Approximate Dynamic Programming
Applied to Biofuel Markets in the Presence of Renewable Fuel Standards

Kevin Lin
Advisor: Professor Warren B. Powell

Submitted in partial fulfillment
of the requirements for the degree of
Bachelor of Science in Engineering
Department of Operations Research and Financial Engineering
Princeton University

April 2014
I hereby declare that I am the sole author of this thesis.

I authorize Princeton University to lend this thesis to other institutions or individuals for the purpose of scholarly research.

Kevin Lin

I further authorize Princeton University to reproduce this thesis by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

Kevin Lin
Abstract

Due to the increased attention on renewable fuel standards to mitigate the effect of greenhouse gas emissions from automobiles, the US government has started implementing new policies that require gasoline distributors to mix biofuel (such as corn-based ethanol or other similar fuels based on biomass) into gasoline. We focus on the problem of managing a biomass plant in the presence of renewable fuel standards in this thesis. We develop a model where we decide how much biomass to sell, how much biomass to produce and how much capacity to add. We solve this policy using a backward Markov Decision Process (bMDP) and Value Function Approximation (VFA) where we approximate the value functions in the latter policy using the Concave, Adaptive Value Estimator (CAVE). We show that under certain assumptions on the approximated value functions, the VFA, is a must faster algorithm and obtains a policy that achieves a profit that is close to that of the bMDP. We survey various machinery commonly utilized in approximate dynamic programming throughout this thesis.
Acknowledgements

This thesis could not have completed without the help of Professor Warren B. Powell, a professor who never seems to run out of enthusiasm and a depth of knowledge and foresight for interesting problems. I also thank Princeton University’s Operations Research and Financial Engineering Department for the wonderful four years I had. The topics covered throughout the undergraduate and graduate classes in this department have truly expanded my creativity and technical skills more than I could have ever imagined. The professors and classmates of this student have been more helpful than I can describe in words. Without them, I would not have been able to find my goal in pursuing a career in academia. Lastly, I thank my family who has always provided support when it matters the most. I am where I am today thanks to their guidance.
To my family and all those who ever told me to follow my dreams
Contents

Abstract . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . iii
Acknowledgements . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . iv
List of Tables . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . ix
List of Figures . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . x

1 Introduction and Background . . . . . . . . . . . . . . . . . . . . . . 1
1.1 US Energy Usage and Global Warming . . . . . . . . . . . . . . . . 1
    1.1.1 Energy Consumption . . . . . . . . . . . . . . . . . . . . . . 3
    1.1.2 Impact on Environment . . . . . . . . . . . . . . . . . . . . 4
    1.1.3 Historical Trends of Oil, Gasoline, and Ethanol Market . . . 6
1.2 US Energy Policy . . . . . . . . . . . . . . . . . . . . . . . . . . . . 7
    1.2.1 Energy Efficiency Standards . . . . . . . . . . . . . . . . . 10
    1.2.2 Carbon Tax and Permits . . . . . . . . . . . . . . . . . . . . 14
    1.2.3 Renewable Portfolio Standard and Renewable Fuel Standard . 16
    1.2.4 Tax Incentives and Subsidies . . . . . . . . . . . . . . . . . 18
1.3 Current Developments of RFS . . . . . . . . . . . . . . . . . . . . . 22
1.4 Our Problem . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 24

2 Model . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 26
2.1 Modeling Introduction . . . . . . . . . . . . . . . . . . . . . . . . . 26
    2.1.1 Five Components of a Model . . . . . . . . . . . . . . . . . 27
3 Algorithm: Backward Markov Decision Process

3.1 Value Function 
3.2 Algorithm Description 
3.3 Algorithm Verification for Model 1 
3.4 Algorithm Verification for Model 2 
3.5 Look-Up Tables and Curses of Dimensionality 
3.6 Closing Remarks

4 Algorithm: Value Function Approximation

4.1 Approximation Methods 
4.1.1 Aggregation Functions 
4.1.2 Post-Decision Value Functions as Gradients
List of Tables

3.1 Time Used to Solve bMDP Policy . . . . . . . . . . . . . . . . . . . 58
4.1 Summary Statistics for Three Policies . . . . . . . . . . . . . . . . . 80
4.2 Summary Statistics for Three Policies . . . . . . . . . . . . . . . . . 82
## List of Figures

1.1 An illustration of the greenhouse effect. ........................................... 2  
1.2 Fossil fuel $CO_2$ emissions based on data of Marland and Boden and  
    British Petroleum ............................................................................. 2  
1.3 Global mean surface temperature change based on surface air measure-  
    ments over land ................................................................................ 2  
1.4 Temperature changes in surface temperatures between the average in  
1.5 A map showing the plants operational in 2006 where the size of the  
    dot is proportional to the plant capacity (output in megawatts). ....... 3  
1.6 A map highlighting the areas along the US Atlantic and Gulf coasts  
    that are vulnerable to inundation resulting from sea-level rise. This  
    simple elevation map does not include additional future impacts from  
    erosion and local sinking of land. .................................................... 5  
1.7 Weekly US Retail Gasoline Prices by Dollars per Gallon. ................. 6  
1.8 US mid-range abatement curve - 2030. Carbon dioxide abatement:  
    estimated removal cost per ton of $CO_2$ in 2005 dollars and removal  
    potential in gigatons/yr for various strategies. ............................... 8  
1.9 Breakdown of energy use for residential and commercial buildings ... 10  
1.10 Emissions of greenhouse gases in the transportation sector. ........... 10
1.11 Diagram to illustrate how to lower the energy consumption of commercial buildings. ......................................................... 12
1.12 Plot showing the energy efficiency and fuel economy of various types of cars over the years. .................................................. 12
1.13 Average household refrigerator energy use, volume and price over time. ................................................................. 12
1.14 Figure depicting the contributing uncertainty of various independent variables ............................................................ 14
1.15 US ethanol capacity and production from January 2000 to November 2013 ................................................................. 17
1.16 Ethanol price, cost and Profits from January 2005 to November 2013 ................................................................. 17
1.17 Hybrid auto sales in the US. ................................................................. 19
1.18 The predicted emissions under different plans. “Moderate action” refers to changing fuel efficiency rates while “strong action” adds hydrogen-powered vehicles. ................................................................. 19

2.1 A chronological illustration on the order of events. ......................... 30
2.2 Mean demand difference $\hat{d}_t$ over time $t$. .................................. 35
2.3 Mean demand $D_t$ over time $t$ ......................................................... 35
2.4 An illustration to motivate Model 2. ................................................. 41

3.1 An algorithmic check for Model 1 but using a deterministic price process where the price is a rescaled value of the expected demand. .... 50
3.2 A bMDP algorithmic check for Model 1 using the correct pricing function. ......................................................... 52
3.3 An illustration of the difference in value measured by the pricing function. ................................................................. 54
3.4 A bMDP algorithmic check for Model 2. ................................................. 55
3.5 A bMDP algorithmic check for Model 2 over 96 months. .................. 56
3.6 A plot of the profit over time based on Figure 3.5. ......................... 57
3.7 A plot of price over time for different capacity levels based on the bMDP policy. ................................................. 57
3.8 A plot of the value as a function of $R^\text{bio}_t$ for different capacity levels for time $t = 65$ based on the bMDP policy. ......................... 58
3.9 A 3D plot of the value as a function of $S_t$ for time $t = 65$ based on the bMDP policy. ................................................. 59

4.1 An example of the assumed form of the value function. .............. 66
4.2 Three-step graphical explanation of the CAVE algorithm .............. 71
4.3 Initial setting of $\bar{V}_t^x$ for all $t$ .............................................. 73
4.4 A VFA algorithmic check for Model 2. ................................. 74
4.5 A VFA algorithmic check for Model 2 over 96 months. ................ 75
4.6 A plot of the profit over time based on Figure 4.5 ..................... 76
4.7 Post-decision value functions with respect to $R^\text{bio,x}_t$ for different months based on Figure 4.5 .............................. 76
4.8 Post-decision value functions with respect to $R^\text{cap,x}_t$ for different months based on Figure 4.5 .............................. 77
4.9 An example of the assumed form of the value function. .............. 78
4.10 An illustration of the stepsizes $\alpha^{(n)}(n)$ per iteration $n$. .................. 80
4.11 A comparison of cumulative profit using different stepsizes. .......... 81
4.12 A comparison of the three different policies. .......................... 82
4.13 A comparison of the three simulations. .................................. 85

xii
Chapter 1

Introduction and Background

We present an overview of global warming and US energy policy in this chapter. Specifically, we focus on techniques utilized in research in this field. We pay close attention to renewable energy policies related to fuel standards in the transportation sector as it will later relate to our problem of managing a biomass plant in the presence of a renewable fuel standard.

1.1 US Energy Usage and Global Warming

Climate change is an unprecedented challenged that our society faces. The Fourth Assessment Report (AR4) produced by the Intergovernment Panel on Climate Change (IPCC) in 2007 indicated a link between the increase of greenhouse gases (GHG) in the atmosphere and global warming. This is evident by the trends in warming of oceans, rising sea levels, and melting of Artic ice. Numerous polices have been made to control GHG emissions from the power and energy-intensive sectors. Global warming and climate change (two terms often used interchangeably) is caused by the “greenhouse effect” whereby certain pollutants accumulate in the ozone layer and trap the energy from the sun within the atmosphere. See Figure 1.1. The trend of global emission of $CO_2$, a common pollutant associated with the greenhouse effect, is shown graphically
in a report by Hansen Figure 1.2. While the rate of emissions has slowed down, the total amount of emissions still grows annually. In this section, we review the basic facts about US energy consumption, greenhouse gases, and historical trends in the oil, gas and ethanol market.

The effects of climate change are commonly associated with the gradual rise in ocean and surface temperatures. Hansen’s report goes on to display these changes in Figure 1.3 and Figure 1.4 respectively.
1.1.1 Energy Consumption

In 2007, the average US household spent $2400 on gasoline for their autos and $1900 on natural gas, electricity and heating oil. The US Bureau of Statistics stated that heating and cooling were the most energy-intensive applications in the household. This energy is supplied from a multitude of power sources scattered throughout the US Figure 1.5.

The power sector faces three major challenges: reforms of power markets to encourage competition, requirements to mitigate greenhouse gas emissions and rising energy prices. Risks and uncertainties often compel investment in flexible power production technologies with short periods of return on investments. Economics of scale require investors to develop larger power facilities to minimize the cost of unit production however. In the power sector, investments often have three important characteristics.

- Investments is partially or completely irreversible. Capital costs usually become
sunk.

- There is always uncertainty over future return from the investment. Future energy prices and carbon prices are always unpredictable, so the cash inflow of a project is always uncertain.

- Investors have a choice to invest at flexible timings. They have the opportunity or option but never obligation to invest in a project in a period of time. They also have the flexibility to abandon, expand, extend or contract the operation of the project after investment.

Many financial models have been developed to understand this tradeoff. Traditionally, companies have used pay-back period or discounted cash flow method. Pay-back period methods estimates the number of years it would take for the income of a particular project to pay back the costs of investment. However, these type of analyses are often flawed because of their finite time horizon biases against investments due to the long-term benefits of most projects. Discounted cash flow methods determine the present value of future cash flows by discounting them appropriately and naturally incorporates opportunity costs and uncertainty over time.

1.1.2 Impact on Environment

Before studies on different policies can be done to prevent climate change, initial studies quantifying the risks and uncertainties associated with the climate change itself must be made. Environmental regulations generally aim at targeting three major pollutants: carbon dioxide ($CO_2$), sulfur dioxide ($SO_2$) and nitrogen oxides ($NO_x$). The latter two pollutants are non-uniformly mixed pollutants meaning their marginal damage vary across location and time. These gradual buildup of greenhouse gases, if left unchecked, will result in rising sea-levels that will destroy valuable land in the US. A study by Wang and Chameides studied area in the US vulnerable to inundation resulting from sea-level rise (Wang and Chameides (2005)). They estimated that
Figure 1.6: A map highlighting the areas along the US Atlantic and Gulf coasts that are vulnerable to inundation resulting from sea-level rise. This simple elevation map does not include additional future impacts from erosion and local sinking of land.

22,400 square miles of land, more than twice the size of Maryland, lie within 5 feet of sea level, and that the sea level can rise by this amount within the next few centuries. This is shown in Figure 1.6.

Other studies on the impact of environment have met difficulties due to the subtle differences in each market. Many models quantify the effects of different environmental policies in the electricity market but disregard the transmission constraints that limit energy flow between subregions. This lack of detail mistakes the electricity market as a single market in each time period. The same necessity of spatial detail holds true for the renewable energy market. Many models assume renewable energy generated in different locations have the same environmental value, but in reality, spatial variation in renewable energy such as wind power is vital to understanding the impact.
Fuel prices have fluctuated from $2.86/gallon to $3.87/gallon in the last five years (Administration (2013)). This is shown in Figure 1.7. The volatility of fuel prices has increased interest to invest in alternative biofuel sources such as ethanol. In recent years, ethanol production capacity has increased 70% from its original value. Most ethanol is blended with gasoline at concentrations of 5 to 10 percent. The ethanol concentration is limited to this range since the current technology in most vehicles does not permit higher percentages of substitution of ethanol for gasoline. There have been some small developments in flexible fuel vehicles using E85, 85% ethanol and 15% gasoline, as the primary fuel though.

Historically, ethanol production has been subsidized by the government. This ensures that ethanol-gasoline blends offer a competitive price compared to traditional gasoline. The ethanol market has been striving to decrease the production and distribution costs. Estimates have shown that ethanol consumption can decrease gasoline prices by 89 cents per gallon which can translate to each American household saving $800 a year on gasoline. The US ethanol industry also sustains 400,677 jobs in the American economy.

Corn-based ethanol has faced many challenges in the US. Some of these issues relate to the fluctuations in the agriculture productions, increase of corn and animal
food prices, and threat to natural wildlife habitats. Historically, corn price volatility was caused primarily by shocks in supply, but the recent ethanol developments have caused shocks in the demand as well. Stochastic gasoline prices have caused stochastic demand in ethanol and uncertain policy changes have only added to its volatility. Cellulosic fuel ethanol have also been considered, but the commercialization of cellulosic is not fully developed yet. While corn ethanol plants have been developing for 30 years with more than 200 plants producing more than 14 billion gallons of corn-based biofuels, no successful commercialization of cellulosic ethanol has been reported yet. Cellulosic ethanol has a smaller carbon footprint than corn ethanol.

The US government has often set ambitious goals to become more energy independent and more ecologically friendly. In the 2011 State of the Union, US President Barack Obama stated, “With more research and incentives, we can break our dependence on oil and bio-fuels and be the first country to have a million electric vehicles on the road by 2015.”

### 1.2 US Energy Policy

Two societal responses to climate change are abatement and energy technology R&D. Abatement responses are direct and try to reduce greenhouse gas emissions. The questions related to this response are the determination of optimal path of emissions in future years, emissions allocations, or level of a carbon tax. The energy technology R&D response focuses on investing in research to make emission abatement less costly in the future. Policy makers who study R&D often talk about the short-term decisions that are made within the next fifty years and the long-term decisions that are to be made when more information about current uncertainties are available. A study by the American Physical Society listed the abatement methods that are unrelated to government polices and ranked them according to their environmental
impact (Physics (2008)). This is shown in Figure 1.8 where the negative horizontal bars refer to savings to the economy thanks to energy efficiency (such as improving residential and commercial electronics, integrating heat and power sources, etc.) while the positive horizontal bars refer to economic spending (such as reforestation, construction of better coal plants, car hybridization). It comes to no surprise that the latter strategies that reduce the amount of gigatons per year have a higher cost associated with them.

Two general categories of abatement policies for combating climate change are command-and-control and market-based instruments. The former focuses of technology and performance standards. Owners can decide which technologies to utilize and tailor the designs towards their particular plant or facility. Market-based instruments take advantage of the competition in polluting industries however. The price
instrument, a “tax”, acts like a cost adder internalizing pollution damage, an “externality.” One such example is the cap-and-trade policy where facilities are allowed a fixed amount of emission quantity via permits or allowances, and facilities need to demonstrate their compliance by “surrendering” sufficient allowances to cover their emissions at the end of each compliance cycle. These permits can be traded freely in the secondary market. In theory, the companies with low control costs would sell their permits while their counterparts would purchase surplus allowances.

There are many studies comparing these two categories of abatement policies. Economists often advocate market-based approaches on the ground of economic efficiency. A general consensus of the advantages of market-based approaches over command-and-control are as follows:

- Market-based instruments result in an equilibrium where the marginal abatement cost is equal across all polluters. More importantly, regulators do not need to know this cost to implement the policy.

- Cap-and-trade and tax instruments are more effective in incentivizing polluters to undertake research and development opportunities for exploring cost-saving abatement options.

- Revenues raised by tax or permits can be recycled. For example, it can be used to lower other taxes or mitigate distortions induced by previous taxes.

Developing new energy sources is critical for the US economy. In the following parts of this subsection, we will cover more specific examples of these policies. As we briefly tour the research that revolves around these policies, we take a note of the various transition functions, objective functions, and exogenous information processes within these models.
1.2.1 Energy Efficiency Standards

Many analysts claim that energy efficiency investments are “win-win” opportunities since it reduces the pollution level and also saves money. A study by McKinsey & Co. in 2009 showed that a holistic approach to energy efficiency can yield a gross savings worth of more than $1.2 trillion. This estimate is far better than the estimated $520 billion needed through 2020 for upfront investment in energy efficiency. Such holistic programs are estimated to reduce the end-energy consumption by 9.1 quadrillion BTUs equating to abating 1.1 gigatons of greenhouse gases annually. A study by Alliance showed that most of energy consumed by commercial and residential users is utilized for lighting and space heating (Alliance (2013)). This is shown in Figure 1.9. The American Physical Society showed that among the energy used in the transportation sector however, passenger cars contribution the most to the emissions (Physics (2008)). This is shown in Figure 1.10.

To reduce the emissions of these primary sources, the US has financially supported many efforts to research and develop new technologies. Many different studies have investigated and documented trends in energy saving. Figure 1.11 illustrates methods to reduce the energy consumed for space heating in commercial buildings while studies
by the American Physical Society and Cavanagh et al. illustrated the theme that both cars and refrigeration units are getting bigger and more energy efficient over the years (Physics (2008)),(Cavanagh and Martinez (2013)). These are shown in Figure 1.12 and Figure 1.13.

However, Allcott published a study that stated many industrial studies measuring the benefits energy efficiency investments tend to be flawed and that the profitable unexploited investment opportunities are much smaller than what the engineering-accounting studies suggest (Allcott (2012)). His analysis relied on two types of market failures that standard models fail to account for. These are uninternalized externalities from energy consumption and forces such as imperfect information that cause consumers and firm to not exploit privately-profitable energy efficiency investments. This “investment inefficiency” is often called the Energy Efficiency Gap, a wedge between the cost-minimizing level of energy efficiency and the level actually realized.

A common aggregate measure of energy efficiency is the ratio of GDP to total energy use with different energy sources. The US “energy productivity” per unit GDP is 2.4 times higher in 2012 than in 1949. Energy prices also influenced factor substitution and technical changes. Many studies have shown that the US is progressively becoming less energy intensive, but these studies show no indication of whether or not the US is near the economically efficient level of energy efficiency.

The US has historically implemented various policies to encourage energy efficiency. Some examples include the following.

- Corporate Average Fuel Economy (CAFE) which required new cars and trucks sold by each manufacturer to meet a minimum average rating based on the miles-per-gallon,

- “Weatherization,” the set of energy efficiency investments primarily for wall and attic insulation, due to the Weatherization Assistance Program and the 2009 American Recovery and Reinvestment Act,
Figure 1.11: Diagram to illustrate how to lower the energy consumption of commercial buildings.

Figure 1.12: Plot showing the energy efficiency and fuel economy of various types of cars over the years.

Figure 1.13: Average household refrigerator energy use, volume and price over time.
• Tax credits of up to $3400 for hybrid vehicle buyers,

• “Tax guzzler taxes” ranging from $1000 to $7000 on sale of passenger cars with low fuel economy,

• National-level energy efficiency standards for household appliances such as refrigerators, air conditioners, and washing machines,

• State-wide energy efficiency building codes such as stipulating the minimum amount of insulation.

Air conditioning equipment is used by more than 100 million homes each year, and accurate energy efficiency metrics are vital in order to have an effective domestic and international energy policy and a confident energy forecast. A study from the Lawrence Berkley National Laboratory measured the uncertainty in the energy efficiency ratio (EER) of split-system air conditioners (Yu (2013)). The study showed that increasing the measurement precision in barometric pressure and power have the largest impact in reducing the overall uncertainty of the EER. Figure 1.14 displays the propagation of uncertainty done in the analysis. Other types of analysis on energy efficiency involves energy technology research and development (R&D) portfolio management. While it is clear that rapid technological growth would limit climate change, new research studies are determining how exactly to allocate financial resources to R&D. One such study was made by Baker and Solak (Baker and Solak (2012)). They utilized probabilistic data derived from expert elicitations to form a stochastic dynamic program and conclude that the optimal technology portfolio for the set of projects is fairly robust to different specifications of climate uncertainty. They also find that it is better to over-invest in R&D in expectation. Informally speaking, their stochastic maximization problem takes the following form.

$$\max_{\text{short, long decisions}} \text{Welfare(Short decisions)} + \mathbb{E}_\Omega \text{Welfare(Long decisions)}$$
Figure 1.14: Figure depicting the contributing uncertainty of various independent variables constrained to Relationships between short and long term decisions and outcomes.

Their model can examined in detail in their paper, and it is an interesting idea that links short and long term decisions where the long decisions are a function of the short term decisions and the observed exogenous variables.

### 1.2.2 Carbon Tax and Permits

Tradable permits and carbon taxes are two market-based instruments commonly used by policymakers to regulate pollution. These two differ though since taxes are usually predetermined by authorities while the prices of permits fluctuate with the price of natural gas and electricity.

Many policies analyzing the effect of greenhouse gases talk about a “tipping point,” an event that will cause the climate system to transition into an alternative state where damages are abrupt and irreversible. These tipping points are high impact-low probability events. A study by Lontzek, Cai, and Judd used a stochastic integrated...
assess assessment model (IAM) that exhibit a ramping structure of climate policies (Lontzek et al. (2013)). The ramping structure, as well as its underlying implication of discounting climate change impacts, occurs because damages rises gradually with global warming. Their study incorporated a stochastic and abrupt tipping point event in the climate system and computed that the optimal additional carbon tax of $5 per ton of CO$_2$ emissions would delay the event in expectation by five years. Their paper uses a Markov process described by

\[ M_{t+1} = \Phi^M M_t + (\epsilon_t, 0, 0) \]

\[ T_{t+1} = \Phi^T T_t + (\xi F_t(M_{t}^{AT}), 0) \]

\[ F(M^{AT}) = \eta \log_2(M^{AT}/M_0^{AT}) + F_t^{EX} \]

where the first equation models evolution of \( M_t = (M_t^{AT}, M_t^{UP}, M_t^{LO}) \), the vector describing the masses of carbon concentration in the atmosphere and upper and lower strata of the ocean for some noise \( \epsilon_t \) based on a transition function \( \Phi^M \). The second equation models the evolution of \( T_t = (T_t^{AT}, T_t^{LO}) \), the vector of temperatures in the atmosphere and ocean, and \( \Phi_t \) represents the diffusion process between the ocean and air. \( \xi_1 \) is a sensitivity parameter to control \( F(M^{AT}) \), the total radiative force affected by the external force \( F_t^{EX} \). This is all built into a dynamic program that minimizes the hazard rate based on the tipping point process.

A study by Chen and Tseng studied the optimal investment policy when a coal-fired plant owner uses clean technologies under an imposed carbon tax or tradable permits setting using a real options analysis (Chen and Tseng (2011)). Their study showed that tradable permits can effectively trigger owners to adopt clean technologies at a much lower carbon price relative to the tax policy. This is because the higher volatility of the permit prices are more likely to encourage suppliers to take early action to hedge their carbon risk. The model assumes that three sources of uncertainty
(price of electricity $X_t$, natural gas $G_t$ and emission permits $Y_t$) all evolve according to correlated geometric Brownian motion, and the coal plant owners are maximizing their profit per MW by

$$\max_{0 \leq \alpha_t \leq \alpha} (1 - \alpha_t)(X_t - C - Y_t) + \alpha_t(X_t - G_tH - 0.5Y_t)$$ (1.1)

where $C$ is a constant cost for firing coal plants. The first term represents using fraction of MW produced from coal plants and the second term represents using clean technology based on natural gas which has half the emission cost.

### 1.2.3 Renewable Portfolio Standard and Renewable Fuel Standard

The US has historically used a federal volumetric ethanol production subsidy, but this was reduced by the 2008 Farm Bill and later removed in 2011. This has hindered the growth of the industry compared to historical growth. The falling crude oil prices in 2008 and economic downturn in 2011 drastically cut the profit margins of many ethanol farms. For example, VeraSun Energy, the largest US ethanol producer at the time, filed bankruptcy in 2008. In 2009, ethanol production capacity expansion averaged less than 1% per month compared to the 4.6% growth between 2008 and the end of 2008. The renewable fuels standard (RFS) was implemented in 2005 and set a goal of 7.6 million gallons of renewable fuels by 2012. It was later extended in 2008 to set a target for 2022. A study done by O’Brien illustrated that the ethanol market, while growing, operates on rather thin margins (O’Brien (2013)). Figure 1.15 shows that while annual production of ethanol grows yearly, the rate of expansion has slowed down due to political setbacks. Figure 1.16 shows the profit margins of ethanol plants where the red line represents the revenue and the red line represents the cost. Their difference is shown as green bars on the bottom (the net profit). Over
the years, the profit margins of ethanol are seen to get smaller. This causes a lack of incentive for owners to start new ethanol plants.

Much analysis revolving around ethanol policies cast the problem within the real options analysis (ROA) framework. This framework searches for the optimal investment behavior and has a clear method to balance the volatility and cost components as well as factoring in the correlation between them.

A paper by Schmidt, Luo, and Conrad showed that within their real options framework, US’s ethanol policy has narrowed the distance between the optimal entry and exit curves (Schmidt et al. (2011)). This shows that there is a more narrow range of inactivity, and this promotes a more volatile evolution for the industry. This research extended the ROA model by incorporating multiple sources of stochastic behavior. Their particular model accounts for variability in the unit revenues and costs. Functional approximation procedures were used to determine the optimal switching conditions to enter or exit the market, and their calculations showed that under equilibrium, the value function for keeping a project active and shutting down are

\[ V_0(x, y) = By^{\beta_2}x^{1-\beta_2} \]  
\[ V_1(x, y) = Ay^{\beta_1}x^{1-\beta_1} + \frac{y}{\delta - \mu_y} - \frac{x}{\delta - \mu_x} \]
where the important variables in question are $x$ (unit cost) and $y$ (unit revenues), both stochastic.

More stochastic models have been published of this sort. One study by McPhail and Babock incorporated five sources of stochasticity into their model for corn prices and showed that the Energy Independence and Security Act (EISA) of 2007 would increase the price of corn by 7.1% and the volatility by 12.1% at the time of their research. Their stochastic elements were

\[
\begin{align*}
\tilde{y}_{2008} & \sim \text{beta}(\bar{y}, \sigma_y^2, p_y, q_y) \\
\tilde{A}_{2008} & \sim \text{beta}(\bar{A}, \sigma_A^2, p_A, q_A) \\
Q_{D,\text{export}}^{2008} & \sim c_1(1 + .1\epsilon_{D,\text{export}}^{2008}) - c_2P_{2008} \\
\tilde{E}_{2008} & \sim \text{beta}(\bar{E}, \sigma_E^2, p_E, q_E) \\
\tilde{\gamma}_{2008} & \sim \text{beta}(\mu_\gamma, \sigma_\gamma^2, p_\gamma, q_\gamma)
\end{align*}
\]

for the distribution of the US corn yield in 2008, planted acreage for corn, export demand curve of corn (as a function of $P_{2008}$, the price of corn, and $\epsilon_{D,\text{export}}^{2008}$, a demand shock), capacity of ethanol production, and the gallons of ethanol produced per bushel of corn. These modified beta distributions with shape and size as the third and fourth parameters are multiplied by the second parameter and shifted over by the first parameter.

1.2.4 Tax Incentives and Subsidies

Fiscal incentives are common to encourage households to undertake energy-efficient renovations or adopt renewable energies. In the height of the economic recession, the US government passed the American Recovery and Reinvestment Act of 2009 which granted tax credit for consumers who purchased electric vehicles. This act boosted many car companies to launch their own versions of electric cars. However,
many of these companies often ran out of products when first released due to demand uncertainty. The sales of these hybrid cars have actually spiked over the last few years shown in Figure 1.17 (Physics (2008)). A study by Hansen predicted the various outcomes of “moderate” and “strong” policy incentives for advocating more electric cars and higher fuel efficiency as seen in Figure 1.18 (Hansen (2006)).

Many studies regarding tax incentives discuss the “free-riding effect” where households that would have made energy-efficiency investments regardless of whether or not there was a tax credit. The study by Dauphine Universte Paris showed that the tax credit was not effective in influencing the decision to renovate, and the free-riding effect reduces the actual effect of the tax credit (Geoffron (2014)).

Another form of incentives is the carbon market. Some studies have considered how to boost the cellulosic ethanol market, a close competitor of the corn-based ethanol market. A study by Luo and Miller showed that using game theory and the
assumption of farmers forming an oligopoly, a carbon market would only encourage
the cellulosic ethanol market when the price of switchgrass is extremely high (Luo and
Miller (2013)). Their model included farmers each having a decision $x$ to utilize part
of their total land by producing either switchgrass ethanol or corn ethanol. There were
also corn ethanol producers and switchgrass ethanol producers. The three respective
expected profit functions were

$$I = (P_2 - C_2 + s_2)xR_2 + (P_1 - C_1)(1 - x)R_1$$

$$o_3 = y_1(1 - x)R_1R_3(P_3 - C_3)$$

$$o_4 = y_2xR_2[R_4(P_3 - C_3 + s_4) - P_2]$$

where the first equation is for the $i$th farmer who receives a profit of $(P_2 - C_2)R_2$ as
well as a subsidy $s_2$ for planting $x_i$ percentage of switchgrass and $(P_1 - C_1)R_1$ for
plant $1 - x_i$ percentage of corn. The second equation is the corn ethanol producer’s
profit function where they buy $y_1$ percentage of the total corn available who receive a
profit of $R_1R_3(P - 5 - C_3)$ per unit of corn ethanol produce, and the third equation
follows a similar logic but with a subsidy of $s_4$. The equilibrium conditions under
varying amounts of subsidies was studied.

One study from the University of California developed a model that accounts
for the dynamic implications of government subsidies and RFS policies (Yi et al.
(2013)). Their study captures the entire cost structure of the industry including the
distributions of fixed entry costs and exit scrap values. In each type of subsidy they
studied, the inclusion of RFS was critical to sustaining the corn-based fuel ethanol
production. Investment subsidies and entry subsidies were shown to be more effect
than production subsidies. Their research formulated the problem as a Markov perfect
equilibrium and assumed each plant optimizes its behavior conditional on the current
state variable. Their model assumes that for a state variable $X_{it}$ for agent $i$ at time
A latent investment variable is calculated by

\[ s^*_t = X^\top_t\xi + u_{it} \]

for Gaussian noise \( u_{it} \).

Cohen, Lobel and Perakis took a different approach to analyzing the effectiveness of subsidies by extending the newsvendor problem to be a two-player problem (Cohen et al. (2014)). Subsidies were shown to encourage cooperation, and their model quantified how demand uncertainty would affect the objective functions of various individuals (government, industry and consumers) while showed the importance of incorporating demand uncertainty into future models in order to accurately represent the probability of meeting desired adoption target levels of policies. In this model, the government imposes a subsidy level \( r \) per product and the suppliers set a price \( p \) and production quantity \( q \) to maximize their profit. The suppliers then maximize

\[ \Pi = \max_{p,q \in \mathbb{R}^+} p\mathbb{E}[\min(q, D(p - r, \epsilon))] - cq \]

for a demand function \( D(p - r, \epsilon) \) where \( p - r \) represents the price consumers face and \( \epsilon \) represents the uncertainty. The government sets \( r \) in order to satisfy meeting the desired adoption target level \( \Gamma \) by at least probability \( \Delta \), or

\[ \mathbb{P}(\min(q, D(p - r, \epsilon)) \geq \Gamma) \geq \Delta. \]

The government is aiming to solve their own optimization problem to minimize expenditures,

\[ E = r\mathbb{E}[\min(q, D(p - r, \epsilon))]. \]

Wibulproprasert approached the analysis of subsidies in a different way by studying the long-run welfare benefits of optimizing designs of the existing renewable en-
nergy subsidy to coordinate efficient patterns of renewable investment in the presence of heterogeneity of offset emissions (Wibulpolprasert (2013)). The market and environmental values exhibit both spatial and temporal heterogeneity. This is due to constraints imposed by the existing thermal generation mix, available renewable resources, the existing transmission grid and demand locations. In the presence of heterogeneity, the effectiveness of renewable energy subsidies depends on their ability to induce quantities and locations of renewable capacity investment that maximize social value. Thus, theoretically, the subsidies rates should be differentiated across time and location and equal to the environmental value of renewable energy generation. This study improves upon traditionally used “uniform subsidies” to become “optimal variable subsidies.” In her model, the author defines various optimization problems for each system operator and potential wind entrant, and the measure of the effectiveness of the subsidy is defined by the following equation for welfare $W$

$$W(K) = \sum_{y,d,h} \int_0^{Q_{ydh}(K)} \left[ \frac{P(q) - m(q) - ce(q)}{(1 + r)^y} - C^\omega \right] dq - C^\omega K$$

where the demand, incumbent’s aggregate marginal cost and aggregate emissions are modeled by $P(q)$, $m(q)$ and $e(q)$ respectively. Here, the summation is over the years, days, and hours $(y,d,h)$ in question and the integration is all values up to $Q_{ydh}(K)$, the hourly supply given the total energy capacity $K$. The $C^\omega$ term refers to the marginal cost per unit capacity.

### 1.3 Current Developments of RFS

The U.S. Congress has recently met multiple times this year to discuss different approaches to implement the Renewable Fuel Standard (RFS). The RFS bills that Congress plan to vote on will give fuel industries two options: 1) to continue producing gasoline with 10% ethanol content and incur substantial penalties, resulting in less
gasoline produced, or 2) produce higher-ethanol blends of gasoline and potentially risk the integrity of millions of cars and large portion’s of the country’s fuel infrastructure. This bill is made to encourage the US economy to move towards renewable energy sources such as biofuels, but a careful assessment of the biofuel market must be made. Failure to correctly anticipate how the biofuel market would react to such a stimulus provided by the new RFS bills would result in higher gasoline prices at the pump without any benefits.

The biofuel markets currently is already estimated to double in size over the next decade. A report from Pike Research states the market will expanded from $82.7 billion in 2011 to $185.3 billion in 2021. This prediction is a result of higher oil prices, emerging mandates, new feedstock availability, and advanced technologies. At least 38 countries around the world have blending mandates. But even with this rapid growth, the predicted industry size will not meet the total market needs in the future. Higher blending requirements for the immense global transportation fuel market will generate much interest in developing the biofuel market, but current technology makes it questionable if this business venture will be profitable. Much of the biofuel market in the U.S. is currently backed by the RFS program created under the Energy Policy Act of 2005. The original program required 7.5 billion gallons of renewable fuel (typically biofuel) to be blended into gasoline by 2012. It was then expanded by the Energy Independence and Security Act of 2007.

Many senators currently wish to repeal RFS or at least lower its stringent mandates. In January 2014, Senators Ben Cardin (D-Md), Susan Collins (R-Maine), Dianne Feinstein (D-Calif), Bill Nelson (D-Fla) and six other senators opening proposed to cut back 1.5 billion gallons to biofuels to be blended into the gasoline. These senators claim that higher blending requirements will push the U.S. to depend more on foreign oil and to destroy American jobs. However, many spokespeople for independent biofuel companies such as Sue Ellspermann, Indiana Lt. Gov., do not wish
these cutbacks to take place.

1.4 Our Problem

We have surveyed various research techniques used to analyze different solutions to confront the growing environmental problem. Most of the models we described previously had an underlying model where the researchers tried to find the optimal decisions within this method for their specific definition of optimality or to find an equilibrium among all the players. Many of these models were stochastic in nature to reflect reality. We utilize these ideas to build our own model motivated by the increasing concern for renewable fuel standards (RFS) that require some percentage biofuels (such as corn-based ethanol) to be mixed into gasoline. Instead of talking about finding optimal decisions, we speak more broadly by finding optimal policies, where a policy is a method for making decisions.

We model the biofuel market as a stochastic process embedded with decisions at each time step. We then apply approximate dynamical programming techniques to solve this stochastic optimization problem. We investigate the performance of our approximate dynamical programming algorithms by comparing its performance against other methods that are known to give optimal solutions in specialized settings.

In this thesis, we present multiple models and algorithms. We separate the two to avoid any confusion. We follow Box’s Loop to iteratively make a model, solve it using an algorithm, evaluate the performance, and repeat (Box and Hunter (1972)). Our goal at the end to construct a complicated model-algorithm pair that can both reflect the actual biofuel market and produce solutions that are within some percentage of optimality. This thesis is focused on developing the machinery needed to study this model and studying the algorithms utilized the solve the model. We leave calibration of the model settings for potential future work.
In Chapter 2, we lay down the foundational notation system of our model and fill in the details. In Chapter 3, we describe the backward Markov Decision Process (bMDP), an algorithm used to solve for the optimal policy in this model. In Chapter 4, we describe the Value Function Approximation (VFA) using Concave, Adaptive Value Estimators (CAVE), a different policy that is faster in terms of computation but only finds an “approximately optimal” policy. We then perform comparisons between bMDP and VFA in different parameter settings. In Chapter 5, we conclude our thesis.
Chapter 2

Model

We now develop the model that we will inspect for the rest of the thesis in this chapter. We first define the five foundational components to modeling and then we utilize this framework to describe our own model. In this thesis, we focus on an inventory problem of managing a biofuel facility. The user is part of a competitive market, and he has to decide when to expand his facility, how much biofuel to sell, and how much biofuel to produce. He also needs to come up with a reasonable price to sell his biofuel at. We first define the foundational framework for modelling and then present two models in this chapter. In the next chapter, we will utilize an algorithm to solve this model.

2.1 Modeling Introduction

In this section, we lay down the five foundational elements of a model: a state variable $S_t$, a decision variable $x_t$, an exogenous information process $W_t$, a transition function $S^M$, and an objective function $\zeta^\pi$. We describe the importance of each element, and in the following sections, we fill in the details as we develop various models.
2.1.1 Five Components of a Model

State Variable: A state variable $S_t$ is the minimally-dimensioned variable that contains all the necessary and sufficient components to fully specify the model at time $t$. In particular, the state variable should contain all the variables needed to calculate the decision function, the transition function, and the objective function. While this explanation of a state variable might seem overly-complicated, the motivation is simple: we put all the “important” variables we need to we simulate the model into the state variable. For example, we do not want to put temperature of Mars into our model unless that variable is relevant.

We make a note that the state variable should make the system Markovian (history independent). While the system might not be Markovian under a underspecified state variable (which is usually a state variable that includes only physical quantities such as in a G/G/1 queue), a fully specified state variable (which includes elements that might not necessary be physical quantities) should make it Markovian.

Decision Variable: A decision variable $x_t$ is the user-controlled variable that is calculated at time $t$ based on the state variable $S_t$. Mathematically, we write

$$x_t = X_t^\pi(S_t)$$

where the function $X_t^\pi$ relates the state $S_t$ to the decision $x_t$. This function is the policy and is indexed by $\pi$ which states what kind policy is being used and what are the settings of the parameters needed to compute the function.

Exogenous Information Process: A exogenous information process $W_t$ contains all the variables that cannot be controlled by the user and is revealed at time $t$. Loosely speaking, $W_t$ shows how all the randomness arises within the model. Mathematically, let $\Omega$ be the set of all possible outcomes $W_1 \ldots W_T$. Let $\mathcal{F}$ be the
sigma-algebra on $\Omega$, and define the sub-sigma algebra $F_t$ using

$$F_t = \sigma(W_1 \ldots W_t)$$

where $F_t \subseteq F_{t+1}$ is a filtration. Finally, let $\mathcal{P}$ be the probability measure on $(\Omega, \mathcal{F})$, giving us the probability space $(\Omega, \mathcal{F}, \mathcal{P})$. Throughout our presentation, we assume that any variable indexed by $t$ is $F_t$-measurable.

Since $W_t$ comes from a probability space $(\Omega, F, \mathcal{P})$, we now define the expectation $E$. We write

$$E[\cdot] = \int_{\Omega} \partial W_1 \ldots \partial W_T$$

where we integrate over the entire set of outcomes. If we want to take a conditional expectation, we write either

$$E[\cdot | F_t] = \int_{\Omega} \partial W_1 \ldots \partial W_T | F_t$$

where the conditional on the filtration $F_t$ denotes a fixed realization of $W_1 \ldots W_t$. We use $E_t[\cdot]$ as a convenient shorthand to writing $E[\cdot | F_t]$.

Note that calculating $E_t[\cdot]$ still requires integrating over the outcomes for $W_{t+1} \ldots W_T$. But by the design of our model (which will be clear in the following chapters as we describe the value function), usually our argument only involves terms that involve $W_{t+1}$.

We tend to put a hat on all the variables part of the exogenous information to remind ourselves that the variable in question is an exogenous variable only revealed at its indexed time.

**Transition Function**: A transition function $S^M$ is a function that relates one time period to the next. Mathematically, we write

$$S_{t+1} = S^M(S_t, x_t, W_{t+1}).$$
Usually, if \( S_t \) contains \( d \) dimensions, the transition function contains \( d \) equations, one for each dimension of \( S_{t+1} \).

We also introduce another very useful concept. A \textbf{post-decision state variable} \( S_t^x \) is a state variable but accounts for the decision \( x_t \) but not the exogenous information \( W_{t+1} \). The dimensionality of the post-decision state variable might not necessarily be the same as the state variable as some of the dimensions containing information on how to make a decision might not needed to transition into the next state. Usually the dimensionality of \( S_t^x \) is smaller than that of \( S_t \), which is why sometimes we choose to use \( S_t^x \). If we choose to utilize post-decision states, then split our transition function into two separate transition functions

\[
S_t^x = S^M(S_t, x_t) \\
S_{t+1} = S^{M,x}(S^x_t, W_{t+1})
\]

where we choose to overload our notation for \( S^M \).

**Objective Function:** A \textbf{objective function} \( \zeta^\pi \) is a function that determines the quality of the decisions \( x_t \)'s, and in turn, the quality of the policy \( X_t^\pi \). This is only used after time \( T \). If our model contains model-parameters, we typically speak of one policy being better than another for a specific set of model-parameters. It is useful to think of calculating the objective function as a computer simulation that runs from \( t = 1 \) to \( t = T \).

Within the calculation of the objective function, usually a term called the \textbf{contribution function} \( C(S_t, X_t^\pi(S_t)) \) is needed. This reflects the instantaneous contribution (such as profit or loss) when decision \( X_t^\pi(S_t) \) is made at state \( S_t \).
2.1.2 Notation and Chronological Order of Events

Now that we have described the five parts of a model, we now see how each portion relates to one another. This is shown visually in Figure 2.1. We are always working in discretized time in our model. We note that all variables indexed by time $t$ are revealed at time $t$. This is consistent with our notation for filtrations mentioned above. Thus, all the exogenous information $W_t$ at time $t$ becomes part of $S_t$, and right as the time interval ends, we make a decision $x_t$. The post-decision state $S_t^x$ is calculated immediately after the decision $x_t$ is made. In this thesis, when we say “time $t$”, we are actually referring to the time interval $t-1$ to $t$.

2.2 Model 1

Our model takes on the perspective of a manager of a biofuel ethanol plant that produces biofuel every month to satisfy the renewable fuel standards (RFS) of gasoline distributions. The manager receives a notice of how much biofuel the gasoline distributors are demanding for each month, and the manager must determine how much biofuel to produce and how much biofuel to store. In addition, the manager also has the decision of determining if he should expand the production facility. The manager also needs to determine a price to charge for his biofuel. We use a market-based
policy where the price of the biofuel is determined by the competition. The model we develop in this section will capture all these ideas.

In our model, we will describe biofuel plants as a single entity but we can choose to expand or shrink its capacity. We do not buy “more plants.” Time $t$ goes from 1 to $T$ and is incremented at 1 at every time step. Here, $t$ represents a month.

2.2.1 Assumptions

Our model follows these assumptions:

- Plants do not decay in productivity over time and are instantly built.

- When plant capacity is built, it are initially turned off (do not produce biofuel). If the plant capacity is shrunk, the amount stored is not lost until the next month.

- The government regulation does not change throughout $t \in [1, T]$.

- Biofuel does not decay. That is, biofuel can be stored indefinitely.

- The biofuel market is competitive. That is, we charge a price according to how much we perceive the future value of storing biofuel is. However, biofuel is always bought at the price we set as long as it is lower than the government regulation price.

- The distributors would never buy biofuel if no regulation took effect.

- The distributors must satisfy total demand for gasoline at each time step.

2.2.2 Parameters

- $\rho$ - The fixed amount of gasoline that distributors are allowed to distribute before they need either pay a penalty for every unit of gasoline sold or buy
biofuel from the biofuel market.

- $\gamma$ - The fixed government price per unit of gasoline that distributors need to pay for each unit of gasoline sold above $\rho$

- $c^{\text{rat}}$ - The ratio between production rate and storage amount for capacity of a plant. We call this the production capacity and storage capacity respectively.

- $c^{\text{salv}}$ - The salvaged percentage of selling capacity.

- $c^{\text{cap}}$ - The fixed cost to build one unit of capacity for a plant

- $c^{\text{pro}}$ - The fixed cost to produce one unit of biofuel

- $c^{\text{bio}}$ - The fixed cost to store one unit of biofuel

For this model to make sense, we insist that $c^{\text{cap}} \gg c^{\text{pro}} > c^{\text{bio}}$.

### 2.2.3 State Variable

The state $S_t$ at time $t$ consists of

- $R_t^{\text{bio}}$ - The amount of biofuel available in storage

- $R_t^{\text{cap}}$ - The amount of production capacity available

- $D_t$ - The amount of gasoline demanded by consumers

All these variables refer to the values in the state variable at the time $t$ before the decision at time $t$ is made. To make this model more concrete, we define plant capacity as a term that characterizes both how much biofuel can be produced and how much biofuel can be stored. This will simply our state decision by one dimension. This is explained in detail with the decision variable.

Similarly, we define a post-decision state $S^x_t$ to reflect the state variable immediately after the decision $x_t$ is made but before the exogenous information of the next time period $W_{t+1}$ is revealed. Here, the post-decisions state $S^x_t$ at time $t$ consists of
• \( R_{t}^{\text{bio},x} \) - The amount of biofuel available in storage after \( x_t \) is made

• \( R_{t}^{\text{cap},x} \) - The amount of plant capacity available after \( x_t \) is made

• \( D_t \) - The amount of gasoline demanded by consumers

We explain how to calculate \( S_t^x \) from \( S_t \) and \( x_t \) in the next subsection. When we talk about the “state variable,” we are referring to the pre-decision state variable. We explicitly write “post-decision state variable” to talk about the post-decision state variable.

### 2.2.4 Decision Variable

When \( t \) represents the start a new year (that is, when \( \text{mod}(t-1, 12) = 0 \)), the decision \( x_t \) at the end of time interval \( t \) consists of

• \( x_t^{\text{cap}} \) - The amount of plants capacity to build (sell)

• \( x_t^{\text{pro}} \) - The amount of plant capacity utilized

• \( x_t^{\text{bio}} \) - The amount of biofuel sold

The decision is made at the end of the time \( t \). A negative value of \( x_t^{\text{cap}} \) indicates shrinking capacity while a positive value indicates building capacity.

We let \( \mathcal{X}_t \) denote the set of possible decisions \( x_t \) at time interval \( t \). As we will see, \( \mathcal{X}_t \) depends on the state variable \( S_t \), but we write \( \mathcal{X}_t \) instead of \( \mathcal{X}_t(S_t) \) for notational convenience. The following are the constraints imposed on \( x_t \)

\[
\begin{align*}
x_t^{\text{cap}} &\geq -R_t^{\text{cap}} \\
x_t^{\text{bio}} &\leq R_t^{\text{bio}} + x_t^{\text{pro}} \\
x_t^{\text{bio}} &\leq (D_t - \rho)^+ \\
x_t^{\text{pro}} &\leq R_t^{\text{cap}}
\end{align*}
\]
Here, \((\cdot)^+ = \max(0, \cdot)\). The constraint \(x_t^{\text{cap}} \geq -R_t^{\text{cap}}\) means we cannot sell off more capacity than we currently hold. The second constraint, \(x_t^{\text{bio}} \leq R_t^{\text{bio}} + x_t^{\text{pro}}\), indicates that the amount of biofuel we sell must be less than the sum of amount of biofuel we have stored and the amount we chose to produce for this time \(t\) while the third constraint, \(x_t^{\text{bio}} \leq (D_t - \rho)^+\), indicates that we cannot sell more biofuel than total gasoline demanded minus the amount allowed by the government regulation. The next two constraints reflect what our definition of capacity explained in the subsection above. The constraint \(x_t^{\text{pro}} \leq R_t^{\text{cap}}\) is a limitation on the amount we can produce since capacity reflects the rate we can produce biofuel. The next constraint \(x_t^{\text{pro}} + R_t^{\text{bio}} \leq c^{\text{rat}} R_t^{\text{cap}}\) is a limitation on the amount we can store since our capacity also reflects the amount of biofuel we can store.

When \(t\) does not represent the start of a year (that is, \(\mod(t - 1, 12) \neq 0\)), the decision only consists of \(x_t = (x_t^{\text{pro}}, x_t^{\text{bio}})\). We think of this as \(x_t^{\text{cap}} = 0\).

Given a current state \(S_t\), we make a decision based on a policy \(\pi\) which dictates the function \(X_t^\pi\) where

\[
x_t = X_t^\pi(S_t, W_t)
\]

We explain how to calculate \(X_t^\pi\) in the next chapter.

2.2.5 Exogenous Information Process

The exogenous information \(W_t\) that is revealed right before \(t\) consists of \(\hat{d}_t\), the difference in demand for gasoline between at time \(t\) and \(t - 1\). That is, \(\hat{d}_{t+1} = D_{t+1} - D_t\).
In our model, we sample values of $W_t$ based on

$$\hat{d}_{t+1} = a_1 [(\cos(a_2(t + 1))) - (\cos(a_2 t))] + N(0, a_4)$$

where $a_1$ is a parameter specifying the amplitude of the cycle, $a_2$ is a parameter specifying the frequency of the cycle, and $a_4$ is the variance for the Gaussian noise.

Note that if we were add $\hat{d}_t$'s together, we can recover the demand per month $t$. We include an additional parameter $a_3$ to represent the constant amount of demand. That is, $a_3 = D_1$. Doing so, we get

$$D_t = D_1 + \sum_{t'=1}^{t} \hat{d}_{t'} = a_1 (\cos a_2 t) + a_3 + N(0, ta_4)$$

for $t > 1$. Both the values of $\hat{d}_t$ and $D_t$ over time $t$ is shown below.

### 2.2.6 Transition Function

The transition function $S_{t+1} = S^M(S_t, x_t, W_{t+1})$ links the state at time $t$ to $t + 1$. It utilizes the following functions

$$R_{t+1}^{\text{bio}} = R_t^{\text{bio}} - x_t^{\text{bio}} + x_t^{\text{pro}}$$
\[ R_{t+1}^{\text{cap}} = R_t^{\text{cap}} + x_t^{\text{cap}} \]
\[ D_{t+1} = D_t + \hat{d}_{t+1}. \]

Since we restricted the set of possible values for \( x_t^{\text{pro}} \) in \( \mathcal{X}_t \) by \( x_t^{\text{pro}} + R_t^{\text{bio}} \leq 2R_t^{\text{cap}} \) and \( x_t^{\text{bio}} \geq 0 \), we do not need to worry about having too much biofuel in our transition function.

We overload the notation of \( S^M \) to include transitions from state to post-decision state variables and from post-decision state variables to state variables. It is obvious how to calculate \( S_t^x \) from \( S_t \) and \( x_t \). That is

\[ S^M(S_t, x_t) = (R_t^{\text{bio}} - x_t^{\text{bio}} + x_t^{\text{pro}}, R_t^{\text{cap}} + x_t^{\text{cap}}, D_t) = (R_t^{\text{bio,x}}, R_t^{\text{cap,x}}, D_t) = S_t^x \] (2.1)

Likewise, to transition from \( S_t^x \) to \( S_{t+1} \), we write

\[ S_{t+1} = S^M(S_t^x, W_t) = (R_{t+1}^{\text{bio,x}}, R_{t+1}^{\text{cap,x}}, D_t + \hat{d}_{t+1}) = (R_{t+1}^{\text{bio}}, R_{t+1}^{\text{cap}}, D_{t+1}) = S_{t+1}. \] (2.2)

### 2.2.7 Objective Function

We take the perspective of the biofuel producers. At the end of each time interval \( t \) (after the decision is made), we calculate the contribution function \( C \) where

\[ C(S_t, x_t) = p_t x_t^{\text{bio}} - \mathbb{I}[x_t^{\text{cap}} \geq 0]c_{\text{cap}} x_t^{\text{cap}} + \mathbb{I}[x_t^{\text{cap}} < 0]c_{\text{salv}} c_{\text{cap}} |x_t^{\text{cap}}| \] (2.3)
\[ - c_{\text{pro}} x_t^{\text{pro}} - c_{\text{bio}} (R_t^{\text{bio}} + x_t^{\text{pro}} - x_t^{\text{bio}}) \] (2.4)

where \( p_t \) denotes the price for biofuel and \( \mathbb{I}[x] \) is an indicator variable with value 1 if \( x \) is satisfied, 0 if not.

The price \( p_t \) at time interval \( t \) is determined by the numerical scalar derivative of the value function \( V_{t+1}(S_{t+1}) \) at the amount of biofuel remaining \( R_t^{\text{bio,x}} \) after the
decision is made. The value function is explained in detail in the next chapter. More formally, the price is a function of the current state $S_t$ and decision $x_t$. Then we write

$$p_t = p_t(S_t', x_t') = \frac{\partial \mathbb{E}_t[V_{t+\delta}(S_{t+\delta}^M(S_t, x_t, W_t))]}{\partial R_t^{\text{bio}}}|_{S_t' = S_t, x_t' = x_t}$$

(2.5)

and for this model to be nontrivial, we require $\gamma \geq p_t \forall t$.

The value function $V_t(S_t)$ can be set to anything we want, but in order to follow the assumption that the biofuel market is competitive and to capture a reasonable sense of optimality, we mandate that $V_t$ captures the expected future value of storing biofuel. This way, the relationship between $p_t$ and $V_t$ captures the interaction of being ambivalent of whether or not the marginal unit of biofuel is sold now or stored later to be sold at another date.

Our objective function then becomes maximizing the expected profits, which is

$$\max_{\pi \in \Pi} \mathbb{E} \left[ \sum_{t=0}^{T} C(S_t, X_{t}^\pi(S_t, W_t)) \right]$$

(2.6)

where the expectation is taken over all exogenous information $W$. We seek to maximize the function by finding an optimal policy $\pi$ in the set of possible policies $\Pi$. This is explained in the next section.

### 2.2.8 Model Justification and Properties

We now verbally defend the formulation of this model. The numerical simulations to defend this model are shown in the next chapter. Our goal is to create a model to simulate the optimal behavior of a biofuel facility’s manager. Clearly our state variable is capable of doing this. Our model captures the details of determining how much to expand to facility at the start of every year and how much ethanol to produce and much to sell in each month. We separate the decisions of producing and selling since we want to capture the notion of producing more than the market currently
demands so the manager can stockpile biofuel for “big shocks.”

We will later explain the theoretical justifications of the value function, but we will briefly expand on our justification of setting the price as the derivative of next month’s value function. The value function of a specific state at time $t$ is, as will see in the next chapter in detail, will be set to be the expected profit to be earned between time $t$ and $T$ following a specific policy. It is equivalent to the following recursive definition:

$$V_t(S_t) = \max_{x_t \in X_t} C(S_t, x_t) + \mathbb{E}_t[V_{t+1}(S^M(S_t, x_t, W_{t+1}))].$$

The equivalence is shown in the next section. If we adopt this value function, then we see that the price function balances out the marginal value of biofuel. By setting the price according to Equation (2.5), we can either earn $p_t$ now by selling this unit of gas, or we can collect $p_t$ sometime in the future by storing it and following the actions dictated by the value function. This matches our intuition.

### 2.3 Model 2

One of the dissatisfying equations in Model 1 is Equation (2.5), where the derivative should reflect a lower bound rather an equality. Also, $\gamma$ hardly factors into the model. Another dissatisfaction is the lack of incentive to store biofuel. Model 1 has no “shock” events, so there is no incentive to store biofuel especially if our price process already accounts for how much biofuel we currently have. These dissatisfactions are made concrete in the simulations in the next chapter. In this section, we propose an alternative model that fixes these problems, and as we will see in the simulations in the next chapter, the changes we propose here give a more satisfactory model.
2.3.1 Motivating Ideas

Based on the background we laid out in Chapter 1, we have good reason to believe that the current ethanol market would not be able to support the suddenly influx of demand for biofuel from gasoline distributors if a more aggressive renewable fuel standard (RFS) were enforced. Thus, it is not unreasonable to believe that there were be certain instances where the assumption of a competitive biofuel market will be violated. This is because the standard assumption of the a competitive market implies that there are many companies who can satisfy the demand, and one company’s decisions will not affect the price on a global scale. However, in our particular case, since there is not enough biofuel producers for the high amount of demand dictated by the RFS, sometimes we (an individual biofuel producer) might be the one of the few company able to supply biofuel. If this were the case, then we would simply charge a price of $\gamma$ since that is the maximum price of biofuel allowed. (If the price were higher than $\gamma$, the gasoline distributors would simply pay the penalty imposed by the government.)

Let us introduce a new variable $\hat{I}_t$ to put into the state variable $S_t$ to capture this idea

$$\hat{I}_t = \begin{cases} 0 & \text{the market is saturated at time $t$} \\ 1 & \text{else} \end{cases}$$

where we mean “saturated” as in there are enough biofuel producers where the distributors could fulfill their biofuel requirements even if we (one particular biofuel producer) did not sell to them.

Thus, we suggest an alternative pricing mechanism as illustrated in Figure 2.4. Here, $D_t$ (in blue) represents the total demand for time $t$ while $\rho$ represents the amount distributors are able to sale before facing regulation. $R_{t}^{\text{bio}}$ (in orange) represents how much biofuel is currently in storage. At time $t$, $\hat{I}_t = 1$, meaning the market is saturated. Note in this illustration, $R_{t}^{\text{bio}} \geq D_t - \rho$, meaning we have enough biofuel to
fulfill the needs of the distributors ourselves. But there are three different possibilities in the future:

- **Future a)** At some time in the future, the market is still saturated. Thus, based on our idea of a competitive market, the price still follows Equation (2.5) regardless of whether or not we can fulfill the needs of the distributors ourselves.

- **Future b)** At some time in the future, the market is no longer saturated, but we have enough biofuel to fulfill the needs of the distributors single-handedly. Since there are not enough other companies to supply biofuel, we collaborate with other companies and set the price to be slightly less than $\gamma$.

- **Future c)** At some time in the future, the market is no longer saturated and we do not have enough biofuel to fulfill the needs of the distributors single-handedly. In this case, the distributors will be desperate to buy biofuel and thus buy biofuel from us for a price slightly less than $\gamma$.

It is worth mentioning that the outcome for “Future b” is not obvious (and a bit arbitrary). Our choice of “Future b” allows us to reduce the three future possibilities into only two. Thus, we circumvent this problem by imposing large enough values of $c^{cap}$, $c^{pro}$, and $c^{bio}$ such that one biofuel producer can never reasonably saturate the entire amount demanded by the gasoline distributors. Another equally valid pricing model would be to set “Future b” to follow Equation (2.5), and we leave future work to experiment with this possibilities.

### 2.3.2 Model Specification

We specify the model by making the following changes to the state variable, exogenous information process, and transition function stated in Model 1.

**State Variable:** The state $S_t$ at time $t$ consists of $(R_t^{bio}, R_t^{cap}, D_t, \tilde{I}_t)$ where $\tilde{I}_t$ is an indicator value taking value 1 if the market is saturated at time $t$ and 0 else. The
Figure 2.4: An illustration to motivate Model 2.

Post-decisions state $S_t^x$ at time $t$ consists of $(R^\text{bio,}x_t, R^\text{cap,}x_t, D_t, \hat{I}_t)$ where $R^\text{bio,}x_t$ and $R^\text{cap,}x_t$ follow the same relationships as before.

**Exogenous Information Process:** The exogenous information $W_t$ that is revealed throughout the time $t$ consists of $(D_t, \hat{I}_t)$. At time $t$, we mandate that $W_{t+1}$ depends on $S_t$. Specifically, we impose a stationary distribution on $\hat{I}_{t+1}$ given $\hat{I}_t$ the following form

$$
\mathbb{P} (\hat{I}_{t+1} | \hat{I}_t = 1) = \begin{cases} 
\alpha^\text{sat} & \text{if } \hat{I}_{t+1} = 1 \\
1 - \alpha^\text{sat} & \text{if } \hat{I}_{t+1} = 0 
\end{cases}
$$

$$
\mathbb{P} (\hat{I}_{t+1} | \hat{I}_t = 0) = \begin{cases} 
\alpha^\text{usat} & \text{if } \hat{I}_{t+1} = 1 \\
1 - \alpha^\text{usat} & \text{if } \hat{I}_{t+1} = 0 
\end{cases}
$$

where $0 \leq \alpha^\text{sat}, \alpha^\text{usat} \leq 1$.

For notational convenience, we continue to write $W_{t+1}$ in the transition function
even though it should be formally written as $W_{t+1}(S_t)$ since (as we have specified here), the exogenous information depends on the state.

**Transition Function:** The transition function $S_{t+1} = S^M(S_t, x_t, W_{t+1})$ utilizes the following (same) functions with the additional transition function for $W_{t+1}$. That is,

$$R_{t+1}^{\text{bio}} = R_t^{\text{bio}} - x_t^{\text{bio}} + x_t^{\text{pro}}$$

$$R_{t+1}^{\text{cap}} = R_t^{\text{cap}} + x_t^{\text{cap}}$$

$$D_{t+1} = D_t + \hat{d}_{t+1}$$

$$W_{t+1} = \begin{cases} 
0 & \text{with probability } \alpha^{\text{sat}} \text{ if } \hat{\mathbb{I}}_t = 1 \\
& \text{or probability } \alpha^{\text{usat}} \text{ if } \hat{\mathbb{I}}_t = 0 \\
1 & \text{with probability } 1 - \alpha^{\text{sat}} \text{ if } \hat{\mathbb{I}}_t = 1 \\
& \text{or probability } 1 - \alpha^{\text{usat}} \text{ if } \hat{\mathbb{I}}_t = 0.
\end{cases}$$

### 2.3.3 Model Justification

We now verbally defend the formulation of the Model 2. Based on our new pricing procedure, we can see that we fixed the problems mentioned in the opening statements of this section. There is a reason to store biofuel rather than producing just enough biofuel to satisfy the current demand. We have also incorporated $\gamma$ more directly into this model, and now the price generated by Model 2 will be higher than the those generated in Model 1. Aside from the direct comparison due to the addition of these shock events, the value function for states with higher amounts of $R_t^{\text{bio}}$ will be valued more than those with lower amounts of $R_t^{\text{bio}}$ since the latter states cannot take advantage of the shock events as much. This difference will lead to a larger value of the derivative and hence a higher price even when the market is still saturated. The possibility of an unsaturated market drives up the prices and encourages storage.
2.4 Closing Remarks

We have developed two models in this section that both capture the intracity of managing a biofuel ethanol production facility. We keep track of how much biofuel is currently stored and how much production capacity (and storage capacity) we have, and we have to make decisions each month about how much biofuel to produce and how much biofuel to store. At the start of each year, we have the option to change our facility capacity. In Model 2, we add on an additional notion of market saturation to further encourage storage of biofuel. This “shock event” of lack of market saturation allows the biofuel facility to charge the highest possible price, $\gamma$, to the gasoline distributors.

It should be noted that there are many slight tweaks we can do to this model. For example, we can change the cost of production, storage and capacity changes by using a cost function that is not linear. In this thesis we focus on the stated setting since we want to keep the contribution function concave in the resources. In the next chapter (Chapter 3) though, we focus primarily on the backward Markov Decision Process, an algorithm that solves for the optimal policy. We will then explore the various quantities produced by the model to convince the reader of the algorithm’s accuracy.
Chapter 3

Algorithm: Backward Markov Decision Process

We now explain the algorithms we use to solve Model 1 and Model 2, the two models described in the previous chapter. Our goal to solve the problem is to determine a policy that figures out the decision $x_t$ at every time $t$ based on the state $S_t$. One method is the backward Markov Decision Process (bMDP) and the other is a Value Function Approximation (VFA) using the Concave, Adaptive Value Estimator (CAVE). While the former algorithm gives provably optimal solutions for a discretized problem, it suffers from an explosive computational expensive as the discretizations become finer and the dimensionality of the problem increases. The latter gives an approximate solution and does not suffer from such a drastic computational drawback. This chapter focuses on the bMDP. We first describe the value function, a critical component of both algorithms. We then explain this algorithm, understand its computational limitations and demonstrate example solutions to Model 1 and Model 2 to both verify the algorithm’s correctness in implementation and further gain intuition about our problem.
3.1 Value Function

The value function comes in two flavors: the value function for state variables and the post-decision value function for post-decision state variables. It usually is clear based on context which value function we are talking about, but regardless, the high-level idea of both flavors of value functions remain the same. We focus on the former here.

The value function of a state \( S_t \), \( V_t(S_t) \), is a function indexed by time interval \( t \) based on a policy \( \pi \) that returns the value of being in \( S_t \). It has a recursive definition where \( V_T^\pi(S_T) \) is

\[
V_T^\pi(S_T) = \max_{x \in \mathcal{X}_t} C(S_t, x)
\]

while for any \( t < T \), the value function is

\[
V_t^\pi(S_t) = \max_{x \in \mathcal{X}_t} C(S_t, x) + \mathbb{E}_t[V_{t+1}^\pi(S_{t+1}(S_t, x, W_{t+1}))]
\]

where the expectation is taken with respect to all the information known at time \( t \). A more careful explanation of this notation is given in Chapter 2. When the policy we are using is the backward Markov Decision Process (bMDP), we denote the value function (VFA) as \( V_t^{bMDP} \). Later when the policy is the value function approximation, we denote the value function as \( V_t^{VFA} \). If we are talking about value functions in general, we use \( V_t \) and drop the reference to the a specific policy \( \pi \).

It is important to notice that once we define the value function this way, \( V_t \) not only encodes the optimal decision \( x_t \) at time interval \( t \) as seen directly in the optimization problem but also for all times \( t' > t \) thanks to its recursive nature. To recover this decision, we can solve

\[
x_t = \arg \max_{x \in \mathcal{X}_t} C(S_t, x) + \mathbb{E}_t[V_{t+1}(S_{t+1}(S_t, x_t, W_{t+1}))]
\]

\[
= \arg \max_{x \in \mathcal{X}_t} C(S_t, x) + \sum_{\omega \in W_{t+1}} p(\omega) V_{t+1}(S_{t+1}(S_t, x_t, \omega))
\]
where our second equality explicitly writes out the expectation as a weighted summation of the value multiplied by the probability $p(\omega)$ of realizing the exogenous information $\omega \in W_{t+1}$.

We can start to see the intuition of the bMDP based on the value function. The first term in the value function, $C(S_t, x_t)$ denotes a deterministic, immediate profit for taking decision $x_t$ in state $S_t$ while the second term, $V_{t+1}(S^M(S_t, x_t, W_{t+1}))$ is a random, future profit. The maximization problem in the definition of the value function balances out these two forces, the instantaneous payoff versus the downstream payoff.

### 3.2 Algorithm Description

We introduce the most commonly-used algorithm to solve stochastic optimization problems, the backward Markov Decision Process (bMDP). A high-level description of the algorithm is as follows: the algorithm works backwards in time starting from $t = T$ to $t = 1$ where on time $t$, it determines the best possible decision $x_t$ for each possible state $S_t$ and calculates the value using $S_t, x_t,$ and $V_{t+1}$. The algorithm then records this value in the current value function, $V_t$. Since $V_{t+1}$ encodes the optimal decision $x_{t+1}$ and onwards, we are guaranteed to have an optimal policy once the algorithm finishes. The bMDP is a brute-force search algorithm since there are four levels of for-loops needed to solve the model. The outermost loop is over all time steps $t$, next the loop over all states $S_t$, next the loop over all decisions $x_t$, and lastly the loop over all possible exogenous information $W_t$ in the future time step. As we will see, these four sets of loops contribute the computational drawback of the bMDP.

We simulate both models in Matlab using the bMDP. In a bMDP, we discretize the state space $S_t$ for each time $t \in \{1 \ldots T\}$ and starting from time $T$, we calculate the optimal decision (and thus, value) for each discretized partition of $S_t$. This is
Algorithm 1: Backward MDP

\begin{algorithm}
\begin{algorithmic}[1]
\STATE 1 Determine a discretization for $S_t$, $x_t$, and $W_t$;
\FOR{$t \in \{T \ldots 0\}$}
\FOR{Each partition $s$ of $S_t$}
\FOR{Each partition $x$ of $x_t$}
\IF{$t \neq T$}
\FOR{Each partition $w$ of $W_{t+1}$}
\STATE Calculate the value of choosing decision $x$ when in state $s$
\STATE and the exogenous information $w$ weighted by the
\STATE probability of $w$ based on the value determined in $t+1$;
\END
\STATE Sum to calculate the expected value.;
\ELSE
\STATE Calculate the value of choosing action $x$ when in state $s$;
\END
\STATE Determine the best action $x$.;
\END
\END
\END
\end{algorithmic}
\end{algorithm}

described in Algorithm 1 where we write it explicitly to show all four sets of for-loops.

When we calculate the contribution term $C(S_t, x_t)$ inside the value $V_t(S_t)$, we will need to first calculate the price $p_t(S_t, x_t)$. Recall the definition of the price function Equation 2.5. The bMDP calculates a numerical derivative which involves taking the expected value of supposed post-decision state and adding the post-decision state with $\Delta$ units less stored biofuel as one the steps. (We get the number $\Delta$ since that is the smallest spacing in our bMDP.) That is, we actually calculate

$$p_t(s_t', x_t') = \sum_\omega p(\omega)[V_{t+1}(S_t + \Delta e_1, x_t, \omega) - V_{t+1}(S_t, x_t, \omega)] \frac{\Delta}{\Delta}$$

(3.2)

where $e_1$ is a vector of all zeros except a 1 in the first position. Recall that $S_t$ is a vector of $(R_{bio}^t, R_{cap}^t, D_t, \tilde{I}_t)$, so $S_t + \Delta e_1$ means we are adding $\Delta$ amount of $R_{bio}^t$ to $S_t$. Since we are using a bMDP, we have already calculated the value of every possible
While this algorithm can suffer from computational inefficiency extremely quickly depending on the way the states or decisions or exogenous information is partitioned, it gives an exact solution. We note that it is possible code the bMDP to have only three sets of for-loops since the inner-most set of for-loops (over each partition \( w \) of \( W_{t+1} \)) is needed to only calculate an expected value. Thus, we can simply do a dot product where one vector is a vector of probabilities instead of using a for-loop.

The bMDP requires the following parameters:

- Model parameters (8 parameters): \( T, \gamma, \rho, c^{\text{rat}}, c^{\text{salv}}, c^{\text{cap}}, c^{\text{pro}}, c^{\text{bio}} \).
- \( D_t \) (4 parameters): \( a_1 \) (Amplitude), \( a_2 \) (Frequency), \( a_3 \) (Base Height), and \( a_4 \) (Noise)
- \( D_t, R^\text{bio}_t, R^\text{cap}_t, x^\text{bio}_t, x^\text{cap}_t, x^\text{pro}_t \) (3 parameters each, 18 total): Minimum value, maximum value, spacing

The last parameter “spacing” refers to the discretization of the state, decision, and exogenous information. The smaller the spacing is, the finer detail the bMDP captures at the cost of a longer computation time.

For our particular set of simulations, we set the following values:

- \( T = 48, \gamma = 30, \rho = 0, c^{\text{rat}} = 3, c^{\text{salv}} = .6, c^{\text{cap}} = 10, c^{\text{pro}} = 3, c^{\text{bio}} = 1. \)
- \( a_1 = 200,000, a_2 = \pi/12, a_3 = 300,000, a_4 = 45,000. \)
- \( \min(D_t) = 500,000, \max(D_t) = 900,000, \max(R^\text{bio}_t) = 320,000, \max(R^\text{cap}_t) = 320,000, \max(x^\text{bio}_t) = 640,000, \max(x^\text{pro}_t) = 320,000, \min(x^\text{cap}_t) = -160,000, \max(x^\text{cap}_t) = 160,000. \)

All the minimums not stated are set to 0, and all of the spacings are set to \( \Delta = 80,000. \) We set \( \rho = 0 \) in our simulations in this section since we are just interested in the
veracity of the algorithm. Our settings of $a_1-a_4$ indicate that in expectation, our demand oscillates between 100,000 and 500,000 where the highest values of demand are larger than the possible amount to generate in a single period. Thus, for our model to satisfy the largest demands, it must store biofuel from the previous period. Our choice of frequency of the cyclical demand implies that one full cycle takes 24 months. This allows our model to change capacity during the “upwards” part of the cycle and the “downwards” part of the cycle. This will be useful to demonstrate the functionality of our model.

It is useful to make a quick calculation of how many inner-most iterations the bMDP does. Here, we have 48 time steps, each of which loops over the state $S_t = (R_{t}^{\text{bio}}, R_{t}^{\text{cap}}, D_t)$, each dimension having 5, 5, and 6 partitions each. For each partition of the state, we must try each decision $x_t = (x_t^{\text{bio}}, x_t^{\text{pro}}, x_t^{\text{cap}})$, each dimension having 9, 5, and 5 partitions each. Of course we only need to consider $x_t^{\text{cap}}$ once per 12 time steps. And as mentioned earlier in this section, we can calculate the expectation over the exogenous information $W_{t+1}$ as a dot product as opposed to a for-loop, so we ignore this last set of partitions. Thus, the bMDP calculates the inner-most iteration

$$\left(\underbrace{44}_{\text{Months without } x_t^{\text{cap}}} + \underbrace{4}_{\text{Months with } x_t^{\text{cap}}} \times \underbrace{5}_{x_t^{\text{cap}}} \times \underbrace{5 \times 6}_{S_t} \times \underbrace{9 \times 5}_{x_t^{\text{bio}}, x_t^{\text{pro}}}\right) = 432,000$$

And it also worth noting that if we made our spacing half the current size, the number of times the inner-most iteration gets calculated increases by a factor of roughly $2^6 = 64$ since all 6 variables (3 dimensions of the state and decision each) get twice as many partitions. This rapid increase in time alludes to why bMDP can only solve small problems.
3.3 Algorithm Verification for Model 1

We now use the bMDP to solve for the optimal decisions in Model 1 under the parameter settings described in the previous subsection. There are specifically two components we wish to check. The first component is the decision and transition functions, the second is the price function. When we check the decision and transition functions, we temporarily fix a basic pricing function that is derived from demand and ignore the definition of the price stated in Equation 2.5. Our goal is simply to see if the decisions outputted by the bMDP match our intuition. Here, the price function is set to be a rescaled value of the expected demand. This is not an unreasonable alteration to our model since we could imagine higher demands represented a higher willingness to pay for biofuel. The simulation results are show in Figure 3.1.

We will be using this plotting style for all future simulations presented in this
thesis. In these simulations, the lines denoted with the upper triangle denote state variables. The red line denotes how much biofuel is left in storage, and the blue line denotes how much biofuel capacity current exists, and the green line denotes the demand at time $t$. The downward triangles denote either the decision variables or the price. The black line denotes how much biofuel was sold, the black line denotes how much biofuel capacity was used for production, and the cyan line denotes how much biofuel capacity was added. Again, adding biofuel capacity only happens once every 12 time steps. The magenta line denotes the biofuel price and is rescaled to be shown on the graph. The price ranges from $0$ to $30$.

Looking at Figure 3.1, we can see that our decisions and transition functions work properly. The following are the observations that support this claim.

- We store biofuel mass $R_t^{bio}$ (red line) all throughout the upward part of the cyclical demand, and we unload all the biofuel to the gasoline distributors once we have reached the peak of the demand cycle. After that point, we are sure that high demands will not occur until the start of the next cycle.

- We see that we are producing the maximum amount of biofuel in most periods that is allowed by both the production rates and storage constraints. The former is seen by the fact that we are producing $x_t^{pro}$ (blue downward triangle) the same amount as our production capacity $R_t^{cap}$ (blue upward triangle), and the latter is seen that we are storing and producing enough so that $R_t^{bio} + x_t^{pro} \leq c^{rat} R_t^{cap}$ for the upward part of each cycle. We accumulate the amount in storage when the demand is relatively low but we still produce at maximum rate.

- The simulation shows that we add some capacity $x_t^{cap}$ (cyan) at the very start of the simulation just so we have enough capacity to last the entire 48 months.

- At each time interval, the amount we sell, $x_t^{bio}$ (black), does not exceed the demand $D_t$ (green).
Thus, we are confident that the bMDP correctly identifies the best decision for each state and transitions correctly. We now check the second component, the pricing function. Instead of using a deterministic price function, we return to the original pricing function described in Equation 2.5. The result is shown in Figure 3.2.

The results seen in Figure 3.2 are both promising and unsatisfactory. They are promising since we see a more reactive decision on when to expand and shrink the plant capacity. Specifically, we see that capacity is added the start of each cycle and shrunk at the end of each cycle. This matches our intuition on how the model should work since the additional capacity will help produce biofuel to serve the rising demand. The results are unsatisfactory though for the following reasons that were outlined in Chapter 2. They are as follows.

- The price $p_t$ (magenta) grows with time. Equivalently, as the bMDP solves the model backwards from $t = T$ to $t = 1$, the prices are decreasing with each
iteration. We notice that the price drops with each low-portion of the upward-half of the cycle, and in addition, the price is constant throughout the middle of each cycle.

• There does not seem to be much incentive to store biofuel.

We explain why, despite these unsatisfactory results, the algorithm is indeed correct. As mentioned in the previous section, the value function $V_t$ encodes the optimal decision from time interval $t$ onwards. Consider the following figure shown in Figure 3.3. Our optimal policy for Model 1 tells us to sell all of the biofuel we have at time $t$ in the same time. As we see in Figure 3.2, sometimes we inevitably produced more biofuel than the gas distributors currently demand, so that amount gets put into storage. (There are some exceptions to this rule is during time $t = 14, 15, 23, 24$ when the optimal policy temporarily stops selling biofuel completely to build up some initial storage amount.) Given this observation, consider the top graphic of Figure 3.3.

When we calculate the price using Equation (3.2), we perturb the post-decision state by $\Delta$ units, $S_t^x + \Delta$, and we compare the optimal policy from our current post-decision state $S_t^x$ to the optimal policy from our $S_t^x + \Delta$. Based on our observation, both policies converge to the same step in the very next time period. This implies that the only difference in value between the $S_t^x$ and $S_t^x + \Delta$ occurs in selling $\Delta$ more units of biofuel at time interval $t$. This logic applies for most of the time periods, and this explains why the price $p_t$ constant for many of the time intervals during the simulation. The price drops near start of each cycle’s upward trend since these are the time intervals where we cannot utilize the extra $\Delta$ units of biofuel. This is because the demand is currently too small and the extra $\Delta$ units are simply stored and sold at the same $p_t$ as before at a later date. The drop in difference in value is therefore due to the extra cost of storage.

We have argued that Figure 3.2 is not calculated incorrectly. However, it is somewhat disappointing since the price increases with each cycle. Ideally, we would like
What We Have:

\[ S_t \]

\[ S_t - \Delta \]

where the difference in value occurs

Time \( t < T - 1 \)  
Time \( t + 1 \)  
Time \( T \)

What We Want:

\[ S_t \]

\[ S_t - \Delta \]

where the difference in value occurs

Time \( t < T - 1 \)  
Time \( t + 1 < t' < T \)  
Time \( T \)

Figure 3.3: An illustration of the difference in value measured by the pricing function.

to see cyclical price pattern to match the cyclical demand. Thus, we need to alter Model 1 slightly so instead of observing the phenomenon described by the top portion of Figure 3.3, we observed the phenomenon described by the bottom portion. Specifically, if we give our model more incentive to store the extra \( \Delta \) units of biofuel so that the time when the optimal polices of both \( S_t^x \) and \( S_t^x - \Delta \) converge is further out in the future, there is more time for the optimal policy of \( S_t^x \) to collect a larger value. This will increase the price at time \( t \), and this thought process is how we came up with Model 2.

### 3.4 Algorithm Verification for Model 2

We now apply the bMDP to Model 2 with the desire to make the price process more cyclical with nature. We set \( \alpha^{sat} = .5 \) and \( \alpha^{unsat} = 0.1 \). This means there’s a 50% chance for the market to remain unsaturated if the market is currently unsaturated, but only a 10% chance for the market to become unsaturated if it’s currently sat-
Figure 3.4: A bMDP algorithmic check for Model 2.

As described in Chapter 2, the only change we made is the addition of $\hat{I}_t$ to the exogenous information process that controls the price process. The result is shown in Figure 3.4. The inclusion of the orange dots to our plots denote when $\hat{I}_t$ is 0 or 1.

The results shown in Figure 3.4 are quite shocking mainly for its quite dynamic price process. We notice that while our simulation does not store as much biofuel as before in Figure 3.2, we now use more production capacity. This is so we can maximize our profits when a shock event ($\hat{I}_t = 1$) occurs. The price seems to drop close to 0 at the tail end of each cycle, an indication that there is the marginal $\Delta$ units of biofuel makes little difference in this scenario since the biofuel is simply stored to be sold at the start of the next cycle. The price is higher during the start of each cycle since having $\Delta$ more units of biofuel in storage makes a big difference especially when a shock event occurs. Model 2 seems to have achieved what we desired since
the price $p_t$ is no longer simply increasing with time.

We can explore the associated quantities to our 96-month simulation using this Model 2 to further confirm its veracity.

In Figure 3.6, we plot the cumulative profit over time. This corresponds with the simulation shown in Figure 3.5. As we can see, the regions where the profit flat-lines correspond exactly with the regions where the price is close to 0. This makes sense since our price is defined to be the derivative of the value function as seen in Equation 2.5, so a price of near-0 implies a constant cumulative profit. Next, in Figure 3.7, we plot the expected prices over time during market saturation ($\tilde{I}_t = 0$) at the various discretizations of $R_t^{cap}$. As we can see, when we do not have any capacity, the price is $30 (the maximum price dictated by $\gamma$) since any additional $\Delta$ units of biofuel is immediately beneficial. Once we have a non-zero capacity, the marginal benefit of gas fluctuates, so the price $p_t$ fluctuates as well. We can once again confirm a cyclical
Figure 3.6: A plot of the profit over time based on Figure 3.5.

Figure 3.7: A plot of price over time for different capacity levels based on the bMDP policy.

nature of the price process with the additional of an end-of-horizon effect. Next, in Figure 3.8, we plot the expected value with respect to the amount of biofuel stored $R_t^{bio}$ at various discretizations of $R_t^{cap}$. As we stated in Chapter 2, we expect our value function to be concave in both $R_t^{bio}$ and $R_t^{cap}$, and we can clearly verify its concavity with $R_t^{bio}$ here. The slopes for all capacities besides $R_t^{cap} = 320,000$ are roughly the same, and the slope for $R_t^{cap} = 320,000$ is considerably lower. In addition, the actual values for $R_t^{cap} = 320,000$ are lower than the values for $R_t^{cap} = 240,000$. This makes sense with respect our plot Figure 3.5 since we end up using capacity $R_t^{cap} = 240,000$. 

57
Figure 3.8: A plot of the value as a function of $R^\text{bio}_t$ for different capacity levels for time $t = 65$ based on the bMDP policy.

in our bMDP policy. Lastly, we show the 3D plot of the value function in Figure 3.9. The value function along the x-axis ("Amount of Biofuel") correspond to the curves shown in Figure 3.8.

Lastly, we mention the empirical time needed to solve these different policies. These are based on one trial. The results are shown in Table 3.1. We notice a roughly-twice increase in time needed from “Model 1 with 48 Months” to “Model 2 with 48 Months” since there are twice as many states in Model 2 due to the inclusion of the shock variable $\hat{I}_t$. There is a roughly-twice increase from “Model 2 with 48 Months” to “Model 2 with 96 Months” due to twice the number of months. Lastly, as expected, there is more than a 64-times increase from “Model 2 with 48 Months, $\Delta = 8000$” to “Model 2 with 48 Months, $\Delta = 4000$” since we halved the discretization level so there are 64 times as many iterations of the inner loop as discussed previously.
3.5 Look-Up Tables and Curses of Dimensionality

While the bMDP policy gives the optimal policy, it suffers from many drawbacks. Based on our calculation of how many inner-loops was required for the bMDP, we can see that simply applying a discretization with half the coarseness will result in
64 times the computation time. If the model had even more dimensions in either
the state variable, the decision or exogenous information, the computational time
required will grow even more.

There are many reasons why the bMDP is unattractive, and these reasons are
often related to the three “curses of dimensionality.” Recall our definition of the
value function in Equation (3.1). The three curses of dimensionality refer to

• 1) Computing the value of a particular state $V_t(S_t),$
• 2) Computing the expectation $E_t(V_{t+1}(S^M(S_t, x_t, W_{t+1}))),$
• 3) Solving the maximization problem for the optimal decision $x_t.$

The bMDP policy is purely a look-up table policy. That is, we literally calculate
the value at every possible state based on how we discretized our state space into a
grid of points so we can later “look up” the value at that state. This alludes to the
first curse of dimensionality especially since we have a multi-dimensional state. The
bMDP policy then calculates the expectation $E_t(V_{t+1}(S^M(S_t, x_t, W_{t+1})))$ by calculating

$$
\sum_{\omega \in W_{t+1}} p(\omega)V_{t+1}(S^M(S_t, x_t, \omega)),
$$

a loop over all possible exogenous information $\omega \in W_{t+1}$ received for time interval
$t+1$. This refers to the second curse of dimensionality especially since we have a multi-
dimensional exogenous information. Finally, the bMDP tries all possible decisions $x_t$
to determine the best decision, and this refers to the third curse of dimensionality
especially since we have a multi-dimensional decision.

In the next chapter, we try a different policy that alleviates the first two curses.
That is, we will choose a different way to represent the state space in order to reduce
the number of dimensions in our look-up table and we choose a different way to
represent our value functions so we avoid calculating a summation over all possible
exogenous information realizations $\omega \in W_{t+1}$.

### 3.6 Closing Remarks

In this section, we developed a powerful algorithm that solve for optimal policies but is extremely computationally intensive. We tried this algorithm on both models developed in Chapter 2 to see what the difference of the indicator variable $\hat{I}_t$ resulted in. As we have seen in this section, the addition of $\hat{I}_t$ satisfies what we hoped to have achieved as stated in Chapter 2, and we have shown detailed results of various simulations backed by verbal arguments to justify that the algorithms were implemented correctly and to gain deeper intuition on our model. Next, we will try a different policy called the value function approximation (VFA) that is much faster but does not get the exact optimal solution.
Chapter 4

Algorithm: Value Function Approximation

We introduce powerful tools used in approximate dynamic programming that allow us to approximate the value of being a state in this chapter. These approximation methods combine to alleviate the curse of dimensionality. In particular, we introduce an algorithm that estimates the value function that relies only on the assumption of convexity. In addition, we avoid the necessity to solve the model backwards in time. Recall that the backward Markov Decision Process (bMDP) solves the model starting from $t = T$ to $t = 1$ and calculates the value of every state along the way. However, many states solved in this procedure are simply never actually visited during the simulation. We take a slightly different approach in this chapter where we step forward in time to solve the model and only update the “relevant” states (the states we more likely to end up visiting during the simulation).

4.1 Approximation Methods

One of the most powerful tools of approximate dynamic programming is the ability to create a policy by approximating the value of being in a state. This is known
as value function approximations (VFA). In this section, we introduce three ways to perform different approximations. If we can somehow obtain an approximated value $\bar{V}_t(S_t)$ for state $S_t$, then our policy can determine its action by solving

$$x_t = \arg \max_{x \in X_t} C(S_t, x) + \mathbb{E}_t[\bar{V}_{t+1}(S^M(S_t, x, W_{t+1}))].$$

The three approximations involve aggregating states to lower the dimensionality of the state-space when we calculate the value, using post-decision value functions to avoid calculating each possible exogenous information and updating our post-decision value functions via the Concave, Adaptive Value Estimator (CAVE). In the next section, we incorporate these ideas into the full algorithm that will show how we can solve for the approximately optimal policy. Note that while we lose strict optimality, we can verify experimentally how close we are to the true optimal.

4.1.1 Aggregation Functions

One popular way to deal with curse of dimensionality has been define the value function over a smaller state space. That is, instead of calculating $V_t(S_t)$, the exact value of at state $S_t$, we calculate $\bar{V}_t(S^{(g)}_t)$, the approximated value at an aggregated state. If we are using a look-up table to store these approximated values, then having a lower-dimensional state $S^{(g)}_t$ means we have a simpler calculation. Note that this aggregation is used only to approximate the value function. Once we have chosen our decision $x_t$, we simulate our way to the next state as before using $S_{t+1} = S^M(S_t, x_t, W_{t+1})$ where $S_t$ is our original, disaggregated state variable. A formal treatment of aggregate functions is given in Powell (2011).

In our particular application, instead of aggregating the state variable $S_t$, we are more interested in aggregating the post-decision state variable $S^x_t$. There are two aggregation functions are interested in this chapter. In the first method, we
disregard the $\hat{I}_t$ variable in the post-decisions state, so if we were to write out the state fully, it would be $S_t^x = (R_t^{\text{bio},x}, R_t^{\text{cap},x}, D_t, \hat{I}_t)$ and the aggregation function would be $g'(S_t^x) = (R_t^{\text{bio},x}, R_t^{\text{cap},x})$. The second aggregation function we are interested in is $g(S_t^x) = R_t^{\text{bio},x} + R_t^{\text{cap},x}$.

When we describe the CAVE and VFA algorithm, we be will using the former aggregation function to deal with value functions. However, extending the CAVE and VFA algorithms to work on second aggregated states is trivial, and we show the comparisons between these two simplifications of the value functions later on in this chapter. For notational convenience, we will omit the superscript $(g)$ in the aggregate state $S_{t}^{x,(g)}$ throughout this entire chapter, and we simply write $S_t^x$ when we describe value at a particular aggregated post-decision state variable.

### 4.1.2 Post-Decision Value Functions as Gradients

Determining the value of a state $V_t(S_t)$ requires $\mathbb{E}_t[V_{t+1}(S^M(S_t, x_t, W_{t+1}))]$, an expectation with respect to the exogenous information $W_{t+1}$ of the value in the next time interval $t + 1$. Previously in the bMDP, we calculated this expectation by calculating $\sum_{\omega} p(\omega)V_{t+\delta}(S^M(S_t, x_t, \omega))$ for $\omega \in W_{t+1}$. Sometimes this expectation cannot be easily calculated for multi-dimensional exogenous information $W_{t+1}$, so we introduce the post-decision value function to alleviate this problem. The post-decision value function of a post-decision state $S_t^x$, $V_{t}^{x,:\pi}(S_t^x)$, is a function indexed by time interval $t$ based on a policy $\pi$ that returns the value of being $S_t^x$. We define this function by stating its relationship with the value function for $t < T$.

\[
V_{t}^{\pi}(S_t) = \max_{x \in X_t} C(S_t, x) + V_{t}^{\pi}(S_t^x)
\]

\[
V_{t}^{x,:\pi}(S_t^x) = \mathbb{E}_t[V_{t}^{\pi}(S^M(S_t^x, W_{t+1}))]
\]
where $S^x_t = S^M(S_t, x_t)$ and $x_t$ refers to the argument that maximizes the first equation. When describe the VFA, it will be clearer why this is useful. As before, we drop the $\pi$ in the notation if we are talking about post-decision value functions in general. Based on Equation (3.1), we can easily see that the relationship between the value and post-decision value functions are valid.

We include an additional twist. We often are not interested in the value of a state itself but the gradient of the value at that state. If the function is piecewise linear, we can replace the estimate of the value of being in a state with a single parameter that estimates the slope of the function in each particular piece. Estimating the constant term is unnecessary since it does not affect our maximization problem. Storing the derivatives has another advantage over storing the actual value since we are no longer constrained to picking a decision that corresponds to a position to the look-up table.

Let $(R_{t}^{\text{bio},x}, R_{t}^{\text{cap},x})$ denote our resource vector within in our post-decision state $S^x_t$. That is, instead of storing $V^x_t(S^x_t)$, we are interested in

$$\nabla V^x_t(S^x_t) = (v_{t}^{\text{bio},x}, v_{t}^{\text{cap},x})$$

where

$$v_{t}^{x,\text{bio}} = \frac{\partial V^x_t(S^x_t)}{\partial R_{t}^{\text{bio},x}}$$

and a similar equation holds for $v_{t}^{x,\text{cap}}$.

We stay in the look-up table framework where we store the derivatives instead of the values themselves at a discretized grid of points. If we deal with a multi-dimensional state, we can pursue two different options as opposed to storing the derivative with to $R_{t}^{\text{bio},x}$ and $R_{t}^{\text{cap},x}$ at a discretized grid of points. We can either assume additional structure to the value function or we can aggregate the states through an aggregation function $g$ to produce a lower-dimensional representation of the state. Let us focus on the aggregate function $g'(S^x_t) = (R_{t}^{\text{bio},x}, R_{t}^{\text{cap},x})$ mentioned
We assume
\[
\begin{align*}
\frac{\partial \bar{V}_t^x((R_t^{\text{bio}, x}, R_t^{\text{cap}, x}))}{\partial R_t^{\text{bio}, x}}_{R_t^{\text{bio}, x} = s, R_t^{\text{cap}, x} = s'} &= \frac{\partial \bar{V}_t^x((R_t^{\text{bio}, x}, R_t^{\text{cap}, x}))}{\partial R_t^{\text{bio}, x}}_{R_t^{\text{bio}, x} = s, R_t^{\text{cap}, x} = t} \\
\frac{\partial \bar{V}_t^x((R_t^{\text{bio}, x}, R_t^{\text{cap}, x}))}{\partial R_t^{\text{bio}, x}}_{R_t^{\text{bio}, x} = s, R_t^{\text{cap}, x} = t'} &= \frac{\partial \bar{V}_t^x((R_t^{\text{bio}, x}, R_t^{\text{cap}, x}))}{\partial R_t^{\text{bio}, x}}_{R_t^{\text{bio}, x} = s, R_t^{\text{cap}, x} = t'} \\
\forall s, s', t \\
\forall s, t, t'.
\end{align*}
\]

We will call this the independent derivatives assumption. These assumptions simply require that the derivative with respect to one variable is independent of the other variable. This assumption allows us to simply store two separate vectors of derivatives instead of a grid of derivatives. A graphical illustration of the assumed structure of the value function is shown in Figure 4.1. On the right, we see that the derivative (in black) with respect to $R_t^{\text{bio}, x}$ is the same at both $(s_1, t_1)$ and $(s_1, t'_1)$ where $t_1 \neq t'_1$. Likewise, the derivative (in red) with respect to $R_t^{\text{cap}, x}$ is the same at both $(s_2, t_2)$ and $(s'_2, t_2)$ where $s_2 \neq s'_2$. On the left, we see that this two-dimensional value function was built from two one-dimension value functions.
4.1.3 CAVE Algorithm

We now describe the algorithm that builds upon the post-decision value function. One of the key ideas that we will utilize the Monte Carlo sampling to approximate the expectation. That is, we would let $\nu$ be our current approximation for $\bar{V}_x(S_t)$, the gradient of the approximate value at $S_t$. Once we calculate a numerical noisy derivative $\bar{v}$, we update our gradient estimate of the expectation with

$$\nu^{(n+1)} = \alpha^{(n)} \bar{v} + (1 - \alpha^{(n)}) \nu^{(n)}$$

where we explicitly write $(n)$ to denote the iteration counter. Of course, if we set $\alpha^{(n)} = \frac{1}{n}$, we are simply averaging our noisy derivatives $\pi^{(n)}$. The factor $\alpha^{(n)}$ is called the step-size. However, since we have a stochastic model, setting $\alpha^{(n)} = \frac{1}{n}$ often underperforms. We set different step-size schedules and hope that over time, we converge upon the true value of $V^x_t(S^x_t)$. We discuss this more in a later subsection.

We now assume that the value function stated in Equation 3.1 is concave with respect to $R^\text{bio}_t$ and $R^\text{cap}_t$ at each time $t$. That is, for any $\lambda \in [0, 1]$ and any two pairs of state variables $R^\text{bio}_t, R^\text{cap}_t$ and $R'^\text{bio}_t, R'^\text{cap}_t$, our value function satisfies

$$\lambda V_t((R^\text{bio}_t, R^\text{cap}_t)) + (1 - \lambda) V_t((R'^\text{bio}_t, R'^\text{cap}_t)) \leq V_t((\lambda R^\text{bio}_t + (1 - \lambda) R'^\text{bio}_t, \lambda R^\text{cap}_t + (1 - \lambda) R'^\text{cap}_t)).$$

This property is not unreasonable if we assume the contribution function $C(S_t, x_t)$ is always concave. This cannot be easily verified due to the definition of the price seen in Equation 2.5. But if $C(S_t, x_t)$ is concave and the value function in the next time step $V_{t+1}$ is also concave, then certainly $\mathbb{E}_t[V_{t+1}]$ is concave, meaning $V^x_t$ is concave. Since the sum of two concave functions is also concave, we would get the desired result that the post-decision value function $V^x_t$ is concave for all time $t$. In reality, this is not too important since we are simply asserting this assumption.
The Concave, Adaptive Value Estimation (CAVE) algorithm, introduced in Godfrey et al., estimates a continuous, concave function $V_x^t(S_t)$ as a concave, piecewise-linear approximation using repeated stochastic gradient samples of the function at different values (Gregory A. Godfrey (2001)). Thanks to the assumption of independent derivatives made in the previous section, we will apply the CAVE algorithm to each one-dimensional value function separately. We state the algorithm to update only the value function associated with $R_{t}^{\text{bio},x}$ here, but the same ideas can be used on the value function of $R_{t}^{\text{cap},x}$.

Let us focus on one particular time $t$. The piecewise-linear approximation of $V_x^t(R_{t}^{\text{bio},x})$, which we denote by $\bar{V}_x^t(R_{t}^{\text{bio},x})$ is defined by a finite set of ordered breakpoints $\{(\nu_k, u_k) : k \in \mathcal{K}\}$ where $\mathcal{K} = \{0, 1, \ldots, k_{\text{max}}\}$. Each breakpoint defines a linear segment with $\nu_k$ as the slope of the segment from $u_k$ to $u_{k+1}$. However, if $k = k_{\text{max}}$, we interpret $\nu_k$ as the slope of the segment from $u_k$ to $\infty$. The breakpoints are ordered so that $0 = u_0 < u_1 < u_2 < \ldots < u_{k_{\text{max}}} \leq \infty$ and $\nu_0 > \nu_1 > \ldots > \nu_{k_{\text{max}}}$. Let $u$ be the vector $\{u_0, u_1, \ldots, u_{k_{\text{max}}}\}$ and $\nu$ be the vector $\{\nu_0, \nu_1, \ldots, \nu_{k_{\text{max}}}\}$. We define $\nu(R_{t}^{\text{bio},x})$ to be slope of $\bar{V}_x^t$ at $R_{t}^{\text{bio},x}$. It follows from the monotonically decreasing slopes that $\bar{V}_x^t(R_{t}^{\text{bio},x})$ is concave.

Next, at a point $R_{t}^{\text{bio},x} = s$ and $R_{t}^{\text{cap},x} = t$ in the domain, we define the gradient of $V_x^t$ with respect to $R_{t}^{\text{bio},x}$ as

$$\bar{v}(s, \Delta) = \frac{\bar{V}_{t+1}(S^M((s + \Delta, t), w)) - \bar{V}_{t+1}(S^M((s, t), w))}{\Delta}$$

(4.1)

where $\Delta \in \mathbb{R}$ is some small positive number and $w$ is a sample realization of $W_{t+1}$. Intuitively, we get a numerical derivative of the post-decision amount of biofuel in storage by perturbing the post-decision state and seeing how our value changes. Since we do not actually store the value functions $\bar{V}_t$ but rather the approximated post-decision value functions $\bar{V}_x^t$, we can calculate $\bar{v}$ by Algorithm 2. We will use $\bar{v}$ as the
Algorithm 2: Numerical Derivative of Post-Decision Value Function.

<table>
<thead>
<tr>
<th>Line</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Sample ( w ) from ( W_{t+1} ).</td>
</tr>
<tr>
<td>2</td>
<td>( S_{t+1} \leftarrow S^M((s, t), w); , S'_{t+1} \leftarrow S^M((s + \Delta, t), w); )</td>
</tr>
<tr>
<td>3</td>
<td>( \bar{V}<em>{t+1}(S</em>{t+1}) \leftarrow \max_{x \in \mathcal{X}<em>{t+1}} C(S</em>{t+1}, x) + \bar{V}<em>{x</em>{t+1}}(S^x_{t+1}); )</td>
</tr>
<tr>
<td>4</td>
<td>( \bar{V}'<em>{t+1}(S'</em>{t+1}) \leftarrow \max_{x \in \mathcal{X}<em>{t+1}} C(S'</em>{t+1}, x) + \bar{V}<em>{x</em>{t+1}}(S'^x_{t+1}); )</td>
</tr>
<tr>
<td>5</td>
<td>( \bar{v} \leftarrow \frac{\bar{V}'<em>{t+1}(S'</em>{t+1}) - \bar{V}<em>{t+1}(S</em>{t+1})}{\Delta}; )</td>
</tr>
</tbody>
</table>

Output: Numerical derivative \( \bar{v} \)

noisy numerical derivative to update the post-decision value function \( \bar{V}^x_{t} \). We see two key ideas in this procedure. First, \( \bar{v} \) is a noisy derivative, but if we update \( \bar{V}^x_{t} \) many times, we achieve the expectation embedded within \( \bar{V}^x_{t} \). Second, the calculation of \( \bar{v} \) requires finding the optimal decision based on \( \bar{V}^x_{t+1} \). Since \( \bar{v} \) will be used to update \( \bar{V}^x_{t} \), we see how information about optimal decision propagates backwards in time in our model. Once \( \bar{V}^x_{t+1} \) is updated to reflect the “true” optimal decision at time \( t + 1 \), our procedure will use \( \bar{V}^x_{t+1} \) to update \( \bar{V}^x_{t} \) to reflect the “true” optimal decision at time \( t \).

Now that we explained how to calculate the derivative \( \bar{v} \), we explain how to use \( \bar{v} \) to update \( \bar{V}^x_{t} \). The CAVE algorithm uses \( \bar{v} \) to update an interval of \( \bar{V}^x_{t} \) of length at least \( 2\beta \) around the current value of \( R^\text{bio},x_{t} \). If the monotonicity of the slopes is violated, the algorithm then uses \( \bar{v} \) to update a wider interval until the monotonicity of the slopes is restored. The full algorithm is explained in Algorithm 3. This procedure is applied from \( t = 1 \) to \( t = T \) for many iterations. If the final objective value at \( t = T \) does not change much within many iterations, we decrease the smoothing interval \( \beta \).

This is graphically explained in Figure 4.2. In the first step, we calculate the derivative \( \bar{v} \) at \( R^\text{bio},x_{t} = s \) which we will later apply over an interval \([k^-, k^+]\). The
Algorithm 3: CAVE Algorithm (for $R_t^{bio,x}$)

**Input:** Post-decision state $S_t = (s, t)$, Post-decision value functions $\bar{V}_t^x, \bar{V}_{t+\delta}^x$, Perturbation factor $\Delta$, Smoothing interval width $\beta$, Step-size $\alpha$

1. Approximate derivative $\bar{v}$

/* Perform Smoothing */
2. $k^- \leftarrow \min(s - \beta, 0); k^+ \leftarrow s + \beta$;
3. Create new breakpoints at $k^-$ and $k^+$ in the $\bar{V}_t^x$ by inserting $k^-, k^+$ into the vector $u$ and adding the corresponding slope values to $\nu$ by borrowing from the preceding segment;.
4. $i \leftarrow$ Index of $k^-$ in $u$; $j \leftarrow$ Index of $k^+$ in $u$;
5. For each segment $k$ of $\nu$ between $i$ and $j - 1$ (inclusive), update the slope according to $\nu^k \leftarrow \alpha \bar{v} + (1 - \alpha) \nu^{k-1}$;

/* Restore monotonicity in slopes */
6. For all segments from $k = \max(i, 0)$ to $k = 0$, check if $\nu_k \leq \nu_{k-1}$. If not, $\nu_{k-1} \leftarrow \alpha \bar{v} + (1 - \alpha) \nu_{k-1}$;
7. For all segments from $k = \min(j + 1, k_{\text{max}})$ to $k = k_{\text{max}}$, check if $\nu_k \leq \nu_{k-1}$. If not, $\nu_k \leftarrow \alpha \bar{v} + (1 - \alpha) \nu_k$.

**Output:** Updated post-decision value function $\bar{V}_t^x$

height of $\nu_0$ and $\bar{v}$ represent the value of each respectively, and each blue rectangle represents $u_k$. Here, $[k^-, k^+]$ intersects $u_3$ through $u_5$. In the second step, we perform our smoothing by forming new breakpoints at $k^-$ and $k^+$ and updating the derivatives by $\nu_k \leftarrow \alpha^{(n)} \pi + (1 - \alpha^{(n)}) \nu_k$ for all $k$ that fall inside the interval $[k^-, k^+]$. In the third step, we restore monotonicity by updating the derivatives of all other $\nu_k$’s until monotonicity in the derivatives is restored.

### 4.2 Algorithm Description

Now that we have the CAVE algorithm, we present the full algorithm to determine the approximately optimal policy. We set a total number iterations $I$. In each iteration, we go through our model from $t = 1$ to $t = T$ where in each time interval $t$, we find the best decision $x_t$, calculate the derivative $\bar{v}$ and update the post-decision value function $\bar{V}_t^x$.

The last component of the algorithm is specifying the step-size schedule $\beta^{(n)}$ and
\( \alpha^{(n)} \). In this version of the CAVE algorithm, we consider a constant \( \alpha^{(n)} = \frac{150}{150+n-1} \). Let \( P^{(n)} \) denote the objective function achieved at time interval \( t = T \) on iteration \( n \) of the algorithm. Let \( P^{(i:i')} = \frac{1}{i'-i+1} \sum_{n=i}^{i'} P^{(n)} \). We set \( \beta^{(n+1)} \) to

\[
\beta^{(n+1)} = \begin{cases} 
\beta^{(n)}/3 & \text{if } \frac{|P^{(n-11:n-1)} - P^{(n-10:n-1)}|}{P^{(n-11:n-1)}} \leq 0.001 \\
\beta^{(n)} & \text{otherwise}
\end{cases}
\]  

(4.2)

where we consider this equation if there have been more 12 iterations. Before then, \( \beta^{(n+1)} = \beta^{(n)} \). This is shown in Algorithm 4. Based on this rule, if the average objective function over the last 10 iterations does not change by more than 0.1%, then we claim that our VFA policy needs to update over a smaller \( \beta \) to hone in on a better decision.
Algorithm 4: VFA Algorithm using CAVE

**Input:** Total number of iterations $N$, Initial post-decision value functions $\bar{V}^{x,(1)}_{1:T}$, Initial state $S_1$, Initial interval width $\beta^{(1)}$, Step-size schedule $\alpha^{(1:N)}$, Perturbation factor $\Delta$

1. for $n \in 1 : N$ do
   2. for $t \in 1 : T$ do
      /* Solve for decision */
      3. $x_t \leftarrow \arg \max_{x_t} C(S_t, x) + \bar{V}^{x,(n)}_{t}(S_t)$;
      /* Calculate the derivative (See Algorithm 2) */
      4. $s \leftarrow R^{bio,x}_{t}$; $t \leftarrow R^{cap,x}_{t}$;
      5. $\bar{v}^{bio} \leftarrow \frac{\bar{V}^{x,(n)}_{t+1}(S^M((s+t)\Delta),w) - \bar{V}^{x,(n)}_{t}(S^M((s,t)\Delta),w))}{\Delta}$;
      6. $\bar{v}^{cap} \leftarrow \frac{\bar{V}^{x,(n)}_{t+1}(S^M((s,t+\Delta)\Delta),w) - \bar{V}^{x,(n)}_{t+1}(S^M((s,t)\Delta),w))}{\Delta}$;
      /* Update post-decision value function (See Algorithm 3) */
      7. $\bar{V}^{x,(n+1)}_{t} \leftarrow$ CAVE update of $\bar{V}^{x,(n)}_{t}$ using $\bar{v}^{bio}$, $\bar{v}^{cap}$, $\alpha^{(n)}$ and $\beta^{(n)}$ at $S^x_{t}$;
      /* Generate exogenous information and transition */
      8. Sample $w$ from $W_t$;
      9. $S_{t+1} \leftarrow S^M(S^x_t, w)$;
   end
/* Update $\beta$ */
10. $\beta^{(n+1)} = \begin{cases} \beta^{(n)}/3 & \text{if } n \geq 12, \frac{|p^{(n-11:n-1)} - p^{(n-10:n)}}{p^{(n-11:n-1)}} \leq 0.01 ; \\ \beta^{(n)} & \text{otherwise} \end{cases}$
11. end

**Output:** Post-decision value functions $\bar{V}^{x,(N)}_{1:T}$

4.3 Algorithm Verification for Model 2

We now run the VFA policy to verify the correctness of the algorithm and to gain intuition on how VFA behaves. We use the following parameter settings.

- All post-decision value functions $\bar{V}^{x,(1)}_{1:T}$ are initialized in such a way where the derivatives with respect to $R^{bio,x}_{t}$ were discretized so $k_{max} = 7$, $\nu_0 = 5$, $\nu_{k_{max}} = -5$, $u_0 = 0$, $u_{k_{max}}$ = 960,000 and all the spacings were equal. Likewise, the derivatives with respect to $R^{cap,x}_{t}$ were discretized so $k_{max} = 5$, $\nu_0 = 5$, $\nu_{k_{max}} = -5$, $u_0 = 0$, $u_{k_{max}} = 320,000$. The resulting post-decision value function is
Figure 4.3: Initial setting of $\bar{V}_t^x$ for all $t$

shown in Figure 4.3.

• $\alpha^{(n)} = 150/(150 + n - 1)$, $\beta^{(1)} = 240,000$ when updating $R_t^{\text{bio},x}$, $\beta^{(1)} = 80,000$ when updating $R_t^{\text{cap},x}$,

• $\Delta = 10$, $N = 10,000$.

The result when we apply this algorithm for only 48 months is shown in Figure 4.4. As we can see, the price curve is not as smooth as in the bMDP counterpart shown in Figure 3.4, but there are strong resemblances. For instance, the price drops towards 0 during the months with low demand, and we observe a period of time where biofuel is accumulated in storage. This accumulated biofuel is sold when there is a shock event ($\hat{I}_t = 1$).

Since we want to ensure that our simulation is not affected by end-of-horizon effects, we show a simulation over 96 months. This is shown in Figure 4.5. First let us make three observations.

• First, storage does not always go to zero after every cycle such as $t = 20$ to $t = 35$. The VFA policy tells us to store biofuel and sell the stored biofuel
whenever there is shock event. In the lack of a shock event, we simply keep the biofuel in storage for the next shock event.

- Second, during the shock event in $t = 22$ to $t = 24$, we do not unload all our stored biofuel onto the market though. This is because the price in these months (which we can infer) are already extremely high during these months disregarding of the occurrence of a shock event. Thus, we choose only sell a portion of our stored biofuel. However, as we can see in $t = 50$, the contribution term in the maximization problem strongly outweighs the value term. Thus, we sell everything we have in storage.

- Third, production is not necessarily always set to a maximum. We can see in $t = 38$ to $t = 40$, we are not utilizing the full production capacity of our plants. This is because we already have a lot of biofuel in storage already, so we do not want to accumulate even more.
Compare Figure 4.5 with Figure 3.5. The most striking difference is the production capacity. In the bMDP policy, we have a high amount of production capacity (240,000 units) towards the beginning of each cycle, but in the VFA policy, we have this level of capacity towards the end of each cycle. It is hard to say which policy achieves a higher profit from these figures since the bMDP finds its solution over a discretized state space. In the next section, we make a more direct comparison.

Finally, we display the cumulative profit over time as well as the post-decision value functions. The profits are shown in Figure 4.6. Once again, we see a plateau-like curve where the flat areas represent months where price is near-zero while the ascent areas represent high prices and high demand. In Figure 4.7, we show the post-decision value function with respect to $R_{t}^{\text{bio},x}$ for months $t = \{61, 63, 65, 67, 69, 71, 73, 75, 77, 79\}$. In Figure 4.8, we show the same except with respect to $R_{t}^{\text{cap},x}$ for months $t = \{1, 13, 25, 37, 49, 61\}$. The plots all have a value of zero at $0 R_{t}^{\text{bio},x}$ or $0 R_{t}^{\text{cap},x}$ since
we are ignoring constant terms in the value function. (Recall that we store only the
derivatives.) The red dots indicate the actual breakpoints of \( u_0, u_1, \ldots, u_{k_{\text{max}}}. \) Not all
curves have the same number of breakpoints due to how we solve our maximization
problem.

Let’s look at the post-decision value functions with respect to \( R_t^{\text{bio},x} \). We choose to
show these particular months in Figure 4.7 since these plots cover an entire demand
cycle. We can see in these curves that for months \( t = \{61, 63\} \), our VFA policy
discourages us from storing anything in storage. We do not want to store anything
during these months since the demand is low. Even if a shock event were to happen, we cannot sell that much. Storing in these months is not worth the additional storage cost. However, as demand starts to rise from months $t = 65$ onwards, our value functions encourage us to store more. However, if we recall how the decisions are made in the maximization problem, usually the contribution term encourage us to sell more biofuel now. We see this tradeoff most clearly in $t = \{61, 63, 65\}$ where we see many red dots indicating a careful search via the CAVE algorithm. Note that we cannot directly compare Figure 4.7 to Figure 3.8 since we restricted $\max(R_{t}^{bio}) = 320,000$ in the bMDP due to our discretization of the state-space.

Let’s look at the post-decision value functions with respect to $R_{t}^{cap,x}$. We choose to show these particular months in Figure 4.8 since we make a decision on whether or not to expand or shrink our plant capacity at these months. We can see in months $t = \{1, 25, 49\}$, our value curves themselves encourage us to adjust our capacity.
levels to be roughly 24,000. Likewise, in months $t = \{13, 37, 61\}$, our value curves themselves encourage us to adjust our capacity levels to be roughly 12,000. This behavior is observed in our actual simulation shown in Figure 4.5.

To finalize our inspection of the value curves, we show the combined value function in both $R_{t}^{\text{bio},x}$ and $R_{t}^{\text{cap},x}$ for $t = \{1, 13, 25, 37, 61\}$. This is shown in Figure 4.9. The left axis (y axis) represents $R_{t}^{\text{bio},x}$ while the right axis (x axis) represents $R_{t}^{\text{cap},x}$. We can first appreciate the assumption of independent derivatives in all six value functions since all the derivatives with respect to $R_{t}^{\text{bio},x}$ are identical for the same value of $R_{t}^{\text{bio},x}$. A similar statement holds true for the derivatives with respect to $R_{t}^{\text{cap},x}$. A comparison between Figure 4.9 and Figure 4.8 will reveal that we are looking at the same curves in the x-direction.
4.4 Numerical Comparisons

In this section, we make comparisons on the different step-size schedule $\alpha^{(n)}$ for the CAVE algorithm and also on the bMDP and VFA policy. In this section, we deal with only $T = 48$ months.

4.4.1 Different Stepsizes

In the previous sections, we described step-size schedule $\alpha^{(n)}$ for the CAVE algorithm. We previously described $\alpha^{(n)} = c/(c+n-1)$, a harmonic step-size schedule. However, it is not clear whether or not this is the best step-size. We try two different step-size schedules as well where the step-sizes are still deterministic. One is $\alpha^{(n)} = 1/n$ where $n$ is the iteration number. This step-size results in averaging all the derivatives of the post-decision value function. The other is $\alpha^{(n)} = \alpha$, a constant step-size. The effect of choosing different step-sizes are shown in Figure 4.10. To get a rough idea of the effect of choosing a harmonic step-size schedule instead of a constant step-size schedule, when $n = 341$, $150/(150+n-1) \approx 0.3$. This means during the first 341 iterations, the harmonic step-size schedule with $c = 150$ is take larger steps than the constant step-size schedule. Afterwards the first 341 iterations, the reverse is true. At $n = 1000$, $150/(150+n-1) \approx 0.13$. Likewise, when $n = 94$, $40/(40+n-1) \approx 0.3$.

We cannot optimize the step-size in this model because of the stochastic nature of the model. Generally speaking, a constant step-size underperforms since the step-sizes are too large in the later iterations of the algorithms. However, a step-size schedule of $\alpha^{(n)} = 1/n$ also underperforms since the step-sizes shrink too fast. This means it will take a very long time for the VFA policy to converge upon the optimal decision.

To compare the resulting performance for each step-size, we run 10 trials our entire algorithm over 48 months for a fixed set of random values where each trial involves 1000 iterations. We then plot the mean cumulative profit at $t = T$ in
Figure 4.10: An illustration of the stepsizes $\alpha^{(n)}$ per iteration $n$.

Figure 4.11. As we can see, the harmonic step-size schedule with $c = 150$ (black) seems to perform the best. We see that it suddenly performs worse than the constant step-size schedule (blue) between iterations $n = 150$ and $n = 350$. This is most likely because the harmonic step-size schedule is taking too large of steps during this period compared to the constant step-size schedule. The $1/n$ step-size schedule (green) is always underperforming since it probably is always taking too little step-sizes to get close to the optimal solution. The summary statistics are shown in Table 4.1 where “Sd Objective” refers to the standard deviation of the objective function.

<table>
<thead>
<tr>
<th>Description</th>
<th>Mean Objective</th>
<th>Sd Objective</th>
</tr>
</thead>
<tbody>
<tr>
<td>.3 (Constant)</td>
<td>$10.02 \times 10^7$</td>
<td>$9.42 \times 10^6$</td>
</tr>
<tr>
<td>$\frac{1}{n}$</td>
<td>$8.43 \times 10^7$</td>
<td>$2.14 \times 10^6$</td>
</tr>
<tr>
<td>$40/(40+n-1)$ (Harmonic)</td>
<td>$9.34 \times 10^7$</td>
<td>$2.52 \times 10^6$</td>
</tr>
<tr>
<td>$150/(150+n-1)$ (Harmonic)</td>
<td>$10.83 \times 10^7$</td>
<td>$3.43 \times 10^6$</td>
</tr>
</tbody>
</table>

4.4.2 bMDP and VFA for Different Aggregation Functions

We have discussed three different polices across the previous chapter and this chapter. These are the bMDP policy, the VFA policy with independent derivatives, and the VFA policy with $g(S_t^x) = R_t^{bio,x} + R_t^{cap,x}$. We compare all three polices to see how our two VFA polices compare with the optimal bMDP policy. This benchmarking
Figure 4.11: A comparison of cumulative profit using different step sizes.

The procedure is based on the procedure described in Powell and Scott (2012). Based on our previous results, we utilize the step-size schedule $\alpha^{(n)} = 150/(150+n-1)$ for both VFA policies. To ensure that the two VFA polices do not have an advantage over the bMDP policy since the bMDP requires a discretized space of states and decisions, we enforce that the VFA polices can only select among the same set of decisions as the bMDP policy. The exogenous information are all discretized according to the bMDP policy, and each policy receives the same set of exogenous information during the evaluation. We set $T = 48$ and utilize the same discretization described in Section 3 with exception to $\max(R^{\text{bio}})$, which we expand to 40,000. We simulate all three polices on 100 trials and average their cumulative profit from $t = 1$ to $t = 48$. The results are shown in Figure 4.12. Furthermore, while the VFA policies still maintain their own specific value functions to make decisions, we always refer to the prices calculated by the bMDP. Since the bMDP has calculated the exact value at each
Figure 4.12: A comparison of the three different polices.

discretized state, we always can calculate this price. As we can see, bMDP (blue) performs the best. This is expected since bMDP applies a brute-force approach to simulate every possible state and exogenous information. This is why it is the most computational-intensive policy. Meanwhile, VFA using independent derivatives (green) performs second-best, and it does not do much worse that bMDP. The VFA using CAVE with the aggregation function $g(S_t^x) = R^{\text{bio},x}_t + R^{\text{cap},x}_t$ (red) seemed to perform the worst. Table 4.2 shows the summary statistics for overall performance and time for each of the three policies.

Table 4.2: Summary Statistics for Three Policies
<table>
<thead>
<tr>
<th>Description</th>
<th>Time Needed to Solve</th>
<th>Mean Objective</th>
<th>Sd Objective</th>
</tr>
</thead>
<tbody>
<tr>
<td>bMDP</td>
<td>13 min. 21 sec.</td>
<td>11.93 * 10^7</td>
<td>3.40 * 10^6</td>
</tr>
<tr>
<td>VFA with CAVE</td>
<td>2 min. 23 sec.</td>
<td>11.33 * 10^7</td>
<td>4.64 * 10^6</td>
</tr>
<tr>
<td>VFA with CAVE (Aggregated)</td>
<td>2 min. 5 sec.</td>
<td>8.71 * 10^7</td>
<td>2.63 * 10^6</td>
</tr>
</tbody>
</table>

We show a full comparison between all three policies in Figure 4.13. The top line (blue) shows the result of the bMDP policy. The middle line (green) shows the result of the VFA policy with independent derivatives. The bottom line (red) shows the result of the VFA policy with the aggregation function \( g(S_t^x) = R_{t}^{\text{bio},x} + R_{t}^{\text{cap},x} \).

As we can see in Figure 4.13, the bMDP and VFA polices differ in how much capacity to maintain (blue upward triangles). The bMDP tends to maintain a higher production capacity at the start of each cycle while the VFA tends to increase the production near the end of each cycle. Regardless, in the first two policies, there is always a period where we make more biofuel than is demanded. This goes into storage which is later utilized whenever there is a shock event or when demand is high. Unfortunately, in the very last policy, the VFA never seems to store any biofuel. This means that the aggregation function \( g(S_t^x) = R_{t}^{\text{bio},x} + R_{t}^{\text{cap},x} \) is not a good one. We surmise that using this aggregate function loses too much information on differentiating how much biofuel is actually in storage and how much biofuel can be produced.

### 4.5 Closing Remarks

In this chapter, we developed the VFA policy which hinged upon the CAVE algorithm, an algorithm that updates the post-decision value functions by exploiting concavity. We showed numerical results for the CAVE algorithm, and we performed experiments for varying step-size schedules and saw that the harmonic step-size schedule with \( c = 150 \) performed the best. We also compared the VFA policy to the bMDP policy.
to see if we lose much by using the faster but approximated VFA policy. We see through experimentation that if we represent the post-decision value function as a gradient of both $R_{t}^{\text{bio},x}$ and $R_{t}^{\text{cap},x}$ using independent derivatives, we achieve near-optimality defined by the bMDP. However, if we aggregated our post-decision state variable by $g(S_{t}^{x}) = R_{t}^{\text{bio},x} + R_{t}^{\text{cap},x}$, our policy suffered quite a bit.
Figure 4.13: A comparison of the three simulations.
Chapter 5

Conclusion

We have laid out the foundational machinery used in many approximate dynamic programming settings and developed a model to simulate the gasoline certificate market from the standpoint of biofuel producers in the thesis. In Section 1, we introduced various methods to promote a smarter usage of energy within the US by citing previous methods and research associated with these methods. In Section 2, we introduced the basics of modeling and presented a gasoline certificate market model. In Section 3, we introduced the backward Markov Decision Process to solve for the optimal policy on a discretized space. To do this, we explained the value function. In Section 4, we used the post-decision value function and adopted the Value Function Approximation (VFA) policy which utilizes the Concave, Adaptive Value Estimator. Using this method, we get an “approximate” optimal policy but the computation is much faster. We see numerically our value function approximation policy under an appropriate representation of value function does almost as well as the backward Markov Decision Process.

Future work in this model could include any of the following tasks.

- Changing the contribution function where the price of storing biofuel, producing biofuel or increasing/decreasing production capacity is not linear: We did
not pursue this direction in this thesis since the linear function kept the value function concave. If we had decreasing marginal costs, the value function would no longer be concave. Recall that we currently have

\[ C(S_t, x_t) = p_t x_t^{\text{bio}} - \mathbb{I}[x_t^{\text{cap}} \geq 0] c^{\text{cap}} x_t^{\text{cap}} + \mathbb{I}[x_t^{\text{cap}} < 0] c^{\text{salv}} c^{\text{cap}} |x_t^{\text{cap}}| \]

\[ - c^{\text{pro}} x_t^{\text{pro}} - c^{\text{bio}} (P_t^{\text{bio}} + x_t^{\text{pro}} - x_t^{\text{bio}})^+. \]

Had we changed this to

\[ C(S_t, x_t) = p_t x_t^{\text{bio}} - \mathbb{I}[x_t^{\text{cap}} \geq 0] (x_t^{\text{cap}})^{\text{cap}} + \mathbb{I}[x_t^{\text{cap}} < 0] c^{\text{salv}} |x_t^{\text{cap}}|^{\text{cap}} \]

\[ - (x_t^{\text{pro}})^{\text{pro}} - [(P_t^{\text{bio}} + x_t^{\text{pro}} - x_t^{\text{bio}})^+]^{\text{bio}}. \]

for \(0 < c^{\text{cap}}, c^{\text{pro}}, c^{\text{bio}} < 1\), we would have achieved decreasing marginal costs. However, since \(x^y\) for \(0 < y < 1\) is concave, \(-x^y\) is not concave.

• Calibrating the model: We did not pursue this direction in this thesis since the cyclical demand was the most effective demand we found to demonstrate the strength of the different polices. In reality, the demand for gasoline is not as cyclical as we have shown in this thesis. Rather, the demand is somewhat stationary. This resulted in somewhat uninteresting results that we have chosen not to include in this thesis. Perhaps the model needs to be more adapted to give rise to more interesting solutions despite a somewhat stationary demand.

• Changing the regulation structure: We presented a very simple regulation structure in this thesis that involved only two parameters, \(\rho\) (the allowed amount of gasoline that was sold before regulation) and \(\gamma\) (the price of the penalty). Other regulation structures could promote different type of solutions. We did not pursue this direction in this thesis since we wanted to try the simplest regulatory structure for this foundational work.
• Using other approximate dynamic programming techniques: We utilized the value function approximation (VFA) with concave, adaptive value estimators (CAVE) in this thesis, but there are many other policies that could present near-optimal policies as well. These include lookahead policies and policy function approximation. We did not pursue this direction in our thesis since the VFA with CAVE was the most intuitive algorithm to use for this setting, and as we have seen in this thesis, it performs quite well in our stated model.

• Expanding the model: Currently, our state is 4-dimensional, our decision is 3-dimensional, and our exogenous information is 2-dimensional. Expanding our model to higher dimensions would give rise to a deeper study on various aggregate functions that are required to solve the policy. We did not pursue this direction in this thesis since we already saw the necessity in VFA with our basic model (Model 2). We saw that finding a finer solution (smaller discretization) in the bMDP required roughly a 64 times increase in computation time, and this made it unreasonable to reliably use bMDP to solve our model.
Appendix A

Code

There is a lot of code used in the creation of this thesis. All the code was written in Matlab. In this chapter, we provide just the code for simulating the backward Markov Decision Process (bMDP) on the discretized state-space described in Chapter 3. The other code can be reproduced upon request.

A.1 Helper Functions

Below is the function to calculate the probability of demand of the next time step.

```matlab
function [probb, midpoints] = calcProbVect(demandmean, demandint, DEMANDMIN, DEMANDNOISE, DEMANDSPACE)
    midpoints = zeros(demandint-1,1);
    probvect = zeros(demandint,1);
    for i = 1:(demandint-1)
        midpoints(i) = DEMANDMIN+(i-1+0.5)*DEMANDSPACE;
    end
    probvect(1) = normcdf(midpoints(1),demandmean,DEMANDNOISE);
    for i = 2:(demandint-1)
        probvect(i) = normcdf(midpoints(i),demandmean,DEMANDNOISE)-normcdf(midpoints(i-1), demandmean, DEMANDNOISE);
    end
    probvect(demandint) = 1-sum(probbvext(1:(demandint-1)));
end
```

Below is the function to calculate the price based on the value functions.

```matlab
function [price] = calcPrice2(demandmean, nextbioidx, nextcapidx, valmatrix, demandint, DEMANDMIN, DEMANDNOISE, DEMANDSPACE, PROB1, PROB1_4, rbioint, RBIOSPACE)
% the price function for model 2 where there's an indicator variable
```
probvect = calcProbVect(demandmean, demandint, DEMANDMIN, DEMANDNOISE, DEMANDSPACE);

if nextbioidx == 1
    diff1a = valmatrix(:, nextbioidx+1, nextcapidx, 1) - valmatrix(:, nextbioidx, nextcapidx, 1);
    diff1b = valmatrix(:, nextbioidx+1, nextcapidx, 2) - valmatrix(:, nextbioidx, nextcapidx, 2);
    price = PROB10*dot(probvect, diff1a)/RBIOSPACE + PROB11*dot(probvect, diff1b)/RBIOSPACE;
else if nextbioidx == rbioint
    diff2a = valmatrix(:, nextbioidx, nextcapidx, 1) - valmatrix(:, nextbioidx-1, nextcapidx, 1);
    diff2b = valmatrix(:, nextbioidx, nextcapidx, 2) - valmatrix(:, nextbioidx-1, nextcapidx, 2);
    price = PROB10*dot(probvect, diff2a)/RBIOSPACE + PROB11*dot(probvect, diff2b)/RBIOSPACE;
else
    diff1a = valmatrix(:, nextbioidx+1, nextcapidx, 1) - valmatrix(:, nextbioidx, nextcapidx, 1);
    diff2a = valmatrix(:, nextbioidx, nextcapidx, 1) - valmatrix(:, nextbioidx-1, nextcapidx, 1);
    diff1b = valmatrix(:, nextbioidx+1, nextcapidx, 2) - valmatrix(:, nextbioidx, nextcapidx, 2);
    diff2b = valmatrix(:, nextbioidx, nextcapidx, 2) - valmatrix(:, nextbioidx-1, nextcapidx, 2);
    diffa = .5*diff1a + .5*diff2a;
    diffb = .5*diff1b + .5*diff2b;
    price = PROB10*dot(probvect, diffa)/RBIOSPACE + PROB11*dot(probvect, diffb)/RBIOSPACE;
end
end

A.2 Backward Markov Decision Process Script

Below is the entire script to run the backward Markov Decision Process (bMDP) using the previously-stated helper functions.

TIMESTEPS = 48;
PARAPRO = 0;
PARAGOV = 30;
COSTCAP = 10;
COSTPRO = 3;
COSTBIO = 1;

PROB01 = .1;
PROB11 = .5;
PROB00 = 1-PROB01;
PROB10 = 1-PROB11;

DEMANDA = 200000;
DEMANDB = pi/12;
DEMANDC = 300000;
DEMANDNOISE = 1500*30;
DEMANDMIN = 100000;
DEMANDMAX = 500000;
DEMANDSPACE = 80000;
RBIOMAX = 320000; %amount bio stored
RBIOMIN = 0;
RBIOSPACE = DEMANDSPACE;
RCAPMAX = 320000; %production capacity (think: storage capacity)
RCAPMIN = 0;
RCAPSPACE = DEMANDSPACE;
ACAPMAX = 160000; %how much capacity to add
ACAPMIN = -160000;
ACAPSPACE = DEMANDSPACE;
APROMAX = RCAPMAX; %how much bio to produce
APROMIN = RCAPMIN;
APROSPACE = DEMANDSPACE;
ABIOMAX = RBIOMAX+APROMAX; %how much bio to sell
ABIOMIN = RBIOMIN;
ABIOSPACE = DEMANDSPACE;

LARGENEG = -9999999;

%%
%% initialization
%%— determine spacing

demandint = round((DEMANDMAX-DEMANDMIN)/DEMANDSPACE)+1;
rbioint = round((RBIOMAX-RBIOMIN)/RBIOSPACE)+1;
rcapint = round((RCAPMAX-RCAPMIN)/RCAPSPACE)+1;
acapint = round((ACAPMAX-ACAPMIN)/ACAPSPACE)+1;
aproint = round((APROMAX-APROMIN)/APROSPACE)+1;
abioint = round((ABIOMAX-ABIOMIN)/ABIOSPACE)+1;

valuenamelist = cell(TIMESTEPS,1);
acapnamelist = cell(floor(TIMESTEPS/12),1);
apronamelist = cell(TIMESTEPS,1);
abionamelist = cell(TIMESTEPS,1);
pricenamelist = cell(TIMESTEPS,1);

for i=1:TIMESTEPS
    valuenamelist{i} = sprintf('valuematrix%d',i);
apronamelist{i} = sprintf('apromatrix%d',i);
abionamelist{i} = sprintf('abiomatrix%d',i);
pricenamelist{i} = sprintf('pricematrix%d',i);
end

for i=1:floor(TIMESTEPS/12)
    acapnamelist{i} = sprintf('acapmatrix%d',i);
end

tmpvalmatrix = zeros(demandint,rbioint,rcapint,2);
tmpapromatrix = zeros(demandint,rbioint,rcapint,2);
tmpbiomatrix = zeros(demandint,rbioint,rcapint,2);
tmppricematrix = zeros(demandint,rbioint,rcapint,2);

%— solve for the last time step

for i = 1:demandint
    currentdemand = DEMANDEMIN+(i-1)*DEMANDSPACE;
    for j = 1:rbioint
        for k = 1:rcapint
            tmpmax = LARGENEG;
            optapro = -1;

            for l = 1:acapint
                tmpmax = max(tmpmax,tmpvalmatrix{i,j,k,l});
            end

            for l = 1:acapint
                tmpmax = max(tmpmax,tmpapromatrix{i,j,k,l});
            end

            for l = 1:acapint
                tmpmax = max(tmpmax,tmpbiomatrix{i,j,k,l});
            end

            for l = 1:acapint
                tmpmax = max(tmpmax,tmppricematrix{i,j,k,l});
            end

            if currentdemand > tmpmax
                optapro = l;
                break;
            end
        end
    end

    tmpvalmatrix{i}(:,:,optapro) = currentdemand;
end

end
\[
\text{optabio} = -1;
\]

%we would never build capacity in the last iteration, so ignore
%the possibility
for \( a = 1 : \text{aproint} \)
\[
\text{atmp} = a;
\]
%you can't produce more than your capacity+stored
%amount allows
if \( (\text{APROMIN}+(a-1)\times\text{APROSPACE})+(\text{RBIOMIN}+(j-1)\times\text{RBIOSPACE}) > 2\times(\text{RCAPMIN}+(k-1)\times\text{RCAPSPACE}) \)
\[
\text{atmp} = \text{floor} \left( \frac{(2\times(\text{RCAPMIN}+(k-1)\times\text{RCAPSPACE})-\text{RBIOMIN}+(j-1)\times\text{RBIOSPACE})}{\text{APROMIN}/\text{APROSPACE}} \right) + 1;
\]
end
if \( (\text{APROMIN}+(\text{atmp}-1)\times\text{APROSPACE}) > (\text{RCAPMIN}+(k-1)\times\text{RCAPSPACE}) \)
\[
\text{atmp} = \text{floor} \left( \frac{(\text{RCAPMIN}+(k-1)\times\text{RCAPSPACE})-\text{APROMIN}}{\text{APROSPACE}} \right) + 1;
\]
end

for \( b = 1 : \text{abioint} \)
\[
\text{btmp} = b;
\]
%you can't sell more than your production + amount in
%storage
if \( (\text{ABIOMIN}+(b-1)\times\text{ABIOSPACE}) > (\text{RBIOMIN}+(j-1)\times\text{RBIOSPACE})+(\text{APROMIN}+(\text{atmp}-1)\times\text{APROSPACE}) \)
\[
\text{btmp} = \text{floor} \left( \frac{(\text{RBIOMIN}+(j-1)\times\text{RBIOSPACE})+(\text{APROMIN}+(\text{atmp}-1)\times\text{APROSPACE})-\text{ABIOMIN}}{\text{ABIOSPACE}} \right) + 1;
\]
end
%you can't sell more than demand
if \( (\text{ABIOMIN}+(\text{btmp}-1)\times\text{ABIOSPACE}) > (\text{DEMANDMIN}+(i-1)\times\text{DEMANDSPACE}-\text{PARAPRO}) \)
\[
\text{btmp} = \text{floor} \left( \frac{(\text{DEMANDMIN}+(i-1)\times\text{DEMANDSPACE}-\text{PARAPRO})-\text{ABIOMIN}}{\text{ABIOSPACE}} \right) + 1;
\]
end

%price = \( \text{DEMANDA}\times\cos(\text{DEMANDB}\times t) + \text{DEMANDC} \)
\[
\text{nextbio} = \max(0,(\text{RBIOMIN}+(j-1)\times\text{RBIOSPACE})+(\text{APROMIN}+(\text{atmp}-1)\times\text{APROSPACE})-(\text{ABIOMIN}+(\text{btmp}-1)\times\text{ABIOSPACE})));
\]
\[
\text{currentpro} = (\text{APROMIN}+(\text{atmp}-1)\times\text{APROSPACE});
\]
\[
\text{contr} = \text{PARACOV}+(\text{ABIOMIN}+(\text{btmp}-1)\times\text{ABIOSPACE}) - \text{COSTBIO}\times\text{nextbio};
\]
if \( \text{contr} > \text{tmpmax} \)
\[
\text{tmpmax} = \text{contr};
\]
\[
\text{optapro} = \text{atmp};
\]
\[
\text{optabio} = \text{btmp};
\]
end
end
\[
\text{tmpvalmatrix}(i,j,k,:) = \text{tmpmax};
\]
\[
\text{tmpapromatrix}(i,j,k,:) = \text{optapro};
\]
\[
\text{tmpbiomatrix}(i,j,k,:) = \text{optabio};
\]
\[
\text{tmppricematrix}(i,j,k,:) = \text{PARACOV};
\]
end
end
end
assignin ('base', char(valuename_list(TIMESTEPS)), tmpvalmatrix);
assignin ('base', char(aproname_list(TIMESTEPS)), tmpapromatrix);
assignin ('base', char(abioname_list(TIMESTEPS)), tmpabiomatrix);
assignin ('base', char(pricename_list(TIMESTEPS)), tmppricematrix);

%%% solve backwards
\[ t = \text{TIMESTEPS} - 1; \]
while \( t > 0 \)
nextdemanddiff = DEMANDA*(cos(DEMANDB*(t+1)) - cos(DEMANDB*(t))) - DEMANDC;
if mod(t-1,12)==0
    tmpvalmatrix = zeros(demandint, rbioint, rcapint);
    tmpapromatrix = zeros(demandint, rbioint, rcapint);
    tmpabiomatrix = zeros(demandint, rbioint, rcapint);
    tmpcapmatrix = zeros(demandint, rbioint, rcapint);
    tmppricematrix = zeros(demandint, rbioint, rcapint);
    for i = 1:demandint
        currentdemand = DEMANDMIN+(i-1)*DEMANDSPACE;
        for j = 1:rbioint
            for k = 1:rcapint
                h=1:2
                tmpmax = LARGENEG;
                optapro = -1;
                optabio = -1;
                optacap = -1;
                optprice = -1;
                for a = 1:apoint
                    atmp = a;
                    if (APROMIN+(a-1)*APROSPACE)+(RBIOMIN+(j-1)*RBIOSPACE) > 2*(RCAPMIN+(k-1)*RCAPSPACE)
                        atmp = floor((2*(RCAPMIN+(k-1)*RCAPSPACE) - RBIOMIN+(j-1)*RBIOSPACE - APROMIN) / APROSPACE) + 1;
                    end
                    if (APROMIN+(atmp-1)*APROSPACE) > (RCAPMIN+(k-1)*RCAPSPACE)
                        atmp = floor(((RCAPMIN+(k-1)*RCAPSPACE) - APROMIN) / APROSPACE) + 1;
                    end
                end
                for b = 1:aboint
                    btmp = b;
                    if (ABIMIN+(b-1)*ABIOSPACE) > (RBIOMIN+(j-1)*RBIOSPACE)+(APROMIN+(atmp-1)*APROSPACE)
                        btmp = floor(((RBIOMIN+(j-1)*RBIOSPACE) + (APROMIN+(atmp-1)*APROSPACE) - ABIOMIN) / ABIOSPACE) + 1;
                    end
                    if (ABIMIN+(btmp-1)*ABIOSPACE) > (DEMANDMIN+(i-1)*DEMANDSPACE-PARAPRO)
                        btmp = floor((DEMANDMIN+(i-1)*DEMANDSPACE-PARAPRO-ABIMIN) / ABIOSPACE) + 1;
                    end
                end
            end
        end
    end
end

93
nextbio = (RBIOMIN + (j-1) * RBIOSPACE) - (ABIOMIN + (btmp-1) * ABIOSPACE) + (APROMIN + (atmp-1) * APROSPACE);

nextj = floor((nextbio - RBIOMIN) / RBIOSPACE) + 1;
if nextj > rbioint
    nextj = rbioint;
end

% NEW: it is possible to sell capacity now
for c = 1:acapint
    ctmp = c;
    currentcap = RCAPMIN + (k-1) * RCAPSPACE;
    nextcap = currentcap + ACAPMIN + (ctmp-1) * ACAPSPACE;
    if nextcap < RCAPMIN
        ctmp = floor((currentcap - RCAPMIN - ACAPMIN) / ACAPSPACE) + 1;
        nextcap = RCAPMIN;
    end
    if nextcap > RCAPMAX
        ctmp = floor((ACAPMAX - (RCAPMAX - currentcap)) / ACAPSPACE) + 1;
        nextcap = RCAPMAX;
    end
end

% can’t have more than RCAPMIN for capacity
nextk = floor((nextcap - RCAPMIN) / RCAPSPACE) + 1;

if h == 2
    price = calcPrice2(nextdemanddiff + currentdemand, nextj, k, eval(char(valuename_list(t+1))), demandint, DEMANDMIN, DEMANDNOISE, DEMANDSPACE, PROB1_0, PROB1_1, rbioint, RBIOSPACE);
    if price > PARAGOV
        price = PARAGOV;
    end
else
    price = PARAGOV;
end

nextbio = max(0, ((RBIOMIN + (j-1) * RBIOSPACE) + (APROMIN + (atmp-1) * APROSPACE)) - (ABIOMIN + (btmp-1) * ABIOSPACE));
costbuild = -COSTCAP * (ACAPMIN + (c-1) * ACAPSPACE);
if costbuild > 0
    costbuild = .6 * costbuild;
end
currentpro = (APROMIN + (a-1) * APROSPACE);

contr = price * (ABIOMIN + (btmp-1) * ABIOSPACE) - COSTPRO * currentpro - COSTBIO * nextbio + costbuild;
probvect = calcProbVect(nextdemandmean, demandint, DEMANDMIN, DEMANDNOISE, DEMANDSPACE);
tmpnextvalmatrix = eval(char(valuename_list(t+1)));
if h == 1 % %market unsaturated
    nextval = PROB0_0 * dot(tmpnextvalmatrix(:, nextj, nextk, 1), probvect) + PROB0_1 * dot(tmpnextvalmatrix(:, nextj, nextk, 2), probvect);
end
else 

nextval = PROB1*dot(tmpnextvalmatrix(:,nextj,nextk,1),
probvect) + PROB2*dot(tmpnextvalmatrix(:,nextj,nextk
,2),probvect);
end

if contr+nextval == tmpmax && price > optprice
    tmpmax = contr+nextval;
    optapro = atmp;
    optabio = btmp;
    optacap = ctmp;
    optprice = price;
end

if contr+nextval > tmpmax
    tmpmax = contr+nextval;
    optapro = atmp;
    optabio = btmp;
    optacap = ctmp;
    optprice = price;
end

end
end
end
end
end

assignin('base',char(valuename_list(t)),tmpvalmatrix);
assignin('base',char(aproname_list(t)),tmpapromatrix);
assignin('base',char(abiomname_list(t)),tmpabiomatrix);
assignin('base',char(pricename_list(t)),tmppricematrix);

tmp = ceil(t/12);
assignin('base',char(acapname_list(tmp)),tmpcapmatrix);

else
    tmpvalmatrix = zeros(demandint,rbioint,rcapint);
    tmpapromatrix = zeros(demandint,rbioint,rcapint);
    tmpabiomatrix = zeros(demandint,rbioint,rcapint);
    tmppricematrix = zeros(demandint,rbioint,rcapint);
    for i = 1:demandint
        for j = 1:rbioint
            for k = 1:rcapint
                for h = 1:2
                    tmpmax = LARGENEG;
                end
            end
        end
    end
end
optapro = -1;
optabio = -1;
optprice = -1;

%we would never build capacity in the last iteration, so ignore
%the possibility
for a = 1:aprint
    atmp = a;
    if (APROMIN+(a-1)*APROSPACE) +(RBIOMIN+(j-1)*RBIOSPACE) > 2*(RCAPMIN+(k-1)*RCAPSPACE)
        atmp = floor((2*(RCAPMIN+(k-1)*RCAPSPACE) - RBIOMIN+(j-1)*RBIOSPACE - APROMIN)/APROSPACE) +1;
    end
    if (APROMIN+(atmp-1)*APROSPACE) > (RCAPMIN+(k-1)*RCAPSPACE)
        atmp = floor((RCAPMIN+(k-1)*RCAPSPACE) - APROMIN)/APROSPACE) +1;
    end
    nextbio = (RBIOMIN+(j-1)*RBIOSPACE) - (ABIOMIN+(atmp-1)*ABIOSPACE) + (APROMIN+(atmp-1)*APROSPACE);
    nextj = floor((nextbio - RBIOMIN)/RBIOSPACE) +1;
    if nextj > rbioint
        nextj = rbioint;
    end
    nextk = k;
    if h == 2
        %price = DEMANDA*(cos(DEMANDB*t)) + DEMANDC;
        price = calcPrice2(nextdemanddiff+currentdemand, nextj, nextk, eval(char(valuename_list(t+1))), demandint, DEMANDMIN, DEMANDNOISE, DEMANDSPACE, PROB1_0, PROB1_1, rbioint, RBIOSPACE);
        if price > PARAGOV
            price = PARAGOV;
        end
    else
        price = PARAGOV;
    end
    nextbio = max(0, (RBIOMIN+(j-1)*RBIOSPACE) + (APROMIN+(atmp-1)*APROSPACE) - (ABIOMIN+(atmp-1)*ABIOSPACE));
    currentpro = (APROMIN+(a-1)*APROSPACE);
\[
\text{contr} = \text{price} \ast (\text{ABIOMIN} \ast (\text{btmp} - 1) \ast \text{ABIOSPACE}) - \text{COSTPRO} \ast \text{currentpro} - \\
\text{COSTBIO} \ast \text{nextbio};
\]

\[
\text{probvect} = \text{calcProbVect}(\text{nextdemandmean}, \text{demandint}, \text{DEMANDMIN}, \\
\text{DEMANDNOISE}, \text{DEMANDSPACE});
\]

\[
\text{tmpnextvalmatrix} = \text{eval}([\text{valuename_list} \{t + 1\}]);
\]

if \( h = 1 \) % %market unsaturated

\[
\text{nextval} = \text{PROB0} \ast \text{dot}(\text{tmpnextvalmatrix}(:, \text{nextj}, \text{nextk}, 1), \\
\text{probvect}) + \text{PROB0} \ast \text{dot}(\text{tmpnextvalmatrix}(:, \text{nextj}, \text{nextk}, 2), \\
\text{probvect});
\]

else % %market saturated

\[
\text{nextval} = \text{PROB1} \ast \text{dot}(\text{tmpnextvalmatrix}(:, \text{nextj}, \text{nextk}, 1), \\
\text{probvect}) + \text{PROB1} \ast \text{dot}(\text{tmpnextvalmatrix}(:, \text{nextj}, \text{nextk}, 2), \\
\text{probvect});
\]

end

if \( \text{contr} \ast \text{nextval} = \text{tmpmax} \&\& \text{price} > \text{optprice} 

\[
\text{tmpmax} = \text{contr} \ast \text{nextval};
\]

\[
\text{optapro} = \text{atmp};
\]

\[
\text{optabio} = \text{btmp};
\]

\[
\text{optprice} = \text{price};
\]

end

if \( \text{contr} \ast \text{nextval} > \text{tmpmax} 

\[
\text{tmpmax} = \text{contr} \ast \text{nextval};
\]

\[
\text{optapro} = \text{atmp};
\]

\[
\text{optabio} = \text{btmp};
\]

\[
\text{optprice} = \text{price};
\]

end

end

\[
\text{tmpvalmatrix}(i, j, k, h) = \text{tmpmax};
\]

\[
\text{tmpapromatrix}(i, j, k, h) = \text{optapro};
\]

\[
\text{tmpabiomatrix}(i, j, k, h) = \text{optabio};
\]

\[
\text{tmppricematrix}(i, j, k, h) = \text{optprice};
\]

end

end

assignin('base', char( valuename_list \{t\} ), tmpvalmatrix);

assignin('base', char( approname_list \{t\} ), tmpapromatrix);

assignin('base', char( abioname_list \{t\} ), tmpabiomatrix);

assignin('base', char( pricename_list \{t\} ), tmppricematrix);

end

if mod(t, floor(TIMESTEPS/10)) == 0

disp('**')
end

t = t - 1;
end
Bibliography


Hansen, J. (2006). Global warming: is there still time to avoid disastrous human-made climate change. Have we passed a ‘tipping point’? *National Academy of Science Discussion Series*.


Powell, W. B. and Scott, W. R. (2012). Approximate dynamic programming for energy storage with new results on instrumental variables and projected Bellman errors. *In Review*.


