Applications of Optimization to Shale Oil and Gas Monetization

by

Siah Hong Tan

B.S., Johns Hopkins University (2011)
S.M., Massachusetts Institute of Technology (2013)

Submitted to the Department of Chemical Engineering
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Chemical Engineering Practice

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2017

Author .....
Department of Chemical Engineering
May 5, 2017

Certified by .....
Paul I. Barton
Lammot du Pont Professor of Chemical Engineering
Thesis Supervisor

Accepted by .....
Daniel Blankschtein
Herman P. Meissner (1929) Professor of Chemical Engineering
Chairman, Committee for Graduate Students

Applications of Optimization to Shale Oil and Gas Monetization

by

Siah Hong Tan

Submitted to the Department of Chemical Engineering
on May 5, 2017, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Chemical Engineering Practice

Abstract

This thesis addresses the challenges brought forth by the shale oil and gas revolution through the application of formal optimization techniques. Two frameworks, each addressing the monetization of shale oil and gas resources at different ends of the scale spectrum, are developed. Importantly, these frameworks accounted for both the dynamic and stochastic aspects of the problem at hand.

The first framework involves the development of a strategy to allocate small-scale mobile plants to monetize associated or stranded gas. The framework is applied to a case study in the Bakken shale play where large quantities of associated gas are flared. Optimal strategies involving the continuous redeployment of plants are analyzed in detail. The value of the stochastic solution with regards to uncertainty in resource availability is determined and it indicates that mobile plants possess a high degree of flexibility to handle uncertainty.

The second framework is a comprehensive supply chain optimization model to determine optimal shale oil and gas infrastructure investments in the United States. Assuming two different scenario sets over a time horizon of twenty-five years, the features of the optimal infrastructure investments and associated operating decisions are determined. The importance of incorporating uncertainty into the framework is demonstrated and the relationship between the stability of the stochastic solution and the variance of the distribution of future parameters is analyzed.

The thesis also analyzes the Continuous Flow Stirred Tank Reactor (CFSTR) equivalence principle as a method for screening and targeting favorable reaction pathways, with applications directed towards gas-to-liquids conversion. The principle is

found to have limited usefulness when applied to series reactions due to an unphysical independence of the variables which allows for the maximization of production of any intermediate species regardless of the magnitude of its rate of depletion. A reformulation which eliminates the unphysical independence is proposed. However, the issue of arbitrary truncation of downstream reactions remains.

Thesis Supervisor: Paul I. Barton
Title: Lammot du Pont Professor of Chemical Engineering

Dedicated to my loving parents.

Acknowledgments

This thesis is the product of a labor of love. Nevertheless, its completion could not have been achieved without the help of several key individuals. First, I would like to thank my research advisor, Professor Paul Barton. What I appreciate most about Paul is that he has given me the freedom to chart my own path in research while ensuring that the quality of research is kept to high standards. To him, no question is too challenging to tackle, and he places his trust and patience in me to find a way. Through this process, I have grown to become a more fearless and thoughtful researcher who takes pride in getting things done right.

I would also like to thank my thesis committee members, Professors Robert Armstrong and William Green, who have provided me with invaluable feedback on my work and gave me advice on how to proceed at each stage of this journey.

My thanks goes out to my fellow lab members at the Process Systems Engineering Laboratory. I appreciate the camaraderie and intellectual discussions that very much make our lab a special place. Special thanks goes out to Harry, Jose, and Rohit, who started the program with me. They are not only great labmates, but also great friends outside of the lab.

My thesis could not have been complete without the support of a close group of friends. Karthick and Sue Zanne, thanks for reminding me of home with our *makan* sessions. Kenneth, Marcus, Qinyi and Sheng Rong, thanks for all the fun and laughter during my early years. My thanks also goes out to Boon, Jon, Mila, Ming Qing and Rushabh for being great company at various stages of my time at MIT. Here's also a shout-out to my core team at Sloan - Gal, Galen, Maddie, Michael, Priyanka and Scott - fine individuals with whom I had the pleasure of working. Finally, my heartfelt thanks goes out to my good buddy Adriano, who taught me more about

life and business than the classroom ever could.

The MIT ecosystem of faculty and students is truly unparalleled in quality. Brilliance can be found in every corner of the campus, be it in the classrooms, labs or lecture halls. Yet, this brilliance is often tempered with kindness and humility. I feel myself very fortunate to have been a part of this community. Definitely, I have grown to become a better version of myself in more ways than one. To all my coursemates, professors, lecturers and administrative and support staff with whom I have ever crossed paths, thank you for all that you have done.

Finally, thank you, Dominique, for being all that you are. To my parents, thank you for all the love and support you have given me not only for these few years, but for my entire life. Thank you for believing in me. This thesis is dedicated to you.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background

The main theme of this thesis is the application of optimization techniques with the aim of optimal monetization of oil and gas from unconventional resources. The discovery of large reserves of shale oil and gas in many locations worldwide and the technological advances that have made it possible to exploit them has presented an unprecedented economic opportunity. This revolutionary development in the global energy arena has been led largely by activity in the United States. From 2008 to 2013, U.S. crude oil production grew from 5.0 to 7.4 million barrels a day, and U.S. dry natural gas production grew from 20.2 to 24.3 trillion cubic feet per year [5].

The work of this thesis began at a time when the shale revolution was transitioning from a focus on upstream activities towards that on mid- and downstream activities. Great strides in increasing the efficiency of shale oil and gas production and the management of production sites were already being made, and the question of the day turned towards analyzing the possible ways in which the growing

abundance of shale oil and gas could be monetized.

The analysis of shale oil and gas monetization involves the simultaneous consideration of numerous factors, including technological options, scales of production, availability of existing infrastructure, nature of production at sources, nature of markets, economic parameters of prices, demand and supply of resources and end products, transportation options, time horizon, and uncertainty of future conditions. With the multitude of options, heuristic approaches to determine the best options for monetization are likely to lead to sub-optimal results.

The aim of this thesis seeks to address this issue through formal optimization techniques. Formalized optimization frameworks can handle complexity with relative ease. The development of efficient algorithms and formulations, combined with the rapid increase in computational power of modern computers over the past decades, has now made it possible to solve once-intractable problem instances containing tens of millions of variables and constraints. The holistic yet granular nature of an optimization framework allows it to uncover optimal solutions which might not be accessible by heuristic-based analyses.

The heart of this thesis' analysis is centered around the twin issues of scale and risk and their associated trade-offs. On one hand, operating at larger scales allows for the benefits of economies of scale. On the other, the large capital outlay and lengthy development times might pose a significant risk for investors due to considerable uncertainty in the future demand, supply and prices at the time at which an investment decision is being made.

As the scale of operation decreases, there is a distinct discontinuity in the conceptualization of the mode of operation of plants. The main difference is that below a certain operational scale, individual plants gain the ability to be moved to different locations during their lifetime. This discontinuity in the mode of operation

leads to distinct formulations of the optimization models and their specific realms of application. As such, the work of this thesis is organized around this distinction.

The application of optimization techniques to solve problems related to shale oil or gas in the research community is still relatively new. A brief overview of the relevant studies performed thus far are given here.

Martín and Grossmann [6] presented a superstructure optimization approach to produce liquid fuels and hydrogen from switchgrass and shale gas in a facility. Cafaro and Grossmann [7] optimized the design and operation of supply chain networks at shale gas drilling sites. Yang *et al.* [8] optimized water management operations during shale gas production to maximize profits. An extension to consider both strategic design decisions and environmental objectives in the optimization of water management during shale gas production was provided by Gao and You [9]. The same authors also optimized a shale gas and water supply chain network from well sites to power plants and performed a life-cycle analysis of electricity generated from shale gas [10].

In a series of papers, Knudsen and Foss [11] optimized the production from a set of late-life wells at a shared production pad to avoid well liquid-loading. The formulation was extended to multiple pads and solved using a Lagrangian relaxation based decomposition scheme [12]. Later, the scheduling problem was extended to consider the operation of the wells to supply electric power [13]. Bistline [14] explored how uncertainties in natural gas prices and future climate policies impacted economic and environmental outcomes in the U.S. power sector with a two-stage stochastic programming formulation.

1.2 Small-scale mobile plants

Large upfront investments preclude the tapping of stranded gas reserves, which are reserves that are either too small or too physically inaccessible to be economically exploitable. A recent survey by Attanasi and Freeman [15] of the gas fields in the world excluding the U.S. estimated that only around 12.2% of the gas fields tabulated were larger than 1.54 tcf in size. In contrast, as indicated by Velocys [16], the remaining fields which would be considered too small to monetize by traditional large-scale technologies might be accessible to medium- to small-scale technologies.

Stranded gas can also arise from the lack of infrastructure access despite the field having a large size. A pertinent example in the United States is the gas associated with the production of shale oil at liquids-rich fields, such as the Bakken shale field in North Dakota. In 2013, Ford and Davis [17] estimated that 33% of the natural gas produced at the Bakken was not marketed, where most gas not marketed was flared.

Recent concepts for implementing gas-to-liquids (GTL) and liquefied natural gas (LNG) technology at a small-scale and modular level have the game-changing potential to shift the paradigm away from large capital expenditures and one fixed location. These proposed plants are currently in the early stages of commercialization by several companies in the oil and gas industry, including GE Oil & Gas [18], CompactGTL [19] and Velocys [20]. These technologies involve pre-manufacturing each process unit as compartmentalized, individual modules which can then be shipped to the site of interest and assembled together in minimal time to form the entire plant. Additionally, plants can be quickly disassembled into their individual modules and redeployed at other sites, affording them the benefit of mobility. This mobility will allow them to respond quickly to changes in conditions that might affect their prof-

itability. This could include economic factors such as large changes in the price of both the raw gas and its associated products, and supply shocks arising from the steep decline curves typically observed with unconventional sources of gas. For example, a study by Hughes in 2013 [21] concluded that wells in the top five U.S. shale plays typically produced 80-95% less gas after three years. Although the commercial availability of modular plants is limited at the time of writing, there has been a growing interest in evaluating them for purposes of monetizing stranded or associated gas from both conventional and nonconventional sources.

In view of these promising claims, it would be both useful and informative for industry players to have access to a framework that optimally utilizes these small-scale, mobile technologies to monetize stranded or associated gas. To this aim, this thesis develops a multi-period strategy for the optimal allocation of these technologies under time-varying supplies of gas in locations where stranded or associated gas is present and time-varying prices of and demand for the various products in their respective markets.

Prior to this work, the application of optimization to analyze small-scale mobile plants has not been investigated in literature, to our knowledge. However, ideas on problem formulation are similar in the context of the unit commitment problem. For example, in the case of a hydro-thermal system, decisions are made to operate thermal units and pumped hydro storage plants, which can be turned off and on at certain time points [22, 23].

1.3 Large-scale infrastructure investments in the United States

With high levels of growth of shale oil and gas production, there has been an accompanying increase in capital spending on midstream and downstream infrastructure [24]. These infrastructure investments primarily aim to provide greater access to shale plays which previously lacked connections and direct the resources to locations where additional demand could be served.

Unfortunately, making infrastructure investments in the oil and gas industry is rarely a straightforward affair. Because of the large sizes of investments, regulatory and geopolitical challenges, and considerable uncertainty in future resource prices, projects often face schedule delays and cost overruns [25]. In severe cases, the original intentions behind the investments might have to be abandoned. A pertinent example would be the over-investment in LNG import terminals in the early 2000s, when the general expectation was that of significant declines in future U.S. natural gas production [26].

To deal with the challenges associated with oil and gas infrastructure investments, a formalized framework in which to analyze these investments is required. This thesis develops a framework which assumes a comprehensive, high-level view of making optimal shale oil and gas investments in the context of current and future projections of supply, demand and prices of various commodities. This framework allows for a systematic study of the optimal types and levels of infrastructure investments, as well as the accompanying operational aspects of the supply chain.

The study continues the rich history of the application of optimization techniques to solve problems related to the management of supply chains. In particular, it can

be broadly associated to the study of facility location problems, where in general, optimal facility locations are to be determined to satisfy customer demand, with the objective of minimizing overall costs or maximizing overall profits. A review of the history of these problems is given by Owen and Daskin [27]. A compilation of significant contributions to the field, as well as mention of potential areas for further research, was given by Melo *et al.* [28]. A review of supply chain optimization specifically applied to the field of energy, in particular, that involving hybrid feedstock processes, was contributed by Elia and Floudas [29]. In these reviews, the importance of the need for more future models to incorporate both stochastic and dynamic aspects was highlighted. Sahinidis [30] provided a review of optimization under uncertainty, which provides hints of how one might extend the traditional facility location problems to their stochastic counterparts.

Although there has been a long history of papers demonstrating the application of supply chain optimization, we believe that the study in this thesis possesses a scope that is unprecedented in existing literature. Table 1.1 shows the different extent in which a supply chain optimization framework can vary in terms of its comprehensiveness, and as indicated, the study in this thesis lies on the high side of comprehensiveness for all of the dimensions of scope explored. To the best of our knowledge, most papers in existing literature typically only possess the high side of comprehensiveness for one or two dimensions of scope at best, while the other dimensions remain on the low side of comprehensiveness.

In addition, to the best of our knowledge, there has been no comprehensive nationwide supply chain optimization model specific to shale oil and gas prior to this work. Our framework integrates the economic dynamics of the upstream, midstream and downstream sectors of the oil and gas industry in the U.S. and select foreign markets and takes into account both the time-varying projections of supply, demand

Scope	less comprehensive ←		**Extent**	→ more comprehensive
Sector	Upstream	Midstream	Downstream	Full
				✓
Structure	Converging	Diverging	Multi-nodal	
			✓	
Time periods	Single (Rate-based)		Multiple	
			✓	
Objective	Cost-based		Profits-based	
			✓	
Uncertainty	Excluded		Included	
			✓	
Geography	State	Region	Country	International
				✓
Technologies	Single		Multiple	
			✓	
Commodities	Single		Multiple	
			✓	
Transportation	Single		Multiple	
			✓	

Table 1.1: Different dimensions of scope in which a supply chain optimization can vary in its extent of comprehensiveness. Having a more comprehensive model is ideal, but has to be balanced with considerations of computational tractability.

and price parameters as well as the different scenario realizations of these parameters. The development of the model would achieve the aim of providing an accurate and timely guide towards making the best investment and operating decisions for monetizing shale oil and gas in the country moving forward.

1.4 Mixed-integer linear programming

The optimization frameworks developed in this thesis are formulated as mixed-integer linear programs (MILP), which have the following structure:

$$\begin{aligned} \underset{x,y}{\text{minimize}} \quad & c^{\mathrm{T}}x + d^{\mathrm{T}}y \\ \text{subject to} \quad & Ax + Dy = b, \\ & x \in X \cap \mathbb{R}_{+}^{n_x}, \\ & y \in Y \cap \mathbb{Z}_{+}^{n_y}, \end{aligned}$$

where x are nonnegative continuous variables and y are nonnegative integer variables. X and Y denote polyhedral sets containing x and y respectively. c, d are cost vectors for x and y, respectively. $Ax + Dy = b$ are the coupling constraints between these variables.

In the context of the thesis, the continuous variables typically refer to the operational decisions, whereas the integer variables typically refer to investment, location, or logical decisions.

Modern commercial solvers like CPLEX [31] and Gurobi [32] typically implement a branch-and-cut method to solve the MILP. Many ways in which the solution procedure can be made more efficient are documented in literature, typically by developing sharp formulations, formulating tight cuts, or decomposing the structure of

the problem into smaller subproblems [33]. However, the effectiveness of these modifications over a direct implementation of the fullspace formulation depends largely on a case-by-case basis among the instances examined.

1.5 CFSTR equivalence principle

Chemical engineers often encounter reaction schemes with varying complexity and have an abundance of reactor and separator designs to choose from. This proves to be a formidable task, and a method for screening reaction schemes is often desired such that more attention can be focused on particular schemes that have comparatively higher productivities of a certain desired species. It was with this in mind that Feinberg and co-workers developed a theory to determine an absolute and computable limit for the achievable production of any species for any arbitrary steady-state reactor-separator design, given the kinetics of the reaction network and a specified commitment of resources [34, 35]. Named the Continuous Flow Stirred Tank Reactor (CFSTR) equivalence principle, it asserts that the effluent of any steady-state reactor-separator design can be achieved arbitrarily closely by another steady-state design with arbitrarily sharp separations but in which the only reactor components are $s + 1$ ideal CFSTRs, where s is the rank of the underlying network of chemical reactions.

The CFSTR equivalence principle is atttractive because it has a simple form yet powerful applicability as a screening tool for potentially very complex reaction networks. In this thesis, distinct from the studies above, we explored the applicability of the principle to screen and target favorable reaction pathways, with applications directed towards gas-to-liquids conversion.

1.6 Thesis outline

The outline of the remainder of the thesis is as follows:

- Chapter 2 discusses an optimization framework for the dynamic allocation of small-scale mobile plants to monetize associated or stranded gas and its application to the Bakken shale play.
- Chapter 3 expands upon the previous chapter by incorporating uncertainty into the framework.
- Chapter 4 discusses a comprehensive supply chain optimization framework to determine optimal shale oil and gas infrastructure investments in the United States.
- Chapter 5 explores the screening of reaction networks using an optimization framework based on the CFSTR equivalence principle.
- Chapter 6 concludes.
- Appendix A provides Supplementary Material for Chapters 2 and 3.
- Appendix B provides Supplementary Material for Chapter 4.
- Appendix C provides the Capstone Paper, written during my studies at the Sloan School of Management.

Chapter 2

Small-Scale Mobile Plants: Bakken Shale Play

Associated or stranded natural gas presents a challenge to monetize due to its low volume and lack of supporting infrastructure. Recent proposals for deploying mobile, modular plants, such as those which perform gas-to-liquids (GTL) conversion or produce liquefied natural gas (LNG) on a small scale, have been identified as possible attractive routes to gas monetization. However, such technologies are yet unproven in the marketplace. To assess their potential, we propose a multi-period optimization framework which determines the optimal dynamic allocation and operating decisions for a decision maker who utilizes mobile plants to monetize associated or stranded gas. We then apply this framework to a case study of the Bakken shale play. Our framework is implemented to determine the optimal net present value (NPV) which would be realized over a twenty-year time frame. Sensitivity studies on the technology costs and conversion inputs conclude that the profitability and viability of mobile technologies remain valid for a wide range of possible inputs.

2.1 Problem description and challenges

We will assume the role of a decision maker whose primary concern is to monetize natural gas in stranded fields or associated with the production of oil. When making decisions, the decision maker would have to consider the production characteristics unique to the field and choose among several technology options. These technologies convert natural gas into either higher-value products or a more transportable form, or both. Among the technology options available, two which have garnered the most interest due to their relative maturity are the gas-to-liquids (GTL) and liquefied natural gas (LNG) technologies.

GTL has recently gained attention due to the increased spread between the price of oil and natural gas, as noted by Hobbs and Adair [36] and Salehi *et al.* [37]. The GTL process converts natural gas into liquid fuel. There are three main parts to this process: 1) syngas generation, 2) Fischer-Tropsch (FT) synthesis, and 3) refining and upgrading.

In syngas generation, natural gas is first cleaned and then converted into syngas, which is a mixture of hydrogen and carbon monoxide. After the syngas has been generated, it undergoes FT synthesis where it is converted into longer chain hydrocarbons. Finally, after the FT synthesis step, the product is sent for refining and upgrading to meet final specifications.

GTL products are attractive not only because they are liquid fuels and can be easily transported, but also because they are virtually sulfur-free, as mentioned in studies by Wood *et al.* [38] and Salehi *et al.* [37]. The most promising product from the GTL process is GTL diesel. Also high in cetane number, it is ideal as a blendstock for refineries to adjust conventional diesel in production to meet specifications.

LNG technology is considered mature and proven. The process involves lique-

faction of gas by cooling to cryogenic temperatures. Prior to cooling, the feed gas undergoes several treatment steps, such as filtration and removal of carbon dioxide, sulfur, mercury and water.

The value of LNG is that it significantly increases the energy density of natural gas, allowing it to be transportable for sale in distant markets. In the U.S., the most promising market for LNG is fuel for heavy-duty trucks or freight rail, as documented in a study by TIAX [39].

In addition, depending on the source of natural gas, there may be a significant presence of natural gas liquids (NGLs), ethane, propane, butane, etc., mixed in the wellhead gas. Such gas is termed "wet gas", and the NGLs are usually separated from the mixture because they possess substantial economic value. NGLs primarily serve as feedstock for the petrochemical industry or as fuel for heating and transportation purposes, as noted by Platts Price Group [40]. Therefore, GTL and LNG technologies which take in wet gas as their feedstock should necessarily have a NGL separation unit.

Applying these technologies to monetize stranded or associated gas poses a challenge. First, the technologies have to be designed to be mobile, since the supply of gas at any fixed location would not last for very long. The mobility of the plants adds a dimension of complexity to the decision-making process. Although the idea of mobility generally allows plants to be more agile and thus suitable for capturing stranded gas, start-up and shut-down costs would be incurred every time a move is made. Thus, the company has to weigh the costs and benefits of continuing operations at a certain location versus redeployment in the context of how the supply profile and the demands of its customers evolve over time. Second, because of the dynamic nature of gas supply and well availability, it is a challenge to determine the optimal number of mobile plants of each technology type to be purchased or sold at

each time point of the time horizon that would maximize profits.

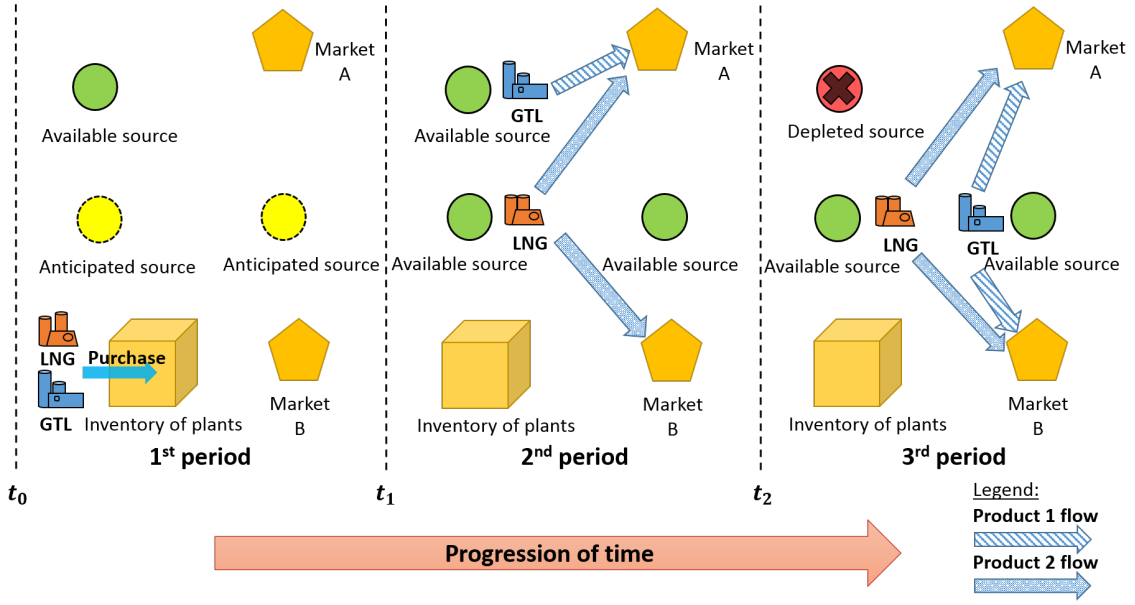


Figure 2-1: General simplified illustration of the dynamic mobile plant allocation problem to monetize associated or stranded gas.

Figure 2-1 portrays a simplified illustration of the decision framework under consideration. In this example, there are three time stages, three gas sources, two technologies for mobile plants (GTL and LNG) and two markets. Depending on the time period, the gas sources might or might not have gas available to monetize in various quantities. At each time stage, the decision maker has to decide how many plants of each technology type and size to purchase or sell and where to locate existing plants purchased in previous time stages, along with the appropriate levels of production to deliver the finished products to the markets. In Figure 2-1, we see that the decision in the first time period to purchase a GTL and an LNG plant and develop gathering systems at the gas sources is made in anticipation of more sources coming online in subsequent time periods. Once the plants have been constructed off-site and shipped to the site of operation in the second time period, they are then

available for the decision maker to deploy to the available gas sources. Subsequently, between the second and third time period, the gas source in which the GTL plant has been located becomes depleted. Hence, the decision is made to move the plant to another source where it can resume its operations profitably. All decisions are made with the goal of maximizing the net present value (NPV) of the project over the entire time horizon.

2.2 Mathematical formulation

The problem is formulated as a multi-period MILP. A decision maker makes decisions on a uniform grid $t \in \{0, \ldots, T\}$ over a given time horizon. Based on the topology of the gas field and the production characteristics, the decision maker has access to a number of gas sources $i \in \{1, \ldots, I\}$, which might or might not supply gas depending on the point in time considered. At each source, the decision maker can choose to deploy a plant of type $j \in \{1, \ldots, J\}$ in order to monetize the gas. The products are then shipped and sold to various surrounding markets $k \in \{1, \ldots, K\}$, each of which has a demand for a particular type of product $l \in \{1, \ldots, L\}$. In order to track the age of the plants at a given time t, we also record the time point at which the plant was purchased $\tau \in \{0, \ldots, t\}$.

The optimization decisions are:

1. Decision to allocate plant of type j to source i at time t, denoted by $y_{ij}^t \in \{0, 1\}$.

2. Indicator of the presence of a gas gathering system at source i at time t, denoted by $z_i^t \in \{0, 1\}$.

3. Gas feed rate to plant of type j at source i at time t, denoted by $x_{ij}^t \in \mathbb{R}_+$.

4. Product delivery rate of product l from source i to market k at time t, denoted by $w_{ikl}^t \in \mathbb{R}_+$.

5. Number of plants of type j purchased at time t, denoted by $Buy_j^t \in \mathbb{Z}_+$.

6. Number of plants of type j which originally arrived in inventory at time $0 \leq \tau < t$, sold at time t, denoted by $Sell_{j\tau}^t \in \mathbb{Z}_+$.

7. Inventory of plants of type j at time t, arriving in inventory at time $0 \leq \tau \leq t$, denoted by $Inv_{j\tau}^t \in \mathbb{Z}_+$.

The decision maker has to make his or her decisions subject to the following constraints.

Once a decision has been made to develop a gathering system at source i at time t, the gathering system is available at that source for all subsequent time points:

$$z_i^t \leq z_i^{t+1}, \quad \forall i, \quad \forall 0 \leq t < T. \tag{2.1}$$

A plant can only be deployed at a particular source if a gathering system had been developed at least $\mathcal{T}_g$ time periods prior, where $\mathcal{T}_g$ denotes the time necessary to construct the gathering system:

$$y_{ij}^t \leq z_i^{t-\mathcal{T}_g}, \quad \forall i, j, \quad \forall t \geq \mathcal{T}_g, \quad \text{and} \tag{2.2}$$

$$y_{ij}^t = 0, \quad \forall i, j, \quad \forall t < \mathcal{T}_g. \tag{2.3}$$

The inventory balance of plants has to be satisfied. Constraints (2.4) describes the inventory balance where $\tau = t$, (i.e., for plants that are brand new). In this case, the inventory of new plants are simply the number bought $\mathcal{T}_j$ time stages ago, where

$\mathcal{T}_j$ denotes the construction time lag between the purchase of plants and their actual arrival in the inventory ready for deployment.

$$Inv_{j\tau}^{t} = Buy_{j}^{t-\mathcal{T}_j}, \quad \forall j, \quad \forall t \geq \mathcal{T}_j, \quad \forall \tau = t, \quad \text{and} \tag{2.4}$$
$$Inv_{j\tau}^{t} = 0, \quad \forall j, \quad \forall t < \mathcal{T}_j, \quad \forall \tau = t.$$

Constraint (2.5) describes the inventory balances where $\tau < t$, which considers plants which are at least one time point old. Here, the current inventory of plants bought at a particular previous time point is simply that carried forward from the previous inventory, less any plants that are sold.

$$Inv_{j\tau}^{t} = Inv_{j\tau}^{t-1} - Sell_{j\tau}^{t}, \quad \forall j, t, \quad \forall \tau < t. \tag{2.5}$$

The number of plants allocated to the sources cannot exceed the number of plants in the inventory. Note that we are indifferent to the purchase date of the plant and treat all plants as equally efficient for our allocation decisions:

$$\sum_{i} y_{ij}^{t} \leq \sum_{\tau=0}^{t} Inv_{j\tau}^{t}, \quad \forall j, t. \tag{2.6}$$

Each gas source is associated with a gas supply s_i^t. If a plant has been deployed at a gas source, the gas feed rate to the plant cannot exceed the gas supply from the source:

$$x_{ij}^{t} \leq s_i^t y_{ij}^{t}, \quad \forall i, j, t. \tag{2.7}$$

At every source, the sum of flows from the gas source to the plants cannot exceed

the gas supply:

$$\sum_j x_{ij}^t \leq s_i^t, \quad \forall i, t. \tag{2.8}$$

Each plant type is associated with a capacity m_j. The feed to a plant cannot exceed the plant's capacity:

$$x_{ij}^t \leq m_j y_{ij}^t, \quad \forall i, j, t. \tag{2.9}$$

In addition, each plant type is associated with a minimum capacity that cannot be violated due to physical constraints on the equipment. This minimum capacity is determined by the turndown ratio, expressed as a fraction γ_j of total capacity. Thus, if allocated, the feed to the plant cannot be below the plant's minimum capacity:

$$x_{ij}^t \geq \gamma_j m_j y_{ij}^t, \quad \forall i, j, t. \tag{2.10}$$

For each technology, we denote a conversion factor α_{jl} to denote the proportion of gas feed which gets converted into an end product. Then, the total flow rate of a product exiting a source must equal the sum over all routes of shipment of that product to its markets:

$$\sum_j \alpha_{jl} x_{ij}^t = \sum_k w_{ikl}^t, \quad \forall i, l, t. \tag{2.11}$$

Each market is associated with a demand d_{kl}^t. Then, if a market is chosen to be served, the total rate of products delivered to that market cannot exceed their

demand:

$$\sum_i w_{ikl}^t \leq d_{kl}^t, \quad \forall k, l, t. \tag{2.12}$$

The objective function of the decision maker is to maximize the net present value (NPV) of the project, given an appropriate discount factor r:

$$\text{NPV} = \sum_t (1+r)^{-t}(Rev^t - Cost^t) \tag{2.13}$$

where Rev^t is the total revenue and $Cost^t$ is the total cost at time t.

Rev^t is defined as the sum of the following terms. The sale of all products, each with a corresponding price p_{kl}^t:

$$\sum_i \sum_j \sum_k \sum_l p_{kl}^t \alpha_{jl} x_{ij}^t. \tag{2.14}$$

The gains obtained by salvaging a plant and its associated equipment. The salvage value of a plant j purchased at time τ, sold at time t, is denoted $c_{salv,j\tau}^t$:

$$\sum_j \sum_{\tau=0}^{t} c_{salv,j\tau}^t Sell_{j\tau}^t. \tag{2.15}$$

$Cost^t$ is defined as the sum of the following terms. The total plant investment costs, where each plant has an associated capital cost $c_{cap,j}$:

$$\sum_j c_{cap,j} Buy_j^t. \tag{2.16}$$

The total gathering system investment costs, where each gathering system has an associated capital cost $c_{gathercap}$. Note that because of Constraint (2.1), we are guaranteed that for each term $c_{gathercap}(z_i^t - z_i^{t-1})$, with i fixed, there will be at most one non-zero term across all values of t:

$$\sum_i c_{gathercap}(z_i^t - z_i^{t-1}). \tag{2.17}$$

The total startup costs of deploying plants to their sources. These costs include installation and assembly costs of the plant to the associated gathering system connecting to the wells, as well as any lost revenue that occurs due to partial operation during the period. The following representation assumes that a plant can be set up within the time elapsed between two time points:

$$\sum_i \sum_j c_{start,j} \max\{y_{ij}^t - y_{ij}^{t-1}, 0\}. \tag{2.18}$$

The total shutdown costs of removing plants from their sources. These costs include disassembly and removal costs of the plant from the associated gathering system connecting to the wells, as well as any lost revenue that occurs due to partial operation during the period. Again, the following representation assumes that a plant can be shut down within the time elapsed between two time points:

$$\sum_i \sum_j c_{shut,j} \max\{y_{ij}^{t-1} - y_{ij}^t, 0\}. \tag{2.19}$$

For modeling purposes, the max functions (2.18) and (2.19) are typically repre-

sented by introducing auxiliary variables $\delta^t_{start,ij}, \delta^t_{shut,ij} \in \mathbb{R}_+$ and the constraints:

$$y^t_{ij} - y^{t-1}_{ij} \leq \delta^t_{start,ij}, \quad \forall i, j, t, \tag{2.20}$$

$$y^{t-1}_{ij} - y^t_{ij} \leq \delta^t_{shut,ij}, \quad \forall i, j, t, \tag{2.21}$$

which then allow (2.18) and (2.19) to be represented by $\sum_i \sum_j c_{start,j} \delta^t_{start,ij}$ and $\sum_i \sum_j c_{shut,j} \delta^t_{shut,ij}$, respectively. Note that when $t = 0$, we set y^{t-1}_{ij} to 0.

The total operating costs of the plants. The operating costs are divided into fixed operating costs $c_{opFixed,j}$ and variable operating costs $c_{opVar,j}$.

$$\sum_i \sum_j c_{opFixed,j} y^t_{ij}, \tag{2.22}$$

$$\sum_i \sum_j c_{opVar,j} x^t_{ij}. \tag{2.23}$$

The total transportation costs of products from the plants to their markets. Shipping a product l from its source i to a market k incurs a per unit cost of $c_{ship,ikl}$. This cost primarily depends on the mode of transportation and the physical distance between the gas source and the market:

$$\sum_i \sum_k \sum_l c_{ship,ikl} w^t_{ikl}. \tag{2.24}$$

The total costs of wellhead gas fed to the plants. The spot price of the wellhead gas is denoted by $p^t_{gas,i}$:

$$\sum_i \sum_j p^t_{gas,i} x^t_{ij}. \tag{2.25}$$

In the case of associated gas otherwise flared if not monetized, $p^t_{gas,i} = 0$.

In short, we seek to maximize (2.13) subject to the constraints of Eqs. (2.1) to (2.12) and (2.20) and (2.21).

2.3 Case study on the Bakken shale play

2.3.1 Overview

The Bakken Formation is a wet shale formation occupying approximately 200,000 square miles within the Williston Basin, extending through various parts of North Dakota, South Dakota, Montana and the Canadian provinces of Manitoba and Saskatchewan. A detailed description of the formation was given by Wocken *et al.* [41].

The rapid growth in oil production has also led to a significant production of associated gas. Despite attempts to develop infrastructure to gather, process and transmit the gas, this has proceeded at a much slower rate than that of production. The result has been a rapid increase in the amount of gas that is unable to be monetized and hence flared.

There are several unique aspects to consider when monetizing associated gas in the Bakken play: first, as analyzed by Mason [42], the production rate of a typical well operating on unconventional resources faces a very steep decline. As a result, the flow rates of associated gas might very quickly drop to levels which make the gas uneconomical to monetize. Second, as indicated by Wocken *et al.* [41], the associated gas arising from the Bakken is typically wet, and exists as a mixture of methane and NGLs. Therefore, when developing technology to be implemented in the Bakken, we

include an NGL separation unit in the process.

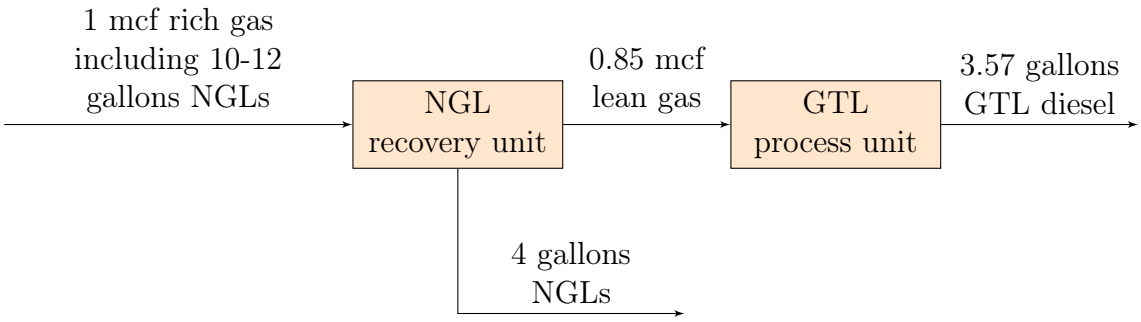
We adopt the role of a decision maker who faces a time horizon of twenty years, divided into quarterly time steps. The choice for twenty years corresponds to the typical lifespan of a mobile plant implementing the technologies that we considered - GTL and LNG, each of three different sizes.

2.3.2 Technologies

Figure 2-2 shows the flow diagrams of the two systems considered, specifying flow rates and compositions of the feed gas and finished products. There are three sizes for each system considered in our study - small, medium and large. The sizes correspond to the capacity to process a feed rate of 500, 1,000, and 1,500 thousand cubic feet (mcf) of rich gas per day, respectively. Although in reality the gas composition can vary between the Bakken wells, we assumed a constant, representative gas composition at 10 - 12 gallons of NGLs per 1 mcf of rich gas. This value was chosen to be in line with a study by Wocken *et al.* [41], where several options for monetizing associated gas in the Bakken were discussed and assessed by a more qualitative approach.

Each technology produces two high-value products: the "GTL process" produces NGLs and GTL diesel, while the "LNG process" produces NGLs and LNG. Although 10 - 12 gallons of NGLs is present per mcf of feed gas, the recovery of NGLs in the product stream can only be achieved at a rather low rate at 4 gallons per mcf of feed gas. This recovery rate was obtained from the Wocken *et al.* [41] study. The low recovery rate was mainly due to the technological simplicity of the extraction unit, which operated as a two-stage compression and chilling process at -20 °F and 1,000 psi. Although not explicitly stated in the report, it could be that the technological

GTL process flow:



LNG process flow:

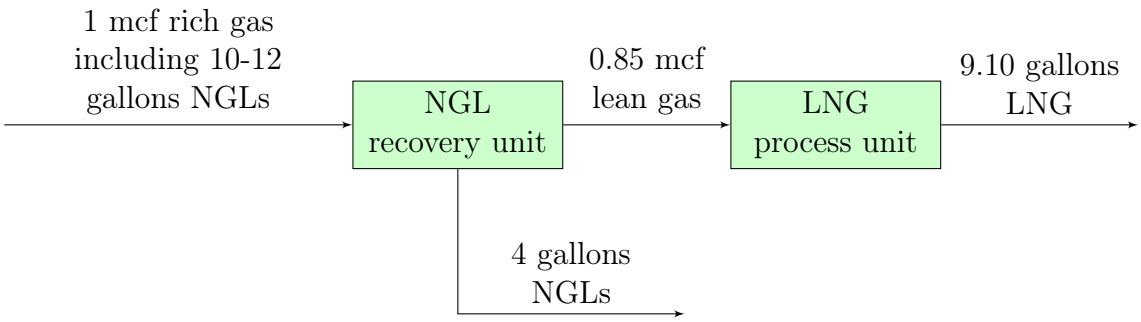


Figure 2-2: Flow diagrams showing compositions of streams for the modular GTL and LNG technologies under consideration. Flow rates are expressed on a 1 mcf feed rich gas basis.

sophistication of the NGL extraction process was limited by size requirements.

For the GTL process, a conversion factor of 1 mcf of lean feed gas to 0.1 barrel of GTL diesel product was used, which matches the values reported by Patel [43] and Wood *et al.* [38]. For the LNG process, we assumed an efficiency rate of 88%, as reported by Patel [43] and Garcia-Cuerva and Sobrino [44]. This corresponds to a conversion factor of 1 mcf of lean feed gas to 10.7 gallons of LNG. We then assumed that 85% of the rich gas flow rate is recovered as lean gas after the extraction of NGLs. This is again in line with the assumptions made by Wocken *et al* [41]. With these inputs, we arrive at 3.57 gallons of GTL diesel product and 9.10 gallons of LNG per mcf of rich gas, respectively.

As the flow rates of associated gas are characterized by steep declines, it is also important to consider the point at which the flow rates of feed gas are too low for the mobile plants to operate. That is, the flow rates must respect the minimum turndown capacity for each technology. A reasonable lower bound for the turndown capacity for both GTL and LNG technologies is set at 50%, which is a value that was quoted by Ballout and Price [45] and Baxter [46], who studied these technologies at the small scale.

These technologies are agile enough such that they could be moved from one area to another within a quarter of a year. To define what constitutes an area of operation, we considered that the technology should serve a maximum area of one square mile, which is a reasonably small size such that the installation or disassembly of equipment could be done well within the allocated time frame. Well spacing in the Bakken was set at four wells per square mile, which corresponded to the average well density in the Bakken determined in a technical report by Continental Resources [47]. Therefore, each technology takes in as its input the associated gas arising from four wells, denoted as a "gas source".

One of the main proposed advantages of modular plants is that they have much shorter construction lead times than that of traditional investments. For example, GE's LNG In a Box system [48], which produces 10,000 to 50,000 gallons of LNG per day has a 6 - 12 month lead time while Oxford Catalysts' 2,500 barrel per day modular GTL plant [49] has a 18 - 24 month lead time. Since the plants under consideration in our case study are on the smaller side of the scale, we assumed a reasonable construction lead time of one year. We assume that the gathering system takes a year to construct.

Current literature on estimation of startup and shutdown costs is scant, but an early study by Bauman [50] estimated the average and median start-up costs of various types of chemical process capital projects to be around 1% of the fixed capital costs. Correspondingly, we assume that both the startup and shutdown costs are 1% of the capital costs of the plant. In addition, we assume that each plant will have an operating lifetime of twenty years, equal to the time horizon under study. Finally, at any point of the time horizon, the plants can be sold at their depreciated costs. We determine the depreciated costs by performing a straight-line depreciation of the original capital costs of the plants by their age, over twenty years. The basic characteristics of the mobile plants under consideration are summarized in Table 2.1.

Capital costs of the GTL and LNG processing units were not readily available for the small sizes that were considered for this study. As such, we collected data on capital costs from existing larger GTL and LNG plants or from studies of hypothetical plants from the literature, as documented in the Supplementary Material. Using the U.S. Inflation Calculator [51], costs were normalized to 2012 dollars in order for valid comparisons to be made across plants. We then plotted cost curves to determine the relationship between capacity and capital costs for both technologies. Figures 2-3 and 2-4 depict the log-log relationship between capital costs and capacity.

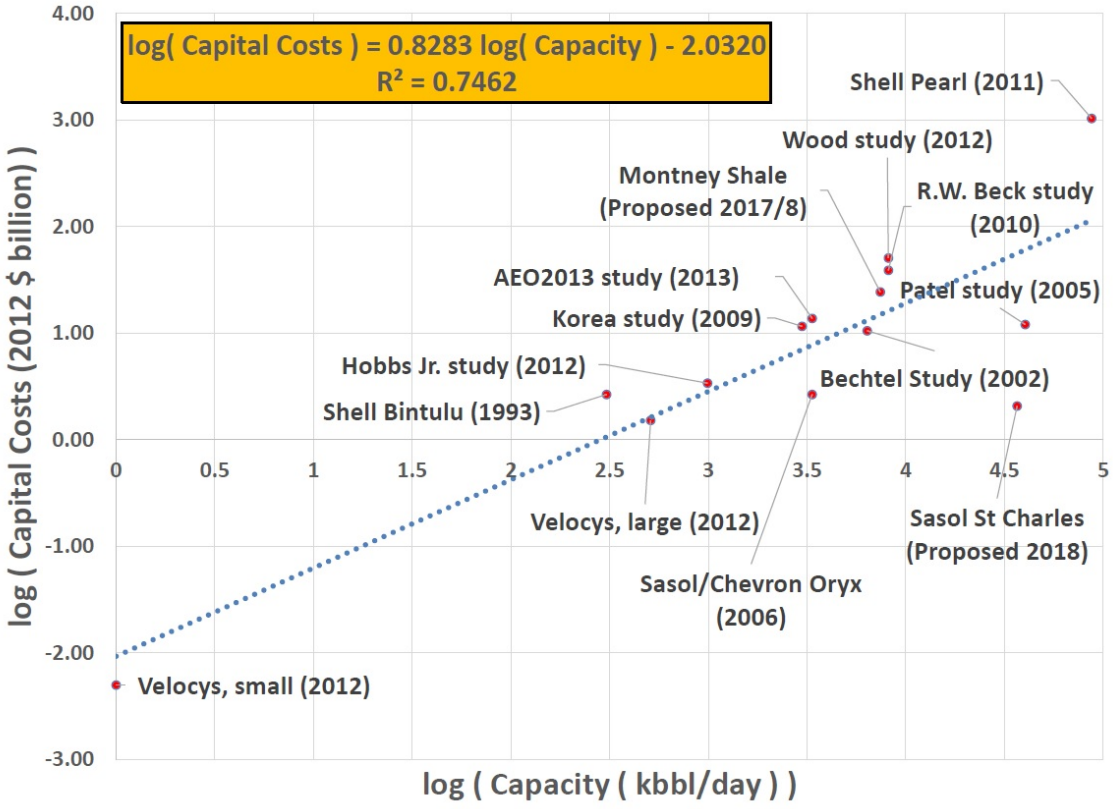


Figure 2-3: Capital cost curve of GTL plants in actual implementation or in literature studies.

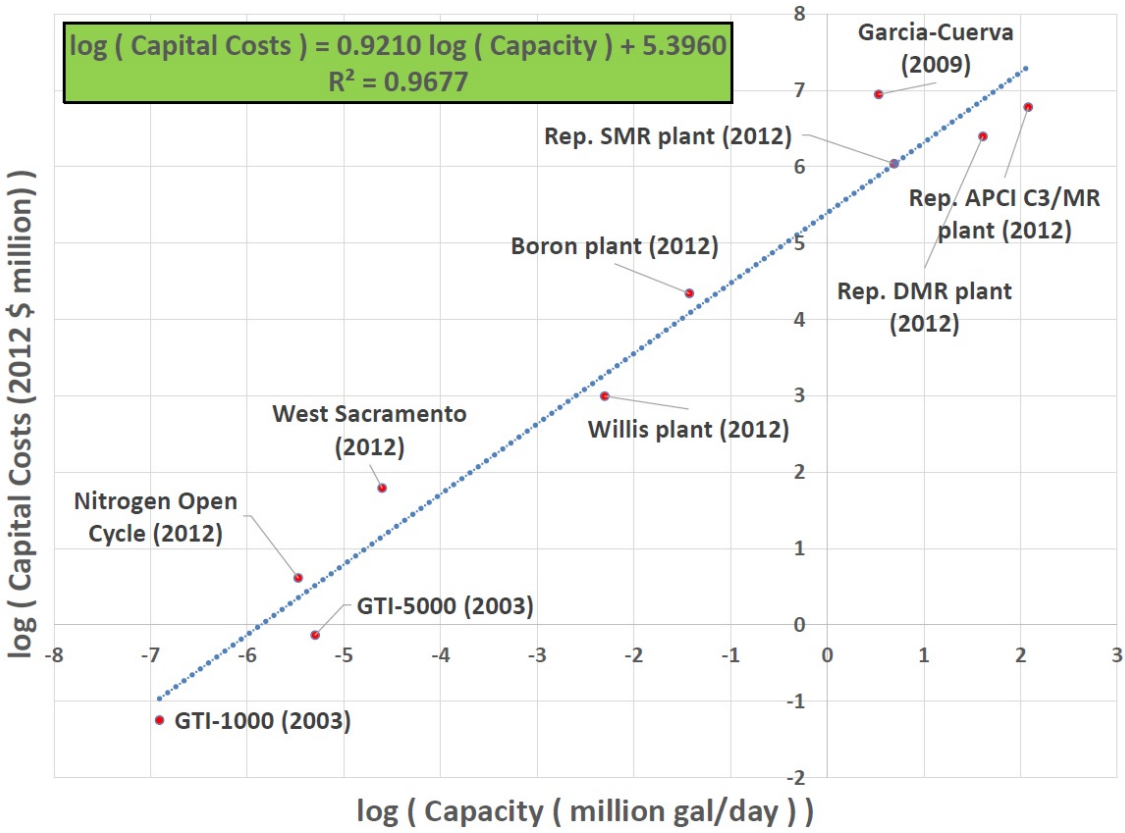


Figure 2-4: Capital cost curve of LNG plants in actual implementation or in literature studies.

Plant life	20 years
Sizes	Small (500 mcfd) Medium (1,000 mcfd) Large (1,500 mcfd)
Turndown capacity	50%
Gathering system construction time	1 year
Plant construction lead time	1 year
Startup costs	1 % of capital costs
Shutdown costs	1 % of capital costs
Depreciation	Straight-line
Technology land requirement	1 sq. mile
Number of wells served per plant	4

Table 2.1: Basic characteristics of the GTL and LNG mobile plants under consideration.

From the fitted equations, we calculated the estimated capital costs required for GTL and LNG units sized to our specifications. The capital costs of the GTL units required extrapolation of the fitted line, as compared to the capital costs of the LNG units, which were interpolated values. Therefore, we have a greater confidence in the accuracy of the capital costs of the LNG units. Nevertheless, since mobile plant technologies are not mature, there would be a large variation in the costs that manufacturers of such plants might quote. As detailed later, we performed sensitivity studies on significant variations of the assumed capital costs to determine how our results might change. Using the Wocken *et al.* study [41] as a guide, we estimated the capital costs of the NGL removal system sized to our specifications using the *six-tenths* rule.

The operating costs for each plant were separated into two components: fixed and variable operating costs. The fixed costs were expressed on a per plant per quarter basis, since fixed costs are incurred every time the decision is made to deploy a plant

at a location. The variable operating costs were expressed in a per unit of feed gas basis, as they naturally scale with the level of production.

Fixed costs comprise the cost of labor, taxes and duties, maintenance costs and general administration costs. For a single plant, our assumptions for labor includes operating personnel at four shifts with one person per shift, maintenance personnel of one person, and administration and support personnel of one person. Salaries were set at \$35,000/year per person. Taxes and duties were set at 0.75% of capital costs, fixed maintenance costs at 1.5% of capital costs, and general administration costs at 20% of personnel wages and fixed maintenance costs, which corresponded to values set by Garcia-Cuerva and Sobrino [44].

The variable operating costs differ depending on the type of technology considered. For a single GTL plant, we considered catalysts and gas processing chemicals and water as the main components of variable operating costs. The catalysts and chemicals were set at \$6 per barrel, after assimilating information from correspondence with a manufacturer of small-scale GTL technologies with reasonable estimates from industry literature. Corresponding to a GTL study by the National Energy Technology Laboratory (NETL) [52], water was costed at \$0.2 per barrel.

For a single LNG plant, the main sources for variable operating costs are the make-up for refrigerant losses and chemicals needed for gas processing. Corresponding to Kohler *et al.* [53], refrigerant make-up was priced at \$800 per ton of refrigerant. Corresponding to Garcia-Cuerva and Sobrino [44], a ratio of 0.0008 refrigerant required per LNG produced was assumed and chemicals were costed at \$0.82 per ton of LNG produced.

Note that both plants do not require an external source of electricity. As noted by NETL [52], in the GTL process, the highly exothermic FT process generates enough steam to power the plant. Similarly, as noted by Garcia-Cuerva and Sobrino [44], in

the LNG process, part of the natural gas feed is used as fuel for the plant, and the conversion value from feed to product has taken the power production process into account.

Lawlor and Conder [54] provided capital costs for installing gathering and processing systems in unconventional gas fields for a range of pipes of different diameters and materials. Such costs included equipment, transportation, survey, installation, supervision and contingency costs. For our study, we set the capital cost to be the average of the quoted cost per mile for 3, 4, 6 and 8-inch diameter steel pipes. In addition, the length of steel pipe required at a gas source was taken to be $2\sqrt{2}$ miles. This value arises from our assumed geometry that plants would be located in the center of a square plot of land one square mile in area, and the four wells in the plot of land are located at each of the corners of the square. Thus, running pipes from the center of the square to its corners amounts to $2\sqrt{2}$ miles of pipeline. This is a conservative estimate, suited to the needs of our analysis. With this assumption, capital costs of the gathering system amounted to \$0.6 million.

We summarize the associated capital costs for each plant in Table 2.2 and the associated operating costs for each plant in Table 2.3. A detailed breakdown of the operating costs is found in the Supplementary Material.

2.3.3 Production curves

Mason [42] collected data of monthly production curves from various strata of the Bakken field and fitted the curve to a hyperbolic decline function. The hyperbolic decline function is given as:

$$q_{oil}(t) = \frac{q_{i,oil}}{(1 + bD_i t)^{1/b}} \tag{2.26}$$

Entity	Capital Costs ($ million)		
	Small	Medium	Large
(1) NGL Recovery Unit	1.6	2.5	3.2
(2) GTL Processing Unit	9.5	17.0	23.8
(3) LNG Processing Unit	1.5	2.9	4.2
Overall Capital Costs:			
GTL System = (1) + (2)	11.1	19.5	27.0
LNG System = (1) + (3)	3.1	5.4	7.4

Table 2.2: Capital costs for GTL and LNG mobile plants.

Plant	Fixed ($ million/Quarter)			Variable ($/mcf feed gas)
	Small	Medium	Large	All Sizes
GTL System	0.136	0.191	0.240	0.53
LNG System	0.085	0.101	0.115	0.02

Table 2.3: Operating costs for GTL and LNG mobile plants.

where $q_{oil}(t)$ denotes the monthly oil production rate for a typical well at time t, $q_{i,oil}$ denotes the initial production rate, D_i is the nominal decline rate and b is the decline exponent.

To determine the amount of associated gas with the production from the well, we referred to the EIA's monthly drilling productivity reports [55] which reported a value of 258 to 292 mcf per day of new-well gas production per rig throughout 2012. A similar value of 300 mcf per day of rich gas flow rate from the average wellhead was reported in the Wocken *et al.* study [41]. Taking these values into account, our study set the initial gas flow rate from a well to be 9,000 mcf per month. These production parameters are summarized in Table 2.4.

Table 2.4: Bakken well production characteristics and assumptions.

Parameters of production	
$q_{i,oil}$ (bbl/month)	14,225
b	1.4
D_i	0.197
$q_{i,gas}$ (mcf/month)	9,000

From this initial value and the fitted parameters associated with the decline curve, we then constructed a similar decline curve for associated gas by assuming that the ratio of associated gas to oil remained constant throughout production. Then, making the assumption that all four wells in an area served by one mobile plant came online at the same time, we determined the associated gas production profile for the area served by the plant, denoted as a "gas source". The resulting production curve for a gas source is shown in Figure 2-5.

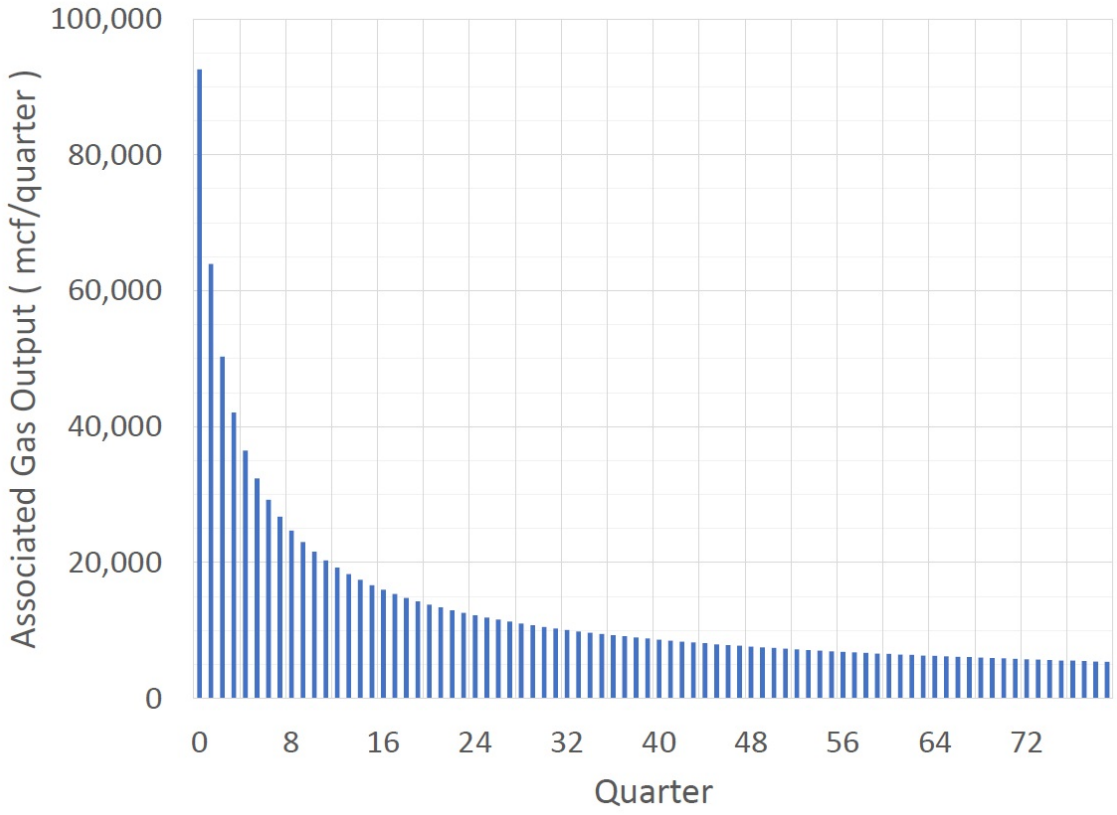


Figure 2-5: Quarterly production curve of associated gas at a gas source under consideration in the Bakken field.

2.3.4 Markets

For purposes of this study, we only considered currently existing domestic markets for GTL diesel, LNG, and the associated NGLs products that allowed for a realistic chance for the products to be sold in the near future. Overseas markets are considered too nascent at this point, and could be analyzed in a subsequent study. Due to the disparate nature of the finished products, the markets corresponding to each product differ from each other. Nevertheless, our proposed general model allows for situations where multiple products get shipped to one market.

Markets for GTL diesel were taken to be the ten closest existing refineries surrounding the Bakken shale play, the locations of which were taken from EIA's U.S. Energy Mapping System [56]. As mentioned earlier, the most immediate market for GTL diesel lies in its use as a refinery blendstock for high quality diesel. To determine the demand for GTL diesel in each of the refineries, the capacities (in barrels of crude oil per day) were obtained. We then multiplied this value by 10 gallons of diesel produced per barrel of crude oil processed, a yield value given by EIA [57], to determine the diesel output for each refinery. Then, assuming that demand for GTL diesel as a blendstock comprise 20% of the diesel output, we arrive at the final values for the demand for GTL diesel at each of the refineries.

Markets for LNG were taken to be the nine closest existing public LNG fueling stations surrounding the Bakken shale play, the locations of which were taken from the U.S. Department of Energy Alternative Fuels Data Center's Alternative Fueling Station Locator [58]. This recognizes LNG's immediate value as a fuel for long distance trucks. Unfortunately, exact capacity values for each station were not available. Therefore, we relied on the value specified in EIA's Natural Gas Transmission and Distribution Module of the National Energy Modeling System (NEMS) [59], where

the capacity of an LNG public retail station was set at 4,000 diesel gallon equivalent (dge) per day.

Finally, markets for the extracted NGLs were taken to be the two main NGLs market hubs as indicated on the EIA's U.S. Energy Mapping System [2] - Conway and Mont Belvieu. The demand for NGLs was set to be the fractionation capacities of the two hubs respectively. In 2013, Van Hull [60] reported that the fractionation capacities of Conway and Mont Belvieu were 550,000 barrels/day and 1,250,000 barrels/day of NGLs respectively.

To determine shipping costs, we first determined the distance from the Bakken shale play to the various markets. Since the gas sources were assumed to be tightly clustered, we made the simplification that the distance from all gas sources to a particular market would be identical, essentially considering the Bakken shale play as a single location. The Google Distance Matrix API [61] was used to map the road distance from Williston, ND to each of the GTL diesel and LNG product markets, since it is assumed that the products will be transported by trucks.

Hadder and McNutt [62] attributed shipping costs for light duty diesel fuel at $0.10 per gallon for several hundred miles by long distance trucking or rail cars. We assumed that this quote would cover two hundred miles, and set shipping costs for GTL diesel at $0.050 per gallon per hundred miles. The 2012 study on LNG infrastructure by TIAX [39] quoted a trucking cost of LNG fuel at $0.20 per gallon of LNG for 300 miles. Thus, we used this rate (i.e., $0.067 per gallon per hundred miles) for LNG shipping costs. Rates were multiplied by the corresponding distance to market to arrive at final shipping costs for the GTL diesel and LNG markets. For NGLs, we assumed delivery by single-car rail to the corresponding NGLs market hub. Aux Sable's 2014 assessment of NGLs transportation costs from North Dakota [63] quoted single-car rail shipping to Conway at $0.28 per gallon NGL and to Mont

Belvieu at $0.39 per gallon NGL.

Product	Market	Number	Shipping Costs
GTL diesel	Petroleum Refinery	10	$0.050/gallon/hundred miles
LNG	LNG Fueling Station	9	$0.067/gallon/hundred miles
NGLs	NGL Market Hub	2	To Conway: $0.28/gallon; To Mont Belvieu: $0.39/gallon

Table 2.5: Type and number of markets for GTL diesel, LNG and NGLs products and associated shipping costs.

Table 2.5 summarizes the market type and number for the various products and the associated shipping costs. The full list of markets and specific shipping costs are detailed in the Supplementary Material.

2.3.5 Supply, price and demand projections

We use data generated from the Energy Information Administration's (EIA) National Energy Modeling System (NEMS) [64] to obtain projections for supply, price and demand parameters for the next twenty years. In particular, we use the results generated for the Reference Case in the EIA's Annual Energy Outlook 2014 [65].

In the Annual Energy Outlook, 2012 served as the base year, from which projections were then made for subsequent years. As such, we use 2012 as the beginning of our time horizon for our case study. Relevant parameters were obtained from the year 2012 up till 2032, which gave the required twenty-year horizon. Each year was divided into four quarters, and linear interpolation was performed to determine the intermediate values for quarters that were lying in between the beginning of each year.

Supply

To determine the number of wells available in the Bakken for monetization, we first obtained the data from the Total Lower 48 Wells Drilled series of the Oil and Gas Supply table for the time horizon, and normalized each value such that the 2012 number corresponded with the actual number of new oil wells which appeared in the Bakken for the year 2012 - a total of 1,773 wells - a statistic obtained from the North Dakota Department of Mineral Resources [66].

The values were then translated into a per quarter basis and linear interpolation was performed to obtain intermediate values. We then divided the values by four to obtain the number of gas sources available per year, where each gas source is defined as four gas wells.

The question of what percentage of these gas sources are actually available for the decision maker to monetize is an open one, as it depends on a variety of factors, such as the level of competition from other entrants into the market, the level of substitute uses of associated gas that the producers might consider, and the build up of infrastructure at several locations which might be able to carry away the gas in long-distance pipelines. Such a value is highly subjective and depends on the specific circumstances of the decision maker. For purposes of our study, we chose a conservative percentage of 1.6%. That is, we assumed that the decision maker has access to 1.6% of these gas sources for monetization purposes, and rounded the resulting number of gas sources down to the nearest integer.

Even with such a seemingly low percentage, informative results can be obtained. This is because of the long time horizon which, as a result of an average of one or two gas sources becoming available for monetization per quarter, leads to a total of 131 gas sources under consideration for the entire time horizon, which is a significant

amount for the decision maker to make allocation decisions on. As will be seen later in the results section, the project is already profitable under these conservative assumptions, and assuming a higher percentage of gas source availability is only expected to increase profitability. Figure 2-6 depicts the number of wells that come online per quarter which are available for monetization. At the point in time at which each source comes online, the supply of gas assumes the profile shown in Figure 2-5.

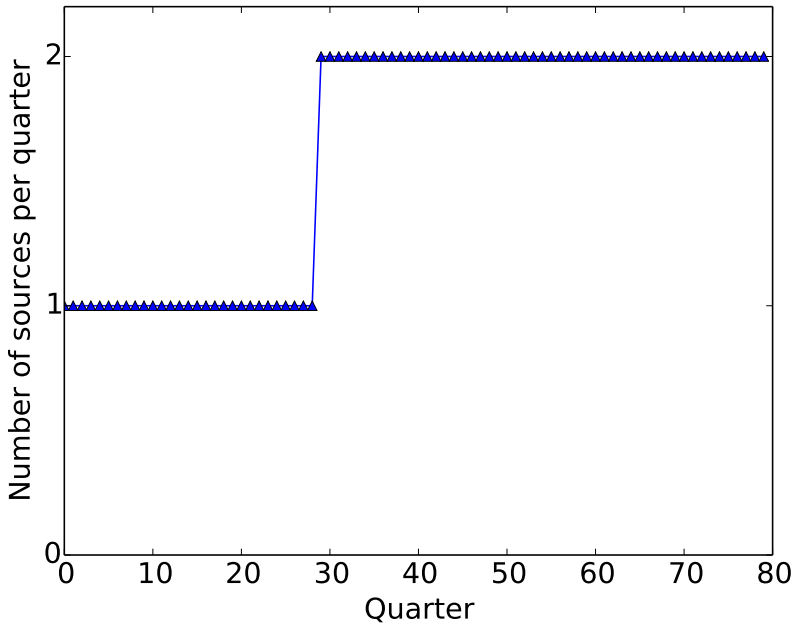


Figure 2-6: Number of gas sources coming online per quarter.

Demand and prices

For each market under consideration, we determined the census division it belonged to (refer to the Supplementary Material for the corresponding mappings), and thereby determined the appropriate demand and price parameters given by the

2014 Annual Energy Outlook's projections.

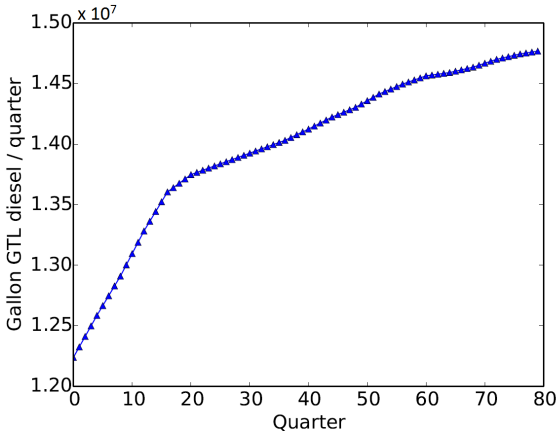
Demand forecasts for GTL diesel was determined by first obtaining the Energy Consumption for Distillate Fuel Oil for Transportation data series in the Energy Consumption by Sector and Source table for the time horizon. Then, all values were normalized such that the value for the base year 2012 corresponded to the baseline demand for GTL diesel at each of the markets. This served as the forecasted demand for GTL diesel for our case study. Price forecasts for GTL diesel were determined by the Wholesale Price of Diesel for the Transportation Sector series, obtained from the Components of Selected Petroleum Product Prices table for every scenario.

Demand forecasts for LNG was determined by first obtaining the Natural Gas Consumption by the Transportation sector data series for the time horizon, and again, all values were normalized such that the value for the base year 2012 corresponded to the baseline demand for LNG at each of the markets. This served as the forecasted demand for LNG for our study. The price forecasts of LNG was determined by first obtaining the Natural Gas Delivered Prices series for the Transportation sector. In order to obtain the wholesale prices for LNG, dispensing costs, federal taxes and state taxes were deducted. These values were obtained from correspondence with an EIA representative. Deducting these costs, we arrived at the projections of LNG wholesale prices used in our study.

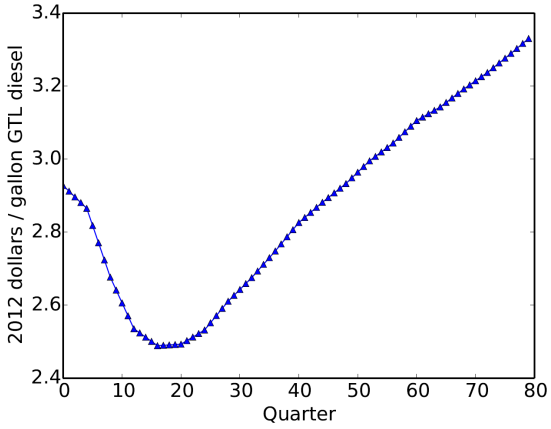
Finally, demand forecasts for NGLs were determined by the Consumption of Liquefied Petroleum Gases and Other for the Industrial Sector data series for the time horizon, and values were normalized such that the value for 2012 corresponded to the baseline demand for NGLs at the markets. Price forecasts for NGLs were set at the prices for Propane in the Industrial Sector.

Figure 2-7 depicts the final values of the demand and price forecasts for a selected market for each of the products that were used in the case study. A complete set of

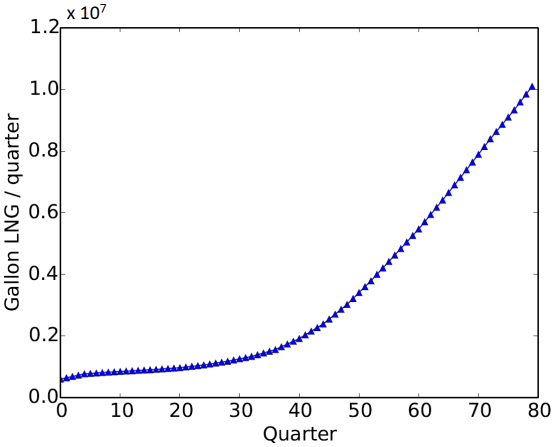
forecasts for all the markets can be found in the Supplementary Material.



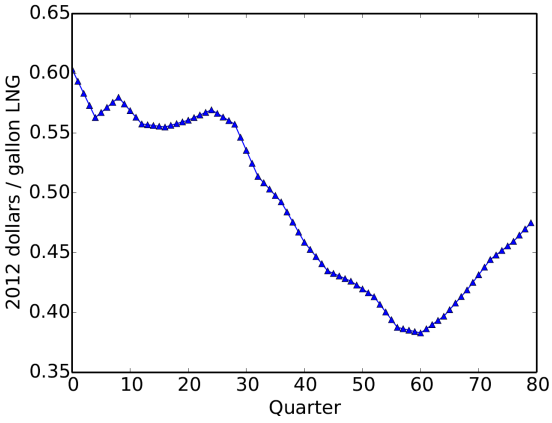
(a) Demand forecast at GTL diesel market 1.



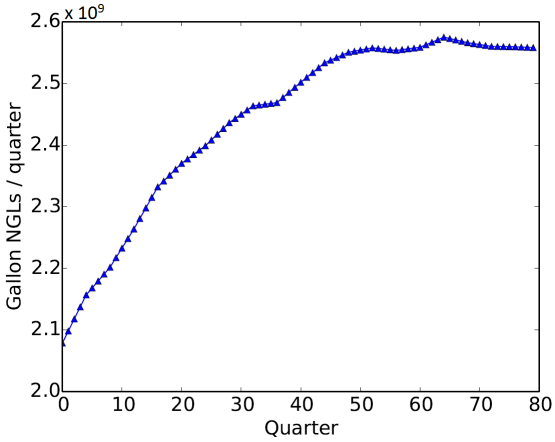
(b) Price forecast at GTL diesel market 1.



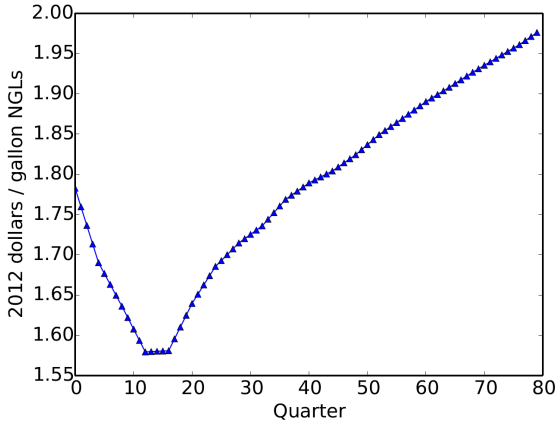
(c) Demand forecast at LNG market 1.



(d) Price forecast at LNG market 1.



(e) Demand forecast at NGLs market 1.



(f) Price forecast at NGLs market 1.

Figure 2-7: Demand and price forecasts generated at a selected market for each product under the Reference case. The complete set of forecasts for every market can be found in the Supplementary Material.

2.3.6 Results and discussion

With all the parameters specified, we solved the dynamic mobile plant allocation problem as presented in the Mathematical Formulation section. In addition, the required annual rate of return was set to 10%, compounded quarterly. This is a slightly conservative estimate over the typical weighted-average cost of capital for projects in the Oil and Gas industry in the world, at 8.6% (estimated by Deloitte) and 9.7% (estimated by Bloomberg), as reported by Deloitte [67].

The resulting instance of the MILP formulation consisted of 971,600 variables, of which 122,720 were integer, and 464,989 constraints. The optimization problem was solved on an Intel Xeon E5-1650 3.20 GHz machine with 12 GB of RAM using IBM ILOG CPLEX 12.6 accessed by the CPLEX Python API [31]. CPLEX is an optimization software which includes a general-purpose MILP solver that employs a branch-and-cut procedure to solve the MILP. Given the finiteness of the set of integers in the problem, a global optimum is guaranteed to be achieved in a finite number of steps. The optimal solution with 0% gap was found in 6.8 seconds.

A maximum net present value (NPV) of $1.73 billion over the twenty-year horizon could be achieved with the optimization framework. This required the overall purchase of 14 small- and 6 medium-sized GTL plants and no LNG plants. Figure 2-8 shows the cumulative NPV of the project. The discounted payback period is 2 years.

The option of using mobile plants of different sizes also allows for an interesting optimal strategy to present itself. Figure 2-9 displays the optimal plant deployment decisions at several time points for the Reference case. Each cell in the grid represents a gas source and the shade intensity reflects the current rich gas output. The presence of a small-sized square (small GTL plant) or medium-sized square (medium GTL

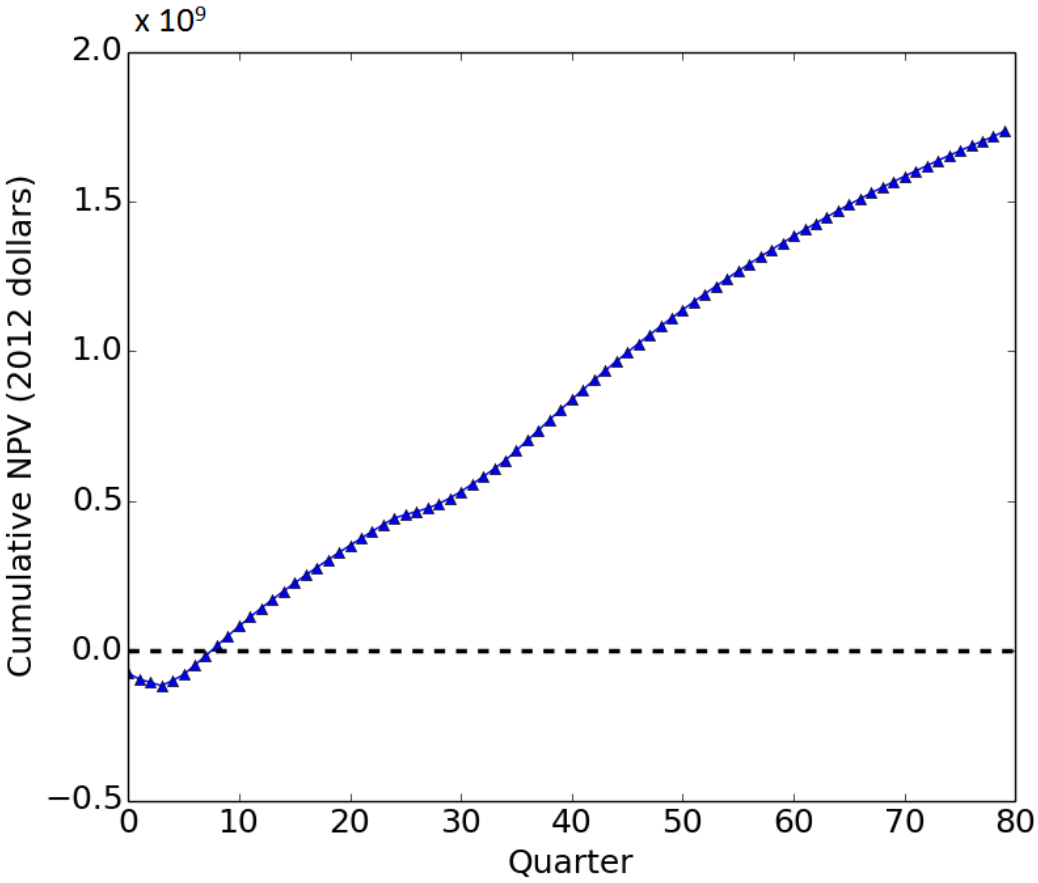
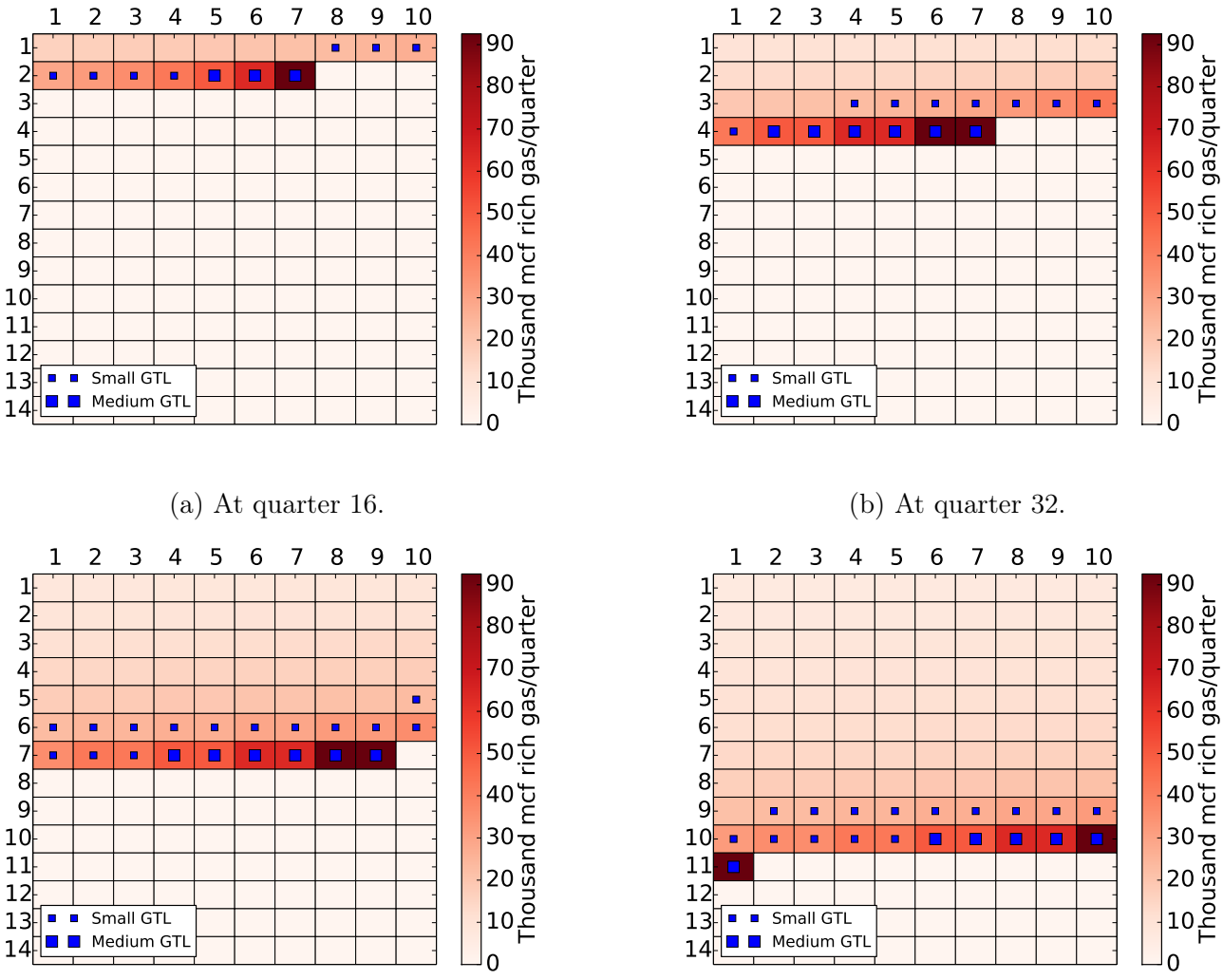


Figure 2-8: Cumulative NPV determined by solving the dynamic allocation of mobile plants optimization model.

plant) indicates that a plant of the corresponding type is currently being deployed at the gas source.



(a) At quarter 16.

(b) At quarter 32.

(c) At quarter 48.

(d) At quarter 64.

Figure 2-9: Optimal plant deployment decisions at several points in the time horizon. Each cell in the grid represents a gas source with its corresponding gas flow rate. The presence of a square in the cell indicates that a plant has been deployed at that gas source for the current point in time.

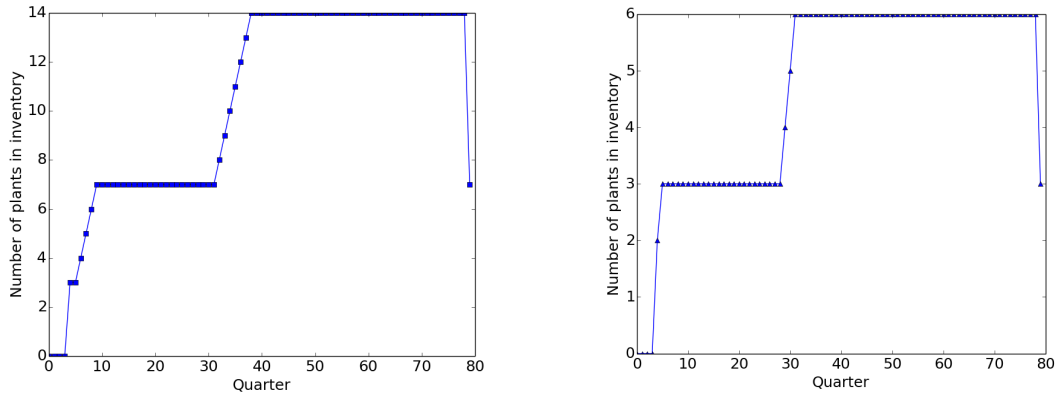
We can see that the optimal strategy involves first allocating the medium-sized plants to the sources with the highest gas outputs. However, once the output of a gas source is reduced below a certain level, the decision is made to move the medium-sized plant away from its current source to a newer source, while at the same time, bring in a small-sized plant to the current source in order to further monetize it. Such a strategy clearly brings to light the great value in the degree of flexibility mobile plants possess when monetizing gas sources of a transient nature.

Figure 2-10 shows the changes in the number of plants of each type and size in the decision maker's inventory over time. Again, the ability of the decision maker to buy and sell plants at any point in time allows him or her to easily expand or contract capacity to accommodate changes in demand, supply and price conditions over time. In this case, overall economic conditions in the Reference scenario dictate that opportunities for profitability become more available further into the time horizon, and the optimal solution reflects this by buying more plants as a response to this change. At the final time point, it is more profitable to sell the newer plants because of their higher salvage values, whereas it would be more profitable to continue operating the older plants because of their lower salvage values.

To further demonstrate that such profitability mainly arises from the ability of mobile plants to be deployed dynamically to counter the effects of steep decline curves, we performed an alternative case study where the plants were not allowed to be moved once they have been deployed. This was done by introducing the following constraint into the optimization model:

$$y_{ij}^{t} \leq y_{ij}^{t+1}, \quad \forall i, j, \quad \forall 0 \leq t < T. \tag{2.27}$$

With Constraint (2.27) included, the NPV declined drastically to \$46.4 million,



(a) Small GTL plants.

(b) Medium GTL plants.

Figure 2-10: Number of plants in inventory over time. Plants of other types and sizes were not purchased.

which is 2.7% of the original maximum NPV. In addition, the purchase decisions were made towards the end of the time horizon, particularly because of the relatively high product prices and demand at those time points. For most of the time horizon, no purchases were made simply because the constraint of fixed locations made it unprofitable to operate the plants. This clearly demonstrates the distinct advantage that mobility brings to the table.

Another interesting area to explore is that the optimal solution favored GTL plants over LNG plants to the point of excluding any purchases of LNG plants. We wanted to see if this was due to LNG plants being uneconomical in as of themselves, or if the constraints of having a fixed number of gas sources to monetize, coupled with the more favorable economics of GTL over LNG led to the resulting exclusion, even though LNG plants might be economical in as of themselves.

To address this question, we conducted another case study where we excluded the availability of GTL plants from the choice of technologies. Thus, the decision maker has only access to small, medium and large LNG plants to purchase, deploy and

operate. Solving this version of the problem led to a maximum NPV of $0.79 billion, and an overall inventory of 14 small- and 6 medium-sized LNG plants. Again, the optimal strategy of allocating the small plants to the gas sources previously occupied by the medium plants was employed. We therefore conclude that the LNG plants are an economical option, and were only not selected in the original formulation because of the more attractive economics of the GTL plants.

Sensitivity studies were also performed in order to see how our conclusions would change with our assumptions of the technology characteristics. In particular, costs assumptions for each plant were varied from their base value and the resulting optimal NPV and overall optimal number of GTL plants and optimal number of LNG plants were recorded.

We varied the following in the sensitivity studies:

1. Capital costs at 50, 75, 100, 125, and 150% of base value.
2. Operating costs at 80, 90, 100, 110 and 120% of base value.
3. Shipping costs at 80, 90, 100, 110 and 120% of base value.
4. Startup and shutdown costs at 50, 75, 100, 125 and 150% of base value.
5. Conversion efficiencies at 95, 97.5, 100, 102.5 and 105% of base value.

These were chosen to reflect a reasonable range in which these parameters could take values. For each item in the list, the parameters were varied independently for each technology. Since there were two technologies under consideration, and five different values for each item, there were $5^2 = 25$ studies for each item in the list. Since there are 5 items on the list, a total of 125 studies were carried out (allowing for the double-counting of the cases where the parameters were at 100% of the base value for both technologies).

The results from the sensitivity studies show that in all 125 cases, the same number and type (14 small GTL and 6 medium GTL) of plants were purchased over the entire time horizon. In addition, the maximum NPV varied within a tight range of $1.64 to $1.83 billion. These results reflect that even allowing for inaccuracies in cost estimates and technology assumptions, the profitability of the project is minimally affected.

2.4 Concluding remarks

We have proposed a framework that allows a decision maker to allocate mobile plants dynamically to monetize stranded or associated gas under time-varying supply, price and demand such that the net present value of the project is maximized. In addition, we have demonstrated how this framework can be used in a real-world case study on the Bakken shale play, where the current amount of untapped associated gas presents a tremendous economic opportunity.

With our current assumptions on GTL and LNG technologies and the nature of supply, demand and prices, we conclude that utilizing mobile plants indeed offers a profitable and flexible method to monetize associated gas in the Bakken, resulting in a NPV of $1.73 billion over a twenty-year time horizon with a 10% annual required rate of return, compounded quarterly. The expected discounted payback period was 2 years. In addition, both technologies are currently attractive and the optimal strategy involves the application of medium-sized GTL plants to gas sources with higher initial outputs, followed by smaller-sized GTL plants to the same sources when the gas outputs have lowered.

The mobility of the plants was identified as a key ingredient leading to the high profitability, as an alternative case study where the plants were forced to remain in

one location dramatically reduced the profitability of the project. In another case study, we found that LNG technologies could also be a profitable technology option in the absence of GTL technologies. Finally, sensitivity studies of technology cost and conversion parameters over a large range indicated that the optimal NPV still remained profitable.

The intention of our study is to be descriptive rather than prescriptive. That is, we hope that our case study provides a guideline as to how a decision maker could apply our framework to a real-world situation. Ultimately, the results and degree of profitability of implementing mobile plants would depend on the unique circumstances of the decision maker (i.e., the specific technologies, gas supply options and potential markets that are accessible to him or her). Although the Bakken is currently the most active shale play where associated gas is present, more opportunities might open up in the future to apply this framework to, as more shale fields are developed in the U.S., or further down the road, in the rest of the world. In addition, many stranded gas fields around the world which possess short-lived gas supplies can now undergo a similar method of analysis to determine their profitabilities. Developers of the mobile plant technologies can also utilize this framework to identify key areas for improvement of their proposed technologies which would offer them significant opportunities to enhance their economic competitiveness.

In practice, our framework should be implemented with a rolling time horizon. For example, although the optimizer determines the optimal decisions for every time point of the twenty-year horizon, in practice the decision maker would only implement the decisions for the first year. At the end of the first year, the optimization problem can then be re-solved with updated initial conditions and new parameters for the twenty-year time horizon now shifted one year ahead. This iterative process will continue indefinitely, stepping forward one year at each iteration. A rolling horizon

implementation would allow decisions to better reflect changes in current conditions and forecasts with the passage of time.

One issue that has not been addressed in this chapter is the presence of uncertainty. That is, would the mobile plants perform as well if we were to consider that our supply, price and demand estimates could vary significantly differently from the values that we had assumed? This is a typical issue which plagues large-scale, fixed-location plants in the oil and gas industry, since it is extremely difficult to obtain accurate forecasts of the future. In the next chapter, we shall explore how the introduction of uncertainty affects our optimization framework and how robust these plants could be in dealing with uncertainty.

Chapter 3

Small-Scale Mobile Plants: Addressing Uncertainty

Using the Bakken shale play as a case study, the previous chapter demonstrated how small-scale mobile plants could be used to monetize associated or stranded gas effectively. Here, we address the issue of uncertainty in future supply, demand and price conditions. To this end, we modified our multi-period optimization framework to a stochastic programming framework to account for various scenarios of different parameter realizations in the future. The maximum expected net present value (ENPV) obtained was $2.01 billion, higher than the NPV obtained in the previous part. In addition, by comparing the value of the stochastic optimal solution to that of the deterministic method, we determine that the flexible nature of mobile plants affords them a great advantage when dealing with uncertainty.

3.1 Motivation

We add to the complexity of the case study in the previous chapter by introducing uncertainty into various parameters that would ultimately affect the profitability of the entire undertaking. First, typically complicating the decision process of investments in the oil and gas industry is the considerable uncertainty in the estimated ultimate recovery (EUR) of gas from wells being drilled. As mentioned by the U.S. Energy Information Administration (EIA) [65], this problem is significantly acute for the case of unconventional resources where the data collected on production patterns thus far have not been enough to estimate production rates far into the future reliably. In addition, this uncertainty in EUR also impacts the predictions of drilling patterns of future wells in the play.

A second major source of uncertainty lies in the prices and demand for the finished products, as they directly impact the revenue generated from the decision maker's activities. As the markets for the finished products are very large, the decision maker is essentially a price taker in these markets. Predicting the future prices and demand for oil and gas-based products is extremely difficult and as can be seen from past experience, can be very wrong. An example would be the flurry of activity a decade ago to increase LNG import capacity with the expectation of future shortages of domestic natural gas production, as reported by White [26]. Hence, the decision maker has to live with the uncertainty with regards to prices and demands and make his or her decisions under such conditions.

To make decisions under uncertainty adequately, a stochastic programming framework should be used. Readers are referred to Shapiro [68] for a comprehensive reference on the subject. In this framework, several price, supply and demand scenarios are projected for the future. The decision variables are partitioned into two sets: the

'here-and-now' decisions, which have to be made before the scenarios are realized, and the 'wait-and-see' (or 'recourse') decisions, which are made once a particular scenario has been realized. The objective function for the stochastic program would typically be to maximize the expected net present value (ENPV) of the project, which is the sum of the net present value of every scenario weighted by its associated probability.

3.2 Stochastic formulation

The modification of the optimization problem from the previous chapter is straightforward - we retain the previous indices: time stages $t \in \{0, \ldots, T\}$, gas sources $i \in \{1, \ldots, I\}$, plant type $j \in \{1, \ldots, J\}$, markets $k \in \{1, \ldots, K\}$ and products $l \in \{1, \ldots, L\}$. We then introduce a new index for scenarios $s \in \{1, \ldots, S\}$.

The optimization decisions are:

1. Decision to allocate plant of type j to source i at time t of scenario s, denoted by $y_{ij}^{ts} \in \{0, 1\}$.

2. Indicator of the presence of a gas gathering system at source i at time t of scenario s, denoted by $z_i^{ts} \in \{0, 1\}$.

3. Gas feed rate to plant of type j at source i at time t of scenario s, denoted by $x_{ij}^{ts} \in \mathbb{R}_+$.

4. Product delivery rate of product l from source i to market k at time t of scenario s, denoted by $w_{ikl}^{ts} \in \mathbb{R}_+$.

5. Number of plants of type j purchased at time t of scenario s, denoted by $Buy_j^{ts} \in \mathbb{Z}_+$.

6. Number of plants of type j which originally arrived in inventory at time $0 \leq \tau < t$ of scenario s, sold at time t, denoted by $Sell_{j\tau}^{ts} \in \mathbb{Z}_+$.

7. Inventory of plants of type j at time t of scenario s, arriving in inventory at time $0 \leq \tau \leq t$, denoted by $Inv_{j\tau}^{ts} \in \mathbb{Z}_+$.

Essentially, to each constraint in the previous chapter, we replace it with its stochastic counterpart simply by requiring that each constraint holds for each individual scenario. For brevity, we directly list the constraints here and refer the reader to the previous chapter for detailed explanation of each set of constraints.

$$z_i^{ts} \leq z_i^{t+1,s}, \quad \forall i, s, \quad \forall 0 \leq t < T. \tag{3.1}$$

$$y_{ij}^{ts} \leq z_i^{t-\mathcal{T}_g,s}, \quad \forall i, j, s, \quad \forall t \geq \mathcal{T}_g, \quad \text{and} \tag{3.2}$$

$$y_{ij}^{ts} = 0, \quad \forall i, j, s, \quad \forall t < \mathcal{T}_g. \tag{3.3}$$

$$Inv_{j\tau}^{ts} = Buy_j^{t-\mathcal{T}_j,s}, \quad \forall j, s, \quad \forall t \geq \mathcal{T}_j, \quad \forall \tau = t, \quad \text{and} \tag{3.4}$$

$$Inv_{j\tau}^{ts} = 0, \quad \forall j, s, \quad \forall t < \mathcal{T}_j, \quad \forall \tau = t.$$

$$Inv_{j\tau}^{ts} = Inv_{j\tau}^{t-1,s} - Sell_{j\tau}^{ts}, \quad \forall j, t, s \quad \forall \tau < t. \tag{3.5}$$

$$\sum_i y_{ij}^{ts} \leq \sum_{\tau=0}^{t} Inv_{j\tau}^{ts}, \quad \forall j, t, s. \tag{3.6}$$

$$x_{ij}^{ts} \leq s_i^{ts} y_{ij}^{ts}, \quad \forall i, j, t, s. \tag{3.7}$$

$$\sum_j x_{ij}^{ts} \leq s_i^{ts}, \quad \forall i, t, s. \tag{3.8}$$

$$x_{ij}^{ts} \leq m_j y_{ij}^{ts}, \quad \forall i, j, t, s. \tag{3.9}$$

$$x_{ij}^{ts} \geq \gamma_j m_j y_{ij}^{ts}, \quad \forall i, j, t, s. \tag{3.10}$$

$$\sum_j \alpha_{jl} x_{ij}^{ts} = \sum_k w_{ikl}^{ts}, \quad \forall i, l, t, s. \tag{3.11}$$

$$\sum_i w_{ikl}^{ts} \leq d_{kl}^{ts}, \quad \forall k, l, t, s. \tag{3.12}$$

One key difference between the stochastic program and our original formulation is that we have to introduce a new set of constraints which are termed non-anticipativity constraints. Non-anticipativity refers to the fact that the decisions at any particular time stage have to be made based only on the information realized up to that particular stage.

Consider a scenario tree where different branches of the tree represent different scenarios of parameter realizations that might occur in the future. At present however, all scenarios share the same realization of parameters, as they can be directly observed. Therefore, in our optimal solution, decisions for the first-stage variables have to be identical for all scenarios to reflect this reality. This condition is imposed by the followng constraints:

$$y_{ij}^{0,s} = y_{ij}^{0,s+1}, \quad \forall i, j, \forall s \in \{1, \ldots, S-1\}. \tag{3.13}$$

$$z_{i}^{0,s} = z_{i}^{0,s+1}, \quad \forall i, \forall s \in \{1, \ldots, S-1\}. \tag{3.14}$$

$$x_{ij}^{0,s} = x_{ij}^{0,s+1}, \quad \forall i, j, \forall s \in \{1, \ldots, S-1\}. \tag{3.15}$$

$$w_{ikl}^{0,s} = w_{ikl}^{0,s+1} \quad \forall i, k, l, \forall s \in \{1, \ldots, S-1\}. \tag{3.16}$$

$$Buy_{j}^{0,s} = Buy_{j}^{0,s+1}, \quad \forall j, \forall s \in \{1, \ldots, S-1\}. \tag{3.17}$$

$$Sell_{j,0}^{0,s} = Sell_{j,0}^{0,s+1}, \quad \forall j, \forall s \in \{1, \ldots, S-1\}. \tag{3.18}$$

$$Inv_{j,0}^{0,s} = Inv_{j,0}^{0,s+1}, \quad \forall j, \forall s \in \{1, \ldots, S-1\}. \tag{3.19}$$

The objective function of the decision maker is to maximize the expected net present value (ENPV) of the project, given an appropriate discount factor r. This is done by

summing up the NPV for each scenario, weighted by its associated probability π_s:

$$\text{ENPV} = \sum_s \pi_s \sum_t (1+r)^{-t}(Rev^{ts} - Cost^{ts}), \tag{3.20}$$

where Rev^{ts} is the total revenue and $Cost^{ts}$ is the total cost at time t for scenario s.

Rev^{ts} is defined as the sum of the following terms:

$$\sum_i \sum_j \sum_k \sum_l p_{kl}^{ts} \alpha_{jl} x_{ij}^{ts}, \tag{3.21}$$

$$\sum_j \sum_{\tau=0}^{t} c_{salv,j\tau}^{t} Sell_{j\tau}^{ts}. \tag{3.22}$$

$Cost^{ts}$ is defined as the sum of the following terms:

$$\sum_j c_{cap,j} Buy_j^{ts}, \tag{3.23}$$

$$\sum_i c_{gathercap}(z_i^{ts} - z_i^{t-1,s}), \tag{3.24}$$

$$\sum_i \sum_j c_{start,j} \max\{y_{ij}^{ts} - y_{ij}^{t-1,s}, 0\}, \tag{3.25}$$

$$\sum_i \sum_j c_{shut,j} \max\{y_{ij}^{t-1,s} - y_{ij}^{ts}, 0\}, \tag{3.26}$$

$$\sum_i \sum_j c_{opFixed,j} y_{ij}^{ts}, \tag{3.27}$$

$$\sum_i \sum_j c_{opVar,j} x_{ij}^{ts}, \tag{3.28}$$

$$\sum_i \sum_k \sum_l c_{ship,ikl} w_{ikl}^{ts}, \tag{3.29}$$

$$\sum_i \sum_j p_{gas,i}^{ts} x_{ij}^{ts}. \tag{3.30}$$

In the case of associated gas otherwise flared if not monetized, $p^{ts}_{gas,i} = 0$.

The max functions (3.25) and (3.26) are typically represented by introducing auxiliary variables $\delta^{ts}_{start,ij}, \delta^{ts}_{shut,ij} \in \mathbb{R}_+$ and the constraints:

$$y^{ts}_{ij} - y^{t-1,s}_{ij} \leq \delta^{ts}_{start,ij}, \quad \forall i, j, t, s, \tag{3.31}$$

$$y^{t-1,s}_{ij} - y^{ts}_{ij} \leq \delta^{ts}_{shut,ij}, \quad \forall i, j, t, s, \tag{3.32}$$

which then allow (3.25) and (3.26) to be represented by $\sum_i \sum_j c_{start,j} \delta^{ts}_{start,ij}$ and $\sum_i \sum_j c_{shut,j} \delta^{ts}_{shut,ij}$ respectively. Note that when $t = 0$, we set $y^{t-1,s}_{ij}$ to 0.

The stochastic program is summarized by maximizing (3.20) subject to the constraints of Eqs. (3.1) to (3.19) and (3.31) and (3.32).

3.3 Scenario generation

Investments in the oil and gas industry are risky because they rely on favorable supply, prices and demand conditions for extended periods of time in order to recoup the initial investment. Depending on the scenarios of parameter realizations assumed, drastically different results could be obtained. The challenge lies in generating reasonably accurate scenarios such that adequate planning can be made. This is a very challenging task, especially in the field of energy, where the correlations between supply, prices and demand arise due to highly complex interactions at many levels - geographical, economical and political, to name a few.

The EIA has developed a computer-based energy-economy modeling system for the U.S., known as the National Energy Modeling System (NEMS) [64]. NEMS has, in recent years, utilized data from public and private sources to generate plausible scenarios of prices, production, trade and consumption of energy which might play

out through 2040, given different assumptions of macroeconomic factors, world energy markets, resource availability, technological choices, demographics and public policy. These projections are reported annually in the EIA's Annual Energy Outlook [65], which is one of the most quoted resources in the U.S. with regards to energy-related data.

In Annual Energy Outlook 2014, all scenarios begin with a common set of data realizations from the year 2012. From 2013 onwards, the data varies depending on which scenario is assumed. For our case study, we will consider the three most pertinent scenarios: the Reference, High Oil and Gas Resource, and Low Oil and Gas Resource cases [65]. These three cases were chosen because of the significant differences in how the supply, demand and prices would change in the future, depending on their assumptions.

In the Reference case, the EIA assumes a business-as-usual scenario, given all known technological and demographic trends. Further, it assumes that current laws and regulations affecting the energy sector are largely unchanged throughout the projection period. In particular, the Reference case estimates the total unproved technically recoverable crude oil resources at 209 billion barrels, and the total unproved technically recoverable dry natural gas resources at 1,932 tcf.

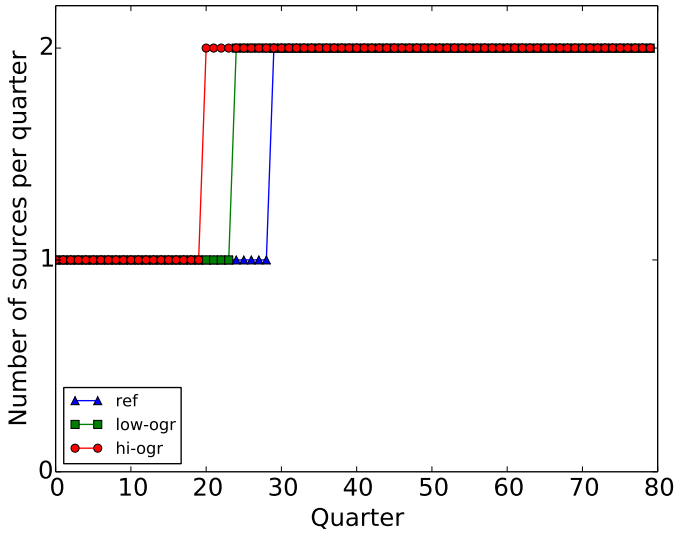
The High and Low Oil and Gas Resource cases differ in their assumptions from the Reference case that result in higher and lower estimates of the technically recoverable crude oil and natural gas resources respectively. For the High Oil and Gas Resource case, assumptions include 50% higher EURs for tight oil, tight gas and shale gas wells, long-term technology improvements resulting in an addition 1% increase in the aforementioned EURs, 50% lower well spacing, diminishing returns on EURs beyond a certain well count limit, more resources in Alaska and areas offshore, and the development of onshore oil shale. The Low Oil and Gas Resource case simply

assumes that EURs for tight oil and gas and shale wells are 50% lower than in the Reference case.

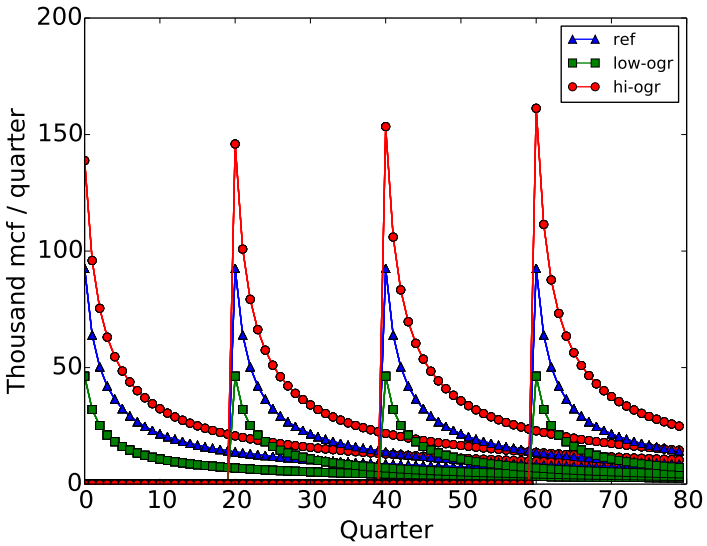
For each of the scenarios, relevant parameters (as explained in detail later) were obtained from the year 2012 (as the base year) up until 2032, which gave the required twenty year horizon for our study. Each year was divided into four quarters, and linear interpolation was performed to determine the intermediate values for quarters that lie between the beginning of each year.

3.3.1 Supply

Using the assumptions detailed in the previous chapter, the time evolution of the number of gas sources coming online per quarter and the corresponding supply curve for each gas source were then created. The supply curve for each gas source will vary according to the assumptions for each scenario. For the Low Oil and Gas Resource case, the supply curve was reduced by 50% from the Reference case. For the High Oil and Gas Resource case, the supply curve was increased by 50% from the Reference case and in addition, depending on the year in which the gas source came online, enjoyed an annual 1% increase in the output of its supply, compounded quarterly, for the time period between the base year (2012) and the time point in which it came online. This took into account the long-term technological growth assumption for the scenario. After these calculations, the final supply conditions used in our case study are depicted in Figure 3-1. Figure 3-1a depicts the number of wells which come online per quarter and Figure 3-1b depicts the associated gas output from gas sources which come online at a few selected time points for the different scenarios under consideration.



(a) Number of gas sources coming online per quarter.



(b) Decline curves for sources coming online at quarters $0, 20, 40$ and 60.

Figure 3-1: Supply characteristics generated for the various scenarios under consideration: Reference (ref), Low Oil and Gas Resource (low-ogr) and High Oil and Gas Resource (hi-ogr).

3.3.2 Demand and prices

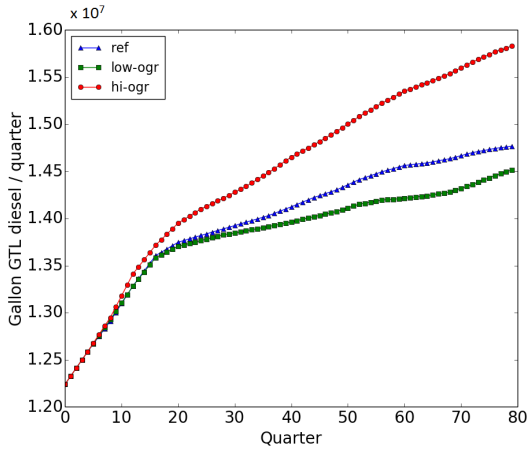
Using the assumptions and methodology detailed in the previous chapter, the corresponding demand and prices for the various products in their respective markets were created for all three scenarios. Figure 3-2 depicts the final values of the demand and price forecasts for a selected market for each of the products that were used in the case study. A complete set of forecasts for all the markets can be found in the Supplementary Material.

3.4 Results and discussion

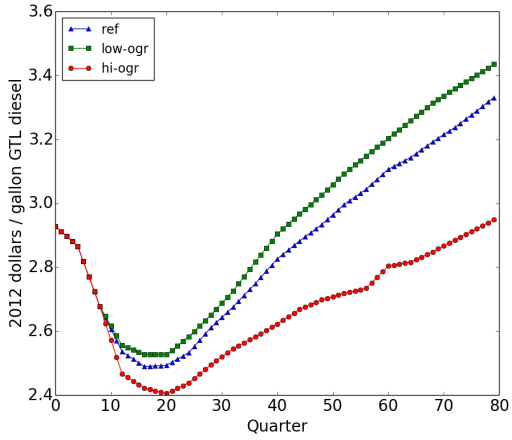
With the introduction of new parameters corresponding to each scenario, we solved the stochastic program with the same inputs for all non-stochastic parameters as in the previous chapter. Since there is no specific bias as to which scenario will materialize, each scenario was assigned an equal probability of 1/3.

The full stochastic program instance was an MILP consisting of 3,014,480 variables, of which 377,120 were integer, and 1,462,309 constraints. The optimization problem was solved on an Intel Xeon E5-1650 3.20 GHz machine with 12 GB of RAM using IBM ILOG CPLEX 12.6 accessed by the CPLEX Python API [31]. The optimal solution with 0% gap was found in 42.1 seconds.

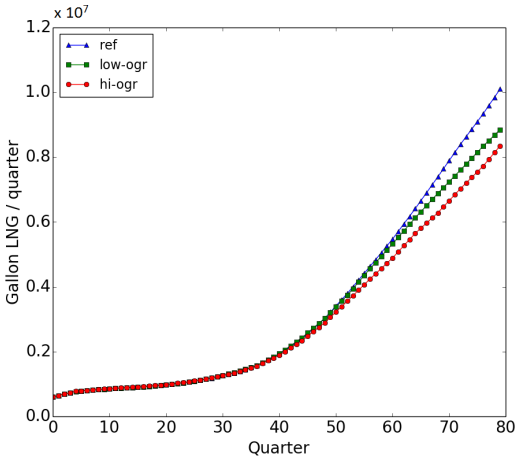
An expected net present value (ENPV) of $2.01 billion over the twenty-year horizon could be achieved with the optimization framework. The expected discounted payback period was slightly over 2 years. For the Reference scenario, the NPV was $1.73 billion, requiring the purchase of 14 small and 6 medium GTL plants and sales of 7 small and 3 medium GTL plants throughout the planning horizon. The corresponding figures for the Low Oil and Gas Resource case was an NPV of $0.44



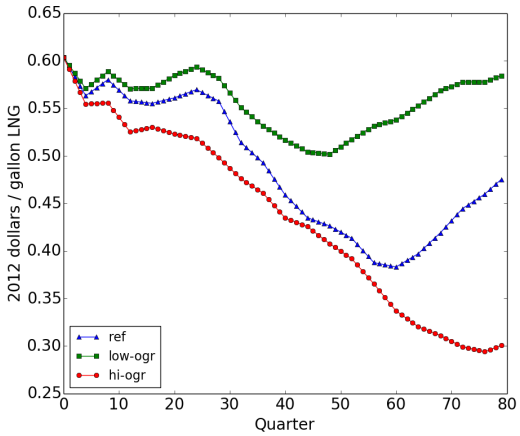
(a) Demand forecasts at GTL diesel market 1.



(b) Price forecasts at GTL diesel market 1.

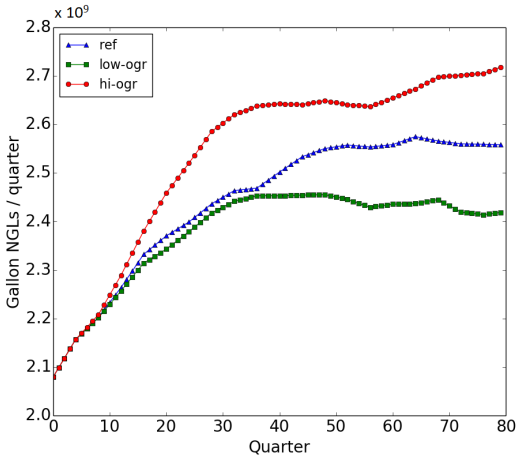


(c) Demand forecasts at LNG market 1.

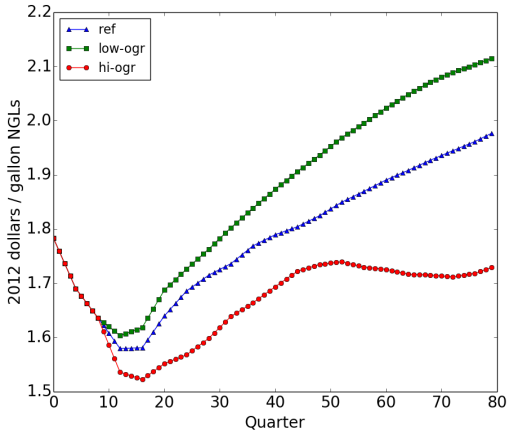


(d) Price forecasts at LNG market 1.

Figure 3-2: Demand and price forecasts generated at selected markets for the various scenarios under consideration: Reference (ref), Low Oil and Gas Resource (low-ogr) and High Oil and Gas Resource (hi-ogr). The complete set of forecasts for every market is found in the Supplementary Material.



(e) Demand forecasts at NGLs market 1.



(f) Price forecasts at NGLs market 1.

Figure 3-2: Demand and price forecasts generated at selected markets for the various scenarios under consideration: Reference (ref), Low Oil and Gas Resource (low-ogr) and High Oil and Gas Resource (hi-ogr). The complete set of forecasts for every market is found in the Supplementary Material.

billion and purchases (sales) of 6 (2) small and 3 (3) medium GTL plants, while that of the High Oil and Gas Resource case was an NPV of $3.85 billion and purchases (sales) of 32 (16) small, 10 (4) medium and 4 (0) large GTL plants. Figure 3-3 shows the cumulative ENPV, and, given the optimal decisions implemented for each scenario, the cumulative net present value (NPV) for each of the scenarios. It is important to note that even in the pessimistic Low Oil and Gas Resource case, the project is profitable. Also, consistent with the findings in the previous chapter, the optimal strategy typically involved the reallocation of smaller plants to gas sources that were occupied by larger plants previously. These strategies are depicted in Figures 3-4 to 3-6. In addition, Figure 3-7 depicts the flexbility of the decision maker to respond to changes in profitability by adjusting his or her inventory of plants over time.

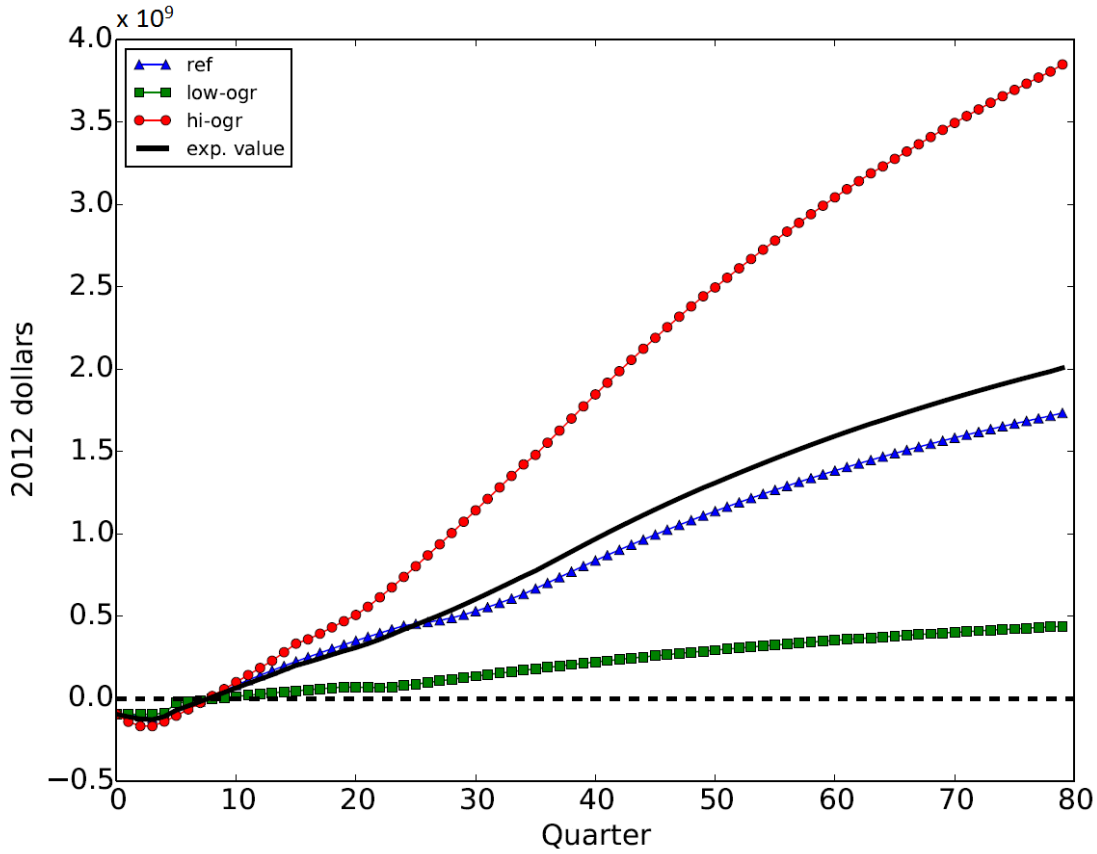
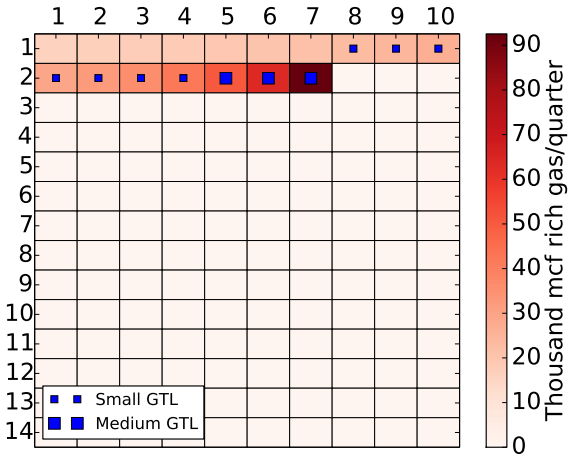
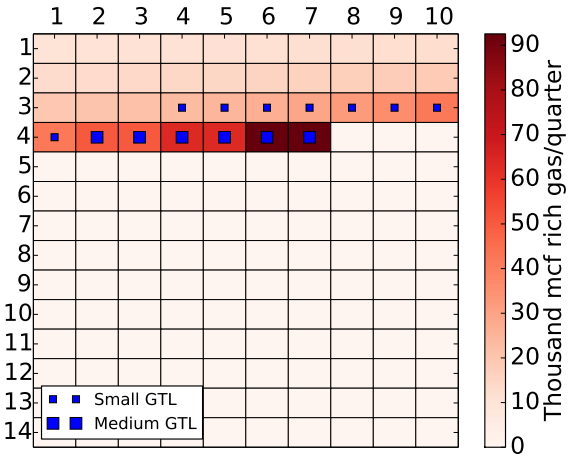


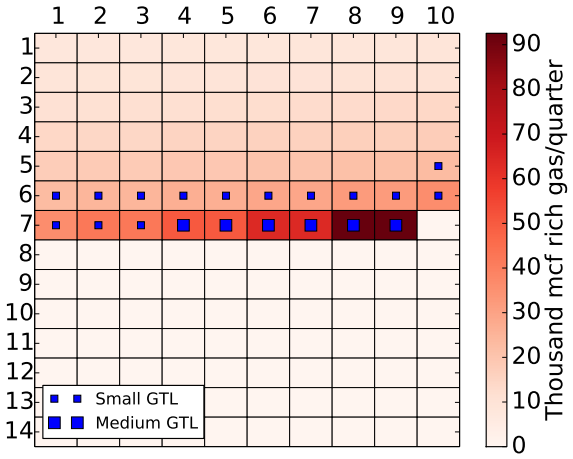
Figure 3-3: Cumulative ENPV and NPV for every scenario, determined by solving the stochastic program for the dynamic allocation of mobile plants. Scenarios: Reference (ref), Low Oil and Gas Resource (low-ogr) and High Oil and Gas Resource (hi-ogr).



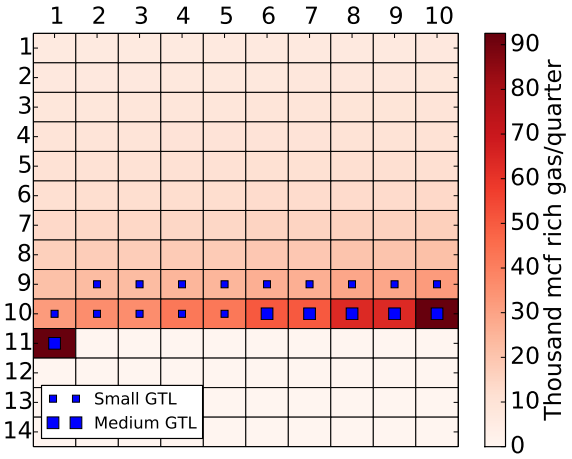
(a) At quarter 16.



(b) At quarter 32.

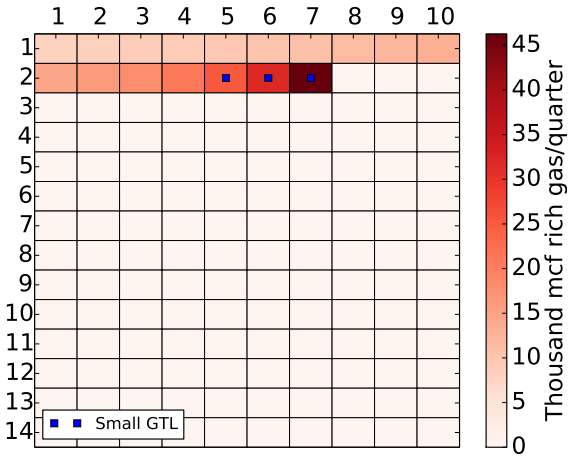


(c) At quarter 48.

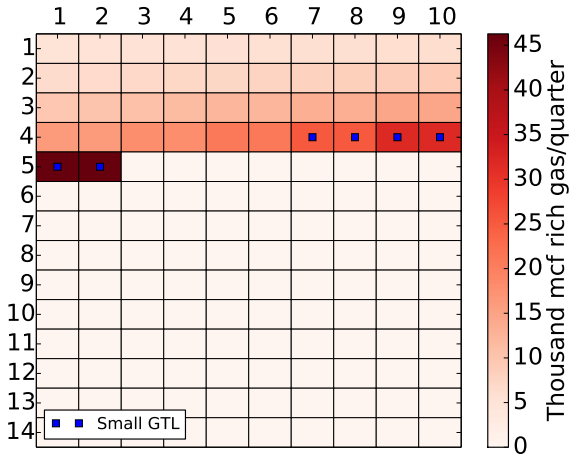


(d) At quarter 64.

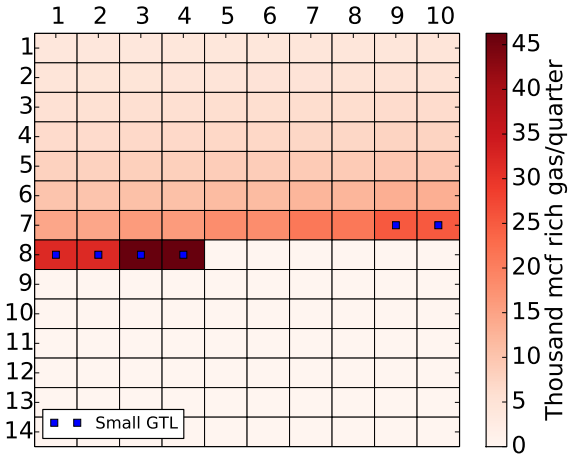
Figure 3-4: Optimal plant deployment decisions at several points in the time horizon for the Reference case. Each cell in the grid represents a gas source with its corresponding gas flow rate. The presence of a square in the cell indicates that a plant has been deployed at that gas source for the current point in time.



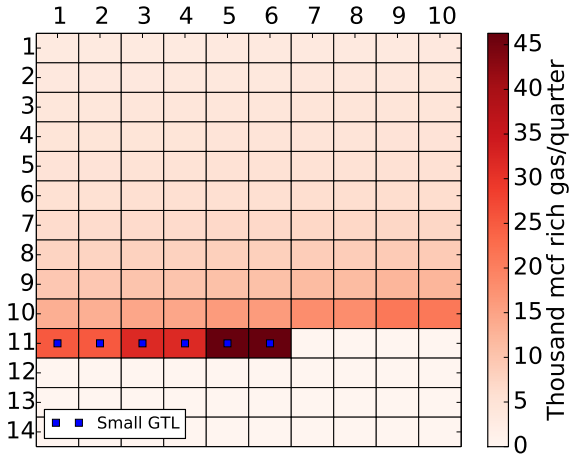
(a) At quarter 16.



(b) At quarter 32.

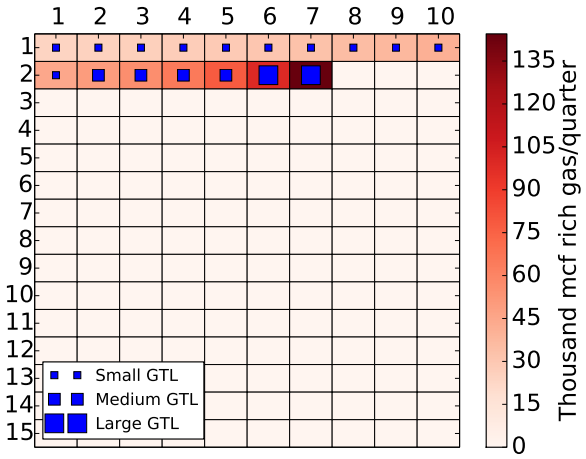


(c) At quarter 48.

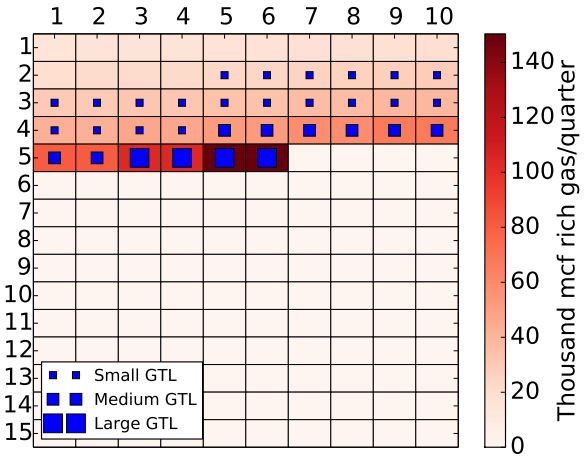


(d) At quarter 64.

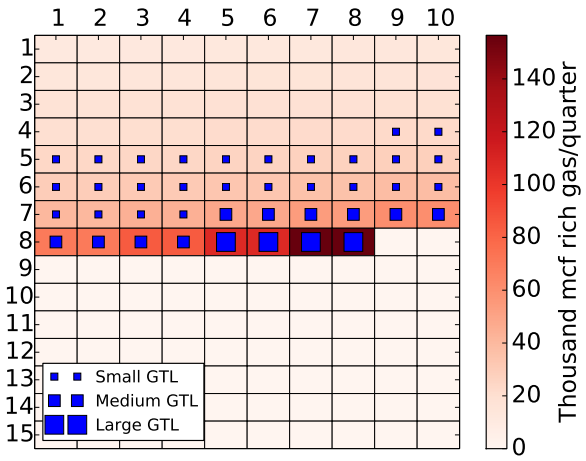
Figure 3-5: Plant deployment decisions at several points in the time horizon for the Low Oil and Gas Resource case. Each cell in the grid represents a gas source with its corresponding gas flow rate. The presence of a square in the cell indicates that a plant has been deployed at that gas source for the current point in time.



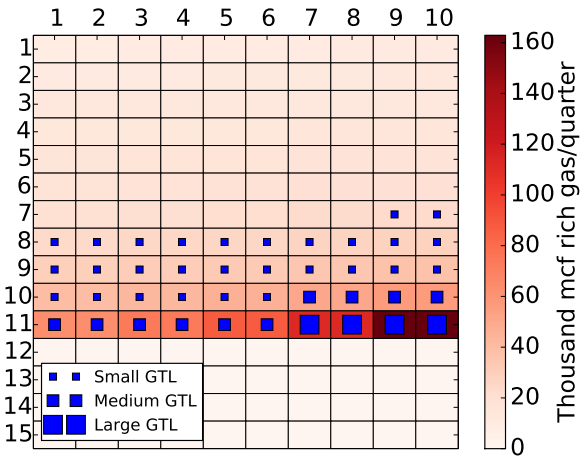
(a) At quarter 16.



(b) At quarter 32.

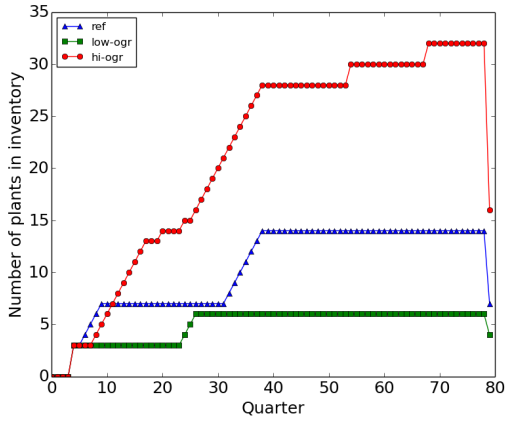


(c) At quarter 48.

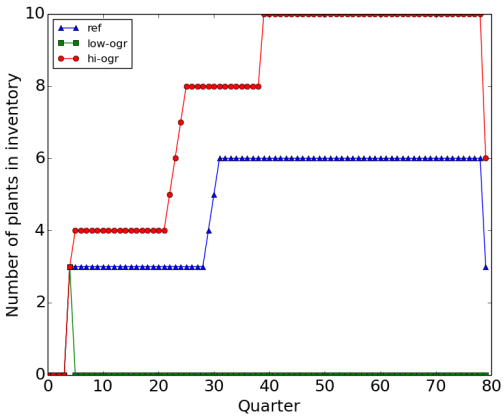


(d) At quarter 64.

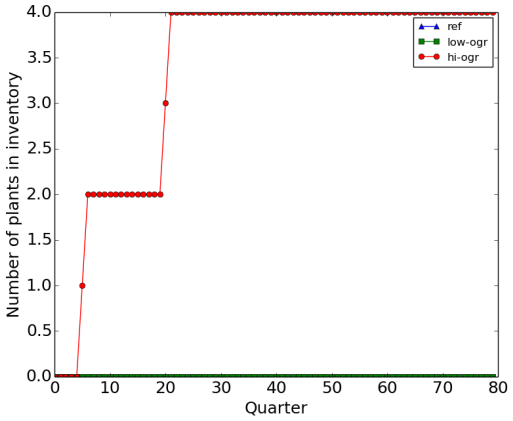
Figure 3-6: Plant deployment decisions at several points in the time horizon for the High Oil and Gas Resource case. Each cell in the grid represents a gas source with its corresponding gas flow rate. The presence of a square in the cell indicates that a plant has been deployed at that gas source for the current point in time.



(a) Small GTL plants.



(b) Medium GTL plants.



(c) Large GTL plants.

Figure 3-7: Number of plants in inventory over time. LNG plants were not purchased.

One of the main oft-claimed advantages of utilizing mobile plants as compared to traditional large-scale, fixed-location investments is their increased agility to better react to uncertain conditions. We will assess this by measuring a quantity known as the value of the stochastic solution.

To motivate this, let us concern ourselves with the issues typically faced with large-scale investments. That is, the profitability of a large-scale investment critically hinges on how accurately future earnings can be predicted. Once the investment decisions have been made, there is very little flexibility in terms of recourse in the operational decisions. Therefore, if supply, price and demand conditions turn out to be drastically different from a prediction on which the design was based, the project would be in jeopardy.

Sophisticated planners who try to mitigate this issue would therefore construct a scenario tree and use a stochastic program to determine the most optimal "here-and-now" decisions one has to make in the face of uncertainty. This solution is guaranteed to maximize the ENPV of the project (over the scenarios considered).

Compare this approach to a less sophisticated planner, who, although aware that different scenarios might play out in the future, does not implement the stochastic programming approach. Instead, he or she identifies a nominal scenario among all possible scenarios and solves for the optimal "here-and-now" decision based on this average scenario. Based on this "here-and-now" decision, he or she then tries his or her best to make the most optimal recourse decisions once the uncertainty has been realized at future time points.

The difference between the ENPV from approach of the sophisticated planner and the less sophisticated planner is termed the value of the stochastic solution. As might be expected, this value is largely determined by how much flexibility the decision maker possesses in the recourse decisions - the lower the flexibility, the

greater the value.

The value of the stochastic solution therefore gives an appropriate metric to judge the degree of flexibility mobile plants have to respond to uncertainty. Our aim is to show that this value is small, demonstrating that mobile plants have a high degree of robustness to uncertainty. The implications of this is that using mobile plants allows for the decision maker to be profitable even if a sub-optimal initial investment decision was made, so long as a reasonable nominal scenario can be constructed beforehand.

To determine the value of the stochastic solution, we first solve the optimization problem only for a single scenario that represents the most likely or average values the uncertain parameters would take. In our case, we chose the Reference case as this scenario.

After solving the problem for this single scenario, we noted the optimal decision variables made for the first time point of the time horizon. Then, fixing these first-stage decision variables, we solved the optimization problem for the High and Low Oil and Gas Resource cases independently, for the entire time horizon. In this way, we obtained what we termed the "deterministic" optimal solution value by summing up the optimal NPVs for each independent scenario, multiplied by their associated probabilities. By subtracting the deterministic optimal solution value from the optimal ENPV as determined from solving the entire stochastic program, we arrived at the value of the stochastic solution.

The value of the stochastic solution for our case study was determined to be \$2.23 million, which is only 0.11% of the optimal ENPV. This is very low compared to other case studies on large-scale, fixed-location investments. For example, a study by Li *et al.* in 2011 [69] on the Sarawak gas production system yielded a value of the stochastic solution of 9.94% of the optimal ENPV.

The robustness of mobile plants operating under uncertainty can be attributed to the great degree of flexibility in the recourse actions. First, their ability to be redeployed over time to different gas sources ensures that they closely track the gas sources which offer the highest flow rates, even if the gas sources by themselves deplete very quickly. Second, the ability to avoid pipeline infrastructure for transportation to markets ensures that the plants are not required to deliver to any specific market in order to recoup the investment made. Third, the feature of plants to be continuously purchased throughout the time horizon, due to the standardized manner in which they are constructed, coupled with short lead times, allows one to react to changing supply, price and demand conditions readily.

Sensitivity studies on our cost and conversion assumptions were also performed with the same ranges as in the previous chapter. The ranges studied were as follows:

1. Capital costs at 50, 75, 100, 125, and 150% of base value.

2. Operating costs at 80, 90, 100, 110 and 120% of base value.

3. Shipping costs at 80, 90, 100, 110 and 120% of base value.

4. Startup and shutdown costs at 50, 75, 100, 125 and 150% of base value.

5. Conversion efficiencies at 95, 97.5, 100, 102.5 and 105% of base value.

For each item in the list, the parameters were varied independently for each technology.

With this range of capital cost inputs, the optimal ENPV remained profitable from a low of \$1.89 billion to a high of \$2.12 billion. In addition, the value of the stochastic solution never exceeded 0.13% of the optimal ENPV.

3.5 Concluding remarks

We have demonstrated the effectiveness of mobile plants in scenarios where estimates of supply, price and demand conditions might differ significantly from one another. With the introduction of uncertainty, utilizing mobile plants to monetize associated gas in the Bakken results in an ENPV of $2.01 billion over a twenty year time horizon assuming a 10% annual required rate of return, compounded quarterly, with an expected discounted payback period of slightly over 2 years.

Analyzing the value of the stochastic solution leads us to conclude that mobile plants offer a robust way to be profitable even in uncertain conditions of supply, prices and demand, and this confirms their advantage over larger-scale, fixed location investments in this regard. Finally, sensitivity studies show that our conclusion remain valid for a reasonably large range of technological costs and conversion inputs.

Chapter 4

Shale Oil and Gas Investments in the United States

We present a comprehensive supply chain optimization framework to determine optimal shale oil and gas infrastructure investments in the United States. The framework encompasses multiple shale plays, commodities, plant locations, conversion technologies, transportation modes and both local and foreign markets. The dynamic evolution of supply, demand and price parameters and the uncertainty in parameter realizations are also fully taken into account. Assuming two different scenario sets over a time horizon of twenty-five years, we analyze the features of the optimal infrastructure investments and associated operating decisions. We also highlight the importance of incorporating uncertainty into the framework and analyze the stability of the stochastic solutions as the degree of uncertainty changes.

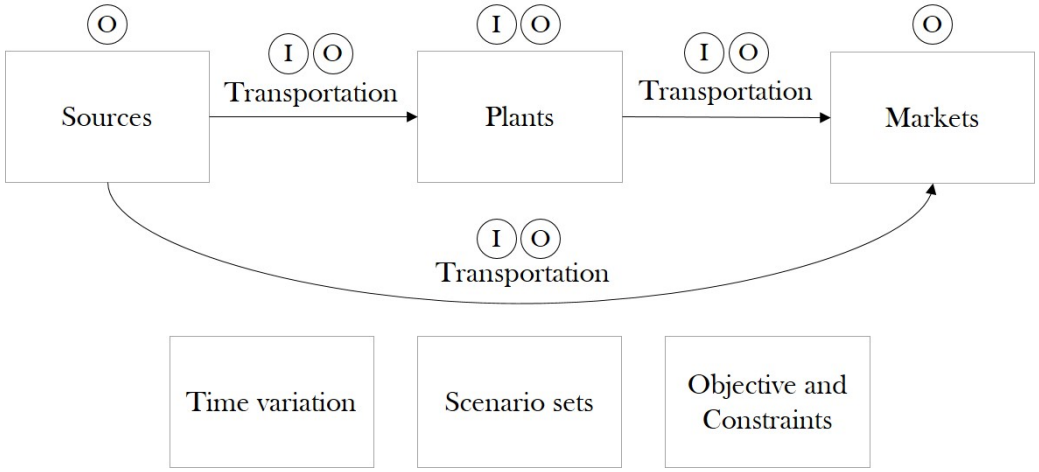


Figure 4-1: General framework of the supply chain in study. The flows of commodities are represented with arrows. Resources from shale oil and gas sources can either be sold directly to the markets or undergo a conversion or upgrading process at different plants, where the resulting products are then sold to the markets. Variation of parameter values in different time periods and the uncertainty of future predictions are also incorporated into the framework. Constraints and an objective function are imposed with the formulation of an optimization problem. The circles containing 'I' and 'O' refer to possible investment decisions and operating decisions respectively, made at either the nodes or arcs of the supply chain.

4.1 Problem description and methodology

4.1.1 General framework

Figure 4-1 illustrates the resulting general framework that was employed in this study. A decision maker seeks to make new oil and gas investments to best capitalize on the current and future projected abundance of shale oil and gas resources in the United States, in order to meet future demand in excess of that currently satisfied by existing infrastructure.

The supply chain begins with the sources of oil and gas available to monetize, namely, the shale plays currently active in the country and projected to be increas-

ingly so in the future. These shale plays produce a combination of oil, natural gas, and natural gas liquids (NGLs), which we term collectively as resources. Here, the decision maker makes the operating decision of how much of each resource would be purchased from the producers each year. Further, he or she then has to decide whether to sell the resources to the respective markets directly, or to convert or upgrade the resources into higher-value products by transporting them to plants of various technologies.

If the conversion/upgrading route is chosen, the decision maker then has to decide what types of technologies to invest in, the appropriate capacities and counts, and locations at which to build the plants. For each plant, the decision maker also has to choose the optimal operating levels every year. Finally, for each product, the decision maker has to select the most profitable markets at which the products should be sold.

The transportation of resources from source to plants or to markets, and the transportation of products from plants to markets adds an additional layer of consideration for the decision maker. Various modes of transportation are available for different products. For each route, the decision maker has to make the operating decision of how much commodity to ship through that route, if any, and which mode of transportation to take. In addition, the decision maker also has the ability to make investments in new pipelines to transport the commodities if necessary.

The sequence of decisions described are made with the complication of time-varying economic parameters and the presence of uncertainty in the parameter values. That is, the decision maker has to make a series of decisions in each time period, over a time horizon of twenty-five years. From one year to the next, the values of the supply, demand and price parameters of each commodity differ. In addition, at each particular time point, aside from the first time point where all parameters are known with certainty, each parameter can assume a different value based on the particular

scenario realized.

We assume a stochastic programming formulation of the problem by partitioning the investment and operating variables into two separate sets. The first set consists of the investment variables which are decided at the beginning of the time horizon, where all parameter values are known with certainty. That is, the decisions are made "here-and-now". In contrast, the second set consists of the operating variables which are decided individually for each scenario, once the parameter values have been realized under a particular scenario of consideration. That is, they are the "wait-and-see" decisions that are made to optimize profitability given the constraints imposed by the investment decisions that have been made [70].

These decisions have to be made satisfying various constraints imposed on the framework. The relevant constraints include operating within the limits of supply and demand of each commodity at the sources and markets, respectively, respecting capacity limits of plants and pipelines, respecting the resource-to-product conversion efficiencies of the various plant technologies adopted, and respecting the conservation of flow of materials through the network structure imposed between sources, plants and markets. Under these constraints, the decision maker then seeks to maximize the expected net present value his or her investments over given a time horizon, assuming an appropriate discount rate.

4.1.2 Time horizon

The time horizon under consideration is twenty-five years, with each year representing a time period in which decisions can be made. Projections through this twenty-five year time frame were obtained from the latest U.S. Energy Information Administration (EIA)'s Annual Energy Outlook Report (2015) [5]. These projections

were generated by the National Energy Modeling System (NEMS) a computer-based energy-economy modeling system [64]. NEMS utilizes data from public and private sources to generate plausible scenarios of prices, production, trade and consumption of energy which might play out through 2040, given different assumptions of macroeconomic factors, world energy markets, resource availability, technological choices, demographics and public policy. The twenty-five year time horizon in our study spanned 2015 to 2039.

4.1.3 Scenario sets

Scenario sets are collections of projections of future realizations of economic parameters (supply, demand and prices) which the decision-maker might feel are important to consider in the process of making investment and operating decisions.

We considered two different scenario sets, termed GDP and Oil Prices. The scenarios in each set have been constructed to align with the main cases from the Annual Energy Outlook 2015 (AEO 2015). Table B.1 from the Supplementary Material describes the assumptions made in generating each case.

Each scenario set contains three scenarios. In particular, the Reference scenario is included in both sets, which assumes a "business-as-usual" outcome, given known technological and demographic trends [5]. The other two scenarios in each scenario set represents the results from perturbing upwards or downwards the inputs to NEMS along a particular dimension (GDP growth or international oil supply and demand conditions) and they are termed as the High and Low scenarios respectively.

Table 4.1 shows the scenarios corresponding to each of the scenario set.

The supply chain optimization procedure is performed for each scenario set separately, potentially yielding different optimal investment decisions. The relative im-

Table 4.1: Scenarios corresponding to each scenario set (GDP and Oil Price).

GDP	Oil Price
• High Economic Growth	• High Oil Price
• Reference	• Reference
• Low Economic Growth	• Low Oil Price

portance of the consideration of either scenario set lies with the judgment of the decision-maker. Ideally, we would like to identify investment decisions which are common to both scenario sets, as this would suggest a good degree of robustness regardless of the specific probability distribution of future economic parameters that was assumed for the study.

4.1.4 Sources

The sources under consideration corresponded to the shale oil and gas plays analyzed by the EIA's Drilling Productivity Report [55, 71]. They are namely the Bakken, Eagle Ford, Haynesville, Marcellus, Niobrara, Permian and Utica Shale regions. When combined, these seven shale regions accounted for 95% of U.S. oil production growth and all of U.S. natural gas production growth during 2011 - 2013. Current trends from the report indicate that drilling activity in most of these regions continues to increase. As such, they provide a reasonable representation of the current and future availability of shale oil and gas resources in the country.

Figure 4-2 shows the locations of these seven shale regions assumed in our framework. We term each region as a source. Although each source was represented as a single point on the map, each is in reality a collection of shale oil and gas producing counties associated with that particular geographical area. This classification was in accordance with how the data was collected and analyzed in the Drilling Produc-

Figure 4-2: Locations of the seven sources in this study.

tivity Report. For reference, the list of counties associated with each source can be found in the report data file accompanying the report [55].

For each source, we were concerned with the production levels of three resources: oil, dry natural gas, and NGLs. We considered two of the sources, namely Haynesville and Marcellus as "dry" sources. That is, only dry natural gas can be produced from them. The other five "wet" sources can produce a combination of all three resources.

The Drilling Productivity Report classified the production of commodities at each source as that of oil or gas. Because data was collected at the well level, any separating processes downstream of the well meter were not accounted for. Therefore, in order to get an estimate of NGLs production at the wet sources, we referred to play level data supplied by the Oil and Gas Supply Module of NEMS [2]. In

particular, plays were assigned to their corresponding wet sources (as documented in Table B.3 in the Supplementary Material) and the respective sums of the estimated ultimate recovery (EUR) of natural gas plant liquids and that of dry natural gas were tabulated. Then, the ratio of the EUR of the natural gas plant liquids to that of dry natural gas was taken at each source, and this ratio was multiplied by the natural gas production data in the Drilling Productivity Report to obtain the production rates of NGLs at the sources. These production values attributed to NGLs production were then subtracted from the oil production values from the Drilling Productivity Report, and the result was taken to be the production rate of crude oil at the sources.

Table 4.2 summarizes the production levels of interest of each resource at each source at the end of January 2015, which corresponds to the beginning of the time horizon considered.

Table 4.2: Initial production rates of resources at the sources at the beginning of 2015.

Source	**Crude Oil (MMB/yr)**	**Dry Gas (BCF/yr)**	**NGLs (MMB/yr)**
Bakken	447.4	554.1	25.9
Eagle Ford	588.4	2,668.8	26.9
Haynesville	-	2,507.6	-
Marcellus	-	5,953.4	-
Niobrara	127.4	1,692.4	22.9
Permian	498.4	2,294.9	191.6
Utica	17.2	648.6	1.7

Supply and price projections of resources

We assumed that for all resources, supply matched demand at the beginning of the time horizon. The decision maker is concerned only with monetizing any supply in the future that is over and above 2015 levels. Additionally, if future projections

indicate that supply of a resource in later years falls below that of 2015 levels, the corresponding supply would be set to zero during that period of time. To construct the projected supplies for each source in each scenario, the sources were first mapped to their corresponding oil and gas producing region in the AEO 2015 projections. Then, the initial values in Table 4.2 were scaled with the projections for each year of the time horizon. Finally, the difference between each resulting value and its initial value was taken in order to arrive at the supply value available for the decision-maker to utilize.

The prices of the resources were mapped in a similar manner, where the price projections of each product were mapped to each source in its corresponding production region. The prices of NGLs were assume to vary in a fixed ratio of $6 per barrel of NGL to $1 per mcf of dry natural gas at each point in time for each scenario. Prices were also adjusted to bring them from 2013 (given format in AEO 2015) to 2015 levels using the U.S. Inflation Calculator [51]. The AEO projections used to construct each time series are indicated in Table B.11 in the Supplementary Material.

4.1.5 Plants

Plants serve to upgrade or convert the resources of oil and dry natural gas to higher-value products, which have varying end-use purposes. In our study, these products are liquefied natural gas (LNG), gasoline, kerosene, diesel and residual fuel oil (RFO). The decision maker has to assimilate competing factors of supply, demand, prices, transportation distances and economies of scale in order to select the optimal locations, technologies, scale and number of plants to construct in order to maximize his or her profitability. Each consideration is explained below in further detail.

Technologies

We consider three technologies - hydroskimming, gas-to-liquids conversion, and natural gas liquefaction. Correspondingly, the plants associated with these technologies are hydroskimming refineries, gas-to-liquids (GTL) plants, and liquefied natural gas (LNG) plants, respectively. These technologies have been chosen because they are relatively mature and have a history of past implementations. As shown in Figure 4-3, each technology accepts different inputs and yields different outputs, thus offering a variety of options to the decision maker.

Hydroskimming refineries Hydroskimming refineries are refineries with relatively low complexity. They typically include primary distillation equipment and downstream units and are suitable for processing oil of a lighter nature. We considered them in this study since there is currently a mismatch between existing U.S. refinery configurations, which were designed to process lower-quality, imported crude oil, and the abundance of light tight oil that shale oil plays produce [72]. In this study, we defined a hydroskimming refinery to consist of a desalter, an atmospheric distillation unit, a catalytic reformer and a catalytic hydrotreater.

The products from the hydroskimming refinery are gasoline, kerosene, diesel, and residual fuel oil (RFO). Because of the relative simplicity of the equipment, RFO represents the combination of all of the heavy fractions of the oil. The relative proportions of each refined product correlates strongly to the natural yield of the crude oil barrel. In our study, we set the barrel yield for each product, per barrel of crude oil fed as follows: gasoline - 0.25, kerosene - 0.10, diesel - 0.20 and RFO - 0.40. Note that there is an assumption of 0.05 losses per unit of input fed. These numbers were reasonable values made with the consideration of typical ranges quoted from

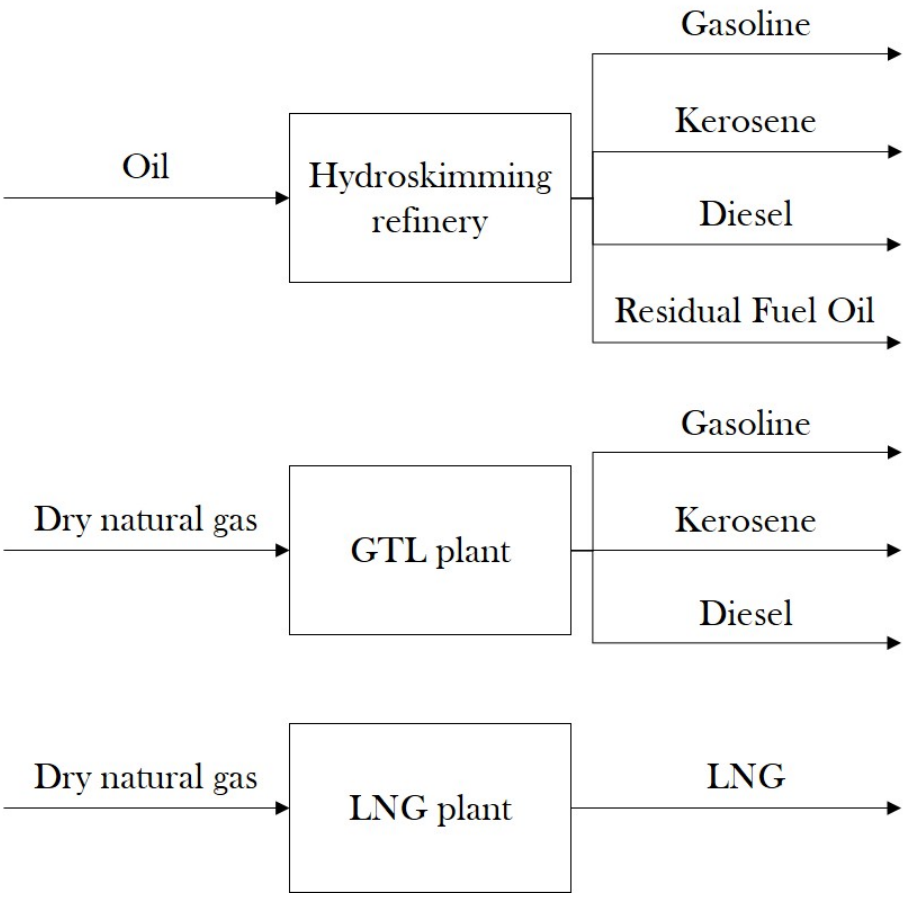


Figure 4-3: Plant technologies under consideration and their associated inputs and outputs.

industry sources [72, 73].

GTL plants The GTL process converts natural gas into liquid fuel, which, depending on the specifications of the process, can be adjusted to lie predominantly in the gasoline, kerosene, diesel or wax range, or a combination of them [38]. There are three main parts to this process: 1) syngas generation, 2) Fischer-Tropsch (FT) synthesis, and 3) refining and upgrading.

In syngas generation, natural gas is first cleaned and then converted into syngas, which is a mixture of hydrogen and carbon monoxide. This is typically done through one of or a mixture of three main technologies: steam methane reforming (SMR), partial oxidation (POX) and autothermal reforming (ATR). After the syngas has been generated, it proceeds to the next step, FT synthesis, where it is converted into longer chain hydrocarbons. Depending on the temperature, the process can be adjusted to optimize production of molecules within a certain chain length range. Finally, after the FT synthesis step, the product is sent for refining and upgrading to meet its final specifications.

In our study, we defined the output of the GTL plants to match the current demand ratios of refined products in the U.S. [5], which, for every thousand cubic feet of dry gas fed into the plant, yields 0.0589, 0.0109 and 0.0302 barrels of gasoline, kerosene and diesel, respectively.

LNG plants The value of LNG technology lies in its ability to increase the energy density of natural gas significantly, thus allowing it to be transportable for sale in distant markets. In the U.S., the most promising market for LNG lies in the transportation sector involving heavy-duty trucks or freight rail, where it can be used as a fuel for long distance travel [65, 39]. In overseas markets, the price of

natural gas is currently higher than that in the U.S., primarily due to the fact that gas in overseas markets has traditionally been pegged closely to the price of oil. As such, a substantial profit margin could be captured if LNG could be shipped to overseas markets, and a flurry of activity has taken place in order to expedite the process of exporting LNG from U.S. shores [74].

Essentially, the LNG process involves the liquefaction of gas by cooling it to an average temperature of approximately -162 °C, primarily by mixed-refrigerant cycles at the larger scales. The finished product is then stored in cryogenic tanks from which it is loaded into trucks or tankers for delivery to markets.

We assumed an efficiency rate of 88% for the LNG process, which translates to 0.0172 tons of LNG produced per thousand cubic feet of dry gas fed to the plant.

Candidate locations

An approximately evenly spaced grid of 35 locations throughout the United States that provided good coverage of the country served as the collection of candidate location sites for the construction of plants. Figure 4-4 shows the layout of these candidate locations.

In brief, the candidate locations were selected through a scoring procedure that took into account different land cover types and land ownership and favored land that was less developed, had lighter vegetation or was barren, and owned by agencies which were deemed to be more likely to accept development projects. Readers should refer to the Supplementary Material where the procedure is explained in detail.

Figure 4-4: Plant technologies under consideration and their associated inputs and outputs.

Capacities, capital costs and operating costs

For each technology, the decision maker can construct plants of different capacities. Traditionally, when building a plant for a single source and a single market, the decision maker will seek to maximize the cost savings that arise from economies of scale and build at the largest scale possible, within the limits of supply and demand, while locating the plant as close as possible to the supply or demand center, depending on the relative transportation costs of the resources and products. However, such a choice is not obvious when one has access to multiple sources, markets and plant locations within a large geographical area. For example, having distributed, smaller plants, might be optimal if transportation costs overpower economies of scale when a significant number of distant sources or markets have to be served. Another example of when operating at a smaller scale might be beneficial would be when there is a need to mitigate some risk of the loss of profitability arising from operating under capacity due to changing future economic conditions.

In our study, we considered the effect of economies of scale in the capital costs of the plants. For each technology, we selected three different plant capacities that the decision maker would be able to choose from in his or her investments, hereby termed "small", "medium" and "large". The range of plant capacities has been chosen to correspond to the typical range found in previous implementations of similar plants in real life or in studies.

The available plant capacities are as follows. The hydroskimming refinery was available at capacities of 25, 50 and 100 million barrels per year of oil feed. The GTL plant and LNG plant were available at capacities of 150, 300, 600 billion cubic feet per year of dry natural gas feed. Converting to commonly-used units in industry, this translated to approximate capacities of 68, 137 and 274 thousand barrels per

day of oil for the refinery, 41, 82 and 164 thousand barrels per day of GTL product for the GTL plant and 2.6, 5.1 and 10.3 million tonnes per annum of LNG product for the LNG plant, respectively.

Because the supply of resources varies from year to year, there might be a mismatch in the amount of actual resource being fed into the plant versus its design capacity. In our model, we allow for a feasible operation of each plant down to a lower bound of half its design capacity for each year. Note that this does not necessarily imply that the plant is required to operate at half its capacity throughout the year, where the feasibility of such a mode of operation would depend on the turndown characteristics of the plant equipment. Rather, it could imply that the plant is operated, for example, at full capacity for half the year, and shut down its operations for the other half, thereby achieving the same effect. Because our study was not concerned with operating decisions with finer granularity than a year, we consider both situations and variations thereof as identical.

The capital costs of the plants were determined from capital cost curves obtained from a combination of industry sources and previous studies [75, 76]. The details of the cost curves and methodology can be found in the Supplementary Material. The resulting capital costs are, in order of size: hydroskimming refineries of $559.5, $921.7 and $1,533.6 million, GTL plants of $2,921.7, $5,184.2 and $9,198.6 million, and LNG plants of $885.0, $1,676.1 and $3,174.5 million. The effects of economies of scale on capital costs can clearly be recognized, since the capital costs are less than double that of the previous size.

Operating costs comprise labor, utility, maintenance and administrative expenses. For this study, we set the operating costs of hydroskimming refineries at $8.50 per barrel, which was a reasonable value set in consultation with an EIA's performance report of major energy producers in refining and marketing in 2009 [77]. We further

assume operating costs of $1.5 per mcf gas fed (translating to $15 per barrel of GTL product) for the GTL plant [38, 78, 37] and $0.2 per mcf gas fed for the LNG plant [43, 79], which are within the ranges reported in various studies in the literature.

Investment decisions

The investment decisions lie in choosing among the various plant locations, technologies, capacities and quantities for an optimal outcome. This planning decision has to be made at the beginning of the time horizon, when uncertainty in economic parameters is present. In addition, although plans have to be made "here-and-now", the points in time to initiate construction can be further decided upon. This is to accommodate possible changes in demand or supply conditions further in the time horizon which necessitate the expansion of plant capacities.

Formally, the decision maker initiates his investment decisions in the year 2015. He has two time points at which he could initiate construction of a plant: at the beginning of the time horizon at 2015, and ten years later at 2025. Once the construction is initiated, each plant has to undergo a construction period before it is fully operational. We have set the construction period to be three years. Once the plant has been constructed, it is fully operational for the rest of the time horizon.

One issue to consider is that our study is limited to a finite time horizon due to the finiteness of the time series of the data. This leads to edge effects which have to be corrected for. In particular, if we assume that each plant can remain fully operational for twenty-five years from the time construction is initiated, then, plants which are constructed in 2025 would still have life remaining beyond the time horizon. To correct for this, we appropriately prorate the capital costs of these plants in accordance to the length of the remaining time horizon from the year of

construction. This implies that the capital costs of plants constructed in 2025 would be scaled by (15 years/25 years) = 0.6. In addition, the time horizon over which the plant is in operation is larger than that over which it would be depreciated. As such, we assume no salvage value can be recovered from the plants at the end of the time horizon.

Summary of plant characteristics

Table 4.3 summarizes the plant characteristics discussed in the previous sections.

4.1.6 Markets

We considered two sets of markets which the decision maker can access - local and foreign markets.

We first describe our representation of the local markets. The local markets comprise the Lower 48 states (i.e., contiguous U.S.). These states, along with Washington D.C., were each considered as a distinct market, yielding 49 local markets in total. The geographical location of each market was taken to be the geographical center of the corresponding state [80]. This level of granularity in the representation was able to describe adequately the geographical variation of flows of commodities and their demand pertinent to the objectives of our study, while at the same time preserving the computational tractability of our endeavor.

We next describe our representation of the foreign markets. The foreign markets were aggregated into three main entities which, based on projections, are likely to face an increasing need for natural gas imports in the future [81]. The three foreign entities were namely Mexico, OECD Asia, and OECD Europe[1]. The aggregation was

[1]We followed the EIA's grouping of the OECD countries. OECD Asia comprises Japan and South

Table 4.3: Summary of plant characteristics. Corresponding units are indicated in brackets. Abbreviations for products: G (gasoline), K (kerosene), D (diesel), R (residual fuel oil), L (liquefied natural gas).

Parameter	**Value**		
Locations	35		
Construction time	3 years		
Technologies	Refinery	GTL	LNG
Capacities (in feed terms)	[MMB/yr] sm: 25 md: 50 lg: 100	[BCF/yr] sm: 150 md: 300 lg: 600	[BCF/yr] sm: 150 md: 300 lg: 600
Capacities (commonly-used units)	[MB/day] sm: 68 md: 137 lg: 274	[MB/day] sm: 41 md: 82 lg: 164	[mtpa] sm: 2.6 md: 5.1 lg: 10.3
Capacity lower limit	50%	50%	50%
Capital costs	[$B] sm: 0.56 md: 0.92 lg: 1.53	[$B] sm: 2.92 md: 5.18 lg: 9.20	[$B] sm: 0.88 md: 1.68 lg: 3.17
Operating costs	[$/bbl] 8.5	[$/mcf] 1.5	[$/mcf] 0.2
Feed	Oil	Dry natural gas	Dry natural gas
Output	[bbl/bbl feed] G: 0.25 K: 0.10 D: 0.20 R: 0.40	[bbl/mcf feed] G: 0.059 K: 0.011 D: 0.030	[tonnes/mcf feed] L: 0.017

performed to accommodate the lower degree of granularity of data available for these markets, while still ensuring a broad and adequate representation of the potential foreign opportunities that the U.S.-based decision maker might have access to.

The nature of the commodities available to be sold directly to these markets vary slightly according to geographical and end-use considerations. Among the mix of resources (oil, dry natural gas, and NGLs) and products (LNG, gasoline, kerosene, diesel, and RFO) available, only a subset of these commodities would be available to each type of market. Table 4.4 describes the type of commodity available to be sold in each market.

Table 4.4: Types of commodities accessible to respective markets.

Lower 48 States	**Mexico**	**OECD Europe and OECD Asia**
• Dry natural gas • NGLs • LNG • Gasoline • Kerosene • Diesel • RFO	• Dry natural gas • NGLs • Gasoline • Kerosene • Diesel • RFO	• NGLs • LNG • Gasoline • Kerosene • Diesel • RFO

We stress that we only considered any additional demand above the initial year when planning for new investments, as we have made the assumption that demand levels up to that of the initial year are satisfied by existing infrastructure in the U.S.

Korea. OECD Europe comprises Austria, Belgium, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Luxembourg, Netherlands, Norway, Poland, Portugal, Slovakia, Spain, Sweden, Switzerland, Turkey, and the United Kingdom.

Demand of commodities

Lower 48 states The commodities that were demanded by the Lower 48 states were: dry natural gas, NGLs, LNG, gasoline, kerosene, diesel, and RFO. Crude oil was excluded from the demand mix as our model accounted for the presence of refineries to convert the crude oil into refined products that would be demanded by end-use customers.

To construct the demand series, consumption data for each commodity in each state was first obtained from the EIA's State Energy Data System (SEDS) [82], corresponding to the year 2013 which was the final year in which data was available. Using these numbers, the proportion of consumption allocated to each state within each census division was determined. These proportions were then multiplied by the corresponding initial value taken from the Reference scenario in the year 2015 in the AEO 2015 projections to arrive at the initial demand levels for each state. These initial demand levels were then evolved with the AEO 2015 projections to obtain the appropriately scaled projections. Finally, with the scaled projections, only the excess demand above initial year levels was taken to arrive at the final demand time series.

Foreign markets For foreign markets, we exclude the direct sale of crude oil in our consideration, as currently there is a legacy ban on oil exports from the U.S., with the exception of a handful of restricted situations [72].

With recent developments transforming the oil supply landscape, there has been interest in understanding the various implications if this ban is potentially lifted [72, 83, 84, 85]. Nevertheless, because we would like this study to be informative for actionable implementations based in the present, we did not consider the possibility

of lifting the oil ban. Future work might incorporate various case studies where the limitation of oil exports is relaxed, depending on the likelihood of future legislation changing in this regard.

In addition, we exclude LNG as a product for sale in Mexico, as the LNG export projections provided by AEO 2015 are based only on exports to Asia and Europe. Any natural gas sold to Mexico would be in the form of dry natural gas shipped by pipeline. In contrast, we exclude dry natural gas from the slate of salable commodities to OECD Asia and OECD Europe because these regions are only accessible across the ocean and therefore any transport of natural gas would have to be in the form of LNG.

The demand time series were determined for each product/region as follows: Dry natural gas exports to Mexico corresponded to pipeline exports to Mexico projections in AEO 2015. LNG exports were taken from the AEO 2015 projections and appropriately scaled to the respective foreign regions using data from the International Energy Outlook 2013 Reference Case projections of world natural gas consumption by region [81]. Initial values for exports of NGLs and refined products were obtained from EIA's Exports by Destination annual data for the year 2014 [86]. These initial values were then evolved with the the Petroleum and Other Liquids Consumption projections in AEO 2015 to obtain the appropriately scaled projections. Finally, only the excess demand above initial year levels was taken to arrive at the final demand time series.

The details of specific calculations for the foreign demand series, as well as a table of the data series used for projections of both the local and foreign demand of commodities can be found in the Supplementary Material.

Prices of commodities

In a similar vein, the initial prices of commodities in the Lower 48 states were taken from the SEDS price data, reported for the year 2013, which was the latest year in which data was available. These prices were then adjusted to 2015 values by adjusting for the trend in prices from 2013 to 2015 as indicated in the Reference case of the AEO 2015 projections, corresponding to the appropriate census division for each state. These initial prices were then evolved with the corresponding price series in the AEO 2015 projections to arrive at the final price time series for each state through the entire time horizon. Prices were also adjusted to bring them from 2013 dollar terms (given format in AEO 2015) to 2015 dollar terms using the U.S. Inflation Calculator [51].

The time series of prices of commodities in the foreign markets required a more involved construction process. We first discuss the handling of Mexico price data. Initial dry natural gas prices in Mexico were obtained from EIA's U.S. Natural Gas Exports and Re-Exports by Point of Exit data as an average of the pipeline gas prices in the first three months of 2015 [87]. These initial prices were then scaled according to the evolution of AEO 2015 border prices of pipeline gas imports from Mexico. The assumption we made was that both import and export prices reflect the prevailing economic conditions in Mexico and would be expected to move in similar fashion.

The evolution of oil prices in Mexico is taken to be the average between the projected Brent and WTI prices from the AEO 2015 data, adjusted for inflation from 2013 dollar terms to 2015 dollar terms, using the U.S. Inflation Calculator [51]. Our assumption is made in light of the lack of better data, and due to Mexico's accessibility to both the U.S. and to the North Sea. This time series, termed as

"Mexico Blend", also served as the trend to determine the evolution of prices for NGLs and the refined products. Adjustments were also made to convert the initial spot price to an import price. This was done by multiplication with a proxy factor of the ratio of the gasoline import price in Mexico to the spot price of price in the U.S. in 4Q2014, obtained from the International Energy Agency (IEA)'s Quarterly Statistics on Energy Prices and Taxes [88].

Initial prices for the refined products were also obtained from the same publication. These prices were reported for 4Q2014, ex-tax (i.e., after excise tax and VAT have been deducted from the final selling prices), which is a good representation of the actual receipts that would accrue to a U.S.-based decision maker. As the price of NGLs was not available from the publication, a price ratio of 1.20 gasoline price/NGLs price was assumed to obtain the corresponding initial NGLs price [89]. Initial prices were also brought forward to 1Q2015 values by multiplication with the average of the ratios of 1Q2015 North Sea and WTI crude oil to their 4Q2014 values. Finally, the initial prices for each commodity were then evolved according to the price series of the "Mexico Blend" to obtain the corresponding final time series.

We next discuss the handling of OECD Asia and OECD Europe price data. Initial LNG prices were taken from the Federal Energy Regulatory Commission World Landed LNG Prices for the beginning of the year 2015 [90]. These prices were then evolved with the Brent prices from the AEO 2015 data, as we assumed that overseas LNG prices are predominantly indexed to the price of oil. Oil prices were assumed to be that of Brent prices from the AEO 2015 data, adjusted for inflation from 2013 dollar terms to 2015 dollar terms, using the U.S. Inflation Calculator [51]. The initial spot oil price was also converted to an import price. This was done by multiplication with a factor of the ratio of North sea crude import costs in IEA Europe to its spot price in 4Q2014 [88].

Initial prices for all refined products and NGLs in OECD Asia were determined by taking an average of ex-tax prices in Japan and Korea of the corresponding commodities from the IEA publication [88]. The initial prices were also brought forward to 1Q2015 values by multiplication with the ratio of 1Q2015 North Sea crude to its 4Q2014 value. Initial prices for gasoline, diesel and kerosene in OECD Europe were set according to average in-Europe EIA country ex-tax prices from 5 January to 16 March 2015 [91, 92]. Residual fuel oil was assumed to be selling at 70.6% of diesel price [93, 94], while NGLs were assumed to be selling at 51.9% of gasoline price [95]. Finally, the initial prices of the products were made to evolve with Brent prices from the AEO 2015 data to obtain the final price series.

Details on the initial prices set at foreign markets for the various commodities, as well as the data used for each price series can be found in the Supplementary Material.

4.1.7 Transportation

To transport the commodities, the decision maker has access to four main modes of transportation: pipeline, road, rail and barge. The possible flows of transportation lie in three categories: source-to-plant, source-to-market, and plant-to-market. We use the term "node" to indicate either a source, plant or market. Depending on the type of commodity carried, some of modes of transportation might or might not be available. Table 4.5 documents the commodities carried by each mode of transportation in our study. Details of the peculiarities of each mode of transportation, methods of routing from node to node, and costs are made specific in the following sections.

Table 4.5: Types of commodities carried by each mode of transportation.

Pipeline	Road	Rail	Barge
• Oil	• Oil	• Oil	• Oil
• Dry natural gas	• NGLs	• NGLs	• NGLs
• NGLs	• LNG	• LNG	• LNG
• Refined Products	• Refined Products	• Refined Products	• Refined Products

Pipeline

The pipeline is the most common mode used for the transportation of oil and gas commodities in the United States today [96]. In our study, pipelines were made available for the transport of oil, dry natural gas, NGLs and refined products. Note that for purposes of transportation, individual refined products (i.e. gasoline, kerosene, diesel and residual fuel oil) were grouped into a single category, since it is usual practice to transport different refined products in the same pipeline, by sequencing the flow of each batch of product through the pipeline [97]. Naturally, LNG was excluded as a possibility for pipeline transport, since the main appeal of liquefaction technology is to act as a substitute for transporting natural gas via pipelines over long distances [98].

Consistent with our assumption that current supply matches current demand for the initial year, we excluded the consideration of utilizing existing pipelines in the U.S. to service the transportation of commodities arising from the operational decisions in our study. This was a fair assumption to make, since many of the existing pipelines currently operate in excess of 95% utilization [99]. Therefore, we required that new pipelines be built in order to enable any future transportation of commodities by pipeline in our study.

Routing from one node to another was performed with ESRI ArcMap 10.2.2 [100]

by determining the straight-line distances between each node. The result was a collection of potential pipeline paths, one for each particular node-to-node connection, upon which pipelines could be built.

For each set of pipelines corresponding to a particular commodity and node-to-node flow, we provided three different capacities for the decision maker to utilize. Typically, the capacities for pipelines for source-to-plant flows were set to be larger than that for source-to-market or plant-to-market flows. This was mainly due to the fact that commodity volumes supplied by the sources were typically larger than volumes demanded at the markets. As a result, nodes located upstream in the supply chain typically deliver commodities to multiple nodes downstream. With this consideration, the pipelines were sized to match the volumes at the destinations rather than at the origins.

Table 4.6 shows the pipeline capacities considered for new investments in this study. Note that for notational convenience, we classified the three different pipeline sizes as "small", "medium" and "large", although the actual capacity would vary depending on the specific type of commodity and flow that the pipeline is associated with.

A lower annual operational limit of 50% of the stated capacity of the pipelines was imposed. Similar to our discussion in the section of plant capacities, we note that this lower bound does not necessarily imply that pipelines can practically be operated at 50% capacity at all times. Rather, it might suggest the case where the pipelines are operated at full capacity for half a year and shut down for the other half, or other reasonable variations thereof. With the lower bound of capacity set in this manner, the decision maker has an *ex ante* continuous range of choice of transportation levels, since the capacity for each pipeline size is set to double that of the previous size.

Table 4.6: Pipeline capacities available for investments. Abbreviations for flows: s-p (source-to-plant), s-m (source-to-market), p-m (plant-to-market).

Commodity	**Flow**	**Capacities**			**units**
		sm	md	lg	
Oil	s-p	25	50	100	MMB/yr
	s-m	25	50	100	
Dry natural gas	s-p	150	300	600	BCF/yr
	s-m	37.5	75	150	
NGLs	s-m	5	10	20	MMB/yr
Refined products	p-m	5	10	20	MMB/yr

Note: Lower operating bounds of pipelines are set to 50% of stated capacity values. Construction time is 3 years.

Similar to plant investments, pipeline investments also exhibit cost savings due to economies of scale. The two main determinants of capital costs for pipeline investments are its capacity and length. To determine the pipeline cost curves, data was collected based on implemented projects of dry natural gas and oil pipelines, and regressions were performed with the following relationship: $\log_e(\text{Capital Costs/Length})_i = \beta_0 + \beta_1 \log_e(\text{Capacity})_i + \epsilon_i$. The cost curve of the oil pipelines also served as a proxy for the costing of NGL and refined products pipelines. Finally, the capital costs of pipelines estimated from the fitted regression were scaled by a curvature multiplier of 1.2 to account for our straight-line assumption of pipeline lengths. Details of the data used for the regressions and resulting coefficient estimates can be found in the Supplementary Material.

Transportation costs for pipelines generally comprise power consumption costs for the operation of pumps and compressors as well as general monitoring and maintenance of the system. We estimated operating costs for dry natural gas pipelines at \$0.78 per mcf per thousand miles [101], for oil and refined products pipelines at

$1.06 per barrel per thousand miles [102, 103] and for NGLs pipelines at $0.75 per barrel per thousand miles [102, 104].

Similar to plant investment decisions, the investment decisions for new pipelines have to be made at the beginning of the time horizon, when the uncertainties in economic parameters are still present. The decision maker can choose to initiate the construction process at the beginning of the time horizon in 2015 and/or ten years into the time horizon in 2025, if there is a need to capture any additional increase in commodity flows arising further down in the time horizon.

Once the construction process has been initiated, the pipelines undergo construction for a period of three years before they become operational. Similar to the case with plant capital costs, we appropriately scale the capital costs of pipelines constructed in 2025 by a factor of 0.6 (i.e., 15 years/25 years) to deal with edge effects arising from the finite nature of the time horizon. In addition, we assume no salvage value at the end of the time horizon.

Road

We considered the transportation of liquid fuels by road through the use of trucks. The commodities thus available to transportation by road are oil, NGLs, LNG and refined products. Node-to-node road distances were computed using the Google Distance Matrix API [61]. Transportation costs were set at $7.30 per barrel per thousand miles for oil and refined products [105, 103], $5.14 per barrel per thousand miles for NGLs [105, 104], and $8.07 per mcf per thousand miles for LNG [39, 106]. Contrary to pipeline transport, we assumed that current roads can support the movement of commodities that arise in our study.

Rail

Rail transport was an alternative method for the transportation of liquid fuels, which were oil, NGLs, LNG and refined products in this study. Node-to-node distances were computed using the Network Analyst tool of ESRI ArcMap 10.2.2 [100] over a rail network dataset. The rail network dataset used was the U.S. National Transportation Atlas Railroads network [107], which was pre-processed to remove any short, isolated lines or loops before routing. Transportation costs were set at $3.45 per barrel per thousand miles for oil and refined products [105, 103], and $2.43 per barrel per thousand miles for NGLs [105, 104]. Transportation costs of $3.82 per mcf per thousand miles for LNG by rail were assumed by appropriate scaling from its road transportation costs. Similar to road transport, we assumed that the existing railroads network can support the movement of commodities that arise in our study.

Barge

Transport by barges on U.S. waterways was the final alternative for transportation of liquid fuels, which were, once again, oil, NGLs, LNG and refined products. Because most nodes were not directly situated beside a petroleum port along a waterway, we defined barge transport between a pair of nodes as a composition of three legs: node-to-port, port-to-port, and port-to-node. The node-to-port and port-to-node legs was carried out either by road or rail transportation, while the port-to-port leg was carried out by barge transportation.

To illustrate this idea, consider the transportation route from a generic node A to a generic node B. For node A, we identified the port which required the lowest transportation cost to it, either by road or by rail, and defined this selected port as port 0. This yielded the node-to-port leg. Similarly, for node B, we identified the

port with the lowest transportation cost to it, and defined this selected port as port 1. This yielded the port-to-node leg. Finally, the port 0 to port 1 leg yielded the port-to-port leg. The cost of the transportation route from node A to B was then determined by the sum of the transportation costs of each leg.

The petroleum ports were obtained from an EIA-filtered list of ports used in the U.S. Energy Mapping System [108], originally taken from data from the U.S Army Corps of Engineers Navigation Data Center [109]. Port-to-port routes and corresponding distances were obtained from the Sea Route & Distance calculator from Ports.com [110], and several adjustments were made to deal with specific ports whenever data could not be obtained from the calculator. The adjustments are documented in the Supplementary Material. Port-to-node and node-to-port legs were routed using the Google Distance Matrix API [61] and ESRI ArcMap 10.2.2 [100] as described in the sections on road and rail transport respectively.

Transportation costs for barge transport were set at $1.32 per barrel per thousand miles for oil and refined products [105, 103], and $0.93 per barrel per thousand miles for NGLs [105, 104]. Transportation costs of $1.46 per mcf per thousand miles for LNG by barge were assumed by appropriate scaling from its road transportation costs. We considered barge transport to be uncapacitated for the purposes of this study.

Table 4.7 summarizes the transportation costs associated with each transportation mode applied to this study.

Adjustments for transport to foreign markets

Characterizing the modes of transportation of commodities to foreign markets, namely Mexico, OECD Asia and OECD Europe, required additional adjustments which we

Table 4.7: Transport costs for transportation modes in the study.

Commodity \ Mode	Pipeline	Road	Rail	Barge
Oil[1]	1.06	7.30	3.45	1.32
Dry natural gas[2]	0.78	N.A.	N.A.	N.A.
NGLs[1]	0.75	5.14	2.43	0.93
LNG[2]	N.A.	8.07	3.82	1.46
Refined products[1]	1.06	7.30	3.45	1.32

1: Units of $ per barrel per thousand miles.
2: Units of $ per mcf per thousand miles.

now describe.

Mexico Road transportation from nodes to Mexico was determined by the Google Distance Matrix API [61], where the destination was set to Mexico City. Pipeline and rail transportation were described by a two-leg setting: the first leg involved determining the routes within the U.S. to one of seven points on the U.S.-Mexico border, whose coordinates can be found in Table B.15 of the Supplementary Material. The routes within this leg were characterized using the techniques described in the previous sections on pipeline and rail transportation. The second leg mapped the route from the border points to Mexico City. Due to the lack of finer information, we approximated the distance of this leg as simply the straight-line distance from the border points to Mexico City, again indicated in Table B.15. Both legs were then combined to obtain the final routes for pipeline and rail transport.

Barge transport was also characterized in a two-leg fashion. The first leg comprised the transport from domestic nodes to domestic petroleum ports, which were filtered to only include ports that processed products associated with foreign movement (i.e., with export and/or import activity). For each node, the port that required

the lowest transportation cost by road or rail was chosen. The second leg then comprised the barge route from this port to one of two ports - Dos Bocas and Salina Cruz, which were representative petroleum ports found on either coast of Mexico [111]. The shorter of the two distances was set to be the second leg. Finally, both legs were then combined to obtain the final routes for barge transport.

OECD Asia and OECD Europe Transportation to OECD Asia and OECD Europe required a two-leg process, the first being within the United States (i.e., source/plant-to-coast) and the second across the oceans (i.e., U.S. coast-to-foreign coast). For the first leg, we allowed three modes of transportation - pipeline, road and rail - from sources/plants to U.S. ports. Intra-U.S. barge transportation was not considered since for most cases it would make better sense to ship commodities directly to the foreign destination once a port could be accessed in the route. Only ports associated with foreign movement of petroleum products were considered. The same routing methods previously discussed were performed to characterize the first leg for each mode of transport.

The second leg involved transport across the oceans via tankers to their final destinations. A coarse mapping of origins and destinations was implemented, where four possible U.S. coasts served as the origin to a final foreign destination either in Asia or Europe. Representative ports for each coast are described in Table B.16 in the Supplementary Material. Distances between each U.S. coast and foreign destination were mapped using the Ports Distances calculator from Sea-Distances.org [112].

For the second leg, we imposed tanker costs for oil, NGLs and refined products of \$0.28 per bbl per thousand nautical miles, and for LNG, \$0.22 per mcf per thousand nautical miles. These values were arrived at by averaging several representative shipping costs found in several shipping-related technical reports [113, 114]. Finally,

both legs were composed together by matching the selected ports in the first leg to their corresponding U.S. coasts in the second leg, to obtain the final, complete paths from sources or plants to the overseas markets.

4.2 Model formulation

This section describes in detail the formulation of the optimization model which was implemented in the study. Readers only interested in the results of the study may proceed to the next section with no problems.

4.2.1 Index sets

- Sources $I = \{$Bakken, Eagleford, Haynesville, Marcellus, Niobrara, Permian, Utica$\}$
- Plant technologies $J = \{$hydroskimming refinery, GTL plant, LNG plant$\}$
- Plant locations $K = \{$35 selected points in the U.S.$\}$
- Plant sizes $N = \{$small, medium, large$\}$
- Local markets $L_0 = \{$geographic centers of U.S. Lower 48 + D.C.$\}$
- Land-connected foreign markets $L_1 = \{$Mexico$\}$
- Overseas foreign markets $L_2 = \{$OECD Asia, OECD Europe$\}$
- Resources $F = \{$crude oil, dry gas, NGLs$\}$
- Products $P = \{$LNG, gasoline, kerosene, diesel, residual fuel oil$\}$
- Commodity classification in transportation
  $V = \{$class crude oil, class dry gas, class NGLs, class LNG, class refined products$\}$

- Modes of transportation $A = \{\text{pipe, truck, rail, barge}\}$
- Pipeline sizes $O = \{\text{small, medium, large}\}$
- Operational time periods $T = \{2016, 2017, \ldots, 2038, 2039\}$
- Scenarios $S = \{\text{High, Reference, Low}\}$
- Construction time points $C = \{2015, 2025\}$

4.2.2 Sets constructed from index sets

Tags

- Resources associated with transportation classification tags
  $\mathcal{F} = \{(\text{crude oil, class crude oil}), (\text{dry gas, class dry gas}), (\text{NGLs, class NGLs})\}$
- Products associated with transportation classification tags
  $\mathcal{P} = \{(\text{LNG, class LNG}), (\text{gasoline, class refined products}), (\text{kerosene, class refined products}), (\text{diesel, class refined products}), (\text{residual fuel oil, class refined products})\}$

Associations between Commodities and Markets

- Markets served by each commodity $\mathcal{L}(v) = \begin{cases} \varnothing, & \text{if } v = \text{class crude oil} \\ L_0 \cup L_1, & \text{if } v = \text{class dry gas} \\ L_0 \cup L_2, & \text{if } v = \text{class LNG} \\ L_0 \cup L_1 \cup L_2, & \text{otherwise} \end{cases}$

Routes

- All routes from sources to plants by pipeline $\mathcal{R}^{\text{pipe}}_{I\to K} = \{(i,k,j,(f,v)) \in I \times K \times J \times \mathcal{F} \setminus \{(\text{NGLs, class NGLs})\}\}$
- All routes from sources to plants by road $\mathcal{R}^{\text{road}}_{I\to K} = \{(i,k,j,(f,v)) \in I \times K \times J \times \mathcal{F} \setminus \{(\text{dry gas, class dry gas}), (\text{NGLs, class NGLs})\}\}$
- All routes from sources to plants by rail $\mathcal{R}^{\text{rail}}_{I\to K} = \{(i,k,j,(f,v)) \in I \times K \times J \times \mathcal{F} \setminus \{(\text{dry gas, class dry gas}), (\text{NGLs, class NGLs})\}\}$
- All routes from sources to plants by barge $\mathcal{R}^{\text{barge}}_{I\to K} = \{(i,k,j,(f,v)) \in I \times K \times J \times \mathcal{F} \setminus \{(\text{dry gas, class dry gas}), (\text{NGLs, class NGLs})\}\}$
- All routes from sources to markets by pipeline $\mathcal{R}^{\text{pipe}}_{I\to L} = \{(i,l,(f,v)) \in I \times \mathcal{L}(v) \times \mathcal{F}\}$
- All routes from sources to markets by road $\mathcal{R}^{\text{road}}_{I\to L} = \{(i,l,(f,v)) \in I \times \mathcal{L}(v) \times \mathcal{F} \setminus \{(\text{dry gas, class dry gas})\}\}$
- All routes from sources to markets by rail $\mathcal{R}^{\text{rail}}_{I\to L} = \{(i,l,(f,v)) \in I \times \mathcal{L}(v) \times \mathcal{F} \setminus \{(\text{dry gas, class dry gas})\}\}$
- All routes from sources to markets by barge $\mathcal{R}^{\text{barge}}_{I\to L} = \{(i,l,(f,v)) \in I \times \mathcal{L}(v) \setminus L_2 \times \mathcal{F} \setminus \{(\text{dry gas, class dry gas})\}\}$
- All routes from plants to markets by pipeline $\mathcal{R}^{\text{pipe}}_{K\to L} = \{(k,l,(p,v)) \in K \times \mathcal{L}(v) \times \mathcal{P}\}$
- All routes from plants to markets by road $\mathcal{R}^{\text{road}}_{K\to L} = \{(k,l,(p,v)) \in K \times \mathcal{L}(v) \times \mathcal{P}\}$

- All routes from plants to markets by rail $\mathcal{R}^{\text{rail}}_{K\to L} = \{(k, l, (p, v)) \in K \times \mathcal{L}(v) \times \mathcal{P}\}$
- All routes from plants to markets by barge $\mathcal{R}^{\text{barge}}_{K\to L} = \{(k, l, (p, v)) \in K \times \mathcal{L}(v) \setminus L_2 \times \mathcal{P}\}$

Infrastructure

- All plants $\mathcal{K} = \{(k, j, n, c) \in K \times J \times N \times C\}$
- All pipelines from sources to plants $\mathcal{Z}_{I\to K} = \{(i, k, v, o, c) \in I \times K \times V \setminus \{\text{class NGLs}\} \times O \times C\}$
- All pipelines from sources to markets $\mathcal{Z}_{I\to L} = \{(i, l, v, o, c) \in I \times \mathcal{L}(v) \times V \times O \times C\}$
- All pipelines from plants to markets $\mathcal{Z}_{K\to L} = \{(k, l, v, o, c) \in K \times \mathcal{L}(v) \times V \times O \times C\}$

4.2.3 Decision variables

Investment decisions

- $\text{Plant}_{kjnc} \in \mathbb{Z}_+, \forall (k, j, n, c) \in \mathcal{K}$: Number of plants at location k of technology j with size n constructed at time point c.
- $\text{Pipe}^{I\to K}_{ikvoc} \in \mathbb{Z}_+, \forall (i, k, v, o, c) \in \mathcal{Z}_{I\to K}$: Number of pipelines connecting source i to plant location k carrying commodity type v with size o constructed at time point c.

- $\text{Pipe}^{I\to L}_{ilvoc} \in \mathbb{Z}_+, \forall (i,l,v,o,c) \in \mathcal{Z}_{I\to L}$: Number of pipelines connecting source i to market l carrying commodity type v with size o constructed at time point c.

- $\text{Pipe}^{K\to L}_{klvoc} \in \mathbb{Z}_+, \forall (k,l,v,o,c) \in \mathcal{Z}_{K\to L}$: Number of pipelines connecting plant k to market l carrying commodity type v with size o constructed at time point c.

Operating decisions

- $\text{Flow}^{I\to K,\text{pipe}}_{ikj(f,v)ts} \in \mathbb{R}_+, \forall ((i,k,j,(f,v)),t,s) \in \mathcal{R}^{\text{pipe}}_{I\to K} \times T \times S$: Flow of material by pipeline from source i to plant at location k and of technology j, carrying resource (f,v) at time point t in scenario s.

- $\text{Flow}^{I\to K,\text{road}}_{ikj(f,v)ts} \in \mathbb{R}_+, \forall ((i,k,j,(f,v)),t,s) \in \mathcal{R}^{\text{road}}_{I\to K} \times T \times S$: Flow of material by road from source i to plant at location k and of technology j, carrying resource (f,v) at time point t in scenario s.

- $\text{Flow}^{I\to K,\text{rail}}_{ikj(f,v)ts} \in \mathbb{R}_+, \forall ((i,k,j,(f,v)),t,s) \in \mathcal{R}^{\text{rail}}_{I\to K} \times T \times S$: Flow of material by rail from source i to plant at location k and of technology j, carrying resource (f,v) at time point t in scenario s.

- $\text{Flow}^{I\to K,\text{barge}}_{ikj(f,v)ts} \in \mathbb{R}_+, \forall ((i,k,j,(f,v)),t,s) \in \mathcal{R}^{\text{barge}}_{I\to K} \times T \times S$: Flow of material by barge from source i to plant at location k and of technology j, carrying resource (f,v) at time point t in scenario s.

- $\text{Flow}^{I\to L,\text{pipe}}_{il(f,v)ts} \in \mathbb{R}_+, \forall ((i,l,(f,v)),t,s) \in \mathcal{R}^{\text{pipe}}_{I\to L} \times T \times S$: Flow of material by pipeline from source i to market l, carrying resource (f,v) at time point t in scenario s.

- $\text{Flow}^{I\to L,\text{road}}_{il(f,v)ts} \in \mathbb{R}_+, \forall((i,l,(f,v)),t,s) \in \mathcal{R}^{\text{road}}_{I\to L} \times T \times S$: Flow of material by road from source i to market l, carrying resource (f,v) at time point t in scenario s.

- $\text{Flow}^{I\to L,\text{rail}}_{il(f,v)ts} \in \mathbb{R}_+, \forall((i,l,(f,v)),t,s) \in \mathcal{R}^{\text{rail}}_{I\to L} \times T \times S$: Flow of material by rail from source i to market l, carrying resource (f,v) at time point t in scenario s.

- $\text{Flow}^{I\to L,\text{barge}}_{il(f,v)ts} \in \mathbb{R}_+, \forall((i,l,(f,v)),t,s) \in \mathcal{R}^{\text{barge}}_{I\to L} \times T \times S$: Flow of material by barge from source i to market l, carrying resource (f,v) at time point t in scenario s.

- $\text{Flow}^{K\to L,\text{pipe}}_{kl(p,v)ts} \in \mathbb{R}_+, \forall((k,l,(p,v)),t,s) \in \mathcal{R}^{\text{pipe}}_{K\to L} \times T \times S$: Flow of material by pipeline from plant k to market l, carrying product (p,v) at time point t in scenario s.

- $\text{Flow}^{K\to L,\text{road}}_{kl(p,v)ts} \in \mathbb{R}_+, \forall((k,l,(p,v)),t,s) \in \mathcal{R}^{\text{road}}_{K\to L} \times T \times S$: Flow of material by road from plant k to market l, carrying product (p,v) at time point t in scenario s.

- $\text{Flow}^{K\to L,\text{rail}}_{kl(p,v)ts} \in \mathbb{R}_+, \forall((k,l,(p,v)),t,s) \in \mathcal{R}^{\text{rail}}_{K\to L} \times T \times S$: Flow of material by rail from plant k to market l, carrying product (p,v) at time point t in scenario s.

- $\text{Flow}^{K\to L,\text{barge}}_{kl(p,v)ts} \in \mathbb{R}_+, \forall((k,l,(p,v)),t,s) \in \mathcal{R}^{\text{barge}}_{K\to L} \times T \times S$: Flow of material by barge from plant k to market l, carrying product (p,v) at time point t in scenario s.

4.2.4 Constraints

The sum of flows of each resource leaving each source cannot exceed its corresponding supply. Denoting supply_{ifts} as the supply at source i for resource f at time t in

scenario s:

$$\sum_{a\in A}\left[\sum_{\{(k,j,v)|\exists(i,k,j,(f,v))\in\mathcal{R}^a_{I\to K}\}}\text{Flow}^{I\to K,a}_{ikj(f,v)ts}+\sum_{\{(l,v)|\exists(i,l,(f,v))\in\mathcal{R}^a_{I\to L}\}}\text{Flow}^{I\to L,a}_{il(f,v)ts}\right]\leq\text{supply}_{ifts},$$
$$\forall(i,f,t,s)\in I\times F\times T\times S. \quad (4.1)$$

The sum of flows through each pipeline route must be within the capacity bounds available for that route. At time t, let the maximum and minimum available pipeline capacity for a pipe of type v and size o constructed at time point c be $\text{cap\_pipe\_max}^{*\to *}_{votc}$ and $\text{cap\_pipe\_min}^{*\to *}_{votc}$, respectively.

Then, we introduce the encoding:

$$\text{cap\_pipe\_(max/min)}^{*\to *}_{votc}=\begin{cases}\text{(max/min) pipeline capacity}, & \text{if } t-c\geq\mathcal{T}_c\\ 0, & \text{otherwise,}\end{cases}$$

where $\mathcal{T}_c$ is the construction time for capital investments.

With the given encoding, we write the pipeline flow capacity constraints in an aggregated form as such: Source-to-plant:

$$\sum_{\{(j,f)|\exists(i,k,j,(f,v))\in\mathcal{R}^{\text{pipe}}_{I\to K}\}}\text{Flow}^{I\to K,\text{pipe}}_{ikj(f,v)ts}\leq\sum_{\{(o,c)|\exists(i,k,v,o,c)\in\mathcal{Z}_{I\to K}\}}\text{cap\_pipe\_max}^{I\to K}_{votc}\text{Pipe}^{I\to K}_{ikvoc},$$
$$\forall(i,k,v,t,s)\in I\times K\times V\setminus\{\text{class NGLs}\}\times T\times S. \quad (4.2a)$$

$$\sum_{\{(j,f)|\exists(i,k,j,(f,v))\in\mathcal{R}^{\text{pipe}}_{I\to K}\}} \text{Flow}^{I\to K,\text{pipe}}_{ikj(f,v)ts} \geq \sum_{\{(o,c)|\exists(i,k,v,o,c)\in\mathcal{Z}_{I\to K}\}} \text{cap\_pipe\_min}^{I\to K}_{votc}\text{Pipe}^{I\to K}_{ikvoc},$$

$$\forall(i,k,v,t,s) \in I \times K \times V \setminus \{\text{class NGLs}\} \times T \times S. \quad (4.2b)$$

Source-to-market:

$$\sum_{\{f|\exists(i,l,(f,v))\in\mathcal{R}^{\text{pipe}}_{I\to L}\}} \text{Flow}^{I\to L,\text{pipe}}_{il(f,v)ts} \leq \sum_{\{(o,c)|\exists(i,l,v,o,c)\in\mathcal{Z}_{I\to L}\}} \text{cap\_pipe\_max}^{I\to L}_{votc}\text{Pipe}^{I\to L}_{ilvoc},$$

$$\forall(i,l,v,t,s) \in I \times \mathcal{L}(v) \times V \times T \times S. \quad (4.3a)$$

$$\sum_{\{f|\exists(i,l,(f,v))\in\mathcal{R}^{\text{pipe}}_{I\to L}\}} \text{Flow}^{I\to L,\text{pipe}}_{il(f,v)ts} \geq \sum_{\{(o,c)|\exists(i,l,v,o,c)\in\mathcal{Z}_{I\to L}\}} \text{cap\_pipe\_min}^{I\to L}_{votc}\text{Pipe}^{I\to L}_{ilvoc},$$

$$\forall(i,l,v,t,s) \in I \times \mathcal{L}(v) \times V \times T \times S. \quad (4.3b)$$

Plant-to-market:

$$\sum_{\{p|\exists(k,l,(p,v))\in\mathcal{R}^{\text{pipe}}_{K\to L}\}} \text{Flow}^{K\to L,\text{pipe}}_{kl(p,v)ts} \leq \sum_{\{(o,c)|\exists(k,l,v,o,c)\in\mathcal{Z}_{K\to L}\}} \text{cap\_pipe\_max}^{K\to L}_{votc}\text{Pipe}^{K\to L}_{klvoc},$$

$$\forall(k,l,v,t,s) \in K \times \mathcal{L}(v) \times V \times T \times S. \quad (4.4a)$$

$$\sum_{\{p|\exists(k,l,(p,v))\in\mathcal{R}^{\text{pipe}}_{K\to L}\}} \text{Flow}^{K\to L,\text{pipe}}_{kl(p,v)ts} \geq \sum_{\{(o,c)|\exists(k,l,v,o,c)\in\mathcal{Z}_{K\to L}\}} \text{cap\_pipe\_min}^{K\to L}_{votc}\text{Pipe}^{K\to L}_{klvoc},$$

$$\forall(k,l,v,t,s) \in K \times \mathcal{L}(v) \times V \times T \times S. \quad (4.4b)$$

The sum of flows entering a plant must be within its capacity bounds. At

time t, let the maximum and minimum available capacity of a plant of type j and size n constructed at time point c for resource type f be $\text{cap\_plant\_max}_{jnftc}$ and $\text{cap\_plant\_min}_{jnftc}$, respectively.

In a similar manner, we introduce the encoding:

$$\text{cap\_plant\_(max/min)}_{jnftc} = \begin{cases} \text{(max/min) plant capacity}, & \text{if } t - c \geq \mathcal{T}_c \\ 0, & \text{otherwise,} \end{cases}$$

where $\mathcal{T}_c$ is the construction time for capital investments.

The plant capacity constraints are then written as such:

$$\sum_{a \in A} \sum_{\{(i,v)|\exists (i,k,j,(f,v)) \in \mathcal{R}^a_{I \to K}\}} \text{Flow}^{I \to K,a}_{ikj(f,v)ts} \leq \sum_{\{(n,c)|\exists (k,j,n,c) \in \mathcal{K}\}} \text{cap\_plant\_max}_{jnftc} \text{Plant}_{kjnc},$$
$$\forall (k,j,f,t,s) \in K \times J \times F \times T \times S. \quad (4.5a)$$

$$\sum_{a \in A} \sum_{\{(i,v)|\exists (i,k,j,(f,v)) \in \mathcal{R}^a_{I \to K}\}} \text{Flow}^{I \to K,a}_{ikj(f,v)ts} \geq \sum_{\{(n,c)|\exists (k,j,n,c) \in \mathcal{K}\}} \text{cap\_plant\_min}_{jnftc} \text{Plant}_{kjnc},$$
$$\forall (k,j,f,t,s) \in K \times J \times F \times T \times S. \quad (4.5b)$$

At the plants, the resources are then converted into products. We denote convert_{fpj} as the conversion ratio to product p from one unit of input resource f, using plant

technology j. The following constraints then describe the plant conversion processes:

$$\sum_{a \in A} \sum_{\{(i,j,(f,v))|\exists (i,k,j,(f,v)) \in \mathcal{R}^a_{I \to K}\}} \text{convert}_{fpj} \text{Flow}^{I \to K, a}_{ikj(f,v)ts} = \sum_{a \in A} \sum_{\{(l,v)|\exists (k,l,(p,v)) \in \mathcal{R}^a_{K \to L}\}} \text{Flow}^{K \to L, a}_{kl(p,v)ts}, \quad \forall (k,p,t,s) \in K \times P \times T \times S. \quad (4.6)$$

At the markets, the sum of flows for each resource or product entering a market cannot exceed its corresponding demand. Denoting demand_res_{lfts} and $\text{demand\_prod}_{lpts}$ as the demand for resource f or product p respectively at market l at time t and scenario s:

$$\sum_{a \in A} \sum_{\{(i,v)|\exists (i,l,(f,v)) \in \mathcal{R}^a_{I \to L}\}} \text{Flow}^{I \to L, a}_{il(f,v)ts} \leq \text{demand\_res}_{lfts}, \quad \forall (l,f,t,s) \in L \times F \times T \times S. \quad (4.7a)$$

$$\sum_{a \in A} \sum_{\{(k,v)|\exists (k,l,(p,v)) \in \mathcal{R}^a_{K \to L}\}} \text{Flow}^{K \to L, a}_{kl(p,v)ts} \leq \text{demand\_prod}_{lpts}, \quad \forall (l,p,t,s) \in L \times P \times T \times S. \quad (4.7b)$$

We now identify the components that comprise the objective function.

Revenues at time t for scenario s, denoted by Revenues_{ts}, are the sum of sales made at the markets. The prices of resources f and products p at market l at time

t for scenario s are denoted by $\text{price\_res}_{l,f,t,s}$ and $\text{price\_prod}_{l,p,t,s}$, respectively:

$$\begin{aligned}&\text{Revenues}_{ts}\\&= \sum_{a \in A} \left[\sum_{(i,l,(f,v)) \in \mathcal{R}^a_{I \to L}} \text{price\_res}_{l,f,t,s} \text{Flow}^{I \to L,a}_{il(f,v)ts} + \sum_{(k,l,(p,v)) \in \mathcal{R}^a_{K \to L}} \text{price\_prod}_{l,p,t,s} \text{Flow}^{K \to L,a}_{kl(p,v)ts} \right], \\ &\qquad \forall (t,s) \in T \times S. \quad (4.8)\end{aligned}$$

Recurring costs at time t for scenario s, denoted by Costs_{ts}, consist of resource purchase costs, plant operating costs and commodity transportation costs, denoted by Res_Costs_{ts} Op_Costs_{ts}, and Trans_Costs_{ts}, respectively:

$$\text{Costs}_{ts} = \text{Res\_Costs}_{ts} + \text{Op\_Costs}_{ts} + \text{Trans\_Costs}_{ts}, \quad \forall (t,s) \in T \times S. \quad (4.9)$$

Resource purchase costs are the sum of purchases of resources made at every source. The cost of resource f at source i at time t for scenario s is denoted by $\text{cost\_res}_{i,f,t,s}$:

$$\begin{aligned}&\text{Res\_Costs}_{ts}\\&= \sum_{a \in A} \left[\sum_{(i,k,j,(f,v)) \in \mathcal{R}^a_{I \to K}} \text{cost\_res}_{i,f,t,s} \text{Flow}^{I \to K,a}_{ikj(f,v)ts} + \sum_{(i,l,(f,v)) \in \mathcal{R}^a_{I \to L}} \text{cost\_res}_{i,f,t,s} \text{Flow}^{I \to L,a}_{il(f,v)ts} \right], \\ &\qquad \forall (t,s) \in T \times S. \quad (4.10)\end{aligned}$$

Plant operating costs are the sum of operating costs scaled by the flow of resources into every plant. The operating cost corresponding to resource f for plant technology

j is denoted by cost_op_{fj}:

$$\text{Op\_Costs}_{ts} = \sum_{a \in A} \sum_{(i,k,j,(f,v)) \in \mathcal{R}^a_{I \to K}} \text{cost\_op}_{fj} \text{Flow}^{I \to K, a}_{ikj(f,v)ts}, \quad \forall (t,s) \in T \times S. \quad (4.11)$$

Commodity transportation costs are the sum of transportation costs of each commodity by each mode of transportation through each possible origin-destination route. The cost for a particular route for origin d_0 to destination d_1 carrying commodity v by mode a is denoted by $\text{cost\_trans}^{* \to *, a}_{d_0 d_1 v}$

$$\begin{aligned} \text{Trans\_Costs}_{ts} = \sum_{a \in A} \Bigg[& \sum_{(i,k,j,(f,v)) \in \mathcal{R}^a_{I \to K}} \text{cost\_trans}^{I \to K, a}_{ikv} \text{Flow}^{I \to K, a}_{ikj(f,v)ts} \\ & + \sum_{(i,l,(f,v)) \in \mathcal{R}^a_{I \to L}} \text{cost\_trans}^{I \to L, a}_{ilv} \text{Flow}^{I \to L, a}_{il(f,v)ts} \\ & + \sum_{(k,l,(p,v)) \in \mathcal{R}^a_{K \to L}} \text{cost\_trans}^{K \to L, a}_{klv} \text{Flow}^{K \to L, a}_{kl(p,v)ts} \Bigg], \forall (t,s) \in T \times S. \quad (4.12) \end{aligned}$$

Investment costs are one-off costs that occur at construction time points c, denoted by Investments_c. They comprise plant investments and pipeline investments, denoted by Invest_Plants_c and Invest_Pipes_c respectively:

$$\text{Investments}_c = \text{Invest\_Plants}_c + \text{Invest\_Pipes}_c, \quad \forall c \in C. \quad (4.13)$$

Investments in plants are the sum of investments across all plants. We denote the capital cost of a plant of technology j of size n scaled by construction time c by

$\text{cap\_cost\_plant}_{jnc}$:

$$\text{Invest\_Plants}_c = \sum_{\{(k,j,n)|\exists(k,j,n,c)\in\mathcal{K}\}} \text{cap\_cost\_plant}_{jnc}\text{Plant}_{kjnc}, \quad \forall c \in C. \quad (4.14)$$

Similarly, investments in pipelines are the sum of investments across all pipelines. We denote the capital cost of a pipeline from origin d_0 to destination d_1, of type v and size o, scaled by construction time c by $\text{cap\_cost\_pipe}^{*\to*}_{d_0 d_1 voc}$:

$$\begin{aligned}\text{Invest\_Pipes}_c = &\sum_{\{(i,k,v,o)|\exists(i,k,v,o,c)\in\mathcal{Z}_{I\to K}\}} \text{cap\_cost\_pipe}^{I\to K}_{ikvoc}\text{Pipe}^{I\to K}_{ikvoc} \\ &+ \sum_{\{(i,l,v,o)|\exists(i,l,v,o,c)\in\mathcal{Z}_{I\to L}\}} \text{cap\_cost\_pipe}^{I\to L}_{ilvoc}\text{Pipe}^{I\to L}_{ilvoc} \\ &+ \sum_{\{(k,l,v,o)|\exists(k,l,v,o,c)\in\mathcal{Z}_{K\to L}\}} \text{cap\_cost\_pipe}^{K\to L}_{klvoc}\text{Pipe}^{K\to L}_{klvoc}, \quad \forall c \in C. \quad (4.15)\end{aligned}$$

Finally, the objective function seeks to maximize the expected net present value (ENPV) of the entire project. Denoting the discount factor at time t' as $\gamma_{t'}$ and the probability of scenario s as π_s, the objective function is:

$$\text{ENPV} = \sum_{c\in C} -\gamma_c \text{Investments}_c + \sum_{(t,s)\in T\times S} \pi_s \left[\gamma_t \left(\text{Revenues}_{ts} - \text{Costs}_{ts}\right)\right]. \quad (4.16)$$

4.3 Results and discussion

4.3.1 Computational results

All instances of the optimization models were solved on an Intel Xeon E5-1560 3.20 GHz machine with 12 GB of RAM using IBM ILOG CPLEX 12.6.1 in deterministic parallel mode, implemented with GAMS Python API 24.4.1. The stochastic

programs and nominal scenario problem instances were solved as Mixed Integer Programs (MIPs) with a relative tolerance of 1% and time limit of 10,000s. The largest full-sized instance of the stochastic program consisted of 494,431 equations and 2,723,803 variables, of which 18,900 were integer. All instances were solved to within the relative tolerance limit.

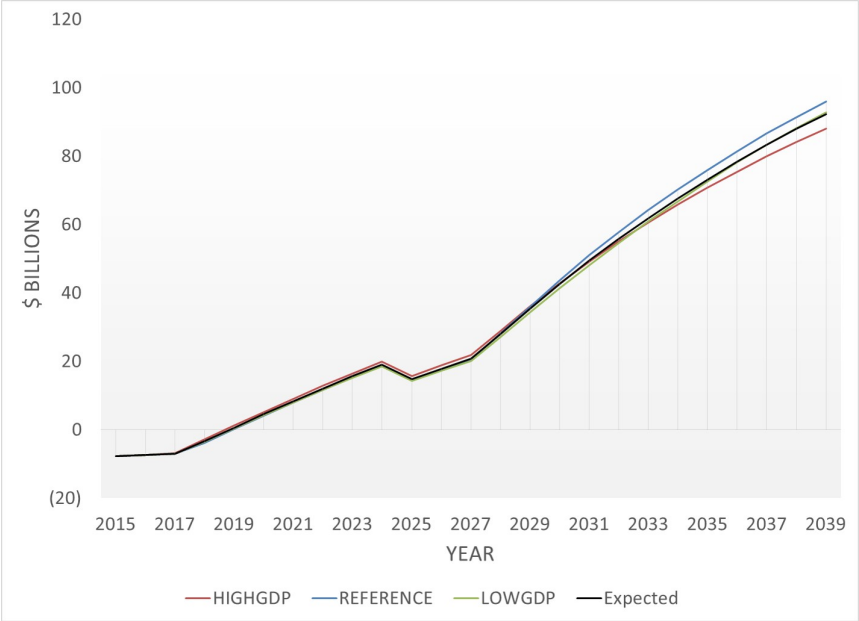
4.3.2 Profitability

Table 4.8 shows the optimal NPVs (in $B) over the full time horizon obtained from solving the stochastic program for each scenario set, while Figure 4-5 shows the cumulative NPVs over time. The optimal NPVs were in a much tighter range across scenarios and at higher levels for the GDP scenario set as compared to the Oil Price scenario set. The differences in levels of the NPVs of the Reference scenario in both scenario sets suggests that more conservative investments were made with the Oil Price scenario set assumptions. From the larger spread in NPV outcomes in the Oil Price scenario set, we infer that the ultimate profitability was much more sensitive to the differences in scenario realizations than in the GDP scenario set. The more conservative investment decision in the Oil Price scenario set is likely a reflection of how the stochastic program adjusts to accommodate this higher degree of sensitivity.

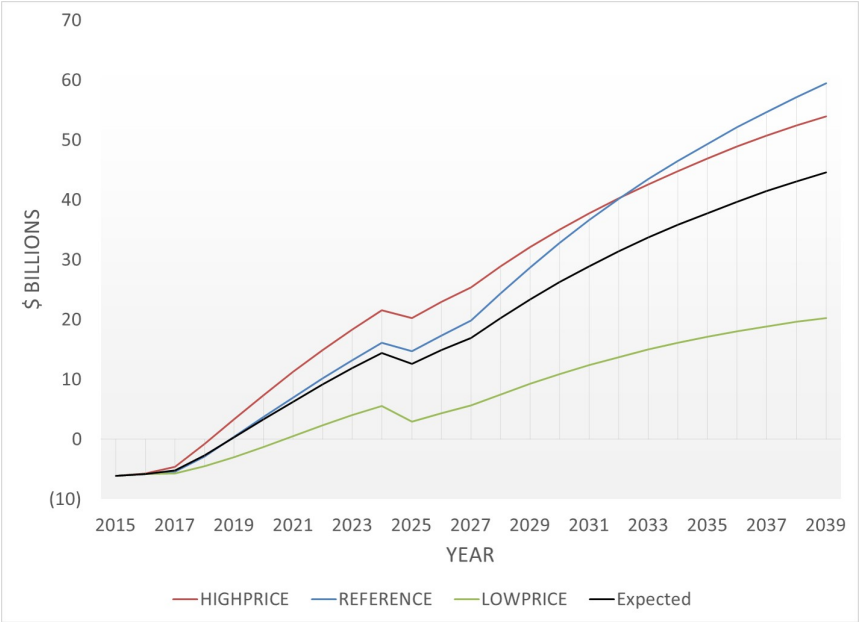
Table 4.8: Optimal NPVs (in $B) over entire time horizon for each scenario and expected value in each scenario set.

Scenario / **Scenario Set**	High	Reference	Low	Expected
GDP	88.1	96.0	92.8	92.3
Oil Price	54.0	59.4	20.2	44.5

Figure 4-6 shows the cumulative expected discounted cash flows for each scenario set. The cash flows were classified into investments, revenues, resource costs, oper-



(a) GDP Scenario Set.



(b) Oil Price Scenario Set.

Figure 4-5: Cumulative Expected Discounted Cash Flows for each scenario set.

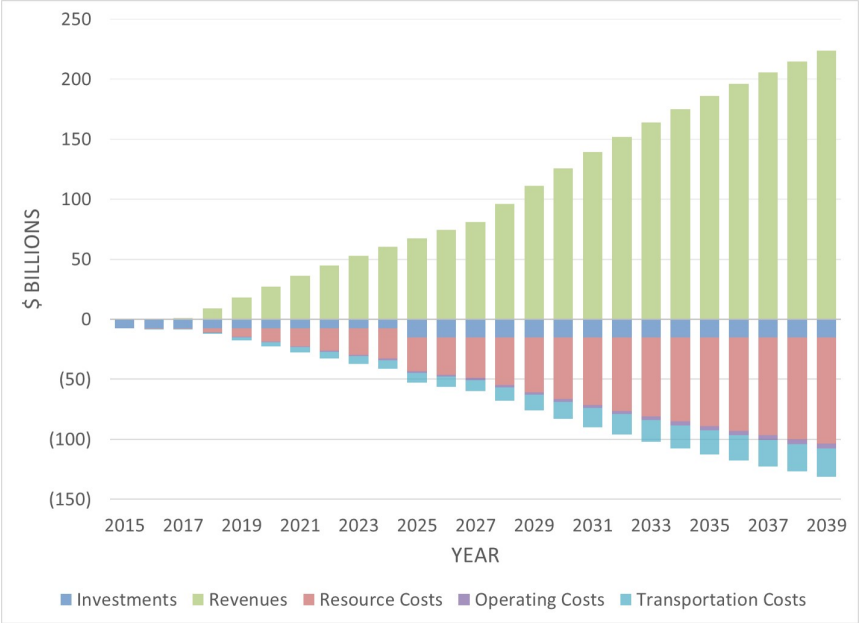
ating costs and transportation costs. Generally, resource costs make up the largest component of costs, followed by investments, transportation costs and operating costs. The respective proportions of each cost component are an indication to the sensitivity of profitability to the respective cost inputs.

4.3.3 Investments

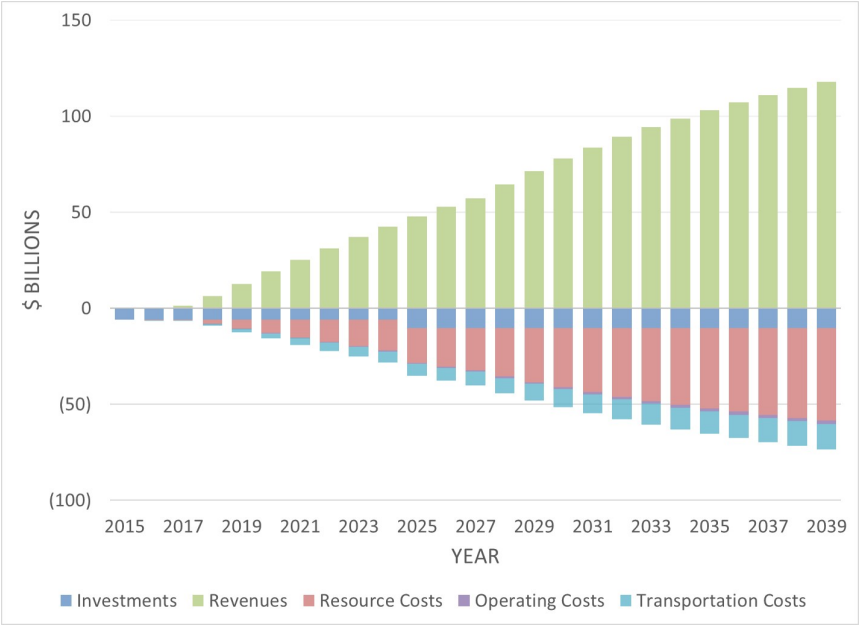
Figure 4-7 shows the optimal plant and pipeline investment types, capacities and construction timing for each scenario set. The optimal plant investments for the GDP scenario set involved all three technologies - GTL, LNG and hydroskimming oil refinery (OR), whereas only GTL and LNG technologies were selected for the Oil Price scenario set.

The most significant difference between the optimal plant investments for both scenario sets were the levels of LNG plant investments, particularly in the year 2025. Much larger LNG investments were made for the GDP scenario set, at a cumulative level of 4,050 BCF/yr (69.5 mtpa output), as compared to 1,350 BCF/yr (23.2 mtpa output) for the Oil Price scenario set. Relating to the discussion in the previous section on NPV, higher levels of investment in the GDP scenario set were likely made due to the lower variance in profitability with the different scenario realizations, as compared to that in the Oil Price scenario set, where a more hedged solution was preferred.

Both GTL and OR technologies were selected for the GDP scenario set, where a cumulative capacity of 150 BCF/yr (41 MB/day output) and 25 MMB/yr (68 MB/day output) respectively was installed. In contrast, only GTL was selected for the Oil Price set, where a cumulative capacity of 300 BCF/yr (82 MB/day output) was installed. In both scenarios, the GTL plants were invested in the year 2025,

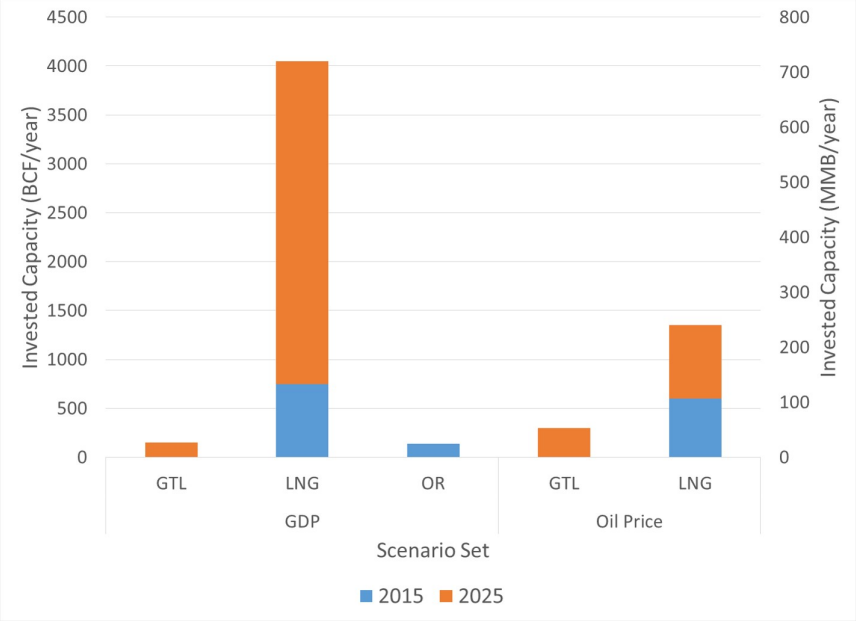


(a) GDP Scenario Set.

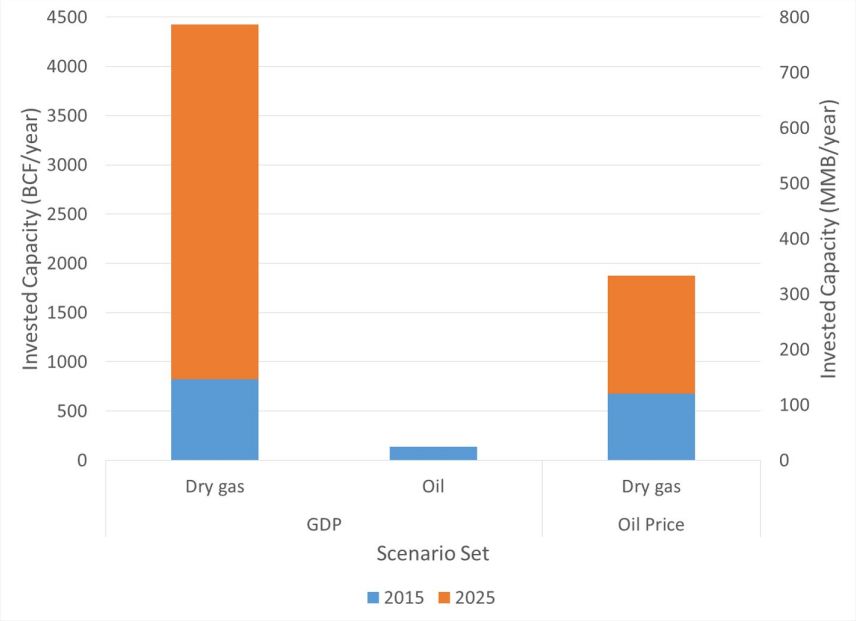


(b) Oil Price Scenario Set.

Figure 4-6: Cumulative Expected Discounted Cash Flows for each scenario set.



(a) Plant investments.



(b) Pipeline investments.

Figure 4-7: Optimal investments for each scenario set.

reflecting relatively lower opportunities for GTL technology in the nearer term. The combined capacity for the output of refined products were at similar levels for both scenario sets, indicating that there is less variation in terms of output of refined products as compared to the case for LNG. The levels of investments in these technologies on an energy-equivalent basis were also much lower than that of LNG.

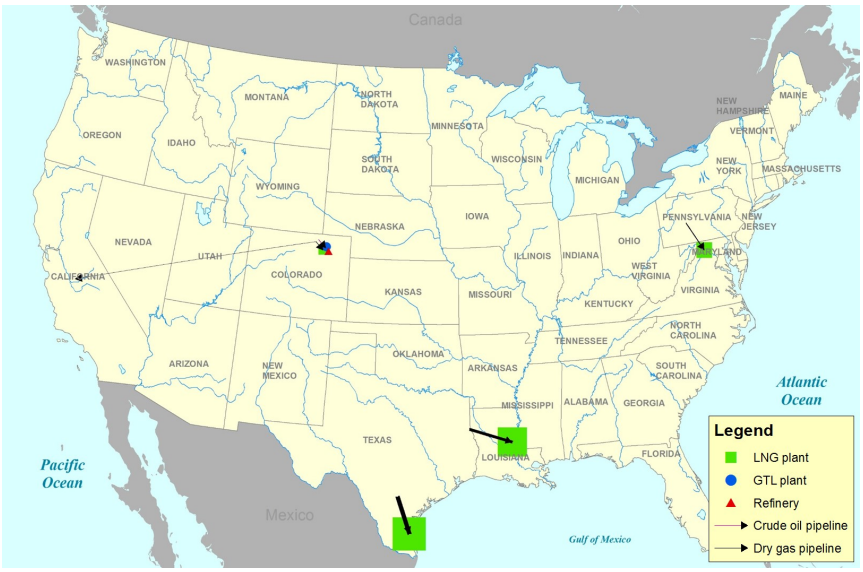
The differences in levels of pipeline investments between the two scenario sets largely reflect the differences in transportation infrastructure required to support the LNG plants. A cumulative total of 4,425 BCF/yr of dry gas pipelines and 25 MMB/yr of oil pipelines were installed for the GDP scenario set, whereas a much lower cumulative total of 1,875 BCF/yr of dry gas pipelines were installed for the Oil Price scenario set. No NGLs or refined products pipelines were installed in either scenario set.

Figures 4-8 and 4-9 depict the geographical locations of the cumulative plant and pipeline investments in 2015 and 2025, for the GDP and Oil Price scenario sets respectively.

The most striking feature of the maps is the small number of locations chosen in which to make these investments, and the similarity of location choices between the two scenario sets. In general, plant locations were chosen to be close to their sources, and only a small number of sources were deemed profitable to obtain resources from.

Common investment location decisions among both scenario sets include the building of LNG plants and associated dry gas pipelines near to the Eagle Ford, Haynesville and Marcellus sources in order to utilize the dry gas output in those areas. In addition, a dry gas pipeline was built from the Niobrara source to supply California with dry natural gas. The selected locations of the plants near the coasts also offered a transportation advantage, since a large proportion of the products were directed to overseas markets.

(a) Investments in 2015.



(b) Investments in 2025.

Figure 4-8: Plant and pipeline investments for GDP scenario set.

(a) Investments in 2015.

(b) Investments in 2025.

Figure 4-9: Plant and pipeline investments for Oil Price scenario set.

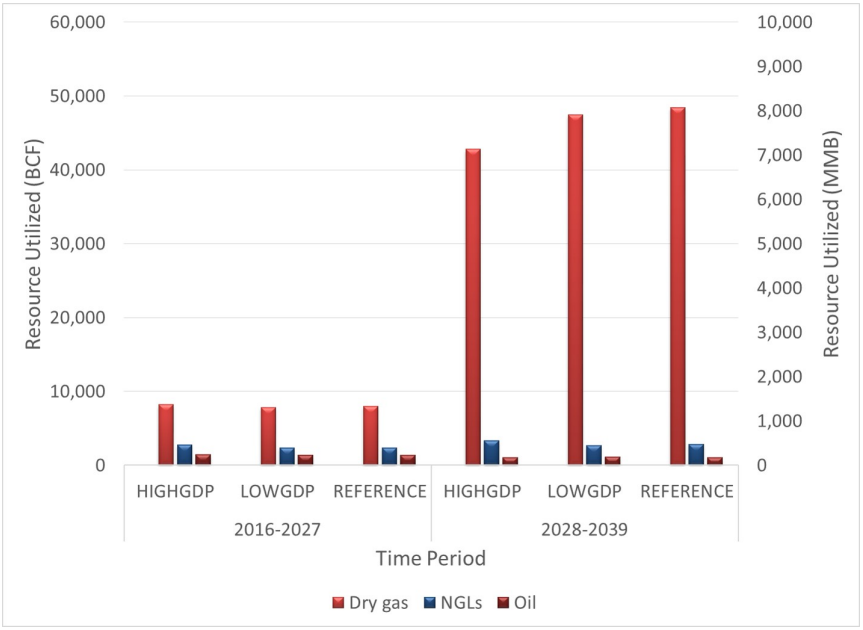
Beyond the common investment location decisions, there were several differences between the two scenario sets. In the GDP scenario set, an oil refinery in 2015, and LNG and GTL plants in 2025 of relatively small capacities and their associated pipelines were built near the Niobrara source. However, these investment decisions were not made in the Oil Price Scenario Set, indicating that the relative attractiveness of the Niobrara source is sensitive with respect to the uncertainty in future realizations of oil prices.

In order to monetize demand for refined products in the Oil Price case, the decision to build a GTL facility near the Eagle Ford source is made, in the absence of Niobrara being a monetizable source. This might also be due to the fact that more excess dry gas over that required to supply the LNG plants was available, since smaller LNG capacities were invested in this scenario set.

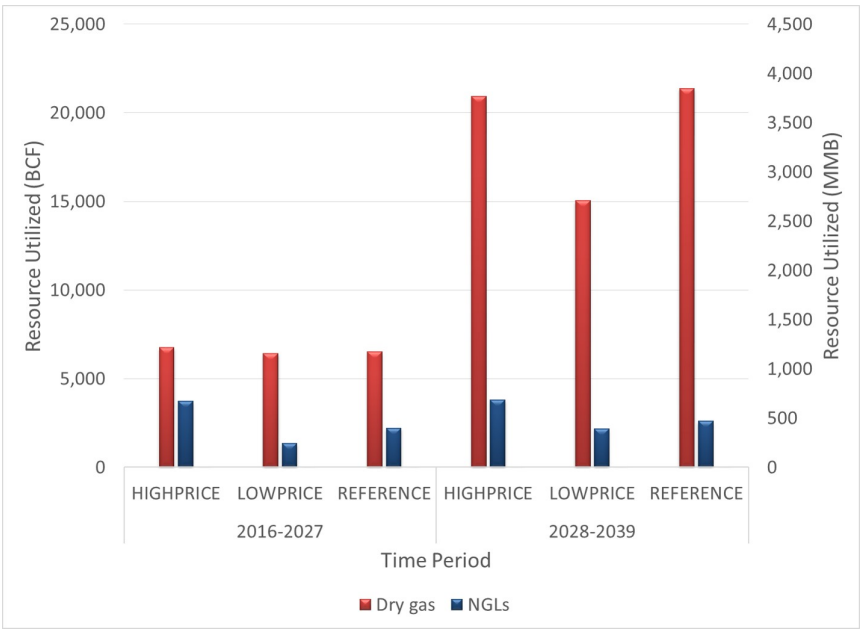
4.3.4 Resources utilized

Figure 4-10 shows the volumes of resources utilized at the sources in two time periods, 2016-2027 and 2028-2039, for each scenario in each scenario set. The resources utilized are either transported to plants for conversion to higher-value products, or transported directly to the markets and sold.

On an energy-equivalent basis, the utilization of dry gas greatly exceeds that of NGLs and crude oil in all scenarios for both scenario sets. In addition, the utilization of dry gas increases greatly in the second half of the time horizon whereas that for NGLs and oil remain relatively constant.



(a) GDP Scenario Set.



(b) Oil Price Scenario Set.

Figure 4-10: Resources utilized for each scenario set.

4.3.5 Commodities delivered

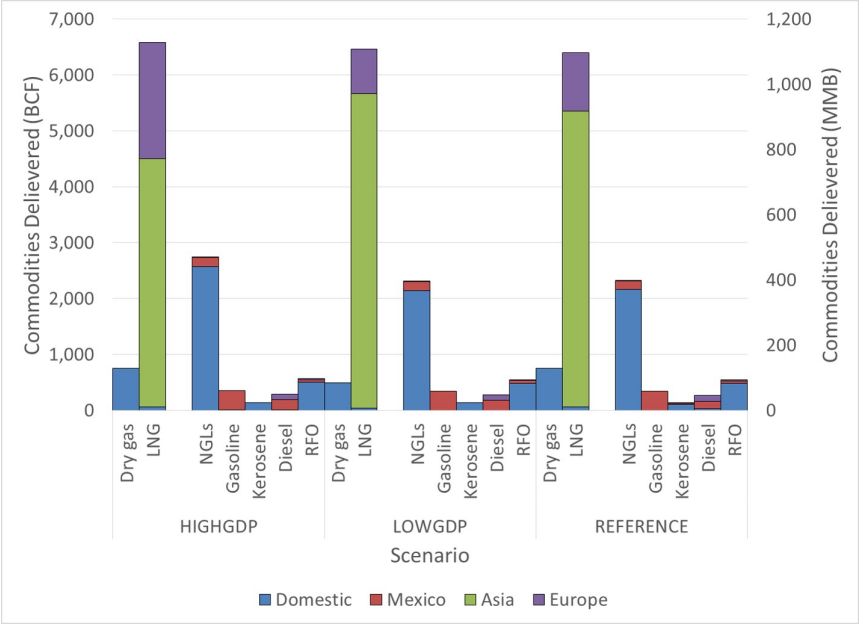
Figures 4-11 and 4-12 show the volumes of commodities delivered at the various markets for two time periods, 2016-2027 and 2028-2039, for each scenario and each scenario set.

In the GDP scenario set, we observe a relatively stable mix of markets for each commodity across all scenarios. In the earlier half of the time horizon, LNG was primarily delivered to overseas markets with a bias towards OECD Asia, whereas the bias changes towards OECD Europe in the later half of the time horizon.

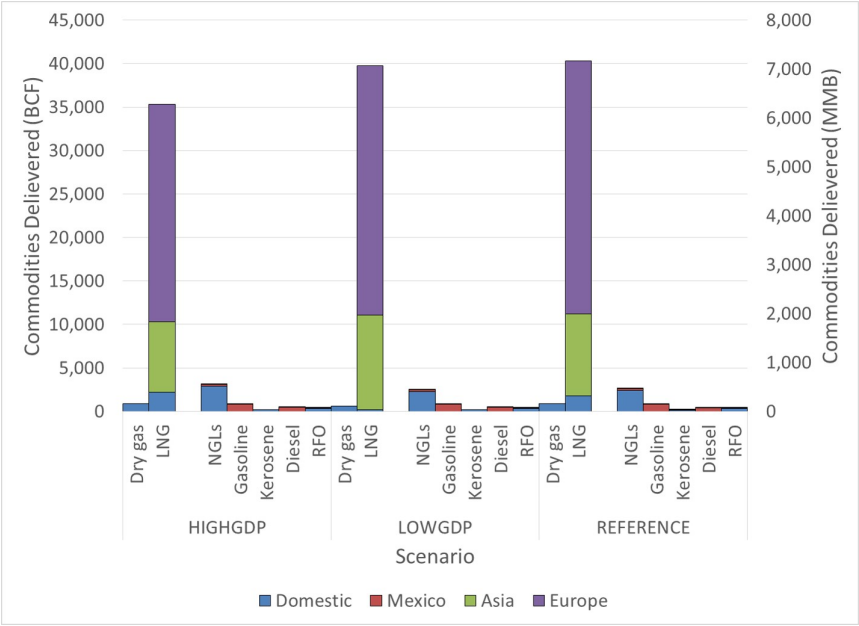
Dry gas, NGLs, Kerosene and RFO were primarily delivered to the Lower 48 States, while gasoline and diesel were primarily delivered to Mexico. The market mix for these products are consistent across scenarios and for both time periods.

In the Oil Price scenario set, the market mix of LNG varied depending on the scenario realized. Across both time periods, LNG was delivered mainly to the Lower 48 and OECD Asia in the High Oil Price scenario, to OECD Asia and OECD Europe with a bias towards OECD Europe in the Low Oil Price scenario case, and to OECD Asia and OECD Europe with a bias towards OECD Asia in the Reference case.

The behavior of varying the LNG market mix depending on the scenario realized lends evidence towards the high sensitivity of LNG profitability in each market to the prevailing oil prices at that point in time and scenario. This serves to explain the more conservative levels of optimal LNG capacity investments in this scenario set, where capacity levels are set to accommodate market-switching behavior while remaining profitable and operationally feasible in each of the realized scenarios. Finally, the market mix patterns for NGLs and refined fuels were similar to the GDP scenario set across all scenarios and time periods.

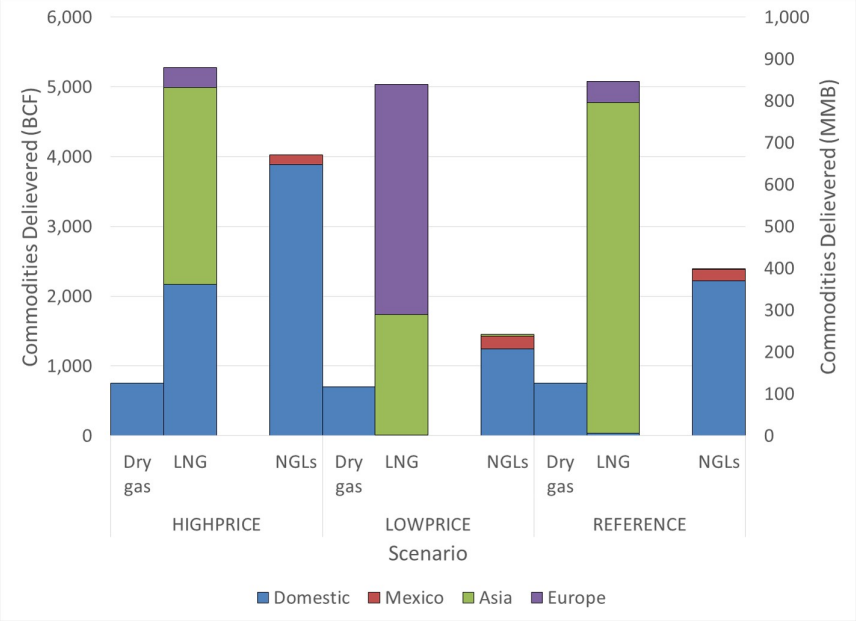


(a) 2016-2027.

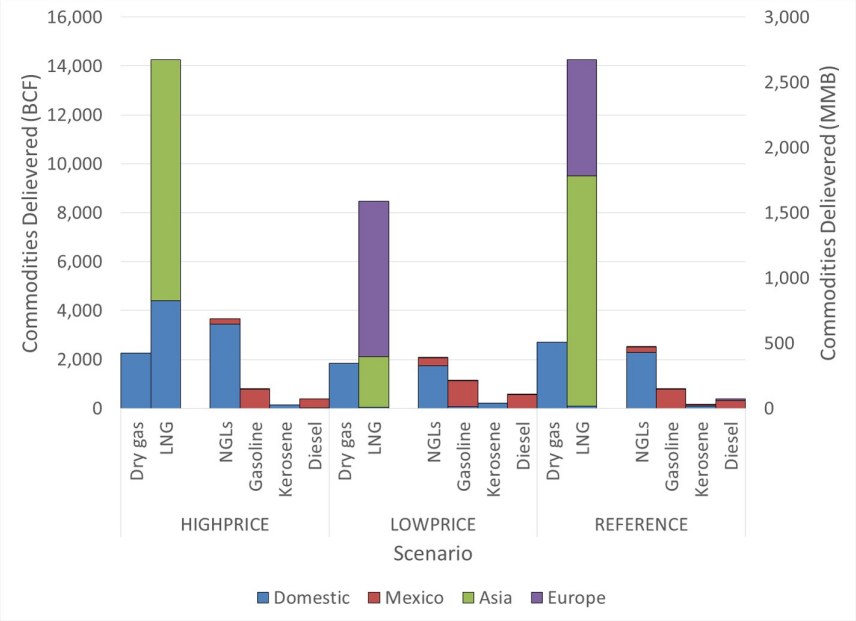


(b) 2028-2039.

Figure 4-11: Commodities delivered for GDP scenario set.



(a) 2016-2027.



(b) 2028-2039.

Figure 4-12: Commodities delivered for Oil Price scenario set.

4.3.6 Transportation utilized

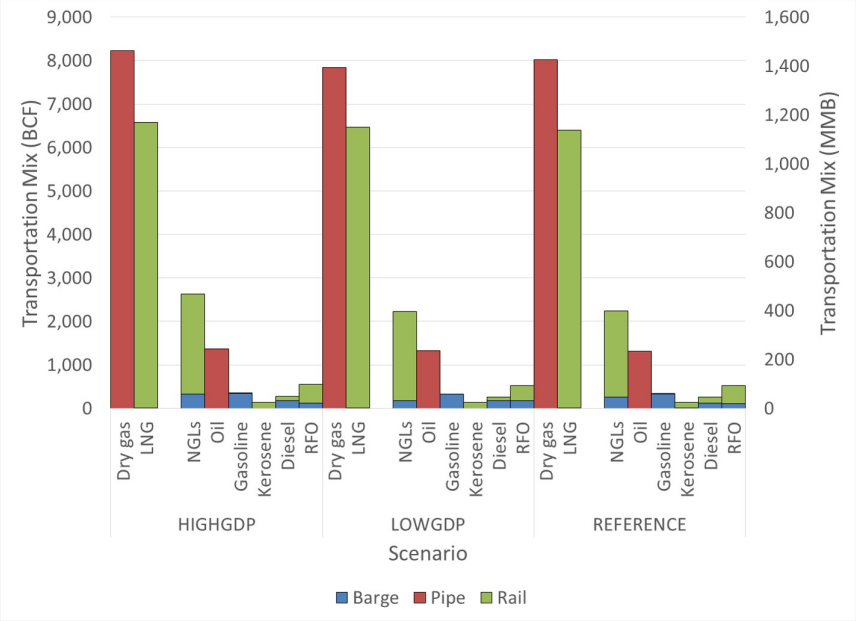
Figures 4-13 and 4-14 depict the various volumes transported for each commodity, partitioned into the various modes, for each scenario in each scenario set. Typically, both rail and barge were favored modes of transportation for the liquid commodities, and road transportation was mostly excluded.

4.3.7 Stochastic versus deterministic optimal investment decisions

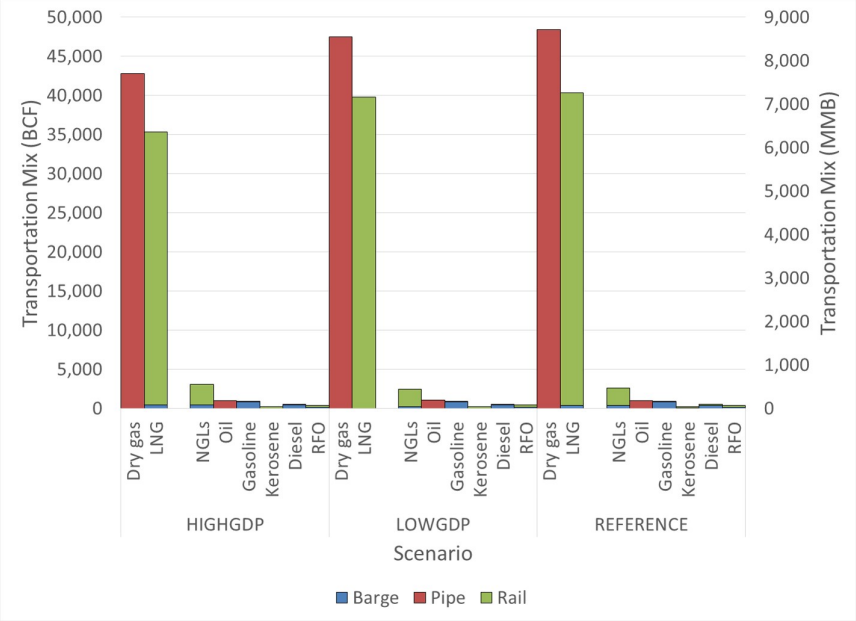
The main advantage of using a stochastic programming framework is to ensure an optimal outcome in the face of uncertainty. Stochastic programming solves two issues by guaranteeing both operational feasibility in all scenario realizations and optimality of the project's expected net present value taken across the scenarios. In other words, it hedges against both operational infeasbility and suboptimality in outcomes.

We can also consider a simpler problem to solve, which we term as the "deterministic" problem. The deterministic problem involves choosing a nominal scenario from the set of possible scenarios and solving an optimization problem with it to determine the optimal investment decisions. The nominal scenario would reflect a "reasonable" scenario which is in line with the decision maker's best guess of how the future would play out. For example, the Reference scenario might be a good choice to set as the nominal scenario. Or, if the decision maker might be more optimistic, he or she could choose a scenario with higher demand or product prices than the nominal scenario.

The deterministic problem serves as a proxy for how many investment projects are analyzed and planned today. That is, investments are made by planning with a nominal scenario projection. Once the investments have been made, a particular

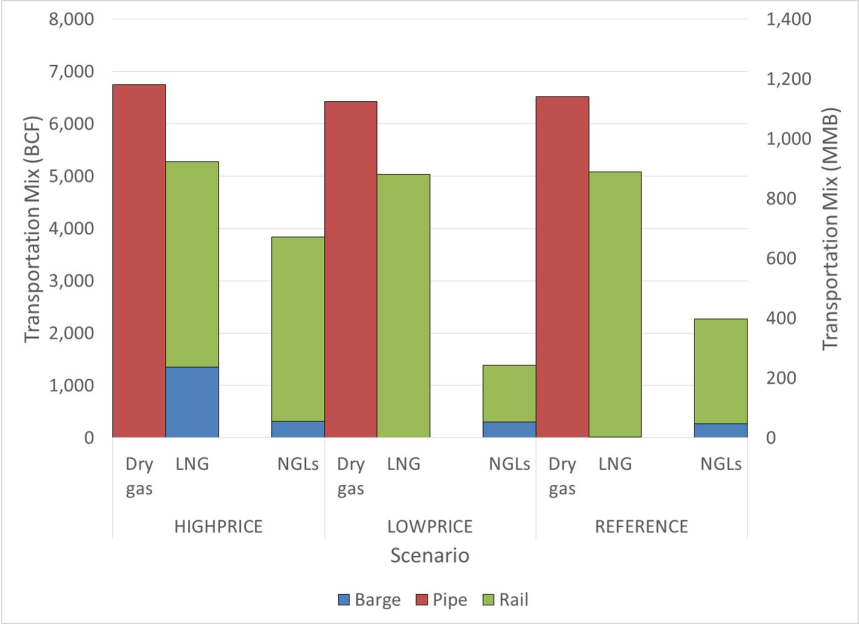


(a) 2016-2027.

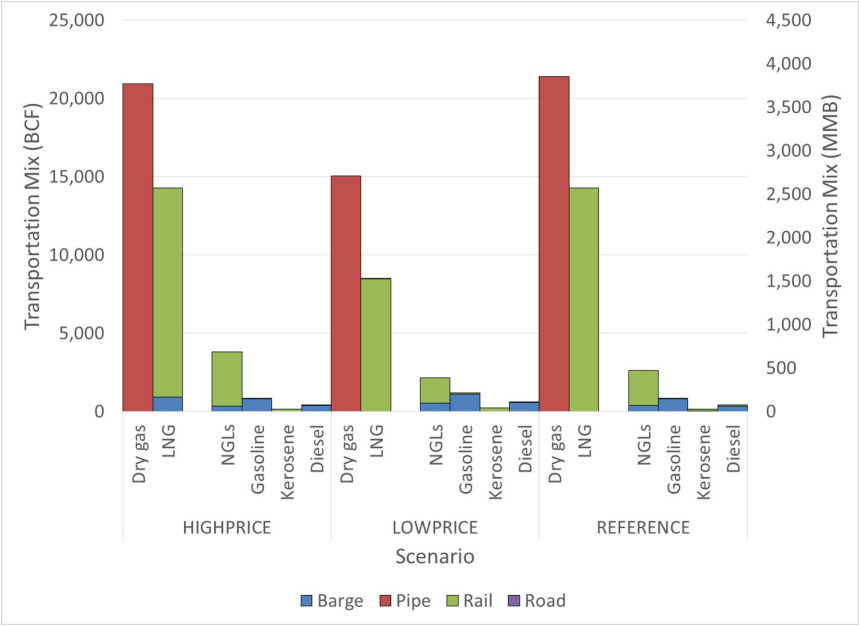


(b) 2028-2039.

Figure 4-13: Transportation utilized for GDP scenario set.



(a) 2016-2027.



(b) 2028-2039.

Figure 4-14: Transportation utilized for Oil Price scenario set.

scenario would then be realized. The decision maker would then have to determine the optimal recourse operational decisions are made in response to the particular scenario realization, constrained by the investment decisions made earlier.

We performed a study to compare the differences in the optimal objective value of the overall project in each realized scenario, conditioned on the investment decisions made either by solving the stochastic problem or the deterministic problem.

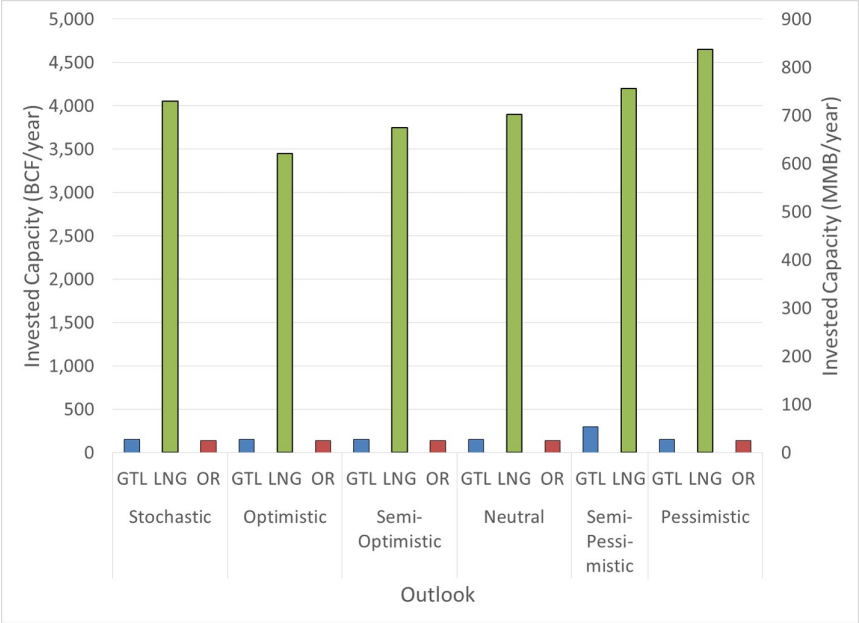
For the deterministic problem, we created five different choices for the nominal scenario, for each scenario set. The five choices were based on five outlooks: optimistic, semi-optimistic, neutral, semi-pessimistic, and pessimistic. Table 4.9 depicts the corresponding nominal scenario for each outlook. Note that by "Avg", we mean that each parameter value at each time point was taken to be the average of its two values in the two selected scenarios at each time point.

Table 4.9: Outlooks and associated nominal scenarios.

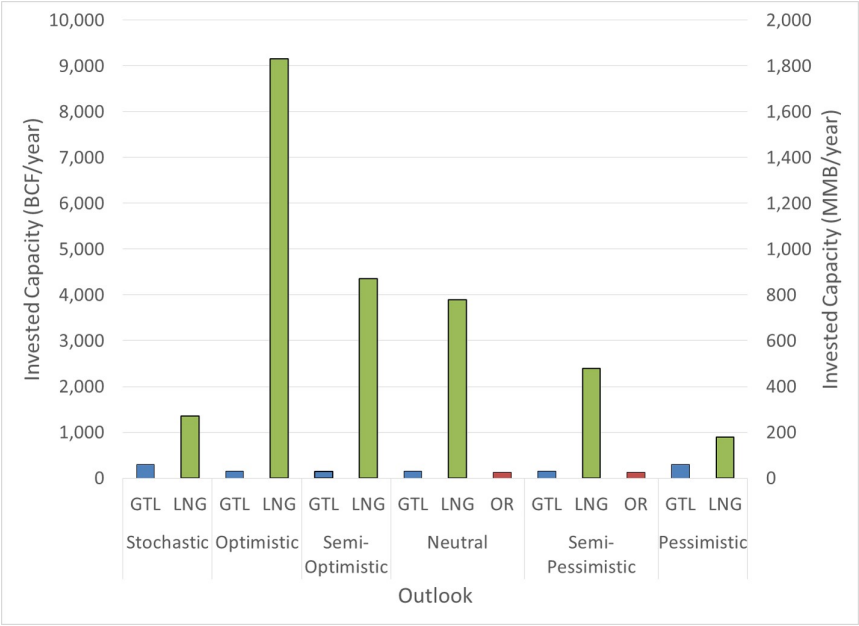
Outlook	**Nominal Scenario**
Optimistic	High
Semi-Optimistic	Avg(High, Ref)
Neutral	Ref
Semi-Pessimistic	Avg(Ref, Low)
Pessimistic	Low

Figure 4-15 shows the variations in optimal plant investments depending on the outlook assumed for each scenario set.

For the GDP scenario set, we observe relatively similar levels and types of investments across all outlooks. There is a slight trend towards building more LNG plants as we transition from an optimistic to a pessimistic outlook, which runs counter to the typical notion that higher GDP is correlated to higher levels of investments. Note that although the levels of capacity investments might be similar for the stochastic



(a) GDP Scenario Set.



(b) Oil Price Scenario Set.

Figure 4-15: Plant investments.

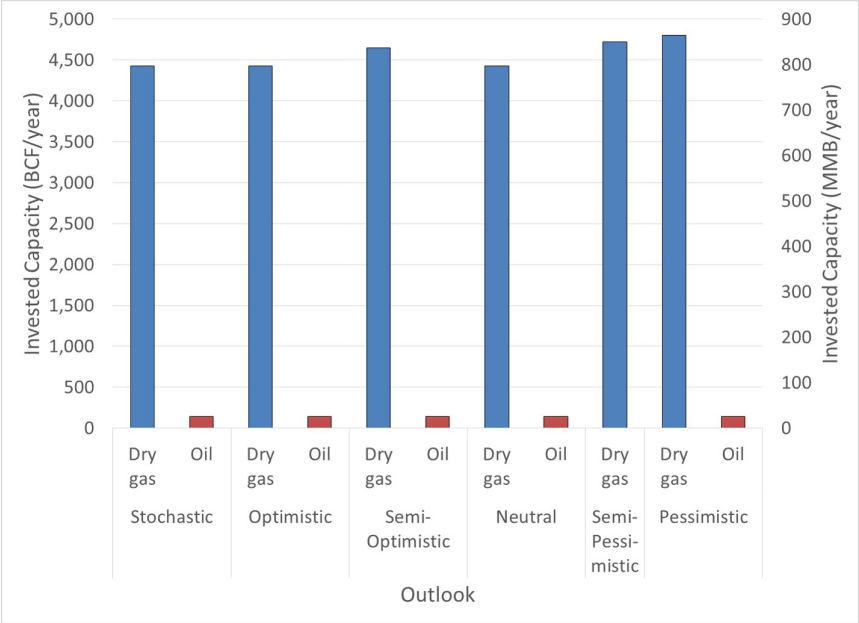
and the neutral or semi-pessimistic case, this does not obviate the need for a stochastic solution. As will be discussed later, the chance of operational infeasibility if an unfavorable scenario is realized is likely under the deterministic solution, even if this might not be obvious from simply considering the aggregate capacity levels of investments.

The Oil Price scenario set is radically different in characterizing the differences in investment levels among the various outlooks. There is a very significant trend towards building more LNG plants as we transition from a pessimistic to optimistic outlook, indicating that high oil prices drive the tendency to invest more in LNG technology. The level of investment in the stochastic case is in between that of the semi-pessimistic and pessimistic case, which indicates a significant amount of over-investment if the decision-maker were to make his or her investment decisions based on optimistic outlooks on the price of oil.

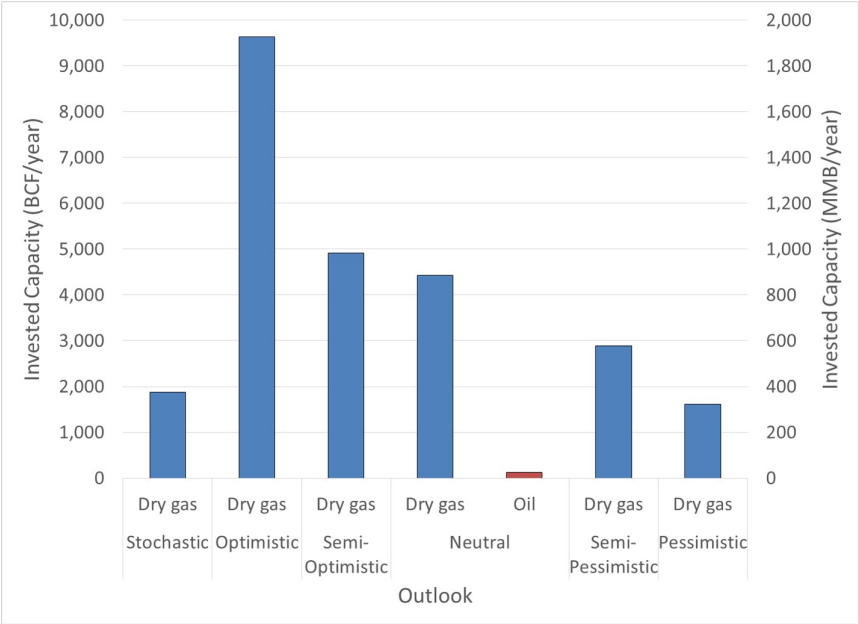
Figure 4-16 shows the corresponding variations in optimal pipeline investments depending on the outlook assumed, for each scenario set. Typically, the investments in dry gas pipelines follow the same trend as that for plant investments across the outlooks within each scenario set, as they adjust proportionally to the different optimal levels of LNG investments.

Given the differences in the optimal investment decisions based on the decision-maker's *ex ante* outlook, we determined how he or she would perform given these investment decisions when a particular scenario is subsequently realized. To simulate this, we solved the recourse problem for each scenario as a linear program, fixing the optimal investment decisions determined earlier from solving either the stochastic or deterministic problem, and determining the optimal operational decisions to maximize profits.

Table 4.10 displays the results from this procedure. We observe that implement-



(a) GDP Scenario Set.



(b) Oil Price Scenario Set.

Figure 4-16: Pipeline investments.

(a) GDP Scenario Set.

Ex-ante Outlook \ **Realized Scenario**	High	Ref	Low
Optimistic	105.9	Infeas.	Infeas.
Semi-Optimistic	100.6	Infeas.	Infeas.
Neutral	Infeas.	102.2	Infeas.
Semi-Pessimistic	89.3	97.8	Infeas.
Pessimistic	88.1	Infeas.	97.0
Stochastic	88.1	96.0	92.8

(b) Oil Price Scenario Set.

Ex-ante Outlook \ **Realized Scenario**	High	Ref	Low
Optimistic	107.9	Infeas.	Infeas.
Semi-Optimistic	85.3	Infeas.	Infeas.
Neutral	Infeas.	102.2	Infeas.
Semi-Pessimistic	Infeas.	Infeas.	Infeas.
Pessimistic	Infeas.	Infeas.	23.4
Stochastic	54.0	59.4	20.2

Table 4.10: Optimal objective values of the recourse problems and initial investments arising from fixing the optimal investment decisions generated from various ex-ante outlooks.

ing the investment decisions from a deterministic solution often leads to operational infeasibility in the majority of scenarios which were not similar to it. Operational infeasibility occurs whenever the flow of material through the invested infrastructure is not sufficient to meet the lower capacity bounds. This might serve as evidence as to why cost and budget overruns are common in large-scale oil and gas investments, which frequently have to contend with significant uncertainty in their decision-making processes.

The relationship between ensuring operational feasibility in all considered scenarios and the foregone profits in a deterministic setting in exchange for this guarantee can be analyzed by comparing the optimal objective values of the stochastic and deterministic implementations for a particular scenario of interest. For example, in the Oil Price scenario set, potentially 53.9 ($= 107.9 - 54.0$) billion dollars of the project's NPV would have to be given up in the High Oil Price scenario, in exchange for the guarantee of operational feasibility should any scenario other than the High Oil Price scenario be realized. Such analyses can guide the decision-makers' choice whether or not to bear risks of operational infeasibility in various scenarios of interest.

4.3.8 Variations in the degree of uncertainty

While the projections from EIA's NEMS serve to be a useful source of future projections, it is also of interest to know how results might change if the degree of uncertainty of the projections increase. As depicted in Figure 4-17, we varied the uncertainty level by adjusting the differences between the High and Low projections for each scenario set. The differences in results which arise from these variations can serve as an important guide for decision makers who would like to assume different degrees of confidence in the variance of future projections.

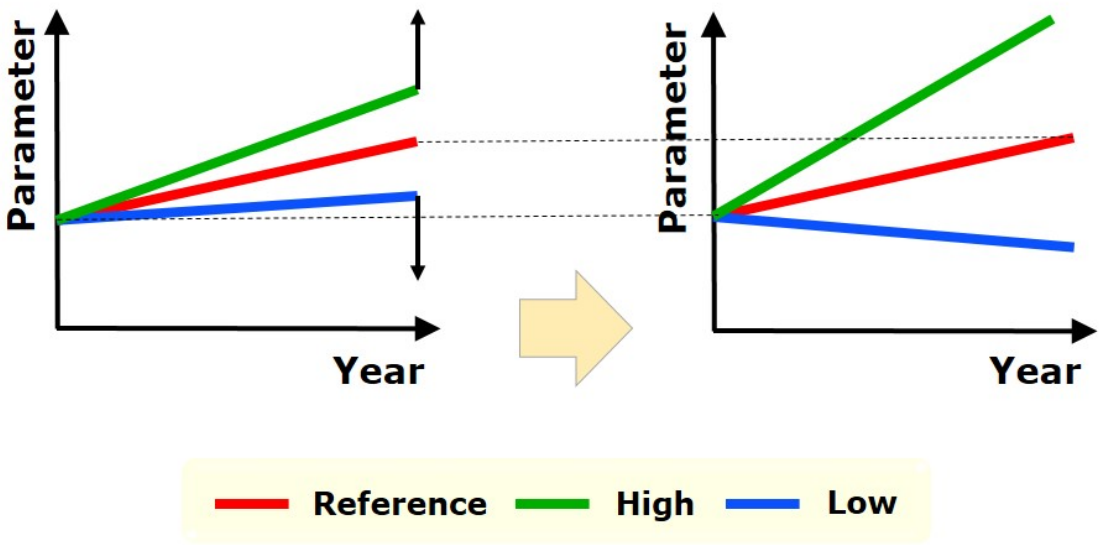


Figure 4-17: The degree of uncertainty of scenario projections were varied by perturbing the parameters of High and Low scenarios from their original values, while keeping the Reference scenario invariant.

For each scenario set, we studied five different degrees of uncertainty, which were measured by the degree of perturbation of the High and Low scenarios from the Reference scenario. The degrees were 0, 20, 40, 60, 80 and 100%. These percentages were expressed as the additional percentage added to the original difference of parameter values from the High or Low scenario to that of the Reference scenario. For example, if a parameter took a value of 30 in the Low scenario, 40 in the Reference scenario and 50 in the High scenario, a 40% perturbation level would imply the adjustment of the parameter value to be 26 in the Low scenario (i.e., $40 + 140\%(30 - 40)$), 40 in the Reference scenario and 54 in the High scenario (i.e., $40 + 140\%(50 - 40)$).

For each perturbation level of the projections of each scenario set, we solved the corresponding stochastic program and analyzed the results. Figure 4-18 depicts how the optimal ENPV changes as a function of the level of perturbation of the

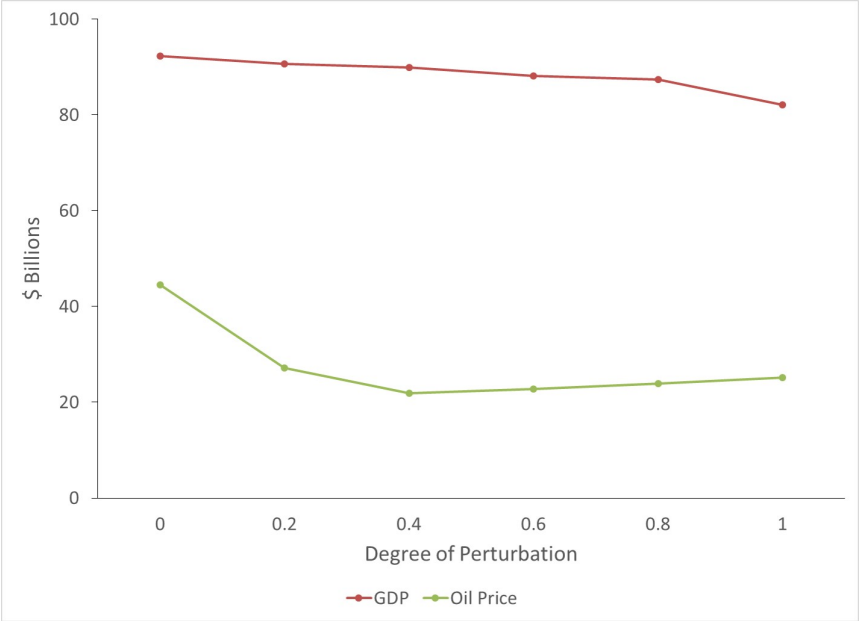


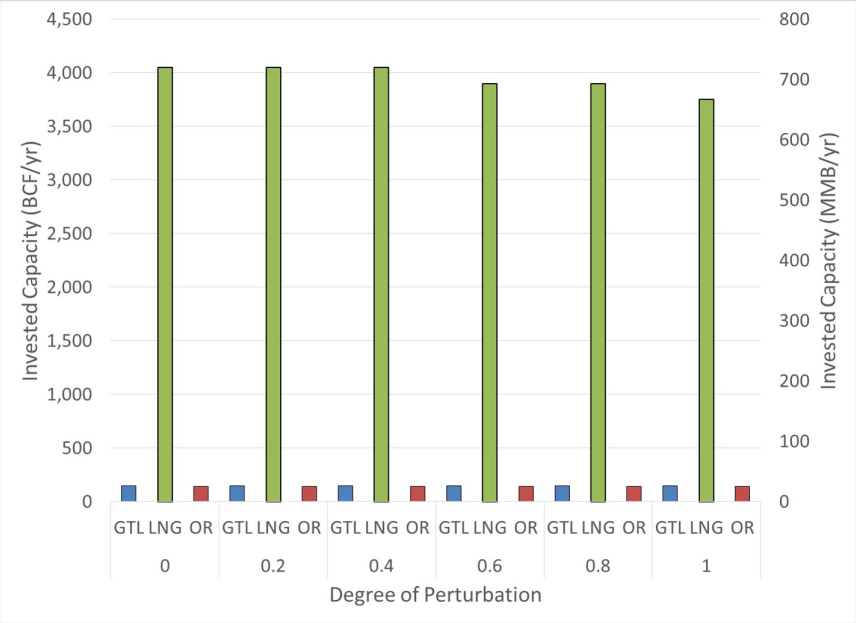
Figure 4-18: Optimal ENPV variations with the level of perturbation of scenario projections from their original values.

projections from their original values.

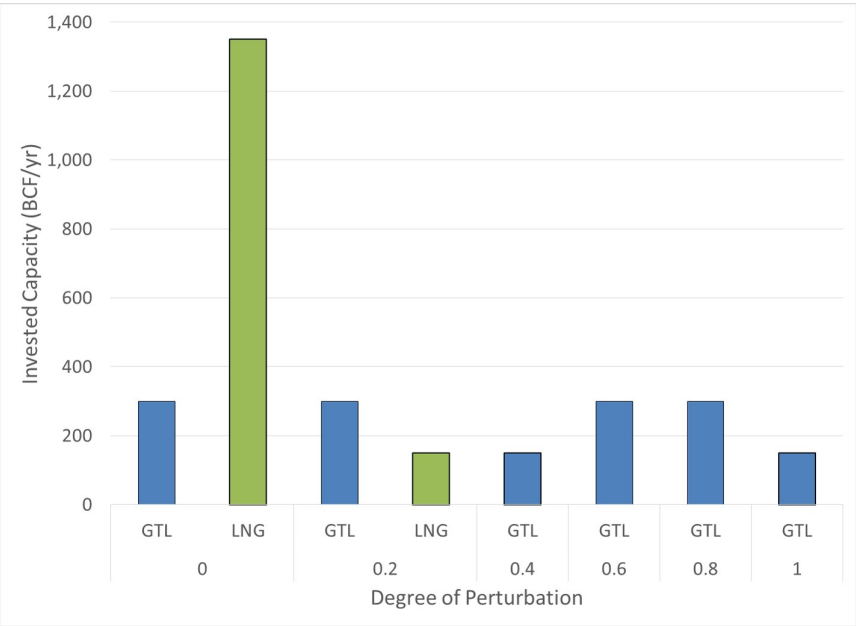
In general, the optimal ENPV decreases with the degree of parameter perturbation for each scenario set. The decrease is gradual for the GDP scenario set. On the other hand, we experience an initial larger decrease, followed by a leveling off and slight increase in the optimal ENPV for the Oil Price scenario set.

The behavior of the optimal ENPV can be explained by referring to Figures 4-19 and 4-20, which show how the optimal investment decisions vary with the degree of parameter perturbation.

The GDP scenario set showed a gradual decrease in the optimal levels of LNG investment capacity with increasing levels of parameter perturbation. This reflected the tendency of the stochastic solution to be more conservative in order to accommodate increasingly disparate scenarios in its consideration. However, the investment

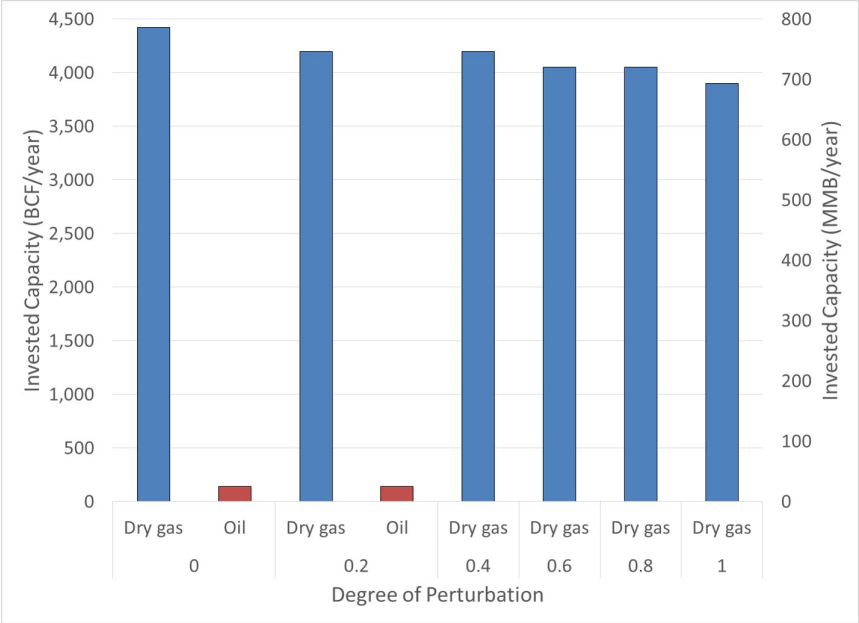


(a) GDP Scenario Set.

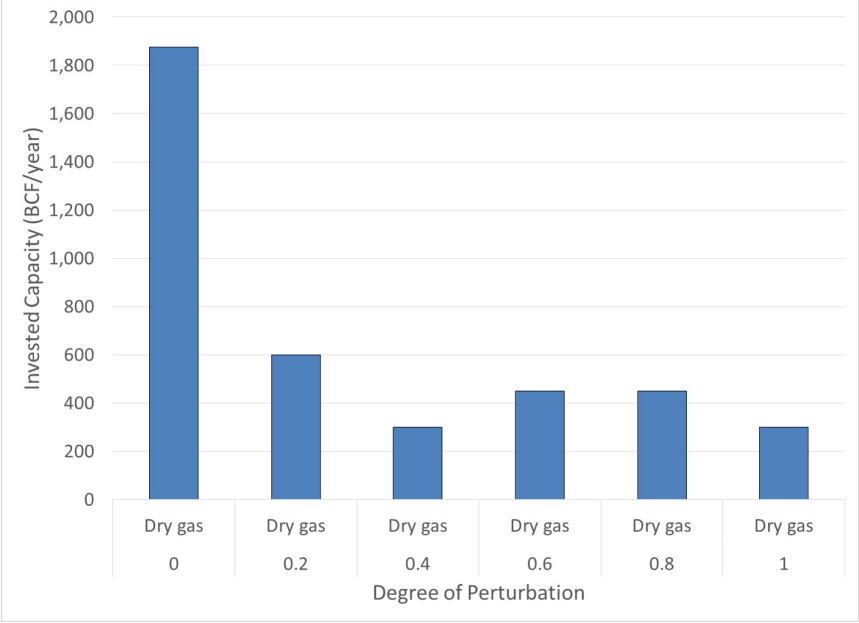


(b) Oil Price Scenario Set.

Figure 4-19: Stochastic plant investments change depending on degree of perturbation.



(a) GDP Scenario Set.



(b) Oil Price Scenario Set.

Figure 4-20: Stochastic pipeline investments change depending on degree of perturbation.

decisions were relatively stable in terms of type and in levels. This corroborated with our observations from previous sections that optimal investment decisions were not very sensitive to the differences in scenarios in this scenario set.

On the other hand, the Oil Price scenario set showed a drastic decrease in LNG investments even at a low degree of parameter perturbation of 20%. In other words, the optimal stochastic solution was highly sensitive to the degree of confidence the decision maker possessed on the potential future evolution of parameters. Assuming too large a spread of possible parameter values arising from the uncertainty in oil prices would cause the stochastic solution to be increasingly conservative and, beyond a certain point, as can be seen from levels of 40% perturbation and beyond, LNG investments were abandoned altogether.

We gain an additional insight that GTL investments were relatively robust towards the degree of uncertainty of future projections, although GTL investments were made at low overall levels. The decrease in LNG investments but relatively stable levels of GTL investments as a function of the degree of parameter perturbation would serve to explain the stabilization of the optimal ENPV with increased parameter perturbation.

This exercise served to demonstrate that although implementing a stochastic solution has its advantages over a deterministic solution in terms of ensuring operational feasibility in the considered scenarios, it also has a disadvantage for this very reason. That is, the optimal stochastic solution might be highly sensitive to the amount of variance present in future projections.

One way to handle the potential instability of the stochastic solution towards the degree of variance in future projections would be to consider the implementation of financial hedging instruments, such as options. For example, cheap out-of-the-money put options on products to be sold might be purchased during the investment phase in

order to effectively "trim" from the scenario tree situations where the product prices are too low. This would serve to control directly the variance of future projections and ensure stability in the stochastic solution at a relatively low cost. The optimal number and characteristics of such options to purchased can be determined with a modification of the optimization problem, although such an exercise is outside the scope of this study but can be implemented in future studies.

4.3.9 Summary of results

In general, we observe a relatively sparse set of types and locations of investments which were common in both scenario sets. However, the main difference between each scenario set was the optimal invested capacity levels of LNG plants. Compared to the GDP scenario set, the Oil Price scenario set had much lower levels of LNG capacity invested and a subsequently lower ENPV over the time horizon. This was mainly attributed to the larger sensitivity of the realized profitability with regards to the different scenario realizations.

The importance of taking into account uncertainty in the investment decisions was also highlighted. Implementing investment decisions based on deterministic optimization of a nominal scenario often led to operational infeasibility if the eventual scenario realized was different. In addition, the degree of variance in the parameter values in the future scenarios also complicated the stability of the stochastic solutions. Our analysis provided a quantitative approach to valuing these trade-offs, which could be used as a guide in the decision-making process.

We briefly comment on how our solution compares with current industry activity with regards to the application to the Federal Energy Regulatory Commission (FERC) for approval to build U.S. LNG export terminals [115]. Strikingly, the opti-

mal locations for the LNG plants matched the areas in which currently approved and proposed LNG export terminals are located. As of June 2015, 9.22 Bcfd of capacity has been approved and is under construction, while an additional 1.40 Bcfd has been approved but not under construction. A further 24.26 Bcfd of capacity is in the proposal pipeline. Comparing these numbers to our results (and assuming a 365 day year), the optimal cumulative levels of investment for LNG plants were 11.10 Bcfd for the GDP scenario set and 3.70 Bcfd for the Oil Price scenario set.

With these numbers, there might be a tendency to believe in a current situation of over-investment in LNG export capacity. However, we state the caveat that our analysis only considered the monetization of any excess supply and demand beyond that of the current year. That is, the higher numbers by FERC might be justified since many applicants are incumbents in the industry. In addition, we assumed that foreign LNG prices were directly indexed to the price of oil. If future conditions indicate a trend towards the delinking of international oil and gas prices, then the optimal investment capacities will change. Nevertheless, the demonstration of the importance of taking into account uncertainty and the resulting conclusions are likely to remain.

4.4 Concluding remarks

We have developed a comprehensive supply chain optimization model of the United States and foreign markets which simultaneously integrates the economic dynamics of the upstream, midstream and downstream sectors of the oil and gas industry which takes into account both the time-varying projections of supply, demand and price parameters as well as the different scenario realizations of these parameters. The model was then used to determine the optimal investment and operating decisions

for monetizing shale oil and gas in the United States moving forward.

We considered two scenario sets in our study which varied in the assumptions of GDP growth and oil prices. Our results show that LNG plants and dry natural gas pipelines were the predominant investments made for both scenario sets. The types and locations of these investments were sparse and common to both scenario sets. The resulting profitability of the investments was much more sensitive to scenario differences in the Oil Price scenario set. As a result, a more hedged solution was optimal for this scenario set, resulting in lower ENPV over the time horizon. Studies were also performed highlighting the importance of including uncertainty in the analysis, and investigating the stability of the stochastic solutions with respect to the degree of variance in future parameters.

Future studies include expanding the current scope of the model, as well as directing it towards further case studies which might be of interest. For example, the model can be made more comprehensive by increasing the degree of granularity in representation or expanding the size of the current sets in the model. This would have to be balanced with the need for a tractable solution of the instance, and could be approached with different algorithmic schemes. Case studies which focus on planning, forecasting or screening could be carried out tailored to the specific needs of the user.

Chapter 5

Examination of the CFSTR Equivalence Principle

The Continuous Flow Stirred Tank Reactor (CFSTR) equivalence principle is a simple and elegant but powerful tool to screen potentially complex reaction networks for their productivities of desired species. Here, while we maintain the correctness of the principle, we present a potential limitation of the principle when applied to a series reaction. Namely, given a reactor of a reasonably moderate size, the principle allows for the maximization of production of any intermediate species regardless of the magnitude of its rate of depletion, thus effectively truncating any reactions downstream from it. This issue arises from the unphysical independence of the variables involving the species molar effluent rate and its concentration.

To eliminate the unphysical independence of the variables, a reformulation was proposed. Nevertheless, this reformulation was unable to solve the issue of the truncation of downstream reactions, but gives the critical additional insight that the target determined by the CFSTR equivalence principle is achieved by allowing infi-

nite flows between the CFSTRs and the omnipotent separator system.

5.1 Introduction

Here, we provide a brief background of the development of the CFSTR equivalence principle.

Assuming any arbitrary reactor design and perfect seperation system, consider the mass balance on the *overall* reactor-separator system:

$$\underline{\mathbf{M}} - \underline{\mathbf{M}}^0 - \underline{\mathbf{R}} = \underline{\mathbf{0}}, \tag{5.1}$$

where $\underline{\mathbf{M}}$ denotes the molar flow rate out of the system, $\underline{\mathbf{M}}^0$ denotes the molar flow rate into the system and $\underline{\mathbf{R}}$ denotes the total molar production rate in the system. $\underline{\mathbf{M}}, \underline{\mathbf{M}}^0, \underline{\mathbf{R}} \in \mathbb{R}^n$, where n is the number of species in the system.

Because we assume any arbitrary reactor design is possible, $\underline{\mathbf{R}}$ would be the sum total of the individual production rate in each *reaction element*. For example, if we consider the design to only consist of three CFSTRs, then:

$$\underline{\mathbf{R}} = V_1 \underline{\mathbf{r}}(\underline{\mathbf{c}}_1, T_1) + V_2 \underline{\mathbf{r}}(\underline{\mathbf{c}}_2, T_2) + V_3 \underline{\mathbf{r}}(\underline{\mathbf{c}}_3, T_3), \tag{5.2}$$

where V_k is the volume of reactor k, T_k is the temperature of reactor k, $\underline{\mathbf{c}}_k \in \mathbb{R}^n$ is the concentration vector in reactor k and $\underline{\mathbf{r}}(\cdot, \cdot) \in \mathbb{R}^n$ is the molar production rate per unit volume.

Another example is if we consider the design to only consist of one PFR, then:

$$\underline{\mathbf{R}} = A \int_0^L \underline{\mathbf{r}}(\underline{\mathbf{c}}(z), T(z)) dz, \tag{5.3}$$

where A is the cross-sectional area of the PFR and z describes the distance from the reactor's entrance, $0 \leq z \leq L$.

Using these two examples as motivation, we now look at the the general case for any arbitrary reactor design. Within the arbitrary reactor, we could discretize the system into reaction elements of volumes ΔV_k, each approximated as a CFSTR, such that the arbitrary reactor is the sum (or integral) of potentially infinitely many reaction elements of different sizes (including those infinitesimally small):

$$\underline{\mathbf{R}} = \Delta V_1 \underline{\mathbf{r}}(\underline{\mathbf{c}}_1, T_1) + \Delta V_2 \underline{\mathbf{r}}(\underline{\mathbf{c}}_2, T_2) + \ldots, \tag{5.4}$$

and

$$\sum_{k=1}^{\infty} \Delta V_k = V^*, \tag{5.5}$$

where V^* is the total volume of the reactor system.

Now if we consider the averaged production rate per unit volume $\underline{\mathbf{r}}^*$, defined by:

$$\underline{\mathbf{r}}^* \equiv \frac{1}{V^*} \underline{\mathbf{R}}, \tag{5.6}$$

then we might write that

$$\underline{\mathbf{r}}^* = \frac{\Delta V_1}{V^*} \underline{\mathbf{r}}(\underline{\mathbf{c}}_1, T_1) + \frac{\Delta V_2}{V^*} \underline{\mathbf{r}}(\underline{\mathbf{c}}_2, T_2) + \ldots, \tag{5.7}$$

and it follows that

$$\sum_{k=1}^{\infty} \frac{\Delta V_k}{V^*} = 1, \quad 0 \leq \frac{\Delta V_k}{V^*} \leq 1 \quad \forall k. \tag{5.8}$$

Immediately, we see that $\underline{\mathbf{r}}^*$ lies in the convex hull of the set $\{\underline{\mathbf{r}}(\underline{\mathbf{c}}_1, T_1), \underline{\mathbf{r}}(\underline{\mathbf{c}}_2, T_2), \ldots\}$. The problem is that this set is potentially infinite and thus computationally in-

tractable because we might have to take infinitesimally small volume elements. This would be the case for example if our reactor includes a PFR among other elements in its design.

To overcome this problem, we appeal to Carathéodory's Theorem:

Theorem 1. *(Carathéodory's Theorem) Let Q be a set in a vector space of dimension n. If $\underline{\mathbf{x}}^*$ lies in the convex hull of Q, then it is possbile to represent $\underline{\mathbf{x}}^*$ as a convex combination of no more than $s+1$ vectors in Q. That is, there are vectors $\left\{\underline{\mathbf{x}}_1, \ldots, \underline{\mathbf{x}}_p\right\}$ in Q and numbers $\{\lambda_1, \ldots, \lambda_p\}$ where $p \leq s+1$, such that*

$$\underline{\mathbf{x}}^* = \lambda_1 \underline{\mathbf{x}}_1 + \ldots + \lambda_p \underline{\mathbf{x}}_p,$$

$$\lambda_1 + \ldots + \lambda_p = 1,$$

and

$$0 \leq \lambda_k \leq 1, \quad k = 1, \ldots, p.$$

In addition, if Q has no more than s topologically connected components, then p can be chosen not to exceed s.

With this, we have converted the potentitally infinite problem into a finite one. Invoking Carathéodory's Theorem, we write:

$$\underline{\mathbf{r}}^* = \lambda_1 \underline{\mathbf{r}}(\underline{\mathbf{c}}_1, T_1) + \ldots + \lambda_{s+1} \underline{\mathbf{r}}(\underline{\mathbf{c}}_{s+1}, T_{s+1}). \tag{5.9}$$

In terms of $\underline{\mathbf{R}}$ we get:

$$\underline{\mathbf{R}} = V^* \underline{\mathbf{r}}^* = \lambda_1 V^* \underline{\mathbf{r}}(\underline{\mathbf{c}}_1, T_1) + \ldots + \lambda_{s+1} V^* \underline{\mathbf{r}}(\underline{\mathbf{c}}_{s+1}, T_{s+1}). \tag{5.10}$$

Denoting $V_k = \lambda_k V^*$, we notice that this representation is identical to that of $s+1$ CFSTRs that exchange materials in a perfect separation system. We thus arrive at the "CFSTR-equivalent" form of equation 5.1:

$$\underline{\mathbf{M}} - \underline{\mathbf{M}}^0 - V_1 \underline{\mathbf{r}}(\underline{\mathbf{c}}_1, T_1) - \ldots - V_{s+1} \underline{\mathbf{r}}(\underline{\mathbf{c}}_{s+1}, T_{s+1}) = \underline{\mathbf{0}}. \tag{5.11}$$

The vector $\underline{\mathbf{r}}^*$ lies in a subspace with dimensions that matches the rank of the stoichiometry matrix of the reaction network. This is because $\underline{\mathbf{r}}^*$ is simply a linear combination of the columns of the stoichiometry matrix. To illustrate, consider the example of the following reaction network:

$$\begin{aligned} \mathrm{A} &\xrightarrow{k_1} \mathrm{B} \xrightarrow{k_2} \mathrm{C} \\ 2\,\mathrm{A} &\xrightarrow{k_3} \mathrm{C} \end{aligned} \tag{5.12}$$

The stoichiometry matrix is then:

$$\underline{\underline{\mathbf{S}}} = \begin{bmatrix} -1 & 0 & -2 \\ 1 & -1 & 0 \\ 0 & 1 & 1 \end{bmatrix}, \tag{5.13}$$

where the rows correspond to the stoichiometric coefficients for each species (A, B, C) while the columns correspond to each reaction (k_1, k_2, k_3).

It can be seen that the rank of $\underline{\underline{\mathbf{S}}}$ is 3, and the vector $\underline{\mathbf{r}}^*$ lies in a vector space of

dim 3, because:

$$\underline{\mathbf{r}}^* = k_1 C_A \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} + k_2 C_B \begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix} + k_3 C_A^2 \begin{bmatrix} -2 \\ 0 \\ 1 \end{bmatrix}. \tag{5.14}$$

So, referring back to equation 5.11, we see that s corresponds to the rank of the stoichiometry matrix of the reaction network.

Based on this principle, Feinberg proposed that the solution of the following nonlinear program will determine the productivity bounds of a certain desired species (eg. M_1) of any given reaction network:

$$\begin{aligned}
&\text{maximize} && M_1 \\
&\text{subject to} && M_i = M_i^0 + V^1 r_i(\underline{\mathbf{c}}_1, T_1) + \ldots + V_{s+1} r_i(\underline{\mathbf{c}}_{s+1}, T_{s+1}),\ i = 1, \ldots, n, \\
& && M_i \geq 0,\ i = 1, \ldots, n, \\
& && V_k \geq 0,\ T^{\text{Max}} \geq T_k \geq T^{\text{Min}},\ P^{\text{Max}} \geq P_k \geq P^{\text{Min}},\ k = 1, \ldots, s+1, \\
& && \underline{\mathbf{c}}_k \geq \underline{\mathbf{0}},\ k = 1, \ldots, s+1, \\
& && V_1 + \ldots + V_{s+1} \leq V^{\text{Max}}, \\
& && g(T_k, P_k, \underline{\mathbf{c}}_k) = 0,\ k = 1, \ldots, s+1 \ \text{(equation of state)}.
\end{aligned} \tag{5.15}$$

The variables are as follows: M_i denotes the molar effluent flow rate of species i. V_k, T_k, P_k denote the volume, temperature and pressure of CFSTR k respectively. $\underline{\mathbf{c}}_k \in \mathbb{R}^n$ denotes the species concentration vector in CFSTR k. The function $r_i(\underline{\mathbf{c}}_k, T_k)$ denotes the rate of formation of species i in CFSTR k per unit volume, while the equation $g(T_k, P_k, \underline{\mathbf{c}}_k) = 0$ describes the equation of state in CFSTR k. The parameters are as follows: M_i^0 denotes the molar supply rate of species i. V^{Max} denotes

the upper bound of the total sum of volumes of all CFSTRs, equivalent to the total volume of the arbitrary reactor, while $T^{\text{Min}}, T^{\text{Max}}, P^{\text{Min}}, P^{\text{Max}}$ denote temperature and pressure bounds of all reactions.

5.2 Applying the CFSTR equivalence principle to a series reaction

As we were examining this principle we came across an interesting observation. Consider the series reaction given as follows:

$$\text{A} \xrightarrow{k_1} \text{B} \xrightarrow{k_2} \text{C}, \tag{5.16}$$

where we would like to maximize the production of species B. Assume that we feed in 40 mol/h of species A into the system.

The reaction network is of rank 2, and the CFSTR equivalence principle allows for three reactors in the optimization problem. However, we restrict the problem to only one reactor, because as will be shown later, using one reactor is enough to fully maximize the production of B.

The optimization problem is formulated as:

$$
\begin{aligned}
\underset{\substack{M_A, M_B, M_C, \\ c_A, c_B, c_C, \\ V, T}}{\text{maximize}} \quad & M_B \\
\text{subject to} \quad & M_A = 40 - V k_1(T) c_A, \\
& M_B = V k_1(T) c_A - V k_2(T) c_B, \\
& M_C = V k_2(T) c_B, \\
& M_A, M_B, M_C \geq 0, \quad c_A, c_B, c_C \geq 0, \\
& 0 \leq V \leq 8000, \quad 373 \leq T \leq 673, \\
& 0 \leq RT(c_A + c_B + c_C) \leq 5.
\end{aligned}
\tag{5.17}
$$

The rate expressions are given as:

$$
\begin{aligned}
k_1 &= 5.12 \times 10^3 \exp(-5535/T), \\
k_2 &= 8.46 \times 10^8 \exp(-10064/T).
\end{aligned}
$$

Because of the nature of the series reaction, we expect that we cannot obtain full conversion of B from A (ie. M_B would not be 40 mol/h), especially if we set $k_2 \gg k_1$. However, from the optimization problem, if we set:

$$
V k_1(T) c_A = 40, \quad c_B = 0,
$$

then the objective function indeed obtains its maximum at $M_B^* = 40$, even if $k_2 \gg k_1$. Thus, any additional CFSTRs added to the formulation are redundant since we are able to maximize M_B without the need for them.

Mathematically, the reason why this occurs is due to the treatment of M_A, M_B, M_C as independent of the concentrations c_A, c_B, c_C. This means that we are able to have an unphysical result where there is maximized molar outflow of M_B even though the concentration c_B in the reactor is zero. While the CFSTR equivalence principle still holds in the sense that the assumption of perfect separation of species implies that we are able to separate a species infinitely fast regardless of the magnitude of its rate of depletion, this implementation would allow for the complete truncation of the reaction downstream of it. Extending this argument for a long chain of series reactions, we see that any intermediate species can be arbitrarily maximized to the feed rate, and any reaction network downstream of that intermediate species can be arbitrarily truncated. Therefore, applying the CFSTR equivalence principle might present a potential limitation when analyzing reaction networks with large depth (ie. having long chains of series reactions).

5.3 A reformulation of the CFSTR equivalence principle

A way to reformulate the problem to express the physicality of the system would be to explicitly define the flows going into and out of each reactor element. Following the development of the original CFSTR equivalence principle, we discretize a reactor of arbitrary design into potentially infinite reaction elements of volume ΔV_k (which might be infinitesimally small). This time however, we explicitly model the flows exchanging between each reaction element and the separator, as shown in Figure 5-1.

A balance around the separator and each individual reaction element yields the

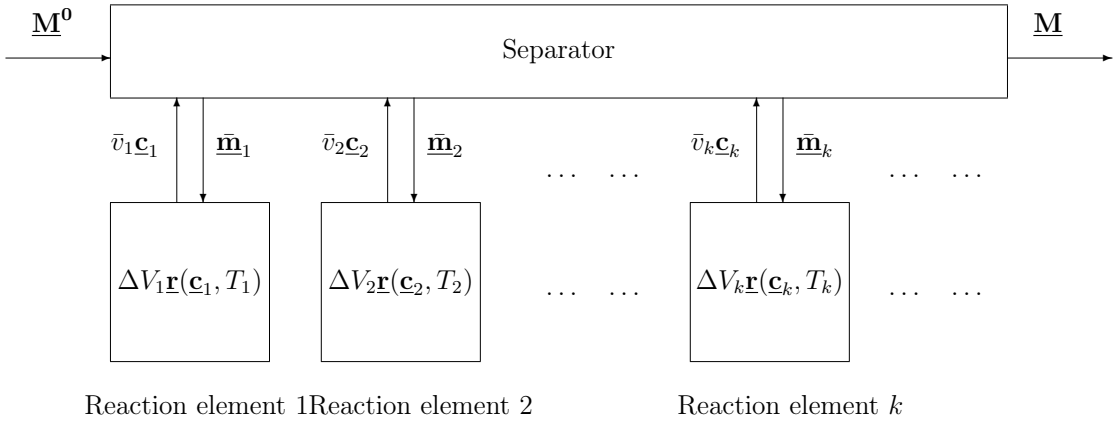


Figure 5-1: Arbitrary reactor formulation.

system of equations:

$$\underline{\mathbf{M}} - \underline{\mathbf{M}}^0 = (\bar{v}_1\underline{\mathbf{c}}_1 - \underline{\bar{\mathbf{m}}}_1) + (\bar{v}_2\underline{\mathbf{c}}_2 - \underline{\bar{\mathbf{m}}}_2) + \ldots, \tag{5.18}$$

$$\bar{v}_k\underline{\mathbf{c}}_k - \underline{\bar{\mathbf{m}}}_k = \Delta V_k\underline{\mathbf{r}}(\underline{\mathbf{c}}_k, T_k) \quad \forall k = 1, 2, \ldots, \tag{5.19}$$

where $\underline{\mathbf{M}} \in \mathbb{R}^n$ denotes the molar flow rate out of the system and $\underline{\mathbf{M}}^0 \in \mathbb{R}^n$ denotes the molar flow rate into the system, where n is the number of species in the system. ΔV_k is the volume of reaction element k, T_k is the temperature of reaction element k, $\underline{\mathbf{c}}_k \in \mathbb{R}^n$ is the concentration vector in reaction element k and $\underline{\mathbf{r}}(\cdot,\cdot) \in \mathbb{R}^n$ is the molar production rate per unit volume.

The only difference between this and the original formulation is that new variables $\underline{\bar{\mathbf{m}}}_k \in \mathbb{R}^n$, the inlet species mass vector flowing into reaction element k, $\bar{v}_k$, the volumetric flow rate out of reaction element k, and additional equations in 5.19 involving mass balances around each individual reaction element have been introduced.

Substituting equations in 5.19 into Equation 5.18, we arrive at

$$\underline{\mathbf{M}} - \underline{\mathbf{M}}^0 = \Delta V_1 \underline{\mathbf{r}}(\underline{\mathbf{c}}_1, T_1) + \Delta V_2 \underline{\mathbf{r}}(\underline{\mathbf{c}}_2, T_2) + \dots . \tag{5.20}$$

Then, we can apply Carathéodory's Theorem to attain

$$\underline{\mathbf{M}} - \underline{\mathbf{M}}^0 = V_1 \underline{\mathbf{r}}(\underline{\mathbf{c}}_1, T_1) + \ldots + V_{s+1} \underline{\mathbf{r}}(\underline{\mathbf{c}}_{s+1}, T_{s+1}), \tag{5.21}$$

where V_k is the volume of CFSTR k and s is the rank of the reaction network.

Using the identity that $V_k = \lambda_k V^*$, where V^* is the total volume of the reactor and λ_k is the coefficient in the convex combination defined in Carathéodory's Theorem, we can define $v_k \equiv \dfrac{\lambda_k V^*}{\Delta V_k} \bar{v}_k$ and $\underline{\mathbf{m}}_k \equiv \dfrac{\lambda_k V^*}{\Delta V_k} \underline{\bar{\mathbf{m}}}_k$ to modify equations in 5.19 to arrive at a CFSTR-equivalent form:

$$v_k \underline{\mathbf{c}}_k - \underline{\mathbf{m}}_k = V_k \underline{\mathbf{r}}(\underline{\mathbf{c}}_k, T_k), \quad \forall k = 1, \ldots, s+1. \tag{5.22}$$

Equations 5.21 and 5.22 together are a valid reformulation such that the CFSTR equivalence principle can be applied in this manner. The reformulation is illustrated in Figure 5-2.

Returning back to the example of the series reaction 5.16, where 40 mol/h of species A was fed into the system, we now reformulate the optimization problem as follows. Again, although we are allowed three CFSTRs, we implement just one for

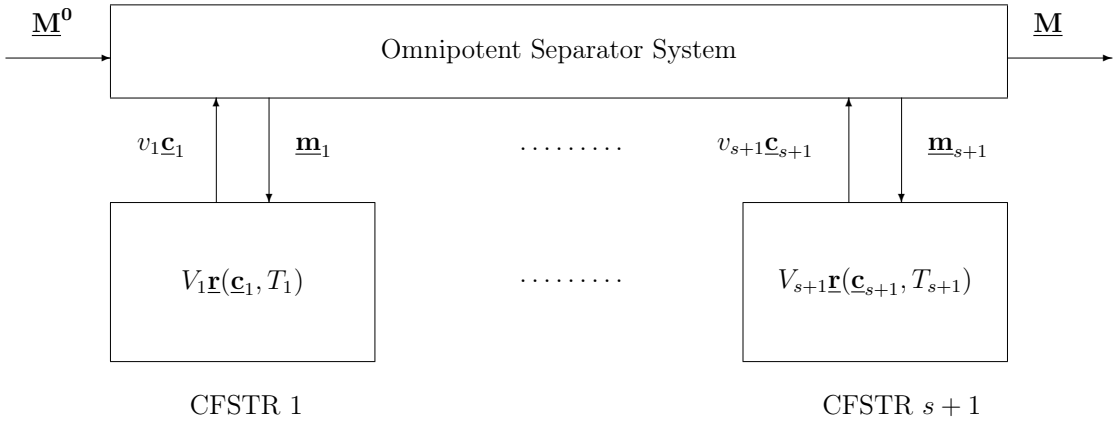


Figure 5-2: CFSTR equivalent formulation.

purposes of illustration:

$$
\begin{aligned}
\text{maximize} \quad & M_B \\
{\scriptstyle M_A, M_B, M_C,\ c_A, c_B, c_C,\ m_A, m_B, m_C,\ v, V, T} \quad & \\
\text{subject to} \quad & M_A = 40 - Vk_1(T)c_A, \\
& M_B = Vk_1(T)c_A - Vk_2(T)c_B, \\
& M_C = Vk_2(T)c_B, \\
& vc_A - m_A = -Vk_1(T)c_A, \\
& vc_B - m_B = Vk_1(T)c_A - Vk_2(T)c_B, \\
& vc_C - m_C = Vk_2(T)c_B, \\
& M_A, M_B, M_C \geq 0, \quad c_A, c_B, c_C \geq 0, \quad m_A, m_B, m_C \geq 0, \\
& v \geq 0, \quad 0 \leq V \leq 8000, \quad 373 \leq T \leq 673, \\
& 0 \leq RT(c_A + c_B + c_C) \leq 5.
\end{aligned}
\tag{5.23}
$$

The idea of this reformulation is to fix the issue of having the situation where $c_B = 0$ but $M_B \neq 0$, which we encountered in the original formulation (and thus M_B could be arbitrarily maximized). To see how this might be fixed, consider the case in the formulation 5.23 where $c_B = 0$. Then we have:

$$-m_B = Vk_1(T)c_A.$$

The LHS is non-positive but the RHS is non-negative, so the only solution is when both sides of the equation are zero. This then forces M_B to be zero as well.

So, consider the case where $c_B \neq 0$. Then, examining the relations

$$0 < Vk_1(T)c_A \leq 40,$$

$$0 \leq M_B = Vk_1(T)c_A - Vk_2(T)c_B \leq 40,$$

$$vc_B - m_B = Vk_1(T)c_A - Vk_2(T)c_B,$$

we see that maximizing M_B is equivalent to maximizing $vc_B - m_B$. Since $m_B \geq 0$, we set $m_B = 0$ so that the expression can be maximized. Now, even though we are maximizing vc_B, we have to minimize c_B because of the relation $M_B = Vk_1(T)c_A - Vk_2(T)c_B$.

This presents a problem, because so long as we set v large enough for any small c_B, we can attain the optimal solution as $M_B^* = 40$. This result is confirmed by global optimizer BARON 11.9.1, where the solution is very close to 40 and is assumed to be at 40, after accounting for the limitations of global optimizers in dealing with unbounded variables.

Therefore, although the reformulation solves the issue of having non-zero species

molar outflows while having zero species concentrations in the reactor, it does not circumvent the issue of allowing for the unphysical maximization of the molar outflow of species B in the given example.

One might wonder if we could solve this issue by setting an upper bound for v. Then we would be unable to arbitrarily set v large enough to compensate for increasingly smaller values of c_B. However, we shall see that if we do so, the CFSTR equivalence principle no longer applies:

Assume by contradiction that there exists an upper bound M such that $v_k \leq M \quad \forall k = 1, \ldots, s+1$, and the CFSTR equivalence principle still applies. Then since $v_k \equiv \dfrac{\lambda_k V^*}{\Delta V_k} \bar{v}_k$ (from the derivation of Equation 5.22), and since V^* and λ_k are bounded, then $\dfrac{\bar{v}_k}{\Delta V_k}$ is bounded by $\dfrac{M}{\lambda_k V^*}$. However, if such a bound exists, then we would exclude reaction elements where for fixed $\bar{v}_k$, $\Delta V_k \to 0$ (a PFR comes to mind), or for fixed ΔV_k, $\bar{v}_k \to \infty$.

Then, the RHS of Equation 5.20 would not lie in the convex hull of all reaction rates in the set $\{\underline{\mathbf{r}}(\underline{\mathbf{c}}_1, T_1), \underline{\mathbf{r}}(\underline{\mathbf{c}}_2, T_2), \ldots\}$ which considers reaction rates found in all possible reaction elements. If so, then Carathéodory's Theorem would no longer apply, and it follows that the CFSTR equivalence principle would not apply. $\square$

To demonstrate our claim, computation simulations were performed on the series reaction example where an upper bound for v_k was set and the number of CFSTRs were varied from one to six. By the CFSTR equivalence principle, since the rank of the reaction network is two, any more than three CFSTRs would be superfluous in attaining a greater production in species B. However, as seen in Table 5.1, the objective function improves as the number of CFSTRs increases beyond three.

Therefore, if we set bounds for v, then the reformulation does not satisfy the CFSTR equivalence principle. On the other hand, if we do not bound v, then we

No. of CFSTRs	1	2	3	4	5	6
M_B^*	9.436	13.859	17.281	20.177	22.681	24.655

Table 5.1: Reformulated problem 5.23 implemented in BARON 11.9.1, where the upper bound of v_k is set at 1×10^3. The optimal value of the objective function is shown as a function of the number of CFSTRs.

still face the problem of the truncation of reactions downstream of the desired species. Although we have not solved the issue of truncation, this exercise in reformulating the CFSTR equivalence principle has granted us the critical additional insight that the Feinberg target is achieved by allowing infinite flows between the CFSTRs and the omnipotent separator system.

5.4 Concluding remarks

The CFSTR equivalence principle as stated with its assumptions gives the correct results. However, when addressing series reactions, there are limitations to obtaining useful information about the productivity of intermediate species. We feel that users seeking to apply the principle should be aware of these additional insights so as to know appropriate situations when the principle can be beneficially applied.

Chapter 6

Conclusions

6.1 Summary and future work

The main contribution of this thesis is the development of novel and comprehensive optimization frameworks for the optimal monetization of shale oil and gas at different scales. At the small scale, a framework that determined the optimal allocation of mobile plants to monetize associated or stranded was developed. At the large scale, a framework to determine optimal shale oil and gas infrastructure investments in the United States was developed.

The development of these frameworks has contributed to the existing literature concerning the application of optimization techniques to the oil and gas sector in two significant ways. First, they are likely to encourage new research activity with the aim towards increasing the depth and breadth of modeling efforts in the field. The small-scale plants study has demonstrated how the traditional unit commitment framework can be applied in a novel area. Opportunities to extend this study are ripe as they can draw on the results on existing literature on unit commitment frameworks.

In addition, we hope that this study spurs further thought on how the traditional unit commitment framework can be applied to other new areas. The nationwide supply chain study has demonstrated, contrary to what has been indicated in most previous studies, that it is now possible to model and solve very large instances of problems that are very comprehensive in their representation of all dimensions of scope commonly considered. We hope that this would encourage greater efforts in seeking a comprehensive representation in future supply chain optimization studies in the field, which would be a large step towards yielding conclusions which can be relied upon or acted on in the real world.

Second, both frameworks have been successfully applied to real-world case studies, the results of which can be readily assimilated by decision makers in industry and government today. The relationship between scale and risk in the studies has been explicitly quantified and allows for a clearer approach to making decisions which require the weighing of trade-offs associated with these two issues. Finally, these frameworks can now be readily applied to further case studies involving the screening of new technologies and the analysis of outcomes with the introduction of new scenarios of interest.

Appendix A

Supplementary Material: Small-Scale Plants

Plant	Capacity (bbl/day)	Capital Costs (2012 dollars / (bbl/day))	Reference
Velocys, small (2012)	1,000	100,000	[38]
Shell Bintulu (1993)	12,000	127,500	[78]
Velocys, large (2012)	15,000	80,000	[38]
Hobbs Jr. study (2012)	20,000	85,000	[36]
Korea study (2009)	32,293	89,773	[78]
Sasol/Chevron Oryx (2006)	34,000	45,000	[78]
AEO2013 study (2013)	34,000	91,800	[78]
Bechtel study (2002)	44,900	61,953	[78]
Montney Shale (Proposed 2017/8)	48,000	83,333	[38]
R.W. Beck study (2010)	50,000	97,964	[78]
Wood study (2012)	50,000	110,000	[38]
Sasol St Charles (Proposed 2018)	96,000	14,280	[78]
Patel study (2005)	100,000	29,500	[43]
Shell Pearl (2011)	140,000	145,714	[78]

Table A.1: Reported capital costs of GTL plants in actual implementation or in literature studies.

Plant	Capacity (gal/day)	Capital Costs (2012 dollars / (gal/day))	Reference
GTI-1000 (2003)	1,000	288	[116]
Nitrogen Open Cycle (2012)	4,200	440	[39]
GTI-5000 (2003)	5,000	175	[116]
West Sacramento (2012)	10,000	600	[39]
Willis plant (2012)	100,000	200	[39]
Boron plant (2012)	240,000	320	[39]
Garcia-Cuerva (2009)	1,700,000	612	[44]
Rep. SMR plant (2012)	2,000,000	210	[39]
Rep. DMR plant (2012)	5,000,000	120	[39]
Rep. APCI C3/MR plant (2012)	8,000,000	110	[39]

Table A.2: Reported capital costs of LNG in actual implementation or in literature studies.

Table A.3: Breakdown of fixed operating costs for small-scale plants. *Rounded to nearest thousand.

	Plant					
Items	GTL, Small	GTL, Medium	GTL, Large	LNG, Small	LNG, Medium	LNG, Large
Operating Personnel ($/Year)	140000	140000	140000	140000	140000	140000
Maintenance Personnel ($/Year)	35000	35000	35000	35000	35000	35000
Administration and Support Personnel ($/Year)	35000	35000	35000	35000	35000	35000
Taxes and Duties ($/Year)	86250	150750	207750	26250	45000	60750
Fixed Maintenance Costs ($/Year)	172500	301500	415500	52500	90000	121500
General Administration Costs ($/Year)	76500	102300	125100	52500	60000	66300
Total Annual Fixed Operating Costs ($/Year)	**545250**	**764550**	**958350**	**341250**	**405000**	**458550**
Total Quarterly Fixed Operating Costs* ($/Quarter)	136000	191000	240000	85000	101000	115000

Table A.4: Breakdown of variable operating costs for small-scale plants

(a) GTL plant variable costs

Items	$/bbl	$/mcf gas feed
Catalysts and Chemicals	6.0	0.51
Water	0.2	0.02
Total	**6.2**	**0.53**

(b) LNG plant variable costs

Items	$/gal LNG	$/mcf gas feed
Chemicals	0.001	0.01
Refrigerant Make-up	0.001	0.01
Total	**0.002**	**0.02**

(a) Markets for GTL diesel under study.

#	Name	Baseline Demand (gallon GTL diesel/day)	City	State	Census Division	Distance (mi)	Shipping Costs ($/gallon)
1	Tesoro West Coast	136,000	Mandan	ND	WNC	225	0.113
2	ExxonMobil Refining & Supply Co.	120,000	Billings	MT	WNC	312	0.156
3	Phillips 66 Co.	118,000	Billings	MT	WNC	316	0.158
4	Cenex Harvest States Coop	119,200	Laurel	MT	WNC	332	0.166
5	Wyoming Refining Co.	28,000	Newcastle	WY	MTN	372	0.186
6	Calumet Montana Refining LLC	20,000	Great Falls	MT	WNC	414	0.207
7	Antelope Refining LLC	7,600	Douglas	WY	MTN	428	0.214
8	Calumet Lubricants Co. LP	76,000	Superior	WI	ENC	605	0.303
9	Flint Hills Resources LP	534,000	Saint Paul	MN	ENC	636	0.318
10	St Paul Park Refining Co. LLC	163,000	Saint Paul	MN	ENC	643	0.322

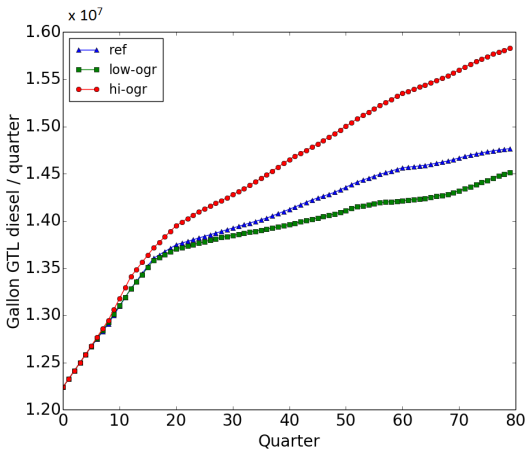
(b) Markets for LNG under study.

#	Name	Baseline Demand (dge/day)	City	State	Census Division	Distance (mi)	Shipping Costs ($/gallon)
1	Blu LNG - Fuel Stop	4,000	Idaho Falls	ID	MTN	660	0.440
2	Blu LNG - Terminal Station	4,000	Pocatello	ID	MTN	708	0.472
3	Blu LNG - Terminal Station	4,000	Raft River	ID	MTN	750	0.500
4	Kwik Trip #870	4,000	La Crosse	WI	ENC	782	0.521
5	Blu LNG - Maverick Travel Plaza	4,000	West Haven	UT	MTN	831	0.554
6	Blu LNG - Blu Travel Plaza	4,000	Myton	UT	MTN	845	0.563
7	Blu LNG - Dunn Travel Plaza	4,000	Salt Lake City	UT	MTN	868	0.579
8	Blu LNG - Flying J Travel Plaza	4,000	Salt Lake City	UT	MTN	869	0.579
9	Blu LNG - Maverick Travel Plaza	4,000	Salt Lake City	UT	MTN	870	0.580

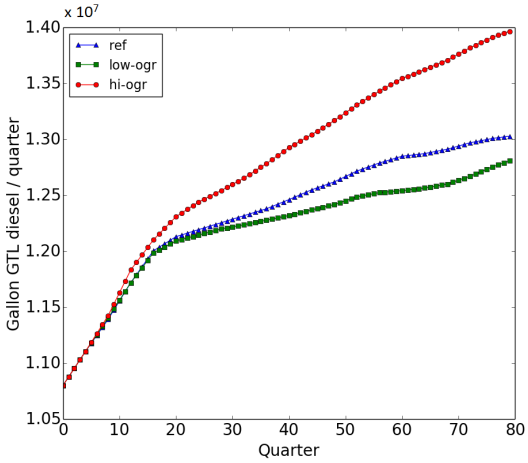
(c) Markets for NGLs under study.

#	Name	Baseline Demand (bbl NGL/day)	City	State	Census Division	Distance (mi)	Shipping Costs ($/gallon)
1	Conway	550,000	Conway	KS	WNC	1,445	0.280
2	Mont Belvieu	1,250,000	Mont Belvieu	TX	WSC	1,614	0.390

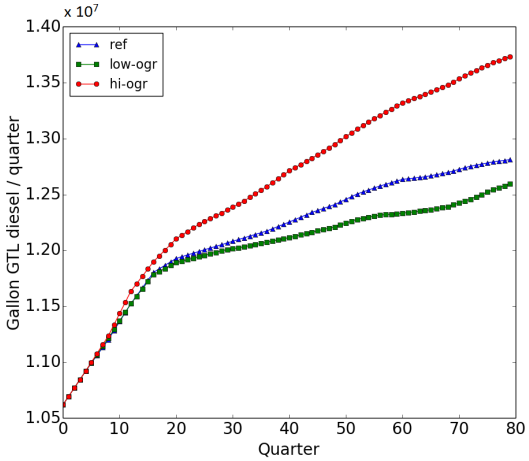
Table A.5: Markets of various products under study. Census Divisions abbreviations: ENC = East North Central, MTN = Mountain, WNC = West North Central, WSC = West South Central.



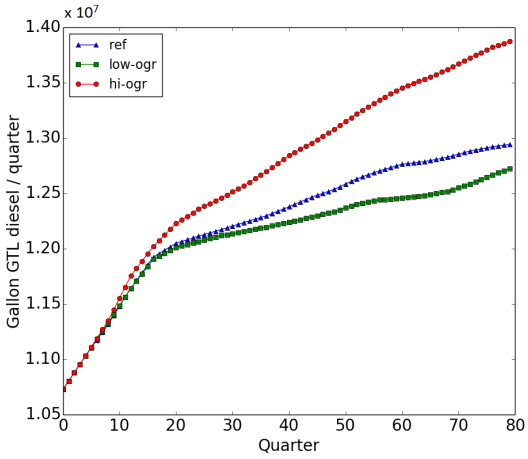
(a) Demand forecasts at GTL diesel market 1.



(b) Demand forecasts at GTL diesel market 2.

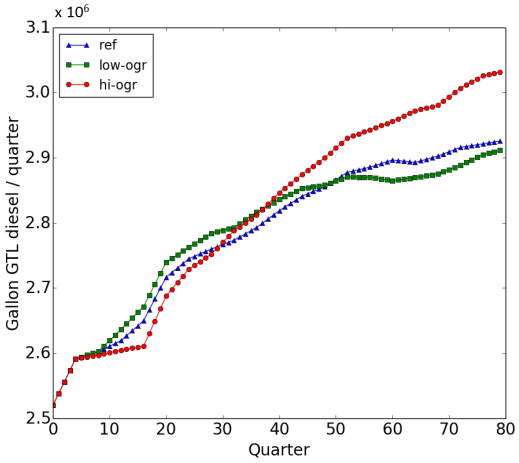


(c) Demand forecasts at GTL diesel market 3.

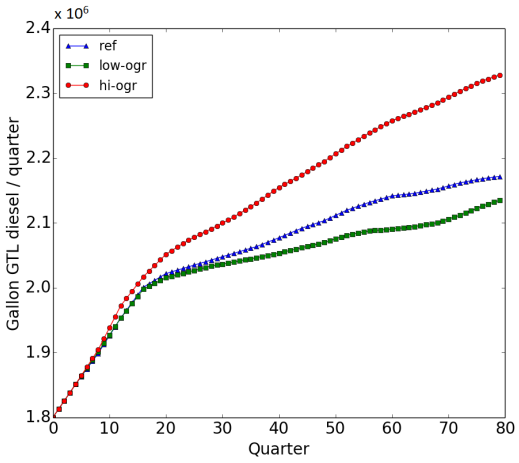


(d) Demand forecasts at GTL diesel market 4.

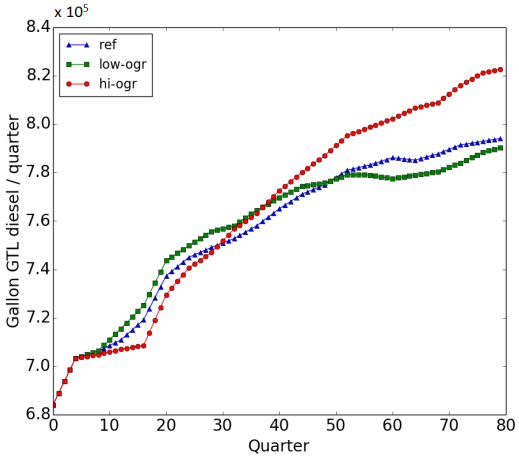
Figure A-1: Demand forecasts for all GTL diesel markets.



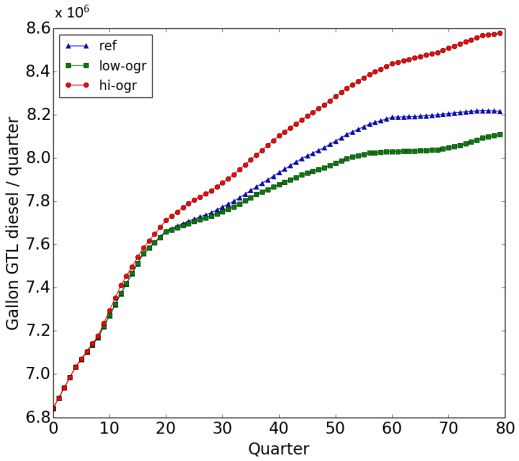
(e) Demand forecasts at GTL diesel market 5.



(f) Demand forecasts at GTL diesel market 6.

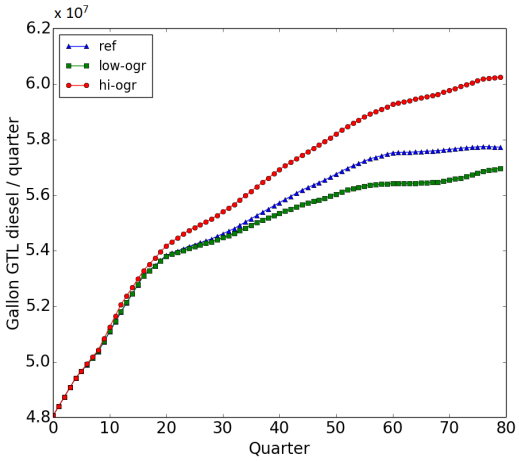


(g) Demand forecasts at GTL diesel market 7.

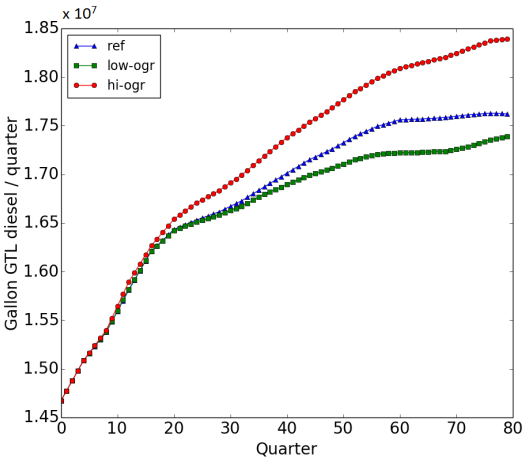


(h) Demand forecasts at GTL diesel market 8.

Figure A-1: Demand forecasts for all GTL diesel markets.

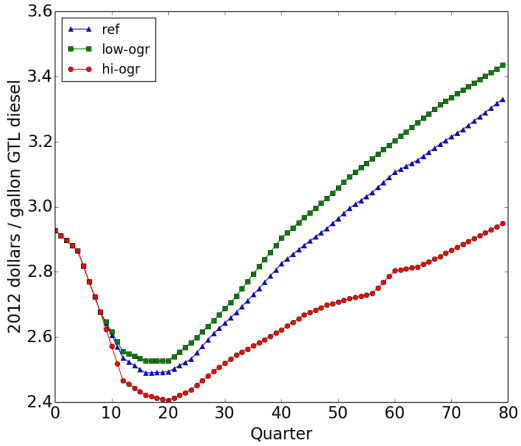


(i) Demand forecasts at GTL diesel market 9.

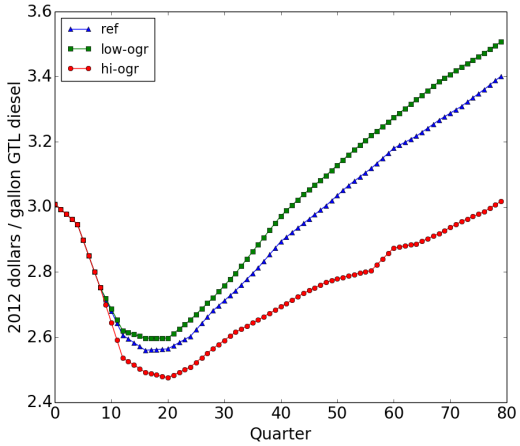


(j) Demand forecasts at GTL diesel market 10.

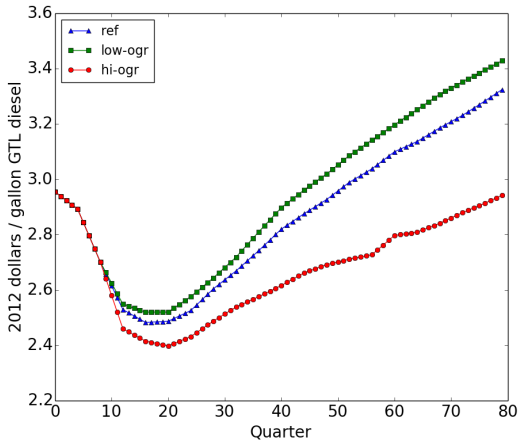
Figure A-1: Demand forecasts for all GTL diesel markets.



(a) Price forecasts at GTL diesel markets 1, 2, 3, 4 and 6.

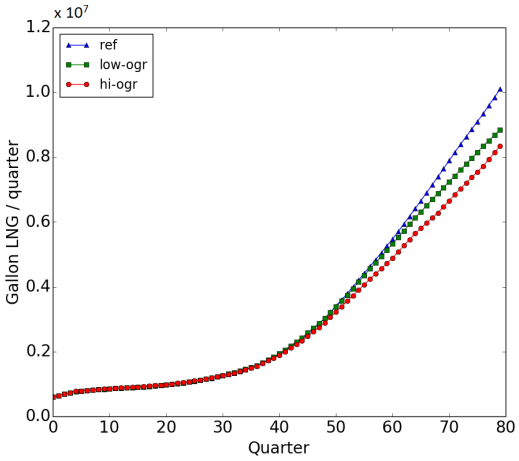


(b) Price forecasts at GTL diesel markets 5 and 7.

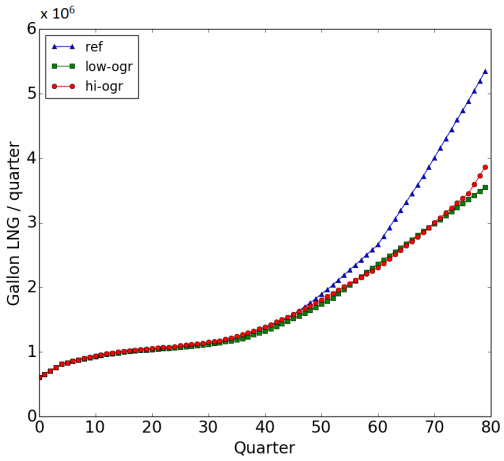


(c) Price forecasts at GTL diesel markets 8, 9 and 10.

Figure A-2: Price forecasts for all GTL diesel markets.

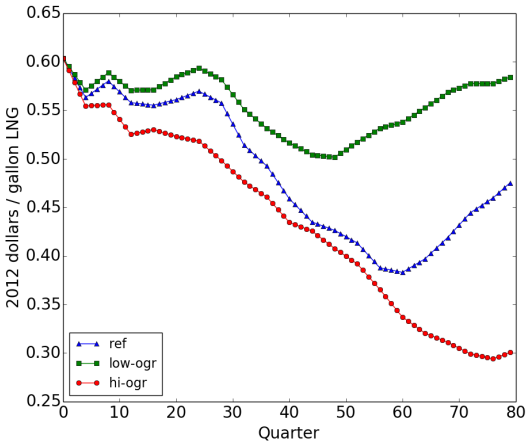


(a) Demand forecasts at LNG markets 1, 2, 3, 5, 6, 7, 8 and 9.

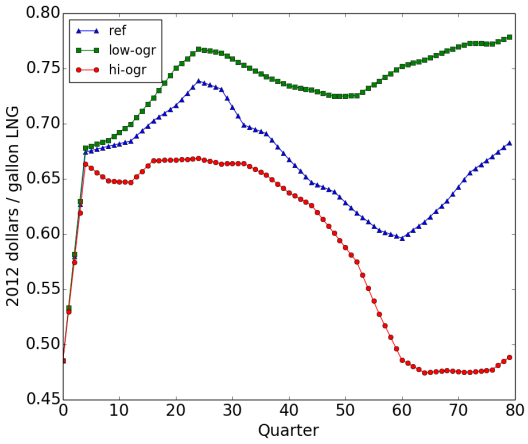


(b) Demand forecasts at LNG market 4.

Figure A-3: Demand forecasts for all LNG markets.

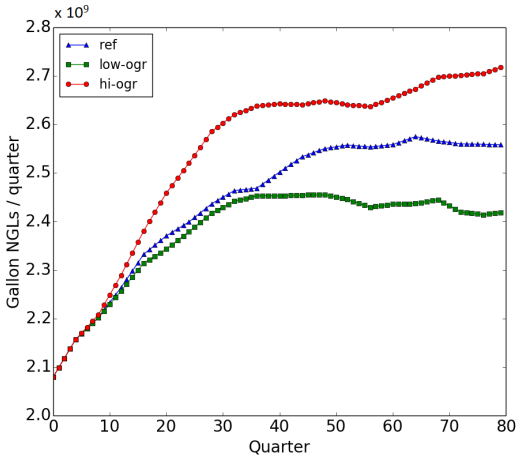


(a) Price forecasts at LNG markets 1, 2, 3, 5, 6, 7, 8 and 9.

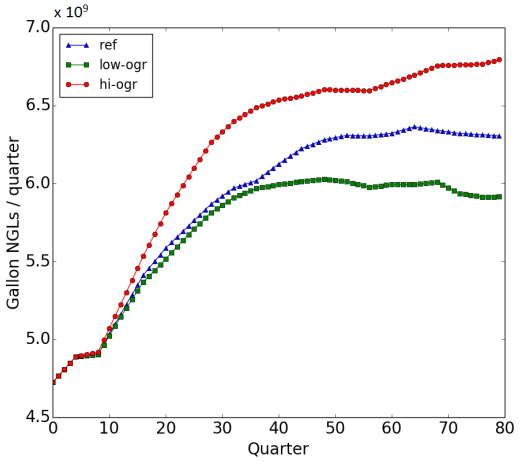


(b) Price forecasts at LNG market 4.

Figure A-4: Price forecasts for all LNG markets.

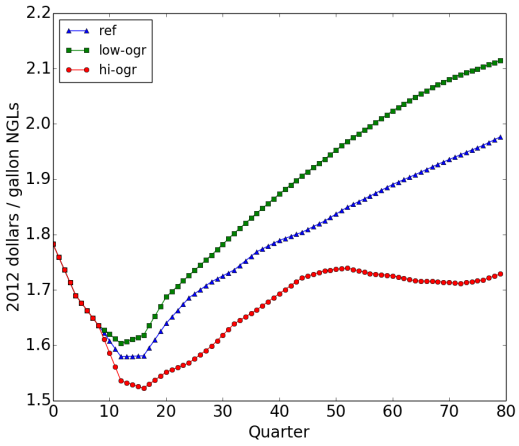


(a) Demand forecasts at NGLs market 1.

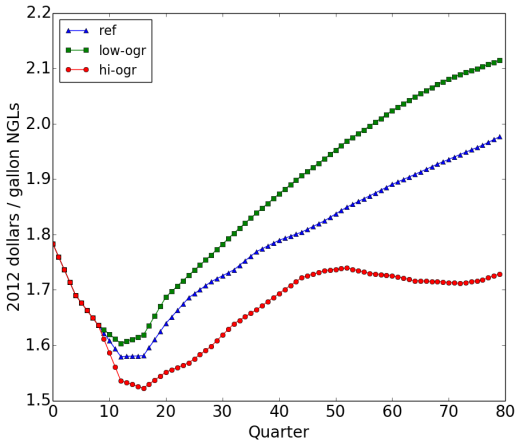


(b) Demand forecasts at NGLs market 2.

Figure A-5: Demand forecasts for all NGLs markets.



(a) Price forecasts at NGLs market 1.



(b) Price forecasts at NGLs market 2.

Figure A-6: Price forecasts for all NGLs markets.

Appendix B

Supplementary Material: U.S. Investments

Assumptions to scenarios in scenario sets

The assumptions for the cases in AEO 2015 are documented in Table B.1. These cases were directly used to generate the corresponding scenarios for this study.

Preparation of sources data

The geographical coordinates of the seven sources used in the study are shown in Table B.2.

Table B.3 shows the correspondence of plays to wet sources that was used to determine NGL content in the wet sources.

Table B.1: Assumptions for the AEO 2015 Cases. Taken from [1]

Case name	Description
Reference	Real gross domestic product (GDP) grows at an average annual rate of 2.4% from 2013 to 2040, under the assumption that current laws and regulations remain generally unchanged throughout the projection period. North Sea Brent crude oil prices rise to $141/barrel (bbl) (2013 dollars) in 2040.
Low Economic Growth	Real GDP grows at an average annual rate of 1.8% from 2013 to 2040. Other energy market assumptions are the same as in the Reference case.
High Economic Growth	Real GDP grows at an average annual rate of 2.9% from 2013 to 2040. Other energy market assumptions are the same as in the Reference case.
Low Oil Price	Low oil prices result from a combination of low demand for petroleum and other liquids in nations outside the Organization for Economic Cooperation and Development (non-OECD nations) and higher global supply. On the supply side, the Organization of Petroleum Exporting Countries (OPEC) increases its liquids market share from 40% in 2013 to 51% in 2040, and the costs of other liquids production technologies are lower than in the Reference case. Light, sweet (Brent) crude oil prices remain around $52/bbl (2013 dollars) through 2017, and then rise slowly to $76/bbl in 2040. Other energy market assumptions are the same as in the Reference case.
High Oil Price	High oil prices result from a combination of higher demand for liquid fuels in non-OECD nations and lower global crude oil supply. OPEC's liquids market share averages 32% throughout the projection. Non-OPEC crude oil production expands more slowly in short- to mid-term relative to the Reference case. Brent crude oil prices rise to $252/bbl (2013 dollars) in 2040. Other energy market assumptions are the same as in the Reference case.

Table B.2: Geographical coordinates of sources.

Source	Longitude	Latitude
Bakken	-102.608	47.984
Eagle Ford	-98.374	28.805
Haynesville	-94.071	32.026
Marcellus	-78.695	40.713
Niobrara	-104.470	40.631
Permian	-101.606	32.108
Utica	-78.655	41.381

Determining candidate plant locations

The generation of the plant locations was performed with ESRI ArcMap 10.2.2 [100]. The candidate plant locations were determined through the use of two datasets, the National Land Cover Database [117], which described land cover, and the U.S. National Atlas Federal and Indian Land Areas [118], which described land ownership.

An evenly-spaced grid of 35 points that provided good coverage of the United States was first generated. Then, each land type dataset was converted into raster form, each with a cell size of 3 km x 3 km. The Weighted Overlay tool from the Spatial Analyst Extension was used to generate a suitability score for each raster cell by equally weighting scores from each input raster. Depending on the land cover or land ownership type, different scores were assigned, with lower scores indicating higher suitability. The scores assigned to both datasets are shown are Tables B.4 and B.5 respectively.

Once the weighted overlay raster had been created, it was then filtered such that only cells with a score of 1 (indicating the most suitable areas) were selected. Then, each plant in the original evenly-spaced grid was then mapped to a cell of score 1 which was located closest to it. The resulting grid of selected cells was then used as

Table B.3: Mapping of plays to wet source for the determination of NGL content. Basins/plays referenced from [2].

Source	Basin	Play
Bakken	Williston	Bakken Central
		Bakken Eastern
		Bakken Elm Coulee-Billings
		Bakken Nesson-Little Knife
		Basin Northwest
		Bakken Three Forks
		Gammon
		Judith River-Eagle
Eagle Ford	Western Gulf	Buda
		Eagle Ford-Dry Zone
		Eagle Ford-Oil Zone
		Eagle Ford-Wet Zone
		Olmos
		Pearsall
		Tuscaloosa
		Vicksburg
		Wilcox Lobo
		Woodbine
Niobrara	Denver	Muddy
		Niobrara
	Greater Green River	Hilliard-Baxter-Mancos
		Tight Oil Plays
	Powder River	Tight Oil Plays
	San Juan	Mesaverde
	Uinta-Piceance	Mancos
		Tight Oil Plays
		Williams Fork
Permian	Permian	Abo
		Avalon/BoneSpring
		Barnett-Woodford
		Canyon
		Spraberry
		Wolfcamp
Utica	Appalachian	Utica-Gas Zone Core
		Utica-Gas Zone Extension
		Utica-Oil Zone Core
		Utica-Oil-Zone Extension

Table B.4: Scores assigned to land cover type from the National Land Cover Database dataset. Detailed descriptions of the land cover classifications can be found at [3].

Classification Key	Description	Score
11	Open Water	Exclude
12	Perennial Ice/Snow	Exclude
21	Developed, Open Space	2
22	Developed, Low Intensity	5
23	Developed, Medium Intensity	8
24	Developed, High Intensity	Exclude
31	Barren Land (Rock/Sand/Clay)	1
41	Deciduous Forest	4
42	Evergreen Forest	4
43	Mixed Forest	4
52	Shrub/Scrub	2
71	Grassland/Herbaceous	2
81	Pasture/Hay	3
82	Cultivated Crops	5
90	Woody Wetlands	Exclude
95	Emergent Herbaceous Wetlands	Exclude
N.A.	Unclassified	1

Table B.5: Scores assigned to land ownership type from the U.S. National Atlas Federal and Indian Land Areas dataset.

Classification Key	Description	Score
DOD	Department of Defense	5
BOR	Bureau of Reclamation	3
BLM	Bureau of Land Management	1
BIA	Bureau of Indian Affairs	1
NPS	National Park Service	Exclude
FWS	Fish and Wildlife Service	Exclude
FS	Forest Service	Exclude
TVA	Tenessee Valley Authority	1
OTHER	Primarily national labs and experimental ranges	2
N.A.	Unclassified	1

the grid of candidate plant locations for the study. Table B.6 shows the geographical coordinates of the candidate plant locations.

Table B.6: Geographical coordinates of candidate plant locations.

Plant	Longitude	Latitude	Plant	Longitude	Latitude
0	-97.638	27.013	17	-96.343	41.120
1	-80.349	26.700	18	-91.376	40.501
2	-102.429	31.553	19	-83.985	39.688
3	-96.943	32.083	20	-77.794	39.286
4	-91.490	31.350	21	-123.010	42.214
5	-85.538	31.951	22	-117.265	43.300
6	-114.498	35.012	23	-109.721	43.576
7	-108.436	35.542	24	-103.764	44.359
8	-103.013	35.692	25	-95.790	45.474
9	-96.963	36.458	26	-90.054	43.913
10	-90.573	36.017	27	-83.954	43.903
11	-84.985	35.953	28	-77.351	42.361
12	-79.720	35.207	29	-71.069	42.140
13	-121.267	38.048	30	-119.090	47.881
14	-114.950	39.255	31	-111.941	48.561
15	-109.324	39.825	32	-104.077	48.907
16	-103.926	40.256	33	-96.298	47.722
			34	-68.996	46.620

Capital costing of plants

The capital costs of hydroskimming refineries were determined by a lookup table which indicated the process cost functions for four main process units that made up the refinery - the desalter, the atmospheric distillation unit, the continuous catalytic reformer and the catalytic resid hydrotreater. Table B.7 shows the parameters used for the process cost functions.

Table B.7: Process cost function parameters for process units in a hydroskimming refinery. Taken from [4].

Process Unit	α	β
Desalter	0.44	0.555
Atmospheric distillation	8.2	0.51
Catalytic reforming, continuous	12.19	0.547
Catalytic hydrotreating, resid desulfurization	8.61	0.834

The capital cost for each process unit was calculated with the formula: Capital Costs (\$M) = $\alpha \cdot$ [Capacity (thousand bbls/stream day)]$^{\beta}$. We assumed an equivalence between a stream day and a calendar day. We arrived at the capital cost for the plant through the summation of the capital costs of the four processes. Finally, we adjusted the capital cost from 2007 to 2015 values by adjusting for the rate of inflation using the U.S. Inflation Calculator [51].

The capital costs of GTL and LNG plants were determined from a previous study that generated cost curves which were fitted to cost data from literature studies and real world implementations of these plants [75, 76]. The costs from the study were updated to 2015 prices using annual inflation estimates from the U.S. Inflation Calculator [51]. The resulting plant capital cost curve for GTL was: ln(Capital Costs (\$B)) = 0.8273 ln(Capacity (thousand bbls/day)) - 2.002, while that for LNG was: ln(Capital Costs (\$M)) = 0.9214 ln(Capacity (million gallons/day)) + 5.425.

Specific calculations for construction of demand series

Foreign LNG demand

The World Natural Gas Consumption by Region (Reference case) data was obtained from the International Energy Outlook 2013 [81] and the relative proportions of Japan, Korea and OECD Europe were determined for each year. OECD Asia was then represented with the sum of the proportions of Japan and Korea. These relative proportions were then multiplied into the Liquefied Natural Gas Export data from AEO 2015 to obtain the LNG demand for each foreign entity.

Foreign NGLs and refined products demand

Initial export values of NGLs and refined products were determined from EIA's Exports by Destination annual data for 2014 [86]. The determination was performed by mapping each commodity in our study to one or several exported products provided by EIA, depicted by Table B.8. The initial export value for each commodity in each foreign market was determined by summing across all export values of products and countries which corresponded to that particular commodity and foreign market respectively.

Table B.9 documents the initial values of liquid commodities demand for each foreign market resulting from these calculations, which were used to multiply into the original International Liquids Use demand series from AEO 2015 to generate individual demand projections for each commodity. Note that these initial values were used as baselines for which the final demand series which represented excess over these baselines were input into the optimization model.

Table B.8: Mapping of commodities in study to exported products classification in EIA's Exports by Destination data.

Commodity	Exported Products by EIA's Classification
NGLs	Pentanes Plus
	Liquefied Petroleum Gases
Gasoline	Finished Motor Gasoline
	Motor Gasoline Blending Components
Kerosene	Kerosene
	Kerosene-Type Jet Fuel
Diesel	Distillate Fuel Oil
RFO	Residual Fuel Oil

Table B.9: Initial demand values of NGLs and refined products for each foreign market. Units are in MMB/year.

Commodity \ Foreign Market	Mexico	OECD Asia	OECD Europe
NGLs	24.215	23.860	36.028
Gasoline	86.717	0.035	1.837
Kerosene	4.585	0.001	6.162
Diesel	46.445	0.424	107.104
RFO	8.416	1.548	14.395

Initial prices used for price series in foreign markets

Following the procedure as described in the main paper, the initial prices of the commodities in foreign markets used in the study are documented in TableB.10.

Table B.10: Initial prices used for price series in foreign markets.

Commodity \ Foreign Market	Mexico	OECD Asia	OECD Europe
Dry gas	3.17	N.A.	N.A.
LNG	N.A.	10.00	8.45
Crude Oil	76.78	61.57	56.86
NGLs	72.64	71.85	44.12
Gasoline	87.17	86.56	84.98
Diesel	92.97	82.38	94.05
Kerosene	92.97	98.23	94.05
RFO	48.05	71.93	66.40

Prices of dry gas and LNG in $ per mcf.
Prices of crude oil, NGLs and refined products in $ per bbl.

Tables for data series

Tables B.11 and B.12 show the corresponding AEO 2015 series used for future projections of the parameters in our study at the sources and markets respectively. The AEO 2015 series were used in one of two ways to determine future parameter values - either they were used as-is, or rather, in the cases where initial values were separately determined, the AEO 2015 series were used to scale the evolution of the initial values over time.

Table B.11: AEO 2015 series used for determining the future evolution of corresponding parameter values at the sources.

Parameter	AEO 2015 Series
Supply of natural gas and NGLs	Natural Gas : Dry Production : Lower 48 Onshore
Supply of crude oil	Crude Oil : Production : Lower 48 Onshore
Price of natural gas and NGLs	Natural Gas : Supply Prices : Lower 48 Onshore
Price of crude oil	Crude Oil : Wellhead Prices : Lower 48 Onshore

Table B.12: AEO 2015 series used for determining the future evolution of corresponding parameter values at the markets. Asterisks represent specially constructed series, as described in the body of the paper. * : Appropriately apportioned quantities of the AEO 2015 Natural Gas : Volumes : Exports : Liquefied Natural Gas Exports demand series, according to relative foreign consumption ratios in IEO 2013 Reference Case. ** : 'Mexico Blend' prices, which are the average of Brent and WTI prices.

Parameter	AEO 2015 Series
Local demand for dry gas	Energy Use : Delivered : All Sectors : Natural Gas - Energy Use : Transportation : Natural Gas
Local demand for NGLs	Energy Use : Delivered : All Sectors : Liquefied Petroleum Gases
Local demand for LNG	Energy Use : Transportation : Natural Gas
Local demand for gasoline	Energy Use : Delivered : All Sectors : Motor Gasoline
Local demand for kerosene	Energy Use : Delivered : All Sectors : Kerosene + Energy Use : Delivered : All Sectors : Jet Fuel
Local demand for diesel	Energy Use : Delivered : All Sectors : Distillate Fuel Oil
Local demand for RFO	Energy Use : Delivered : All Sectors : Residual Fuel Oil
Mexico demand for dry gas	Natural Gas : Volumes : Exports : Pipeline Exports to Mexico
Mexico demand for NGLs, gasoline, kerosene, diesel and RFO	International Liquids : Use : OECD : Mexico and Chile
Mexico, OECD Asia and OECD Europe demand for LNG	*
OECD Asia demand for NGLs, gasoline, kerosene, diesel and RFO	International Liquids : Use : OECD : Japan + International Liquids : Use : OECD : South Korea
OECD Europe demand for NGLs, gasoline, kerosene, diesel and RFO	International Liquids : Use : OECD : OECD Europe
Local prices of dry gas	Energy Prices : Average Price to All Users : Natural Gas
Local prices of NGLs	Energy Prices : Average Price to All Users : Propane
Local prices of LNG	Energy Prices : Transportation : Natural Gas
Local prices of gasoline	Energy Prices : Average Price to All Users : Motor Gasoline
Local prices of kerosene	Energy Prices : Average Price to All Users : Jet Fuel
Local prices of diesel	Energy Prices : Average Price to All Users : Distillate Fuel Oil
Local prices of RFO	Energy Prices : Average Price to All Users : Residual Fuel Oil
Mexico prices of dry gas	Natural Gas : Border Prices : Pipeline Import Prices : From Mexico
Mexico prices of NGLs, gasoline, kerosene, diesel and RFO	**
OECD Asia and OECD Europe prices of NGLs, LNG, gasoline, kerosene, diesel and RFO	International Liquids : Crude Oil Prices : Brent

Capital costing of pipelines

For dry natural gas pipelines, an aggregated list of natural gas pipeline and expansion projects in operation or slated to commence operations in coming years, compiled by the EIA, was used as the sample set [119]. The data was compiled from FERC, trade press, company websites and other industry sources. The sample set was filtered to exclude samples with zero cost or zero length. The costs were further adjusted to 2015 prices by using annual inflation estimates from the U.S. Inflation Calculator [51].

For oil pipelines, a sample set was created using a Reuters-compiled list of U.S. crude oil pipeline projects either fully or partially in service, nearing completion, under construction, or planned [120]. Only projects with the full set of cost, capacity and length data, either procured directly from the Reuters source or through other industry reports, were included in the sample set. The capital costs were assumed to be appropriate at 2015 levels, since most projects are relatively new. Table B.13 documents the final list of projects used for the regression.

Table B.14 shows the estimated regression coefficients for the cost curves of the pipelines associated with the different commodity types. Due to a lack of data for NGLs and refined products pipelines, we used the cost curve for the oil pipelines to determine their capital costs.

Note that the regression coefficients were less than one and significant, providing evidence for economies of scale. The estimated regressions were then used to determine capital costs for every possible node-to-node pipeline generated in the study.

Table B.13: List of U.S. oil pipeline projects used for determining oil pipeline cost curve.

Project	Capacity (KB/day)	Length (Miles)	Capital Costs ($M)
Eagle Ford Joint-Venture Pipeline First Expansion	120	140	120
Double Eagle Pipeline	100	140	150
Knight Warrior Pipeline	100	160	300
Double H Pipeline	84	488	375
Cactus Pipeline	330	310	450
Line 67 (Alberta Clipper) Expansion	350	700	450
Kinder Morgan Crude and Condensate Pipeline	400	147	555
Pony Express Pipeline	320	260	725
Permian Basin Projects	860	182	800
Southern Access Extension	300	165	800
Saddlehorn Pipeline	400	600	850
Diamond Pipeline	200	440	900
BridgeTex Pipeline	300	450	1000
Seaway Pipeline	700	500	2000
Gulf Coast Pipeline Project	700	535	2300
Flanagan South Pipeline	600	600	2600
Sandpiper Pipeline	600	608	2600
Dakota Access Pipeline	320	1134	5000
Keystone XL Northern Leg	830	1179	5300

Table B.14: Regression coefficient estimates for log-log relation between pipeline capital costs per length and capacity. Standard errors are shown in parentheses.

Commodity	$\hat{\beta}_0$	$\hat{\beta}_1$
Dry natural gas	-0.468	0.304
	(0.148)	(0.034)
Oil, NGLs and refined products	-2.306	0.663
	(0.772)	(0.160)

Units: Capital Costs ($M), Capacity (oil, NGLs and refined products) (MMB/yr), Capacity (dry natural gas) (BCF/yr), Length (miles).

Transportation Routes

Adjustments made to port-to-port routing calculations

Port-to-port distances were mainly calculated using the Sea Route & Distance calculator from Ports.com [110]. However, in some instances, routing information was not available. This section documents the additional calculations made in order to obtain a complete set of port-to-port distances for the study.

Adjustments were made to ports located along the Delaware River (New Castle, DE, Wilmington, DE, Marcus Hook, PA, Paulsboro, NJ, Camden-Gloucester, NJ, and Philadelphia, PA). First, pairwise distances among these ports were determined by the Distances Between United States Ports publication by the Office of Coast Survey, National Ocean Service, National Oceanic and Atmospheric Administration (NOAA) [121]. An additional assumption that Paulsboro was located 1/3 and Camden-Gloucester was located 2/3 the way from Marcus Hook to Philadelphia was made. The distance between a port not in the Delaware River system and one in it was calculated as the sum of the distance between the former port and New Castle (as determined from Ports.com) and the distance between New Castle and the latter port.

Distances to Freeport, TX and Tacoma, WA were determined by first selecting ports located close-by to each, which were Galveston, TX and Anacortes, WA respectively. Port-to-port distances from any other port to Freeport or Tacoma were then determined by summing up the distance first to the corresponding ports located close-by (as determined from Ports.com) with the distance from these corresponding ports to either Freeport or Tacoma (as determined by NOAA's publication).

Distances to Hempstead, NY and Port Jefferson, NY were determined in a similar

manner. The closest port located to both ports was Bridgeport, CT. The distance from Bridgeport to Port Jefferson was 15 nautical miles (as determined from NOAA's publication), and distances from Bridgeport to Hempstead and from Port Jefferson to Hempstead were estimated at 30 and 15 nautical miles respectively.

Adjustments were also made for ports in the Ohio River system (Mount Vernon, IN, Louisville, KY, Cincinnati, OH, Huntington, WV, and Pittsburgh, PA). The distances between each of these ports were determined. In addition, the distance from each of these ports to a reference port, Memphis, TN, was determined by the sum of distance from each of these ports to the Ohio River mouth, and the distance from the Ohio River mouth to Memphis, TN (as determined by NOAA's publication). An additional assumption that Mount Vernon was located midway between the Wabash river mouth and Evansville, IN was made. The distance between a port not in the Ohio River system and one in it was calculated as the sum of the distance between the former port and Memphis and the distance between Memphis and the latter port.

Intermediate information used for calculating transport to foreign markets

This section documents the tables referenced when describing the adjustments made to determine the transportation routes to foreign markets. Table B.15 shows the geographical coordinates of the seven border points for the characterization of pipeline and rail transportation to Mexico.

Table B.16 shows the representative ports for each coast of the U.S. and for foreign markets which were used to determine the cross-ocean distances from the U.S. to foreign markets.

Table B.15: U.S.-Mexico Border Points used in determining pipeline and rail transportation routes to Mexico.

Border Point	Latitude	Longitude	Straight-Line Distance to Mexico City (miles)
B0	32.611	-116.229	1,389.31
B1	31.642	-112.044	1,158.92
B2	31.792	-107.626	994.44
B3	30.573	-104.886	834.40
B4	29.765	-101.536	719.41
B5	26.069	-97.945	455.65

Table B.16: Representative ports of U.S. and Foreign Coasts.

Location	Representative Port
East Coast	New York
West Coast	Long Beach
Great Lakes	Chicago
Gulf Coast	Houston
OECD Asia	Nagoya
OECD Europe	Rotterdam

Appendix C

Capstone Paper: A Literature Survey of Portfolio Optimization

Introduction

We document, through a survey of the literature, the uses and limitations of optimization in the construction of investment portfolios. The framework of mean-variance optimization was introduced by Markowitz [122] in 1952. Namely, given a set of securities $i \in \{1, \ldots, N\}$, their expected returns $\bar{r}_i$, and the covariance matrix of future returns $[\sigma_{ij}]$ (where $\sigma_{ii} \equiv \sigma_i^2$), we seek to determine weights w_i that maximize the portfolio's expected return, given an upper limit on its variance, as shown in Formulation C.1, or equivalently, minimize its variance, given a lower limit on its expected return, as shown in Formulation C.2.

$$\begin{aligned} \underset{\mathbf{w}}{\text{maximize}} \quad & \sum_{i=1}^{N} \bar{r}_i w_i \\ \text{subject to} \quad & \sum_{i=1}^{N} \sum_{j=1}^{N} \sigma_{ij} w_i w_j \leq \gamma, \\ & \sum_{i=1}^{N} w_i = b. \end{aligned} \tag{C.1}$$

$$\begin{aligned} \underset{\mathbf{w}}{\text{minimize}} \quad & \sum_{i=1}^{N} \sum_{j=1}^{N} \sigma_{ij} w_i w_j \\ \text{subject to} \quad & \sum_{i=1}^{N} \bar{r}_i w_i \geq \lambda, \\ & \sum_{i=1}^{N} w_i = b. \end{aligned} \tag{C.2}$$

Budget constraints, where the sum of all security weights must equal b, are typically included in the formulation. For the most common cases, $b = 1$ for the unlevered portfolio, greater or less than 1 for a levered or de-levered portfolio, respectively, and 0 for the zero-investment portfolio. The tuning parameters γ or λ are specified by the investment manager as a measure of the level of his or her risk aversion. An efficient frontier of optimal portfolios can then be traced out on the return-risk axes, giving an exact representation of the trade-off between risk and return.

This mean-variance optimization framework is both a simple and powerful way to approach construction of investment portfolios. Yet, portfolios constructed through this approach are usually far from optimal when implemented in real life. One key issue is the large estimation error which complicates an accurate determination of the parameters which go into the optimizer. In particular, expected returns are in practice not straightforward to determine. Yet, this optimization framework is known

to be highly sensitive to the inputs of expected returns, heavily skewing weights towards securities with the highest expected returns. Clearly, if *ex-post* returns do not match *ex-ante* expected returns, we would expect the "optimal" portfolio to perform poorly. It was with this observation that Michaud, perhaps in jest, referred to mean-variance optimizers as "error maximizers" [123].

Besides issues of uncertainty or estimation errors, the standard mean-variance framework does not take into account other considerations such as liquidity, transaction costs, trading restrictions and alternative measures of risk. As such, extensions to the framework which address such issues have been developed. This paper explores such extensions of the standard mean-variance framework, with the purpose of serving as a compact reference for practitioners. Finally, we return to the issue of determining meaningful model inputs.

Practical considerations

Transaction costs

In reality, transaction costs are incurred whenever a portfolio is formed or rebalanced, and should be taken into account. In general, they are assumed to be separable, that is, the total transaction costs $\phi(\mathbf{w})$ can be expressed as the sum of transaction costs associated with each security traded.

$$\phi(\mathbf{w}) = \sum_{i=1}^{N} \phi_i(w_i),$$

where $\phi_i(w_i)$ is the transaction cost function for security i. The transaction cost function can be included in the objective or in the constraints.

Linear transaction costs

As documented in Lobo *et al.* [124], piecewise linear transaction costs are easily handled since the constraints they generate are linear. A typical example of piecewise linear costs is:

$$\phi_i(w_i) = \begin{cases} \alpha_i^+ w_i, & \text{if } w_i \geq 0, \\ -\alpha_i^- w_i, & \text{otherwise.} \end{cases}$$

where $\alpha_i^+, \alpha_i^- \geq 0$. This can be better represented by introducing new variables $w_i^+, w_i^- \geq 0$ such that

$$w_i \equiv w_i^+ - w_i^-,$$

resulting in

$$\phi_i = \alpha_i^+ w_i^+ + \alpha_i^- w_i^-.$$

Alternatively, the epigraph reformulation could be used. In general, any piecewise linear convex transaction cost function could be handled with similar techniques.

Fixed or concave transaction costs

Fixed or concave transaction costs also appear in practice. However, incorporating them into the formulation introduces nonconvexities in the objective function or constraints, which significantly increases the difficulty of solution. Fixed costs are costs which occur conditioned on the event that a trade is placed, independent of of the size of the trade. Concave costs occur when transaction costs per share decreases as the number of shares traded increases.

As documented in Speranza [125], fixed costs can be modeled using binary vari-

ables. Namely, for each security i, we introduce binary variables z_i such that

$$\phi_i(w_i, z_i) = \alpha_i |w_i| + \beta_i z_i,$$

where β_i is the fixed costs that are incurred for every non-zero trade made for security i.

The relationship between w_i and z_i is handled by introducing the constraints

$$|w_i| \leq M_i z_i, \quad i = 1, \ldots, n,$$

where M_i is the upper bound and $-M_i$ the lower bound on w_i if security i is traded. Auxiliary variables ξ_i would then be introduced to represent $|w_i|$, accompanied with the constraints:

$$\begin{aligned} w_i &\leq \xi_i, \quad i = 1, \ldots, n, \\ -w_i &\leq \xi_i, \quad i = 1, \ldots, n. \end{aligned}$$

This framework can be extended to accommodate situations where the fixed and linear costs are different depending on whether one goes long or short the security.

For concave costs, we have that $\phi_i(w_i)$ is a continuous concave function of w_i. Concave costs could be handled by generating convex relaxations of the cost function, then solving the system to global optimality with a branch-and-bound approach [126]. Alternatively, one could linearize the concave cost function into a continuous piecewise linear concave function, and apply integer programming methods to solve the problem.

Given a generic piecewise linear concave function $f(x)$, we denote its breakpoints

by $a_1 < a_2 < \ldots < a_k$. Any $x \in [a_1, a_k]$ can be expressed in the form:

$$x = \sum_{i=1}^{k} \lambda_i a_i,$$

and $f(x)$ can be expressed as

$$f(x) = \sum_{i=1}^{k} \lambda_i f(a_i),$$

where $\lambda_1, \ldots, \lambda_k \geq 0$ and $\sum_{i=1}^{k} \lambda_i = 1$.

These representations of $f(x)$ for a given x are not unique in terms of the λ_i's unless we impose the additional constraint that at most two consecutive coefficients λ_i can be zero. This condition can be imposed through the introduction of binary variables δ_i, which are equal to 1 only if $a_i \leq x \leq a_{i+1}$, and 0 otherwise.

Then, we can represent the problem of minimizing the piecewise linear concave

function with the following formulation:

$$
\begin{aligned}
\underset{\boldsymbol{\lambda},\boldsymbol{\delta}}{\text{minimize}} \quad & \sum_{i=1}^{k} \lambda_i f(a_i) \\
\text{subject to} \quad & \sum_{i=1}^{k} \lambda_i = 1, \\
& \lambda_1 \leq \delta_1, \\
& \lambda_i \leq \delta_{i-1} + \delta_i, \ , i = 2, \ldots, k-1, \\
& \lambda_k \leq \delta_{k-1}, \\
& \sum_{i=1}^{k-1} \delta_i = 1, \\
& \lambda_i \geq 0, \\
& \delta_i \in \{0, 1\}.
\end{aligned}
$$

Minimum transaction units

Analogous to the issue of fixed costs is the situation of minimum transaction units. That is if security i is traded, at least a minimum of m_i units must be transacted, or multiples of m_i units must be transacted. This can be addressed by the introduction of binary, or in general, integer variables z_i, with the constraints

$$m_i z_i = \frac{W}{p_i} |w_i|, \quad i = 1, \ldots, n,$$

where p_i is the current price of security i, W is the total wealth of the portfolio, and z_i is binary for the former situation and integer for the latter. As with the example in fixed costs, auxiliary variables are used to linearize the constraints.

Sizing

Setting bounds on the security weights or number of securities may help to restrict the "error maximization" tendencies of a naive implementation of mean-variance optimization. The simplest case involves limiting the absolute weights of each security to an upper bound M_i:

$$|w_i| \leq M_i, \quad i = 1, \ldots, n.$$

We could also have the case where we could like a lower bound on the number of securities to achieve diversification, where we introduce binary variables z_i such that:

$$\sum_i z_i \geq N,$$
$$m_i z_i \leq |w_i| \leq M_i z_i, \quad i = 1, \ldots, n.$$

Considering that this integer formulation significantly increases the difficulty of solution, we could look at an alternative approach that only requires continuous variables. As documented in Lobo *et al.* [124], consider the case that we require no more than a weight γ be invested in fewer than r securities. If we denote $\mathbf{f}_{[i]}$ as the ith largest component of the vector $\mathbf{f}$, this constraint can be expressed as

$$\sum_{i=1}^{r} \mathbf{w}_{[i]} \leq \gamma.$$

It is shown that this constraint can be reformulated into the following system, re-

quiring only $1 + 2n$ linear inequalities:

$$\begin{aligned}
&\gamma \geq rt + \sum_{i=1}^{n} y_i, \\
&t + y_i \geq w_i, \quad i = 1, \ldots, n, \\
&y_i \geq 0, \quad i = 1, \ldots, n,
\end{aligned}$$

where y_i and t are auxiliary variables.

To see why the reformulation is valid, note that $\sum_{i=1}^{r} \mathbf{w}_{[i]}$ is the solution of the linear program

$$\begin{aligned}
\underset{\mathbf{x}}{\text{maximize}} \quad & \sum_{i=1}^{n} w_i x_i \\
\text{subject to} \quad & 0 \leq x_i \leq 1, \quad i = 1, \ldots, n, \\
& \sum_{i=1}^{n} x_i = r,
\end{aligned}$$

where the variables are x_i. Assuming the problem is feasible and bounded, its optimal value is also the same as that of its dual:

$$\begin{aligned}
\underset{t,\mathbf{y}}{\text{minimize}} \quad & rt + \sum_{i=1}^{n} y_i \\
\text{subject to} \quad & t + y_i \geq w_i, \quad i = 1, \ldots, n, \\
& y_i \geq 0, \quad i = 1, \ldots, n.
\end{aligned}$$

The optimal value of the dual is $\sum_{i=1}^{r} \mathbf{w}_{[i]}$, and is less than γ if and only if there exists a feasible solution $t, \mathbf{y}$ with $rt + \sum_{i=1}^{n} y_i \leq \gamma$.

There are extensions to this type of constraint, for example, if we divide the n securities into m classes, and require that no more than a fraction γ of the portfolio

be invested in fewer than R of these classes.

Institutional investment managers often also face constraints on short selling. Individual bounds s_i on short selling can naturally be set:

$$-w_i \leq s_i, \quad i = 1, \ldots, n.$$

The total amount of short selling can also be bounded by S:

$$\sum_{i=1}^{n} \max\{-w_i, 0\} \leq S.$$

The max function is naturally replaced by using auxiliary variables with linear constraints, as had been documented earlier.

Finally, the total weight of short positions can also be expressed as a fraction γ of the total weight of long positions:

$$\sum_{i=1}^{n} \max\{-w_i, 0\} \leq \gamma \sum_{i=1}^{n} \max\{w_i, 0\}.$$

Risk measures

The variance of portfolio returns has served as the classical measure of risk in the mean-variance framework. However, in a more general scheme, a risk measure $R(\mathbf{w})$ is minimized subject to a minimum required expected portfolio return. Here, we document alternative risk measures found in the literature. In the sections following below, we denote the random variable r_i as the return of security i in the following period.

Mean absolute deviation

The relative difficulty in solving a quadratic program in the early 1990s led to the proposal of mean-absolute deviation (MAD) as a risk measure by Konno and Yamazaki [127], where a linear program formulation was available. MAD is defined as:

$$R(\mathbf{w}) = \mathrm{E}\left[\left|\sum_{i=1}^{n} r_i w_i - \mathrm{E}\left[\sum_{i=1}^{n} r_i w_i\right]\right|\right].$$

Under conditions where returns are multivariate normally distributed, the minimum MAD portfolio coincides with the minimum variance portfolio. Admittedly, these assumptions do not hold in practice.

In order to reformulate the problem into a linear program, a sample of returns data would have to be available. This could be generated by the user, or obtained from historical data. Denoting each observation of returns as $(r_{1t}, \ldots, r_{nt})$, $t = 1, \ldots, T$, we can rewrite the MAD measure as:

$$R(\mathbf{w}) = \frac{1}{T}\sum_{t=1}^{T}\left|\sum_{i=1}^{n} a_{it} w_i\right|,$$

where $a_{it} \equiv (r_{it} - \bar{r}_i)$ and $\bar{r}_i \equiv \frac{1}{T}\sum_{t=1}^{T} r_{it}$.

Introducing auxiliary variables v_t and u_t, minimizing $R(\mathbf{w})$ is equivalent to the linear program [128][129]:

$$\begin{aligned}
\underset{\mathbf{w},\mathbf{v},\mathbf{u}}{\text{minimize}} \quad & \sum_{t=1}^{T}(v_t + u_t) \\
\text{subject to} \quad & v_t - u_t - \sum_{i=1}^{n} a_{it} w_i = 0, \quad t = 1, \cdots, T, \\
& v_t, u_t \geq 0, \quad t = 1, \cdots, T.
\end{aligned}$$

As evident in the reformulation, we introduce $2T$ new variables and T new constraints. Naturally, one might want to explore how the difficulty of solution increases as we increase the number of observations included in our sample. In addition, little comment is made about how to choose the sample in order to achieve a useful solution. These are very important practical issues to consider. In addition, the necessity of replacing variance with MAD might be diminished today with much improved performance of modern-day second-order cone program solvers.

Maximum loss

We can alternatively define the maximum loss of the portfolio as the appropriate risk measure to minimize. Similar in spirit to MAD, this leads to a linear program formulation [130]. Given observations of returns $(r_{1t}, \ldots, r_{nt}),\ t = 1, \ldots, T$, let $R(\mathbf{w})$ be the maximum loss of the portfolio over the time period, defined as:

$$R(\mathbf{w}) = \max_t \Big\{ -\sum_{i=1}^{n} r_{it} w_i \Big\}.$$

Minimizing $R(\mathbf{w})$ involves the introduction of a single auxiliary variable, z, in the linear program:

$$\begin{aligned} &\text{minimize} \quad z \\ &\text{subject to} \quad -\sum_{i=1}^{n} r_{it} w_i - z \leq 0, \quad t = 1, \cdots, T. \end{aligned}$$

Skewness

One drawback of using variance as the risk measure is that it neglects the consideration of skewness in the portfolio's return distribution. All else being equal, investors

would naturally prefer positive skewness over negative skewness. Konno *et al.* extended their work on MAD to account for skewness as a risk measure [129]. They proposed the negative of the lower semi-third moment of the portfolio's return as the risk function:

$$R(\mathbf{w}) = \mathrm{E}\left[g\left(\sum_{i=1}^{n} r_i w_i - \mathrm{E}\left[\sum_{i=1}^{n} r_i w_i\right]\right)\right],$$

where

$$g(u) = \begin{cases} 0, & \text{if } u \geq 0, \\ -u^3, & \text{otherwise.} \end{cases}$$

Although $g(\cdot)$ is convex, it might introduce computational difficulties when the size of the problem is large. As such, the authors proposed replacing it with a piecewise linear convex function $G(\cdot)$, given as

$$G(u) = |u - \rho_1|_- + \alpha\,|u - \rho_2|_-\,,$$

where $\alpha > 0, \rho_1, \rho_2 < 0$ are tuning parameters and $|x|_- \equiv \max\{0, -x\}$.

Similar to the MAD method, we generate a sample of observations to evaluate the expectation terms. In addition, we ignore the mean term in the skewness. This leads to a rewrite of the skewness risk measure as:

$$R(\mathbf{w}) = \frac{1}{T-1}\left(\sum_{t=1}^{T}\left|\sum_{i=1}^{n} r_{it} w_i - \rho_1\right|_- + \alpha \sum_{t=1}^{T}\left|\sum_{i=1}^{n} r_{it} w_i - \rho_2\right|_-\right).$$

Introducing auxiliary variables u_t and v_t, minimizing $R(\mathbf{w})$ is equivalent to the

linear program:

$$\begin{aligned}
\underset{\mathbf{w},\mathbf{u},\mathbf{v}}{\text{minimize}} \quad & \frac{1}{T-1}\left(\sum_{t=1}^{T} u_t + \alpha \sum_{t=1}^{T} v_t\right) \\
\text{subject to} \quad & u_t + \sum_{i=1}^{n} r_{it} w_i \geq \rho_1, \quad t = 1, \cdots, T, \\
& v_t + \sum_{i=1}^{n} r_{it} w_i \geq \rho_2, \quad t = 1, \cdots, T, \\
& v_t, u_t \geq 0, \quad t = 1, \cdots, T.
\end{aligned}$$

$2T$ new variables and $2T$ new constraints have been introduced in the reformulation. As with the MAD reformulation, careful attention needs to be paid on the choice of the sample.

Value-at-risk

An alternative measure of risk is value-at-risk (VaR). Given a confidence level $1-\alpha$, VaR is the minimum value of loss that occurs no more than $\alpha * 100\%$ of the time. For example, if $\alpha = 0.05$, a VaR of -0.10 indicates that in the future realizations of portfolio returns, 5% of them will have returns less than or equal to -10%, while 95% of them will have returns greater than -10%. Defining $R(\mathbf{w})$ as the negative of VaR, which we intend to minimize, we formalize its definition as:

$$R(\mathbf{w}) = -Q_\alpha\left(\sum_{i=1}^{n} r_i w_i\right),$$

where $Q_\alpha(\cdot)$ is the α-quantile of return:

$$Q_\alpha(Z) = \inf\{z : F_Z(z) > \alpha\},$$

where $F_Z(\cdot)$ is the cdf of the random variable Z.

Studies have been done to evaluate $R(\mathbf{w})$ given a sample of return observations and working with the sample cdf [131]. However, $R(\mathbf{w})$ turns out to be non-differentiable and non-convex in $\mathbf{w}$. Following which, smoothing techniques can be used to address the non-differentiability.

An alternate way to consider VaR is not to work with minimizing $R(\mathbf{w})$ directly, but instead expressing the VaR requirement in a constraint, and perform the portfolio optimization for different values of VaR, given a certain confidence level $1-\alpha$ [124]. With this method, we could include one or multiple VaR constraints in a single optimization model. A tractable expression of this constraint makes the assumption that portfolio returns are Gaussian, i.e., $\sum_{i=1}^{n} r_i w_i \sim \mathcal{N}(\tilde{\mu}, \tilde{\sigma}^2)$, where $\tilde{\mu}$ and $\tilde{\sigma}^2$ are the mean and variance of the portfolio's returns, respectively.

Following which, we can express the VaR requirement with the following constraint:

$$\Pr\left(\sum_{i=1}^{n} r_i w_i \leq VaR\right) \leq \alpha.$$

Under the assumption of normality, we have:

$$\Phi\left(\frac{VaR - \tilde{\mu}}{\tilde{\sigma}}\right) \leq \alpha,$$

where $\Phi(\cdot)$ is the cdf of the standard normal distribution. Using the identity $\Phi^{-1}(\alpha) = -\Phi^{-1}(1-\alpha)$, we have:

$$\tilde{\mu} - VaR \geq \Phi^{-1}(1-\alpha)\tilde{\sigma}.$$

Substituting $\tilde{\mu} = \sum_{i=1}^{n} \bar{r}_i w_i$ and $\tilde{\sigma}^2 = \sum_{i=1}^{N} \sum_{j=1}^{N} \sigma_{ij} w_i w_j$, we have:

$$\Phi^{-1}(1-\alpha)\sqrt{\sum_{i=1}^{N}\sum_{j=1}^{N}\sigma_{ij}w_i w_j} \leq \sum_{i=1}^{n}\bar{r}_i w_i - VaR.$$

Provided that $\Phi^{-1}(1-\alpha) \geq 0$, and therefore $\alpha \leq 0.5$, this is a second-order cone constraint, which is easily handled. Since investment managers are often interested in VaR values for which α is small, this constraint is useful when constructing portfolios.

Conditional value-at-risk

Related to VaR is conditional value-at-risk (CVaR). A CVaR at confidence level $1-\alpha$ is the expected losses of all return realizations which are at or below the corresponding VaR. In contrast to VaR, CVaR has the ideal properties of being a coherent risk measure and its minimization can be formulated as a linear program, as first introduced by Uryasev *et al.* [132][133]. Further, by its definition, the VaR is never more than the CVaR, and so portfolios with low CVaR necessarily have low VaRs too. Using the definition of the α-quantile of return, $Q_\alpha(\cdot)$, as above, we define the CVaR risk measure $R(\mathbf{w})$ as:

$$R(\mathbf{w}) = \frac{1}{\alpha}\int_{\{\mathbf{r}:\sum_{i=1}^{n} r_i w_i \leq Q_\alpha(\sum_{i=1}^{n} r_i w_i)\}} \left(-\sum_{i=1}^{n} r_i w_i\right) p(\mathbf{r}) d\mathbf{r},$$

where $p(\mathbf{r})$ is the pdf of the return vector $\mathbf{r}$. It can be seen that this expression indeed gives us our definition of CVaR: the integral is over all realizations that give a total loss greater than our α-quantile of return, weighted by their respective probabilities, and the normalizing factor $1/\alpha$ correctly rescales the unconditional probabilities to attain the required conditional expectation of losses.

Minimizing $R(\mathbf{w})$ as defined would be a little tricky, since we would have to also deal with the inner minimization $Q_\alpha(\cdot)$. However, the key finding by Uryasev *et al.* was that one could more easily directly work with the risk measure $R(\mathbf{w}, z)$, defined as:

$$R(\mathbf{w}, z) = z + \frac{1}{\alpha} \int_{\mathbf{r} \in \mathbb{R}^n} \left[-\sum_{i=1}^{n} r_i w_i - z \right]^+ p(\mathbf{r}) d\mathbf{r},$$

where $[x]^+ \equiv \max\{0, x\}$.

It was shown that the joint minimization of $R(\mathbf{w}, z)$ leads to the same optimal value as the minimization of $R(\mathbf{w})$, and further, in the case where the set $Q_\alpha(\sum_{i=1}^n r_i w_i^*)$ is a singleton (which is typically the case), an optimal solution $(\mathbf{w}^*, z^*)$ from the joint minimization of $R(\mathbf{w}, z)$ will yield a $\mathbf{w}^*$ which minimizes $R(\mathbf{w})$.

Again, similar to some of the previous risk measures, we generate a sample of observations that can help to approximate the integral. Denoting each observation of returns as $(r_{1t}, \ldots, r_{nt}),\ t = 1, \ldots, T$, we can rewrite $R(\mathbf{w}, z)$ as:

$$R(\mathbf{w}, z) = z + \frac{1}{\alpha} \frac{1}{T} \sum_{t=1}^{T} \left[-\sum_{i=1}^{n} r_{it} w_i - z \right]^+.$$

By using auxiliary variables u_t to replace the max function, this can be expressed as a linear program, where minimizing $R(\mathbf{w}, z)$ is equivalent to:

$$\begin{aligned} \underset{z, \mathbf{w}, \mathbf{u}}{\text{minimize}} \quad & z + \frac{1}{\alpha} \frac{1}{T} \sum_{t=1}^{T} u_t \\ \text{subject to} \quad & u_t \geq -\sum_{i=1}^{n} r_{it} w_i - z, \quad t = 1, \cdots, T, \\ & u_t \geq 0, \quad t = 1, \cdots, T. \end{aligned}$$

To more finely shape the profit/loss distribution, we can introduce use multiple

CVaRs in the objective (or constraints) by using a weighted sum of CVaRs at different α's [134]. Standard cross-validation methods can then be used to select the appropriate weights. Finally, the use of CVaR minimization would be much more relevant in the cases where the return distribution is highly non-normal, exhibits a negative skew, and if the portfolio manager's main priority is to protect against heavy losses, such as in a credit portfolio [135].

Drawdown

While the above risk measures are concerned with losses occurring at a point in time, drawdown concerns itself with the specific accumulative sequence of losses over time [136]. Given a sequence of return realizations over time $\mathbf{r}_t \equiv (r_{1t}, \ldots, r_{nt}),\ t = 1, \ldots, T$, and a static portfolio allocation $\mathbf{w}$, we define $v(\mathbf{w}, t) \equiv \sum_{\tau=1}^{t} \mathbf{r}_\tau \cdot \mathbf{w}$ as the uncompounded portfolio value at time t. The drawdown function $d(\mathbf{w}, t)$ at time t is defined as the difference between the maximum of $v(\mathbf{w}, t)$ over all preceding observations and its current value, that is:

$$d(\mathbf{w}, t) = \max_{0 \leq \tau \leq t} \{v(\mathbf{w}, t)\} - v(\mathbf{w}, t).$$

Similar in spirit to CVaR, conditional drawdown (CDD) at a confidence level $1-\alpha$ is defined as the average drawdown of the worst $\alpha * 100\%$ of drawdowns in the sample. Specifically, the CDD risk function $R_{CDD}(\mathbf{w})$ is defined as:

$$R_{CDD}(\mathbf{w}) = \frac{1}{\alpha T} \sum_{t \in \Omega} d(\mathbf{w}, t) \quad \Omega = \{t \in \{1, \cdots, T\} : d(\mathbf{w}, t) \geq z(\mathbf{w}, t)\},$$

where $z(\mathbf{w}, t)$ is defined as the threshold such that $\alpha * 100\%$ of drawdowns exceed this threshold.

Similar to the method of implementation in CVaR, it was noted that we could instead define the function $R_{CDD}(\mathbf{w}, z)$ and perform the minimization jointly over $(\mathbf{w}, z)$, in order to bypass the inner minimization implied by $z(\mathbf{w}, t)$, where:

$$R_{CDD}(\mathbf{w}, z) = z + \frac{1}{\alpha T} \sum_{t=1}^{T} \left[d(\mathbf{w}, t) - z \right]^{+},$$

where $[x]^{+} \equiv \max\{0, x\}$.

If we consider the limits of α, we also get two special cases which might serve as alternative risk measures. If $\alpha \to 0$, we attain the maximum drawdown, $R_{MDD}(\mathbf{w})$, defined as:

$$R_{MDD}(\mathbf{w}) = \max_{1 \leq t \leq T} \{ d(\mathbf{w}, t) \}.$$

If $\alpha \to 1$, we attain the average drawdown, $R_{ADD}(\mathbf{w})$, defined as:

$$R_{ADD}(\mathbf{w}) = \frac{1}{T} \sum_{t=1}^{T} d(\mathbf{w}, t).$$

The problems defined by minimizing $R_{CDD}(\mathbf{w}, z)$, $R_{MDD}(\mathbf{w})$ and $R_{ADD}(\mathbf{w})$ can be expressed as linear programs, by introducing auxiliary variables to represent the max function in $d(\mathbf{w}, t)$, as well as the outer max functions in $R_{CDD}(\mathbf{w}, z)$ and $R_{MDD}(\mathbf{w})$.

A naive LP transformation of minimizing $R_{CDD}(\mathbf{w}, z)$ would introduce $O(T^2)$ additional constraints, since for each $d(\mathbf{w}, t)$ term we have a max function over t values, and we have $O(T)$ $d(\mathbf{w}, t)$ terms in total. However, it is possible to arrive at a formulation which only adds $O(T)$ constraints through the elimination of intermediate variables [137]:

$$
\begin{aligned}
\underset{z,\mathbf{w},\mathbf{u},\mathbf{v}}{\text{minimize}} \quad & z + \frac{1}{\alpha}\frac{1}{T}\sum_{t=1}^{T} u_t \\
\text{subject to} \quad & u_t \geq v_t - z, \quad t = 1, \cdots, T, \\
& v_t \geq v_{t-1} - z, \quad t = 2, \cdots, T, \\
& v_0 = 0, \\
& u_t, v_t \geq 0, \quad t = 1, \cdots, T.
\end{aligned}
$$

Varying α in $[0, 1]$ allows one to fine-tune required performance of the portfolio with respect to drawdowns. One major drawback of using a single drawdown risk function is that the solution is very dependent on the exact sequence of realizations fed into the optimizer. As such, the solution is not expected to be robust once implemented out-of-sample. One might mitigate this by generating S different sequences of return realizations, and then either add S individual drawdown constraints, or add a single constraint which limits the average CDD among these different sequences. The former would lead to a total of $O(ST)$ constraints added, whereas the latter preserves the number of constraints at $O(T)$. However, the former would lead to more conservative portfolios, whereas the latter might lead to solutions that have extreme drawdowns clustered in a few scenarios. Naturally, one has to consider the trade-offs between computational tractability and the usefulness of the solution in practice.

Robust optimization

In contrast to many of the above approaches, which generally apply stochastic optimization techniques, requiring the generation of many scenarios of return realizations as model inputs, robust optimization is an alternative approach to optimization un-

der uncertainty that is deterministic and set-based. That is, it seeks to construct a solution that is feasible for any realization of uncertainty in a given set of possible parameter values. This set is termed as an 'uncertainty set'. Intersections of uncertainty sets for a given parameter, or uncertainty sets defined separately for different parameters also fall under the general setting of robust optimization.

A potential advantage of using robust optimization over stochastic optimization methods is the preservation of computational tractability while obtaining solutions that are fairly robust to uncertainty [138]. However, some care is needed in the choice of uncertainty sets in order to ensure that tractability is maintained.

We consider the robust minimum variance problem, which is the robust version of Formulation C.2:

$$\begin{aligned} \underset{\mathbf{w}}{\text{minimize}} \quad & \max_{\boldsymbol{\Sigma} \in \mathcal{S}} \quad \mathbf{w}^\top \boldsymbol{\Sigma} \mathbf{w} \\ \text{subject to} \quad & \min_{\boldsymbol{\mu} \in \mathcal{M}} \quad \boldsymbol{\mu}^\top \mathbf{w} \geq \lambda, \\ & \mathbf{1}^\top \mathbf{w} = b, \end{aligned} \tag{C.3}$$

where $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are the mean return vector and covariance matrix of the securities in the universe respectively, and $\mathcal{M}$ and $\mathcal{S}$ are their corresponding uncertainty sets.

The disadvantage of Formulation C.3 is that it might be overly conservative. Further, one would have to be cautious in determining the appropriate structure of $\mathcal{S}$ in order to arrive at a tractable formulation. We might consider instead, a less conservative formulation, where we treat $\mathcal{S}$ as a singleton and only consider uncertainty in $\boldsymbol{\mu}$. Thus, we define the *linear* robust minimum variance problem,

where the uncertainty sets are only defined on the coefficients of the linear constraint:

$$\begin{aligned} \underset{\mathbf{w}}{\text{minimize}} \quad & \mathbf{w}^\top \boldsymbol{\Sigma} \mathbf{w} \\ \text{subject to} \quad & \min_{\boldsymbol{\mu} \in \mathcal{M}} \quad \boldsymbol{\mu}^\top \mathbf{w} \geq \lambda, \\ & \mathbf{1}^\top \mathbf{w} = b. \end{aligned} \tag{C.4}$$

For Formulation C.4, well-known results that lead to computationally tractable and easily expressed robust counterparts are available [138]. In particular, these include cases where the uncertainty set $\mathcal{M}$ is a polyhedron, or can be described as a ball with a general norm.

We can also express Formulation C.4 equivalently as:

$$\begin{aligned} \underset{\mathbf{w}}{\text{minimize}} \quad & \mathbf{w}^\top \boldsymbol{\Sigma} \mathbf{w} \\ \text{subject to} \quad & -\boldsymbol{\mu}^\top \mathbf{w} \leq -\lambda \quad \forall \boldsymbol{\mu} \in \mathcal{M}, \\ & \mathbf{1}^\top \mathbf{w} = b. \end{aligned} \tag{C.5}$$

In general, for variable $\mathbf{x}$, we have the set-up:

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & \mathbf{a}_i^\top \mathbf{x} \leq b_i \quad \forall \mathbf{a}_i \in \mathcal{U}_i, \quad i = 1, \ldots, m, \\ & \mathbf{x} \in \mathcal{X}, \end{aligned} \tag{C.6}$$

or equivalently:

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & \max_{\mathbf{a}_i \in \mathcal{U}_i} \quad \mathbf{a}_i^\top \mathbf{x} \leq b_i, \quad i = 1, \ldots, m, \\ & \mathbf{x} \in \mathcal{X}, \end{aligned} \tag{C.7}$$

where $f(\mathbf{x})$ is a potentially non-linear objective function and the set $\mathcal{X}$ define additional constraints on $\mathbf{x}$.

Given the general form above, we present the results of the robust counterpart for different assumptions on the structure of $\mathcal{U}_i$:

Polyhedral uncertainty

Given polyhedral uncertainty sets of the form:

$$\mathcal{U}_i = \{\mathbf{a}_i | \mathbf{D}_i \mathbf{a}_i \leq \mathbf{d}_i\},$$

we consider the subproblem of maximizing $\mathbf{a}_i$:

$$\begin{aligned} \underset{\mathbf{a}_i}{\text{maximize}} \quad & \mathbf{a}_i^\top \mathbf{x} \\ \text{subject to} \quad & \mathbf{D}_i \mathbf{a}_i \leq \mathbf{d}_i, \end{aligned}$$

and its dual:

$$\begin{aligned} \underset{\mathbf{p}_i}{\text{minimize}} \quad & \mathbf{p}_i^\top \mathbf{d}_i \\ \text{subject to} \quad & \mathbf{p}_i^\top \mathbf{D}_i = \mathbf{x}, \\ & \mathbf{p}_i \geq 0. \end{aligned}$$

Thus the robust counterpart becomes [138]:

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & \mathbf{p}_i^\top \mathbf{d}_i \leq b_i, \quad i = 1, \ldots, m, \\ & \mathbf{p}_i^\top \mathbf{D}_i = \mathbf{x}, \quad i = 1, \ldots, m, \\ & \mathbf{p}_i \geq 0, \quad i = 1, \ldots, m, \\ & \mathbf{x} \in \mathcal{X}. \end{aligned}$$

Note that in contrast to the original, non-robust problem which contained m constraints, the robust counterpart contains $3m$ constraints and introduces m additional variables.

Ellipsoidal uncertainty

Given ellipsoidal uncertainty sets of the form:

$$\mathcal{U}_i = \{\mathbf{a}_i | \mathbf{a}_i = \bar{\mathbf{a}}_i + \boldsymbol{\Delta}_i^\top \mathbf{u}_i, \quad \|\mathbf{u}_i\|_2 \leq \rho\}, \quad \boldsymbol{\Delta}_i : k_i \times n, \quad \mathbf{u}_i : k_i \times 1,$$

where $\bar{\mathbf{a}}_i$ is a nominal value, the robust counterpart is [138]:

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & \bar{\mathbf{a}}_i^\top \mathbf{x} + \rho \|\boldsymbol{\Delta}_i \mathbf{x}\|_2 \leq b_i, \quad i = 1, \ldots, m, \\ & \mathbf{x} \in \mathcal{X}. \end{aligned}$$

We observe that the robust counterpart is an SOCP, which can efficiently be solved.

Using ellipsoidal uncertainty sets are in general useful when the underlying data has covariance structure, which is the case for security returns. Given a vector

of nominal expected returns $\boldsymbol{\mu}_0$ and the covariance matrix of returns $\boldsymbol{\Sigma}$, we can construct an ellipsoidal uncertainty set for $\boldsymbol{\mu}$ of the form:

$$\mathcal{U} = \{\boldsymbol{\mu} | \boldsymbol{\mu} = \boldsymbol{\mu}_0 + \boldsymbol{\Sigma}^{-\frac{1}{2}}\mathbf{u}, \ \ \|\mathbf{u}\|_2 \leq \rho\},$$

where ρ is determined from the appropriate critical value for the confidence interval corresponding to a given size of the uncertainty set. As we shall discuss in detail later, $\boldsymbol{\mu}_0$ is best represented through the influence of a practitioner's forward-looking views, and $\boldsymbol{\Sigma}$ should be estimated using a factor model or shrinkage techniques.

General norm uncertainty

We can generalize the above to consider general norm uncertainty sets of the form:

$$\mathcal{U}_i = \{\mathbf{a}_i | \mathbf{a}_i = \bar{\mathbf{a}}_i + \boldsymbol{\Delta}_i^\top \mathbf{u}_i, \ \ \|\mathbf{u}_i\|_p \leq \rho\}, \quad \boldsymbol{\Delta}_i : k_i \times n, \quad \mathbf{u}_i : k_i \times 1,$$

where the L_p norm $\|x\|_p \equiv \left(\sum_{j=1}^n |x_j|^p\right)^{1/p}$.

The robust counterpart in this general case is [138]:

$$\begin{array}{ll} \underset{\mathbf{x}}{\text{minimize}} & f(\mathbf{x}) \\ \text{subject to} & \bar{\mathbf{a}}_i^\top \mathbf{x} + \rho \|\boldsymbol{\Delta}_i \mathbf{x}\|^* \leq b_i, \quad i = 1, \ldots, m, \\ & \mathbf{x} \in \mathcal{X}, \end{array}$$

where $\|s\|^*$ is the dual of the L_p norm, with $\|s\|^* = \|s\|_q$ where $q = 1 + \frac{1}{p-1}$.

We are particularly interested in the cases of the L_1, L_2 and L_∞ norms. Note that the dual norm of the L_1 norm is the L_∞ norm, and vice versa. In contrast, the dual norm of the L_2 norm is L_2. Uncertainty sets with L_1 and L_∞ norms reduce

the robust counterpart to a linear optimization problem, while those with L_2 norms reduce the robust counterpart to a SOCP, as we have seen earlier.

Factor models

Factor models are commonly used to model security returns. Given n securities, it is common to find that only a small number k of underlying factors drive their movement. Specifically, a factor model is described as:

$$\mathbf{r}_t = \boldsymbol{\alpha} + \boldsymbol{\beta}^\top \mathbf{f}_t + \boldsymbol{\epsilon}_t,$$

where $\mathbf{r}_t \in \mathbb{R}^n$ is the single-period return vector of the securities, $\mathbf{f}_t \in \mathbb{R}^k$ is the single-period return vector of the factors, $\boldsymbol{\beta} \in \mathbb{R}^{k \times n}$ is the matrix of factor loadings for each security, $\boldsymbol{\alpha} \in \mathbb{R}^n$ is the vector of excess returns, and $\boldsymbol{\epsilon}_t \in \mathbb{R}^n$ is the vector of residual returns.

We assume that the factors $\mathbf{f}_t$ have mean returns $\boldsymbol{\mu}_f \in \mathbb{R}^k$ and covariance matrix $\boldsymbol{\Sigma}_f \in \mathbb{R}^{k \times k}$, while the residual returns $\boldsymbol{\epsilon}_t$ have zero mean and covariance matrix $\boldsymbol{\Sigma}_\epsilon \in \mathbb{R}^{n \times n}$.

Given this setup, the single-period expected return of the securities is $\boldsymbol{\alpha} + \boldsymbol{\beta}^\top \boldsymbol{\mu}_f$ and the covariance matrix is $\boldsymbol{\beta}^\top \boldsymbol{\Sigma}_f \boldsymbol{\beta} + \boldsymbol{\Sigma}_\epsilon$.

With a weight vector $\mathbf{w}$, the portfolio's single-period return is:

$$\mathbf{w}^\top \mathbf{r}_t = \mathbf{w}^\top \boldsymbol{\alpha} + \mathbf{w}^\top \boldsymbol{\beta}^\top \mathbf{f}_t + \mathbf{w}^\top \boldsymbol{\epsilon}_t,$$

with mean $\mathbf{w}^\top \left(\boldsymbol{\alpha} + \boldsymbol{\beta}^\top \boldsymbol{\mu}_f \right)$ and variance $\mathbf{w}^\top \left(\boldsymbol{\beta}^\top \boldsymbol{\Sigma}_f \boldsymbol{\beta} + \boldsymbol{\Sigma}_\epsilon \right) \mathbf{w}$.

A computational advantage of using factor models in mean-variance optimization

is that it reduces the number of cross-terms in the expression for the portfolio's variance [127]. That is, we introduce additional k additional variables $\bar{\mathbf{w}} = [\bar{w}_1, \ldots, \bar{w}_k] \equiv \boldsymbol{\beta}\mathbf{w}$, such that

$$\begin{aligned}\sigma_p^2 &= \mathbf{w}^\top \boldsymbol{\Sigma}_\epsilon \mathbf{w} + \bar{\mathbf{w}}^\top \boldsymbol{\Sigma}_f \bar{\mathbf{w}} \\ &= \sum_{i=1}^{n} w_i^2 \sigma_{\epsilon_i}^2 + \sum_{i=1}^{k}\sum_{j=1}^{k} \bar{w}_i \bar{w}_j \sigma_{f_{ij}},\end{aligned}$$

where σ_p^2 is the variance of the portfolio, $\sigma_{\epsilon_i}^2$ denotes the ith diagonal element of $\boldsymbol{\Sigma}_\epsilon$, $\sigma_{f_{ij}}$ is the (i,j)th entry of $\boldsymbol{\Sigma}_f$, with the additional constraints that

$$\bar{w}_j = \sum_{i=1}^{n} \beta_{ji} w_i, \quad j = 1, \ldots, k.$$

Expressing security returns in terms of factor returns also allow constraints that restrict the portfolio's overall exposure to particular factors. Note that by definition of $\bar{w}_j$ above, this is precisely the sensitivity or exposure of the portfolio to the jth factor. Thus, we could set the portfolio's target exposure (beta) to each factor, β_j^{target}, as such:

$$\beta_j^{target} - \gamma_j \leq \bar{w}_j \leq \beta_j^{target} + \gamma_j, \quad j = 1, \ldots, k,$$

where γ_j is a small number.

Resampling

Portfolio resampling is a heuristic method to determine portfolio weights which might be more resilient to estimation errors in the model parameters [139]. Given nomi-

nal values of the mean return vector $\boldsymbol{\mu}_0$ and covariance matrix $\boldsymbol{\Sigma}_0$, a sample of T observations can be generated by bootstrapping from the nominal distribution. Alternatively, we can draw bootstrap samples directly from an empirical distribution of returns.

Either way, we generate n bootstrap samples and estimate their corresponding mean return vectors and covariance matrices, $(\hat{\boldsymbol{\mu}}_1, \hat{\boldsymbol{\Sigma}}_1)$ to $(\hat{\boldsymbol{\mu}}_n, \hat{\boldsymbol{\Sigma}}_n)$. These estimates then each serve as inputs to the mean-variance optimizer. For each optimization, we choose m portfolios along the frontier by varying the risk-aversion parameter and save the corresponding allocation vectors $\mathbf{w}_{11}, \ldots, \mathbf{w}_{1m}$ to $\mathbf{w}_{n1}, \ldots, \mathbf{w}_{nm}$.

The resampled weight for a portfolio of risk level m' is given by:

$$\bar{\mathbf{w}}_{m'}^{resampled} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{w}_{im'},$$

which simply takes the average weight across all samples. The advantage of this method is that it preserves any constraints on the weights in the optimization, for example, constraints on the total weight or constraints on individual weights.

Despite the simplicity of resampling, some caveats must be seriously noted. Resampled weights are known to suffer from the problem of "optionality". This is a situation where securities with higher volatilities generate a wider distribution of expected returns through the bootstrapping procedure. The result is that there exists a minority of cases where the expected return of this security is exceptionally high and for which the optimal allocation heavily weights this security, whereas for the majority of the cases, low or zero weights are given to this security. However, the averaging process during the final determination of the resampled weights obscures this behavior. As as result, the resampled weights might exhibit a bias towards higher volatility securities even with a deterioration of their Sharpe ratio. Further, resam-

pling might help to mitigate the "error maximization" effect of a naive mean-variance optimization in-sample, but its usefulness out-of-sample over other methods in this regard is not clear.

Crafting the inputs

Using sample means and covariances of returns directly into the optimizer usually leads to poor results due to the large amount of estimation error present. Further, historical returns are usually not a good predictor of future returns in the short to middle term, and even in the long term (defined as several decades, using the human lifespan as a yardstick) this might not be true.

In this section, we document methods that somewhat address these weaknesses.

Expected returns

The Black-Litterman model is a systematic way to combine either expected returns estimated through sample means, or implied by an existing portfolio allocation, with an investor's own forward-looking views on expected returns [140]. Both the former and latter models of returns are described by probability distributions characterized by their means and covariances, and are mixed together in a way that assigns weights inversely proportional to their degree of uncertainty to arrive at the final vector of expected returns.

There are advantages to using the Black-Litterman update to systematically incorporate investors' views. First, investors do not need to provide a full set of expected returns in their projections. Rather, they could only provide their views on a subset of securities. Further, they have the flexibility of providing their views both

in a relative ("A outperforms B by 2%") or absolute sense ("A will return 6%, while B will return 4%."). Second, investors are able to express the degree of confidence in their views, and this information is used to derive the appropriate weighting between the existing and new views. Third, the update incorporates the current correlation structure between the securities, such that the resulting expected returns vector will reflect this information, even if only partial views are provided. This might also contribute to the observation that the inputs are generally robust, that the new optimal weights are not drastically different from the old ones if the investor provides only weak views in his or her update. In a sense, practitioners should be comforted by the observation that the new allocation should vary from the old one in a somewhat direct proportion to how different the new and old expected return vectors are.

We briefly document the inputs and outputs of the Black-Litterman model here. Readers should refer to Satchell and Scowcroft [141] and Idzorek [142] for greater detail.

The inputs to the model include:

- $\boldsymbol{\pi} \in \mathbb{R}^n$: A vector consisting of the existing implied expected returns for the n securities.
- $\boldsymbol{\Sigma} \in \mathbb{R}^{n \times n}$: The covariance matrix of returns for the n securities.
- $\tau \in \mathbb{R}$: A scaling prefactor for $\boldsymbol{\Sigma}$, commonly set to 1.
- $\mathbf{P} \in \mathbb{R}^{k \times n}$: A matrix that describes the linear relationships among the n securities for each of the k investor views.
- $\mathbf{q} \in \mathbb{R}^k$: A vector consisting of the k views of the investor.
- $\boldsymbol{\Omega} \in \mathbb{R}^{k \times k}$: A diagonal matrix of the variance of error terms for each view, reflecting the investor's degree of uncertainty in each view.

In practice, $\boldsymbol{\pi}$ is typically inferred via 'inverse optimization' given existing weights of a portfolio or benchmark. Alternatively, it could adopt the assumption that we are living in a CAPM world where all investors are holding global assets in a proportion that seeks to maximize their overall expected return. As such, the current market capitalization of global assets can be used to infer what the world assumes the assets' expected returns would be. In this situation, the implied excess expected return in an unconstrained portfolio setting is given by:

$$\boldsymbol{\pi} = \delta \boldsymbol{\Sigma} \mathbf{w}_m,$$

where $\delta = (\mu_m - r_f)/\sigma_m^2$, μ_m is the return on the global market in domestic currency, r_f is the risk-free rate, σ_m^2 is the variance of the rate of return on the world market, and $\mathbf{w}_m$ are the weights on the global market, as determined by market values.

Under the assumption that both returns and investor views are normally distributed, Bayes rule is applied to arrive at the formula for the new combined return vector $\mathbf{R}$. In fact, $\mathbf{R} \sim \mathcal{N}(\boldsymbol{\mu}_R, \boldsymbol{\Sigma}_R)$, where:

$$\begin{aligned}
\boldsymbol{\mu}_R &= [(\tau\boldsymbol{\Sigma})^{-1} + \mathbf{P}^\top \boldsymbol{\Sigma}^{-1} \mathbf{P}]^{-1} [(\tau\boldsymbol{\Sigma})^{-1} \boldsymbol{\pi} + \mathbf{P}^\top \boldsymbol{\Sigma}^{-1} \mathbf{q}], \\
\boldsymbol{\Sigma}_R &= [(\tau\boldsymbol{\Sigma})^{-1} + \mathbf{P}^\top \boldsymbol{\Sigma}^{-1} \mathbf{P}]^{-1}.
\end{aligned}$$

Covariances

There is a serious issue of estimation error when estimating the true covariance matrix using sample covariances. Consider the fact that for N securities, we require the estimation of $\frac{N(N+1)}{2}$ parameters, which is a large number. If we use data observations over T time periods, we have a total of NT observations. Naturally, if the latter

quantity is less than the former, we are unable to do the estimation, while if it is larger, but not by much, the estimation error of the covariance matrix will be very large. Consider a somewhat not uncommon case where $N = 100$ and $T = 120$ (i.e., monthly return observations over ten years). For every estimated parameter we have only slightly more than 2 observations - not a very promising situation.

As such, one should never simply use the sample covariance directly to represent the covariances. Rather, techniques have been proposed to "shrink" the sample covariance matrix, in the sense of reducing its estimation error. At the heart of the approach is the tuning of the trade-off between bias and variance in estimation. The sample covariance matrix has no bias, but at the expense of a huge variance. We can thus consider substituting or blending it with alternative estimators with higher bias but lower variances.

A common alternative estimator for the covariance matrix of returns is derived from a factor model. Given the general form of the factor model as introduced earlier in this paper:

$$\mathbf{r}_t = \boldsymbol{\alpha} + \boldsymbol{\beta}^\top \mathbf{f}_t + \boldsymbol{\epsilon}_t,$$

where $\mathbf{r}_t \in \mathbb{R}^n$ is the single-period return vector of the securities, $\mathbf{f}_t \in \mathbb{R}^k$ is the single-period return vector of the factors, $\boldsymbol{\beta} \in \mathbb{R}^{k \times n}$ is the matrix of factor loadings for each security, $\boldsymbol{\alpha} \in \mathbb{R}^n$ is the vector of excess returns, and $\boldsymbol{\epsilon}_t \in \mathbb{R}^n$ is the vector of residual returns, the covariance matrix of returns is given by:

$$\boldsymbol{\Sigma}_{FM} = \boldsymbol{\beta}^\top \boldsymbol{\Sigma}_f \boldsymbol{\beta} + \boldsymbol{\Sigma}_\epsilon,$$

where $\boldsymbol{\Sigma}_f \in \mathbb{R}^{k \times k}$ is the covariance matrix of the factors, and $\boldsymbol{\Sigma}_\epsilon \in \mathbb{R}^{n \times n}$ is the diagonal matrix of residual return variances.

Factor models involve the projection of security returns into a lower dimensional space, and this naturally reduces the variance of the covariance matrix estimator, at the expense of increasing its bias. They have shown to perform better in out-of-sample forecasts of covariances than using the sample covariance matrix. However, one sobering fact is that the correlation between predicted and future covariances is not large. For instance, in a study on AMEX and NYSE stocks, the correlation between past and future sample covariances is 34% at the 36-month horizon and much less at 18% at the 12-month horizon. Nevertheless, they are especially useful in cases where tracking error towards a benchmark is sought to be minimized [143].

Instead of directly using a substitute for the sample covariance matrix, Ledoit and Wolf provide a way to combine the sample covariance matrix with alternative, more structured estimators to produce a blended estimator $\mathbf{\Sigma}^*$ by the following linear combination:

$$\mathbf{\Sigma}^* = \delta\mathbf{F} + (1-\delta)\mathbf{S},$$

where $\mathbf{S}$ is the sample covariance matrix, $\mathbf{F}$ is a highly structured estimator, and δ is a number between 0 and 1. Ledoit and Wolf showed how to determine δ systemically by choosing δ such that the expected error of the combined matrix with the true covariance matrix is asymptotically minimized.

Examples of $\mathbf{F}$ given by Ledoit and Wolf are the identity matrix [144], the single-factor covariance matrix [145], and the constant correlation matrix [146]. The constant correlation matrix is constructed such that all pairwise correlations between securities are set to the global average pairwise correlation. That is, given entries s_{ij} of the sample covariance matrix, we construct $\mathbf{F}$ with entries $f_{ii} = s_{ii}$ and $f_{ij} = \bar{\rho}\sqrt{s_{ii}s_{jj}}$, where $\bar{\rho} = \frac{2}{N(N-1)}\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\rho_{ij}$ and $\rho_{ij} = \frac{s_{ij}}{\sqrt{s_{ii}s_{jj}}}$.

Interestingly, several papers have shown that constraining the weights of the

portfolio to be non-negative, or constraining the norm of the portfolio weights have a similar effect of shrinking the covariance matrix [147][148]. This gives insight as to why practitioners typically prefer bounds to be set on portfolio weights, where the practice is usually justified as a way to achieve "diversification". Here, we understand it to be a way to diversify against the presence of large estimation errors.

Conclusion

In this day and age, the main challenge in successfully applying optimization to portfolio selection lies in the generation of high quality input parameters. This trumps all other intellectual challenges commonly associated with the study of optimization - namely, new formulations, algorithmic schemes and the need to overcome computational tractability. Because the usefulness of optimization is of first-order tied to the quality of inputs, we see that useful optimization can be most easily achieved when the uncertainty regarding parameters are low or non-existent. These include cases involving the minimization of transaction costs, the construction of portfolios to adhere to certain trading requirements (such as minimum lot requirements), optimizing a portfolio to minimize tracking error to a benchmark, or the tuning the portfolio's overall exposure to underlying factors in a factor model.

On the other end of the uncertainty spectrum, when uncertainty is large, matters require much more care in order to preserve the usefulness of optimization. Whenever the problem involves expected returns, covariances, or scenarios of return realizations as numerical input, the majority of the value added derives from the practitioner's ability to provide these inputs, which require the ability to look forward reasonably accurately in time. Precluding any major advances in artificial intelligence, this skill is currently only able to be acquired through many years of practical working

experience in the field, requiring a judicious combination of human intuition with the support of hard data from a variety of sources to make sound inferences about the future.

Supplementary note on leverage and short selling

Here, we discuss the concepts of leverage and short selling. These are concepts which are common in the practice of finance, but might be obscure to members of the engineering community. In fact, the definition of leverage implies that the total portfolio weight is greater than 1, while the definition of short selling implies a negative weight on whichever security is "sold short". As a result, these concepts possess a certain "unphysicality" which understandably might unsettle an engineer, whose attendance is towards that of the physical world, upon first encounter. In reality, the ability to leverage or sell short is enabled by the fact that investors in the financial markets are able to borrow either cash or securities from other parties in order to achieve their financial aims.

Leverage implies that the investor is borrowing cash over and above what he or she possesses, in the hopes of generating returns in excess of the borrowing costs. A total portfolio weight of $1 + \Delta$ where $\Delta > 0$ implies that for every one dollar that the investor owns, he or she borrows Δ dollars. A typical example that most people would be familiar with would be taking out a mortgage loan when purchasing a house. Here, owners are commonly expected to contribute cash equivalent to 20% of the value of the house, while borrowing the remaining 80%. In this case, if the investor's portfolio consists only of the house, then $\Delta = 4$ and the total portfolio weight is 5.

Additionally, we could also have a situation where the total portfolio weight is

between 0 and 1, say $1 - \Delta$, where $0 < \Delta < 1$. This is when the investor decides to keep a proportion Δ of his total wealth in cash, while investing the rest.

Short selling is a technique where the investor borrows a security that he or she does not own from another party and sells it in the market, paying borrowing fees in order to do so. The investor is expected to return the security at a later date. The reason why investors are motivated to sell short is because they expect the security to go down in price in the future, where they can buy it back at a cheaper price and book a profit, net of borrowing costs. The result of short selling is a negative weight on the security in the investor's portfolio while it is being held short. Intuitively, this makes sense since a negative return on the security sold short would lead to a positive return contribution to the portfolio.

Finally, the ability to go long and short on securities lead to the interesting concept of a "zero-investment portfolio", where the total portfolio weight is zero. This implies the situation where every dollar on the long side is balanced with a dollar on the short side. This is a useful construct theoretically for studying the sources of risk in trading strategies, and also in practical situations, for example, when the portfolio manager would like to reduce certain systematic risks in the portfolio. For example, by going long on one stock and short another within the same sector, risks specifically related to the sector can be reduced, and profits are made only through the relative performance of these stocks.

Bibliography

[1] U.S. Energy Information Administration, "Annual Energy Outlook 2015 - Introduction," 2015. [Online]. Available: http://www.eia.gov/forecasts/aeo/chapter_intro.cfm [Accessed: 30 June 2015]

[2] ——, "Assumptions to the Annual Energy Outlook 2014 - Oil and Gas Supply Module," Office of Integrated and International Energy Analysis, U.S. Department of Energy, Washington, DC, USA, Tech. Rep., 2014. [Online]. Available: http://www.eia.gov/forecasts/aeo/assumptions/pdf/oilgas.pdf [Accessed: 26 May 2015]

[3] Multi-Resolution Land Characteristics Consortium, "Product Legend," 2011. [Online]. Available: http://www.mrlc.gov/nlcd11_leg.php [Accessed: 26 May 2015]

[4] M. J. Kaiser and J. H. Gary, "Study updates refinery investment cost curves," 2007. [Online]. Available: http://www.ogj.com/articles/print/volume-105/issue-16/processing/study-updates-refinery-investment-cost-curves.html [Accessed: 5 August 2014]

[5] U.S. Energy Information Administration, "Annual Energy Outlook 2015," Office of Integrated and International Energy Analysis, U.S. Department of Energy, Washington, DC, USA, Tech. Rep., 2015. [Online]. Available: http://www.eia.gov/forecasts/aeo/pdf/0383(2015).pdf [Accessed: 26 May 2015]

[6] M. Martín and I. E. Grossmann, "Optimal use of hybrid feedstock, switchgrass and shale gas for the simultaneous production of hydrogen and liquid fuels," Energy, vol. 55, pp. 378–391, 2013.

[7] D. C. Cafaro and I. E. Grossmann, "Strategic planning, design, and development of the shale gas supply chain network," AIChE Journal, vol. 60, no. 6, pp. 2122–2142, 2014.

[8] L. Yang, I. E. Grossmann, and J. Manno, "Optimization models for shale gas water management," AIChE Journal, vol. 60, no. 10, pp. 3490–3501, 2014.

[9] J. Gao and F. You, "Optimal design and operations of supply chain networks for water management in shale gas production: mixed-integer linear fractional programming model and algorithms for the water-energy nexus," AIChE Journal, vol. 61, no. 4, pp. 1184–1208, 2015.

[10] ——, "Shale gas supply chain design and operations toward better economic and life cycle environmental performance: Minlp model and global optimization algorithm," ACS Sustainable Chemistry & Engineering, vol. 3, no. 7, pp. 1282–1291, 2015.

[11] B. R. Knudsen and B. Foss, "Shut-in based production optimization of shale-gas systems," Computers & Chemical Engineering, vol. 58, pp. 54–67, 2013.

[12] B. R. Knudsen, I. E. Grossmann, B. Foss, and A. R. Conn, "Lagrangian relaxation based decomposition for well scheduling in shale-gas systems," Computers & Chemical Engineering, vol. 63, pp. 234–249, 2014.

[13] B. R. Knudsen, C. H. Whitson, and B. Foss, "Shale-gas scheduling for natural-gas supply in electric power production," Energy, vol. 78, pp. 165–182, 2014.

[14] J. E. Bistline, "Natural gas, uncertainty, and climate policy in the US electric power sector," Energy Policy, vol. 74, pp. 433–442, 2014.

[15] E. Attanasi and P. Freeman, "Role of stranded gas in increasing global gas supplies, Open-File Report 2013-1044," U.S. Geological Survey, Reston, VA, USA, Tech. Rep., 2013. [Online]. Available: http://pubs.usgs.gov/of/2013/1044 [Accessed: 11 June 2015]

[16] Velocys, "Velocys - Gas-to-liquids," 2015. [Online]. Available: http://www.velocys.com/our_business_markets_gtl.php [Accessed: 11 June 2015]

[17] M. Ford and N. Davis, "Nonmarketed natural gas in North Dakota still rising due to higher total production," 2014. [Online]. Available: http://www.eia.gov/todayinenergy/detail.cfm?id=15511 [Accessed: 11 June 2015]

[18] GE Oil & Gas, "Small-Scale LNG Liquefaction Plant," 2015. [Online]. Available: https://www.geoilandgas.com/distributed-gas/processing/small-scale-lng-liquefaction-plant [Accessed: 11 June 2015]

[19] CompactGTL, "Technology Overview," 2015. [Online]. Available: http://www.compactgtl.com/technology/overview/ [Accessed: 11 June 2015]

[20] Velocys, "Smaller Scale GTL," 2013. [Online]. Available: http://www.velocys.com/arcv/press/ppt/CERAWeek%20v9%20March%202013.pdf [Accessed: 11 June 2015]

[21] J. D. Hughes, "Energy: A reality check on the shale revolution," Nature, vol. 494, no. 7437, pp. 307–8, February 2013.

[22] M. P. Nowak and W. Römisch, "Stochastic Lagrangian relaxation applied to power scheduling in a hydro-thermal system under uncertainty," Annals of Operations Research, vol. 100, no. 1-4, pp. 251–272, 2000.

[23] N. Gröwe-Kuska, K. C. Kiwiel, M. P. Nowak, W. Römisch, and I. Wegner, "Power management in a hydro-thermal system under uncertainty by Lagrangian relaxation," in Decision Making Under Uncertainty: Energy and Power. Springer, 2002, pp. 39–70.

[24] IHS Global Inc., "Oil & Natural Gas Transportation & Storage Infrastructure: Status, Trends, & Economic Benefits," American Petroleum Institute, Washington, DC, USA, Tech. Rep., 2013. [Online]. Available: http://www.api.org/~/media/Files/Policy/SOAE-2014/API-Infrastructure-Investment-Study.pdf [Accessed: 22 June 2015]

[25] EY Global Oil & Gas Center, "Spotlight on oil and gas megaprojects," Ernst & Young Global Limited, London, UK, Tech. Rep., 2014. [Online]. Available: http://www.ey.com/Publication/vwLUAssets/EY-spotlight-on-oil-and-gas-megaprojects/$FILE/EY-spotlight-on-oil-and-gas-megaprojects.pdf [Accessed: 26 May 2015]

[26] B. White, "North American LNG industry looks for survival through exports," Office of the Federal Coordinator, Alaska Natural Gas Transportation Projects, Tech. Rep. January 2013, 2013. [Online]. Available: http://www.arcticgas.gov/north-american-lng-industry-looks-survival-through-exports [Accessed: 22 June 2015]

[27] S. H. Owen and M. S. Daskin, "Strategic facility location: A review," European Journal of Operational Research, vol. 111, no. 3, pp. 423–447, 1998.

[28] M. T. Melo, S. Nickel, and F. Saldanha-da Gama, "Facility location and supply chain management–A review," European Journal of Operational Research, vol. 196, no. 2, pp. 401–412, 2009.

[29] J. A. Elia and C. A. Floudas, "Energy Supply Chain Optimization of Hybrid Feedstock Processes: A Review," Annual Review of Chemical and Biomolecular Engineering, vol. 5, pp. 147–179, 2014.

[30] N. V. Sahinidis, "Optimization under uncertainty: state-of-the-art and opportunities," Computers & Chemical Engineering, vol. 28, no. 6, pp. 971–983, 2004.

[31] IBM, "IBM ILOG CPLEX Optimizer," 2014. [Online]. Available: http://www-01.ibm.com/software/commerce/optimization/cplex-optimizer/index.html [Accessed: 11 June 2015]

[32] Gurobi Optimization, Inc., "Gurobi optimizer reference manual," 2015. [Online]. Available: http://www.gurobi.com [Accessed: 7 September 2015]

[33] M. Conforti, G. Cornuéjols, and G. Zambelli, Integer programming. Springer, 2014, vol. 271.

[34] M. Feinberg and P. Ellison, "General kinetic bounds on productivity and selectivity in reactor-separator systems of arbitrary design: Principles," Industrial & Engineering Chemistry Research, vol. 40, no. 14, pp. 3181–3194, 2001.

[35] Y. Tang and M. Feinberg, "Carnot-like limits to steady-state productivity," Industrial & Engineering Chemistry Research, vol. 46, no. 17, pp. 5624–5630, 2007.

[36] H. O. Hobbs Jr. and L. S. Adair, "Analysis shows GTL viable alternative for US gas producers," 2012. [Online]. Available: http://www.ogj.com/articles/print/vol-110/issue-8/refining-report/analysis-shows-gtl-viable-alternative.html [Accessed: 11 June 2015]

[37] E. Salehi, W. Nel, and S. Save, "Viability of GTL for the North American gas market," Hydrocarbon Processing, vol. 92, no. 1, pp. 41–48, Jan. 2013.

[38] D. A. Wood, C. Nwaoha, and B. F. Towler, "Gas-to-liquids (GTL): A review of an industry offering several routes for monetizing natural gas," Journal of Natural Gas Science and Engineering, vol. 9, pp. 196–208, November 2012.

[39] TIAX, "U.S. and Canadian Natural Gas Vehicle Market Analysis : Liquefied Natural Gas Infrastructure," America's Natural Gas Alliance, Lexington, MA, USA, Tech. Rep., 2012. [Online]. Available: https://www.aga.org/tiax-natural-gas-vehicle-market-analysis [Accessed: 11 June 2015]

[40] Platts Price Group, "The North American Gas Value Chain: Developments and Opportunities," Platts, McGraw Hill Financial, New York, NY, USA, Tech. Rep., 2012. [Online]. Available: http://www.platts.com/im.platts.content/insightanalysis/industrysolutionpapers/gasvaluechain.pdf [Accessed: 11 June 2015]

[41] C. A. Wocken, B. G. Stevens, J. C. Almlie, and S. M. Schlasner, "End-Use Technology Study - An Assessment of Alternative Uses for Associated Gas," Energy & Environmental Research Center, Grand Forks, ND, USA, Tech. Rep., 2013. [Online]. Available: http://www.undeerc.org/bakken/pdfs/CW_Tech_Study_April-2013.pdf [Accessed: 11 June 2015]

[42] J. Mason, "Bakken's maximum potential oil production rate explored," 2012. [Online]. Available: http://www.ogj.com/articles/print/vol-110/issue-4/exploration-development/bakken-s-maximum.html [Accessed: 11 June 2015]

[43] B. Patel, "Gas Monetisation: A Techno-Economic Comparison Of Gas-To-Liquid and LNG," in 7th World Congress of Chemical Engineering, Glasgow, Scotland, 2005, pp. 1–11.

[44] E. D. Garcia-Cuerva and F. S. Sobrino, "A New Business Approach to Conventional Small Scale LNG, Paper No. 599.00," in IGU 24th World Gas Conference, Buenos Aires, Argentina, 2009, pp. 1–42.

[45] N. C. Ballout and B. C. Price, "Comparison of Present Day Peakshaving Liquefaction Technologies," in AIChE Spring National Meeting, 8th Topical Conference on Natural Gas Utilization, New Orleans, LA, USA, 2008, pp. 30–47.

[46] I. Baxter, "Small-scale GTL: back on the agenda," pp. 84–85, 2012. [Online]. Available: http://www.compactgtl.com/wp-content/documents/pe_worldgas_2012.pdf [Accessed: 11 June 2015]

[47] Continental Resources Inc., "Bakken Field Recoverable Reserves," Enid, Oklahoma, 2011. [Online]. Available: www.contres.com/download/file/fid/269 [Accessed: 11 June 2015]

[48] GE Oil & Gas, "Accelerating adoption of LNG fuelling infrastructure," 2013. [Online]. Available: http://www.gastechnology.org/Training/Documents/LNG17-proceedings/Transport-18-Ujjwal-Kumar-Presentation.pdf [Accessed: 11 June 2015]

[49] Oxford Catalysts, "Oilbarrel Conference," 2013. [Online]. Available: http://www.velocys.com/arcv/press/ppt/ppt130613_2_OCG%20Oil%20Barrel%20June13%20FINAL_rev2.pdf [Accessed: 11 June 2015]

[50] H. Bauman, "Cost of Starting Up the Chemical Process Plant," Industrial & Engineering Chemistry, vol. 52, no. 3, pp. 52–53, 1960.

[51] Coinnews Media Group, "U.S. Inflation Calculator," 2015. [Online]. Available: http://www.usinflationcalculator.com/ [Accessed: 26 May 2015]

[52] National Energy Technology Laboratory, "Analysis of Natural Gas-to Liquid Transportation Fuels via Fischer-Tropsch," U.S. Department of Energy, Tech. Rep., 2013. [Online]. Available: https://netl.doe.gov/File%20Library/Research/Energy%20Analysis/Publications/Gas-to-Liquids_Report.pdf [Accessed: 11 June 2015]

[53] T. Kohler, M. Bruentrup, R. Key, and T. Edvardsson, "Choose the best refrigeration technology for small-scale LNG production," Hydrocarbon Processing, vol. January 2014, pp. 45–52, 2014.

[54] K. A. Lawlor and M. Conder, "Gathering and Processing Design Options for Unconventional Gas," Midstream Business, vol. 3, no. 4, pp. 54–58, 2013.

[55] U.S. Energy Information Administration, "Drilling Productivity Report," 2015. [Online]. Available: http://www.eia.gov/petroleum/drilling/ [Accessed: 26 May 2015]

[56] ——, "U.S. Energy Mapping System," 2015. [Online]. Available: http://www.eia.gov/state/maps.cfm [Accessed: 11 June 2015]

[57] ——, "Frequently Asked Questions - How Many Gallons of Diesel Fuel does One Barrel of Oil Make?" 2014. [Online]. Available:

http://www.eia.gov/tools/faqs/faq.cfm?id=327&t=9 [Accessed: 11 June 2015]

[58] U.S. Department of Energy Office of Energy Efficiency & Renewable Energy, "Alternative Fueling Station Locator," 2014. [Online]. Available: http://www.afdc.energy.gov/locator/stations/ [Accessed: 28 April 2014]

[59] U.S. Energy Information Administration, "Assumptions to the Annual Energy Outlook 2013," U.S. Department of Energy, Washington, DC, USA, Tech. Rep., 2013. [Online]. Available: http://www.eia.gov/forecasts/aeo/assumptions/pdf/0554(2013).pdf [Accessed: 11 June 2015]

[60] K. Van Hull, "Big Surge Comes to Whoville - Northeast NGLs to Increase Six Fold," 2013. [Online]. Available: https://rbnenergy.com/big-surge-comes-to-whoville-northeast-ngls-increase-six-fold [Accessed: 11 June 2015]

[61] Google Inc., "The Google Distance Matrix API," Mountain View, CA, USA, 2015. [Online]. Available: https://developers.google.com/maps/documentation/distancematrix/ [Accessed: 26 May 2015]

[62] G. R. Hadder and B. D. McNutt, "Ultra-Clean Diesel Fuel: U.S. Production and Distribution Capability," Oak Ridge National Laboratory, Oak Ridge, TN, USA, Tech. Rep., 2000. [Online]. Available: http://www.osti.gov/bridge/product.biblio.jsp?osti_id=777709 [Accessed: 11 June 2015]

[63] Aux Sable, "Examining Local and Export NGL Markets: An Assessment of NGL Takeaway Capacity and Transportation Costs," 2014. [Online]. Available: http://www.auxsable.com/Documents/Tim_Stauft_Bakken%20Product%20Markets%20%20Takeaway%20Capacity_30Jan2014_FINALVersion.pdf [Accessed: 11 June 2015]

[64] U.S. Energy Information Administration, "The National Energy Modeling System: An Overview 2009," Office of Integrated Analysis and Forecasting, U.S. Department of Energy, Washington, DC, USA, Tech. Rep., 2009. [Online]. Available: http://www.eia.gov/oiaf/aeo/overview/pdf/0581(2009).pdf [Accessed: 11 June 2015]

[65] ——, "Annual Energy Outlook 2014," Office of Integrated and International Energy Analysis, U.S. Department of Energy, Washington, DC, USA,

Tech. Rep., 2014. [Online]. Available: http://www.eia.gov/forecasts/aeo/pdf/0383(2014).pdf [Accessed: 26 May 2015]

[66] North Dakota Department of Mineral Resources, "ND Monthly Bakken Oil Production Statistics," 2015. [Online]. Available: https://www.dmr.nd.gov/oilgas/stats/historicalbakkenoilstats.pdf [Accessed: 11 June 2015]

[67] Deloitte Financial Advisory S.r.l., "Overview of business valuation parameters in the energy industry," Deloitte Touche Tohmatsu Limited, Italy, Tech. Rep., 2013.

[68] A. Shapiro, D. Dentcheva, and A. Ruszczynski, Lectures on Stochastic Programming: Modeling and Theory. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics and the Mathematical Programming Society, 2009.

[69] X. Li, E. Armagan, A. Tomasgard, and P. I. Barton, "Stochastic Pooling Problem for Natural Gas Production Network Design and Operation Under Uncertainty," AIChE Journal, vol. 57, no. 8, pp. 2120–2135, 2011.

[70] J. R. Birge and F. Louveaux, Introduction to Stochastic Programming. Springer Science & Business Media, 1997, vol. 3.

[71] U.S. Energy Information Administration, "Drilling Productivity Report - Report Background and Methodological Overview," U.S. Department of Energy, Washington, DC, USA, Tech. Rep., 2014. [Online]. Available: http://www.eia.gov/petroleum/drilling/pdf/dpr_methodology.pdf [Accessed: 26 May 2015]

[72] IHS Global Inc., "US Crude Oil Export Decision," IHS Global Inc., Houston, TX, USA, Tech. Rep., 2014. [Online]. Available: https://www.ihs.com/Info/0514/crude-oil.html [Accessed: 10 June 2014]

[73] K. Hauge, "Refining ABC." [Online]. Available: http://www.statoil.com/en/InvestorCentre/Presentations/Downloads/Refining.pdf [Accessed: 12 March 2015]

[74] P. Sankey, L. Hermann, D. T. Clark, S. Micheloto, and W. Nip, "Gorgon & the Global LNG Monster," Deutsche Bank Securities Inc., New York, NY, USA, Tech. Rep., 2012.

[75] S. H. Tan and P. I. Barton, "Optimal Dynamic Allocation of Mobile Plants to Monetize Associated or Stranded Natural Gas, Part I: Bakken Shale Play Case Study," Energy, vol. Manuscript in Review, 2015.

[76] ——, "Optimal Dynamic Allocation of Mobile Plants to Monetize Associated or Stranded Natural Gas, Part II: Dealing with Uncertainty," Energy, vol. Manuscript in Review, 2015.

[77] U.S. Energy Information Administration, "Performance Profiles of Major Energy Producers 2009," Office of Energy Statistics, U.S. Department of Energy, Washington, DC, USA, Tech. Rep., 2011. [Online]. Available: http://www.eia.gov/finance/performanceprofiles/pdf/020609.pdf [Accessed: 5 December 2014]

[78] U.S. Energy Information Administration Biofuels and Emerging Technologies Team, "Gas-To-Liquid (GTL) Technology Assessment in Support of AEO2013," 2013. [Online]. Available: http://www.eia.gov/forecasts/documentation/workshops/pdf/AEO2013_GTL_Assessment.pdf [Accessed: 11 June 2015]

[79] S. Paltsev, F. O'Sullivan, N. Lee, A. Agarwal, M. Li, X. Li, and N. Fylaktos, "Natural Gas Monetization Pathways for Cyprus: Interim Report - Economics of Project Development Options," MIT Energy Initiative, Massachusetts Institute of Technology, Cambridge, MA, Tech. Rep. August, 2013. [Online]. Available: http://mitei.mit.edu/system/files/Cyprus_NG_Report.pdf [Accessed: 26 May 2015]

[80] About.com, "Geographic Centers of the Fifty States." [Online]. Available: http://geography.about.com/library/weekly/aa120699a.htm [Accessed: 26 May 2015]

[81] U.S. Energy Information Administration, "International Energy Outlook 2013," U.S. Department of Energy, Washington, DC, USA, Tech. Rep., 2013. [Online]. Available: http://www.eia.gov/forecasts/ieo/pdf/0484(2013).pdf [Accessed: 26 May 2015]

[82] ——, "State Energy Data System," 2013. [Online]. Available: http://www.eia.gov/state/seds/ [Accessed: 26 May 2015]

[83] ICF International, "The Impacts of U.S. Crude Oil Exports on Domestic Crude Production, GDP, Employment, Trade, and Consumer Costs,"

ICF International, Fairfax, VA, USA, Tech. Rep., 2014. [Online]. Available: http://www.api.org/~/media/Files/Policy/LNG-Exports/LNG-primer/API-Crude-Exports-Study-by-ICF-3-31-2014.pdf [Accessed: 26 May 2015]

[84] T. J. Duesterberg, D. A. Norman, and J. F. Werling, "Lifting the Crude Oil Export Ban: The Impact on U.S. Manufacturing," The Aspen Institute, Washington, DC, USA, Tech. Rep., 2014. [Online]. Available: http://www.aspeninstitute.org/sites/default/files/content/upload/FINAL_Lifting_Crude_Oil_Export_Ban_0.pdf [Accessed: 26 May 2015]

[85] American Petroleum Institute, "U.S. Crude Oil Exports - Benefits for America's Economy and Consumers," American Petroleum Institute, Washington, DC, USA, Tech. Rep., 2015. [Online]. Available: http://www.api.org/~/media/files/policy/exports/crude-oil-exports-primer/us-crude-oil-exports-low-res.pdf [Accessed: 27 March 2015]

[86] U.S. Energy Information Administration, "Exports by Destination," 2015. [Online]. Available: http://www.eia.gov/dnav/pet/pet_move_expc_a_EP00_EEX_mbbl_a.htm [Accessed: 11 May 2015]

[87] ——, "U.S. Natural Gas Exports and Re-Exports by Point of Exit," 2015. [Online]. Available: http://www.eia.gov/dnav/ng/ng_move_poe2_dcu_nus-nmx_a.htm [Accessed: 26 May 2015]

[88] International Energy Agency, "Energy Prices and Taxes Quarterly Statistics, First Quarter 2015," International Energy Agency, France, Tech. Rep., 2015. [Online]. Available: http://www.oecd-ilibrary.org/energy/energy-prices-and-taxes_16096835 [Accessed: 26 May 2015]

[89] Secretaria de Energia, "Liquefied Petroleum Gas Market Outlook 2008-2017," Secretaria de Energia, Mexico, Tech. Rep., 2008. [Online]. Available: http://www.energia.gob.mx/res/pe_y_dt/pub/lpg%20outlook%202008-2017.pdf [Accessed: 26 May 2015]

[90] Federal Energy Regulatory Commission, "LNG Archives," 2015. [Online]. Available: http://www.ferc.gov/market-oversight/othr-mkts/lng/archives.asp [Accessed: 26 May 2015]

[91] U.S. Energy Information Administration, "Retail Premium Gasoline Prices, Selected Countries," 2015. [Online]. Available: http://www.eia.gov/countries/prices/gasolineextax.cfm [Accessed: 26 May 2015]

[92] ——, "Retail Diesel Prices, Selected Countries," 2015. [Online]. Available: http://www.eia.gov/countries/prices/dieselextax.cfm [Accessed: 11 May 2015]

[93] ——, "Residual Fuel Oil Prices by Sales Type," 2015. [Online]. Available: http://www.eia.gov/dnav/pet/pet_pri_resid_dcu_nus_a.htm [Accessed: 26 May 2015]

[94] ——, "No. 2 Distillate Prices by Sales Type," 2015. [Online]. Available: http://www.eia.gov/dnav/pet/pet_pri_dist_dcu_nus_a.htm [Accessed: 26 May 2015]

[95] MyLPG.eu, "Chart of fuel prices in United Kingdom," 2015. [Online]. Available: http://www.mylpg.eu/stations/united-kingdom/prices [Accessed: 11 May 2015]

[96] D. Furchtgott-Roth, "Pipelines are safest for transportation of oil and gas," Manhattan Institute for Policy Research, New York, NY, Tech. Rep., 2013. [Online]. Available: http://www.manhattan-institute.org/pdf/ib_23.pdf [Accessed: 26 May 2015]

[97] C. J. Trench, "How Pipelines Make the Oil Market Work - Their Networks, Operation and Regulation," Allegro Energy Group, Dallas, TX, USA, Tech. Rep. December, 2001. [Online]. Available: http://www.iatp.org/files/451_2_31375.pdf [Accessed: 26 May 2015]

[98] S. Cornot-Gandolphe, O. Appert, R. Dickel, M.-F. Chabrelie, and A. Rojey, "The Challenges of Further Cost Reductions for New Supply Options (Pipeline, LNG, GTL)," in 22nd World Gas Conference, Tokyo, Japan, 2003, pp. 1–17.

[99] D. J. Hammett, "Pipeline Capacity: Ensuring the Resiliency the US Requires," National Defense Transportation Association, Alexandria, VA, Tech. Rep. February, 2014. [Online]. Available: http://www.ndtahq.com/documents/PipelineCapacity.pdf [Accessed: 26 May 2015]

[100] ESRI (Environmental Systems Resource Institute), "ArcMap 10.2.2," Redlands, CA, USA, 2014.

[101] M. M. Foss, "Introduction to LNG," Center for Energy Economics, Houston, TX, USA, Tech. Rep., 2012. [Online]. Available: http://www.beg.utexas.edu/energyecon/INTRODUCTION%20TO%20LNG%20Update%202012.pdf [Accessed: 26 May 2015]

[102] Arkansas Oklahoma Port Operators Association, "McClellan-Kerr Arkansas River Navigation System - Waterway Facts," 2015. [Online]. Available: http://www.aopoa.net/history/facts.htm [Accessed: 26 May 2015]

[103] CME Group, "Conversion Calculator," 2015. [Online]. Available: http://www.cmegroup.com/tools-information/calc_crude.html [Accessed: 26 May 2015]

[104] U.S. Energy Information Administration, "International Energy Statistics - Units," 2015. [Online]. Available: http://www.eia.gov/cfapps/ipdbproject/docs/unitswithpetro.cfm [Accessed: 26 May 2015]

[105] Arkansas Waterways Commission, "Education: Why Waterways?" 2015. [Online]. Available: http://waterways.arkansas.gov/education/Pages/whyWaterways.aspx [Accessed: 26 May 2015]

[106] Prometheus Energy, "LNG Quick Facts," 2015. [Online]. Available: http://www.prometheusenergy.com/_pdf/LNGQuickFacts.pdf [Accessed: 26 May 2015]

[107] Federal Railroad Administration, "U.S. National Transportation Atlas Railroads," Redlands, CA, USA, 2013. [Online]. Available: http://resources.arcgis.com/EN/HELP/MAIN/10.1/index.html#//001z0000002z000000 [Accessed: 26 May 2015]

[108] U.S. Energy Information Administration, "Layers Information for Interactive State Maps," 2014. [Online]. Available: http://www.eia.gov/maps/layer_info-m.cfm [Accessed: 26 May 2015]

[109] U.S. Army Corps of Engineers Navigation Data Center, "Ports and Waterways Facilities," 2014. [Online]. Available: http://www.navigationdatacenter.us/ports/ports.htm [Accessed: 20 June 2014]

[110] Ports.com, "Sea route & distance," 2015. [Online]. Available: http://ports.com/sea-route/ [Accessed: 1 August 2014]

[111] U.S. Energy Information Administration, "Mexico - Analysis," 2014. [Online]. Available: http://www.eia.gov/countries/cab.cfm?fips=MX [Accessed: 26 May 2015]

[112] Sea-Distances.org, "Ports Distances," 2015. [Online]. Available: http://www.sea-distances.org/ [Accessed: 26 May 2015]

[113] J. Frittelli, "Shipping U.S. Crude Oil by Water: Vessel Flag Requirements and Safety Issues," Congressional Research Service, Washington, DC, USA, Tech. Rep., 2014. [Online]. Available: https://www.fas.org/sgp/crs/misc/R43653.pdf [Accessed: 26 May 2015]

[114] K. B. Medlock III, "U.S. LNG Exports: Truth and Consequence," James A. Baker III Institute for Public Policy of Rice University, Houston, TX, USA, Tech. Rep., 2012. [Online]. Available: http://bakerinstitute.org/media/files/Research/da5493d4/US_LNG_Exports_-_Truth_and_Consequence_Final_Aug12-1.pdf [Accessed: 26 May 2015]

[115] Federal Energy Regulatory Commission, "LNG," 2015. [Online]. Available: http://www.ferc.gov/industries/gas/indus-act/lng.asp [Accessed: 22 June 2015]

[116] GTI Commerical and Investments Group, "Small Scale Liquefier Development," 2003. [Online]. Available: http://www1.eere.energy.gov/cleancities/pdfs/gti_liquefier.pdf [Accessed: 11 June 2015]

[117] S. Jin, L. Yang, P. Danielson, C. Homer, J. Fry, and G. Xian, "A comprehensive change detection method for updating the National Land Cover Database to circa 2011," Remote Sensing of Environment, vol. 132, pp. 159–175, 2013.

[118] National Atlas of the United States and the United States Geological Survey, "U.S. National Atlas Federal and Indian Land Areas," Redlands, CA, USA, 2013. [Online]. Available: http://help.arcgis.com/en/arcgisdesktop/10.0/help/index.html#//001z00000031000000.htm [Accessed: 26 May 2015]

[119] U.S. Energy Information Administration, "U.S. Natural Gas Pipeline Projects," Washington, DC, USA, 2014. [Online]. Available: http://www.eia.gov/naturalgas/pipelines/EIA-NaturalGasPipelineProjects.xls [Accessed: 16 March 2015]

[120] K. Hays, "U.S. crude oil pipeline projects: Kinder Morgan acquiring Hiland Crudge," 2015. [Online]. Available: http://www.reuters.com/article/2015/01/21/us-usa-pipeline-oil-factbox-idUSKBN0KU2SX20150121 [Accessed: 16 March 2015]

[121] Office of Coast Survey, National Ocean Service, and National Oceanic and Atmospheric Administration, "Distances Between United States Ports,"

U.S. Department of Commerce, Washington, DC, USA, Tech. Rep., 2012. [Online]. Available: http://www.nauticalcharts.noaa.gov/nsd/distances-ports/distances.pdf [Accessed: 12 January 2015]

[122] H. Markowitz, "Portfolio selection," The journal of finance, vol. 7, no. 1, pp. 77–91, 1952.

[123] R. O. Michaud, "The markowitz optimization enigma: Is optimized optimal?" ICFA Continuing Education Series, vol. 1989, no. 4, pp. 43–54, 1989.

[124] M. S. Lobo, M. Fazel, and S. Boyd, "Portfolio optimization with linear and fixed transaction costs," Annals of Operations Research, vol. 152, no. 1, pp. 341–365, 2007.

[125] M. G. Speranza, "A heuristic algorithm for a portfolio optimization model applied to the milan stock market," Computers & Operations Research, vol. 23, no. 5, pp. 433–441, 1996.

[126] H. Konno and A. Wijayanayake, "Portfolio optimization problem under concave transaction costs and minimal transaction unit constraints," Mathematical Programming, vol. 89, no. 2, pp. 233–250, 2001.

[127] H. Konno and H. Yamazaki, "Mean-absolute deviation portfolio optimization model and its applications to tokyo stock market," Management science, vol. 37, no. 5, pp. 519–531, 1991.

[128] C. D. Feinstein and M. N. Thapa, "Notes: A reformulation of a mean-absolute deviation portfolio optimization model." Management Science, vol. 39, no. 12, 1993.

[129] H. Konno, H. Shirakawa, and H. Yamazaki, "A mean-absolute deviation-skewness portfolio optimization model," Annals of Operations Research, vol. 45, no. 1, pp. 205–220, 1993.

[130] M. R. Young, "A minimax portfolio selection rule with linear programming solution," Management science, vol. 44, no. 5, pp. 673–683, 1998.

[131] A. A. Gaivoronski and G. Pflug, "Value-at-risk in portfolio optimization: properties and computational approach," The Journal of Risk, vol. 7, no. 2, p. 1, 2004.

[132] R. T. Rockafellar and S. Uryasev, "Optimization of conditional value-at-risk," Journal of risk, vol. 2, pp. 21–42, 2000.

[133] P. Krokhmal, J. Palmquist, and S. Uryasev, "Portfolio optimization with conditional value-at-risk objective and constraints," Journal of risk, vol. 4, pp. 43–68, 2002.

[134] R. Mansini, W. Ogryczak, and M. G. Speranza, "Conditional value at risk and related linear programming models for portfolio optimization," Annals of operations research, vol. 152, no. 1, pp. 227–256, 2007.

[135] F. Andersson, H. Mausser, D. Rosen, and S. Uryasev, "Credit risk optimization with conditional value-at-risk criterion," Mathematical Programming, vol. 89, no. 2, pp. 273–291, 2001.

[136] A. Chekhlov, S. P. Uryasev, and M. Zabarankin, "Portfolio optimization with drawdown constraints," 2000.

[137] A. Chekhlov, S. Uryasev, and M. Zabarankin, "Drawdown measure in portfolio optimization," International Journal of Theoretical and Applied Finance, vol. 8, no. 01, pp. 13–58, 2005.

[138] D. Bertsimas, D. B. Brown, and C. Caramanis, "Theory and applications of robust optimization," SIAM review, vol. 53, no. 3, pp. 464–501, 2011.

[139] B. Scherer, "Portfolio resampling: Review and critique," Financial Analysts Journal, pp. 98–109, 2002.

[140] F. Black and R. Litterman, "Global portfolio optimization," Financial analysts journal, vol. 48, no. 5, pp. 28–43, 1992.

[141] S. Satchell and A. Scowcroft, "A demystification of the black–litterman model: Managing quantitative and traditional portfolio construction," Journal of Asset Management, vol. 1, no. 2, pp. 138–150, 2000.

[142] T. M. Idzorek, "A step-by-step guide to the black-litterman model," Forecasting expected returns in the financial markets, vol. 17, 2002.

[143] L. K. Chan, J. Karceski, and J. Lakonishok, "On portfolio optimization: Forecasting covariances and choosing the risk model," Review of Financial Studies, vol. 12, no. 5, pp. 937–974, 1999.

[144] O. Ledoit and M. Wolf, "A well-conditioned estimator for large-dimensional covariance matrices," Journal of multivariate analysis, vol. 88, no. 2, pp. 365–411, 2004.

[145] ——, "Improved estimation of the covariance matrix of stock returns with an application to portfolio selection," Journal of empirical finance, vol. 10, no. 5, pp. 603–621, 2003.

[146] ——, "Honey, i shrunk the sample covariance matrix," The Journal of Portfolio Management, vol. 30, no. 4, pp. 110–119, 2004.

[147] R. Jagannathan and T. Ma, "Risk reduction in large portfolios: Why imposing the wrong constraints helps," The Journal of Finance, vol. 58, no. 4, pp. 1651–1684, 2003.

[148] V. DeMiguel, L. Garlappi, F. J. Nogales, and R. Uppal, "A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms," Management Science, vol. 55, no. 5, pp. 798–812, 2009.