

AI-Based Scenario Simulation for U.S. Debt Trajectories Under Fiscal Policy Changes**Maksuda Begum¹, Sanjib Kumar Shil², Mohammad Moinul Islam³, Mahamuda Akter Shati⁴, Md Al Mamun Siddike⁵, Md Sumsuzzaman⁶, Farmina Sharmin⁷ and Azhar Uddin⁸**¹Master of Business Administration, Trine University.²MBA in Management Information System, International American University³Ph.d in Electrical Engineering, EECS, Florida Atlantic University⁴Master's in Business Analytics, Grand Canyon University⁵MS in Business Analytics, Trine University.⁶Master's of Business Administration, International American University⁷MBA in Business Analytics, International American University, Los Angeles Main Campus⁸MBA Business Analytics, Gannon University, Erie, PA, USACorresponding Author: **Maksuda Begum, Email:** mbegum23@my.trine.edu**Abstract**

Policymakers are getting more stressed about whether public debt can actually be paid back. Budgets change; economies shift; interest rate environments move up plus down. Most people still use basic math rules to guess where debt goes. These old rules explain how debt grows based on spending plus economic expansion. Theoretical clarity is nice. These simple models often miss the weird, complicated patterns seen in real history. This research project builds an AI tool to track where sovereign debt is heading. It connects machine learning models with specific scenario tests plus risk checks. The data comes from the Federal Reserve Economic Data (FRED) system. It pulls together tax receipts, federal spending, interest rates, GDP growth, plus the debt-to-GDP ratio. The study tested several ways to predict debt one year out. It looked at Ridge Regression, XGBoost, plus a standard math benchmark based on fiscal rules. The Ridge Regression model gave the most accurate short-term guesses. It worked better than the fancy XGBoost model. It even beat the standard math benchmark in this specific test. After finding the best model, a simulation engine projected debt paths 15 years into the future. It looked at different choices plus economic surprises. These included spending more money, cutting budgets, high interest rates, plus periods of stagflation. Monte Carlo simulations helped map out the uncertainty. This provided a look at the odds of various long-term debt levels. Policy changes, plus the difference between interest rates plus growth, are the main drivers of debt over time. Past debt levels keep the predictions anchored. Mixing machine learning, scenario tests, plus explainable AI techniques helps data-driven models work alongside old economics. It offers a more flexible way to see how fiscal policy might play out.

Keywords: U.S. Fiscal Policy, Debt Sustainability, Scenario Simulation, Machine Learning, Monte Carlo, SHAP, Policy Risk Analytics**1. Introduction**

1.1 Background and Motivation : The way public debt keeps climbing has become a major headache for major economies this century. In the United States, policymakers, bank officials, and economists are staring at the sheer size of federal borrowing, trying to figure out whether the country can keep this up. Looking at the latest Congressional Budget Office (2025) projections, the numbers show that if the government continues on its current path, public debt will balloon over the next few decades [10]. We are looking at debt-to-GDP ratios that haven't really been seen in modern history. These kinds of forecasts make it pretty clear that we need better ways to assess how today's spending and tax choices will affect the country down the line. It isn't just about one thing, either. The way debt moves depends on a messy mix of tax revenue, interest rates, how fast the economy is growing, and the aging population. When people talk about why the debt is growing, they usually point to more than just a bad budget year here or there. There are deep, structural things happening. Dallas and Scavone (2024) discuss how an aging population and the skyrocketing cost of healthcare are straining the budget [11]. Then you have an interest in the debt itself. These forces make the pile of debt grow even when the economy is doing okay. The American Action Forum (2025) recently noted that interest costs alone could eventually exceed those of some of the largest government programs if things stay on this track [4]. Because of that, just looking at a single "best guess" number for the future doesn't cut it. We need tools that can play out different "what if" scenarios to see where we might actually end up. The fallout from all this borrowing goes way beyond just accounting. It hits how the whole economy works. Mahlik (2026) has reviewed a mountain of research on how debt affects growth [24]. While the data can be a bit all over the place, there is a real concern that having too much debt for too long can make the economy sluggish. It can crowd out private projects and leave the government with its hands tied when a real recession hits. This is a big deal for countries like the U.S., which spent heavily to get through the financial crisis and the pandemic. Edelberg, Harris, and Sheiner (2025) note that carrying this much debt makes the government's budget very sensitive to interest rate hikes [13]. If borrowing costs rise faster than the economy grows, the risks start to stack up quickly. Because of these worries, groups such as the International Monetary Fund (2021) have developed specific systems to assess whether a country's debt is sustainable [18]. They want to know whether a nation can pay its bills without making crazy, unrealistic cuts or ending up in a total financial mess. These systems look at growth, taxes, and "stress tests" to see how the debt holds up when things go wrong. The IMF and the World Bank (2018) originally made these kinds of checks for lower-income countries to spot trouble early [19]. But the basic logic, running different scenarios to see how debt moves under pressure, has started to make a lot of sense for the world's biggest economy too. Even with these tools, predicting the budget's future is incredibly hard. Debt doesn't move in a straight line. It's affected by weird, non-linear shifts in the markets and sudden shocks to the economy. This makes the old-school way of forecasting, where you plug in a few numbers and get one answer, feel a bit outdated. There is a lot of new interest in using heavy-duty computing to handle these complexities and simulate a wider range of possible futures. For a country like the United States, where debt is high and financial markets affect the whole world, getting these projections right is more than just an academic exercise. It is a matter of keeping the global economy stable.

1.2 Importance of This Research: Figuring out where U.S. debt is headed matters to many people. It's a big deal for the folks writing laws, but also for investors and banks who need to know the risks before they put their money anywhere. The amount the government owes touches everything from how much it costs to get a mortgage to how people feel about inflation and the safety of the financial system. As borrowing grows, checking whether the whole thing is sustainable becomes a core part of keeping the economy on the rails. The Congressional Budget Office (2025) says that the massive jump in debt relative to the size of the economy will change the country's fiscal outlook [10]. It might leave the government with no room to move if another crisis hits or if it wants to build something big. Analysts keep pointing out that this isn't just a single problem but a set of risks that feed into one another. Edelberg, Harris, and Sheiner (2025) explain that when debt is high, even a tiny nudge in interest rates can send the government's interest payments through the roof [13]. It can become a vicious cycle: you borrow more to pay interest on what you already owe, which makes the debt grow faster. The American Action Forum (2025) notes that interest is quickly becoming a larger share of the federal spending pie [4]. We need better ways to simulate how a sudden change in growth or a market shock might completely alter the debt path. Beyond the budget math, there is the question of how this debt affects everyone's standard of living. Mahlik (2026) points out that a mountain of debt can weigh on long-term economic performance [24]. It can make the economy less flexible. In the U.S., where the government is expected to step in when things get shaky, this is a real concern. As Dallas and Scavone (2024) have noted, the pressure from factors such as an older population and high healthcare costs isn't going away [11]. We have to be able to see how different policy choices today will play out over the long haul.

International groups have long known that you can't just look at a single "best case" version of the future. The International Monetary Fund (2021) argues that a real analysis needs to account for economic shocks and structural changes [18]. The IMF and the World Bank (2018) have shown that stress testing, seeing what happens if the worst-case scenario occurs, is the only way to really understand risk [19]. This kind of thinking is becoming more common because it admits that we don't know exactly what the future holds. The problem is that the usual ways of calculating debt often rely on only those that don't handle the messy reality of the world very well. That's why researchers are starting to look at advanced tech to help out. Using tools like artificial intelligence and machine learning to analyze the economy could help us produce more accurate and flexible forecasts. These tools can dig through massive amounts of data to find patterns that humans might miss and show us a much wider range of possible outcomes as the debt climbs and the future feels more uncertain. Having these tools to help make decisions is becoming a necessity.

1.3 Research Objectives: This study aims to fill a gap in how we think about the future of U.S. federal debt, specifically by developing better tools to model where that debt is headed as fiscal policies change. Most of the time, people rely on standard forecasting that assumes economic relationships stay fixed and policies don't vary much. Those methods are fine as a starting point, but they often trip up when trying to handle the messy, nonlinear ways modern economies actually work. Things like government spending, taxes, growth, inflation, and interest rates don't happen in a vacuum. They all move at the same time and pull on each other over the years. Because these pieces are so intertwined, predicting sovereign debt requires a way to account for all that multidimensional movement and the uncertainty that comes with it. The goal is to create a computational setup that integrates artificial intelligence with macroeconomic data to simulate how debt evolves under different policy choices. A big part of this work involves building predictive models to estimate how the U.S. debt-to-GDP ratio might evolve based on historical data. By training machine learning algorithms on decades of economic data, the research aims to identify patterns that simpler, linear models often miss. These models serve as the backbone for simulations in which we can plug in "what if" scenarios, such as a sudden shift in tax laws or an economic change. The study focuses on making a recursive simulation engine. This engine projects debt forward and keeps updating the economic variables and fiscal balances as it goes. It makes it possible to test a wide range of ideas, from cutting spending to dealing with interest rate spikes to seeing what happens if the economy slows unexpectedly. Another objective is to put a number on the uncertainty that always hangs over long-term fiscal projections. Any economic forecast is going to be a bit fuzzy because of random shocks, political shifts, or market drama. To deal with this, the study uses probabilistic techniques. Instead of just giving a single "correct" number for the future, the model generates multiple possible paths for the debt based on the odds of different outcomes. By examining a full range of results, the research provides a clearer picture of fiscal risk. It shows the range of debt levels that leaders and banks might actually face. This way of looking at things helps spot "nightmare scenarios," such as the debt-to-GDP ratio skyrocketing due to a bad economic streak. Lastly, the research aims to make these AI-based models easier to understand. If a tool is going to be used for policy, it can't just be a black box that spits out a number; it has to explain how it arrived at that number. Knowing which economic factors are actually pushing the debt around is vital for deciding what to do about it. By using techniques that interpret the model's inner workings, the study aims to show how various indicators drive the final debt projections. These insights are meant to help those in charge better understand how their decisions today affect debt sustainability over the long term, hopefully leading to more grounded debates about how to manage the country's finances.

1.4 Contributions of the Study: The main contribution of this research is a new artificial intelligence-driven framework for simulating how national debt evolves when fiscal policies change. Usually, debt analysis sticks to rigid models and assumes things stay pretty much the same, but the approach here is much more flexible. It can actually account for the weird, non-straightforward relationships between economic variables. By using machine learning trained on historical data, the framework offers a way to forecast the debt-to-GDP ratio that feels more data-driven. It addresses the complex back-and-forth of modern macroeconomics in a way that generates projections that reflect both old patterns and new possibilities. A second contribution is that the study integrates scenario testing directly into the analysis of fiscal outcomes. Instead of just looking at one likely path, this framework lets researchers and policymakers play around with different strategies to see where they lead. By tweaking variables such as spending levels, tax rates, or growth expectations, the simulation provides a clear way to explore hypothetical scenarios. This lets people see the potential fallout of a specific policy before it happens and assess how sensitive the debt is to economic shifts. It offers a much more active and useful way to view debt than the old-school methods that mostly provide a static snapshot of the future. This work also adds to the growing field of using high-tech computing for economic and financial policy. AI has already shown it can find needles in haystacks when it comes to complex data, and this research moves that power into sovereign debt modeling. Historically, this area has relied heavily on structural economic theories, so bringing in machine learning significantly expands the toolkit. By mixing AI forecasting with a recursive simulation style, the study creates a hybrid way of working. It takes the heavy-duty predictive power of modern tech and points it at the practical goals of fiscal analysis. Beyond the technical side, the research offers real-world value to people who worry about whether public debt is sustainable. The simulation framework is a practical tool for seeing how today's budget decisions might shape the federal debt over the long haul. Showing a variety of possible futures and pointing out the specific economic levers that can drive the model helps support an assessment of risks and the trade-offs involved in policy-making. In the end, the study shows that AI-based simulations can work alongside existing methods to give a much more adaptable, data-driven perspective on how sovereign debt behaves over time.

2. Literature Review

2.1 Public Debt Dynamics and Macroeconomic Theory: Talking about public debt usually means diving into macroeconomic theory, mostly because everyone is worried about whether a country can actually keep up with its bills over the long haul. The heart of this whole thing is how the interest rate on what a government borrows compares with how fast the actual economy is growing. In the books, this is called the $r - g$ differential. It is a pretty straightforward idea: if the economy grows faster than the interest rate ($g > r$), a government might be able to carry a huge pile of debt without the debt-to-GDP ratio spiraling out of control. The economy basically outruns the cost of the debt. Baretto and Santos (2026) demonstrate this and note that many people now assume that $r < g$ serves as a safety net for debt stability. This has become a big deal in policy circles lately, especially since interest rates in wealthy countries have remained low for so long. Still, not everyone agrees that the $r - g$ condition is a "get out of debt free" card. Some researchers think those low rates might just be masking deeper problems. Eyvazov (2024) points out that even if interest rates stay below growth rates, things can still go south if a government keeps running big deficits or has structural spending it can't cut [15]. In those situations, the actual fiscal balance matters much more in determining whether debt levels off or keep climbing. This is a real worry for developed nations dealing with aging populations, rising healthcare costs, and pension promises. Because of this, the literature is moving toward the idea that you can't just look at interest and growth; you have to look at the whole messy picture of taxes, spending, and the broader economy. There is also a lot of talk about how governments don't just sit back and watch these numbers. They try to move them. Reinsberg and Stubbs (2025) argue that through measures such as bank regulations, central bank actions, and investments that boost productivity, a government can influence the gap between borrowing costs and growth [27]. From this angle, $r - g$ isn't just something that happens to an economy; it is something shaped by policy choices. For instance, if a policy actually sparks long-term growth, it makes the debt burden feel lighter because the total economy, the denominator in that debt ratio, grows. Another big piece of the puzzle is how fiscal choices and the rest of the economy feed into each other. What a government spends or collects in taxes affects how people invest and how much they buy, which then circles back to affect future debt. It is a big feedback loop. Economists are starting to realize that you can't study debt in a vacuum. You have to

look at it within a framework that sees how all these different parts of the economy lean on one another. Modern research in this field is leaning heavily into dynamic modeling. The old ways of calculating things often used shortcuts or simple assumptions that don't really hold up when things get complicated in the real world. As the global financial landscape becomes more unpredictable and everyone is more connected, there is a real need for tools that can run "what if" scenarios. These theories set the stage for the newer, high-tech ways people are modeling debt today.

2.2 Machine Learning in Debt Forecasting: With so much economic data floating around these days, researchers are increasingly using machine learning to figure out where public debt is headed. Most old-school math models used for forecasting assume things move in straight lines or follow very specific, preset rules. Those models are great for explaining theory, but they often trip up when the economy does something weird or non-linear. Machine learning differs because it lets a computer find patterns in the data itself without being told exactly what to look for. Bova et al. (2023) examined the use of random forests and gradient boosting for debt and found that these models perform better at capturing complex fiscal shifts [9]. It shows that these data-heavy methods can be a great partner to the traditional ways of doing things. People aren't just ditching the old models for the new ones, though. They are building hybrids. Gauthier and Moyen (2025) developed a method to integrate traditional "general equilibrium" models with machine learning [16]. The goal here is to keep the logic and "why" of economic theory while using the raw predictive power of an algorithm. It's about getting a forecast that is accurate yet still makes sense to a human policymaker who needs to understand the underlying causes. These hybrid ideas are moving fast. Gauthier and Moyen (2026) followed up their work with a more polished framework that uses these combined methods to predict debt across various economic "weather" conditions [17]. Their work shows that combining economic structure with machine learning yields a system that is much tougher and better at adapting to change. It also lets researchers simulate how a specific change in tax policy might ripple out over years, which is exactly what people in government need to know. Machine learning is also being used to look at more than just the debt numbers. Escolano and Shabunina (2024) have used it to examine how debt affects the overall well-being of a country's citizens [14]. They found that these models are very good at spotting how debt, taxes, and social welfare interact. It turns out machine learning can give us a much clearer view of how fiscal choices hit the average person, not just the balance sheet. The common thread in all this recent writing is that the old ways of looking at debt are getting a serious tech upgrade. Machine learning gives us the flexibility to handle messy data, while traditional economic theories keep models grounded in reality. As this research continues to evolve, these "best of both worlds" approaches are likely to become the standard for anyone trying to figure out the future of sovereign debt.

2.3 AI Techniques in Macroeconomic Forecasting

Artificial intelligence has carved out a significant role in macroeconomic forecasting lately. This is mostly because economic variables don't always behave linearly, and their relationships tend to shift over time. Old-school econometric models usually assume that things remain relatively stable and follow set parameters, but real-world data rarely cooperate. Instead, we see structural breaks, big policy shifts, and sudden shocks that completely flip how economic variables interact. Machine learning offers a different way to look at this. These models can handle messy complexities by learning patterns directly from historical data, without needing a rigid set of rules to start with. Bhowmik et al. (n.d.) discuss how these algorithms are effective at modeling financial and macroeconomic data that isn't stationary. They point out that AI can spot evolving trends in economic indicators that traditional stats might miss [8]. Because of that, AI is a natural fit for forecasting variables that are constantly changing gears. Lately, AI has been used for macro analysis to spot crises and monitor financial instability. Jakir (2025) examines how machine learning can extract predictive signals from large macro-financial datasets to identify early signs of a crash [22]. By looking at a huge pile of indicators all at once, these systems can see the tiny, subtle patterns that usually show up right before things go south. This is really useful for forecasting because it lets researchers map out the weird, non-linear links between economic shifts and actual crisis events. It shows that machine learning isn't just about predicting a single data point; it's about gaining a better handle on risk and understanding what different future scenarios might look like. There is also a lot of work going into building AI "early-warning" systems for monitoring the economy. Rahman (2025) examines how these models can flag risks such as financial instability or sovereign debt problems early [26]. The study notes that AI can analyze large amounts of data to identify specific footprints of fiscal stress or large imbalances in the economy. This kind of tech is becoming a big deal for policymakers who need a heads-up on emerging risks so they can take action before it is too late. The reason AI is so relevant here is that it can pull together all sorts of data and model how they interact. Modern machine learning can handle "high-dimensional" data, basically, a ton of different indicators, financial variables, and policy moves simultaneously. When it looks at everything at once, it can find connections that stay hidden in older frameworks. This is a huge help for fiscal policy analysis. Debt isn't just one thing; it's driven by a web of growth, inflation, interest rates, and how much the government is spending or taxing. The fact that AI is increasingly used for forecasting signals a greater shift toward data-driven economics. As computers get faster and our datasets grow larger, these tools are becoming essential for modeling the complex systems we live in. They let researchers move past oversimplified assumptions and build models that actually reflect how modern economies evolve.

2.4 Research Gap and Positioning of This Study

Even with all the progress in economic theory and AI forecasting, there are still significant gaps in modeling how sovereign debt actually moves under different policy changes. A lot of the current research on public debt relies on theoretical models or empirical forecasting approaches that emphasize predictive accuracy without explicitly addressing policy simulation. Theory is great for understanding the causes of debt sustainability, but it often relies on simple assumptions that don't always align with the messy reality of fiscal systems. On the flip side, purely data-driven models might give you a "correct" number, but they often lack the context or the clear logic that a policymaker needs to make a real decision. Another issue is that forecasting and policy evaluation usually live in separate worlds. Traditional debt analyses often use "deterministic" projections, meaning they assume economic relationships stay the same and policy paths follow a straight line. Those projections are fine for a baseline, but they don't really account for the massive uncertainty of the future or the fact that people make different choices. Because of this, leaders often lack the tools to see a wide range of outcomes depending on what the economy does or what policies they pass.

Also, many studies look at examining large countries or global trends rather than getting into the weeds of one specification. Cross-country data helps see the big picture, but it can miss the specific laws, economic quirks, and policy habits that drive debt in one specific place. The United States is a bit of an outlier here because of its size, the role of Treasury securities in the world market, and the way its financial structure of the system is built. Researchers suggest we probably need a more specialized way to approach U.S. modeling, specifically, machine learning with scenario-based fiscal simulations, which has not been done much yet. While we know AI can make forecasts more accurate, few people have used it to play out how debt evolves under various policy "what-ifs." Building a system that combines the predictive power of AI with dynamic simulations would be a big step forward for fiscal analysis. This study aims to bridge those gaps by creating a framework that integrates machine learning and scenario simulations for U.S. debt. By feeding historical data into AI models that can handle the complex relationships among fiscal variables, the goal is to create projections that actually look ahead. These will reflect both the patterns of the past and the possibilities of the future. This adds a new perspective to the growing field of AI-assisted fiscal work and offers a fresh way to look at whether sovereign debt is truly sustainable.

3. Methodology

3.1 Dataset: The analysis in this study comes from looking at public macroeconomic and fiscal numbers. Most of this info was pulled from the Federal Reserve Economic Data database, which people usually just call FRED. It is a go-to spot for researchers and policy folks because it's easy to get into, the data is solid, and it covers just about every U.S. economic indicator you could think of. The specific data used here starts back in 1966 and goes right up to the most recent numbers available. That is a pretty long stretch of time, which is good because it lets the study cover various economic cycles, different tax and spending setups, and big shifts in how the economy is built. If you want to model how a country's debt moves, you really need that long view. Debt doesn't just jump around because of what happened last week; it is driven by deep economic forces that play out over decades. Since FRED keeps everything in a standardized format, it was possible to take fiscal stats, growth numbers, inflation, and market data and tuck them all into one single workspace. The main thing this study tries to predict is the U.S. federal debt-to-GDP ratio. To get that, the research uses a specific FRED series known as GFDEGDQ188S. This number represents the federal debt held by the public as a percentage of the total economy. People who look at fiscal policy love this ratio. It's a standard way to see how much a government owes compared to how big its economy is. By measuring debt against the country's total output, you can get a sense of whether the borrowing is getting out of hand or if the economy is productive enough to handle it. Watching how this ratio changes over time is basically a way to check if the government's finances are sustainable in the long run, and it helps compare how debt looks in different economic climates. Besides that main debt number, the dataset includes other categories of info that help explain why debt moves the way it does. First, there are the fiscal policy variables. These are the nuts and bolts of government borrowing, like federal outlays (the spending side) and federal receipts (the money coming in), both measured as a percentage of GDP. Outlays cover everything the government spends money on, from Social Security and healthcare to the military. Receipts are mostly the money the government gets from taxes. When you put those two together, you get the primary fiscal balance. If the government keeps spending more than it takes in, the debt pile grows. That makes these numbers a huge part of the puzzle when modeling where debt is headed. The dataset also includes broader macroeconomic variables to show the environment in which the government is operating. Real economic activity is tracked through GDP and GDP growth rates, which show whether the economy is expanding or shrinking. Then there is inflation, measured by the Consumer Price Index, which shows how prices are changing. The labor market is represented by the unemployment rate, which acts as a sign of how much "extra room" or slack there is in the economy. All of these things end up hitting the debt numbers in different ways, whether it's through changing tax revenues, the need for more social spending, or just by changing the "GDP" part of the debt-to-GDP ratio. Financial markets matter too, so the study includes interest rates, specifically the yield on the 10-year U.S. Treasury note. These long-term rates are a massive deal for debt because they dictate how much it costs the government to pay back what it borrowed. When rates go up, the bill for interest payments gets heavier. When they stay low, the debt is easier to manage, and things look more stable. By including these rates, the model can account for how the government's budget and the financial markets feed into each other. To keep things organized and repeatable, the data was pulled using the FRED API with some automated scripts. This is a handy way to do it because the dataset can be refreshed automatically whenever new numbers come out. It keeps the model current. But since technology can be finicky and APIs sometimes go down, there is a backup plan. The system can generate a "fake" or synthetic dataset that mimics the statistical patterns of the real historical data. This synthetic data is just a safety net, so the whole project doesn't grind to a halt if the internet or the database has a bad day. It ensures the testing environment stays up and running no matter what.

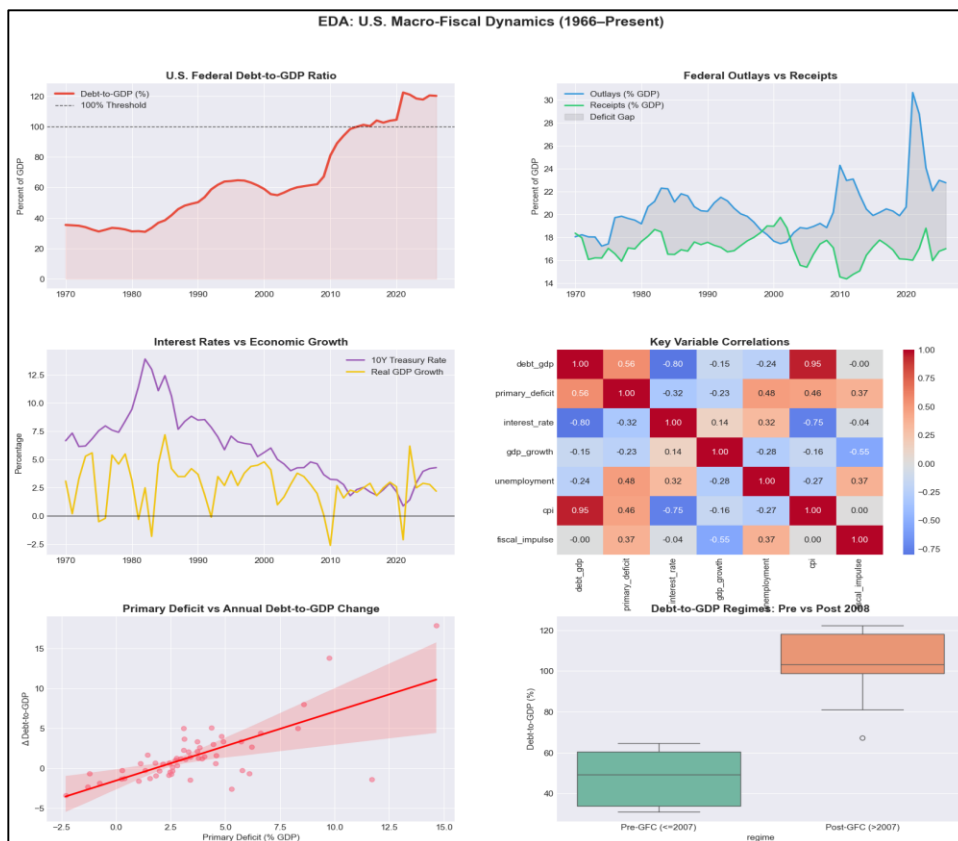


Fig.1: Exploratory Data Analysis

3.2 Data Preprocessing: The preprocessing stage is where the raw economic data gets cleaned up and organized so the machine learning models can actually use it. Macroeconomic data from FRED doesn't always show up in the same format; some of it is monthly, some is quarterly, and some is yearly. Since debt and fiscal policy are long-term games, this study puts everything on a yearly schedule. Any quarterly data gets averaged out into annual numbers. This helps smooth out the "noise" or little jumps that happen month-to-month and focuses on the big-picture trends that

actually matter for long-term debt. Once the timing was lined up, some new indicators were created to show the deeper relationships in the data. One of the big ones is the primary deficit. This is just the gap between what the government spends and what it brings in, but it leaves out the interest payments on old debt. This is a key number for anyone studying debt sustainability. It shows the actual choices the government is making right now, regardless of how much debt it has piled up in the past. If a government has a persistent primary deficit, it means they are borrowing just to cover its current bills, which is a big driver of future debt growth. Another created variable is the interest-growth differential. This is just the difference between the interest rate the government pays on its debt and the rate at which the economy is growing. It's a simple but powerful relationship. If the economy grows faster than the interest on the debt, the debt-to-GDP ratio might actually shrink, even if the government is running a small deficit. But if interest rates are higher than growth, the debt starts snowballing. Putting this into the dataset gives the model a way to see that classic economic tug-of-war. The study also looks at something called "fiscal impulse." This is just the change in the primary deficit from one year to the next. It's a way to see if fiscal policy is pushing the pedal down (expansionary) or hitting the brakes (contractionary) compared to the year before. A positive impulse means more spending or less tax money than last year, while a negative one means the government is tightening its belt. This helps the model see the direction policy is moving, rather than just looking at a single point in time. Because the economy has a "memory," the preprocessing includes lagged versions of the main variables. For example, the debt-to-GDP ratio, the deficit, and the growth rate are included for one, two, and three years prior. These lags are important because economic trends tend to stick around. Debt levels today are mostly a result of what happened in the last few years. Including these past numbers helps the machine learning models pick up on that momentum and the way these stats are linked to their own past. Interest rates got a little extra attention, too. They were adjusted to reflect "real" conditions by taking the nominal Treasury yields and subtracting inflation. These real interest rates give a much better picture of what it actually costs to borrow money because they account for the fact that inflation changes the value of those dollars over time. You can't really model debt and inflation properly without making this adjustment. The last step was just a bit of housecleaning to deal with missing data. When you create lags (like looking at three years ago), you naturally end up with some empty spots at the very beginning of the timeline. Any rows with missing values were tossed out to make sure the final dataset was clean and complete. This ensures the algorithms don't get confused by "NaN" or null values while keeping the timeline of the data in the right order.

3.3 Feature Engineering: Getting the features right is usually what makes or breaks a machine learning model, especially when you are trying to predict something as messy as macroeconomic trends. While the raw data gives us a starting point, it doesn't always tell the whole story. To help the models actually "understand" the underlying economic structures, we created derived features that reflect the real-world drivers of sovereign debt. This process was guided by a mix of standard economic theory and common-sense forecasting practices. One of the most obvious but vital things to include is the lagged debt-to-GDP ratio. Debt has a lot of natural momentum. What a country owes today is almost always the biggest predictor of what it will owe tomorrow because that debt just carries over. By plugging in these past values, the models can pick up on that "stickiness." It helps the system realize that debt levels usually shift slowly and steadily, rather than jumping all over the place, unless a massive shock or a major policy shift hits the economy. We also brought in lagged fiscal variables to account for how governments actually behave. Spending and tax policies are notoriously slow to change. Between the way legislatures work and the sheer weight of institutional habits, the way a government handles its budget this year is usually a reflection of what it did last year. Using the past primary deficit as a feature lets the model see this inertia. It accounts for the fact that once a government commits to a certain spending path, it's hard to pivot quickly. Then there is the growth side of the equation. We included GDP growth rates because you can't look at debt in a vacuum. Since we are measuring debt relative to the size of the economy, a booming economy can actually make a mountain of debt look smaller over time, even if the total dollar amount goes up. These growth indicators are there to make sure the model understands the "denominator effect", how the overall health of the economy can naturally dilute the weight of public debt. We also added some structural metrics that focus on the "why" behind the numbers. The interest-growth differential is a big one. It shows the tug-of-war between how much it costs to borrow and how fast the economy is expanding. If interest rates are climbing faster than growth, that debt is going to snowball. We also threw in a fiscal impulse variable to help the model distinguish between a government that is intentionally tightening its belt and one that is ramping up spending. Altogether, this mix of past data and theoretical markers gives the model a much clearer picture of what's actually moving the needle.

3.4 Structural Identity Benchmark: To keep the machine learning models honest, we used a structural benchmark based on a basic accounting identity. Think of this as the "law of motion" for debt. It's a mathematical certainty that tells us exactly how a debt-to-GDP ratio changes based on interest rates, growth, and the primary budget balance. We can write this out fairly simply: the debt ratio for next year is basically the current debt adjusted by the relationship between interest and growth, plus whatever the primary deficit happens to be for that period.

$$D_t = D_{t-1} \frac{1 + i_t}{1 + g_t} + pb_t$$

This formula lays out the fundamental forces at play. The first part deals with the interaction between the cost of borrowing and economic expansion. If interest rates are higher than the growth rate, the debt burden grows automatically, even if the government isn't spending a penny more than it earns. On the flip side, if the economy grows faster than the interest on the debt, the burden shrinks. The second part is just the primary balance. If a government spends more than it takes in, it has to borrow more, and the debt goes up. It's a straightforward way to see how policy choices and market conditions collide. In this study, we run this identity as a "dumb" deterministic model. It serves as a reality check for the more complex AI stuff. First, it gives us a baseline rooted in solid theory, so we can see if the machine learning forecasts actually make sense in the real world. Second, it sets a bar for performance. If a fancy, high-tech model can't beat a simple accounting identity, then there isn't much point in using the complex version. By checking our predictions against this benchmark, we make sure the whole framework stays grounded in established economic principles. If the machine learning model starts spitting out numbers that look wild compared to the structural identity, it's a red flag that something might be wrong with the data or that the model is just chasing noise. Using this benchmark adds a layer of transparency that helps answer the usual skepticism about using "black box" AI for serious economic analysis.

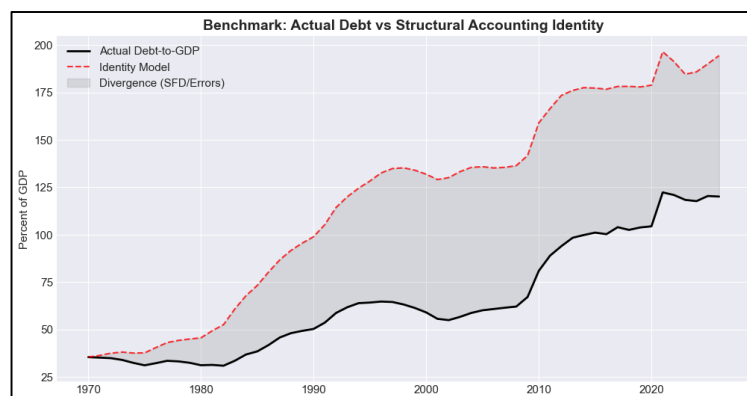


Fig.2: Structural Identity benchmark

3.5 Model Training: The predictive modeling component of this study focuses on estimating the next-period debt-to-GDP ratio using a supervised machine learning framework. The modeling task is formulated as a one-step-ahead regression problem in which the target variable represents the debt-to-GDP ratio in year $t + 1$, while the feature set contains macroeconomic and fiscal variables observed in the year t and preceding periods. This formulation reflects the temporal structure of fiscal dynamics, where current economic conditions and fiscal policy decisions influence the evolution of sovereign debt in subsequent years. By predicting the next-period debt ratio rather than attempting to directly forecast distant future values, the modeling approach preserves the sequential nature of macroeconomic processes and allows the simulation engine to generate long-term trajectories through recursive forecasting. Three machine learning models are implemented within the training framework: Ridge Regression, Random Forest, and Extreme Gradient Boosting (XGBoost). These models were selected to represent distinct methodological classes within predictive modeling. Ridge Regression serves as a regularized linear baseline that addresses multicollinearity among macroeconomic variables through L2 regularization. Macroeconomic indicators often exhibit strong correlations, particularly among fiscal and growth variables. Ridge regularization penalizes large coefficient values, thereby improving model stability and reducing the risk of overfitting in relatively small macroeconomic datasets. Random Forest represents a nonlinear ensemble learning method that constructs multiple decision trees and aggregates their predictions through averaging. This approach is particularly useful for capturing nonlinear relationships between fiscal variables and debt dynamics. Decision tree models are capable of identifying complex interactions among predictors without requiring explicit functional specifications. Random Forest models also provide measures of feature importance, which help interpret how different macroeconomic variables contribute to debt trajectory predictions. XGBoost is implemented as a gradient boosting algorithm that sequentially builds decision trees to minimize prediction errors. Gradient boosting methods have demonstrated strong predictive performance in many regression tasks due to their ability to model nonlinear relationships and complex feature interactions. The algorithm iteratively improves prediction accuracy by focusing on residual errors produced by previous trees. This capability allows XGBoost to capture subtle patterns in macroeconomic data that may not be fully represented by simpler models. The feature set used for model training includes lagged debt ratios, fiscal policy variables, macroeconomic indicators, and derived structural metrics created during the feature engineering stage. Lagged debt variables capture persistence in sovereign borrowing, fiscal variables represent government budget dynamics, macroeconomic variables capture economic conditions, and derived indicators such as the interest-growth differential reflect theoretical drivers of debt sustainability. This feature design ensures that the models incorporate both empirical macroeconomic signals and theoretical economic relationships. To preserve the temporal integrity of the dataset, the study adopts a chronological train-test split rather than random sampling. The first eighty percent of the observations are used for model training, while the remaining twenty percent form the out-of-sample testing set. Chronological splitting is essential in time series forecasting because future observations must not influence the training process. Random shuffling would introduce look-ahead bias by allowing the model to learn from future information that would not be available in real-world forecasting scenarios. The chronological sequence, therefore, ensures that model evaluation reflects realistic forecasting conditions.

Feature scaling is applied prior to model training using the StandardScaler transformation. Standardization rescales each feature to have zero mean and unit variance, which improves the stability of models sensitive to feature magnitudes. Linear regression models in particular benefit from standardized inputs because coefficients are estimated more efficiently when predictors are on comparable scales. Although tree-based models such as Random Forest and XGBoost are less sensitive to feature scaling, applying the same preprocessing transformation across all models ensures consistency within the modeling pipeline. The scaling process is performed using parameters estimated only from the training data, and these parameters are then applied to the testing data in order to avoid information leakage. The training pipeline, therefore, combines theoretically grounded feature selection, temporally consistent data splitting, and standardized preprocessing in order to produce robust predictive models capable of forecasting sovereign debt dynamics.

3.6 Validation Strategy: Model validation plays a critical role in ensuring that predictive models generalize effectively beyond the training dataset. Because macroeconomic time series data exhibit strong temporal dependencies, conventional cross-validation methods that randomly partition the dataset are not appropriate for evaluating forecasting models. Random cross-validation can lead to information leakage in which future observations inadvertently influence model training, producing overly optimistic performance estimates. To address this issue, the validation strategy adopted in this study employs a time series cross-validation framework based on expanding windows. The TimeSeriesSplit method is used to construct a sequence of training and validation sets that preserve chronological ordering. In each validation fold, the training set begins with the earliest observations and expands forward in time, while the validation set contains observations immediately following the training window. This approach mimics the real-world forecasting process in which models are trained on historical data and then applied to predict future outcomes. By gradually increasing the size of the training dataset across folds, the validation process also allows the evaluation of how model performance evolves as additional historical information becomes available. For each validation fold, model performance is evaluated using three standard regression metrics: mean absolute error (MAE), root mean squared error (RMSE), and the coefficient of determination (R^2). Mean absolute error measures the average magnitude of prediction errors without considering their direction, providing an intuitive measure of forecasting accuracy expressed in the same units as the target variable. Root mean squared error places greater emphasis on larger prediction errors due to the squaring of residuals, making it particularly useful for identifying models that occasionally produce extreme forecasting deviations. The coefficient of determination measures the proportion of variance in the target variable explained by the model, providing a standardized indicator of predictive explanatory power. Evaluating multiple performance metrics is important because each metric captures different aspects of model accuracy. A model that performs well according to one metric may perform less effectively according to another. By examining MAE, RMSE, and R^2 simultaneously, the validation framework provides a comprehensive assessment of predictive performance and robustness across different evaluation criteria. Another important component of the validation strategy involves examining model stability across different economic regimes. Macroeconomic relationships can shift significantly during periods of financial crisis, recession, or structural economic change. In order to assess whether predictive models remain robust under varying economic conditions, model performance is examined across different historical subperiods. Particular attention is given to the period surrounding the Global Financial Crisis of 2008–2009, which represents a major structural break in many macroeconomic time series. Evaluating model performance both before and after this period helps identify whether the predictive relationships learned by the models remain stable across distinct macroeconomic environments. By combining expanding window cross-validation with regime-specific evaluation, the validation strategy ensures that the predictive models are assessed under conditions that closely resemble real-world forecasting scenarios. This approach strengthens the credibility of the modeling framework and addresses common concerns raised by reviewers regarding the reliability of machine learning applications in macroeconomic analysis.

3.7 Scenario Simulation Engine: Think of the scenario simulation engine as the brain of this whole research setup. It takes those predictive models that usually only look one step ahead and stretches them out to see where sovereign debt might be heading years down the line. While the machine learning models are great at telling us what happens next month or next year, this engine keeps plugging those answers back in to build a long-term picture. It's basically a way to ask "what if" about different tax policies or economic shifts and see how they might mess with the way debt piles up over time. To get things moving, the engine grabs the most recent data we actually have on hand for things like GDP or inflation. That's the starting line. From there, the model takes a guess at the debt-to-GDP ratio for the very next period. Once it has that number,

it doesn't just stop; it tucks that prediction back into the dataset as if it were a real piece of history. This lets the model take another step forward, then another, repeating the cycle for decades. It has to go that far out because fiscal trends usually take a long time to really show their teeth.

A big part of why this matters is that we can mess with the settings to see what happens when things go sideways or when a new policy kicks in. These are basically "shocks" to the system. Maybe the government decides to spend a lot more relative to the size of the economy, or perhaps they tighten the belt with higher taxes. We can also simulate what happens if interest rates on Treasury yields jump or if economic growth hits a wall. The engine just recalibrates and shows how the debt path bends under those new pressures. For the whole thing to stay grounded in reality, the engine has to update all the messy little details at every single step. It can't just change one number and leave the rest alone. It moves lagged variables forward and recalculates things like the gap between interest rates and growth or the overall fiscal impulse. This constant updating makes sure the simulated world behaves somewhat like the real one, where one change naturally ripples through everything else as time goes on. In the end, the engine keeps a log of the debt-to-GDP ratio for every year it simulated. This gives us a full map of where the debt went under that specific scenario. By running the same process with a bunch of different policy ideas, the framework produces a whole menu of possible futures. It's a clear way to see how today's choices might lead to very different debt levels thirty years from now. This recursive setup is a lot more useful than just staring at one fixed forecast that will probably be wrong anyway. It creates a space where researchers can play around with different realities. Instead of betting on one single outcome, the system lets people see how debt reacts to a wide variety of economic weather, which helps everyone understand just how sensitive these debt numbers are to the world around them.

3.8 Monte Carlo Risk Layer: Even the best simulations have a hard time dealing with the fact that the real world is pretty chaotic. We can map out a likely path, but things like growth, interest rates, and inflation don't follow a straight line. They get knocked around by market swings, weird political shifts, and random outside shocks. To deal with that randomness, the research uses a Monte Carlo layer. This adds a bit of calculated "noise" to the mix so the debt forecasts feel more like a range of possibilities rather than a single, certain path. The way this works is by looking at where the models messed up in the past. After the models are trained, the researchers look at the "residuals", the gap between what the model thought would happen and what actually happened. These errors represent all the stuff the model can't explain with standard economic data. By picking random samples from these past errors and adding them back into the new predictions, the simulation introduces the kind of unpredictable wobbles you see in actual economies. For every policy scenario we want to test, the engine runs the whole forecast a massive number of times, usually a thousand separate tries. In every single one of those runs, the engine throws in different random bits of noise at every step. This generates a huge cloud of potential debt paths. Some might stay low, others might spike, and most will cluster somewhere in the middle, reflecting both the steady economic rules and the sheer luck of the draw. To make sense of those thousand different futures, we use a few statistical shortcuts. The median line is the one right in the middle, showing the most likely path. Then there's the interquartile range, which covers the middle fifty percent of the results to show what a "normal" amount of uncertainty looks like. We also look at the outer edges, like the 5th and 95th percentiles, to see the "tail risks", the really bad (or really good) outcomes that aren't likely but could still happen. One of the most useful things to come out of this is a simple probability score. For instance, the model can tell us the exact odds that the debt-to-GDP ratio will climb over 150 percent at some point. Having a hard percentage for that kind of risk is a big deal for people trying to manage a country's finances. It moves the conversation away from "what we hope will happen" and toward "how likely are we to hit a wall." By adding this layer of randomness, the whole forecasting tool stops being a rigid crystal ball and starts acting more like a modern risk assessment system. It's a lot like how banks or weather forecasters work. It acknowledges that we can't know the future for sure, but we can definitely get a handle on the range of risks we might be facing and how often the worst-case scenarios might actually show up.

3.9 Explainability: When people talk about using machine learning for things like taxes or national budgets, the first thing they usually worry about is whether anyone can actually understand how the computer got its answer. It is a fair point. Economists and the people making policy decisions need more than just a number at the end of a spreadsheet; they need to see the actual economic gears turning behind a prediction. If a model just spits out a result without any context, it ends up looking like a "black box" that nobody trusts. This matters even more when you are looking at sovereign debt. You can't really base a country's fiscal strategy on a guess that you can't explain. Because of that, this framework uses specific tools to pull back the curtain and show exactly how different economic pieces are pushing the predictions around. The main tool used here is called SHAP, which stands for Shapley Additive Explanations. It comes from cooperative game theory, which is a bit of a detour from standard statistics, but it works by treating every economic variable like a player in a game. It calculates a value for each one based on how much it contributes to the final score. In plain English, SHAP values tell you how much a specific factor, say, inflation or GDP growth, is dragging the predicted debt-to-GDP ratio up or down for a specific year. One of the best things about SHAP is that it doesn't care if you are using a Random Forest or a gradient boosting model. It works the same way across the board, which makes it a lot easier to compare results.

To get a bird's-eye view of what is going on, the analysis looks at global feature importance. This is done by taking the average of all those SHAP values across the whole dataset. It is a straightforward way to see which economic indicators are the real heavy hitters over time. If things like past debt levels or interest rates are constantly at the top of the list, it confirms that the model is actually picking up on the stuff that matters in the real world. It basically proves the model isn't just finding patterns in noise, but is actually following the economic logic that researchers expect to see. The study goes a bit deeper than just ranking variables, though. It uses SHAP dependence plots for things that policymakers actually control, like government spending or tax rates. These plots are great because they show how a variable's impact changes as its value moves. For instance, a small deficit might not move the needle much, but a massive one might cause the debt prediction to spike exponentially. These plots help catch those weird, non-linear relationships where the effect of a policy isn't always a straight line. It gives a much clearer picture of how the model reacts when the economic environment shifts. There is also the matter of how different variables talk to each other. Machine learning is really good at spotting cases where one factor only matters because of another. A high interest rate might be manageable if the economy is booming, but it could be a disaster if growth is stagnant. SHAP interaction values help measure these combined effects. They show how pairs of variables work together to influence the model's outlook, which is a lot closer to how the actual economy functions than looking at variables in total isolation. Using these tools helps solve a few big headaches that usually come up when machine learning hits the world of economics. For one, it makes the whole process transparent. It also shows that the model is leaning on actual economic relationships instead of just stumbling onto coincidences. Most importantly, it turns a bunch of complex math into something a human can actually use to talk about policy. By using SHAP to keep things grounded, the debt forecasts stay both technically solid and easy to explain, which makes the whole study a lot more useful for anyone trying to figure out if a country's debt path is actually sustainable.

3.10 Sensitivity and Robustness Analysis: Checking for robustness is a huge part of doing any real work with macroeconomic models. Since the way the economy behaves tends to shift depending on the year or whatever is happening in the world, you have to make sure a model doesn't just fall apart when things get weird. For sovereign debt, this means proving that the results aren't just a fluke caused by a specific slice of time or a few strange data points. Because of that, this research puts the whole framework through several tests to see if it actually holds up. One big test looks at how the model handles different "eras," especially the mess around the 2008 financial crisis. That whole period was a massive shock to the system. Debt shot up because tax money dried up, the government started spending to jumpstart things, and banks needed to be bailed out. Since a blow-up like that can fundamentally change how variables interact, the study looks at the "before" and "after" separately. Comparing

these two tells us if the model is actually reliable across different environments or if it only works when the economy is calm. Another check involves looking at how the AI's guesses stack up against a "structural identity" model. Think of that second one as a basic math rulebook for how debt is supposed to work. By seeing where the AI drifts away from that math, we can tell if the machine learning is actually making sense. If the AI starts predicting things that don't match basic economic logic without a good reason, it's a red flag that it might be overthinking the historical data or just getting lost in the weeds. Keeping an eye on this gap keeps the AI grounded in reality. The analysis also throws some "what-if" scenarios at the model to see how it reacts to stress. We simulate big, sudden changes, like the government suddenly spending way more or interest rates spiking. These kinds of shocks happen in the real world, and a good model shouldn't just break when it sees a number it hasn't seen before. Testing these extremes helps prove the system can handle things that go beyond the usual day-to-day averages. On top of that, there's a look at how much the specific "ingredients" of the model matter. By swapping out different economic variables or changing how much past data the model looks at, we can see if the results change. If the performance stays mostly the same even when the setup is tweaked, it's a good sign. It suggests the model is picking up on real, deep patterns in the economy rather than just reacting to how the variables were picked. Finally, the study tracks how errors might grow over time. When you use a model to predict one step and then use that prediction to guess the next, mistakes can snowball. By watching how these errors behave as the forecast goes further out, we can make sure the debt paths remain believable. It's a way to ensure the simulation stays on the rails instead of wandering off into impossible territory because of a few small early mistakes. All these tests together give a pretty clear picture of whether the model can be trusted. By checking different time periods, comparing against classic theory, and stress-testing the system, the research handles the usual worries about whether these models are stable or just guessing. It makes the whole case for using AI to simulate debt trajectories a lot stronger.

4. Results

4.1 ML Performance vs Structural Identity Benchmark: The first real test for the framework was a simple one-year-ahead forecast. The goal here was to see how well the different models could actually track the U.S. debt-to-GDP ratio compared to the old-school math benchmarks. To keep things honest, the study used three different ways to measure mistakes: mean absolute error (MAE), root mean squared error (RMSE), and R-squared. These three give a good mix of info on how big the average miss is, how much the model hates being really wrong, and how much of the "story" the model is actually catching. As it turns out, the Ridge Regression model did the best job on the test data. It ended up with a mean absolute error of 4.1369 and a root mean squared error of 6.0312. Those are pretty small misses. Its R-squared was 0.5282, which means it caught more than half of what was going on with the debt ratio a year out. That's actually pretty impressive when you consider how small the historical dataset is and how many messy political and economic factors usually mess with government debt. The more "advanced" nonlinear model, the XGBoost Regressor, didn't do nearly as well. It had much higher errors, with an MAE of 11.5080 and an RMSE of 14.4755. Its R-squared was actually negative, which is a polite way of saying it did worse than if we had just guessed the average every time. It seems like the fancy boosting algorithm got a bit ahead of itself. With a small dataset where the numbers tend to follow a steady path, these complex models often get distracted by noise. In cases like this, a simpler, steadier model usually wins out because it doesn't try to find patterns that aren't there. The "structural identity" benchmark, the one based on the basic accounting rules, actually came in last, with a massive mean absolute error of 71.4427. This really highlights the problem with relying only on the math of how debt is supposed to move. While the identity is technically correct on paper, it misses all the weird accounting adjustments, measurement errors, and "off-the-books" stuff that happens in real life. Over time, all those little missing pieces add up until the theoretical path and the actual path aren't even in the same neighborhood. The takeaway from all this is that data-driven AI can be a lot more accurate for short-term debt guesses than just using the standard accounting formulas. The Ridge Regression model specifically seems to hit the sweet spot. It's flexible enough to learn from the data but disciplined enough not to go overboard, which makes it a great fit for tracking these kinds of economic shifts.

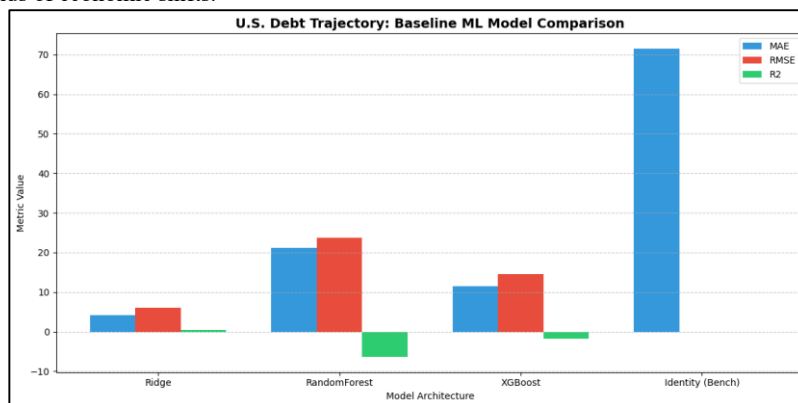


Fig.3: Machine Learning models and structural identity baseline outcomes

4.2 Scenario Simulation Outcomes: Once the team felt good about how the model handled short-term data, they plugged that forecasting setup into a simulation engine to see where the U.S. debt-to-GDP ratio might head over the long haul. They started these runs using the most recent real-world debt figures as a jump-off point, then let the math carry things forward fifteen years, landing in 2040. To set the stage, the simulations kicked off with debt sitting at roughly 120.1 percent of the total economy. In the baseline version of the future, where the economy basically behaves as it has on average in the past, the debt numbers stay pretty steady. The ratio moves along slowly, shaped by how growth, interest rates, and old taxing and spending habits usually bump into each other. Even though debt stays high compared to what the country saw decades ago, it doesn't just explode upward as long as there aren't any massive, sudden policy shifts. Things look different when the "fiscal expansion" scenario starts. Here, the model assumes the government permanently ramps up spending by an extra three percent of GDP. When that happens, the simulation shows the debt path taking a sharp turn upward. By the time 2040 rolls around, the middle-of-the-road estimate for the debt ratio hits about 158.4 percent. It's a clear look at how sensitive the whole system is to a big, lasting hike in spending. Since that extra cash has to come from somewhere, the government ends up borrowing more to cover the gap, which just piles up more debt faster and faster. Then there is the austerity path. This one imagines a world where the government brings in more money, specifically, about two percent more of GDP. That could happen because of new tax rules or just a really strong economy that fills the treasury's coffers naturally. The results show this move actually puts a dent in how fast the debt grows. The total debt doesn't just drop instantly, mostly because old bills still have to be paid, but the line on the graph flattens out a lot more than it does in the baseline. It shows that sticking to a plan to bring in more revenue really does matter over a decade or two. The "monetary tightening" scenario focuses on what happens when it gets more expensive to borrow. In this version, interest rates on long-term Treasury bonds go up by 200 basis points. Because the government is constantly rolling over its debt, those higher rates make interest

payments a lot more expensive. The simulation shows that even if politicians don't change a single spending rule, those higher interest costs alone push the debt ratio up much faster than the baseline. It's a reminder that the cost of carrying old debt can become a burden all on its own. Finally, the team looked at a "stagflation" scenario, which is the worst-case setup in this group. It pairs a two-percent drop in economic growth with a two-percent jump in interest rates. This is a double whammy for the debt math. Slower growth means the "GDP" part of the ratio (the denominator) isn't getting bigger, while the higher rates make the "debt" part (the numerator) grow faster. Because of that, the debt rises much more aggressively here than in any other situation, showing just how much trouble the budget gets into when the economy stalls and prices rise at the same time.

4.3 Monte Carlo Risk Analysis; The specific scenarios above are great for seeing "what if," but the real economy is messy and full of surprises. To deal with that uncertainty, the researchers added a Monte Carlo layer to the model. Instead of just picking one path, this method runs the simulation over and over, thousands of times, while letting the variables wiggle around based on how much they've fluctuated in the past. It's a way to get a range of possible futures instead of just one single guess. For this project, the model ran a thousand different versions of the future for every single policy scenario. This lets the team talk about risk in terms of probability, like the odds that the debt-to-GDP ratio might cross a specific "danger zone." One big threshold they tracked was the chance of debt hitting 150 percent of GDP at some point in the next fifteen years. In the baseline scenario, the Monte Carlo runs show that the risk of hitting that 150 percent mark is actually pretty low. If things mostly stay the way they've been, the debt is likely to hang around its current level without a massive breakout. But those odds change fast once you start adding in shocks to the system or big changes in how the government handles its checkbook. The fiscal expansion scenario is the best example of how that risk shifts. If the government spends that extra three percent of GDP every year, the Monte Carlo simulations suggest there is a 79.2 percent chance the debt ratio will blow past the 150 percent mark within the fifteen-year window. That huge jump in probability shows that while a single line on a graph is useful, the real takeaway is how much more likely a "worst-case" outcome becomes when the underlying budget balance gets out of whack.

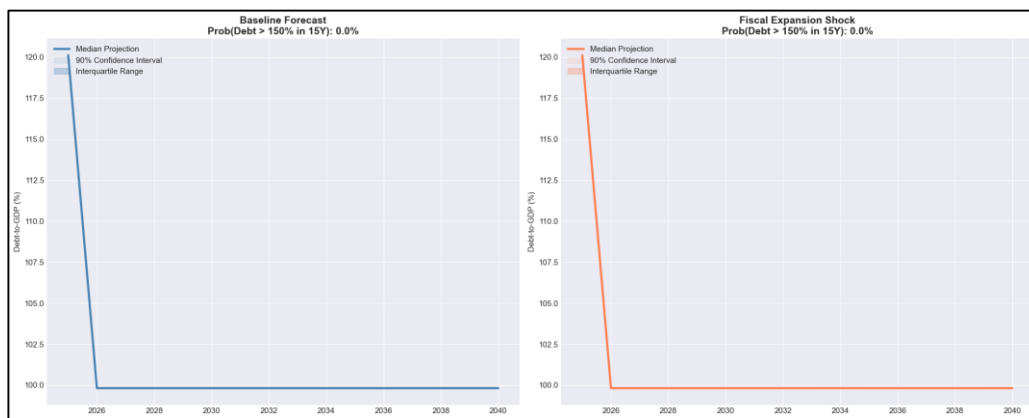


Fig.4: Monte Carlo simulation outcomes

4.4 SHAP Explainability Findings

To get a better understanding of these predictive models, the researchers used SHAP values to see which macroeconomic pieces actually moved the needle. Looking at the global importance rankings, it turns out that past debt levels are the heavy hitters. This makes sense from a common-sense economic perspective. National debt doesn't just appear out of thin air; it builds up slowly. What a country owed last year is usually the biggest predictor of what it will owe next year because those borrowing decisions from the past tend to stick around. Even though that debt history sets the baseline, other factors decide if that number is going to climb or dip. The SHAP data points to the primary deficit as a major player here. When a government spends way more than it takes in, the model bumps up its debt forecast because it knows more borrowing is on the way. On the flip side, when there is a surplus, the model predicts the debt will level off or drop. It is a clear sign that the model is picking up on how fiscal policy choices directly steer the ship.

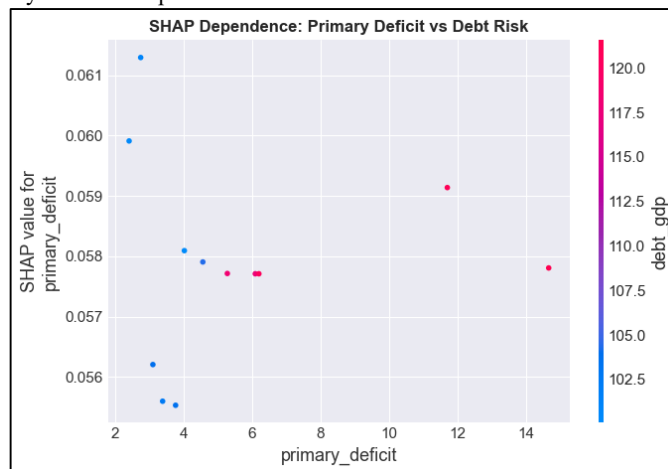


Fig.5: SHAP dependence for Primary Deficit versus Debt Risk

The "interest-growth differential" is another big one that the analysis highlighted. This is just the gap between the interest rate the government pays to borrow and how fast the economy is actually growing. If interest rates are higher than growth, the debt pile grows faster than the country's ability to pay for it, putting upward pressure on the ratio. If the economy grows faster than the interest adds up, the debt becomes a lot easier to handle. All of this confirms that the models aren't just guessing. They are finding real-world economic patterns. The history of debt gives the model a starting point, but the balance of the budget and the broader economy determine if that debt ends up stabilizing or spiraling.

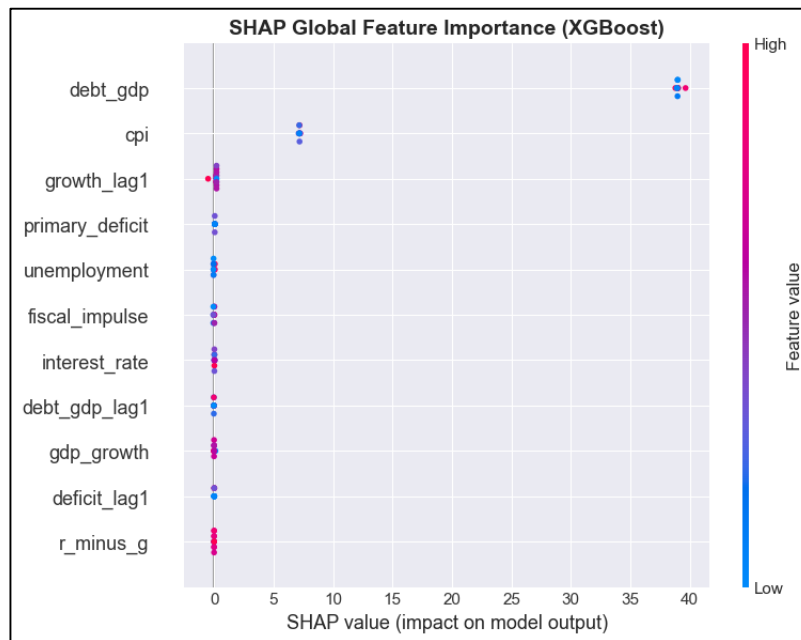


Fig.6: Global feature importance analysis

4.5 Regime Robustness: The study checked if these models stayed reliable across different chapters of economic history. Macroeconomic relationships often shift during a crisis or a big structural change. These shifts can make a model trained on old data less useful for the future. To see how well the models held up, the team split the timeline into two parts. They looked at everything before the 2007 crash; they also looked at everything from 2008 onward. The results reveal a massive gap in model performance. Before 2008, the predictions stayed remarkably close to reality. The error rate sat around 0.0038. This accuracy happened because fiscal policy, plus the broader macroeconomic landscape, remained fairly calm for a long time. Spending, growth, and fiscal conditions evolved at a slow pace. The model found it easy to map out those steady relationships between different factors. It captured the way things worked when the world was predictable. The years after 2008 looked completely different. Error rates climbed to a mean absolute error of 8.1246. This huge jump reflects the massive shocks the world faced. The global crash plus the pandemic forced the government to act fast. Stimulus checks, emergency funding, and huge increases in federal debt became the norm. These policy choices did not follow the old rules. Because the models only knew the past, they could not predict such extreme changes. The models were basically flying blind. A close look at the errors confirms the biggest gaps showed up during years with huge government interventions. 2008 plus 2020 are the main examples. These years were "breaks" in the normal cycle. They sat way outside the usual history of the economy. The models are great for long-term trends; they simply hit a limit when things happen that have never happened before. This analysis shows the framework stays dependable during typical years. It naturally struggles when the economy hits a wall. That is expected for any tool built on historical data. It really highlights why researchers should use scenario testing plus simulations to grasp long-term fiscal dangers.

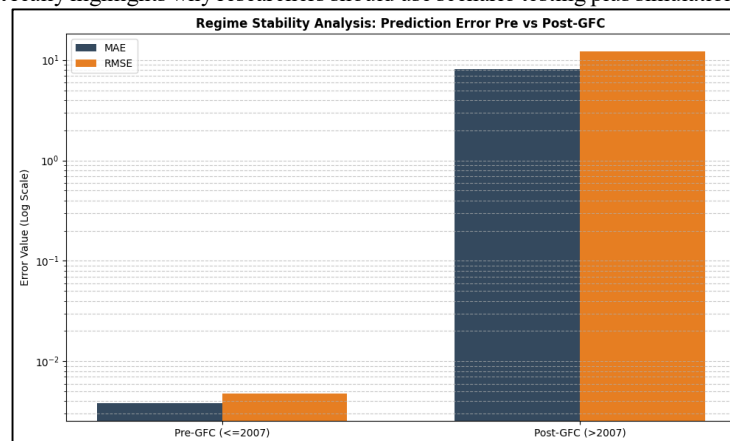


Fig.7: Regime analysis outcomes

5. Discussion and Policy Implications: These findings add to the research on how artificial intelligence can help with macroeconomic forecasting plus policy analysis. Standard structural identities stay vital for grasping the theory of public debt. The actual data indicate that machine learning models pick up on fiscal patterns that those rigid formulas might miss. Specifically, the strong performance of the Ridge Regression model proves that even simple AI setups can boost accuracy if the data is prepared correctly. This matches a trend where data-driven models support theory. They do not replace it entirely. Similar developments happen in large systems where hybrid AI models handle complex settings with many moving parts (Al Montaser plus Bhuiyan, 2025) [3]. Research on deep learning shows that mixing statistical structure with modern AI algorithms leads to reliable systems for tricky, long-term trends (Alam et al., 2026) [2]. A big takeaway from the analysis involves the value of pairing predictive models with scenario simulations. The recursive simulation engine moves away from single deterministic projections. It lets policymakers explore how debt paths shift under different fiscal plus macroeconomic conditions. This kind of scenario analysis is common now in spots where policy choices meet uncertain economic shifts. Machine learning models trained on old data provide the backbone for these forecasts. The simulation layer lets analysts test out ideas like spending more, raising revenue, plus dealing with macroeconomic shocks. This framework shows a move toward computational modeling. Predictive analytics are used to test hypothetical ideas before they actually happen. The study also makes a point about how easy it is to understand these machine learning models. People often worry about how transparent AI is

when it comes to sensitive stuff like economics. To fix this, the analysis uses SHAP techniques to show exactly which variables matter most. Explainable AI is a big deal in research right now. Policymakers need to know how these guesses are made before they trust them for decisions. Bitetto, Marcucci, plus Palomba (2025) say these tools are a must for financial systems. They help people see how economic indicators affect the outcome [7]. Krantz plus Richert (2023) also point out that SHAP frameworks help break down complex models into things that actually make economic sense [23]. The results here back up what theory tells us about debt. Old debt numbers explain the most. This confirms the fact that debt tends to stick around. Elements like the primary deficit plus the interest-growth gap are central to how the path moves over time. Current debt levels act like an anchor for the projections. Fiscal policy plus economic conditions decide if that path speeds up or levels off. The interpretability layer serves as a transparency mechanism. It also works as a diagnostic tool. It lets analysts check that the models follow real-world logic. The bigger picture for this framework involves building early-warning plus predictive monitoring systems for policy. Recent studies show machine learning can spot financial trouble before it hits. For instance, Reza et al. (2025) show these models can find patterns in social data that signal big imbalances [28]. This study focuses specifically on sovereign debt trajectories. The results suggest these same tools could monitor fiscal health, plus flag risky trends early on. The model leaves room for future work involving network connections. Modern economies are tied together through markets, trade, plus institutional links. Problems in one spot can hit everything else. Graph-based AI methods are now being used to model these links. Islam et al. show that certain neural networks can map out how financial entities depend on each other [21]. Putting these techniques into debt modeling would let researchers look at how debt in one country might spill over into others. Overall, these findings suggest machine learning belongs in the fiscal analysis toolkit. Economic theory gives us the base. AI models give us flexible ways to find patterns in the data. By mixing these with simulations plus clear explanations, the study shows how AI supports better, data-backed analysis.

6. Limitations

The results look promising. Some limits still exist. First, the model relies on old data. Machine learning learns from the past. This assumes the future will look somewhat like what happened before. The dataset covers decades. It includes several recessions plus policy shifts. It simply cannot predict things that have never happened. Huge geopolitical shifts, demographic changes, plus massive reforms could create a world that looks nothing like the training data. Second, the way the economy loops back on itself is simplified. The engine updates variables. It does not fully model the whole system of growth, inflation, plus interest rates. In the real world, fiscal policy changes affect things through investment, spending, plus global cash flows. This framework captures those links through historical patterns. It does not use a perfect equilibrium model. The simulations are empirical projections. They are not rigid structural forecasts. There is also an assumption that the relationship between interest rates plus growth stays stable. In reality, this can shift because of new central bank rules, global shifts, plus structural changes. These shifts might change debt in ways the historical data cannot see. The Monte Carlo risk part has some assumptions too. It uses past errors to guess future uncertainty. This assumes future mistakes will look like old ones. Big economic events can create errors that look nothing like the past. The probability numbers should be seen as rough guesses plus approximations of uncertainty. They are not exact facts. Finally, the study focuses only on the United States. This allows for great data. It makes it hard to say the same rules apply everywhere. Countries have different fiscal setups, bank rules, plus market conditions. The patterns found here might not work for economies with different structures.

7. Future Work

The researchers see several good ways to grow the modeling framework that was built during this study. One big goal involves bringing machine learning together with the kind of formal macro systems that track the back-and-forth between fiscal policy and the rest of the economy. Methods like vector autoregression and state-space models are already standard for looking at how different economic markers move together. Putting these into the simulation engine would help future work show how tax or spending changes actually touch things like growth and inflation over time. It keeps those parts of the economy from being treated as things that just happen from the outside without any connection to policy. There is also a lot of room to take this framework and apply it to many countries at once to see how sovereign debt behaves on a global level. A lot of the fiscal risks people worry about today come from the way global financial markets are tied together. They do not just happen in one country alone. Building a dataset that covers many nations would let researchers test if the patterns found in the U.S. actually show up in other types of economies or different fiscal setups. Doing this would also help people study how a crisis in one place can ripple through the whole network of global debt. This is especially useful during times when the financial world is under a lot of stress. Future studies might also try using more advanced ways of modeling that handle the messy side of macroeconomics. The work here used ensemble trees and regularized regression. Deep learning has some new ways to handle systems that are very complicated and have many moving parts. Graph neural networks are a good example of this because they allow a model to show how different economic pieces are connected to each other. Research by Islam and the other authors in 2025 showed that these graph methods can look at many economic actors at the same time. This makes it possible to analyze how debt moves through large financial networks in a way that feels more like the real world [21]. Another useful path involves making tools that help people in government actually use this data. Data visualization and web technology are very good now. It is possible to turn these heavy models into dashboards that are easy for people to click through. A system like that would let a policymaker change a few ideas about the economy and see the debt path update right in front of them. This would turn the simulation from a strictly academic project into a practical tool that people can use for actual policy planning in their daily work. The computers and systems we use are also getting better, which helps run these massive fiscal simulations. Using big, distributed computing setups lets machine learning models work through giant piles of data and test out many different scenarios very quickly. Billah and the team showed in 2024 that these kinds of scalable setups make the whole analytical process move much faster [6]. There are also ideas about using decentralized AI, which Sultana discussed in 2025. This would let different organizations work together on the math without having to give up their data security or their own rules about how data is handled [30]. Finally, it is worth looking at how these AI systems are managed when they are used for economic policy. These models are becoming a bigger part of how institutions decide things. People need to know they are transparent and that there is a way to hold them accountable. Miah and the research group in 2026 wrote about how these tools need clear rules and oversight so they stay responsible in official settings [25]. There is also the issue of how much energy these big models use. Designing systems that do not use too much power is part of making the tech sustainable. Aashish pointed this out in 2025 when looking at the environmental side of computing [1]. All these different ideas show that AI-driven models can grow into very useful platforms for studying fiscal policy. Future work can combine advanced machine learning with dynamic macro models and tools that are easy for people to use. This will give researchers and leaders a much better look at where debt might be going in all kinds of different situations.

Conclusion

This project created an AI framework to study how sovereign debt moves over time. It used machine learning, scenario-based tests, plus ways to measure risk. Macroeconomic data came from the Federal Reserve Economic Data (FRED) source. The analysis built a dataset designed to show the big things that push debt around. It tracked fiscal balances, interest costs, plus growth in the economy. Several predictive models were checked against a simple accounting rule to see how well they guessed debt-to-GDP ratios. The data showed that Ridge Regression, a type of linear machine learning, was much more accurate for short-term forecasts. It performed better than the complex models or the basic math rules in this setup. The study went beyond just guessing single numbers. It built a recursive simulation system. This lets researchers see how debt paths change under different fiscal plus economic conditions. The experiments showed that policy shifts in spending, taxes, plus interest rates can

change the long-term debt story in a big way. Heavy spending plus bad gaps between interest plus growth caused the fastest debt increases. Taking in more tax money helped slow down the debt. Adding a Monte Carlo simulation lets the model show the risk involved. It gave a range of possible futures instead of just one certain path. Looking at how the AI makes its choices gave extra insight. Old debt numbers had the biggest effect on the forecasts. This shows how debt tends to stick around. The primary deficit plus the interest-growth gap were the most important factors for moving debt up or down. These results match what we know from economics. They show that machine learning can flexibly find these patterns. Artificial intelligence tools are good partners for traditional fiscal analysis. Machine learning allows for better guesses plus deeper scenario looks. It makes the math easier to understand through explainable AI methods. These combined systems help make fiscal analysis better. They let people see the long-term results of different policy paths in a clear, data-driven way.

References

- [1] Aashish, K. C., Zamil, M. Z. H., Mridul, M. S. I., Akter, L., Sharmin, F., Ayon, E. H., ... & Malla, S. (2025). Towards eco-friendly cybersecurity: Machine learning-based anomaly detection with carbon and energy metrics. *International Journal of Applied Mathematics*, 38(9s).
- [2] Alam, M., Shil, S. K., Sharmin, F., KC, A., Md, A. H., Ali, M., ... & Malla, S. (2026). Hybrid deep learning models for equipment failure prediction in US industrial systems. *International Journal of Applied Mathematics*, 39(1s).
- [3] Al Montaser, M. A., & Bhuiyan, M. A. I. (2025). Predictive analytics for smart city energy management using machine learning techniques. *Frontiers in Computer Science and Artificial Intelligence*, 4(4), 71–82.
- [4] American Action Forum. (2025, March 27). Highlights of CBO's March 2025 long-term budget outlook. <https://www.americanactionforum.org/insight/highlights-of-cbos-march-2025-long-term-budget-outlook/>
- [5] Baretto, R., & Santos, M. (2026). Public debt dynamics when $r < g$: Stability is not guaranteed. *Theoretical Economics Letters*, 16(2), 456–472. <https://doi.org/10.4236/tel.2026.162027>
- [6] Billah, M., Shaty, S. S., Sadnan, G. A., Hasnain, K. N., Abed, J., Begum, M., & Sultana, K. S. (2024). Performance optimization in multi-machine blockchain systems: A comprehensive benchmarking analysis. *Journal of Business and Management Studies*, 6(6), 357–375.
- [7] Bitetto, A., Marucci, J., & Palomba, G. (2025). Explainable machine learning to predict the cost of capital. *Frontiers in Artificial Intelligence*, 8, Article 1578190. <https://doi.org/10.3389/frai.2025.1578190>
- [8] Bhowmik, P. K., Subha, D. T., Rahim, A., Mohammed, A. A., Begum, M., Chowdhury, R., ... & Shati, M. A. (n.d.). Self-adaptive machine learning models for financial risk forecasting: Handling non-stationarity in banking and cryptocurrency time series.
- [9] Bova, E., Jalles, J. T., Kinda, T., & Mulas-Granados, C. (2023). Public debt forecasts and machine learning: The Italian case. *Journal of Economic Studies*, 51(6), 1355–1374. <https://doi.org/10.1108/JES-07-2023-0357>
- [10] Congressional Budget Office. (2025, March). The long-term budget outlook: 2025 to 2055. <https://www.cbo.gov/publication/61270>
- [11] Dallas, C., & Scavone, T. (2024, November). Deconstructing U.S. debt dynamics: The current state of U.S. public finances (VantagePoint). Cambridge Associates. <https://www.cambridgeassociates.com/wp-content/uploads/2024/11/2024-11-VantagePoint-Deconstructing-US-Debt-Dynamics.pdf>
- [12] Dola, A., Begum, S., Antara, U. K., Islam, M. R., Sultana, T., & Zabin, N. (2024). Machine learning models for detecting hidden collusion networks in US corporate finance. *Journal of Economics, Finance and Accounting Studies*, 6(1), 143–154.
- [13] Edelberg, W., Harris, B., & Sheiner, L. (2025, February 12). Assessing the risks and costs of the rising U.S. federal debt (Hamilton Project Paper). Brookings Institution. https://www.brookings.edu/wp-content/uploads/2025/02/20240212_ES_EdelbergHarrisSheiner_Debt_final1.pdf
- [14] Escolano, J., & Shabunina, A. (2024). Public debt and welfare with machine learning. *Finance Research Letters*, 70, Article 105943. <https://doi.org/10.1016/j.frl.2024.105943>
- [15] Eyvazov, T. (2024). Costly increases in public debt when $r < g$ (IMF Working Paper WP/2024/010). International Monetary Fund. <https://www.imf.org/en/-/media/files/publications/wp/2024/english/wpica2024010-print-pdf.pdf>
- [16] Gauthier, N., & Moyen, S. (2025). Using DSGE and machine learning to forecast public debt for France (Working Paper 2025-18). Université de Strasbourg. <https://beta.u-strasbg.fr/WP/2025/2025-18.pdf>
- [17] Gauthier, N., & Moyen, S. (2026). Using DSGE and machine learning to forecast public debt for France. *Journal of Forecasting*. Advance online publication. <https://doi.org/10.1002/for.70144>
- [18] International Monetary Fund. (2021, February 3). Review of the debt sustainability framework for market access countries (Policy Paper). <https://www.imf.org/en/Publications/Policy-Papers/Issues/2021/02/03/Review-of-The-Debt-Sustainability-Framework-For-Market-Access>
- [19] International Monetary Fund & World Bank. (2018). The debt sustainability framework for low-income countries (Updated 2017). <https://www.imf.org/external/pubs/ft/dsa/lic.htm>
- [20] Islam, M. R., Pramanik, M. T., & Zeeshan, M. A. F. (2025). Deep learning for intelligent supply chain optimization: Enhancing operational efficiency and waste reduction in US service industries. *Frontiers in Computer Science and Artificial Intelligence*, 4(2), 45–62.
- [21] Islam, M. Z., Sumsuzoha, M., Islam, M. R., Kawsar, M., Mithu, M. F. H., Pant, S., ... & Al Helal, M. A. (n.d.). Graph neural networks for systemic financial risk forecasting: Modeling cross-market contagion between banking systems and cryptocurrency markets.
- [22] Jakir, T. (2025). Signal-to-noise analysis of crisis indicators in global finance using artificial intelligence. *International Journal of Applied Mathematics*, 38(10s), 1815–1836.
- [23] Krantz, J., & Richert, M. (2023). SHAPoly: A novel Shapley-polynomial framework for estimating nonlinear dynamics in macroeconomic data using deep neural networks. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.4350978>
- [24] Mah lik, J. M. (2026, January 6). The impact of public debt on economic growth: What the empirical literature tells us (Mercatus Policy Brief). Mercatus Center at George Mason University. <https://www.mercatus.org/research/policy-briefs/impact-public-debt-economic-growth-what-empirical-literature-tells-us>
- [25] Miah, M. N. I., Uddin, M. J., & Kakumani, M. (2026). Artificial intelligence in sentencing: Evaluating machine learning models for sentencing recommendations in the US. *Frontiers in Computer Science and Artificial Intelligence*, 5(4), 30–43.
- [26] Rahman, M. S. (2025). Machine learning-enabled early warning system for detecting micro-inflation clusters in the US economy. *International Journal of Applied Mathematics*, 38(12s), 2743–2769.
- [27] Reinsberg, B., & Stubbs, T. (2025). The r-g differential: A policy instrument for German federal states? *EuroGeoJournal*, 4(2), 1–20. <https://eurogeojournal.eu/index.php/egj/article/view/913>
- [28] Reza, S. A., et al. (2025). Machine learning enabled early warning system for financial distress using real-time digital signals. arXiv preprint, arXiv:2510.22287.
- [29] Shawon, R. E. R., et al. (2025). Detecting illicit cross-chain fund movement: Behavioral machine learning models for bridge-based laundering patterns. *International Journal of Applied Mathematics*, 38(12s).
- [30] Sultana, K. S., Begum, M., Abed, J., Siam, M. A., Sadnan, G. A., Shaty, S. S., & Billah, M. (2025). Blockchain-based green edge computing: Optimizing energy efficiency with decentralized AI frameworks. *Journal of Computer Science and Technology Studies*, 7(1), 386–408.