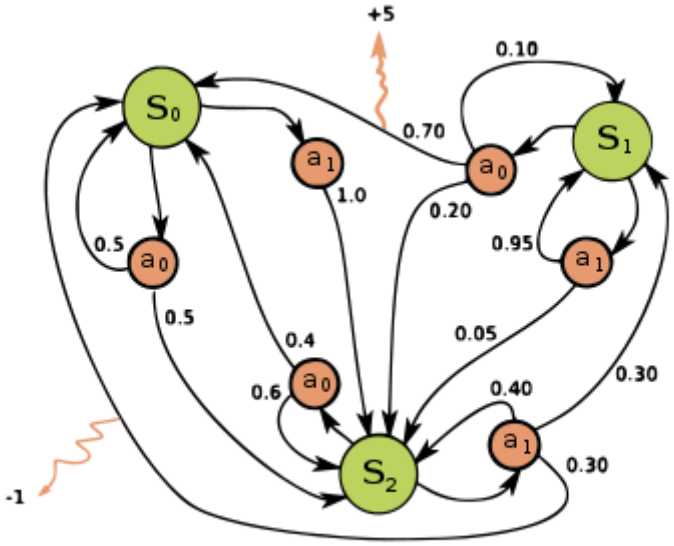

Figure 2: Simple MDP Representation [6]

MDPs use one key assumption called the Markov property which defines all previous states within the process as independent events. When each previous state is independent, this makes the agent effectively memoryless, so the past states have no influence on the present state. RL utilizes the MDP by estimating the transition probabilities when taking an action and choosing the optimal path to maximize the reward returned by the agent. This reward optimization estimation is derived from the Bellman Equation.

$$V(s) = \max_a \left(R(s,a) + \gamma \sum_{s'} P(s,a,s')V(s') \right)$$

Figure 3: The Bellman Equation

The Bellman Equation is used to rate agent decisions, or policies, and the rewards returned by those policies. The objective when training agents is to adjust the policy and value functions such that the reward function is maximized, enabling the agent to learn in an optimal manner. Reward functions are integral to RL, such that the goal is to maximize the reward signal in the RL model. Actions that are positive and move the agent towards the goal are given a positive reward value, while actions that that are detrimental toward the goal are given a negative reward value. MDPs involve

actions that will have immediate rewards, as well as ones that may yield rewards in the future [4].

2.3 Artificial Intelligence Bots in Games

Not all AI in games utilize RL to solve them. Tik Tak Toe is one instance of a mathematically solved game as demonstrated by Peter Baum [7]. When a game has a mathematically sound solution, a bot can be programmed to follow the algorithm outlined in the solution. Other games have also been mathematically solved; Monopoly uses Markov Chains to estimate the expected return landing on any square [8]. Some AI is designed to be intentionally bad; these kinds of AI are useful in games where players may not be able to perform at the highest level. Bad AI can be designed to intentionally play un-optimal moves as a way of teaching new or inexperienced players how to play the game. Bad AI is commonly seen in phone applications for games like chess, where there are different skill levels of AI to compete against.

What separates RL from ML (Machine Learning) is that there is no training data on which the model is tuned. In the RL environment the agent is an intelligent program which works to continuously learn based on rewards and punishments for given actions. [9]. One other feature that distinguishes RL from ML is that RL is less prone to overfitting when compared to ML algorithms. Each of these factors may be good to take into consideration when selecting a predictive model algorithm.

Backgammon is one of the early games modeled by the Markovian Decision Process (MDP) specifically using a method called temporal difference (TD) learning. TD-Gammon, created by Gerald Tesauro from IBM in the late 1980s, was almost able to beat the Backgammon World Champion using TD learning [10]. Unsupervised learning is utilized in Temporal difference. TD learning aims to accurately predict a target variable. Reinforcement learning (RL) builds on this by learning from the current state and the subsequent actions which in turn changes the state of the environment [11].

2.5 Reinforcement Learning in Games

Kasparov, Carlsen, Lee Sedol, and Go Seigen are players whose names would be recognized by the communities built around their games of choice, however, not even these players can defeat the modern agents developed within the last few decades. The agent Deep Blue defeated chess grand master (highest achievable rank in chess) Gary Kasparov in a 5-game match in February 1996 [12]. Since then, AI like Deep Blue has been defeated many times by more modern AI agents such as Stockfish and RL-trained AlphaZero. More recently, in March 2016, the agent AlphaGo defeated 9 Dan Go professional Lee Sedol 4-1 [13].

Even though the games mentioned above have been solved, there are still many open problems in the RL space as applied to games. Research has been done to solve Rocket League, a continuous space game in which teams of cars with rocket boosters play a variation of soccer, with limited success [14]. Games with continuous action spaces are notoriously difficult to optimize since the action space for any given state is unlimited. This property essentially renders these games unsolvable from a mathematical standpoint. Other unsolved games often include multi-agent operations

(teams of agents that must work together), or imperfect information as mentioned before

2.6 Reinforcement Learning in Economic and Finance Research

Reinforcement learning framework can be applied in economics. As a basic example, how will a customer's actions change when the price of an item is increased? The change in price is equivalent to changing the environment, influencing the actions of the consumer. This type of sequential decision making is well suited for RL models [15].

Salesforce has developed "AI Economist" built on a RL framework with the goal of optimizing tax policy in a simulated economy. It has been found that the framework was able to develop a policy to such that productivity was maximized while minimizing income inequality. The results showed that it was able to develop policies that were more fair than academic economists were able to develop. Interestingly, the model found that higher taxes on the rich as well as the poor, and lower taxes on the middle class led to a smaller gap between the rich and the poor [16]. The results show the promise of RL in economics as the recommended policy yields better results than traditional theory.

This paper builds from the research done by Alina Tanzer and Natascha Hinterlang in the "Optimal monetary policy using reinforcement learning" research paper. It investigates the optimal policy function of the German central bank respective of inflation and output gap targets. This field of research is relatively unexplored with regards RL.

2.7 Algorithms Used in this Research

The two RL algorithms being compared in this paper are the Twin-Delayed Deep Deterministic Policy Gradient (TD3) and Soft Actor-Critic (SAC).

TD3 is an off-policy algorithm, used in a continuous domain. TD3 fundamentally uses Double-Q learning, delayed policy updates, and target policy smoothing. In Double-Q learning two Q functions are used. The Bellman error loss uses the lesser of the two Q values to form the target when the policy is updated. In the TD3 algorithm, target policy smoothing refers to adding distortion to the target action. The goal in doing this is to smooth out Q with changes in action, in order to prevent the policy from exploiting Q function errors when there are incorrect sharp peaks. The policy is learned by maximizing Q [17].

$$\max_{\theta} \mathrm{E}_{s\sim\mathcal{D}}\left[Q_{\phi_1}(s,\mu_\theta(s))\right]$$

Figure 4: TD3 Policy [17]

The SAC algorithm maximizes entropy as well as reward to determine the optimal policy. The algorithm randomly selects actions based on estimates of standard deviation and mean of a Gaussian probability distribution [18]. Similar to the TD3

algorithm, SAC learns two Q-functions. The difference being while TD3 only uses lesser of the two Q values in the policy, SAC uses the minimum of the two Q approximations [18]. The optimal policy calculation is shown below in Figure 5.

$$\max_{\theta} \mathop{\mathrm{E}}_{\substack{s\sim\mathcal{D}\\ \xi\sim\mathcal{N}}} \left[\min_{j=1,2} Q_{\phi_j}(s, \tilde{a}_\theta(s,\xi)) - \alpha \log \pi_\theta(\tilde{a}_\theta(s,\xi)|s) \right]$$

Figure 5: SAC Policy [18]

2.8 Hypothesis for this research

Economic research which utilizes RL methodologies is somewhat limited. Thus far, predicting the output gap has been difficult, and with the small amount of data utilized for this experiment there will likely be some unsuccessful models. This research expects to develop at least one model which successfully predicts whether the output gap will be positive or negative in the upcoming quarter.

3 Methods

3.1 Data

The data used in this research is gathered from the Federal Reserve Economic Data (FRED) made publicly available. The output gap is calculated as the difference between current GDP and potential GDP, which was obtained through FRED. GDP and potential GDP is normalized to dollar values from 2012.

- CPIAUCSL – Consumer Price Index for all Urban Consumers: All Items in U.S. City Average. This data from FRED measures inflation compared to the previous year for [19].

- INDPRO – Industrial Production: Total Index. This data set from the Federal Reserve Economic Data source measures the output for gas, electric, mining and manufacturing facilities located in the US [20].

- MABMM301USM189S – M3 for the United States. This data set from the Federal Reserve Economic Data source is a measure of the total money supply from January of 1960 to June of 2022. M3 is an aggregate of M1, M2, and liquid assets which include bank deposits, money markets and repurchase agreements [21].

- MORTGAGE30US – 30 Year Fixed Rate Mortgage Average in the United States. This data set from the Federal Reserve Economic Data source is provided

by Freddie Mac company and represents the interest rates for a 30-year fixed mortgage. The data set ranges from April 1971 to June 2022 [22].

- TDSP – Household Debt Service Payments as a percent of Disposable Personal Income. This data set from the Federal Reserve Economic Data source is a ratio of the total required household debt payments to total disposable income [23].

- FPCPITOTLZGUSA – Inflation, consumer prices for the United States. This data from the Federal Reserve Economic Data source measures the change in prices of a fixed basket of goods to one year prior. It is given as a percentage and ranges from January 1960 to June 2022 [24].

- GDP – Gross Domestic Product. This measures the total US output of goods and services [25].

- GDPPOT – Real Potential Gross Domestic Product. The Congressional Budget Office's estimate of the output of the economy [26].

3.2 Reward Function

The reward function used for training the models is described by the following equation:

$$R_t = \begin{cases} 0.01(y - \hat{y}) + 100 & where\ y - \hat{y} < 0 \\ -0.01(y - \hat{y}) + 100 & where\ y - \hat{y} \geq 0 \end{cases}$$

In the case where the model is predicting the output gap correctly, the difference between these values is 0, and the maximum reward is 100. The reward function value of 100 is the best-case scenario while rewards closer to 0 occur when the predicted values are farthest from the actual value in either the positive or negative direction. The maximum possible reward is 100 times the number of steps in the data set, in the case where the agent predicts the output gap perfectly at each step. For this scenario, if the model is 100 percent accurate, the total reward will be 16800 (100*168). Visually, this reward looks like Figure 5.

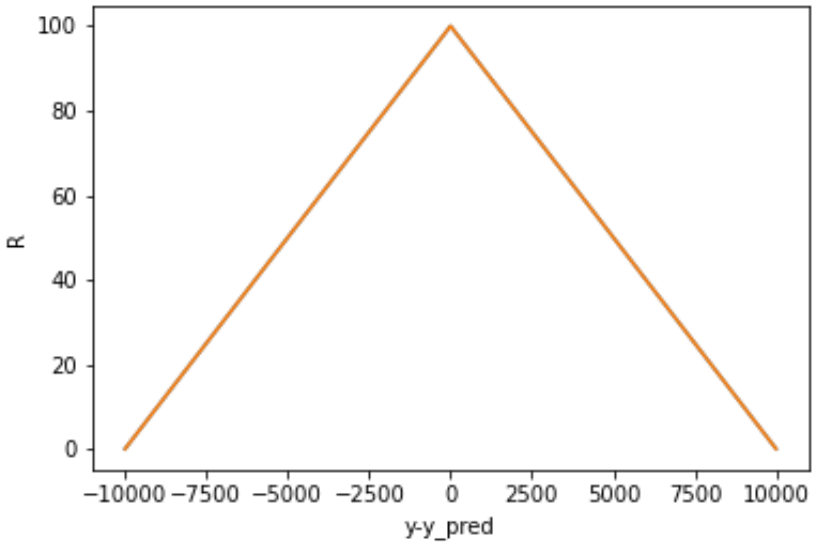

Figure 5: Reward Function

3.3 Environment

The environment is the space in which the agent can see current states and interact via actions to get a reward. Our state is represented by a vector s where each element in s represents the data described in section 3.1. Its format is as follows.

$$h = number\ of\ previous\ quarters\ to\ include$$
$$t = current\ time\ step\ of\ the\ environment$$
$$s =$$

$INDPRO_{t-h}$	$MABMM301USM189S_{t-h}$	$MORTGAGE30US_{t-h}$	$TDSP_{t-h}$	$OUTPUT\_GAP_{t-h}$
…				
$INDPRO_t$	$MABMM301USM189S_t$	$MORTGAGE30US_t$	$TDSP_t$	

The amount of history to include in the input vector has some tradeoffs. On one hand, more historic data allows the agent to infer more complex seasonal behavior. On the other hand, increasing the history available to the agent decreases the number of observations it has overall. For instance, if the model has 10 years of history in the input, it has 40 less observations (10 years * 4 quarters) to learn from.

The output vector, also known as the action space, is represented by the vector a where:

$$a = x \quad \in \quad \mathbb{R} \quad where \quad x \geq -10000 \quad and \quad x \leq 10000$$

This vector represents billions of US dollars adjusted to 2012 values. These are the predictions, or actions, which the agent can make. Since the model can only

predict within -$10000 billion and $10000 billion, if the actual output gap is outside of these boundaries, the agent would not be able to infer that amount. The benefit of reducing the action space to be bounded like this is that it increases the likelihood that the model will converge on a solution. With arbitrarily large action spaces, the training time also increases.

Reinforcement learning environments must also have state transitions. The states for this problem are as follows:

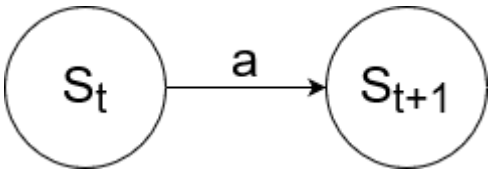

Figure 6: Environment State Transitions

In this diagram, the current state s_t as described before will transition to the state at the next time step s_{t+1} after taking some action a. Action a is the agent's prediction of the output gap in billions of US dollars within the action space described above. This state transition process repeats after each action until the available data has reached its end.

4 Results

This paper hopes to find that RL can generate effective AI agents which successfully predict the output gap. This section will have charts comparing the

performance of each algorithm based on prediction accuracy, reward system, and a summary of those results.

The models produced were overall, very unsuccessful in predicting the GDP output gap. The data visualization below is a summary of the models created using SAC and TD3 for varying lengths of history.

Model Name	Description	Mean Reward at Step 20000
SAC_20_MlpPolicy	SAC model with 20 quarters of historic data in the environment state using a MlpPolicy	-387.4
SAC_40_MlpPolicy	SAC model with 40 quarters of historic data in the environment state using a MlpPolicy	-156.4
SAC_8_MlpPolicy	SAC model with 8 quarters of historic data in the environment state using a MlpPolicy	485.5
TD3_20_MlpPolicy	TD3 model with 20 quarters of historic data in the environment state using a MlpPolicy	387.4
TD3_40_MlpPolicy	TD3 model with 40 quarters of historic data in the environment state using a MlpPolicy	472.8

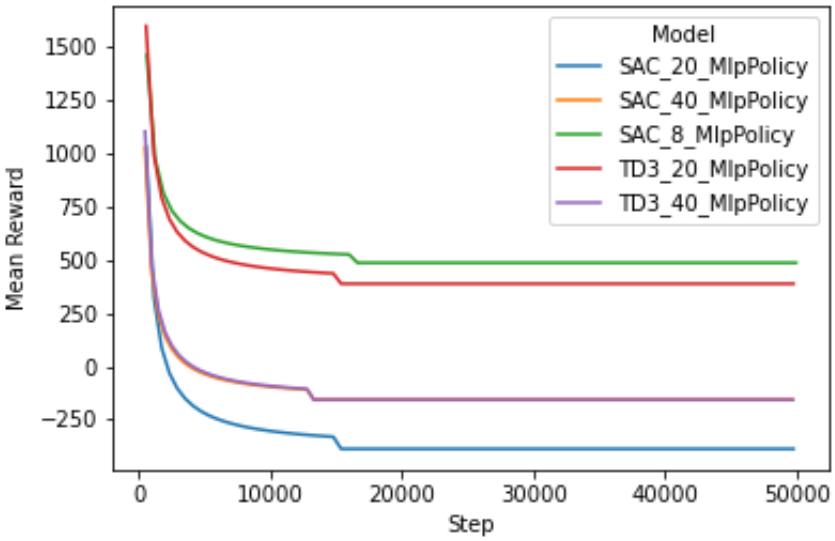

Figure 7: Policy Rewards

As seen in Figure 7, the best model (SAC_8_MlpPolicy) performed with an average reward of only 485.5 across all time steps. For perspective, the best possible mean reward would be 16000 (168 observations – 8 historic observations * 100) if the model could accurately predict the output gap at every time step. This means, on average, SAC_8_MlpPolicy gained a reward of 3.03 at each step.

To get the reward in terms of the GDP output gap, we can put the reward back into the reward equation to get an average prediction difference of around 9697 billion US dollars. For reference, the US GDP in 2022 is around 20,000 billion indicating that these predictions are very far from a practically useful threshold.

There did not appear to be significant differences in performance between the modeling techniques, the policies used, or the history length for these models.

5 Discussion

5.1 Interpretation of Results

These results are not promising in terms of predicting the GDP output gap. The results do, however, introduce a very important concept of gamification in spaces which are not games. Not every machine learning project has an impactful outcome, and many times, data scientists can find themselves using a wide variety of techniques before landing on a final solution. Reinforcement learning may not have worked on this space with this data set, but that does not mean it is not a viable strategy for many other

problems. If a data problem can be framed as an agent interacting with an environment, it could be worth the attempt to turn it into a reinforcement learning problem and enable a new variety of modeling techniques to experiment with.

5.2 Limitations

There are many limitations in this research when applying RL to economic research. One major limitation was the lack of data. Economic data generally comes out on a monthly, quarterly, and annual basis. Most of the data sets utilized from the US federal reserve go back to 1960 at best. The more data that is available, the more likely a reinforcement learning model will be successful in predicting correct outcomes. If economic data had been available for thousands of years, the models would have a better chance at making more accurate predictions. Currently, one of the most common use cases for RL models is with self-driving cars. In this scenario, there can be many millions of data points collected from multiple sensors collected frequently every second. For example, Amazon's DeepRacer self-driving car has a camera that collects 15 frames per second, in addition to multiple other sensors [27]. GDP and the indicator variables used in this research simply do not compare to the magnitude or frequency of data points in typical problems that are used in the RL spaces.

In addition to the lack of observations, it is suspected that the input parameters may not be providing enough information for the model to accurately predict GDP. It is possible that adding more indicator variables could yield better results.

Another large challenge when using RL is defining a good reward function. The agents use reward functions to optimize their decision making, so creating a function which incentivizes better decisions by the model is extremely important. Usually, it is best to start with a very simple reward and work towards more sophisticated ones to help tune the models. In the case of this research, even the most simplistic reward, a linear reward, was not enough.

5.3 Ethics

There are many implications to consider when RL is used to predict GDP. Since the models will be used to influence public policy and the setting of interest rates, it is important to understand that recommendations of the model will have real consequences on the economy and the livelihoods of the entire population. The consequences of incorrect predictions can very well exacerbate issues in the economy and lead to recessions and inflation. Banks will depend on these predictions to set interest rates, while institutions will use the predictions to make decisions on capital and operational investments. At the individual level, predictions can influence if one chooses to save more or consume more. The ramifications of incorrect predictions are far and wide. It is very important to the model is not put into practice until the researchers are highly confident in its ability to correctly predict GDP.

5.4 Future Research

Future research would aim to address the limitations noted above. Specifically, a more extensive data set with more observations would be sought out. This research focused on data made available through the Federal Reserve Economic data which had a limited number of observations. Other sources of data will be used in future research. Future research would also include more input parameters to be included in the data set. Lastly, additional algorithms and further tuning of the reward function should be included in future research.

This research focuses solely on the output gap of US quarterly GDP data. There are many other countries whose output gap prediction can be explored, and many other input variables which can be utilized in this kind of research. Any future work should be done to create an RL agent which interacts with additional input fields or utilizes different algorithms.

6 Conclusion

This research should be used to help guide any future attempts at predicting the GDP output gap using reinforcement learning. There are many more things to test in this space, and so many things not considered by this research; what policies to use, new reinforcement learning algorithms to try, huge amounts of economic data to incorporate, and the list goes on. Any researchers looking to build on this should consider these factors and the attempts made by this paper to save time and energy.

Hopefully this paper also inspires others to try out these new techniques on other data problems which may not initially stand out as candidates for reinforcement learning. While these models were not successful in predicting the GDP output gap, there is still a lot of potential for other areas of economic research and many other problems to apply RL to.

Acknowledgments

Special thanks to our Faculty advisors at SMU (Southern Methodist University): Dr. Cheun and Nibrhat Lohia.

References

1. Newborn, M. (1996). Kasparov versus Deep Blue - computer chess comes of age.
2. DeepMind Technologies. *AlphaGo: The Story So Far.* https://deepmind.com/research/case-studies/alphago-the-story-so-far
3. Vinyals, O., Babuschkin, I., Czarnecki, W. M., Mathieu, M., Dudzik, A., Chung, J., Choi, D. H., Powell, R., Ewalds, T., Georgiev, P., Oh, J., Horgan, D., Kroiss, M., Danihelka, I., Huang, A., Sifre, L., Cai, T., Agapiou, J. P., Jaderberg, M. Silver, D. (2019). *Grandmaster level in StarCraft II using multi-agent reinforcement learning. Nature* (London), 575(7782), 350–354. https://doi.org/10.1038/s41586-019-1724-z

4. Sutton, R., & Barto, A. (2018). *Reinforcement Learning An Introduction* (2nd ed.). The MIT Press.
5. Hinterlang, N., Tänzer, A. (2021). Optimal Monetary Policy Using Reinforcement Learning.
6. Example of a Simple MDP [Online Image]. Wikipedia. https://en.wikipedia.org/wiki/Markov_decision_process#/media/File:Markov_Decision_Process.svg
7. Baum, Peter. (1975). *Tic-Tac-Toe*. 10.13140/RG.2.2.24939.54569.
8. Parker, Matt. [Stand-up Maths]. (2016, December 8). *The Mathematics of Monopoly* [Video]. YouTube. https://www.youtube.com/watch?v=ubQXz5RBBtU
9. Nandy, A., Biswas, M. (2018). *Reinforcement Learning With Open AI, TensorFlow and Keras Using Python*. Apress
10. Tesauro, G. (1992). Practical Issues in Temportal Difference Learning. *Machine Learning,* 8(3-4), 257-277. https://doi.org/10.1007/BF00992697
11. Stanford University. (n.d.). Temporal Difference Learning. https://web.stanford.edu/group/pdplab/pdphandbook/handbookch10.html
12. Newborn, M. (1996). Kasparov versus Deep Blue - computer chess comes of age.
13. DeepMind Technologies. *AlphaGo: The Story So Far.*
14. Ibrahim, D., Stacy, J., Stroud, D., Zhang, Y. (2020). *Rocket Learn.*
15. Charpentier, A., Elie, R., Remlinger, C. (2020). Reinforcement Learning in Economics and Finance
16. Heaven, Will. (2020). *An AI can simulate an economy millions of times to create fairer tax policy.*
17. *Twin delayed DDPG*. Retrieved June 30, 2022, from https://spinningup.openai.com/en/latest/algorithms/td3.html
18. *Soft Actor-Critic*. Retrieved June 30, 2022, from https://spinningup.openai.com/en/latest/algorithms/sac.html
19. U.S. Bureau of Labor Statistics, Consumer Price Index for All Urban Consumers: All Items in U.S. City Average [CPIAUCSL], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/CPIAUCSL, June 19, 2022.
20. Board of Governors of the Federal Reserve System (US), Industrial Production: Total Index [INDPRO], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/INDPRO, June 19, 2022.
21. Organization for Economic Co-operation and Development, M3 for the United States [MABMM301USM189S], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/MABMM301USM189S, June 19, 2022.
22. Freddie Mac, 30-Year Fixed Rate Mortgage Average in the United States [MORTGAGE30US], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/MORTGAGE30US, June 19, 2022.
23. Board of Governors of the Federal Reserve System (US), Household Debt Service Payments as a Percent of Disposable Personal Income [TDSP],

retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/TDSP, June 19, 2022.
24. World Bank, Inflation, consumer prices for the United States [FPCPITOTLZGUSA], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/FPCPITOTLZGUSA, June 19, 2022.
25. U.S. Bureau of Economic Analysis, Gross Domestic Product [GDP], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/GDP, June 19, 2022.
26. U.S. Congressional Budget Office, Real Potential Gross Domestic Product [GDPPOT], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/GDPPOT, June 19, 2022.
27. *What is AWS DeepRacer?.* Retrieved June 15, 2022, from https://docs.aws.amazon.com/deepracer/latest/developerguide/what-is-deepracer.html