

NHH



---

# Uncertainty modeling and spatial positioning in tramp shipping

---

Vít Procházka

Dissertation submitted to Norwegian School of Economics  
for the degree of Philosophiae Doctor (PhD)

11 September 2018

# Acknowledgment

I would like to express my gratitude to several people that played a big part in my journey towards the completion of this thesis. It was a real privilege to work on my PhD under the supervision of Stein W. Wallace. His teaching vastly shaped my thinking about decision-making under uncertainty, which has not just directed my scientific career, but also had an overlap into my everyday's life. (What else do we do than sequentially making decisions in the presence of the uncertain future?) His enthusiasm and positive attitude combined with his ability to bring up challenging questions created a pleasant, but at the same time very stimulating, environment that every PhD candidate could wish for.

I am also extremely grateful to Roar Ådland for introducing me (in)to a fascinating world of shipping. His knowledge of the industry – combining both theoretical and practical perspective – was a great source of inspiration for many ideas that appear in this thesis. Roar has always been incredibly supportive throughout my years as a PhD candidate, always willing to dedicate his precious time to me – no matter whether on discussing “big ideas for future shipping” or correcting silly grammatical errors in my writings.

I would also like to thank François-Charles Wolff for his invaluable input into the first chapter of the thesis. A massive thank goes to Western Bulk company for funding a research project that this thesis is a part of, and for providing me with an opportunity to spend several weeks at their offices. A special thank belongs to Western Bulk's Chief Risk Officer, Egil Husby, and his team for the hours spent in fruitful discussions.

The working atmosphere in the Department of Business and Management Science was incredibly friendly thanks to my always-in-a-good-mood colleagues who managed to turn a quick coffee refill into a long enriching discussion on far-from-just-rare occasions.

My vacation was always spent on traveling to the Czech Republic, where my roots are. I would like to thank my parents, brothers and other members of my family and all my friends from the country, who supported me – by kind words or occasional visits to the far North. Above all, there are two special Czech ladies who shared my struggles, but many more joyful moments, standing by my side through the whole period of my studies – my wife Gabriela and our daughter Faustýna. I am forever indebted for their support and patience, especially in the last weeks of writing. They created a home I needed in Bergen.

Vít Procházka  
Bergen, 11 September 2018

# Contents

<b>Acknowledgment</b>	<b>2</b>
<b>Contents</b>	<b>3</b>
<b>Introduction</b>	<b>5</b>
<b>1 Contracting decisions in the oil transportation market: Evidence from fixtures matched with AIS data*</b>	<b>12</b>
1.1 Introduction . . . . .	13
1.2 Literature review . . . . .	16
1.3 Data and descriptive statistics . . . . .	20
1.3.1 AIS data and VLCC trading . . . . .	20
1.3.2 Spot fixture data . . . . .	23
1.4 The geography of VLCC fixing . . . . .	25
1.5 The determinants of fixture location . . . . .	31
1.6 Concluding comments . . . . .	37
<b>2 The value of foresight in the drybulk freight market</b>	<b>43</b>
2.1 Introduction . . . . .	44
2.2 Methodology . . . . .	48
2.2.1 Model . . . . .	48
2.2.2 Perfect foresight . . . . .	50
2.2.3 Limited foresight for two-region case . . . . .	56
2.3 Numerical results . . . . .	58
2.3.1 Perfect foresight . . . . .	59
2.3.2 Limited foresight horizon . . . . .	62
2.4 Conclusion . . . . .	65

<b>3</b>	<b>Stochastic programs with binary distributions: Structural properties of scenario trees and algorithms</b>	<b>69</b>
3.1	Introduction . . . . .	70
3.2	Stochastic knapsack problem . . . . .	71
3.3	Out-of-sample evaluation . . . . .	72
3.3.1	Recursive implementation of o-o-s . . . . .	73
3.3.2	Example . . . . .	74
3.3.3	Scenario contribution to the objective value . . . . .	76
3.4	Exact reformulation of the stochastic knapsack problem . . . . .	76
3.5	Dependent case . . . . .	78
3.6	Other applications . . . . .	79
3.6.1	Sampling . . . . .	80
3.6.2	Integer random variables . . . . .	80
3.7	Conclusion . . . . .	80
<b>4</b>	<b>Scenario tree construction driven by heuristic solutions of the optimization problem</b>	<b>85</b>
4.1	Introduction . . . . .	86
4.2	Framework . . . . .	87
4.2.1	Properties of a good scenario tree . . . . .	89
4.2.2	Loss function . . . . .	92
4.2.3	Minimization of the loss function . . . . .	93
4.2.4	Binary distributions . . . . .	100
4.2.5	Feasibility . . . . .	104
4.3	Applications . . . . .	108
4.4	Conclusion . . . . .	110

# Introduction

*“God does not play dice with the universe.”*

Albert Einstein (1879 – 1955), physicist

*“God must have been a shipowner. He placed the raw materials far from where they were needed and covered two thirds of the earth with water.”*

Erling Næss (1901 – 1993), shipping tycoon

Although the presented quotes might give an impression that this thesis deals with some religious topic, in particular God’s aversion to gambling or His potential interest in creating an environment that He can subsequently exploit as a market participant, it is not the case. In both quotes, the main message is carried by the second part and “God” is used only to emphasize the importance that authors assign to the topics they talk about. The topics that are also the main themes of this thesis: *uncertainty* and *shipping*.

Albert Einstein’s statement was written in a letter addressed to Max Born. In the mail correspondence, the two great physicists discussed a, back then, new arising field of quantum mechanics. Einstein could not accept the idea that a fundamental randomness is embedded in natural laws of our universe. That contradicted with his belief in strict causality, which implies certainty and determinism (Isaacson, 2007). He believed that randomness could appear as some form of statistical behavior but could not be a part of the natural law (Natarajan, 2008).

Acceptance of quantum mechanics, and thus perceiving uncertainty as an inherent property of our universe, belongs to today’s mainstream of physics. *Uncertainty* presented in this thesis has a slightly different meaning. It expresses mainly limited capabilities of an individual to predict future events, which might be partly caused by the inherent randomness, but mainly due to limited knowledge and abilities to precisely capture all casual relationships. The individual is aware of these limitations and wants to include

them into a decision-making process<sup>1</sup>. Thus, uncertainty is highly subjective and reflects the individual’s perception of reality rather than an objective truth. As such, it does not contradict with neither belief in causal determinism nor in randomness as a fundamental property of our universe.

To illustrate the subjectivity of uncertainty and its consequences in a decision-making process, let us consider an example of a roulette wheel with standard 37 fields. Without any other prior knowledge, except for basic understanding of the principle of the roulette, one would naturally assign  $\frac{1}{37}$  probability to each field. Then, betting on a single (arbitrarily chosen) number with a potential reward 36 times of the bet is a strategy with a negative expected value. And thus, it is not profitable to bet from the long-run perspective.

An approach that could turn the game into profitability would be to observe the roulette for some time and report the numbers with the hope that after collecting a large data sample, it shall be possible to spot some statistical biases that could turn the edge to the bettor’s side. Such a bias was once discovered by a famous statistician Karl Pearson (1857 – 1936) in roulette numbers reported in Monte Carlo’s newspapers. Eventually, what had first looked like an exploitable opportunity, turned out to be just an inability of humans to construct truly random numbers when making them up – which was easier than observing and reporting true roulette outcomes (Kucharski, 2016). Statistical tests for verifying whether there is enough confidence that an experiment’s outcomes are not based purely on luck, remain incredibly useful in many areas until the present.

A third strategy applicable to the roulette wheel would construct the probability distribution for each trial individually from observing the momentum of the wheel and the ball in the beginning of the spin. That requires a fast calculation of the movements, hardly processable by a human mind. However, a computer can do such a job, as it was conducted by none other than the father of information theory Claude Shannon (1916 – 2001). For this purpose, he constructed the first wearable computer (1960) and together with another mathematician, Edward O. Thorp<sup>2</sup> (1932), analyzed the game and came to a conclusion that the prediction made by the computer could give them approximately 40% edge over the casino (Thorp, 2017).

The main point of the preceding reflection on uncertainty is to demonstrate how the (generally overloaded) term “uncertainty” is perceived throughout this thesis – that is, as an subjective view of some process for which we cannot determine the exact outcome, either due to an inherent randomness or, rather, due to our limited capabilities.

---

<sup>1</sup>Some authors, for example Knight (1921), make a distinction between risk, which can be quantified, and uncertainty, which represents truly unknown outcomes. Our notion of uncertainty corresponds to Knight’s risk.

<sup>2</sup>Edward O. Thorp became famous for description of a winning strategy in blackjack (Thorp, 1966) or foundation of one of the first hedge funds. The fund exploited so-called statistical arbitrage (Thorp and Kassouf, 1967).

*Shipping* has been closely connected with the history of mankind for thousands of years. It played a role in the colonization of continents, as well as in the exchange of goods among them. Even though the industry is more than 5,000 years old, it contributes significantly also to today's economy as it carries more than 90% of global trade (Stopford, 2009). There is no other mode of transportation that can compete with shipping in terms of cost efficiency when transporting cargo from one continent to another. Up to almost 20,000 standard containers can be loaded on the largest container vessels, 400,000 metric tons of iron ore on the largest bulk carriers or 550,000 metric tons of crude oil on the largest tankers.

Shipping segments differ also in the way the vessels are operated. Containers are usually carried by vessels that follow regular schedules. This mode is called *liner shipping*: it is a sea-based analogy to bus lines. In contrast, bulk cargo – either wet (oil and its products) or dry (ore, coal, grain, etc.) – is transported by vessels that are dynamically matched with demand. Thus, the vessels do not follow regular schedules, but satisfy the current needs of the market, similar to taxis on land. This mode is called *tramp shipping*. A third category involves specialized cargo – industry products (for example cars or chemicals) – that do not fit in either of the modes and must be handled in a specialized way.

Stochastic optimization deals with the problem of how to incorporate uncertainty into the modeling of a decision-making process. This thesis focuses only on the most classical approach, where the uncertainty is represented by a set of scenarios (King and Wallace, 2012b). However, how to construct these scenarios from an underlying probability distribution is not always clear and this problem is also addressed in this thesis.

I share the view of, for instance, Deng et al. (2018), that before proceeding to formulate a stochastic optimization model, three steps should precede it. In particular one should obtain (i) hindsight, (ii) insight and (iii) foresight for the problem of interest. Rolling backwards, no matter how complex or sophisticated the optimization model is, the results (obtained optimal decisions) are not very useful if the wrong model inputs are used<sup>3</sup>. That is, we need to have a good foresight. A good (probabilistic) forecast should be built on strong understanding of a process, that is, having good insight into the process. We should determine the key variables and analyze the causal relationships among them. Even in cases where a forecast is constructed only from historical data by some “black box” type of model, there is an additional assumption that patterns observed in the past will repeat in the future. Justification of such an assumption (in some cases it is reasonable, in some not) can also be considered as insight into the process. Insight is almost always gained by analyzing the past (hindsight). Thus, we need to start this complex process by observing

---

<sup>3</sup>Less formal but logically identical version of this sentence, which is often heard among practitioners, is: Garbage in → garbage out.

the historical behavior and analyzing the system.

The general objective of this thesis is the modeling of short-term operational planning in tramp shipping, with a focus on handling uncertainty. The short-term decisions are made either directly by a shipowner operating his own vessel(s) or by a commercial operator, who hires vessels from shipowners and collects margins from sub-letting the vessel. Thus, we use the name “ship operator’s problem” to encapsulate the short-term decisions. The name expresses the isolation from the long-term market exposure that is associated with shipowning<sup>4</sup>.

A commercial ship operator manages a fleet of vessels on different types of contracts varying from a single voyage vessel to time charters of different lengths (and different time/spatial options on redelivery vessels to shipowners). Then, it is up to the operator to find a cargo on the market, negotiate a price for carrying it with the cargo owner (usually through a shipbroker) and make sure that all the requirements (laycan period<sup>5</sup>, capacity constraints, commodity type) to match the cargo with a particular vessel are met. The operator must secure bunker (fuel) for the vessel and communicate with the captain of each vessel regarding many aspects of a voyage (route choice that might depend on the weather, speed choice that depends on the next planned voyage and/or general market condition).

All the described subprocesses require several decisions to be made. Naturally, in order to make good decisions, their consequences need to be taken into account. For example, when selecting a cargo the operator needs to think about the position of the ship after completing the voyage and discharging the cargo: Is there (or will there probably be) another cargo to carry in that region? Or will the ship need to ballast (i.e., going empty) somewhere else? Or will she have to wait? How can regional freight rates and bunker prices change in the future in different regions?

The consequences of all these decisions are influenced by many uncertain parameters (weather, future freight rates, demand, port queues) that are, to some extent, dependent on each other. In order to be profitable in this highly competitive market with relatively low-margins, it is necessary to operate extremely efficiently and make good decisions.

We are not able to address all the subprocesses that an operator deals with and combine them into one comprehensive optimization model with all the uncertain parameters taken into consideration. In fact, this is similar to the real-life situation, where the operator does not deal with all of them in a comprehensive way. Usually different business units take responsibility for particular tasks – chartering managers deal with vessels hiring and cargo

---

<sup>4</sup>But naturally, an operator can own (at least partially) some of the vessels, or a shipowner can operate his own fleet.

<sup>5</sup>A period within which a ship must present herself as ready for loading.



fixing, operations managers with the smooth operation of the ship during the voyage, risk managers keep an eye on the financial exposure, etc.

In this thesis, we focus mainly on the aspects related to spatial positioning of vessels and its relation to the geography of the world. As the main source of uncertainty we consider regional freight rates, trip durations and potential cargo availability. The thesis consists of four papers (chapters) and their order follows the hindsight  $\rightarrow$  insight  $\rightarrow$  foresight  $\rightarrow$  optimization logic.

Chapter 1 provides insight into the contracting behavior of freight market participants in crude oil transportation in the VLCC<sup>6</sup> segment. Our analysis is based on two datasets – satellite positions of ships and a database of commercial fixtures. In the first part of the chapter, we demonstrate trading patterns of VLCCs and different strategic choices that a shipowner or an operator faces. These discussions are based on visualization of geographical positions of vessels in the moment they were fixed (hired for a voyage). In order to determine the positions of fixtures, we match the dataset of fixtures with satellite positions of ships. From the matched datasets, we also extract features that are used in the second part, where an econometric model<sup>7</sup> is introduced. The model empirically assesses the relationship between the distance from the fixture location to the loading port and market conditions and vessel specifications.

This chapter belongs to the hindsight and insight part of the process as it provides a description of the market obtained from the past. It highlights key features, for instance, market conditions, vessel specifications (age, size, ownership), that need to be taken into account when forecasting the behavior of market participants and formulating a model of their strategic choices.

Creating a good predictive model for a process is usually a difficult task. Thus, it makes sense to first estimate in some simple way the potential contribution that the model can bring before it is developed. Such a logic is applied in Chapter 2 of this thesis, where we establish the upper bound for the increase in vessel earnings obtained from decisions about the relocation of a vessel between regions. The upper bound can be estimated by assuming perfect knowledge of future regional freight rates, instant cargo availability and optimization of spatial repositioning of a vessel. This means that any realistic forecast (which should be provided in a probabilistic way) can only lead to worse economic results. Moreover, in real life, cargo does not have to appear at every desirable point in time. Our analysis is performed on the drybulk freight market for three vessel segments (by size).

Except for establishing the upper bound, our analysis also provides insight into spatial efficiency of the market. If the upper bound was too close to average earnings, it would

---

<sup>6</sup>Very Large Crude Carriers (VLCCs) are the largest tankers with capacity exceeding 200,000 DWT (deadweight tonnage). They are only involved in the transportation of crude oil.

<sup>7</sup>The empirical model (Section 1.5) is mainly formulated and fitted by our French co-author François-Charles Wolff. My input in this part is very little, mostly just providing the data for the model.

suggest that the market is spatially so efficient in the pricing of transportation that investing into sophisticated predictive analysis and optimization models is pointless as there are basically no exploitable opportunities. This is, however, not what we find and we believe there is space for improvement of economic performance, and thus making the market more efficient.

Therefore, a natural step would be to build a more complex model of the operator's decision-making process that takes into account several features that are omitted in Chapter 2. In particular, we should consider different scenarios for future freight rates, include scenarios for bunker prices (which also differ across regions), include more regions (which is important especially for smaller segments) and model cargo availability. Whether a cargo of some particular attributes is available at certain region at a future time point (in the case the time is discretized as in Chapter 2) can be modeled by a Bernoulli (binary) distribution. This means to combine several random variables, each of them with a different character (continuous, binary), in one application. We found a lack of scientific literature dealing with the handling of binary distributions within stochastic programming models. Specifically, the issue of generating scenarios from binary distributions (let alone a combination with other distributions) is not addressed in the literature. Chapters 3 and 4 contribute to this part of the literature in a general way.

Chapter 3<sup>8</sup> provides two simple, but effective, procedures for stochastic programs with binary distributions. The procedures are designed for problems with a special structure using penalization, which is quite common within stochastic programming. A typical example is a stochastic knapsack problem, which is also used in the chapter for demonstration. The first procedure enables effective out-of-sample evaluation of a solution by using only necessary scenarios – those that produce a penalty. The second procedure enables reduction of scenarios (before the problem is solved) to a minimal number of scenarios needed. Both procedures are based on the same principle – we generate scenarios in a recursive manner and stop when a penalty is no longer generated. We prove that by this procedure, it is possible to reformulate the problem into a version with a minimal number of scenarios. Thus, this is an exact approach as the optimal solution of the reduced problem is also the solution of the original problem.

Even if the number of scenarios is reduced to the minimal number of scenarios for the exact reformulation, the number of scenarios could still be too large to find a solution within a given time (or even to store all the necessary scenarios). In such a case, we search for a representation of the uncertainty by a smaller set of scenarios. The process of choosing this subset is called scenario generation and even though there are many methods for generating scenarios for continuous distributions, we are not aware of any method suitable for binary distributions due to their specific statistical properties.

---

<sup>8</sup>The paper is published in *Computational Management Science* journal (May 2018).

In Chapter 4, we introduce an original framework for scenario generation that is not based on any statistical measures of the distributions that many other methods use, but on minimizing discrepancy between in-sample and out-of-sample performance of a pool of heuristically obtained solutions. The framework is not, in principle, limited to any distribution, and thus, can be applied on binary distributions, as it is demonstrated in one of the sections, or on a mix of continuous and binary distributions. Thus we see a potential in usage the framework for generating scenarios for the ship operator’s problem (and many other areas that are listed in the chapter), if we wanted to model uncertain future cargo availability and future freight rates at the same time.

## References

- Deng, Y., Liu, J., and S., S. (2018). Coalescing data and decision sciences for analytics. *INFORMS, Tutorials in Operations Research*.
- Isaacson, W. (2007). *Einstein : His Life and Universe*. Simon & Schuster Ltd.
- King, A. J. and Wallace, S. W. (2012). *Modeling with Stochastic Programming*. Springer Series in Operations Research and Financial Engineering. Springer.
- Knight, F. H. (1921). *Risk, Uncertainty and Profit*. The Riverside Press, Boston and New York, Houghton Mifflin Company.
- Kucharski, A. (2016). *The Perfect Bet: How Science and Math Are Taking the Luck Out of Gambling*. Basic Books.
- Natarajan, V. (2008). What Einstein meant when he said "God does not play dice...". *Resonance*, page 655–661.
- Stopford, M. (2009). *Maritime Economics, third ed.* Routledge, London.
- Thorp, E. (1966). *Beat the Dealer: A Winning Strategy for the Game of Twenty-One*. Vintage.
- Thorp, E. (2017). *A Man for All Markets: From Las Vegas to Wall Street, How I Beat the Dealer and the Market*. Random House.
- Thorp, E. and Kassouf, S. T. (1967). *Beat the Market: A Scientific Stock Market System*. Random House.

# Contracting decisions in the oil transportation market: Evidence from fixtures matched with AIS data\*

Vit Prochazka<sup>□</sup>    Roar Adland<sup>□</sup>    François-Charles Wolff<sup>△</sup>

<sup>□</sup>Department of Business and Management Science  
Norwegian School of Economics, Bergen, Norway

<sup>△</sup>LEMNA, University of Nantes, Nantes, France

## Abstract

In this paper, we investigate the contracting behavior of participants in the spot freight market for tankers by analyzing the positioning of vessels at the time of fixture. For that purpose, we create a new dataset obtained by merging spatial ship positions, commercial fixtures and technical vessel specifications. Using quantile and quantile fixed effect regressions, we show how market conditions, vessel characteristics and charterers' preferences affect the fixture location. Our main result is that oil buyers secure tonnage earlier during strong tanker markets. We also find that the geography of trade creates natural decision points that dominate in the spatial distribution of fixtures.

**Keywords:** international shipping, spot market, AIS data, oil transportation

---

\*We are indebted to participants to the 25th Annual conference of the International Association of Maritime Economists (IAME) held in Kyoto, Japan (June 2017). We are grateful to the UCL Energy Institute and exactEarth for the provision of downsampled AIS data. The spot fixture data used in this paper may be purchased through Clarkson Research. Any remaining errors are ours. This research was partly financed by the Research Council of Norway under the project "Smart digital contracts and commercial management", project number 280684.

## 1.1 Introduction

International transport of crude oil takes place mainly on oil tankers, with annual seaborne crude flows totaling approximately 14.2 billion barrels per annum (Clipper Data, 2016). This compares to annual global crude oil supply of about 36 billion barrels IEA (2018). The only notable competing transportation mode is pipeline, which is important for certain exporters (for instance for Canadian and Mexican exports into the United States), although it can also be used in parts of the international supply chain in combination with seaborne transport (Adland et al., 2017).

According to Clipper Data (2016), the key exporting areas for seaborne crude oil transport are the Middle East (Saudi Arabia, UAE, Iraq, Kuwait, Iran, Qatar, Oman) with approximately 6.1 billion barrels per year (bbpy), Russia (1.4 bbpy), West Africa (Nigeria, Angola) with approximately 1.3 bbpy and Venezuela (0.7 bbpy)<sup>1</sup>. Because of the natural dominance of large national and international oil companies, the crude oil seller side is relatively consolidated, with the top 10 shippers in the seaborne crude oil market accounting for 55.4% of global seaborne shipments by volume. The oil buyer side is somewhat less consolidated, with the top 10 names accounting for 34.8% of seaborne volumes, though we note that the large players in the market can appear as both crude oil buyers and sellers due to regional imbalances in their internal supply chain or to opportunistically take advantage of spatial price arbitrages<sup>2</sup>.

Most crude oil is still purchased on a Free-on-Board (FOB) basis, which means that the oil buyer is in charge of arranging and paying for transportation. As opposed to the oil market itself, the global market for the chartering (i.e. hiring) of tanker vessels is highly fragmented and decentralized. To illustrate this, we note that the approximately 14.2 billion barrels of crude oil transported internationally by sea in 2015 were shipped on 2260 unique vessels owned by 536 different shipping companies (Clipper Data, 2016; Clarkson Research, 2018). Overall, the crude oil tanker fleet exceeds 428 million deadweight tonnes (DWT) and comprises roughly 21% of the overall fleet of ocean going vessels by DWT (Clarkson Research (2018)). Oil tankers are matched with cargoes in a decentralized voice-brokered market using specialized shipbrokers. Contracts for transportation (charters) are entered into (fixed) for either single voyages in a spot market or a longer period of time under a timecharter. Despite the dominance of long-term offtake agreements for the supply of crude oil, with an estimated 90% of volumes traded on long-term contracts (Schofield, 2007), most of the transportation demand is served by spot voyage charters.

---

<sup>1</sup>All numbers in this section refer to 2015 and are based on aggregating reported cargoes in the Clipper Data database.

<sup>2</sup>According to Clipper Data (2016), the top ten crude oil shippers in 2015 were Saudi Aramco, Transneft, PdVSA, Somo, Petrobras, KPC, ExxonMobil, Botas, Total and BP, with the top ten buyers comprising Sinopec, Petrobras, Shell, Reliance, EGPC, PdVSA, Vopak, Valero, ExxonMobil and Repsol.

The tanker spot market for crude oil transportation has some characteristics which makes it an interesting candidate for the analysis of timing and location decisions. Firstly, there are obvious separate producing and consumption regions, typically located far apart, leading to a stable and well-defined international trading pattern with a clear fronthaul-backhaul structure. In particular, tankers will typically sail fully laden in one direction and empty (in ballast) when returning to one of the loading areas. As a consequence of the limited number of trading routes and typical lead times between contracting and loading, there will be certain points in time and space where ships are more likely to be looking for new employment and, subsequently, be fixed on a new contract.

Secondly, transportation is arranged in a decentralized spot market with a large number of buyers and sellers, with imperfect information about the availability of cargoes, ships, transactions and the contemporaneous choices of competitors, at least from the viewpoint of a single participant. As a consequence, the market functions like an unsupervised logistics system, where choices about where to locate your fleet and when to enter into new contracts have to be made under uncertainty in what is effectively a “matching game”. Thirdly, the transportation cost is a very small share of the value of the oil cargo and the cost of disruption in the crude supply chain (shutting down a refinery) is very high, leading to highly inelastic demand for both the crude oil and its transportation (Stopford, 2009). In addition, the short-run production of crude oil is rather insensitive to variations the oil price as the marginal cost of production is very low, with additional constraints on shore-side storage facilities in loading areas (i.e. a “supply push”). As a consequence, the global trade in crude oil, both in terms of volume and timing of cargoes, is largely exogenous to the tanker freight market.

While crude oil flows may be largely exogenous to the freight market, the supply and demand for transportation in the freight market is definitely not. Zannetos (1966) was the first author to describe the importance of expectations in the interplay between charterers and shipowners, introducing the concept of “intertemporal substitution”. This principally refers to the observation that both shipowners and charterers have some flexibility in when to enter the market with a ship and cargo, respectively, which is also the ultimate source of the strategic value of timing and positioning in our setup. While an oil buyer (charterer) must commence loading the cargo in a certain time window (termed “laydays”), he can secure the corresponding transportation service from any of the ships that are positioned such that they can meet this loading deadline and that are not already under contract (an “open” ship). Conversely, a ship should generally obtain a new contract by the time it reaches the loading area, but long ballast voyages (23 days from Japan to Saudi Arabia at prevailing speeds, for instance) gives the owner considerable flexibility in terms of when to approach the market for new business.

Zannetos (1966) further describes how mere expectations about future prices (i.e. spot

freight rates) can lead to “self-fulfilling prophecies” and contribute to the wild gyrations observed in spot freight rates for tankers. In the presence of intertemporal substitution, the joint expectation that short-run rates will increase will induce profit maximizing owners to delay negotiating (i.e. reducing immediate supply) and cost-minimizing charterers to enter negotiations early (i.e. increase demand), with the effect of increasing freight rates, all else equal. It follows that we may see very large short-run price fluctuations with no fundamental change in the overall number of cargoes to be shipped or the positioning of the fleet, based on expectations only. Consequently, the timing of freight negotiations and subsequent fixtures can have a large impact on the earnings for a voyage and the cost of oil transportation and is, as such, a crucial choice for both shipowners and charterers.

While Zannetos (1966) referred to the timing of contracts, such timing is by definition interlinked with the geographical location of the vessel. Aside from the obvious one-to-one relationship between the remaining distance and time to the loading port at constant sailing speeds, there are at least three more reasons for this.

Firstly, whenever there is a choice between competing load areas (between West Africa and the Middle East for a ship returning from Asia, for instance), there are constraints on when the final decision has to be made. These constraints are a result of the economic trade-off between the cost of deviating from the most direct routing to the load ports (i.e. additional fuel consumption and the alternative cost of time) and the additional profit from delaying the decision and gaining from a higher freight rate in one of the loading regions. Secondly, there may be geographical constraints on where ships can elect to wait for orders during the return ballast voyage (i.e. safe, low-cost anchorages, ideally with access to refueling and crew repatriation services). Thirdly, once the ship has returned to the loading area and become a “prompt” ship (i.e. it is unemployed and waiting, albeit possibly voluntarily in the hope of higher rates in the near future), the cost of relocating elsewhere at the owners expense becomes prohibitively expensive due to the long distances at play, and so the ship becomes part of a captive idle fleet for as long as it takes to clear out the local oversupply. The latter point also highlights the relationship between charterers’ preferences and the location of fixtures, as attractive vessels should on average be “first picks” and therefore fixed early (in space and time).

Until now, the hypotheses discussed above have not been empirically tested due to the unavailability of accurate ship tracking data. In this paper, we make use of the improved availability and coverage of ship position data from the global Automated Identification System (AIS) as supplied by satellite-based receivers to study the location of vessels at the time of fixture. We combine AIS data with fixtures data in order to get the geographical positions of vessels when they are reported as fixed. To the best of our knowledge, our paper represents the first ever empirical analysis of decisions concerning both the timing and location of fixtures in the chartering market. Our main contributions are threefold.

Firstly, we explicitly test a version of Zannetos’ intertemporal substitution hypothesis according to which strong (weak) freight markets are associated with earlier (later) fixtures. Secondly, we show that certain geographical locations naturally dominate in the spatial distribution of fixtures because they represent either safe anchorages near loading areas or strategic “decision points”. Thirdly, we test the hypothesis that certain ship characteristics make the vessel more attractive in the spot market and increase the probability that the vessel is fixed early.

Our findings are important for the improved modeling of spot freight markets on the basis of spatial data, particularly forecasting applications, in both a commercial and academic setting. Principally, freight market models that use ship positions as a proxy for supply and cargo shipments as a proxy for demand fail to recognize the impact of intertemporal substitution and geography on short-run freight rate formation. Similarly, our research improves the understanding of spill-over effects between regional markets and provides a building block for more advanced modeling of discrete routing choices in bulk shipping chartering.

The remainder of this paper is structured as follows. Section 1.2 summarizes the relevant literature. Section 1.3 describes the data. The geography of VLCCs fixing is presented in Section 1.4. We investigate the determinants of fixture location in Section 1.5. Finally, Section 1.6 concludes and suggests areas of future research.

## 1.2 Literature review

There are mainly three concepts from maritime economics that are relevant for our work: risk preferences and risk premia, spatial efficiency in the freight market, and models of spot freight rate formation.

There is clearly a close relationship between charterers’ and shipowners’ risk preferences and their behavior in the spot market. For instance, a risk-averse charterer who is exposed to large potential costs of disruptions in the crude oil supply chain may want to secure tonnage as early as practicable<sup>3</sup>. Conversely, a risk-loving tanker owner may take a calculated risk and prefer to fix its vessel as late as possible in order to catch a short-term peak in rates. Kavussanos and Alizadeh (2002) and Adland and Cullinane (2005) point out that this risk of transportation shortage is time-varying and exists only during strong freight markets. While there are no academic studies on the risk preferences of oil buyers, it seems reasonable to postulate that most are risk averse. Having to withdraw a cargo from the market, delay the shipment or charter a sub-standard vessel because there is no alternative could potentially have large monetary consequences for a charterer, for

---

<sup>3</sup>Anecdotally, major charterers care mostly about paying a market rate that is no higher than their main competitors. See Adland et al. (2016) for a study on the impact of charterer identity on VLCC spot freight rates.



instance due to a resulting breach of contract, refinery downtime, or a loss of reputation (Adland and Cullinane, 2005). As a consequence, we expect charterers to try to fix earlier when spot freight rates are high.

The time-varying nature of risk preferences is well established in the literature, with Friedman and Savage (1948) first presenting the theoretical argument that investors' attitudes towards risk are governed by their liquidity situation. A survey of the risk attitudes of Norwegian shipowners by Lorange and Norman (1973) suggested that the majority of the players were risk loving or risk neutral when their liquidity was good (strong freight markets), but risk averse otherwise. Eckbo (1977) repeated the analysis in a market that had been depressed for a few years and found that there had been some changes in the attitudes towards risk. In the case of good liquidity, half the group showed risk-loving behavior whereas the second half was risk averse. The group was mainly risk averse, however, if the participants were exposed to a liquidity shortage. Overall, the results imply that shipowners as a group have decreasing absolute risk aversion with respect to wealth, assuming that the state of the freight market is a good proxy for the latter (Cullinane, 1991).

Because of the potential for intertemporal substitution (Zannetos, 1966), the time-varying and, more importantly, market-dependent risk preferences among shipowners and charterers are likely the main source of heterogeneity in the timing of fixtures. Based on the theoretical arguments presented and findings in the literature, (Alizadeh and Talley, 2011a,b), we expect a general shift towards early negotiations and fixing activity during strong freight markets, and later fixing in weak markets where oversupply of open ships allow even highly risk-averse charterers to have patience and secure cheap transportation closer to the loading window. In addition, we expect that heterogeneous risk attitudes translate into company-specific fixed effects in chartering behavior (as proxied by the timing/location of fixtures). Finally, the geography of the trading routes, with established waiting areas and "decision points" for the routing of vessels, is expected to lead to singularities in the distribution of fixing activity (when measured by remaining distance).

The modeling of spot freight rate formation in bulk shipping has long been a fascination of maritime economists. Early research focused on structural or reduced-form equilibrium models (see Tinbergen (1934); Koopmans (1939); Zannetos (1966); Eriksen and Norman (1976); Hawdon (1978); Norman and Wergeland (1939); Wergeland (1981); Charemza and Gronicki (1981); Beenstock (1985); Strandenes (1986); Beenstock and Vergottis (1989); Evans (1994)). Later studies adapted univariate stochastic models typically borrowed from the finance literature (see Bjerksund and Ekern (1995), Tvedt (1997), Adland and Cullinane (2005)). The former group of models were deterministic and, thus, at best able to capture some of the long-run variations in market conditions (typically trying to explain annual average spot rates, for instance). The second group ignores the underlying

fundamental market information altogether and considers the spot freight rate history only. Adland and Strandenes (2007) propose a stochastic extension of the classical partial equilibrium models of the spot freight market in an attempt to bridge the two approaches.

However, the notion of “intertemporal substitution” (Zannetos, 1966) of cargoes and ships creates dynamics that these models are unable to account for. Once we introduce flexibility (in time and space) in when to enter the spot market for matching a cargo or a ship, fundamental information such as fleet size and loaded cargo volumes need not be a major influence on short-term spot rate dynamics. Instead, heterogeneous and time-varying expectations, risk preferences, and the relative bargaining power of shipowners and charterers start to play a role. As an example, a Markovian stochastic spot freight rate model cannot account for the well-known presence of short-term positive autocorrelation in spot freight rate dynamics (Benth and Koekebakker, 2016).

The literature dealing with this behavioral part of the spot freight market is limited and mainly theoretical. Tvedt (2011) considers the psychological aspects of the Very Large Crude Carrier (VLCC) market using an assignment model with an exogenous freight rate that varies according to the bargaining power between the shipowner and charterer. Parker (2014) develops a comprehensive simulation model for matching in the VLCC market and finds that agents’ opportunity cost and future expectations influence the matching and contract prices. Moreover, ships’ physical characteristics affect both costs and the charterers’ willingness to pay. Also, varying location and physical characteristics show that ships which are the most favored by physical characteristics cannot compete as strongly with less preferred ships located closer to the loading area (Parker, 2014).

Our empirical work is based in part on ship positions recorded from AIS data, a system which was originally conceived for collision avoidance. The use of AIS data in maritime economic research is still in its early stages, with applications mainly limited to emission accounting Smith et al. (2014) and studies of vessel speeds (Aßmann et al., 2015; Adland and Jia, 2016, 2018). Adland et al. (2017) compare global crude oil trade statistics derived from customs data with those derived from the bottom-up AIS tracking of crude oil shipments. They find that overall there is good alignment in volumes, suggesting that AIS-based trade volumes are reliable, but that temporal and spatial deviations occur due to pipeline transport, temporary storage and transshipment. In related work, Jia et al. (2017) propose an automatic algorithm for generating seaborne transport pattern maps based on AIS data.

The literature on microeconomic analysis of fixture data can assist in identifying the physical characteristics that make ships more or less attractive. Notable papers in this area include, for instance, Tamvakis and Thanopoulou (2000), Köhn and Thanopoulou (2011), Agnolucci et al. (2014), Adland et al. (2016, 2017). While researchers have almost exclusively dealt with the impact of vessel and voyage determinants on contracted rate

levels, many of the variables proposed in the literature, such as vessel age, ownership, flag, and carrying capacity, should also proxy the vessel’s overall attractiveness in the chartering market. An important point here is that of simultaneity, that is, whether there is a relationship between the position at which a vessel is fixed and the rate it obtains as highlighted by the simulation model of Parker (2014). Using reported fixtures, Alizadeh and Talley (2011a,b), model the interrelationship between the lead time (between fixture and loading) and the spot freight rate in a system of simultaneous equations. The results for the tanker market in Alizadeh and Talley (2011b) suggest that ships are fixed earlier during times of high freight rates and lower volatility<sup>4</sup>. The latter is somewhat counterintuitive, but may be related to the expectedly high correlation between volatility and levels (see Adland and Cullinane (2006)) and resulting issues with multicollinearity.

In the present paper, we use the recent availability of AIS data to build on the empirical results in Alizadeh and Talley (2011b). Specifically, estimating the average effects of vessel and voyage determinants in the time dimension only, as done in Alizadeh and Talley (2011a,b), will ignore the finer details in how the full distribution of lead times depends on market conditions and company risk preferences. Perhaps more importantly, such an approach also ignores the link between the distribution of vessel fixing in both space and time and the geography of seaborne trade (such as anchorages and routing decision points). There are also important data quality challenges when relying solely on fixture data for empirical work as in Alizadeh and Talley (2011a,b). The main problem is that fixture data reports the intended loading dates for a vessel, which may be several weeks ahead. There is no assurance that the actual loading takes place during the expected time window, for instance due to cargo delays or bad weather affecting ship arrivals, in which case the reported lead time is incorrect. Nor are all fixtures realized as intended and may be canceled shortly after reporting if the contract was still “on subjects”.

By matching fixtures and the physical position and subsequent routing of the vessel, we are able to verify that a contract was performed as intended and improve the quality of the estimates using realized data points. In general, the impact of geography and trade patterns on market dynamics has received comparatively little attention in the maritime economic literature, with notable exceptions being the works of Laulajainen (2007, 2008, 2011). Our study clearly contributes to link the research streams of maritime geography and economics.

---

<sup>4</sup>The lead time is positively related to the spot freight rate level, hull type (double hull vs single hull) and DWT utilization, and negatively related to vessel age and freight rate volatility.

## 1.3 Data and descriptive statistics

### 1.3.1 AIS data and VLCC trading

The tanker fleet is divided into several size classes based on vessels' carrying capacity, measured in deadweight tonnes (DWT). For crude oil tankers, the main segments are referred to as VLCCs (above 200,000 DWT), Suezmax (around 130,000 – 160,000 DWT), Aframax (around 70,000 – 120,000 DWT) and Panamax (around 50,000 – 70,000 DWT) (Stopford, 2009). In this paper, we focus on spot fixtures of the largest size class of tankers (VLCCs) as they dominate the long-haul routes of interest here. For instance, Adland et al. (2017) report that VLCCs loaded 88% of all crude oil exports from Saudi Arabia, 58% of Angolan exports and 35% of Venezuelan exports in 2015. Suemax and Aframax tankers are typically used on regional trades (in particular in ex-Russia and the North Sea) where they are not subject to the same location choices. The trade of VLCCs is further restricted to only a few main routes due to restrictions in ports regarding the size and maximal allowed draught of a vessel.

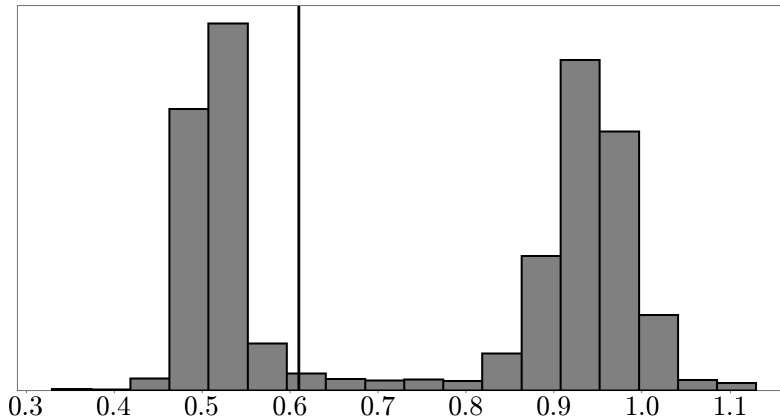


Figure 1.1: Histogram of draught ratios

Source: Authors' calculations, AIS data.

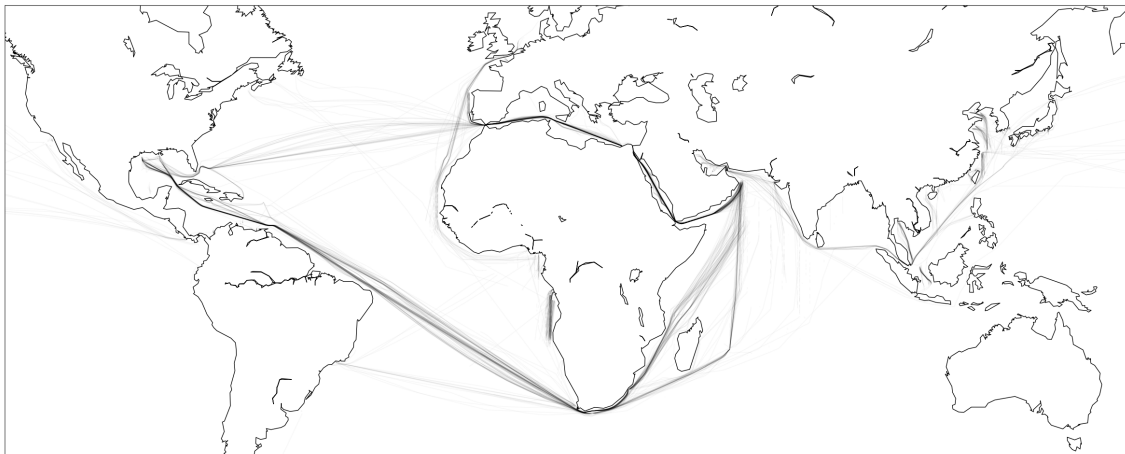
To determine the geographical position of a fixed vessel and illustrate main trading routes, we use AIS data from the period January 2013 – July 2016. AIS data include the location (longitude and latitude) of the ship, its unique IMO (International Maritime Organization) number, vessel name as well as the current draught and speed<sup>5</sup>. However, since the AIS-reported draught is not always reliable (due to manual input), we first assign positions to single trips and then compute the average draught over the trip. The average draught for the voyage is divided by the design draught for each vessel. Figure 1.1 shows the bimodal distribution of resulting draught ratios across our fleet. We classify the voyage as laden if the average draught ratio exceeds 0.6 and note that this will include some part-

<sup>5</sup>The positions in the AIS dataset are reported every three hours, but there are a few gaps in the data due to poor coverage, typically due to signal disruptions in high-traffic areas such as the South China Sea.

laden voyages. However, since we use this classification only for visualization purposes in Figures 1.3 and 1.4, we do not add more sophisticated features that could improve the accuracy of the classification<sup>6</sup>. Note also that the laden/empty classification is not used in the econometric analysis presented in Section 1.5, as ships can be fixed in both conditions.



(a) Laden Eastbound



(b) Laden Westbound

Figure 1.2: Worldwide laden voyages for VLCCs.

Source: authors' calculations, AIS data.

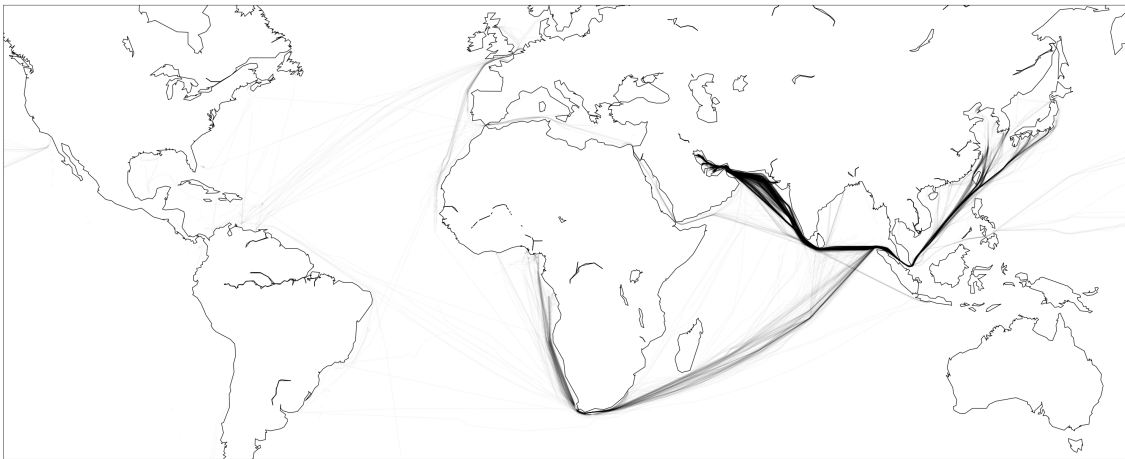
Figure 1.3 illustrates the trading pattern of the 670 VLCCs in our AIS database in 2014 by direction (westbound/eastbound) and loading condition. The routes are highlighted by drawing a thin partly transparent line between subsequent observations, with the direction of the vessel movement determined by comparing longitudes of consecutive positions<sup>7</sup>.

<sup>6</sup>For example, we could assess whether a port of origin (destination) belongs to loading or discharging ports for each trip. However, in our sample ships are fixed both when they are known to be laden and in ballast.

<sup>7</sup>The scale of the transparency is chosen for each figure independently.



(a) Ballast Eastbound



(b) Ballast Westbound

Figure 1.3: Worldwide ballast voyages for VLCCs.

Source: authors' calculations, AIS data.

In Figure 1.2a, we can recognize Persian Gulf (PG) as the main exporting area of crude oil, which is transported mainly to east Asia (China, Singapore, Japan), west coast of India, Pakistan or to the US Gulf (Figure 1.2b). Additional eastbound cargo flows appear from West Africa, Venezuela and, to a much lesser extent, the North Sea. From the US Gulf, vessels often ballast to Venezuela, which is basically the only significant eastbound flow of empty vessels (Figure 1.3a), where crude oil is loaded and transported to a similar set of destinations as in the PG case. The third exporting area that we focus on is West Africa, to which empty vessels mainly ballast from East Asia (Figure 1.3b). More detailed analysis of movements and the different strategies of shipowners and operators are considered after merging AIS data with the dataset of fixtures.

Figure 1.3b depicting ballast voyages points to the importance of geography in the

choice of where to wait. The key decision point in our observations is clearly the Singapore area because of its strategic location. The shortest path from East Asia (Japan, China, Singapore), where most vessels discharge their cargo, to all main exporting areas go through Singapore. Thus, for a ship that is not already fixed, it is possible to wait here for a contract that can come from all exporting areas and still ballast the minimum distance. An alternative decision is to not waste time (and potential earnings) by waiting, but continue to a particular loading area even without a contract. However, this strategy comes with the risk that no contract becomes available before arrival or that it turns out that an alternative loading area would have provided higher voyage profits for the subsequent voyage. Of course, the location decision is reversible, but only at a cost of additional fuel burn and lost alternative revenue.

### 1.3.2 Spot fixture data

We analyze the dataset of spot fixtures for VLCCs provided by Clarkson Shipping Intelligence Network for the same period that we have AIS data (from January 2013 to July 2016)<sup>8</sup>. The data include basic vessel characteristics like name, deadweight tonnage, reported fixture date, the laycan period within which a ship must present herself as ready for loading, origin and destination of the transportation service, etc.

Without filtering, the database of fixtures includes 6179 transactions for VLCCs during this time period. However, as the IMO number is not included in the Clarkson's fixtures dataset, we must rely on the name of a ship for the matching with AIS ship position data. Since the vessel name can change multiple times during the lifetime of a ship, typically when it changes ownership, we are not able to match all records. In addition, though ignored in the literature, not every reported fixture is eventually realized. For further analysis, we consider only realized fixtures, by verifying whether the particular vessel appeared in a loading area within the stated laycan period, with a tolerance of  $\pm 3$  days. One of the aspects we consider in our analysis is the waiting time before loading (for example a vessel might spend several days in Fujairah anchorage), meaning that we must filter out cases with a significant gap between the first appearance in the loading area and the preceding position, since we are not always able to establish (or estimate with a reasonable accuracy) the exact time a vessel enters the loading area<sup>9</sup>.

After filtering, we focus on the three main export areas for which we match AIS and fixture data. Our final dataset includes 2029 fixtures for Persian Gulf, 464 fixtures for West Africa and 217 fixtures for Venezuela. Our main variable of interest is the distance to loading port. Due to the gaps in the AIS dataset, it is sometimes impossible to determine

---

<sup>8</sup>See [www.clarksons.net](http://www.clarksons.net).

<sup>9</sup>For example, a ship may be observed close to the west coast of India and the next AIS position is in PG, where it is supposed to load a cargo. Such an observation is excluded from the dataset.

the exact position of the vessel at the moment of fixture. In such a case, we interpolate the nearest positions before and after the date of the fixture by approximation of the vessel movement. We assume that the vessel is moving along the shortest path on a purpose-made mesh of the ocean (see Appendix for more details). Simple approximation by a line between two consecutive positions would typically not work since it might cross land.

The remaining distance to the load port is measured directly from the AIS data by calculating the cumulative sum of distances between two consecutive positions in the AIS dataset from the date of the fixture until the ship enters the corresponding loading area. For Venezuela and West Africa, we add an approximate distance from the first point of entry to the loading area to the “spatial centre of gravity” of the loading terminals. This approach of “approximating the last step” is applied because it is generally difficult to determine the exact terminal for loading. For instance, it is hard to distinguish between an offshore loading area (buoy), ship-to-ship transfer at sea, and an anchorage area where ships wait for contracts or terminal allocation. Moreover, gaps in the AIS dataset might make it impossible to observe the period when the ship was stationary. In the case of PG the loading area depicted by the rectangle is relatively small and so we measure the distance from the point of fixture to a first entry to the rectangle<sup>10</sup>.

Table 1.1 presents the descriptive statistics of the matched and verified fixture data. The average distance is 2897.2 miles in the Persian Gulf, with a standard deviation of 1535.8 miles. The average distance is 7272.9 miles in West Africa and 4656.6 miles in Venezuela, respectively. We include in Table 1.1 vessel age and deadweight for each geographic area. The average age of VLCCs fixed for the Persian Gulf (8 years) is higher than in West Africa and Venezuela (around 6.5 years). The average deadweight is around 300,000 in all areas.

An important indicator in our analysis is the general state of the tanker spot freight market. In line with previous studies (Alizadeh and Talley, 2011b), we consider the Baltic Dirty Tanker Index (BDTI), which is a daily indicator produced by Baltic Exchange that includes several of the most traded routes across different sizes of vessels<sup>11</sup>. In our specific analysis of the Persian Gulf, we use the BDTI TD2 indicator which is the index for the route Middle East-Singapore.

---

<sup>10</sup>There are, however, some exceptions that we need to treat in a special way. In general, more complicated behavior is observed inside the Persian Gulf. For example, we observe visits to multiple terminals or a relocation to Fujairah anchorage and back to PG for loading. We check for the possibility that a ship is fixed in PG, but manage to make a short trip within the laycan period, for example, to Karachi (Pakistan) and back. In such a case, the distance is measured and added.

<sup>11</sup>More information on Baltic Indices, including the methodology of the calculation, can be found in Laulajainen (2008). See also <https://www.balticexchange.com>.



Variables	Persian Gulf	West Africa	Venezuela
<i>Distance to loading port</i>			
Average	2897.2	7272.9	4656.6
Standard deviation	1535.8	1837.4	2605.7
Median	3168.6	7568.8	4443.0
<i>Control variables</i>			
Vessel age	8.0	6.6	6.5
Vessel deadweight	307,270	306,641	311,508
Baltic tanker index	54.6	54.1	54.4
Number of observations	2029	464	217

Table 1.1: Descriptive statistics of fixtures, by geographic area.

Source: authors' calculations, 2013 – 2016 data from Clarksons and AIS data.

## 1.4 The geography of VLCC fixing

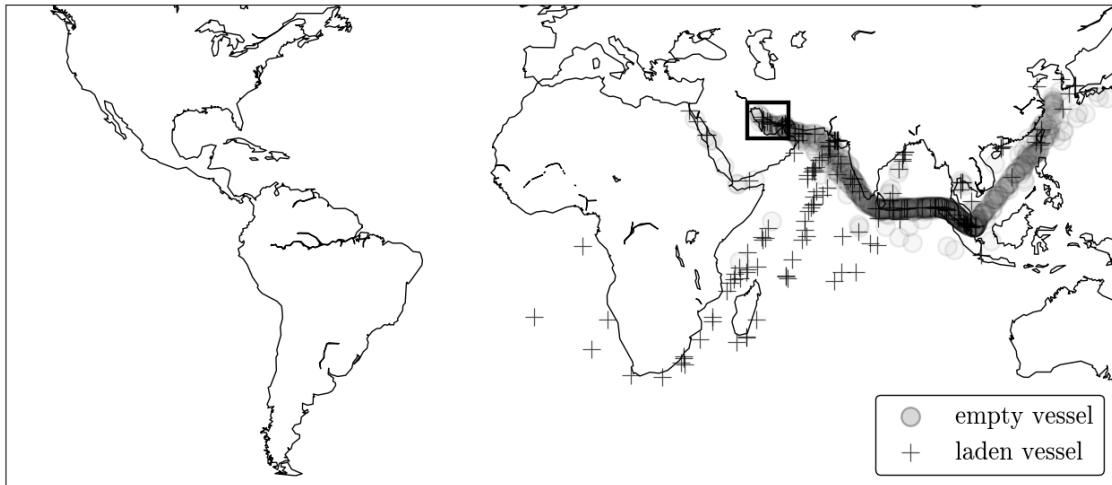
To get a visual overview of geographical distributions of fixtures, we display all assigned fixtures in a map in Figures 1.4a, 1.4b and 1.4c. Fixtures that are made while a ship is laden are denoted by a cross symbol, and fixtures done by ballasting ships by a circle. The circles are partly transparent, thus an increased intensity of the color corresponds to a higher number of fixtures in a certain place<sup>12</sup>.

From a distance point of view, vessels in the laden condition are still sailing away from the loading area and towards the discharge port, or they are potentially fixed while at anchorage or alongside the terminal in the discharge port. Although fixing the vessel while still under contract with the previous cargo represents some risks for the shipowner, mainly in case of unexpected delays leading to a missed laycan in the next loading port, our data suggests that it is fairly common in the ex-Venezuela trade (Figure 1.4c). This is because the ballast trip from the discharge area, typically in the United States, to Venezuela is very short.

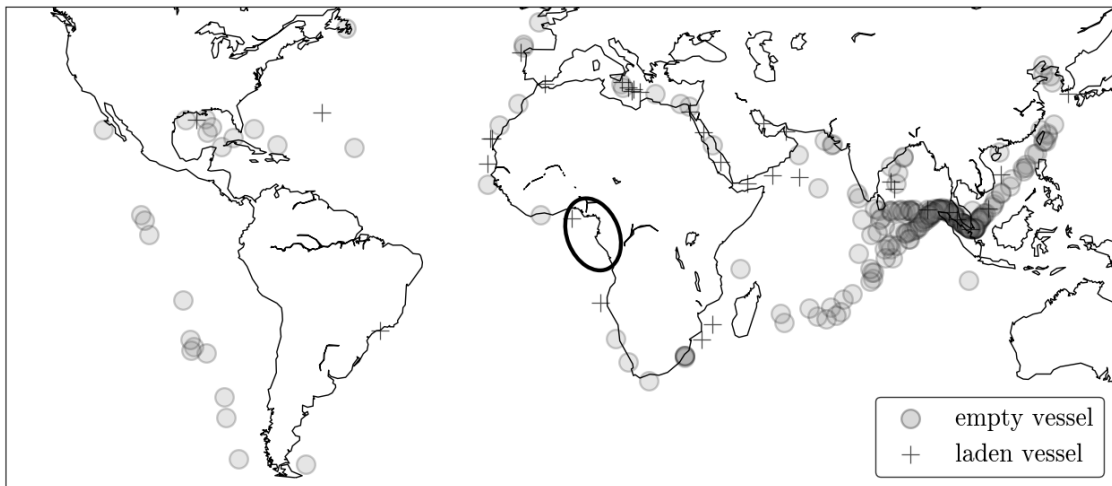
As shown in Figure 1.4a, most fixtures for cargoes in the Persian Gulf area are done during the ballast trip back from Asia. We observe two main mass points in the number of fixtures. The first peak is observed in Persian Gulf, mainly at Fujairah anchorage. It consists of unfixed vessels that had to ballast all the way back to Persian Gulf and wait for a contract. The second peak is in the Singapore area, which is a result of several phenomena. First, Singapore as such is an important destination for crude oil so that many vessels get a new contract while discharging a previous cargo. Second, if the ship does not have a new contract and the discharging process is finished, it can either ballast

<sup>12</sup>The scale of the transparency is chosen for each figure independently, i.e., it is specific to each area.

back to the Persian Gulf or West Africa with the hope of getting a contract later, or it can wait in the Singapore area. The advantage of waiting is that the ship does not lose the option of going to all other areas (mainly West Africa or Venezuela). Clearly, these waiting ships contribute to the peak of fixtures in the Singapore area. Third, even for vessels returning from Japan or China, it makes sense to wait in the Singapore area on the ballast trip back to the Persian Gulf. Again, in doing so, the vessel keeps the options of going to other export areas without wasting fuel<sup>13</sup>.

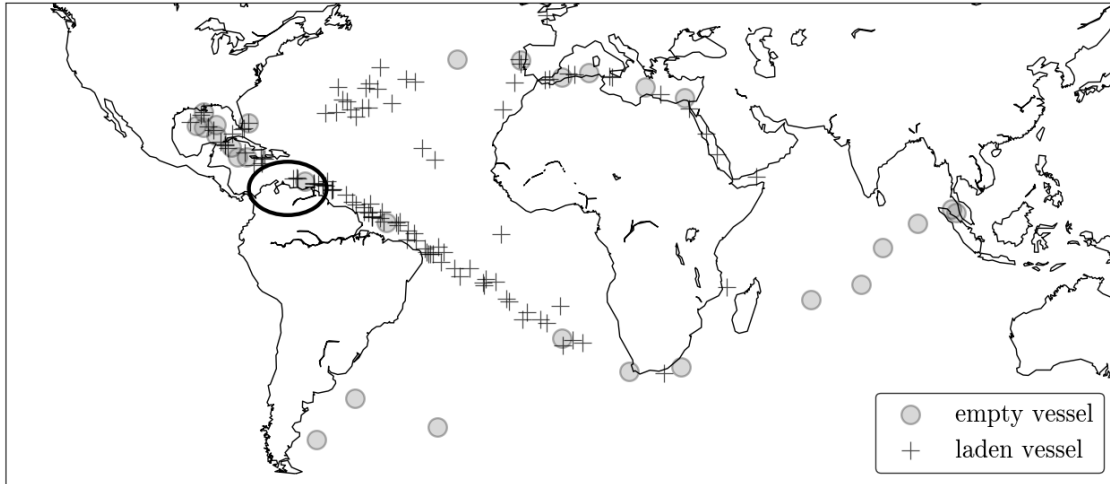


(a) Persian Gulf



(b) West Africa

<sup>13</sup>Such waste is observed in Figure 1.4b according to which some ships decided to ballast without a contract and were later fixed for West Africa.



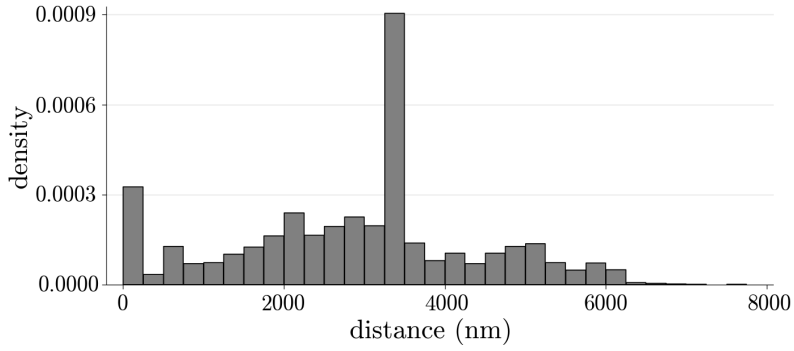
(c) Venezuela

Figure 1.4: Geographic positions of fixtures by export area

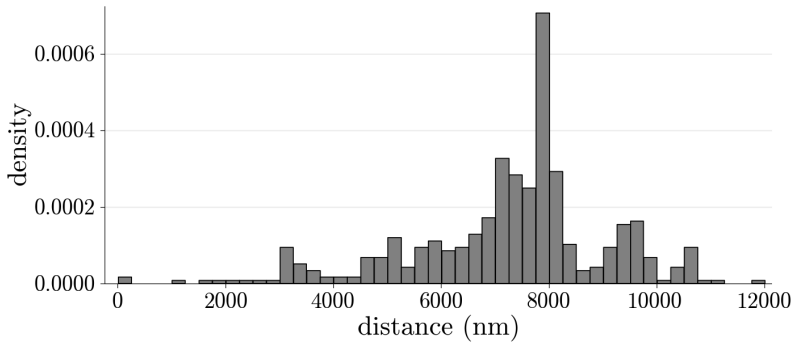
Source: authors' calculations, data from Clarksons matched with AIS 2013 – 2016.

In the case of West Africa (Figure 1.4b), we observe similar phenomena as for the Persian Gulf. Most vessels get fixed on the ballast trip from East Asia. Again, the Singapore area plays an important role, because ships in this area keep the options of going to both main export areas, in particular Persian Gulf and West Africa, without wasting fuel due to deviation. Still, a fair number of vessels are fixed for West Africa while ballasting to the Persian Gulf and therefore incur some deviation costs. However, a profit-maximizing shipowner will only reverse such a routing decision if there are gains in vessel earnings, i.e. because spot rates (or cargo availability) are better out of West Africa. Finally, most fixtures for Venezuela (Figure 1.4c) are recorded during laden trips to the US Gulf or immediately after discharge. Most cargoes head to the US Gulf from Persian Gulf via Cape of Good Hope, but some cargoes go through the Suez Canal<sup>14</sup>.

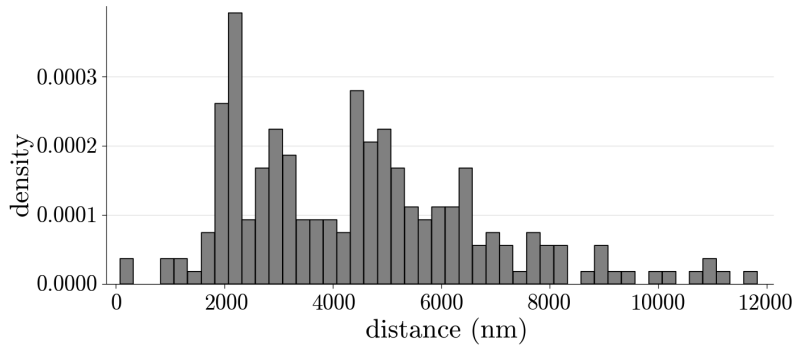
<sup>14</sup>VLCCs can only go part laden through the Suez Canal. Fully laden VLCCs can ship part of their cargo through the Suez-Mediterranean pipeline (Sumed pipeline). This oil pipeline located in Egypt runs from the Ain Sukhna terminal on the Suez Gulf to Sidi Kerir on the Mediterranean Sea.



(a) Persian Gulf



(b) West Africa



(c) Venezuela

Figure 1.5: Distribution of distance from point of fixture to loading by export area

Source: authors' calculations, data from Clarksons matched with AIS 2013 – 2016.

To further highlight the existence of spatial “decision points”, Figure 1.5 shows the distribution of estimated distances from the point of fixture to the loading area. The width of bins in the histograms is set to 250 nautical miles (nm). The distribution for Persian Gulf fixtures is characterized by two modes. The first one is for a category of “prompt” ships which are fixed near the loading area (less than 250 nm distance). The second one corresponds to vessels fixed in the Singapore area (between 3,250 and 3,500 nm distance). The proportion of vessels located more than 4,000 nm away when obtaining

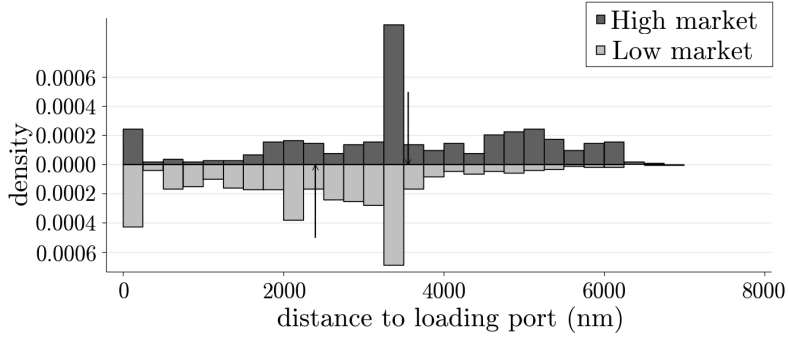
a fixture remains low (around 20%). The distribution for West Africa shows a similar local maximum near Singapore at approximately 8,000 nm, and around one-half of vessels (49.4%) are fixed in a distance ranging between 7,000 and 8500 nm. Conversely, very few ships (1.7%) are fixed within a distance of 3,000 nm near the loading area. Finally, ships fixed for Venezuelan loadings tend to be fixed with an increasing probability nearer the loading area, the mode being at around 2,000 nm.

This pattern is quite revealing in terms of strategic behaviour in the VLCC chartering market. Notably, shipowners generally do not send their vessels speculatively to West Africa, but will fix for this destination latest mid-Indian Ocean. On the other hand, the Persian Gulf acts as the main “sink” for unfixed tonnage, which appears like a natural consequence of the greater cargo volumes being exported here. Venezuela, the only main VLCC loading area in the Americas, also sees its share of fixing of prompt tonnage as owners take the opportunity to wait for a backhaul cargo to India or Asia on the way back from the US Gulf. In all cases, Singapore stands out as a key geographical area from where tonnage is fixed for the reasons already mentioned.

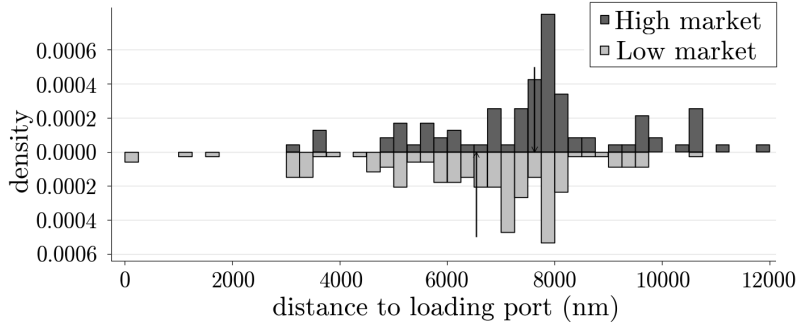
Next, we study the potential effect of market dependency of the fixture location. For that purpose, we show in Figure 5 the distribution of distance to loading area by market conditions. Specifically, we separate high and low freight markets which are defined here as when the BDTI is at least 20% above or below its average, respectively<sup>15</sup>. The arrows indicate the average distance to loading port in each market condition.

---

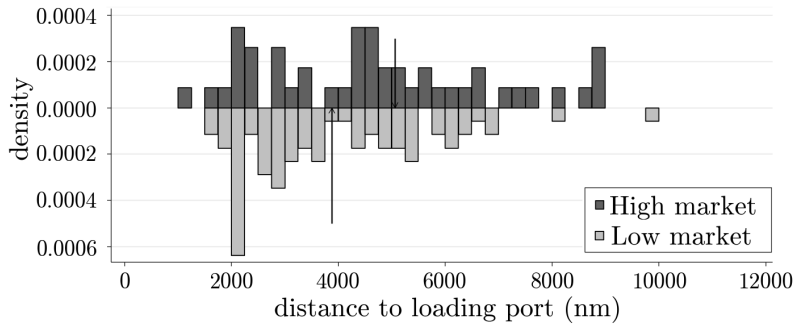
<sup>15</sup>We have checked the robustness of our results to the definition of the high and low freight markets by considering alternative percentages.



(a) Persian Gulf



(b) West Africa



(c) Venezuela

Figure 1.6: Distribution of distance from point of fixture to loading by export area

Source: authors' calculations, data from Clarksons matched with AIS 2013 – 2016.

A first finding is that the average distance to loading area is much higher in high markets than in low market: 3,698 nm against 2,828 nm in the Persian Gulf, 7,624 nm against 7183 nm in West Africa, and 6,061 nm against 4,542 nm in Venezuela. A second finding is that the distribution of distances is not simply shifted. The constraints of geography, i.e. the location of waiting areas and routing decision points, imply that the main modes remain at the same distances. For the Persian Gulf loading area, we still observe a peak at around 3,500 nm corresponding to the Singapore area (and also another one near the loading area). When the freight market is high, relatively more fixtures are done before a ship arrives to the Singapore area from East Asia. As a consequence,

fewer fixtures are done after Singapore since it is more likely that ships are already under contract. This highlights the importance of looking at spatial distributions and not only changes in average lead times as in (Alizadeh and Talley, 2011a,b) where such nuances are lost. The market dependency is less obvious in the case of West Africa, while in Venezuela we observe a larger proportion of vessels fixed when located around 2,000 nm from the loading area in a high market.

In Figure 1.7, we further illustrate the relationship between market conditions and the average distance from the location of fixing and the Persian Gulf loading area by plotting monthly averages. The correlation between distance and the Baltic index is large (0.678) and statistically significant ( $p=0.000$ ). Both trends are highly correlated and the relationship is contemporaneous. This supports Zannetos (1966) intertemporal substitution hypothesis that increasing rates will pull demand forward, with earlier fixing leading to longer average distance.

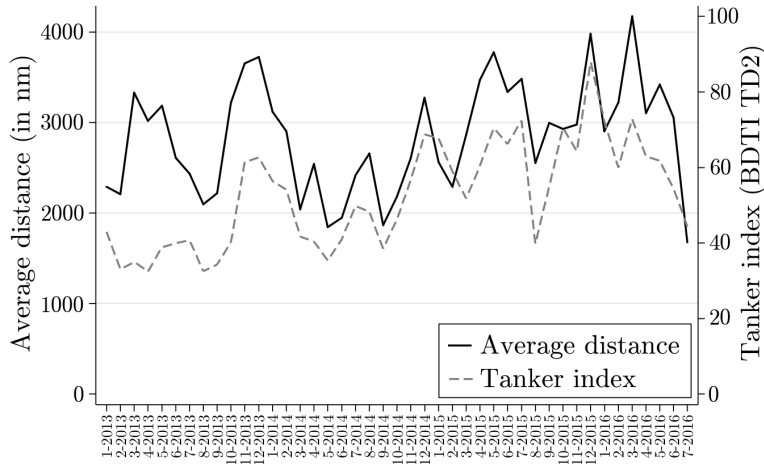


Figure 1.7: Distance to fixtures and Baltic tanker index in Persian Gulf

Source: authors' calculations, 2013 – 2016 data from Clarksons and AIS data.

## 1.5 The determinants of fixture location

Having established that the state of the freight market is a driver of the distribution of fixture locations, we proceed with formal statistical testing of the empirical relationships. A key observation thus far, which has not yet been acknowledged in the related literature (Alizadeh and Talley, 2011a,b), is the fact that the entire distribution of fixture locations depends on the state of the freight market. In other words, the elasticity of distance to the loading port with respect to the spot freight rate may depend on where we are in the market cycle.

To investigate this aspect, we rely on regression models explaining the distance to

loading area for each fixture as a function of market conditions and vessels characteristics. We rely on quantile regressions which provide estimates of the marginal effect of covariates at various locations in the distribution of the dependent variable (Koenker and Bassett, 1978). Quantile regressions will provide robust estimates of the determinants of distance values, particularly with respect to misspecification errors related to heteroskedasticity. Let  $q_\theta(D_{vi})$  be the  $\theta$ th conditional quantile of the distance  $D_{vi}$  for fixture  $i$  concerning vessel  $v$ . The fixture  $i$  occurs at time  $t$  so that  $i$  as index refers in fact to  $i(t)$ . The model that we estimate is:

$$q_\theta(D_{vi}) = \delta(\theta) * FR_t + X_{vi} * \beta(\theta) + \epsilon_{vi} \quad (1.1)$$

where  $\delta(\theta)$  provides the effect of the spot freight rate  $FR_t$  at the  $\theta$ th quantile of the distance distribution,  $X_{vi}$  is a set of control variables dealing with vessel characteristics for fixture  $i$ ,  $\beta(\theta)$  is a vector of coefficients to estimate, and  $\epsilon_{vi}$  is a residual perturbation. In our empirical analysis, we estimate the distance from the fixture location to the load port as a function of spot freight rates at the 10th, 25th, 50th, 75th and 90th percentile. We also estimate the OLS version of (1.1), in which case the estimates relate to the mean value of the dependent variable.

In our dataset, we have repeated observations for many vessels which are observed all over the period<sup>16</sup>. As a consequence, we are able to account for the influence of time-invariant unobserved characteristics of vessels in our estimations by adding a fixed effect. Vessel-specific heterogeneity may arise because it is a dedicated ship on a particular route (so that we pick up a route effect) or because a particular human charterer is charge of the vessel. In a linear model, the residual  $\epsilon_{vi}$  is expressed as  $\epsilon_{vi} = \theta_v + \xi_{vi}$  with  $\theta_v$  the vessel fixed effect and  $\xi_{vi}$  the residual perturbation. The vessel fixed effect is then eliminated by demeaning all observations at the vessel level. Let  $D_{vi} = \delta * FR_t + X_{vi} * \beta + \theta_v + \xi_{vi}$  be the corresponding linear model. The coefficients  $\delta$  and  $\beta$  are obtained by estimating  $(D_{vi} - \bar{D}_v) = \delta * (FR_t - \overline{FR}_t) + (X_{vi} - \bar{X}_v) * \beta + (\xi_{vi} - \bar{\xi}_v)$ . However, the problem is more complex in a non-linear framework since differencing techniques cannot be implemented. Over the last years, a few solutions have been proposed to estimate quantile regression for panel data (Koenker, 2004; Lamarche, 2010).

In this paper, we turn to the two-step estimator proposed in Canay (2011) in a setting where the fixed effects are considered as location shift variables. The fixed effect  $\theta_v$  is eliminated in the first step by estimating the linear model  $D_{vi} = \delta * FR_t + X_{vi} * \beta + \theta_v + \xi_{vi}$  from which we obtain the estimated fixed effects  $\hat{\theta}_v$  such that  $\hat{\theta} = \mathbb{E}[D_{vi} - \hat{\delta} * FR_t - X_{vi} * \hat{\beta}]$ . In the second step, we define the new dependent variable  $\tilde{D}_{vi}$  such that  $\tilde{D}_{vi} = D_{vi} - \hat{\theta}_v$ .

---

<sup>16</sup>The proportion of vessels observed several times is 96.0% in Persian Gulf, 75.6% in West Africa and 56.7% in Venezuela.



Variables	OLS	Quantile regression				
		Q10	Q25	Q50	Q75	Q90
<i>Panel A. Persian Gulf</i>						
A1. Without fixed effect						
Baltic tanker index	28.84*** (13.19)	32.53*** (5.36)	33.40*** (7.99)	23.77*** (11.05)	34.21*** (10.41)	43.64*** (9.74)
Number of observations	2,029	2,029	2,029	2,029	2,029	2,029
R <sup>2</sup>	0.079	0.018	0.033	0.037	0.061	0.073
A2. With fixed effect						
Baltic tanker index	32.03*** (13.43)	28.28*** (6.91)	30.87*** (11.35)	30.58*** (16.12)	34.65*** (13.91)	40.35*** (12.42)
Number of observations	1,947	1,947	1,947	1,947	1,947	1,947
R <sup>2</sup>	0.102	0.046	0.061	0.069	0.086	0.099
<i>Panel B. West Africa</i>						
B1. Without fixed effect						
Baltic tanker index	31.87*** (5.65)	40.50*** (2.73)	33.71*** (3.25)	21.84*** (5.64)	11.89 (1.34)	40.62*** (4.95)
Number of observations	464	464	464	464	464	464
R <sup>2</sup>	0.065	0.039	0.037	0.032	0.012	0.048
B2. With fixed effect						
Baltic tanker index	16.25** (2.23)	33.23*** (3.21)	16.92** (2.50)	7.70 (1.62)	10.19 (1.51)	10.74* (1.72)
Number of observations	351	351	351	351	351	351
R <sup>2</sup>	0.021	0.066	0.024	0.005	0.006	0.012
<i>Panel C. Venezuela</i>						
C1. Without fixed effect						
Baltic tanker index	42.84*** (4.05)	-0.17 (-0.03)	14.44 (1.19)	42.55*** (3.20)	51.88*** (3.17)	59.27* (1.92)
Number of observations	217	217	217	217	217	217
R <sup>2</sup>	0.071	0.002	0.005	0.033	0.057	0.072
C2. With fixed effect						
Baltic tanker index	67.70*** (3.53)	52.33** (2.38)	70.20*** (4.64)	49.32*** (3.41)	66.98*** (4.22)	108.02*** (5.14)
Number of observations	123	123	123	123	123	123
R <sup>2</sup>	0.143	0.111	0.105	0.106	0.102	0.180

Table 1.2: Effect of Baltic Tanker Index on distance to fixtures.  
Source: authors' calculations, 2013 – 2016 data from Clarksons and AIS data.  
Note: Significance levels are equal to 1% (\*\*\*), 5% (\*\*) and 10% (\*).

The fixed effect quantile estimators are then obtained by estimating standard quantile regressions with  $\tilde{D}_{vi}$  as dependent variable instead of  $D_{vi}$ . Canay (2011) shows that this two-step estimator is consistent and asymptotically normal as both the number of vessels and fixtures grow. Both the standard and fixed effect quantile estimates are reported in Table 1.2 for the three geographic area (Persian Gulf, West Africa, Venezuela).

We observe that there is substantial variation in the elasticity of fixture location along the distribution of freight rates. Of particular interest is the tendency towards greater sensitivity at very high freight rates. This pattern makes sense in the context of the discussion of risk aversion and risk attitudes. Compared to the coefficient found in the first quartile of the distance distribution (around 33 without fixed effects and 30 with fixed effects), the coefficient is 30% higher at the 90th percentile. In a strong market, there is a greater risk of transportation shortage, i.e. not finding a vessel for the cargo when it is needed. This implies that risk-averse charterers are forced to fix vessels earlier, resulting in a greater sensitivity to changing market conditions. While the situation in Venezuela seems very similar to that found for the Persian Gulf, the pattern is different in West Africa with some U-shaped profile for the impact of the market index.

The state of the market is not the only factor that influences the distance between fixture and loading area. Some vessels are more desirable than others due to their characteristics. We next investigate the influence of age, deadweight tonnage (DWT) and flag. Additionally, we introduce year and month time fixed effects to account for seasonality. Seasonality in fixture behavior could be influenced by seasonal refinery maintenance periods (common in Asia around April) or seasonal weather patterns such as the monsoon. In Table 1.3, we report the results of both OLS and quantile regressions which are estimated at the 10th, 25th, 50th, 75th and 90th percentile.

Again, we find that the quantile coefficients associated to the market index tend to increase along the distance distribution. There is some evidence of seasonality in fixture behavior, with fixing occurring earlier during March, April and May: the coefficients associated to those months are positive and larger both in the lower and upper parts of the distance distribution. This coincides with the main annual maintenance period for Asian refineries, which ought to reduce shipment volumes temporarily. What we are seeing may therefore simply be the effect of owners seeking earlier employment for their ships during this expected and temporary market lull. We note here that the volume of public fixture data need not be indicative of overall trading volume in the oil market, as coverage is unknown and likely varies over time. Another finding is that younger tonnage (lower age) are fixed earlier (longer distance). This is in line with what we expect from the literature on microeconomic analysis of freight rate premium (Alizadeh and Talley, 2011a,b). Regarding the impact of flag state, we merely note that vessels flagged in Hong

Variables		OLS	Quantile regression				
			Q10	Q25	Q50	Q75	Q90
Baltic tanker index		34.21***	34.49***	37.96***	28.27***	40.97***	42.89***
Month	February	178.69	282.56	-154.96	-57.88	274.84	564.64**
(ref: January)		(1.16)	(0.82)	(-0.55)	(-0.34)	(1.48)	(2.15)
	March	639.24***	774.17**	371.35	524.30***	766.83***	846.83***
		(4.39)	(2.36)	(1.38)	(3.28)	(4.37)	(3.40)
	April	669.81***	948.52***	533.12**	552.52***	826.30***	716.51***
		(4.56)	(2.86)	(1.97)	(3.42)	(4.67)	(2.85)
	May	537.87***	589.85*	398.68	396.29**	763.68***	670.84***
		(3.77)	(1.83)	(1.51)	(2.52)	(4.44)	(2.75)
	June	376.59**	337.24	334.77	312.91*	407.67**	352.35
		(2.54)	(1.01)	(1.22)	(1.92)	(2.28)	(1.39)
	July	146.97	140.12	-169.38	47.50	236.57	307.30
		(0.91)	(0.38)	(-0.57)	(0.27)	(1.21)	(1.11)
	August	431.30**	491.86	41.78	473.44**	648.73***	433.56
		(2.38)	(1.20)	(0.12)	(2.37)	(2.96)	(1.40)
	September	250.27	390.95	-190.34	181.16	490.96**	438.93
		(1.54)	(1.07)	(-0.63)	(1.01)	(2.51)	(1.58)
	October	269.38*	321.19	-3.02	300.15*	406.18**	496.83*
		(1.72)	(0.91)	(-0.01)	(1.75)	(2.16)	(1.86)
	November	235.15	-114.62	-1.68	369.37**	300.58	498.72*
		(1.44)	(-0.31)	(-0.01)	(2.06)	(1.53)	(1.79)
	December	428.01**	444.81	151.61	344.80*	585.12***	637.91**
		(2.42)	(1.11)	(0.46)	(1.77)	(2.74)	(2.11)
Year	2014	-710.14***	-862.27***	-760.07***	-505.65***	-602.39***	-843.91***
(ref: 2013)		(-7.30)	(-3.93)	(-4.23)	(-4.73)	(-5.14)	(-5.08)
	2015	-568.55***	-983.90***	-698.35***	-274.34**	-571.27***	-549.29***
		(-4.67)	(-3.59)	(-3.11)	(-2.05)	(-3.90)	(-2.64)
	2016	-396.03***	-787.10**	-559.41**	-156.83	-245.59	-346.73
		(-2.84)	(-2.50)	(-2.17)	(-1.02)	(-1.46)	(-1.45)
Vessel age		-41.17***	-25.87	-53.58***	-39.38***	-35.60***	-45.95***
		(-5.62)	(-1.57)	(-3.96)	(-4.89)	(-4.03)	(-3.67)
Vessel deadweight		-4,749	-467	-4,932	-4,637	-2,386	-10,754*
		(-1.38)	(-0.06)	(-0.78)	(-1.23)	(-0.58)	(-1.83)
Flag	Hong Kong	305.18***	374.76	468.67**	360.87***	431.08***	-127.73
(ref: Other Flag)		(2.81)	(1.53)	(2.33)	(3.02)	(3.29)	(-0.69)
	Greece	331.61***	179.24	533.02***	366.42***	404.94***	354.57*
		(3.02)	(0.73)	(2.63)	(3.04)	(3.06)	(1.89)
	Marshall Islands	-82.01	8.24	-12.34	-182.77	-11.88	-137.73
		(-0.71)	(0.03)	(-0.06)	(-1.44)	(-0.09)	(-0.70)
	Panama	294.08**	70.09	400.54*	383.37***	337.90**	347.61*
		(2.43)	(0.26)	(1.79)	(2.88)	(2.32)	(1.68)
	Liberia	145.20	421.38	258.39	80.87	231.62	-150.15
		(1.20)	(1.55)	(1.16)	(0.61)	(1.59)	(-0.73)
	Singapore	206.61	393.69	335.89	59.43	268.49*	39.79
		(1.58)	(1.33)	(1.39)	(0.41)	(1.70)	(0.18)
	Malta	-275.91*	-359.11	-150.03	-324.20*	-157.81	-236.34
		(-1.77)	(-1.02)	(-0.52)	(-1.89)	(-0.84)	(-0.89)
Constant		2,790.92**	-529.31	2,100.61	3,055.31**	2,309.39*	5,947.69***
		(2.52)	(-0.21)	(1.03)	(2.51)	(1.73)	(3.15)
Number of observations		2,029	2,029	2,029	2,029	2,029	2,029
R <sup>2</sup>		0.153	0.071	0.080	0.77	0.111	0.127

Table 1.3: Estimates of distance to fixtures in Persian Gulf.

Source: authors' calculations, 2013 – 2016 data from Clarksons and AIS data.

Note: Significance levels are equal to 1% (\*\*\*), 5% (\*\*) and 10% (\*).

Kong, Greece and Panama are fixed earlier than the other flag states<sup>17</sup>.

In a final step, we introduce the identity of the charterer in addition to the state of the market and vessel characteristics. The identity of charterers, estimated here as fixed effects, might influence fixing behavior through the mechanism of risk preferences as discussed earlier, but also whether the charterer is an integrated oil company, refinery or a commodity trading house. Intuitively, integrated oil companies, which are simultaneously oil producers and refinery operators, will treat ocean transportation as an integral part of their supply chain, naturally resulting in longer planning horizons. Conversely, refinery operators and traders may occasionally act opportunistically and fix cargoes on short notice if it makes financial sense. However, distances will naturally also be affected by the location of the company’s assets, and so there will always be a large fixed distance (or time) component in fixture behavior. Figure 1.8 shows the estimated effects of charter identity on the fixture location.

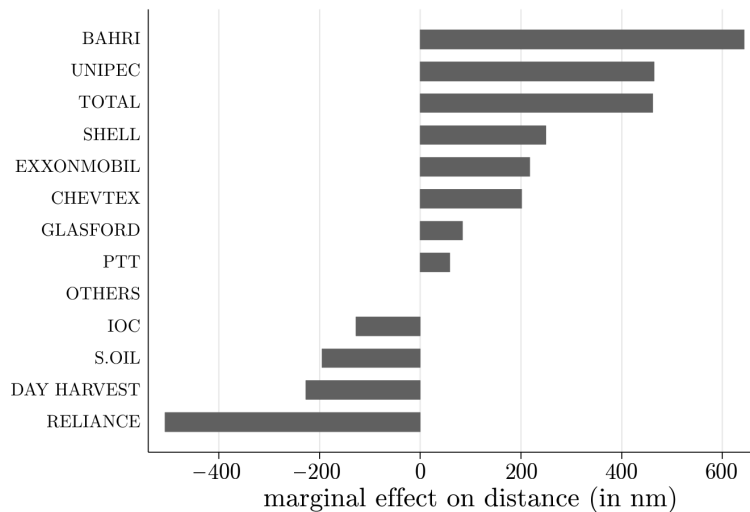


Figure 1.8: Effects of charterers on distance to fixtures in Persian Gulf

Source: authors’ calculations, 2013 – 2016 data from Clarksons and AIS data.

The most obvious point to make here is the relationship between the geographical location of the charterer’s refineries (where applicable) and the distance from the load port at which vessels are fixed. For instance, IOC and Reliance are Indian refiners, which by definition means the voyages for their cargoes will be among one of the shortest in our fixture database and, thus, vessels will never be fixed far away. Similarly, Unipecc is a Chinese refining company, leading to long average voyages from the PG and therefore also higher probability of early (by distance) fixing. This highlights the feedback loop between

<sup>17</sup>As the top flag states in our sample comprise both large national registries and open registries, we are unable to observe an obvious relationship between, for instance, the perceived quality signaling effect of the flag and fixture behavior.

a company’s physical supply chain and its fixing behavior, determined in part by risk attitudes and the risk of stockouts etc. However, we also see that the large international oil majors with global operations (like Total, Shell, Chevtex or ExxonMobil) tend to fix earlier than the reference group. Once again there is reason to believe that this reflects their integrated operations, where oil transportation is less speculation and more a basic cog in their internal supply chain.

## 1.6 Concluding comments

In this paper, we have provided new insight into the behavior of participants in the spot market for tankers as represented by the spatial pattern of contracting (fixtures). We have shown that this is influenced by market conditions, vessel age, vessel flag, seasonality and charterer’s identity. Importantly, we have shown that the presence of key decision points along global trading routes generates a certain “stickiness” in the spatial distribution and that location decisions naturally depend on the location of nodes in the supply chain of the charterer such as refineries. We have, in this regard, contributed to the literature which deals only with contracting lead times. Our work also serves as an illustration of how high-resolution AIS data for ship positions can be used in shipping market analysis. Our research provides the foundation for more advanced modeling of fleet allocation and routing decisions, for instance, where and for how long a vessel should optimally wait for a new contract.

We acknowledge that our empirical work has some limitations. Firstly, only a fraction of all spot fixtures are reported publicly, and we cannot be assured that those reported represents an unbiased sample of the entire population. Consequently, our estimates may be biased and affected by changing data coverage over time. However, this critique affects all related empirical research using micro-level fixture data and, as the remaining data are private and/or unobservable, it is hard if not impossible to address it. Secondly, we do not assume that the distance from the load port at which a vessel is fixed and the rate at which it is fixed are set jointly. This is a point of contention, and our approach differs from the joint modeling of rates and lead times in Alizadeh and Talley (2011a,b). Conceptually, one could envisage the existence of a “term structure” of spot freight rates where, at any given point in time, the going market rate for a vessel depends on where it is located (in time and space) relative to the loading area. However, its shape would likely be non-linear and market dependent, thus making it less suitable for the linear framework currently proposed in the literature. Given that we have shown that a large part of the chartering decision (in terms of location) is governed simply by the geography of trade and refinery locations, we do not believe that a joint modeling of rates and location or lead times is required.

There are several interesting areas for future research within the topic of spatial freight market modeling. Importantly, chartering decisions are influenced by external market conditions, but reflexively, these decisions also influence the future market since the available ship capacity that is dynamically reallocated represents the supply side of the market. Therefore, better understanding of this process can improve regional freight rates forecasts. Further work is also required to understand optimal waiting strategies, and global vessel and fleet positioning.

## References

- Adland, R., Cariou, P., and Wolff, F.-C. (2016). The influence of charterers and owners on bulk shipping freight rates. *Transportation Research Part E: Logistics and Transportation Review*, 86(Supplement C):69 – 82.
- Adland, R. and Cullinane, K. (2005). A time-varying risk premium in the term structure of bulk shipping freight rates. *Journal of Transport Economics and Policy*, 39(2):191–208.
- Adland, R. and Cullinane, K. (2006). The non-linear dynamics of spot freight rates in tanker markets. *Transportation Research Part E: Logistics and Transportation Review*, 42(3):211 – 224.
- Adland, R. and Jia, H. (2016). Vessel speed analytics using satellite-based ship position data. In *2016 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)*, pages 1299–1303.
- Adland, R. and Jia, H. (2018). Dynamic speed choice in bulk shipping. *Maritime Economics & Logistics*, 20(2):253–266.
- Adland, R., Jia, H., and Strandenes, S. P. (2017). Are AIS-based trade volume estimates reliable? The case of crude oil exports. *Maritime Policy & Management*, 44(5):657–665.
- Adland, R. and Strandenes, S. P. (2007). A discrete-time stochastic partial equilibrium model of the spot freight market. *Journal of Transport Economics and Policy*, 41(2):189–218.
- Agnolucci, P., Smith, T., and Rehmatulla, N. (2014). Energy efficiency and time charter rates: Energy efficiency savings recovered by ship owners in the panamax market. *Transportation Research Part A: Policy and Practice*, 66:173 – 184.
- Alizadeh, A. H. and Talley, W. K. (2011a). Microeconomic determinants of dry bulk shipping freight rates and contract times. *Transportation*, 38(3):561–579.

- Alizadeh, A. H. and Talley, W. K. (2011b). Vessel and voyage determinants of tanker freight rates and contract times. *Transport Policy*, 18(5):665 – 675.
- Aßmann, L. M., Andersson, J., and Eskeland, G. (2015). Vessel speed analytics using satellite-based ship position data. In *Discussion Paper 2015/13. Department of Business and Management Science, Norwegian School of Economics*.
- Beenstock, M. (1985). A theory of ship prices. *Maritime Policy & Management*, 12(3):215–225.
- Beenstock, M. and Vergottis, A. (1989). An econometric model of the world tanker market. *Journal of Transport Economics and Policy*, 23(3):263–280.
- Benth, F. E. and Koekebakker, S. (2016). Stochastic modeling of supramax spot and forward freight rates. *Maritime Economics & Logistics*, 18(4):391–413.
- Bjerksund, P. and Ekern, S. (1995). Contingent claims evaluation for mean-reverting cash flows in shipping. In *Real Options in Capital Investment: Models, Strategies, and Applications*, pages 207–220. Praeger, Westport, CT.
- Canay, I. A. (2011). A simple approach to quantile regression for panel data. *The Econometrics Journal*, 14(3):368–386.
- Charemza, W. and Gronicki, M. (1981). An econometric model of world shipping and shipbuilding. *Maritime Policy & Management*, 8(1):21–30.
- Clarkson Research (2018). Shipping Intelligence Network, [www.clarksons.net](http://www.clarksons.net).
- Clipper Data (2016). [www.clipperdata.com](http://www.clipperdata.com).
- Cullinane, K. (1991). The utility analysis of risk attitudes in shipping. *Maritime Policy & Management*, 18(3):157–169.
- Eckbo, B. (1977). Risikopreferanser blant noen Skandinaviske tankrederier før og under krisen i tankmarkedet, Institute for Shipping Research, Norwegian School of Economics and Business Administration, Bergen, Norway.
- Eriksen, I. and Norman, V. (1976). Ecotank, Modell for analyse av tankmarkedenes virkemåte, Institute for Shipping Research, Norwegian School of Economics and Business Administration, Bergen, Norway.
- Evans, J. J. (1994). An analysis of efficiency of the bulk shipping markets. *Maritime Policy & Management*, 21(4):311–329.
- Friedman, M. and Savage, L. J. (1948). The utility analysis of choices involving risk. *Journal of Political Economy*, 56(4):279–304.

- Hawdon, D. (1978). Tanker freight rates in the short and long run. *Applied Economics*, 10(3):203–218.
- IEA (2018). IEA Oil Market Report, [www.iea.org/oilmarketreport/omrpublic/](http://www.iea.org/oilmarketreport/omrpublic/).
- Jia, H., Lampe, O. D., Solteszova, V., and Strandenes, S. P. (2017). An automatic algorithm for generating seaborne transport pattern maps based on AIS. *Maritime Economics & Logistics*, 19(4):619–630.
- Kavussanos, M. G. and Alizadeh, A. H. (2002). The expectations hypothesis of the term structure and risk premiums in dry bulk shipping freight markets. *Journal of Transport Economics and Policy*, 36(2):267–304.
- Koenker, R. (2004). Quantile regression for longitudinal data. *Journal of Multivariate Analysis*, 91(1):74 – 89. Special Issue on Semiparametric and Nonparametric Mixed Models.
- Koenker, R. W. and Bassett, G. (1978). Regression quantiles. *Econometrica*, 46(1):33–50.
- Koopmans, T. (1939). Tanker Freight Rates and Tankship Building. Haarlem, Netherlands.
- Köhn, S. and Thanopoulou, H. (2011). A gam assessment of quality premia in the dry bulk time-charter market. *Transportation Research Part E: Logistics and Transportation Review*, 47(5):709 – 721.
- Lamarche, C. (2010). Robust penalized quantile regression estimation for panel data. *Journal of Econometrics*, 157(2):396 – 408.
- Laulajainen, R. (2007). The geographical foundations of tanker (dirty) shipping. *Maritime Policy & Management*, 34(6):553–576.
- Laulajainen, R. (2008). Operative strategy in tanker (dirty) shipping. *Maritime Policy & Management*, 35(3):313–339.
- Laulajainen, R. (2011). Oil product tanker geography with emphasis on the handysize segment. *Fennia - International Journal of Geography*, 189(1):1–19.
- Lorange, P. and Norman, V. D. (1973). Risk preference in scandinavian shipping. *Applied Economics*, 5(1):49–59.
- Norman, V. and Wergeland, T. (1939). Nortank – A Simulation Model of the Freight Market for Large Tankers. Centre for Applied Research, Norwegian School of Economics and Business Administration, Report 4/81, Bergen, Norway.



- Parker, S. (2014). Matching in the oil tanker industry: implications for energy efficiency, Doctoral dissertation, UCL (University College London).
- Schofield, N. (2007). *Commodity derivatives: Markets and applications*. John Wiley and Sons Ltd: Chichester, England.
- Smith, T. W. P., Jalkanen, J. P., Anderson, B. A., Corbett, J. J., and Faber, J. (2014). Third IMO GHG Study, IMO: London, UK.
- Stopford, M. (2009). *Maritime Economics, third ed.* Routledge, London.
- Strandenes, S. (1986). NORSHIP – A Simulation Model for Bulk Shipping Markets. Center for Applied Research, Norwegian School of Economics and Business Administration, Bergen, Norway.
- Tamvakis, M. N. and Thanopoulou, H. A. (2000). Does quality pay? the case of the dry bulk market. *Transportation Research Part E: Logistics and Transportation Review*, 36(4):297 – 307.
- Tinbergen, J. (1934). In: Klaasen, L.H., Koyck, L.M., Witteween, H.J. (Eds.), Jan Tinbergen – Selected Papers. North-Holland Publishing Company, 1959, Amsterdam.
- Tvedt, J. (1997). Valuation of VLCCs under income uncertainty. *Maritime Policy & Management*, 24(2):159–174.
- Tvedt, J. (2011). Short-run freight rate formation in the vlcc market: A theoretical framework. *Maritime Economics & Logistics*, 13(4):442–455.
- Wergeland, T. (1981). Norbulk: A Simulation Model of Bulk Freight Rates. Working Paper 12, Norwegian School of Economics and Business Administration, Bergen, Norway.
- Zannetos, Z. S. (1966). *The Theory of Oil Tankship Rates*. MIT Press, Cambridge, MA.

## Appendix

In the case where there is a large gap between two adjacent positions in the AIS data and a fixture appears between them, we approximate the movement of a vessel between the two positions and estimate a probable location of the fixture. For the approximation of the movement, we have created a network over the ocean (a part of it is shown in Figure 1.9). We assume that the vessel is moving along the shortest path on the network. This assumption is problematic in the case of too large gaps - a vessel might choose a different path or needs to finish a different voyage first. Thus, we filter out observations if the gap is above a certain threshold.

If the gap is large enough so the approximation of the movement along a straight line might cross land, but not so large that we filter out the data, we proceed in the following way. For any two positions in the ocean, we find the nearest nodes in the network (mesh) of the ocean and assume that the vessel is moving along the shortest path between them. Then we estimate the position of the fixture by assuming a constant speed along the path. That is, the proportional distance between the estimated fixture point and each of the AIS positions is the same as the time differences between the timestamp of the fixture and timestamps of the neighboring AIS positions. Then, when measuring the distance to the loading port, we add the distance from the “future” (i.e. after the estimated fixture point) part of the shortest path.

The whole procedure is illustrated in Figure 1.9. We can see that the interpolation of the vessel positions by a straight line would cross the land and the fixture position would be estimated there. Our procedure overcomes this issue, even though there are still limitations in accuracy and the necessary assumptions - movement along the shortest path and the constant speed (especially if there is an important port along the way, where the vessel might stop).

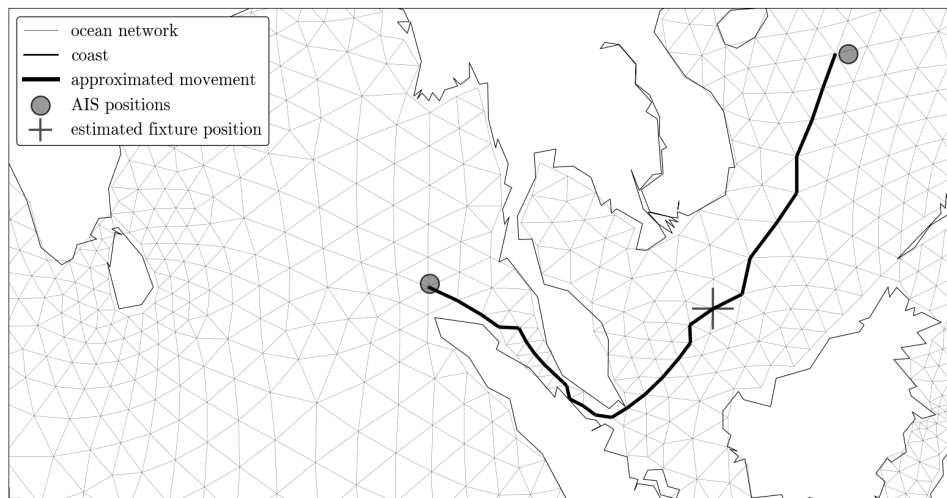


Figure 1.9: Demonstration of movement approximation

Source: Authors' calculations, AIS data.

# The value of foresight in the drybulk freight market

Vit Prochazka<sup>□</sup>    Roar Adland<sup>□</sup>    Stein W. Wallace<sup>□</sup>

<sup>□</sup>Department of Business and Management Science  
Norwegian School of Economics, Bergen, Norway

## Abstract

We analyze the value of foresight in the drybulk freight market when repositioning a vessel through space and time. In order to do that, we apply an optimization model on a network with dynamic regional freight rate differences and stochastic travel times. We evaluate the value of the geographical switching option for three cases: the upper bound based on having perfect foresight, the baseline set by a random strategy, and the case of perfect foresight but only for a limited time horizon. By combining a neural network with optimization, we can assess the impact of varying foresight horizons on economic performance. In a simple but realistic two-region case, we evaluate empirically that the upper bound for large vessels can be as high as 25% cumulative outperformance, and that a significant portion of this theoretical value can be captured with limited foresight of a few weeks. Our research sheds light on the important issue of spatial efficiency in global ocean freight markets and provides a useful upper bound for the value of investing in predictive analysis.

**Keywords:** dry bulk market, dynamic programming, neural network, foresight

## 2.1 Introduction

As in most business sectors, the ocean freight industry has recently seen a surge of interest in the digitalization of business processes and the adoption of machine learning and “big data” applications to improve economic performance. Such investments often take place in the belief that there are market inefficiencies that can be exploited, without any verification of the actual economic potential.

In this paper we propose a method to estimate potential financial benefits of having a perfect forecast of future regional freight rates. Since any realistic forecast is not 100% accurate, the real achievable economic benefits will be lower. In other words, we establish an upper bound for the increase in earnings based on freight rate forecasts and optimal decision making in the spot chartering market.

Naturally, if we were able to predict the market significantly better than other participants, we could extract economic gains in many ways. For example, we could charter vessels in before a predicted increase in rates and then charter them out once the rates are higher. However, in this paper we limit our analysis to spatial inefficiencies in rates across different regions. That is, we focus only on decisions that all operators of vessels in the bulk shipping freight markets face: How to optimally reallocate a ship through space and time by sequentially accepting freight contracts (charters) for spot market cargoes between port pairs in a transport network, often called tramp shipping. We note here that tramp shipping differs from the liner shipping networks usually considered in the optimization literature (see Christiansen et al. (2013) for a comprehensive literature review) in that there are no fixed routes, schedules or contract cargoes. Instead, a large number of ships and full loads (cargoes) are continuously matched by shipbrokers in a perfectly competitive market where the future trading pattern and employment of the vessel is largely unknown.

We analyze two cases of this problem that differ in the type of perfect forecast we use. In the first case, future freight rates are completely known for the entire future<sup>1</sup> (perfect foresight). In the second case we also assume perfect knowledge of future rates, but the period of this foresight is limited (we compare different horizons ranging from 20 to 80 days).

The optimization task to be solved in the first case can be classified as *dynamic assignment problem with stochastic travel times*. Dynamic assignment problems on networks often arise in logistics applications, for instance Powell (1996) in the context of truck logistics or Topaloglu and Powell (2006) in business jet repositioning. In both these cases, the authors use a dynamic programming framework (see Bellman (1957)).

Uncertainty comes naturally in dynamic environments, where new information arrives

---

<sup>1</sup>The entire future that is considered. It is still represented by a finite horizon, but the horizon is sufficiently large.

over time, and usually demand is the main source of uncertainty in logistics applications. Handling stochasticity in the optimization problems is a big topic itself, see for instance Wallace (1986) for a pioneering work on stochasticity in network problems, or Lium et al. (2009) for more recent work. In both cases of our analysis, we work with stochastic travel times. However, we use the uncertainty as a desirable feature of the model that makes exact planning of the distant future impossible, rather than aiming for the most precise description of the underlying probability distribution.

In the case of limited foresight, the main challenge is to handle the “end of the planning horizon” issue. That is, since we optimize over a finite and short horizon, we must make sure the model “knows” that the end of the planning horizon does not mean the “end of the world” and does not simply maximize earnings over this period. In the spot freight market, there is almost always a trade-off between capturing immediate profit and getting a ship into a better location (a region with higher rates). The step that prioritizes location over profit would never be chosen as the last step (which further influences the choice of the previous step, etc.) by simple maximization of the earnings.

Such an issue arises in other classes of problems as well. It naturally appears in reinforcement learning tasks, see Sutton and Barto (1998); Wiering and van Otterlo (2012) for an introduction into the topic and algorithms. In classical optimization tasks, we can find the issue in inventory models (van der Laan et al., 2004), where we need to treat carefully the conditions at the end of the planning horizon (no condition on the inventory level at the end of the planning period leads to an empty inventory at the end).

Less often, we see the issue in deterministic optimization. If a problem is formulated as a standard linear program, it is usually solvable for large instances, therefore it is possible to use a long enough horizon (with discounting) without a special treatment of the end of the planning horizon (as in our perfect foresight setting). It is a different story when facing a difficult complex problem and/or a stochastic environment. Powell (2014) discusses different conceptual approaches that can be applied in stochastic programming. One of the commonly used approaches is an approximation of the value function (Simão et al., 2009; Bertsekas and Tsitsiklis, 1996). That is, the earnings coming from the “after the end of the planning horizon” periods are approximated by features describing the state of the system at the end of the planning horizon. Another approach is to approximate the policy by some known features of the system directly (Buşoniu et al., 2012).

We introduce an algorithm that conceptually belongs to the latter class, but it differs from approaches described in the literature in the way the policy is constructed. We use the knowledge from the perfect foresight case to assess the quality of the policy instead of the direct evaluation of the policy (that is, simulating its performance) during the policy search. The algorithm represents the policy by a neural network that is trained to predict correct decisions by using only information from the limited foresight horizon.

As a learning objective for the neural network, the optimal decisions from the perfect foresight model are used. Such an approach belongs to the *policy approximation* methods.

In practice, the outcome of an operators' decisions is a market where the world fleet of vessels is slowly reallocated around the globe in an attempt to match overall supply and demand. However, in the short run, the bulk shipping freight market can be thought of as several regional markets with their own supply and demand dynamics. As described in Adland et al. (2016), each matching of a cargo and a ship can be thought of as a micro-market, where only those ships commercially available and able to physically meet the loading window can offer transportation. Accordingly, the spot freight rate in one loading region is set by the immediate equilibrium of cargoes offered and vessels available to load within a particular future time window. As vessels can only move slowly between regions, shortages or oversupply of tonnage – and the corresponding high or low regional freight rates – can persist for weeks. This inertia in regional supply – which implies a slow move towards equilibrium - is present even though the market is considered transparent and perfectly competitive, purely because there is as a cost of relocating tonnage. This cost can be implicit in terms of time (which has an alternative value) or explicit (fuel and other voyage costs) if an owner is moving the ship speculatively at his own expense. The solution to our optimization problem is therefore closely related to the macroeconomic question of whether the ocean freight market is spatially efficient: If there are no economic gains to be had from spatial optimization – even with the benefit of limited foresight – then the market is efficient<sup>2</sup>.

The issue of spatial market efficiency has been approached in several parts of the literature. Berg-Andreassen (1996, 1997); Glen and Rogers (1997); Veenstra and Franses (1997) investigate the statistical properties of regional freight rates. The general finding is that regional rates are non-stationary and co-integrated, that is, there exists a stable long-run relationship between them. This is consistent with the observation that regional spot freight rates must revert towards some common (global) stochastic trend because of the ability of ships to continuously move from regions of oversupply towards regions of undersupply. Koekebakker et al. (2006) question the findings of non-stationarity for spot freight rates and argue that non-linear dynamics and weak power of the linear unit root tests used results in a failure to reject non-stationarity. Adland et al. (2017a) emphasize that statistical co-integration of regional spot freight rates is a necessary but not a sufficient condition for an efficient freight market, as it does not preclude the possibility that short-term regional differentials are large and persistent enough to enable operators to take advantage through well-informed chartering decisions.

---

<sup>2</sup>We note here that short-term predictability of the regional spot freight itself does not preclude market efficiency, as spot rates cannot be traded or stored (Adland and Cullinane (2006); Benth and Koekebakker (2016)).

Other studies assess the impact of spatial differences in freight rates directly. Adland et al. (2017b) employ a real option framework to calculate the value of the geographical switching option in the drybulk market. They find that the main source of option value is the observed persistent premium in Atlantic freight rates over Pacific freight rates and that proactive switching does not add further value. However, they acknowledge that certain assumptions imposed by their analytical model (such as immediate switching and constraints on switching costs) are not realistic. The issue of spatial freight market efficiency has also been investigated in the economic geography literature. For instance, Laulajainen (2007) studies systematic geographical differences in drybulk freight rates and proposes that the difference between regions can be explained through a Revenue Gradient, equivalent to the ratio of demanded and available tonnage weighted by sailing distance to a discharge or loading region. Laulajainen (2006) fits route-specific freight rates in a static gravity-type model and finds the sailing distance to be the most important explanatory variable. Laulajainen (2010) argues that operational and tactical decisions are governing at a regional level, while dynamic inter-regional allocation of the fleet represents a strategic decision with an inherent risk.

The contributions of this paper are twofold. Firstly, we assess empirically the upper bound for a vessel’s earnings based on freight rate forecasts and spatial optimization in the drybulk shipping freight markets. This analysis also provides insight into the dynamic of spatial market (in)efficiency by showing the evolution of excess earnings (above the market average) over time. In other words, we answer the following question: Is it possible to persistently outperform the market by repositioning of the vessel in a better way than competitors or are there only occasional opportunities that can be exploited while most of the time it does not matter which decision is made due to a high degree of market efficiency?

Secondly, we propose an algorithm to handle the end of planning horizons problem. Contrary to the related operations research literature, our model does not use constant valuation (imposed by constraints, penalization of different positions, etc.) of a state of the system at the end of the planning horizon that is set before the optimization itself. Instead, we let the algorithm learn to make better decisions from observing the past. This is done in a policy approximation fashion, where a simple neural network is used to predict optimal decisions after the learning phase.

Our findings are important for vessel operators and academic researchers alike. We show that the value of optimization conditional on our foresight assumptions is dependent on vessel size, with large vessels showing greater potential than small- and mid-size vessels. We also show that most of the economic benefit depends on correctly predicting short-term trends in regional freight rates, offering some hope to current investors in predictive analysis.

The remainder of this paper is structured as follows: Section 2.2 outlines our methodology, Section 2.3 presents our numerical results and Section 2.4 concludes.

## 2.2 Methodology

In this section, we formulate a general model of freight trading that we use for our analysis. We should emphasize in the beginning that our model is not meant to capture actual operation of a vessel as we do not include several subprocesses that need to be taken into consideration. For example, we ignore attributes related to specific cargoes (loading requirements, laycan period, time lag between entering the spot market and the first day of the loading window, etc.). In addition, our optimization model assumes perfect knowledge of freight rates, either for the entire future or a limited period. In the real world, future freight rates, as well as other attributes of cargoes, are naturally uncertain. Our aim is to establish an upper bound for any realistic strategy by assuming the perfect knowledge of rates and exploiting inefficiencies in regional freight rates.

The motivation for formulating such an optimization model for vessel allocation is to take into account the time lag needed to switch between regions when analyzing spatial efficiency. This significant time lag is not a regular property in other markets, where changes can be implemented almost instantly, for example, the allocation of money to different assets in the stock market. In shipping, it takes approximately two months to relocate a vessel from the Pacific to the Atlantic region. So even if an opportunity arises in a different region, the vessel may not be able to reposition in time to take advantage of it. Moreover, if more participants spot the opportunity at the same time and independently decide to go for it, they change the balance between supply and demand in that region, thus, not only making the opportunity disappear, but potentially further reducing its value to a costly mistake.

As some studies (for instance Adland et al. (2017b)) on spatial efficiency in the drybulk market use the methodology from financial markets enabling immediate switching between regions, our approach therefore adds an important dimension to the limited literature on spatial efficiency.

### 2.2.1 Model

Let us assume  $\mathcal{T} = \{1, 2, \dots, T\}$  is a finite time (planning) horizon (days) and  $\mathcal{I}$  is a set of ports/regions. If not stated otherwise we use indices  $i, j \in \mathcal{I}$  and  $t \in \mathcal{T}$ . The task is to reallocate a single ship through this region-time space. The tripcharter rate (price) for a single trip<sup>3</sup> on a route from origin  $i$  to destination  $j$  at time  $t$  is  $r_{ijt}$  for every day of

---

<sup>3</sup>A single trip usually comprises one ballast leg to the origin of a cargo, loading the cargo and then a laden leg to a destination port, where it is discharged.



the duration of the trip, which takes from  $\tau_{ij}^{\min}$  to  $\tau_{ij}^{\max}$  days. We assume that the trade duration is unknown before the commencement of the trip (due to weather, port queues, etc.), but is observed at the moment of arrival, and, thus, the consecutive decision (i.e. choosing a destination for the next trip) can differ for different arrival days. We assume that the fixture for the next trip takes place only upon discharging and completion of the previous trip. We assume a discrete uniform distribution  $\mathcal{U}\{\tau_{ij}^{\min}, \tau_{ij}^{\max}\}$  of trip duration as in Adland and Jia (2017). In the real world, the uncertain sailing duration is obviously not uniformly distributed. However we do not try to make a perfect model for a real decision-making process, so rather than capturing the real distribution of trip durations, the cumulative impact of the stochasticity of travel times is used to decrease the capability of exact planning the further we go into the future. This is a required feature of the model. If we did not include it and assumed deterministic trip duration, the decisions would be basically driven by a random few-days spike in rates that happens in a distant future. Thus, it serves a similar purpose as a discount rate in other economic models. However, the uncertain trip duration also reflects the real situation (weather, port queues, etc. are not artificial constructs) and the boundaries  $\tau_{ij}^{\min}$  and  $\tau_{ij}^{\max}$  can be estimated quite realistically by discussing with industry experts. The choice of a proper discount rate would be a more difficult task.

We simulate and compare two types of fleet allocation strategies, the *oracle* and *coin* trading strategy, respectively. Since we will use most of the variables and parameters in both strategies, we distinguish them by superscript  $X \in \{o, c\}$  denoting oracle or coin strategy, respectively.

For the *oracle* one, we assume that all future rates are known. We analyze two cases; first with known rates for the whole horizon  $\mathcal{T}$ , and second the case of limited foresight for which rates are known only for a given number of future days  $\Delta$  ( $\Delta \ll \mathcal{T}$ ). The task in both cases is to reposition a vessel to maximize the overall expected earnings.

The *coin* based strategy does not exploit any knowledge of future rates. Instead we define quantities  $x_{ijt}^c$ , which can be interpreted as probabilities<sup>4</sup> of making a decision “go to  $j$ ” for a vessel positioned in region  $i$  at time  $t$ , where superscript  $c$  refers to the coin strategy and  $\sum_j x_{ijt}^c = 1 \quad \forall i, t$ . The probabilities  $x_{ijt}^c$  are set upfront and disregard the future rates. If only two options for the next destination were available and probabilities were 50-50, the operator could flip a coin to determine where to go next, hence the name. Another possible interpretation, with respect to further modelling, is that vessels are divisible assets and  $x_{ijt}^c$  represents a fraction of a vessel capacity located in  $i$  at time  $t$  that is allocated to  $j$ . A similar thought process can be applied to stochastic trip durations (different fractions sail different amount of days).

---

<sup>4</sup>For simpler presentation,  $x$  represents both probabilities in the coin flip case and decisions in the oracle case, even though most readers may be accustomed to a probability being denoted by  $p$ .

In Adland and Jia (2017), a similar approach is used for trade simulation, and the values  $x_{ijt}^c$  are set proportionally to real tradeflows. In this special case, the expected profit of such a strategy reflects the average earnings of an agent in the market. As we do not have data of tradeflows for all sectors, we set the probability such that the expected utilization of all routes is equal. Based on our numerical test, there is not a significant difference in the obtained results (earnings) between this setting and probabilities proportional to tradeflows. Therefore, the results obtained by the coin strategy approximate reasonably well the average earnings of market participants.

### 2.2.2 Perfect foresight

In the perfect foresight case, we assume that the rates  $r_{ijt}$  for the *oracle* strategy are known for the whole time horizon and ship capacity is positioned optimally to maximize the expected profit at time  $T$  (the end of the planning horizon).

The assumption of known rates for the oracle strategy is not realistic and results obtained by this strategy are therefore an optimistic estimate (upper bound) to any realistic achievable earnings. It is worth noting that some additional features that could increase the profit are not assumed here, for instance speed optimization (with respect to freight market condition, weather etc.) or exercising the option to wait between contracts. Here it is assumed that the vessel must enter into a new contract immediately after transporting the previous cargo. However, we argue that these additional operating options have a smaller impact than the knowledge of future rates with corresponding optimization. Waiting between contracts in a region with the hope of getting a higher freight rate will be profitable just in very rare situations as we lose utilization of the ship every day of waiting. More likely, operators will face the opposite problem: there is no available cargo to carry on the spot market in the immediate vicinity of the empty vessel. Naturally, cargo availability differ across regions (and time) and should be taken into account when operating a real vessel. The assumption of instant cargo availability can be justified by assuming that the problem is being solved for a relatively modern ship. New ships are usually more energy efficient and their lower marginal cost therefore means they will remain employed in low freight markets (Stopford, 2009). Cargo availability is also influenced by the quality of the network of shipbrokers, which would be difficult to model as the information is private and differ for individual operators. Hence, we assume the network is of sufficiently high quality to secure a cargo to carry at any point in time.

Speed choice does not play such a significant role, since the regional freight markets are co-integrated (Adland et al., 2017a) and we focus on regional differences rather than on overall results. Specifically, it might be profitable to speed up in the case of a high market, but then it is very likely that the optimal strategy is to speed up in all regions. Thus, enabling dynamic but correlated speeds would not change the results qualitatively.

It is easy to apply the *coin* strategy in real life: all the operator needs to do is to flip a coin (generate a random number) and accept the decision. Therefore, results achieved by this strategy can be considered as our baseline for comparison with the oracle strategy.

In real life, we can work only with limited predictions, where we have some partial, imperfect, knowledge. For instance, we may have a probability distribution constructed for future rates, and such a distribution will have higher variance (will be less accurate) the further we look into the future. Accordingly, any smart strategy based on a realistic prediction should achieve results above the baseline, but below the upper bound, established by the coin and oracle strategy, respectively.

Generally with this type of model, we need to be aware of the modeling issues with regards to the end of the planning horizon - naturally we do not expect that the real trade ends that day. This repositioning game often has a fronthaul-backhaul dynamic, which means that we often accept a cargo with lower immediate profit (i.e. a backhaul trip) in exchange for having our vessel (resource) positioned in a better region to increase future profits. With a finite horizon, a backhaul trip should never be chosen at the end, and any such (non-optimal) decision may propagate back through time and affect all the previous decisions in a bad (non-optimal) way. In our case, decisions do not have irreversible consequences and the stochasticity of trip duration makes exact planning in the future impossible. Thus, with a sufficiently large time horizon, the impact of non-optimal decisions that are made towards the end of the planning horizon vanishes.

## Solution

We first introduce the variables used in our procedures for obtaining and evaluating solutions. We remind the reader that the superscript  $X \in \{o, c\}$  distinguishes between the oracle and coin strategy in the notation.

To store the optimal decisions for the oracle strategy, we introduce variables  $x_{ijt}^o$ , which have the same interpretation as  $x_{ijt}^c$  in the coin setting, that is, the probability of repositioning the ship from  $i$  to  $j$ , given that she is located in region  $i$  at time  $t$ . However, in the oracle case  $x_{ijt}^o$  are the decisions obtained as a solution to the maximization of expected earnings problem, not probabilities set upfront as in the coin strategy.

To track the expected allocation of the vessel, we define a parameter  $R_{it}^X$  denoting the expected ship capacity allocated in region  $i$  at time  $t$ . Further, we define the expected ship capacity  $f_{ijt}^X$  going from  $i$  to  $j$  and starting at time  $t$ . It can be interpreted as the expected flow on the  $ij$  arc. The relation between decisions and the flow is then:

$$f_{ijt}^X = x_{ijt}^X R_{it}^X \quad (2.1)$$

We assume that only trips that are finished within the planning horizon are considered. To simplify the notation in the algorithms, we use  $\pi_{ij}$  for the probability of a particular trip duration. Since we assume the discrete uniform distribution of trip duration with boundaries  $\tau_{ij}^{\min}$  and  $\tau_{ij}^{\max}$ , that is, we have  $\tau_{ij}^{\max} - \tau_{ij}^{\min} + 1$  particular realizations, probability  $\pi_{ij}$  of each realization is

$$\pi_{ij} = \frac{1}{\tau_{ij}^{\max} - \tau_{ij}^{\min} + 1} \quad (2.2)$$

If the operator decides to take a cargo (perform a trip) from  $i$  to  $j$  at time  $t$ , by this move he earns in expectation (over all realizations of possible trip durations  $s$ )  $e_{ijt}$  given by

$$e_{ijt} = \sum_{s=\tau_{ij}^{\min}}^{\min\{\tau_{ij}^{\max}, T-t\}} \pi_{ij} r_{ijts} \quad (2.3)$$

The solution for the oracle strategy can be obtained by solving the following optimization program determining the optimal flows  $f$  (for better readability the superscript  $X = o$  is omitted):

$$\max_f \sum_{ijt} e_{ijt} f_{ijt} \quad (2.4)$$

$$\text{s.t.} \quad R_{it} = \sum_j f_{ijt} \quad \forall i, t \quad (2.5)$$

$$R_{jt} = R_{jt}^0 + \sum_{s=t-\tau_{ij}^{\max}}^{t-\tau_{ij}^{\min}} \pi_{ij} f_{ijs} \quad \forall j, t \quad (2.6)$$

where  $R_{jt}^0$  is used as an initializing vector. That means it is not a variable to be set by the optimization procedure, but input data that determines the initial state of the system (i.e., initial position of the vessel). For example, if the vessel is positioned in region  $a$  in the first period ( $t = 1$ ), then  $R_{a1}^0 = 1$  and  $R_{jt}^0 = 0$  for any other  $j$ - $t$  node.

The objective is to maximize expected earnings given by (2.4). Constraints (2.5) and (2.6) ensure proper flow balance at every  $i$ - $t$  node. This model is a version of the minimum-cost flow problem. Therefore, the problem can be solved by a dynamic programming scheme instead of standard linear programming. Both procedures are fast, so the choice depends on personal preferences.

In dynamic programming, we iterate backwards through time and reconstruct the performance in a retrospective manner. Notice that a decision made for a ship positioned in region  $i$  at time  $t$  does not depend on any of the previous decisions.

We define values  $V_{it}^X$  to store expected earnings generated from time  $t$  till the end of the planning horizon by a vessel positioned in region  $i$  at time  $t$ , when the corresponding strategy  $X$  ( $o$  or  $c$ ) is applied. Values  $V_{it}^o$  can be obtained as dual variables to constraints (2.5) if classical linear programming is used to solve the optimization model (2.4) – (2.6). The values are used in further analysis (sections 2.2.2 and 2.2.3).

One decision of the destination will almost always be more profitable than others, that is, it will yield higher expected earnings till the end of the planning horizon, than others. Thus, we obtain a binary solution, where all the decisions  $x_{ijt}^o$  are either 1 or 0; 1 in the case of the most profitable destination, 0 otherwise. In the rare case that more destinations lead to the same expected earnings, we can choose arbitrarily among them. Accordingly, it is possible to have fractional decisions, but we do not lose any value by considering only binary ones.

The procedure for the calculation of values  $V_{it}^o$  and the determination of optimal decisions at the time is summarized in Algorithm 1. As a temporary variable, we introduce  $W_{ijt}^o$  denoting expected earnings generated from time  $t$  till the end of the planning horizon by a vessel positioned in region  $i$  at time  $t$  and making the consecutive decision “go to  $j$ ”.  $W_{ijt}^o$  is used to store the earnings of all the available alternative destinations  $j$  for the ship located in  $i$  at time  $t$ . The optimal decision is obtained by simple comparison of these candidate values (rows 4 – 7 in Algorithm 1).

A similar approach is used for simulation of the coin strategy. In fact, the calculation of  $V_{it}^c$  is simpler, since it does not include an optimal destination decision. The expected earnings are constructed by summing expected contributions of all alternatives with non-zero probability of choice  $x_{ijt}^c$ . This procedure is described in Algorithm 2.

The simulation of trade is the same for both strategies, once the values  $V_{it}^X$  are computed. To observe the development of the expected earnings over time, we define variable  $M_t^X$  where we store the sum of expected earnings from time period  $t$ . We assume that the payment is spread over all days of the trip duration, not concentrated in the first day of the trip as in (2.3). That is, every day of the trip duration the operator receives the rate  $r_{ijt}$ , where  $t$  is the first period of the trip. Notice that this interpretation is simply for the sake of results visualization and has no impact on the chosen decision. The simulation of trade is summarized in Algorithm 3.

---

**Algorithm 1** Values for oracle strategy

---

```
1: Set  $W_{ijt}^o = 0, x_{ijt}^o = 0, V_{it}^o = 0 \forall i, j, t$ 
2: for  $t := T$  to 1 do
3:   for  $i \in \mathcal{I}$  do
4:      $j = \arg \max_j W_{ijt}^o$ 
5:      $V_{it}^o = W_{ijt}^o$ 
6:     if  $W_{ijt}^o > 0$  then
7:        $x_{ijt}^o = 1$ 
8:   for  $j \in \mathcal{I}$  do
9:     for  $i \in \mathcal{I}$  do
10:      for  $\tau := \tau_{ij}^{\min}$  to  $\tau_{ij}^{\max}$  do
11:         $s = t - \tau$ 
12:        if  $s > 0$  then
13:           $W_{ijs}^o = W_{ijs}^o + \pi_{ij} (r_{ijs} \tau + V_{jt}^o)$ 
```

---

---

**Algorithm 2** Values for coin strategy

---

```
1: Set  $W_{ijt}^c = 0, x_{ijt}^c = 0, V_{it}^c = 0 \forall i, j, t$ 
2: for  $t := T$  to 1 do
3:   for  $i \in \mathcal{I}$  do
4:     for  $j \in \mathcal{I}$  do
5:       for  $\tau := \tau_{ij}^{\min}$  to  $\tau_{ij}^{\max}$  do
6:          $s = t - \tau$ 
7:         if  $s > 0$  then
8:            $V_{is}^c = V_{is}^c + \pi_{ij} x_{ijs}^c (r_{ijs} \tau + V_{jt}^c)$ 
```

---

---

**Algorithm 3** Trade simulation (same for both type of strategies)

---

```

1: Set  $M_t = 0, R_{it} = 0 \quad \forall i, j, t$ 
2: Set the initial position of the ship, for example  $R_{i_01} = 1$ 
3: for  $t := 1$  to  $T$  do
4:   for  $i \in \mathcal{I}$  do
5:     for  $j \in \mathcal{I}$  do
6:       for  $\tau := \tau_{ij}^{\min}$  to  $\tau_{ij}^{\max}$  do
7:         if  $t + \tau \leq T$  then
8:            $R_{js} = R_{js} + \pi_{ij} x_{ijt} R_{it}$ 
9:           for  $s := t$  to  $t + \tau$  do
10:             $M_s = M_s + \pi_{ij} x_{ijt} r_{ijt} R_{it}$ 

```

---

**Value of geographical switching**

By the procedure described in Algorithm 1, we obtain optimal decisions  $x_{ijt}^o$  indicating the next destination  $j$  given that a vessel is positioned in  $i$  at time  $t$  in the case of known rates. In this section, we introduce the methodology used for the evaluation of how important it is to have the optimal decision (at a particular point in space and time) compared to its alternatives.

We denote by  $Z_{ijt}^X$  the expected earnings generated from time  $t$  till the end of the planning horizon by a vessel positioned in region  $i$  at time  $t$  that makes the consecutive decision “go to  $j$ ” and trade according to the corresponding strategy  $X$  ( $o$  or  $c$ ) after that. We can calculate the values  $Z_{ijt}^X$  from  $V_{it}^X$  by the simple formula:

$$Z_{ijt}^X = \sum_{s=\tau_{ij}^{\min}}^{\tau_{ij}^{\max}} \pi_{ij} (r_{ijt} s + V_{js}^X). \quad (2.7)$$

We evaluate  $Z_{ijt}^X$  for all possible destinations  $j$  and compare the results. From these values, we can observe how valuable it was to choose the best decision, compared to the second best, etc. to the worst one. If the difference between some decisions is close to zero, the operator is (almost) indifferent between the decisions at that time and place. Such comparisons provide insight into the dynamics of the market, highlights its spatial (in)efficiencies and the (in)capability of the market to anticipate future rates. Further, it reveals some exploitable opportunities in freight mispricing that have occurred in the past.

It is also useful to study the discrepancy between  $Z^o$  and  $Z^c$ , since it reveals the impact of planning several steps ahead. This is discussed in Section 2.3, where an empirical analysis of the market is presented.

### 2.2.3 Limited foresight for two-region case

In this section, we describe the methodology used for cases where rates are known for  $\Delta$  future consecutive days and this horizon of known rates is very short – usually allowing us to evaluate just a couple of future steps. We describe our approach for the case with only two regions<sup>5</sup>. The operator can either undertake an intra-region (in short  $A$ ) trip  $ii$  or inter-region (in short  $E$ ) trip  $ij$ , where  $i \neq j$ .

Since we can evaluate only a few future trips, we can not purely maximize expected earnings over the horizon of known rates to find the optimal decisions due to the different states (both in time and space) in which a vessel may occur at the end of the horizon. To illustrate this point in a simple example, let us assume trip  $A$  takes 25 days and trip  $E$  45 days. When making a decision in region  $i$  with 40 days of foresight (known rates), we can evaluate three possible future plans: either to take two intra-region trips  $AA$ , the inter-region trip  $E$ , or first one intra-region followed by the inter-region trip  $AE$ . With the  $AA$  sequence we have guaranteed earnings for 50 trading days, with  $E$  for 45 days and with  $AE$  for 70 days. Moreover, the vessel might end up in a different region at the end of each partial sequence; for  $A$  the vessel stays in  $i$ ; for  $E$  and  $AE$  the vessel relocates to  $j$ . Region  $j$  might be a better (or worse) location with higher (lower) rates, or better in general, but relatively poor 40 days from now. Thus, it is clear that it does not make much sense to purely maximize earnings obtainable over the short horizon without taking into consideration consequences of the decisions in the unforeseen future.

Let us point out that even though  $E$  takes 45 days, hence we cannot evaluate  $EA$  and  $EE$  exactly, the knowledge of rates in 40 days provides a useful hint for the estimation of rates in 45 days. Hence, it is advantageous to take it into account when making a decision.

In the real-life decision-making process, market participants (operators, shipowners) think about the consequences of the decisions. Even though they do not have available exact information about the future rates, they have other observations of the market’s condition, for instance forward curves, or the knowledge of seasonal effects (for example the harvest season in South America). These different inputs are usually processed in an intuitive way (“gut feeling”) based on years of experience. We mimic this natural way of decision-making in shipping by introducing a learning algorithm, based on a neural network, that aims to make good decisions by learning from the past. Instead of the diverse input streams that real participants consider, we rely only on rates-based inputs that are obtainable from the foresight horizon. In particular, we take into account immediate

---

<sup>5</sup>It should be possible to expand the idea into multi-region setting. However, since our method is essentially a heuristic (That is, we cannot guarantee any theoretical performance, we can only provide numerical tests.), we describe only the two-region case that is evaluated on real data (Section 2.3).



earnings from the partial sequences that can be evaluated (as in the example above) and changes of the rates between “today” and the most distant day in the foreseeable future (since they carry information about the market development).

To implement this approach, we split the overall time horizon  $\mathcal{T}$  into two parts:  $\mathcal{T}_\alpha = \{1, 2, \dots, T_\alpha\}$  and  $\mathcal{T}_\beta = \{T_\alpha + 1, \dots, T\}$ , where  $\mathcal{T}_\alpha$  represents the past (training horizon in machine learning vocabulary), from which the algorithm learns, and  $\mathcal{T}_\beta$  is used for evaluation of learned behavior (testing horizon).

The learning process is realized through a function  $S_{it}$ , evaluated on  $\mathcal{T}_\alpha$ , showing the value of switching from region  $i$  (to  $j$ ) at time  $t$ . The function is defined as follows:

$$S_{it} = Z_{ijt}^o - Z_{iit}^o, \text{ where } i \neq j. \quad (2.8)$$

That is, the function expresses how much more (or less if the value is negative) of the expected earnings an operator makes when he decides to undertake an inter-region trip (i.e., switching region) compared to performing the intra-region trip. In both cases, it is assumed that after such a decision, the operator continues trading optimally (according to the oracle strategy, see Section 2.2.2) till the end of the planning horizon  $\mathcal{T}_\alpha$ . Note that  $S_{it}$  can be evaluated only looking backwards (offline), since future consequences of particular decisions are included in the value.

The function carries two important pieces of information: Firstly, it implicitly says, which decision was the optimal one in the past. If the value is positive, it was optimal to switch market (undertake the inter-region trip), and if it is negative, it was optimal to stay in the current region (and undertake the intra-region trip). Secondly, it expresses how important (valuable) it was to make the optimal decision compared to its alternative. Our aim is to make our algorithm able to predict (as well as possible) the value of  $S_{it}$  by using only information available up to  $(t + \Delta)$  day (online). Let us denote the predicted values  $\hat{S}_{it}$ .

We calculate expected earnings obtained from the *partial sequences*, i.e. combination of possible moves that start in the foresight period. Even though it is theoretically possible to assume even sequences where a consecutive decisions differ for different dates of arrival  $\tau_{ij} \in \{\tau_{ij}^{\min}, \tau_{ij}^{\max}\}$ , we ignore that option, since it leads to too many input parameters (many of them highly correlated) that would make the training process difficult. So only the sequences of  $A$ s and  $E$ s, where each consecutive decision is fixed for all arrival days are considered. Expected earnings obtained by each sequence together with rates’ level at the time  $t + \Delta$  are the inputs of the learning algorithm.

After the training phase, we let the algorithm predict  $\hat{S}_{it}$  on the time horizon  $\mathcal{T}_\beta$ . That

is, we construct the predicted switching function by using only information available in the limited foresight horizon  $\{t, t + 1, \dots, t + \Delta\}$ . Subsequently, we reconstruct the decisions  $\hat{x}_{ijt}$  from the function  $\hat{S}_{it}$  as follows: if  $\hat{S}_{it} \geq 0$ ,  $\hat{x}_{ijt} = 1$  for  $i \neq j$  and  $\hat{x}_{iit} = 0$ . Vice versa, if  $\hat{S}_{it} < 0$ ,  $\hat{x}_{iit} = 1$  and  $\hat{x}_{ijt} = 0$  for  $i \neq j$ .

We can then simulate the trade for decisions  $\hat{x}_{ijt}$  by using Algorithm 3 and compare the results with the oracle and coin strategies. Since no information from  $\mathcal{T}_\beta$  was used to train the network, the results obtained on this interval are considered as out-of-sample values that can be compared to results obtained by the coin and oracle strategy. We perform this analysis for different values of  $\Delta$  to assess the impact of the length of the foresight horizon.

The specifications of the neural network used in our case are described in the numerical part of the paper (Section 2.3). We prefer to use a general term “learning algorithm” in this section to emphasize that the methodology is not limited to usage of neural networks, but in principle, any regression algorithm can be used. We chose the neural network scheme because of its ability to handle the non-linearities that occur in our problem.

It is fair to note here that there are plenty of modifications of the neural network and other methods from the machine learning field or statistics that have not been tested and might perform better at this specific task than the presented approach. For example, it is possible to let the model learn the decisions directly instead of learning the switching function value and reconstructing the decisions afterwards. The disadvantage of such approach would be that it does not take into consideration the different importance (value) of optimal decisions at different points. In other words, a higher number of correctly forecasted optimal decisions does not have to lead to higher earnings.

However, the aim of this paper is not to search for the best possible method for this particular task. The problem itself is still unrealistic in real life due to the requirement of exact predictions of rates. The goal is to demonstrate whether some significant portion of the theoretical gains from perfect knowledge can be captured with only a few future steps taken into account.

## 2.3 Numerical results

In the methodology part for perfect foresight, we formulated a general model for an arbitrary number of regions. In our numerical analysis, we work with a simplified model of the world with only two regions – Atlantic and Pacific. This creates four possible routes for trade: trans-Atlantic (TA), trans-Pacific (TP), from the Atlantic to the Pacific basin, also called fronthaul (FH) in the shipping literature, and finally the backhaul (BH) trip from the Pacific to the Atlantic (see for instance Adland and Jia (2017)). For each of the routes we have rates (prices) for the period July 2005 – April 2017 provided by the

Baltic exchange and obtained by Clarkson Research (2017) for three segments of dry bulk shipping according to the standardized size of vessels: Supramax (52,000 metric tonnes (mt) carrying capacity), Panamax (74,000 mt) and Capesize (172,000 mt).

For all sectors we assume the same discrete uniform distribution of trip durations with the following ranges: TA: 30 – 45 days, TP: 30 – 40 days, FH: 60 – 70 days, BH: 60 – 70 days. For the coin strategy, we set probabilities  $x_{TA}^c = 0.634$ ,  $x_{FH}^c = 1 - x_{TA}^c$ ,  $x_{TP}^c = 0.65$ ,  $x_{BH}^c = 1 - x_{TP}^c$ . These probabilities are set such that the vessel spends an equal number of days on each route (in expectation). Although it is not shown here, the results would not be significantly different if probabilities were 0.5 for each route according to our numerical tests.

### 2.3.1 Perfect foresight

Cumulative earnings for the perfect foresight case in the Capesize market, computed at every time period  $t$  as  $\sum_{s=1}^t M_s^X$ , are shown in Figure 2.1. The steeper growth of cumulative earnings till the second half of 2008 is caused by higher rates in this period. After the rates dropped sharply following the financial crisis (for instance, the BCI<sup>6</sup> dropped from over \$200,000/day in June 2008 to less than \$5,000/day in October 2008), the absolute differences decrease as there are lower regional rate differences in a competitive and oversupplied market (Adland et al., 2017b), but the relative differences increase as there remains some inability to correctly anticipate future prices.

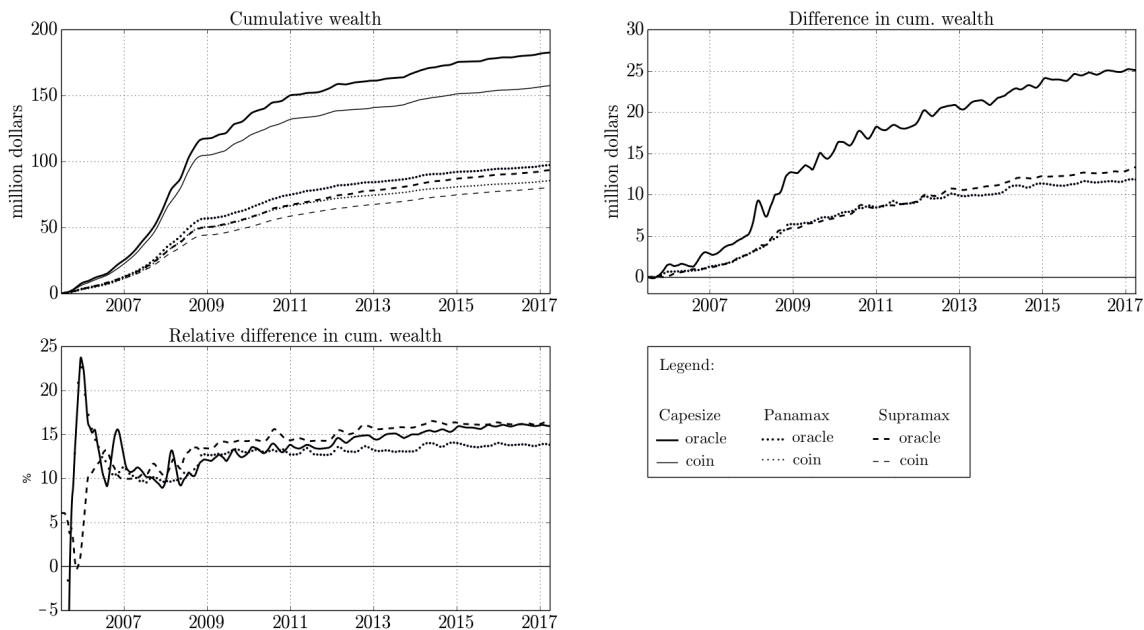


Figure 2.1: Cumulative earnings for the oracle and coin strategy in Capesize market.

<sup>6</sup>Baltic Capesize Index

Table 2.1 compares different segments of bulk shipping by vessel size. We see that the relative gains for the Capesize and Supramax segments are around 16%, while the mid-size Panamax sector shows approximately 14% outperformance over the period 2006 – 2016. If we exclude the boom period and compare cumulative earnings obtained from 2009 to 2016, relative differences are higher, namely 19.5% for Supramax sector, 15.5% for Panamax, and ca 24% for the Capesize sector. In Table 2.1 we can also see that the gains from perfect knowledge differ from year to year. However, it is necessary to emphasize here that the comparison across years is not entirely precise. The values shown in Table 2.1 for every year are calculated as the sum of  $M_s^X$  over that particular year, but the optimization problem (2.4) - (2.6) is solved over the whole period (July 2005 – April 2017). This means that the decisions are taken such that the cumulative expected earnings are maximized at the end of the period (April 2017). Therefore, we may have situations where the earnings in one year is sacrificed for a better position of the ship in the following year. Since the start/end conditions differ across years, the comparison is imperfect. However, detailed analysis of the performance in individual years is not the aim of this paper.

Year	Supramax			Panamax			Capesize		
	<i>o</i>	<i>c</i>	%	<i>o</i>	<i>c</i>	%	<i>o</i>	<i>c</i>	%
2006	9.09	8.05	12.85	9.08	8.34	8.88	17.15	15.82	8.39
2007	18.29	16.54	10.6	21.48	19.74	8.84	43.14	40.06	7.69
2008	19.57	16.51	18.53	22.90	19.51	17.37	49.27	42.50	15.94
2009	7.07	5.86	20.54	7.43	6.43	15.52	17.02	14.29	19.14
2010	9.62	8.37	14.84	10.47	9.43	11.03	15.55	12.76	21.84
2011	6.18	5.32	16.28	5.86	5.17	13.52	6.26	5.65	10.9
2012	4.95	3.60	37.5	3.73	2.99	24.64	4.94	3.30	49.89
2013	4.21	3.59	17.51	3.59	3.28	9.48	6.19	4.84	28.04
2014	4.71	3.72	26.55	4.13	2.98	38.48	7.57	5.58	35.73
2015	3.10	2.65	16.73	2.40	2.11	13.77	3.44	2.73	26.12
2016	2.36	2.18	8.27	2.00	1.90	4.93	3.18	2.62	21.18
$\sum_{2006-2016}$	89.15	76.39	16.70	93.07	81.88	13.67	173.71	150.15	15.69
$\sum_{2009-2016}$	42.20	35.29	19.58	39.61	34.29	15.51	64.15	51.77	23.91

Table 2.1: Overview of annual results obtained by oracle (*o*) and coin (*c*) strategy (all in millions \$), % is the relative difference.

## Value of geographical switching

In our two-region world, we construct the switching function  $S_{it}^X$  for both strategies and each region  $i$  according to (2.8). The functions are shown in Figure 2.2 for the Capesize segment. We observe quite low discrepancy between the functions for the coin and oracle strategies. There also appears to be a correlation between market level and volatility of the switching function. In the post-crisis period, we see asymmetry in the function for the Atlantic market. Although not shown here, it is observed across all sectors. This suggests that an incorrect decision to go to the Pacific basin is potentially more costly than to choose incorrectly to stay in the Atlantic basin. In other words, the vessel should remain in the Atlantic basin once it is located there, unless there is strong evidence in favour of a FH trip. This corresponds to findings in Adland et al. (2017a) suggesting that a persistent Atlantic premium is the main contributor to outperformance.

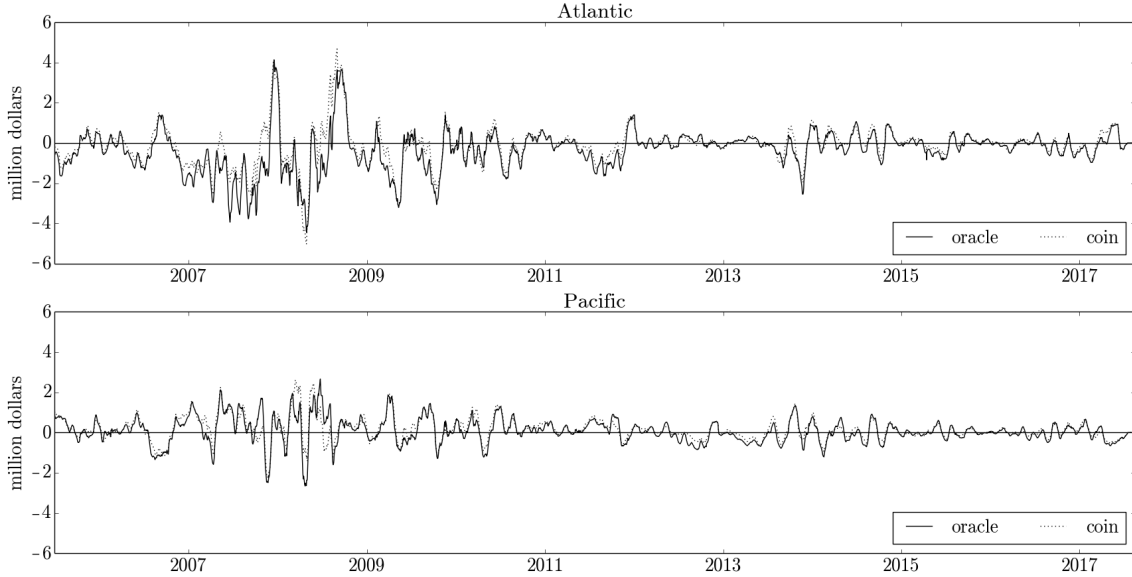


Figure 2.2: Value of geographical switching function - Capesize sector

To understand the low discrepancy between the switching functions for the oracle and coin strategies and its implications, it is important to realize the following: The switching function is defined as a difference of expected earnings generated by two possible alternative decisions and subsequent trade. We can decompose the value of the switching function into two parts. The first part is the contribution of the very first decision, that is, the difference in earnings defined as:

$$\sum_{s=\tau_{ij}^{\min}}^{\tau_{ij}^{\max}} r_{ijt}s - \sum_{z=\tau_{ii}^{\min}}^{\tau_{ii}^{\max}} r_{iit}z \quad (2.9)$$

The second part comes from the difference in earnings from subsequent trading after

the first decision is made. In the coin strategy case, the randomized choice of the next destinations is applied from initial positions  $j$ - $s$  and  $i$ - $z$ , respectively, and its difference contributes to the value of switching function. In the oracle case, the second part is given as the difference between results of optimal positioning from initial positions  $j$ - $s$  and  $i$ - $z$ .

We see that the contribution of the first decision (2.9) is shared by both strategies. Therefore, the discrepancy between the switching functions of the two strategies is caused by the second part. Since the discrepancy is low, the difference between randomized and optimal sequences, starting from the same initial positions  $j$ - $s$  and  $i$ - $z$ , is very similar. That suggests a great importance of the first decision as the main source of extra earnings, if it is optimally chosen.

### 2.3.2 Limited foresight horizon

We use  $\Delta \in \{20, 50, 80\}$  days long horizons of foresight for evaluating the proposed strategy. These lengths are chosen such that the number of possible trips that can be exactly evaluated is expanding. With 20 days of foresight, we know the earnings from the current trip, and this would be the same if we had no future information, but we may yet gain some benefit from knowing regional price trends in the next 20 days. With 50 days of foresight and assumed trip durations, we can exactly evaluate earnings from two intra-region (A) trips, i.e. two TAs or TPs - depending on the position of the vessel, only one inter-region (E) trip, FH or BH), or a combination of one intra-region trip followed by the inter-region trip. In addition, the operator is in a better position to estimate future rates because of the observation of price changes in the next 50 days. In the case of 80 days foresight, it is possible to at least partly evaluate the following combinations of trips:  $AAA^7$ ,  $AAE^7$ ,  $AE$ ,  $EA$ ,  $EE$ .

The training phase of the model should not depend on any information from the testing period. However, we believe it is realistic to assume that high freight rates as they were observed in the boom period prior to financial crisis in 2008 will not be repeated any time soon. Rather, we expect similar order of magnitude in freight rates and their differences as observed in the post-crisis period. Therefore, we exclude the boom period from the training horizon  $\mathcal{T}_\alpha$  and consider only the period 1.1.2009 - 10.11.2013. Since the last decisions are impacted by the end of the period, we do not consider last 120 days of the training period for actual training.

We use a neural network with an input layer and one hidden layer of the same size as the input vector. Both layers use rectified linear units as activation function (Nair and

---

<sup>7</sup> In the case of TA trip, we can evaluate this sequence only partially. That is, the duration of first two trips exceeds 80 days in some scenarios.

Hinton, 2010). Then, a single neuron output layer with sigmoid activation function follows. As a loss function, the standard mean squared error is applied. The output is assumed to be in the interval  $(0, 1)$ . Thus, we need to transform the switching function into that interval. We test two different methods. The first is simple *linear* transformation, that is, the minimum of the switching function on the training horizon is projected to 0 and maximum to 1 with linear scaling between them. This approach has a weakness when the maximum and minimum of the switching functions on the training horizon are asymmetric with respect to the zero axis, in other words, one of the values is significantly larger in absolute value than the other. Then, the zero value, which is our main interest since the decision is changing there, would be projected too far from the middle point, which would not be optimal for the output layer. Therefore, we introduce a second *adjusted* linear approach, where we first transform the negative part of the switching function into the  $(0, 0.5)$  interval and the positive part into the  $(0.5, 1)$  interval. This approach projects 0 correctly to 0.5, which is desired but, on the other hand, it distorts the loss function applied on the evaluation of predictions during the training period.

We also experiment with the input values, which can be divided into trading contributions, i.e., the set of earnings obtained by selected strategies, and freight rates at the end of the foresight horizon. We note that the optimal decisions would be the same if all rates are increased by a constant value, in other words, it is not the absolute value of freight rates, but the regional differences that are crucial for making a decision.

We consider four different settings of inputs by creating all combinations of absolute values/differences of contributions/freight rates. For each combination we apply both transformations, so in total we test 8 instances with different settings for each sector.

After the training phase, we let the model predict values of the switching function on the testing horizon, which is Oct 2013 - Dec 2016. From the value of switching, we reconstruct decisions for each  $i, t$  position and simulate the trade according to Algorithm 3.

We compare these results with the results obtained by the oracle and coin strategies. In Table 2.2, we report which fractions (in percentage) of the oracle strategy gains (compared to the coin strategy) can be captured by this approach for different input settings. We show only the final numbers reached at the end of 2016. On average, we observe the expected result that a longer foresight horizon brings better performance both in average gains and stability of the results. We see, for example, that the results for 20 days foresight in the Capesize segment differ from -20% up to +38% for different settings. This gives the impression of a great deal of randomness involved in the computation. We cannot conclude on the comparison across different vessel size segments, since we performed this analysis only on one date, which splits the period into the training and test samples. Thus, it is possible that such a splitting was more favorable to one sector than the other. For

instance, it might happen that some pattern from the training horizon repeated in the testing sample, but cannot be generalized.

		LF 20			LF 50			LF 80		
		S	P	C	S	P	C	S	P	C
lin	ar-ac	38.34	47.19	35.3	89.15	75.33	48.7	61.56	96.27	82.42
	ar-dc	54.64	37.87	34.47	81.74	48.98	53.28	87.86	95.29	79.05
	dr-ac	62.34	-9.89	11.42	87.72	69.45	68.7	61.19	86.12	81.99
	dr-dc	62.11	49.94	-20.35	70.01	72.09	74.98	89.06	83.3	77.09
adj	ar-ac	62.96	21.84	38.19	92.06	72.18	65.26	85.04	96.46	74.13
	ar-dc	72.27	12.85	38.36	84.27	75.03	50.85	83.6	90.77	77.75
	dr-ac	59.67	27.23	15.47	68.19	51.15	67.38	86.91	92.8	80.91
	dr-dc	65.24	38.78	-10.99	90.1	41.32	71.05	78.26	78.82	73.15

Table 2.2: Percentage of the oracle gains over the coin trade for limited foresight horizon. LF X - limited foresight for X future days; S - Supramax, P - Panamax, C - Capesize; lin - linear, adj - adjusted linear transformation of output value; ar - absolute rates, dr - difference of rates, ac - absolute contribution, dc - difference of contributions as input.

In Figure 2.3 we show the simulation of the trade throughout the whole testing period and all sectors with simple linear transformation of output and differences in contributions (but not in rates) being used as input setting.

We see in this figure that for all foresight (perfect or limited) periods across all sectors, the highest outperformance of the market (the steepest curves in the middle row figures) is realized in the first part of the testing horizon, that is, mainly the year 2014. This is consistent with the shape of the Value of geographical switching function (Figure 2.2). The function is quite volatile in 2014 in both regions, which means greater opportunities for exploitation. We explain the ups and downs by the time needed to relocate a vessel from one region to another. That is, once an opportunity arises in one of the region and too many market participants decide to relocate their vessels into the more profitable region, they create a new (and opposite) imbalance between supply and demand which is reflected in changes in freight rates by the time of arrival. Thus, the participants receive (a negative) feedback on their decisions with a significant delay. If a participant was endowed with the knowledge of freight rates for a consecutive decision, he could avoid this “following the herd” mentality and resist a temptation to make a decision based on a belief that “today’s spatial rate differentials” will continue.

We see in Figure 2.2 that after the volatile year 2014, the market stabilized and exploitable opportunities significantly decreased. Thus, even with the foresight of the freight rates it is difficult to outperform the market, especially if the foresight horizon is short –



there are almost no gains (for this setting of the policy) for 20 and 50 days long horizons of foresight in 2015 and 2016, see the middle row in Figure 2.3. This reflects the significant oversupply of ships during this period, meaning that even minor spatial opportunities could be captured immediately by waiting tonnage.

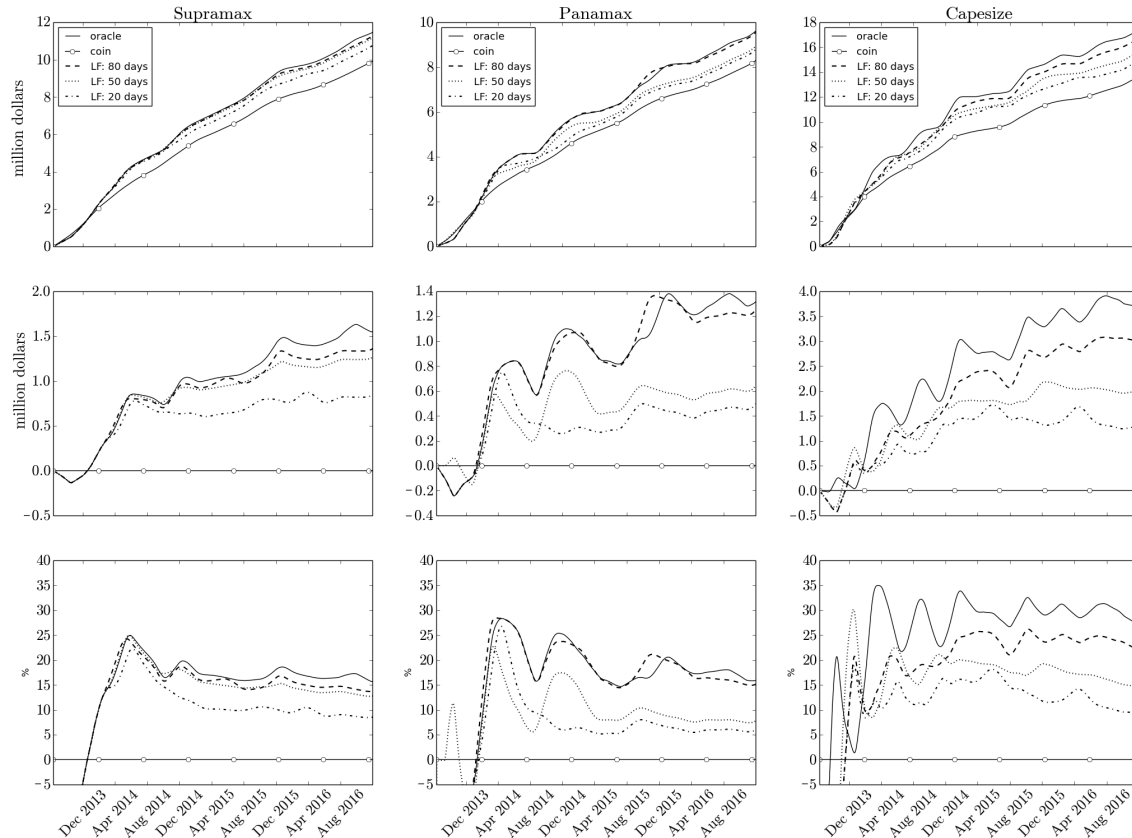


Figure 2.3: Demonstration of limited foresight gains.

## 2.4 Conclusion

In this paper, we have evaluated the degree of spatial efficiency in the drybulk freight market. The main objective was to determine the upper bound of earnings obtained by optimal vessel positioning in space and time by assuming perfect knowledge of future regional freight rates for a given time period. The upper bound is a first indicator of whether or not it makes sense to continue developing chartering strategies based on imperfect forecasts of future rates.

To achieve this, we have formulated a simple model for optimal vessel repositioning through space and time that takes into account uncertainty in travel times. Although we work with a two-region world in the numerical section, the model is formulated in a general way for an arbitrary number of regions. Empirically, we have shown that with

perfect knowledge it is possible to achieve approximately 15% higher cumulative earnings in the years 2006 - 2016. This number is distorted by exceptionally high freight rates up to late 2008, after which the financial crisis hit the shipping market and freight rates dropped sharply. However, with the lower market level in the post-crisis period, the relative outperformance is higher, at around 20% for Supramax, 15% for Panamax, and 23% for Capesize sector in the period 2009 - 2016.

We have also extended the concept of perfect knowledge to the case where we assume perfect knowledge only on a limited (and relatively short) time horizon. This required a different methodology to overcome the so-called end of the planning horizon issue. For that purpose, we have introduced a new approach based on a neural network model that learns to predict the optimal decisions obtained in the perfect foresight section, but in this case, using only information from the foresight horizon. This model naturally produces better results the longer the foresight horizon is. The assumption of perfect foresight on a limited horizon is still unrealistic, but our results demonstrate that it is possible to capture a large portion of the theoretical “perfect foresight” value with only a few future moves taken into consideration.

It is also possible to use the optimization model to provide further insight into the market. For instance, we have calculated the value of each decision at every point in space and time. With that, we have observed an asymmetry in the geographical switching function for the Atlantic region. That is, an incorrect decision “go to Pacific” is potentially more costly if applied at the wrong time than an incorrect decision to “stay in the Atlantic”.

In our view, the empirical findings reveal a big potential for exploiting spatial inefficiencies by a sophisticated chartering strategy. A natural continuation of this research would be to apply stochastic programming to handle uncertainty in freight rates. For instance, to use a scenario tree for describing the future development of freight rates instead of the assumption of perfect foresight. In contrast to many other applications of stochastic programming, the difficult part of such an approach would be to construct the scenarios (and build predictive models to do that), not the subsequent policy search.

## References

- Adland, R., Benth, F. E., and Koekebakker, S. (2017a). Multivariate modeling and analysis of regional ocean freight rates. *Transportation Research Part E: Logistics and Transportation Review*.
- Adland, R., Bjercknes, F., and Herje, C. (2017b). Spatial efficiency in the bulk freight market. *Maritime Policy & Management*, 44(4):413–425.
- Adland, R., Cariou, P., and Wolff, F.-C. (2016). The influence of charterers and owners

- on bulk shipping freight rates. *Transportation Research Part E: Logistics and Transportation Review*, 86(Supplement C):69 – 82.
- Adland, R. and Cullinane, K. (2006). The non-linear dynamics of spot freight rates in tanker markets. *Transportation Research Part E: Logistics and Transportation Review*, 42(3):211 – 224.
- Adland, R. and Jia, H. (2017). Simulating physical basis risks in the capesize freight market. *Maritime Economics & Logistics*, 19(2):196–210.
- Bellman, R. (1957). *Dynamic Programming*. Princeton University Press, Princeton, N.J.
- Benth, F. E. and Koekebakker, S. (2016). Stochastic modeling of supramax spot and forward freight rates. *Maritime Economics & Logistics*, 18(4):391–413.
- Berg-Andreassen, J. (1997). Efficiency and interconnectivity in international shipping markets. *International Journal of Transport Economics / Rivista internazionale di economia dei trasporti*, 24(2):241–257.
- Berg-Andreassen, J. A. (1996). Some properties of international maritime statistics. *Maritime Policy & Management*, 23(4):381–395.
- Bertsekas, D. P. and Tsitsiklis, J. N. (1996). *Neuro-Dynamic Programming*. Athena Scientific, 1st edition.
- Bușoni, L., Lazaric, A., Ghavamzadeh, M., Munos, R., Babuška, R., and De Schutter, B. (2012). *Least-Squares Methods for Policy Iteration*, pages 75–109. Springer Berlin Heidelberg, Berlin, Heidelberg.
- Christiansen, M., Fagerholt, K., Nygreen, B., and Ronen, D. (2013). Ship routing and scheduling in the new millennium. *European Journal of Operational Research*, 228(3):467 – 483.
- Clarkson Research (2017). Shipping intelligence network, [www.clarksons.net](http://www.clarksons.net).
- Glen, D. and Rogers, P. (1997). Does weight matter? a statistical analysis of the sscapsize index. *Maritime Policy & Management*, 24(4):351–364.
- Koekebakker, S., Adland, R., and Sødal, S. (2006). Are spot freight rates stationary? *Journal of Transport Economics and Policy*, 40(3):449–472.
- Laulajainen, R. (2006). A static theory of dry bulk freight rates by route. *Maritime Policy & Management*, 33(5):447–461.
- Laulajainen, R. (2007). Dry bulk shipping market inefficiency, the wide perspective. *Journal of Transport Geography*, 15(3):217 – 224.

- Laulajainen, R. (2010). Geography sets the tone to tramp routing. *International Journal of Shipping and Transport Logistics*, 2(4):364–382.
- Lium, A.-G., Crainic, T. G., and Wallace, S. W. (2009). A study of demand stochasticity in service network design. *Transportation Science*, 43(2):144–157.
- Nair, V. and Hinton, G. E. (2010). Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th International Conference on International Conference on Machine Learning, ICML’10*, pages 807–814, USA. Omnipress.
- Powell, W. B. (1996). A stochastic formulation of the dynamic assignment problem, with an application to truckload motor carriers. *Transportation Science*, 30(3):195–219.
- Powell, W. B. (2014). Clearing the jungle of stochastic optimization. In *Bridging Data and Decisions*.
- Simão, H. P., Day, J., George, A. P., Gifford, T., Nienow, J., and Powell, W. B. (2009). An approximate dynamic programming algorithm for large-scale fleet management: A case application. *Transportation Science*, 43(2):178–197.
- Stopford, M. (2009). *Maritime Economics, third ed.* Routledge, London.
- Sutton, R. S. and Barto, A. G. (1998). *Introduction to Reinforcement Learning*. MIT Press, Cambridge, MA, USA, 1st edition.
- Topaloglu, H. and Powell, W. B. (2006). Dynamic-programming approximations for stochastic time-staged integer multicommodity-flow problems. *INFORMS Journal on Computing*, 18(1):31–42.
- van der Laan, E. A., Kiesmüller, G., Kuik, R., Vlachos, D., and Dekker, R. (2004). *Stochastic Inventory Control for Product Recovery Management*, pages 181–220. Springer Berlin Heidelberg, Berlin, Heidelberg.
- Veenstra, A. W. and Franses, P. H. (1997). A co-integration approach to forecasting freight rates in the dry bulk shipping sector. *Transportation Research Part A: Policy and Practice*, 31(6):447 – 458.
- Wallace, S. W. (1986). Solving stochastic programs with network recourse. *Networks*, 16(3):295–317.
- Wiering, M. and van Otterlo, M. (2012). *Reinforcement Learning: State-Of-The-Art*, volume 12. Springer-Verlag, Berlin Heidelberg.

# Stochastic programs with binary distributions: Structural properties of scenario trees and algorithms

Vit Prochazka<sup>□</sup> Stein W. Wallace<sup>□</sup>

<sup>□</sup>Department of Business and Management Science  
Norwegian School of Economics, Bergen, Norway

## Abstract

Binary random variables often refer to such as customers that are present or not, roads that are open or not, machines that are operable or not. At the same time, stochastic programs often apply to situations where penalties are accumulated when demand is not met, travel times are too long, or profits too low. Typical for these situations is that the penalties imply a partial order on the scenarios, leading to a partition of the scenarios into two sets: Those that can result in penalties for some decisions, and those that never lead to penalties. We demonstrate how this observation can be used to efficiently calculate out-of-sample values, find good scenario trees and generally simplify calculations. Most of our observations apply to general integer random variables, and not just the 0/1 case.

**Keywords:** stochastic programming, scenarios, binary random variables, algorithms

### 3.1 Introduction

Stochastic programming is getting more attention in operations research and has many applications in real processes where uncertainties are present, see for example Wallace and Ziemba (2005); King and Wallace (2012b). Most approaches assume that the uncertainty is represented by a set of scenarios. In this paper we study the case of discrete random variables, with a focus on the multivariate Bernoulli (0/1) distribution. We shall do this using a stochastic version of the knapsack problem. However, the point of the paper is not to study the knapsack problem as such, rather it serves as a simple example suitable for demonstrating our issues that are more general. In this sense we use the same idea as Fortz et al. (2013).

In the case of  $n$  random variables, each taking the value 0 or 1, we have  $2^n$  possible combinations. Every combination represents a scenario with a certain probability. Even though the number of possible scenarios is finite, it rises exponentially with the number of random variables and it quickly becomes numerically infeasible to solve most optimization problems using all scenarios. When we are able to solve a problem only for a limited number of scenarios, we need to choose a subset, i.e., construct a scenario tree. The most common approach is to sample the required number of scenarios. However, in this paper we introduce a heuristic based on iterative scenario selection that provides better results than pure sampling for a given number of scenarios. We also show that pure sampling can be strengthened by taking into account that many scenarios cannot lead to penalties caused by the knapsack being too small; the scenarios cannot lead us into the tail of the outcome distribution.

Technically, binary random variables are a subset of the discrete random variables. So in a sense, if we knew how to handle the general discrete case for generating scenarios, we would also have an approach for the binary case. In fact, binary should be simpler than the general case. But when considering what binary random variables represent in a real setting, it becomes clear that having a representative set is critical in a different way than some demand, for example, that appear in integer amounts. Typical applications are the existence of a customer in a routing or facility location problem, the presence of an edge in a network, whether a machine is up in a scheduling problem, and whether the weather allows a ship to sail. And it is often the combinatorial effects in the scenarios that drive the system. Just to give a few simplistic examples: if we by accident find ourselves in a situation where two customers never appear together in a scenario for a stochastic vehicle routing problem, the optimal solution will very likely take that into account and (if meaningful) put the two on the same route, since it knows that it will only need to visit one of them (hence making route length or total demand less varying). If two edges in a network design problem cannot (according to the scenarios) be down at the same time, the optimal solution will take that into account, for example to make

sure that the probability of two given nodes being connected is rather high (by placing the two in parallel). Optimal solutions will always “look for” such anomalies and utilize them to create what seems to be very good solutions. So, in a sense, scenario generation for binaries becomes a combinatorial fight against such lack of good representation.

The quality of a decision is given by the true objective function value, that is, the evaluation of the solution over all scenarios. In order to stay coherent with the literature, we call it the out-of-sample value. In the case of the stochastic knapsack problem, and also many other problems, the objective function is given by some value term which is easily computed, and some penalty term related to for example financial risk or lack of capacity, what we might call *tail effects*. For the knapsack problem, the penalty occurs when the items we put into it are larger than the capacity of the knapsack. For such cases we introduce a procedure that computes the penalty in the minimal number of steps.

Despite the wide range of real applications, we are not aware of any paper dealing with the scenario generation process for binary random variables. For a general discussion of scenario generation, see for example Chapter 4 of King and Wallace (2012b).

## 3.2 Stochastic knapsack problem

We consider a stochastic version of the knapsack problem, where  $\mathcal{I}$  is the set of  $n$  items, item  $i$  having a value  $c_i$  and size  $w_i$ , with the knapsack having a capacity of  $W$ . All items we want to use must be picked in the first stage and if the total size is too large, we pay a unit penalty of  $d$  in the second stage. The aim is to maximize the expected profit.

Let  $S$  be the set of scenarios, having  $|S|$  elements. A scenario  $s \in S$ , with elements  $s_i = 1$  if the  $i^{\text{th}}$  item is present, 0 if not, is associated with a probability  $\pi_s$ .

For a decision vector  $x \in \{0, 1\}^n$ , where  $x_i = 1$  if item  $i$  is picked, 0 otherwise, we formulate the optimization problem

$$\max_{x \in \{0,1\}^n} \sum_s \pi_s \left( \sum_i s_i c_i x_i - d \left( \sum_i s_i w_i x_i - W \right)^+ \right), \quad (3.1)$$

where  $X^+$  takes value  $X$  if  $X \geq 0$ , 0 otherwise. This problem can be rewritten as a standard stochastic integer linear problem by introducing a new variable  $e_s$  for exceeded capacity of the knapsack in scenario  $s$ . Problem (3.1) then becomes

$$\max_{x \in \{0,1\}^n} \sum_s \pi_s \left( \sum_i s_i c_i x_i - d e_s \right) \quad (3.2)$$

$$\text{s.t.} \quad e_s \geq \sum_i s_i w_i x_i - W \quad \forall s \in S \quad (3.3)$$

$$e_s \geq 0 \quad \forall s \in S \quad (3.4)$$

If we have marginal probabilities of appearances  $p_i$  for each item, either as a starting point or computed from the identity  $\sum_s \pi_s s_i = p_i$ , we can reorganize terms in the objective function to obtain a simplified form:

$$\sum_s \pi_s \left( \sum_i s_i c_i x_i - d e_s \right) = \sum_i p_i c_i x_i - d \sum_s \pi_s e_s. \quad (3.5)$$

In the independent case, we can compute the probability of a scenario by a formula:

$$\pi_s = \prod_i \left( s_i p_i + (1 - s_i)(1 - p_i) \right). \quad (3.6)$$

In the general case we need  $2^n - 1$  parameters to describe a multivariate Bernoulli distribution. This may of course cause practical difficulties, but is not the focus of this paper; Our goal is to understand scenarios and algorithms. For simplicity, we shall therefore in the following assume that items appear independently. The general case with dependencies is discussed in Section 3.5.

If not stated otherwise, we use the notation  $\prod_i$  for  $\prod_{i \in \mathcal{I}}$  (as already used above). The same holds for the summation  $\sum$  and indices of scenarios  $s$  belonging to  $S$ .

### 3.3 Out-of-sample evaluation

Out-of-sample (o-o-s) evaluation is a way to provide the true (or approximately true) value of the objective function for a given solution, principally using the whole distribution. Only occasionally can that be done exactly for continuous distributions (for a case when that was possible, see Zhao and Wallace (2014)). Therefore, the o-o-s value is normally obtained using a large number of sampled scenarios which approximates the true distribution extremely well. With discrete distributions, the number of possible scenarios is finite. However, to stay consistent with the literature, we shall call the evaluation o-o-s, whether we use a very large sample or the full discrete distribution when calculating the true value. For the complete scenario set  $S$ , and a given a solution  $x$ , the o-o-s value  $f(x)$  is given by:

$$f(x) = \sum_s \pi_s \left( \sum_i s_i c_i x_i - d e_s \right) = \sum_i p_i c_i x_i - d \sum_s \pi_s e_s. \quad (3.7)$$

The evaluation of  $f(x)$  in (3.7) is much faster than solving (3.2) – (3.4). That means that there exist many cases where it is impossible to find the solution  $x^*$  for a given  $S$ , but it is possible to evaluate  $f(x)$  by simple enumeration. However, sometimes a straightforward evaluation over all scenarios is still too time consuming. For those cases,



we introduce a method to find the o-o-s value in a recursive manner using a minimal number of calculations.

In (3.7), the expected profit is split into two terms. The first is simply the expected profit from the selected items. This is given by the sum of expected contributions from the selected items. The second term represents the penalty for exceeding the capacity of the knapsack.

As we can see from the last term, we do not pay a penalty in all scenarios. Let us denote by  $Q(x)$  the set of scenarios for which we pay a penalty given a decision  $x$ , that is

$$Q(x) = \{s \mid \sum_i s_i w_i x_i > W\}. \quad (3.8)$$

For a decision  $x$  we define a relation  $\geq_x$  for two scenarios  $s^1$  and  $s^2$  and we state the obvious lemma that is used later in this section.

**Definition 3.3.1** *We say that  $s^1 \geq_x s^2$  if  $s^1_i x_i \geq s^2_i x_i$  for all  $i = 1, \dots, n$ .*

**Lemma 3.3.1** *If  $s^1 \geq_x s^2$  and  $s^1 \notin Q(x)$  then  $s^2 \notin Q(x)$ .*

This lemma says that if a scenario  $s^1$  does not generate a penalty for a particular decision  $x$ , then a scenario  $s^2$ , which does not include any item that is not present in the scenario  $s^1$ , also does not generate a penalty.

Vice versa, we could say that if a scenario  $s^1$  generates a penalty, then a scenario  $s^2$  which includes all selected items from  $s^1$  (that is  $s^2 \geq_x s^1$ ) also generates a penalty. This result has, however, not been used in the paper.

### 3.3.1 Recursive implementation of o-o-s

A straightforward way to obtain the o-o-s value  $f(x)$ , for a given  $x$ , is to evaluate the function over all scenarios, that is, to perform a full enumeration of the scenarios. However, this might be impossible or very time-consuming if the number of scenarios is large. Therefore, we implement a procedure that only visits those scenarios that result in penalties.

We shall utilize the simple observation that to evaluate the penalty term for a solution  $x$ , it is sufficient to compute the penalties for scenarios that belong to  $Q(x)$ . That is,

$$\sum_{s \in S} \pi_s e_s = \sum_{s \in Q(x)} \pi_s e_s. \quad (3.9)$$

Let us notice that if an item  $i$  is not chosen in the solution  $x$ , that is,  $x_i = 0$ , two scenarios that differ only in the appearance of item  $i$  will generate the same exceeded capacity. Thus we aggregate these two scenarios, this practically means that for o-o-s evaluations we ignore the random variables corresponding to items that are not selected.

Let us denote the set of selected items  $\mathcal{I}_x = \{i \in \mathcal{I} | x_i = 1\}$  and the set of all aggregated scenarios for the selected items  $S_x$ . Therefore, a scenario  $s \in S_x$  is described as a vector of length  $|\mathcal{I}_x|$ . Let us further denote by  $\bar{1} = (1, 1, \dots, 1)$  the scenario with all items from  $\mathcal{I}_x$  present.

For items that are selected, the relation  $\geq_x$  is a partial order over the set of scenarios, i.e., a relation which is reflexive, antisymmetric, and transitive. We shall use Algorithm 4 to generate a tree, based on this partial order, with root node  $\bar{1}$ . We can search over this tree in a recursive manner to visit all scenarios in  $Q(x)$  and thus compute the basis for the penalty, namely  $P_x = \sum_{s \in Q(x)} \pi_s e_s$ . Lemma 3.3.1 gives the condition for stopping the depth first search in a particular branch. By  $s = s_{i \rightarrow 0}^I$  in Algorithm 4 we mean that scenario  $s$  is obtained from scenario  $s^I$  by changing the  $i^{\text{th}}$  item in the vector from 1 to 0, while the rest of the items remain unchanged.

The procedure described in Algorithm 4 provides the o-o-s value in the minimal number of steps – we evaluate all scenarios for which the capacity of the knapsack is exceeded, and we stop the search in the branch immediately after discovering that all remaining scenarios in that branch are not penalized for exceeding the capacity.

In order to minimize the number of nodes (scenarios) that we need to consider, we sort the items according to size from largest to smallest. Even though the number of scenarios in  $Q(x)$  remains the same regardless of the sorting, we reduce the number of cases where a scenario is explored, but turns out not to exceed the capacity. For a proof that sorting is indeed optimal in this procedure, see Appendix B.

It may happen that the stopping criterion (row 8 in Algorithm 4) is never met, in other words, all scenarios lead to the penalty, and thus, we need to enumerate all of them. In such a case, the procedure EXCEEDED CAPACITY is called  $2^n$  times. Since other operations are trivial, the computational complexity of the o-o-s evaluation is  $\mathcal{O}(2^n)$  with respect to  $n$  binary random variables ( $n$  selected items).

### 3.3.2 Example

Let us demonstrate the procedure on a simple example. We assume we have selected four items with ordered sizes  $w = (7, 5, 2, 1)$  and that the capacity of the knapsack is  $W = 9$ . In Figure 3.1, we show the tree that is created based on the partial order  $\geq_x$ .

Each box starts with a scenario, for example  $(0, 1, 1, 1)$  (implying that items with sizes 5, 2 and 1 are selected), and their sum of weights in the second row. The scenarios that are explored are those with a gray background, the scenarios from  $Q(x)$  are framed by thick line. The order in which scenarios are explored is shown above the edges. In this example, we need to investigate eight scenarios in order to find the set  $Q(x)$ . Let us note, that if items were sorted in reverse order, the procedure would have to explore 13 scenarios to reveal the set  $Q(x)$ .

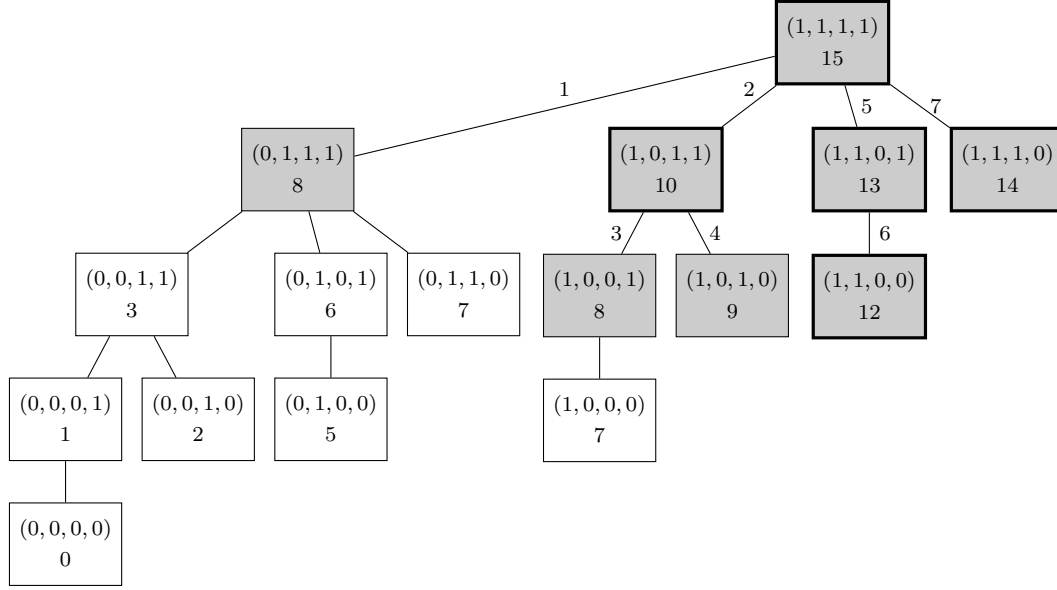


Figure 3.1: Demonstration of the o-o-s recursive procedure.

---

**Algorithm 4** Computing the out-of-sample value

---

1: **sort** the items such that  $w_i \geq w_j$  for all  $i < j$

2: **set** initially:

a:  $j = 1$

b:  $s = \bar{1}$

c:  $\pi_s = \prod_{i \in \mathcal{I}_x} p_i$

d:  $P_x = \pi_s e_s$

3:  $P_x = \text{EXCEEDED CAPACITY}(j, s, \pi_s, P_x)$

4:  $f(x) = \sum_{i \in \mathcal{I}_x} p_i c_i - d P_x$

5: **function** EXCEEDED CAPACITY( $j, s^I, \pi^I, P_x$ )

6:     **for**  $i = j$  to  $|\mathcal{I}_x|$  **do**

7:          $s = s_{i \rightarrow 0}^I$

8:         **if**  $e_s > 0$  **then**

▷  $e_s$  is the exceeded capacity at scenario  $s$

9:              $\pi_s = \frac{(1-p_i)}{p_i} \pi^I$

10:              $P_x = P_x + \pi_s e_s$

11:             **if**  $i < |\mathcal{I}_x|$  **then**

12:                  $P_x = \text{EXCEEDED CAPACITY}(i + 1, s, \pi_s, P_x)$

13:     **return**  $P_x$

---

### 3.3.3 Scenario contribution to the objective value

Another outcome of the procedure that might be useful is the knowledge of the contribution of each scenario to the objective function. This contribution is defined as

$$\gamma_s(x) = \sum_{i \in \mathcal{I}_x} s_i c_i p_i + \pi_s e_s \quad (3.10)$$

for all  $s \in Q(x)$ . For scenarios that do not belong to the set  $Q(x)$ , the contribution is simply  $\gamma_s(x) = \sum_{i \in \mathcal{I}_x} s_i c_i p_i$ . The overall o-o-s value is then  $f(x) = \sum_s \gamma_s$ .

This can be useful, for example, in a setting where we have solved a problem with a limited set of scenarios, and want to find out what non-included scenarios contribute to the o-o-s value, in order to possibly include them in our scenario set.

We can easily compute the contribution from the scenarios in  $Q(x)$  in Algorithm 4, for example between rows 9 and 10.

## 3.4 Exact reformulation of the stochastic knapsack problem

Let us define the set  $Q$  to be the set of all scenarios that lead to a penalty for some feasible decision  $x$ . This set is therefore given as  $Q = \cup_x Q(x)$ . To get this set, let us use the following:

**Definition 3.4.1** *We say that  $x^1 \geq x^2$  if  $x_i^1 \geq x_i^2$  for all  $i = 1, \dots, n$ .*

**Lemma 3.4.1** *If  $x^1 \geq x^2$  then  $Q(x^2) \subseteq Q(x^1)$ .*

Let us take the decision  $\bar{x}$ , where all items are selected, that is,  $\bar{x}_i = 1$  for all  $i = 1, 2, \dots, n$ . It holds that  $\bar{x} \geq x$  for all  $x \in \{0, 1\}^n$ . Therefore,  $Q(x) \subseteq Q(\bar{x})$  for all  $x$ , and hence  $Q = Q(\bar{x})$ . This gives us a way to generate all scenarios that may lead to a penalty in a minimal number of steps - simply by a recursive procedure built just as Algorithm 4, and initiated with the decision  $\bar{x}$ . See Algorithm 5 for the detailed description.

Scenarios not in  $Q$  never lead to a penalty. Therefore, we do not have to evaluate scenarios outside  $Q$ , and we do not need variables  $e_s$  for these scenarios since they are never positive. Thus we reduce the number of constraints and input data and solve the following reduced form of the problem:

$$\max_{x \in \{0, 1\}^n} \sum_i p_i c_i x_i - d \sum_{s \in Q} \pi_s e_s \quad (3.11)$$

$$\text{s.t.} \quad e_s \geq \sum_i s_i w_i x_i - W \quad \forall s \in Q \quad (3.12)$$

$$e_s \geq 0 \quad \forall s \in Q, \quad (3.13)$$

---

**Algorithm 5** Determination of  $Q$ 

---

```
1: sort the items such that  $w_i \geq w_j$  for all  $i < j$ 
2: set initially:
   a:  $j := 1$ 
   b:  $s := \bar{1}$ ;
   c:  $\pi_s := \prod_{i \in \mathcal{I}} p_i$ 
   d:  $Q = \{\bar{1}\}$ 

3:  $Q := \text{FINDSET}(j, s, \pi_s, Q)$ 

4: function  $\text{FINDSET}(j, s^I, \pi^I, Q)$ 
5:   for  $i := j$  to  $n$  do
6:      $s := s^I_{i \rightarrow 0}$ 
7:     if  $e_s > 0$  then  $\triangleright e_s$  is the exceeded capacity at scenario  $s$ 
8:        $\pi_s := \frac{(1-p_i)}{p_i} \pi^I$ 
9:        $Q := Q \cup s$   $\triangleright$  we also store the corresponding value  $\pi_s$ 
10:      if  $i < n$  then
11:         $Q := \text{FINDSET}(i + 1, s, \pi_s, Q)$ 
12:   return  $Q$ 
```

---

The significance of the reduction of constraints and input data depends on the size of  $Q$  relative to the size of  $S$ . We present a numerical experiment showing the dependency in Figure 3.2. We have looked at a large number of cases, all with 10 items. As we move right on the horizontal axis, the variation in size increases, but the average item size remains at 10. We show the results for knapsack capacities from  $W = 10$  (on the top) to  $W = 90$  (on the bottom).

The symmetry around the central ( $W = 50$ ) case is striking. First notice that the points for  $W = 50$  are perfectly lined up. For every scenario in  $S$ , there exists an “inverse scenario”, which has the exact opposite combination of 1’s and 0’s:  $(0, 0, 1, 0, 1)$  versus  $(1, 1, 0, 1, 0)$ . These two scenarios combined have a total size of 100, exactly twice that of the knapsack. Therefore, one of the scenarios must go into  $Q$  and the other one must stay outside. The only exception is when both scenarios have the same weight, namely the knapsack capacity. Then both stay outside. However, as we generated weights randomly from a continuous distribution when creating the figure, this has zero probability of happening (in the ideal case).

Similarly, by using these “inverse scenarios”, we can explain the symmetry between the  $|Q|/|S|$  ratios of 0.9 and 0.1, of 0.7 and 0.3, etc, observed in the figure. If we look at

the knapsack capacities 40 and 60, and the total sum of item weights is 100, we observe the following: if a scenario goes into  $Q$  for 40, its inverse goes into  $S \setminus Q$  for 60 and so on. Again this is distorted only by some special cases when weights of the selected items equals the knapsack capacity. But as we sampled from continuous distributions this never happens.

As we can see from Figure 3.2, the reduction in the number of scenarios can be substantial if we utilize  $Q$  rather than  $S$ , especially when the size of the knapsack is close to the total sum of weights of all items. Then, only a small number of scenarios leads to a penalty. This number is further reduced when the variance in sizes is small.

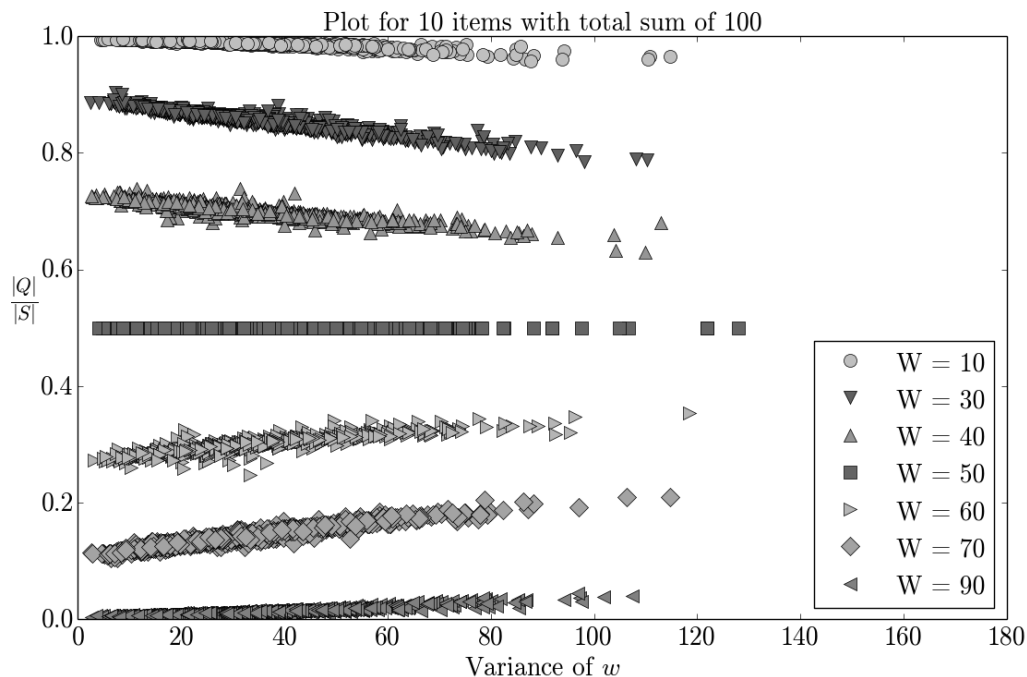


Figure 3.2: The number of scenarios in  $Q$ , relative to the number of scenarios in  $S$  for different knapsack capacities as the variation in item sizes increases.

### 3.5 Dependent case

For simplicity, we assumed that items appear independently of each other. This is, however, rare in real applications. In this paragraph we discuss the case of a general multivariate Bernoulli distribution where interactions between random variables play a role and the probability of a specific scenario is not given by a simple multiplication of probabilities associated with (non-)appearances of single items. We do not provide a comprehensive description of the multivariate Bernoulli distribution, but we discuss the properties that need to be taken into account when dealing with correlated variables.

The crucial observation is that the partial order from Definition 3.4.1 and Lemma 3.4.1 holds regardless of the underlying probability distribution. Since this lemma provides the stopping criterion for Algorithms 4 and 5, the frameworks of both algorithms remain unchanged. However, we need to adjust the computation of probabilities when creating new scenarios, denoted by  $s = s_{i \rightarrow 0}^I$ . One possible description of the multivariate case, with demonstrated modifications in the procedures, is shown in Appendix A.

Further derivations and properties of the multivariate Bernoulli distribution can be found in Dai et al. (2013) or Whittaker (2009).

### 3.6 Other applications

The structures just observed for the simple stochastic knapsack problem also appear in other settings. Consider a network design problem with stochastic existence (breakdown) of edges, and where the task is to satisfy demand at some nodes. If demand is not satisfied, we pay a penalty proportional to the unsatisfied demand. The first-stage decision is to invest in new edges. Assume for the moment that there are no flow costs in the second stage, just penalties for unsatisfied demand. The scenario used to initiate Algorithm 4 is “All edges down”, while the decision that we denoted as  $\bar{I}$  in Algorithm 5, that is, the decision that will always lead to a penalty, is “not to buy any additional edges”. Then if we have a scenario that leads to a penalty with this decision, every scenario where no further edges are present, automatically must lead to a penalty as well. Vice versa, if some scenario does not lead to a penalty, every other scenario that includes the same edges (and possibly some additional) are also penalty free. With this in mind, we end up with a partial order of the scenarios and a subset  $Q$  just as for the knapsack case. Calculating the penalty is somewhat more complicated, but can still be done efficiently. For a given feasible solution, the root node in the tree corresponding to Figure 3.1 is the case when no edges are working. Then as we proceed through the evaluation, edges are added, one by one. In the root node of the tree we shall have to solve a min cost network flow problem (for a given first-stage decision and no edges working). Then one additional edge is opened. The previous optimal flow is now feasible in this new looser problem, and a warm start can take place. So it will never be necessary to start from scratch when calculating the penalties; warm starts for min cost network flow problems will suffice.

If there are flow costs in addition to penalty costs in the second stage, a scenario set  $Q$  will still exist if the penalty is so high that ordinary flow is always preferred over penalties. If not, there is no clear-cut set  $Q$  (as there is no clear-cut tail), but the partial order remains, and so does the recursive way of solving the second-stage problem using warm starts.

Just as for the knapsack problem, the difference between  $S$  and  $Q$  (when it exists)

can be both negligible or substantial based on the input data. For stochastic network design,  $Q$  is small when the network provides high flexibility. That is, the situation where even if some, but not all, edges break, it is still possible to satisfy demand at nodes by using other edges. For instance, highly connected graphs, like city streets, provide such flexibility. On the other hand, poorly connected networks, where one single break might lead to a penalty, no matter what happens to the other edges, provide little flexibility. Therefore, the number of scenarios in  $Q$  is high (all combinations of 1's and 0's of remaining edges). So when dealing with such networks, it is not particularly beneficial to use the presented methodology – the ratio  $|Q|/|S|$  is high.

This way of recursively visiting nodes in a tree is of course common in many other parts of OR. Examples are the use of factoring methods in Ball et al. (1995, Section 3.3) or finding feasible solutions in the progressive hedging algorithm as outlined in Wallace and Helgason (1991). In the first case there is a stopping rule as in our case, in the latter, the recursiveness is used to structure the calculations.

### 3.6.1 Sampling

In case we wish to sample scenarios we can either first generate  $Q$  and then sample from  $Q$  or  $S \setminus Q$  in a controlled way. Or we can sample and then check where the sample belongs (which is equivalent to checking if a sample generates a penalty). This way we can achieve different densities of scenarios in the tail and outside the tail.

### 3.6.2 Integer random variables

Most of what we have said applies also to integer random variables if we have a model where there are tail effects; random integer demand that ought to be satisfied, and where there is a penalty for not doing so; network flow problems where edge capacities are random and integer. The scenarios will again be split into two sets, one that cannot lead to penalties and one that can. The algorithms have obvious extensions in this case.

## 3.7 Conclusion

The main point of this paper has been to show that many two-stage stochastic programs with special tail costs on the output distribution (which could correspond to tail risk measures) possess a specific structure that allows the scenario set to be split into two sets: those scenarios that may lead to a penalty (depending on the optimal solution) and those that may not, whatever happens. We demonstrate the situation on a stochastic knapsack problem, but indicate how many related problems, such as network design, are similar.

The structure and the split of the scenario set allows for an efficient way to calculate the out-of-sample value for the problem at hand, and can also be used in a setting of



sampling if we wish more scenarios in the tail of the output distribution, rather than evenly spread all over as normal sampling would do.

Our procedures are more powerful when the number of scenarios that cannot go to the tail is large. We show by numerical testing as well as verbal arguments when that will be the case, and hence, when our approach is particularly useful.

So a practical way of proceeding whenever our approach potentially applies would be

- Try to solve the problem by using full enumeration of scenarios. If this is possible, there is no need to use anything more advanced.
- If not, try to use our overall procedure, thus still solving the problem exactly. If this is numerically feasible, all is well.
- If not, use some heuristic to find feasible solutions, possibly using ideas from this paper in terms of sampling and scenario contributions. Check the quality of these solutions by full enumeration. If this is possible, there is no need to be more advanced.
- If not, use the out-of-sample evaluation procedure from this paper in combination with the heuristic.
- If not, use some out-of-sample estimation approach.

## References

- Ball, M. O., Colbourn, C. J., and Provan, J. S. (1995). Network reliability. In Ball, M. O., Magnanti, T. L., Monma, C. L., and Nemhauser, G. L., editors, *Network Models*, volume 7 of *Handbooks in Operation Research & Management Science*, chapter 11. North-Holland, Amsterdam.
- Dai, B., Ding, S., and Wahba, G. (2013). Multivariate Bernoulli distribution. *Bernoulli*, 19(4):1465–1483.
- Fortz, B., Labbé, M., Louveaux, F., and Poss, M. (2013). Stochastic binary problems with simple penalties for capacity constraints violations. *Mathematical Programming*, 138:199–221.
- King, A. J. and Wallace, S. W. (2012). *Modeling with Stochastic Programming*. Springer Series in Operations Research and Financial Engineering. Springer.
- Wallace, S. W. and Helgason, T. (1991). Structural properties of the progressive hedging algorithm. *Annals of Operations Research*, 31(1):445–456.

Wallace, S. W. and Ziemba, W. T., editors (2005). *Applications of Stochastic Programming*. MPS-SIAM Series on Optimization. SIAM, Philadelphia.

Whittaker, J. (2009). *Graphical Models in Applied Multivariate Statistics*. Wiley Publishing.

Zhao, Y. and Wallace, S. W. (2014). Integrated facility layout design and flow assignment problem under uncertainty. *INFORMS Journal on Computing*, 26(4):798–808.

## Appendix A Multivariate Bernoulli distribution

Let  $s = (s_1, s_2, \dots, s_n)$  be a scenario, that is a realization of an  $n$ -dimensional random vector from the Cartesian product space  $\{0, 1\}^n$ . There are several ways how to uniquely describe the desired distribution. The most common way is by providing the joint probability mass function. For the case of  $n = 2$  and a random vector  $Y = (Y_1, Y_2)$  with realizations  $y = (y_1, y_2)$  we have:

$$P(Y = y) = p(y_1, y_2) = p_{00}^{(1-y_1)(1-y_2)} p_{10}^{y_1(1-y_2)} p_{01}^{(1-y_1)y_2} p_{11}^{y_1 y_2},$$

where naturally  $p_{ij} = P(Y_1 = i, Y_2 = j)$ .

It is possible to write the logarithm of the density function.

$$\begin{aligned} \log p(y_1, y_2) &= \log p_{00} + y_1 \log \frac{p_{10}}{p_{00}} + y_2 \log \frac{p_{01}}{p_{00}} + y_1 y_2 \log \frac{p_{11} p_{00}}{p_{01} p_{10}} \\ &= u_\phi + y_1 u_1 + y_2 u_2 + y_1 y_2 u_{12}. \end{aligned}$$

The coefficients  $u_\phi, u_1, u_2, u_{12}$  are called *u-terms* and are widely used for studying interactions among random variables in the multivariate Bernoulli distribution. Therefore, they can be considered as an alternative way to define the distribution.

From the u-terms we can compute the logarithms of probabilities:

$$\begin{aligned} \log p_{00} &= u_\phi \\ \log p_{10} &= u_\phi + u_1 \\ \log p_{01} &= u_\phi + u_2 \\ \log p_{11} &= u_\phi + u_1 + u_2 + u_{12} \end{aligned}$$

In the case of  $n = 3$ , the logarithm of the density function by using u-terms notation is

$$\log p(y_1, y_2, y_3) = u_\phi + y_1 u_1 + y_2 u_2 + y_3 u_3 + y_1 y_2 u_{12} + y_1 y_3 u_{13} + y_2 y_3 u_{23} + y_1 y_2 y_3 u_{123}$$

so we can derive expressions for selected u-terms (similarly for  $u_2, u_3, u_{13}$  and  $u_{23}$ ):

$$\begin{aligned} u_\phi &= \log p_{000} \\ u_1 &= \log \frac{p_{100}}{p_{000}} \\ u_{12} &= \log \frac{p_{000}p_{110}}{p_{100}p_{010}} \\ u_{123} &= \log \frac{p_{100}p_{010}p_{001}p_{111}}{p_{000}p_{011}p_{101}p_{110}} \end{aligned}$$

We omit the general notation of the probability density function for a case with  $n$  random variables. We hope that the idea of how the density would look is evident from the given examples. It can be also observed that for the general distribution, there are  $2^n$  parameters (either probabilities or u-terms) with the additional condition that the sum of the probabilities must equal one. So we need  $2^n - 1$  parameters to fully describe the distribution.

When the probabilities are known, there is no need to recompute the probabilities of scenarios when we deal with the transition  $s = s_{i \rightarrow 0}^I$ . A different situation is in the case of describing the distribution by u-terms. Let us assume we know all u-terms, each u-term is associated with one element of the power set of  $\mathcal{I}$ , denoted  $\mathcal{P}(\mathcal{I})$ . Further we denote  $\mathcal{F}_s \subseteq \mathcal{P}(\mathcal{I})$ , which includes all subsets of appearing items.  $\mathcal{F}_s$  plays a role as a collection of indices of items that appear in the scenario  $s$  and are used for computing the scenario probability as is shown in Algorithm 6 in the function COMPUTEPROBABILITY. As initial values for the scenario  $s = \bar{1}$  in algorithms 4 and 5 we set  $\mathcal{F}_s = \mathcal{P}(\mathcal{I})$  and compute the corresponding probability as it is described in the function INITIALVALUES.

---

**Algorithm 6** Modification of computing probabilities and initial values

---

```

1: function COMPUTEPROBABILITY( $i, s_{i \rightarrow 0}^I, \mathcal{F}_{s^I},$ )
2:    $s := s_{i \rightarrow 0}^I$ 
3:    $\mathcal{F}_s = \{F \in \mathcal{F}_{s^I} \mid i \notin F\}$ 
4:    $\pi_s = \exp\left(\sum_{F \in \mathcal{F}_s} u_F\right)$ 
5:   return  $s, \pi_s$ 

6: function INITIALVALUES( $\mathcal{I},$  u-terms)
7:    $s = \bar{1}$ 
8:    $\mathcal{F}_s = \mathcal{P}(\mathcal{I})$ 
9:    $\pi_s = \exp\left(\sum_{F \in \mathcal{F}_s} u_F\right)$ 
10:  return  $s, \pi_s$ 

```

---

## Appendix B

We prove that sorting of items in Algorithm 4 is optimal, i.e. requires the minimal number of recursive calls of the procedure.

Let us assume two orderings of the same set of items, represented by vectors of weights  $w^A$  and  $w^B$  which differ only at neighboring positions  $k$  and  $k + 1$ , which are swapped. That is,  $w_k^A = w_{k+1}^B$  and  $w_{k+1}^A = w_k^B$ . Let us assume  $w_k^A \geq w_{k+1}^A$ . Then, the recursive procedure EXCEEDED CAPACITY, described in Algorithm 4, generates for both vectors  $w^A$  and  $w^B$  fundamentally equal scenarios (the same items are present) if  $i < k$  or  $i > k$ . In the case of  $i = k$  ( $s_k = 0$ ), we have  $\sum_{m=k+1}^n w_m^A s_m \leq \sum_{m=k+1}^n w_m^B s_m$  for every scenario  $s$ .

Therefore,  $s \in Q$  for  $w^A$  implies  $s \in Q$  for  $w^B$  at every branch. However, it may happen that  $s \in Q$  for  $w^B$ , but  $s \notin Q$  for  $w^A$ . In such a case, the recursive procedure is called again for  $w^B$  in the particular branch, but not for  $w^A$ . Thus the total number of times the procedure is called for  $w^B$  is higher or equal to the number of times it is called for  $w^A$ .

Let us take some arbitrary ordering and apply, for instance, the bubble sort method where two neighboring items are swapped whenever  $w_i < w_{i+1}$ . Such a method results in a descending order of items, and since for every swap we can only improve (decrease) the number of times the procedure EXCEEDED CAPACITY is called, working with an ordered vector of items in Algorithm 4 is optimal in this respect.

# Scenario tree construction driven by heuristic solutions of the optimization problem

Vit Prochazka<sup>□</sup>    Stein W. Wallace<sup>□</sup>

<sup>□</sup>Department of Business and Management Science  
Norwegian School of Economics, Bergen, Norway

## Abstract

Many methods for generating scenarios for stochastic programs aim to ensure a good fit (in a sense of some statistical measure) between the scenario tree and the underlying probability distribution. We offer an alternative approach where the scenario generation process is driven purely by the out-of-sample performance of a pool of solutions, obtained by some heuristic procedure. In order to do that, we formulate a loss function that measures the discrepancy between out-of-sample and in-sample (in-tree) performance of the solutions. By minimizing such a (usually non-linear, non-convex) loss function for a given number of scenarios, we receive an approximation of the underlying probability distribution with respect to the optimization problem. This approach is especially convenient in cases where the optimization problem is solvable only for a very limited number of scenarios, but an out-of-sample evaluation of the solution is reasonably fast. Another possible usage that we demonstrate in this paper is the case of binary distributions, where classical scenario generation methods based on fitting the scenario tree and the underlying distribution do not work.

**Keywords:** stochastic programming, scenario generation, binary distributions, heuristic

## 4.1 Introduction

Most methods for solving stochastic programs require discrete scenarios as input. Exceptions would be simple (often inventory) models that have closed-form solutions and methods such as stochastic decomposition (Higle and Sen, 1991), where the discretization takes place inside the method. The simplest way to find scenarios that can be used as input is normally to sample (Cario and Nelson, 1997; Lurie and Goldberg, 1998). If sampling leads to numerically solvable problems with high enough accuracy in short enough time (see for example the discussion in Kaut et al. (2007)), there is no good reason to do anything more complicated. But, if for some reason, sampling is not acceptable, there is a need for something more advanced – and usually more complicated. Very often, the alternatives require some rather expensive offline computations, but lead to a much more efficient online performance. For an overview, see for example King et al. (2012). The methods fall into two major classes; those that try to approximate the probability distribution itself, and those that focus on the quality of the solutions that emerge from using the scenarios. We can call these methods *distribution-oriented* and *problem-oriented*. Both use a metric to measure distance, the first uses measures from probability theory (such as the Kantorovich-Rubinstein or Wasserstein metric (Pflug, 2001)), the second uses the optimization problem itself as metric. This paper falls into the second category, and is hence connected to the principal thinking in Høyland and Wallace (2001), and the methodology set out by Fairbrother et al. (2017). But contrary to the latter work, we do not need to analyze the optimization problem itself, rather we need to be able to produce a set of feasible solutions and to perform, rather efficiently, out-of-sample objective function evaluation.

Hence, in this paper, we introduce a framework that enables generating scenarios in a problem-oriented fashion, but without analyzing the problem explicitly. Our general approach is based on a pool of solutions and a measure of discrepancy between in-sample (in-tree) and out-of-sample performance of the solutions, which we aim to minimize. We assume that the pool of solutions is generated by some heuristic with a reasonable trade-off between speed and accuracy. Every solution from the pool can be evaluated out-of-sample, that is, we can determine its true value by using the true distribution (or a very large sample), and in-sample (in-tree) by using the corresponding scenario tree. We define a loss function that measures the discrepancy between out-of-sample and in-sample performance of the pool of solutions. We search for a tree that minimizes the loss function.

Since we offer a general framework that requires several problem-specific subroutines, a direct comparison with other methods for generating scenarios is not easy. Such comparisons can always be distorted in (or out of) favor, by applying, for example, a different heuristic. Our framework also may require a significant time on development. Both

the subroutines - the heuristic for obtaining solutions and the loss function minimization procedure - must be tailored-made for a specific problem. That is the main disadvantage compared to some other methods for scenario generation, for instance copula-based heuristics (Kaut, 2014), which can be applied immediately by using a publicly released code, and which requires either historical data or specifications of the probability distribution.

Thus, rather than directly competing with these methods on solvable problems, we try to identify their limits, for example how they handle binary random variables, such as a random appearance of customers in vehicle routing problems. We offer an approach that can overcome some of the limits, and which can be applied on a larger spectrum of applications. We also hope to offer an original perspective on the relationship between uncertainty and its representation by scenarios within optimization problems.

## 4.2 Framework

Since this paper is primarily conceptual, and not technical, we will not introduce exact mathematical definitions of all elements of our procedure in order to keep the work easy to follow. In general, we assume there is a true<sup>1</sup> random vector  $\xi$  that enters a process to be optimized. The distribution is either parametrically described or empirically given, for example by historical data. For simplicity, we focus on two-stage problems to avoid the complications that multi-stage problem formulations and conditional distributions between stages bring. See Birge and Louveaux (1997); Kall and Wallace (1994) for proper definitions, if needed.

Without loss of generality we assume a maximization<sup>2</sup> problem throughout the text

$$\max_{x \in \mathcal{X}} f(x, \xi) \tag{4.1}$$

where  $\mathcal{X}$  is the feasible region for decisions  $x$ . The problem that often arises is that (4.1) cannot be solved when using  $\xi$  directly due to its size (in the case of an empirical distribution) or its computational intractability (in the case of a parametrically defined distribution). Thus, we search for a representation of the original distribution by a so-called scenario tree  $\mathcal{T}$  consisting of particular scenarios, which are vectors of realizations of random variables (for example volumes of the items in the stochastic knapsack problem), and probabilities  $p$  associated with each scenario. Then, we solve an optimization problem

---

<sup>1</sup>In a large majority of applications, this still means highly subjective descriptions of the random phenomena.

<sup>2</sup>It is more standard to work with minimization problems in the literature. However, we decided to go with maximization, since our approach is demonstrated on a knapsack problem, for which maximization is natural. In the end, each maximization problem can be converted to a minimization one and vice versa.

$$\max_{x \in \mathcal{X}} f(x, \mathcal{T}) \tag{4.2}$$

and hope that the solution of this program is also a good solution to the original problem (4.1). The quality of the solution and its relation to the original program can be tested, see Kaut and Wallace (2007).

Although it is almost always impossible to solve problem (4.1), it is often possible to evaluate the quality of a fixed solution  $\hat{x}$  by using the whole distribution. That means to determine the value  $f(\hat{x}, \xi)$ . If it is not possible to do this exactly, then, most of the time, it can be done approximately, with very high accuracy, by using a very large sample from the true distribution. We call this procedure *out-of-sample* evaluation of the solution. Similarly, we can get an *in-sample* value  $f(\hat{x}, \mathcal{T})$  for the solution  $\hat{x}$  by inserting the scenario tree into the model.

Our framework to construct a scenario tree consists of two steps:

1. Heuristically generate a pool of solutions for the optimization problem; evaluate the solutions out-of-sample.
2. Construct a scenario tree in such a way that the discrepancy between in-sample and out-of-sample performance of the solutions is minimized.

The first step of our approach is to generate a pool of solutions  $\mathcal{A}$  by a problem-specific heuristic and evaluate the solutions out-of-sample. This is not always possible, since for some problems even finding a feasible solution or its out-of-sample evaluation can be too difficult<sup>3</sup>. But our approach applies to a large class of problems, for which reasonable heuristics exist. To make the right choice of heuristic, one needs to take into consideration a trade-off between the number of solutions in the pool, accuracy of the heuristic and the time spent in this phase.

The heuristically obtained solutions are evaluated out-of-sample, either as a part of the heuristic procedure itself or afterwards, for example in the case the heuristic is using just a small sample of the original distribution. This step, again, assumes that the out-of-sample evaluation can be performed within a reasonable time.

To measure discrepancy between in-sample and out-of-sample performance of the solutions from the pool, we define a *loss function*. The function is derived from our requirements on a good scenario tree, which are discussed in the following section. For a given pool of solutions, the loss function is a function of the scenario tree. Thus we search

---

<sup>3</sup>For example the problem of minimum Hamiltonian cycle, where even finding a Hamiltonian cycle in a given graph is NP-complete (Garey and Johnson, 1979).



for such parameters of the scenario tree (scenarios and their probabilities) that the loss function is minimized. In Section 4.2.3, we discuss different settings of the minimization procedure, but in the case both scenarios and their probabilities are used, we deal with non-linear and non-convex functions to be minimized. This is not a trivial task, and even though there are some solvers for non-linear optimization available, they often require some problem-specific adjustments.

### 4.2.1 Properties of a good scenario tree

In this section, we discuss our views on *what makes a scenario tree good*. We list a set of requirements on the scenario tree and its relation to the original optimization problem (4.1). These requirements are used for the formulation of the loss function. Let us first focus on relatively complete recourse; handling potential infeasibilities in the second-stage problem is discussed in Section 4.2.5.

In theory, to call a scenario tree  $\mathcal{T}$  (almost) perfect, we would simply need that (4.2) returns the same optimal solution for  $\mathcal{T}$  as it would for the true distribution  $\xi$  — had it been obtainable. There are, however, two problems. First, there are cases for which even the optimization program (4.2) is computationally intractable and we need to settle with some non-optimal solution, which can be reasonably good for  $\mathcal{T}$ , but arbitrarily bad for  $\xi$ . Hence “(almost)”. Second, and more importantly, to assure that this requirement will hold is not achievable. It would require the knowledge of the optimal solution of the program for  $\xi$ . If we were able to solve the problem for  $\xi$  and get the optimal solution, generation of the scenario tree is, obviously, of more marginal interest (though it depends on the requirements on CPU time).

Let us, then, discuss what we expect from a good tree, not a perfect one. Here we summarize our requirements on a good scenario tree  $\mathcal{T}$  in relation to the true distribution  $\xi$  when we perform out-of-sample and in-sample evaluation of a pool of solutions.

1. The *ranking is “more or less” preserved*<sup>4</sup>. That is, if one solution  $x_1$  is better than  $x_2$  when evaluated out-of-sample, it is going to be “very likely” better when evaluated in-sample.
2. We *do not observe overconfident outliers*. We argued in Requirement 1 that it is impossible to have a guarantee of the perfect ranking, so we expect it may happen that  $f(x_1, \xi) > f(x_2, \xi)$ , but  $f(x_1, \mathcal{T}) < f(x_2, \mathcal{T})$  for some  $x_1$  and  $x_2$ . Such a case is acceptable when the values  $f(x_1, \xi)$  and  $f(x_2, \xi)$  are close to each other. But we

---

<sup>4</sup>We would naturally prefer the perfect ranking. Then, for any subset of solutions, solving  $\max_x f(x, \mathcal{T})$  and  $\max_x f(x, \xi)$  would be equivalent tasks with respect to our objective. But having a requirement on reaching the perfect ranking of solutions is meaningless, since it implies the ability to solve the problem  $\max_x f(x, \xi)$ . Usage of the scenario tree is, then, redundant.

want to avoid the case where a particularly bad solution (out-of-sample) performs well in-sample, that is, its in-sample value is a (massive) overestimation of the true value. The opposite case – a truly good (out-of-sample) solution performs really badly in-sample – is not so critical, if it does not hold for many solutions. We call these outliers acceptable.

3. There is a greater emphasis on Requirements 1 and 2 to be satisfied for better solutions than for worse solutions. In other words, the scenario tree approximates better the underlying distribution in the space of higher-quality solutions, where an optimization algorithm, either an exact or a heuristics one, will search for the best solution to the program (4.2).
4. *In-sample values approximate well out-of-sample values*, that is  $f(x_a, \xi) \approx f(x_a, \mathcal{T})$ . In theory, this requirement is not needed at all. A scenario tree that can produce in-sample values of solutions that are totally off, but ranks the solutions “more or less” correctly, is still a very useful tree, because it enables us to find a very good solution. Then, the real value of the solution can be found by an out-of-sample evaluation.

However, we still have this requirement on our list, not just because it is a nice (but not necessary) property of the scenario tree, but because it implies, to some extent, other requirements. If the in-sample values approximated out-of-sample values perfectly, it would also preserve the perfect ranking. Thus, we use this requirement as a starting point for the loss function formulation in 4.2.2.

### Example

We demonstrate the relationship between in-sample and out-of-sample values over the pool of solutions in Figure 4.1. We show four examples of different scenario trees, on which we comment some of our views formulated above. We assume the pool consists of six solutions, sorted in ascending order according to their out-of-sample values (y-axis). We assume a maximization problem, thus the higher the out-of-sample value, the better the solution is.

Let us assume we have a scenario tree that returns out-of-sample and in-sample values as in 4.1a). If these six solutions were the only feasible solutions of the optimization problem, solving the program (4.2) would return the true optimal solution, which is fine. However, we see the tree is not reliable for other solutions and their ranking, for example solution no 1 (worst out-of-sample) is ranked as the second best one in-sample. Sometimes, problem (4.2) can be solved only heuristically, so even if we had guaranteed the basic property that the optimal solution in tree is optimal out-of-sample, the heuristic could miss it and return an arbitrarily bad solution (out-of-sample).

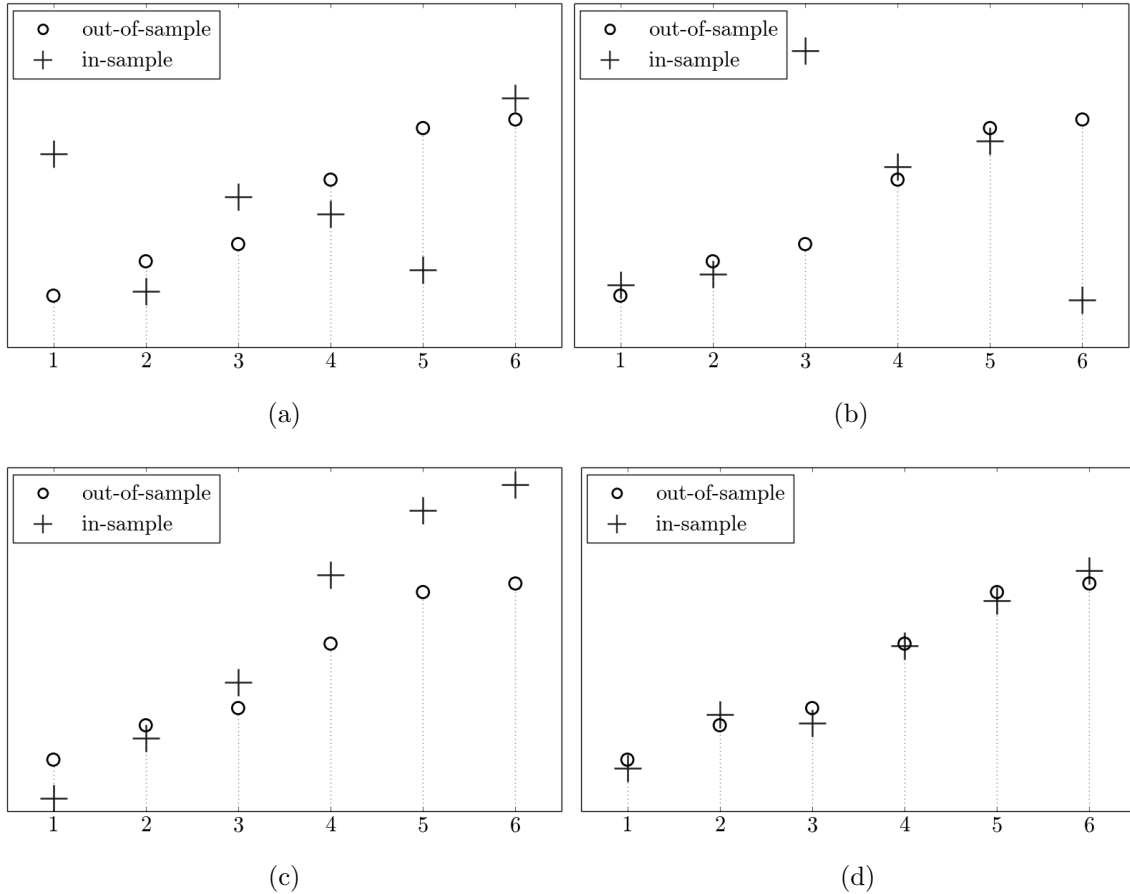


Figure 4.1: Demonstration of different scenario trees and their properties with respect to in-sample and out-of-sample performance.

In 4.1b), we demonstrate the concept of overconfident (solution no 3) and acceptable (solution no 6) outliers. If we solve  $\max_x f(x, \mathcal{T})$  over this set of solutions, we determine solution no 3 as the optimal one, but it performs quite poorly in reality (out-of-sample). Had it not been for solution no 3, we would end up with solution no 5 as the optimal one. That means we would miss the true optimal solution no 6 and some related value, but the error is not as significant as for overconfident outliers, whose in-sample value is overestimated.

In 4.1c), we show a good ranking of solutions, but with the wrong approximated values. Such a tree would be good for the optimization process, which would correctly find the optimal solution. Its out-of-sample value can be determined afterwards.

In 4.1d), the in-sample values approximate the out-of-sample value reasonably well. That implies that the ranking is more or less correct. We see that solutions nos 3 and 2 are ranked incorrectly, but since their out-of-sample values are similar, that would not cause a big error even if we solved the maximization problem over the solutions  $\{1, 2, 3\}$ .

## Note

In this paper, we introduce a whole framework for scenario generation. But even if some other method (mentioned in the introduction) is used, visualization of the tree performance as in Figure 4.1 can offer a fast and intuitive way to assess how good a scenario tree is. In other words, we use just the first step of the proposed framework: we apply some heuristic to generate a pool of solutions for the problem (4.1) and visualize their out-of-sample and in-sample performance for a given tree.

### 4.2.2 Loss function

We introduce a loss function to measure the discrepancy between the out-of-sample and in-sample performance of a pool of solutions. The formulation of the loss function is derived from our requirements on a good scenario tree. The smaller the value of the function, the better the tree (in our view) for the subsequent optimization procedure. We define the loss function in the following way:

$$L(\mathcal{T}, \mathcal{A}) = \sum_{x_a \in \mathcal{A}} \left( z_1^a (1 + z_2 \mathbb{1}_{[f(x_a, \mathcal{T}) > f(x_a, \xi)]}) (f(x_a, \mathcal{T}) - f(x_a, \xi))^2 \right) \quad (4.3)$$

where  $\mathbb{1}_{[condition]}$  takes the value 1 if the *condition* is met, 0 otherwise. The loss function is fundamentally the weighted sum of squares between in-sample and out-of-sample values that captures a basic fit between them (Requirement 4). A good fit between the in-sample and out-of-sample values implies that the ranking is more or less correct (Requirement 1).

Each square is further weighted with the term  $(1 + z_2 \mathbb{1}_{[f(x_a, \mathcal{T}) > f(x_a, \xi)]})$ , which penalizes approximations from above. Together with a potential high difference between in-sample and out-of-sample value, it penalizes the overconfident outliers (Requirement 2). If we dealt with a minimization problem, we would penalize the approximation from below as it leads to overconfident outliers.

Each term is also weighted with  $z_1^a$  which takes a higher value, the better (out-of-sample) a solution is. Thus, contrary to the previous weighting term, it does not depend on the scenario tree. The weights  $z_1^a$  put more emphasis on higher-quality solutions, for which it is more crucial that the scenario tree approximates better the underlying distribution.

Loss function (4.3) is formulated generally by using weights  $z_1^a$  and  $z_2$ . The weights are considered as hyper-parameters of the overall procedure, and they allow us to adjust the loss function based on problem specifications such as the heuristic for solution generation and the loss function minimization procedure (see the following section). For instance, if the heuristic procedure at times produces some bad solutions,  $z_1^a$  should be more pro-

gressive (adding more weights) towards better solutions compared to the case where all solutions from the heuristic phase are relatively good.

### 4.2.3 Minimization of the loss function

Our aim is to construct the scenario tree. Since we defined our measurement of discrepancy between the tree and the underlying distribution, we will naturally look for a tree, for which the discrepancy is as small as possible for a given pool of solutions.

In general, we search for a set of scenarios (realizations of a random vector) and associated probabilities that enter the optimization program (4.2) as input data. But for this task - minimization of the loss function - the roles are swapped. Scenarios and probabilities are free variables to be set, whereas decision variables of the optimization problem (heuristic solutions) are input data.

Specifications of the minimization procedure depend on requirements of the scenario tree and the optimization problem we solve. For instance, it is possible that an application requires equiprobable scenarios. In such a case, probabilities are no longer free variables but parameters in the in-sample evaluation function. This phenomena appears and is discussed also in Høyland and Wallace (2001). When using both probabilities and scenarios as free variables, we deal (most likely) with a non-convex problem. In Section 4.2.4, we show a requirement on integrality of scenarios, which makes the problem constrained.

But in general, we do not want to impose constraints on scenarios and probabilities unless it is forced by the problem itself. The main reason is that it could only worsen the value of the loss function, that is, cause larger discrepancies between the trees and the underlying distributions according to our definition. For the same reason, we do not require the sum of probabilities to be 1, thus, our tree does not have to form a properly defined probability distribution. This is demonstrated and discussed in the example below.

Deriving a list of all possible properties of the loss function minimization procedure (convexity, differentiability, types of constraints, etc.) for different classes of optimization problems goes beyond the scope of this paper. But we point out several phenomena that a reader might face. We kindly encourage the reader to analyze his own problem in order to choose a suitable method.

In most applications, there will be a given number of scenarios based on how many of them can be computationally handled by the subsequent optimization procedure, for instance within a certain time limit. Then, the presented approach returns scenarios with the most similar performance to the original distribution with respect to the optimization task. However, it is possible to use the same framework to find scenarios that ensure some predefined loss function value (in our definition). The problem is to set such a threshold

given several hyperparameters (weights  $z_1^a$  and  $z_2$ ) and solve a constrained optimization (non-linear, non-convex) program.

### Stochastic knapsack problem

We demonstrate the usage of our framework on the classical stochastic knapsack problem where the items have uncertain volumes. The aim is to find  $K$  scenarios to represent a true distribution, which is given by discrete observations (historical data). That means the uncertainty is in both cases represented by the set of scenarios  $\mathcal{S}$ , each scenario has its probability  $p_s$ . In the case of the true distribution  $p_s = \frac{1}{|\mathcal{S}|}$ .

Let  $\mathcal{I}$  be the set of items, item  $i$  having a value  $c_i$  and size  $w_{si}$  in scenario  $s$ , with the knapsack having a capacity  $W$ . All items we want to choose must be picked in the first stage and if their total size exceeds the capacity of the knapsack, we pay a unit penalty of  $d$  in the second stage. The objective is to maximize the expected profit. We formulate the optimization problem for decision variables  $x \in \{0, 1\}^n$ , where  $x_i = 1$  if item  $i$  is picked, 0 otherwise;

$$\max_{x \in \{0,1\}^n} \sum_{i \in \mathcal{I}} c_i x_i - d \sum_{s \in \mathcal{S}} p_s e_s \quad (4.4)$$

$$\text{s.t.} \quad e_s \geq \sum_{i \in \mathcal{I}} w_{si} x_i - W \quad \forall s \in \mathcal{S} \quad (4.5)$$

$$e_s \geq 0 \quad \forall s \in \mathcal{S} \quad (4.6)$$

where  $e_s$  denotes exceeded capacity of the knapsack in scenario  $s$ . The in-sample and out-of-sample evaluation (depending on scenarios we use) of a given solution  $\hat{x}$  is straightforward and fast:

$$f(\hat{x}, p, w) = \sum_{i \in \mathcal{I}} c_i \hat{x}_i - d \sum_{s \in \mathcal{S}} p_s \left( \sum_{i \in \mathcal{I}} w_{si} \hat{x}_i - W \right)^+ \quad (4.7)$$

where  $X^+$  takes value  $X$  if  $X \geq 0$ , 0 otherwise.

The optimization model (4.4) – (4.6) and the evaluation function (4.7) hold for the scenario tree as well as for the true distribution (historical data also form a scenario tree). To distinguish between the two in the following text, we denote the true distribution  $\xi = \{\hat{p}, \hat{w}\}$  and the desired scenario tree, consisting of  $K$  scenarios, as  $\mathcal{T} = \{p, w\}$

Our main focus is on the minimization of the loss function, so we choose a very simple heuristic to generate a pool of solutions  $\mathcal{A}$ . In every solution from the pool, an item  $i$  is randomly picked with probability  $q$ .

Since the pool of solutions is not changing during the subsequent minimization of the loss function, the out-of-sample evaluation is performed only once. Then the list of out-of-

sample values, denoted  $V$ , enters the loss function minimization procedure as input data. We deal with a unconstrained problem

$$\min_{p,w} L(p, w) \quad (4.8)$$

where the loss function is

$$L(p, w) = \sum_{x_a \in \mathcal{A}} \left( z_1^a (1 + z_2 \mathbf{1}_{[f(x_a, p, w) > V_a]}) (f(x_a, p, w) - V_a)^2 \right) \quad (4.9)$$

The loss function is non-linear and non-convex in decision variables  $p$  and  $w$ . That makes the problem difficult to solve due to the existence of many local minima. Thus, we propose a heuristic that explores the search space in an efficient way to find a high-quality solution.

The core of the heuristic is the sub-gradient method that works in an iterative manner. It is designed for solving non-linear convex problems. As the name suggests, it is based on sub-gradients of the loss function with respect to the decision variables, denoted  $g_p$  and  $g_w$ . Sub-gradients express how much the loss function will change, if the inputs ( $w$  and  $p$ ) are changed by a very small step. Sub-gradients can be, therefore, computed numerically by the definition of derivative. That is, we add a small number  $h$  to each element of the scenario tree and probability vector and recompute the value of the loss function for each element. After dividing the change by  $h$ , we get the numerical sub-gradients.

This procedure is, obviously, computationally expensive, as it requires multiple evaluation of the loss function at every iteration of the overall method. Fortunately, it is possible to derive the sub-gradients analytically. It is advantageous to apply the chain rule. It means to derive  $\frac{d\mathcal{L}}{df}$  – the gradient of the loss function with respect to the in-sample evaluation function, further  $\frac{df}{de}$  – the gradient of the in-sample evaluation function with respect to the exceeded capacity, and so on. By doing so, we get a chain of simple operations (square, multiplication, addition,  $\max()$ <sup>5</sup>, etc.), for which we have standard differentiation rules. By simple applications of the chain rule we get the desired sub-gradients  $g_p$  and  $g_w$ .

Let us note that in other optimization problems it might not be so straightforward to derive sub-gradients as in our case. More complicated functions might come into play. Thus it could require more effort in analytical derivation or the use of numerical sub-gradients, which should always be available, at least in in principle. Or it is possible to choose a different method from non-linear optimization theory, that is not based on

---

<sup>5</sup>The function  $\max(\cdot, \cdot)$  is non-differentiable at the point where the two arguments are equal. Hence we use sub-gradient instead of gradient to be precise. It has no practical impact from the computational point of view.

gradients (see some textbook on non-linear optimization, for instance Hendrix and Tóth (2010); Boyd and Vandenberghe (2004)).

With the sub-gradient method, we take a small step in the direction of the negative sub-gradient, that is, in the direction of the steepest descent, at every iteration. Such an approach converges to a local minimum, which is also a global minimum in the case of convex minimization. However, not in our case, so we add two features to enhance exploration of the search space in order to avoid termination of our procedure at some low-quality local minimum.

The first feature is to use multiple starts of the procedure from different initial points. The second feature is the recognition and replacement of “useless” scenarios. We recognize these scenarios by evaluating their impact on the loss function. The impact is defined as the change of the loss function values if we remove a particular scenario from the scenario tree. If the change is very small, it means that the particular scenario is not very useful, and we replace the scenario by a new one (chosen randomly). The heuristic is summarized in Algorithm 7.

---

**Algorithm 7** Minimization of the loss function

---

```

1: for  $m = 1$  to  $M$  do
2:   initialize  $p$  and  $w$ 
3:   for  $j = 1$  to  $J$  do
4:     compute sub-gradients  $g_p$  and  $g_w$ 
5:      $p = p - \alpha_j^p g_p$ 
6:      $w = w - \alpha_j^w g_w$ 
7:     compute  $L_{jm}(p, w)$  according to (4.9)
8:     for  $k = 1$  to  $K$  do
9:       if  $\|g_{w_k}\|_\infty < \epsilon^w$  then
10:          $\tilde{p} = \{p_l : l \neq k\}$ 
11:          $\tilde{w} = \{w_l : l \neq k\}$ 
12:         if  $|L(\tilde{p}, \tilde{w}) - L(p, w)| < \epsilon^L$  then
13:           replace  $w_l$  with a new scenario
14: choose  $p$  and  $w$  that correspond to the smallest  $L_{jm}$  value.

```

---

The parameters  $p$  and  $w$  are updated (rows 5 and 6 in Algorithm 7) by small steps in the direction of the steepest descent. The step sizes, denoted  $\alpha^p$  and  $\alpha^w$ , are decreased with the increasing number of iterations. They are hyper-parameters that enter the procedure and need to be carefully set (too small steps lead to slow convergence, too large to oscillation



or divergence). We set these steps based on some trial tests.

The routine of scenario usefulness assessment is computationally expensive. It would require computation of the loss function  $K$  times at every iteration. To avoid it, we run a pre-test, where we check the  $\infty$ -norm of the sub-gradient related to each scenario (row 9), which allows us to break the test once we find one of its element greater than  $\epsilon^w$ . Only if the sub-gradient is small, do we proceed to the evaluation of the impact on the loss function.

Initialization and replacement of scenarios is performed by a random draw from the original data  $\hat{w}$ , probabilities are randomly set. Algorithm 7 is just a pseudo-code, not the most efficient implementation. Obviously, storing all  $L_{jm}$  is not necessary, since at any time, we can store just two best values - a local one for the  $j$  cycle and a global one for the  $m$  cycle. We can also compute the value of the loss function while evaluating the sub-gradients of the function.

Hence, we do not claim that this is the most effective heuristic for the problem. Most likely it is not. It is based on a simple sub-gradient method. If the main focus was on developing the most efficient algorithm for this task, it would be possible to build it on more sophisticated algorithms for non-linear optimization, such as the adaptive gradient method, possibly with momentum, or methods based on the second sub-derivative.

We believe this can still serve as an inspiration for developing more efficient algorithms, if needed. We pointed out some issues and suggestions how to overcome them. But for some applications, the presented algorithm is “good enough”, as it was in our case. It is important to realize that even if we could guarantee the global optimum of the loss function, the whole framework would still be a heuristic in the sense that there are no guarantees relative to other feasible solutions that are not included in the pool.

### Computational test

We create two pools of solutions. One pool, called a *training pool*, is used to run Algorithm 7 to obtain a scenario tree  $\mathcal{T} = \{p, w\}$ . The second pool, called a *testing pool*, is used for evaluation of the obtained tree. The testing pool serves as a proxy for the entire search space. In real usage, the heuristic from the first phase generates substantially weaker solutions (unless it gets very lucky) than what is assumed to be achievable with the subsequent optimization (otherwise there is no need of searching for a tree and solving the optimization problem). To mimic this property, we exclude the best 10% of the solutions from the training pool.

In our computational experiment, we generate 1000 scenarios for 19 items to represent the true distribution  $\hat{w}$ . Our aim is to find a scenario tree consisting of only 3 scenarios. We use 200 solutions in the training pool and 400 in the testing pool.

To demonstrate the advantage of problem consideration in the scenario generation

process, we apply our method on two cases. In the first case (case 1), we choose a capacity of the knapsack such that approximately half of all scenarios satisfy

$$\sum_i \hat{w}_{si} > b. \quad (4.10)$$

In other words, only half of the scenarios may lead to a penalty. The rest of the scenarios do not have to be taken into account, since they do not lead to a penalty even for the decision  $x = (1, 1, \dots, 1)$  (all items are picked).

In the second case (case 2), we transform the original scenarios  $\hat{w}_{si}^B = q - \hat{w}_{si}$ , where  $q > \max_{si} \hat{w}_{si}$ . That is, small items become large and large items become small. We also set a new capacity of the knapsack  $b^B$  such that those scenarios, for which the condition (4.10) is not met, satisfy  $\sum_i \hat{w}_{si}^B > b^B$ . In other words, all the scenarios that may lead to a penalty in case 1 can be ignored in case 2. And vice versa, scenarios that may lead to a penalty in case 2, can be ignored in case 1.

We present numerical results for these two cases in Figure 4.2. In 4.2a, we show the sum  $\sum_i \hat{w}_{si}$  for all scenarios. We highlight the capacity of the knapsack  $b$ , which divides the scenarios into two sets - those that may lead to a penalty and those that never do (those can be ignored) in case 1, and oppositely in case 2. We shall see that this property was “discovered” and exploited by our framework and all scenarios were set such that they may lead to a penalty for some solution. We did not have to incorporate such a rule explicitly. It comes from the simple fact that more scenarios enable better fit (lower value of the loss function). Thus, the minimization procedure, if designed properly, should use a maximal number of scenarios and not place any of them into the region that never leads to a penalty. This is a numerical counter-part of the analytical results in Fairbrother et al. (2017).

In 4.2b) (case 1) and 4.2c) (case 2), we show the discrepancy between in-sample and out-of-sample performance of the training and testing pool when the three final scenarios are used for in-sample evaluation. We zoom in on the best solutions from the testing pool where our main focus is. For comparison, we show in 4.2d) - 4.2f) results for 15, 30 and 50 scenarios if they are randomly drawn from the original distribution. We show the results when the seed for the random draw is 1, so we were not cherry-picking some specific output. We can find much lower, as well as much higher, discrepancy if we choose different samples, especially in the case of smaller number (15) of scenarios.

The main point of this test is to demonstrate how our framework can shape the scenarios according to needs of the optimization problem, but without an explicit analysis of the problem. With such an approach we can tremendously reduce the number of scenarios. In

our numerical example, three scenarios perform as well as fifty randomly picked scenarios, maybe even better (depending on the chosen metric for comparison).

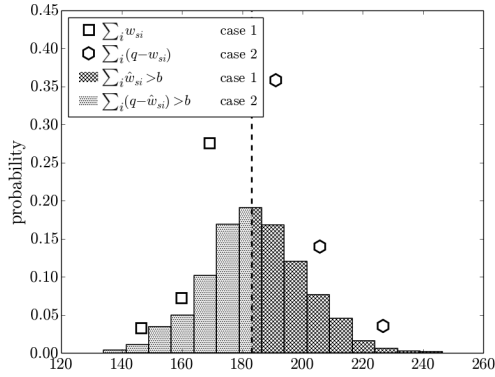
Let us point out that any method for scenario generation, which is based on a good fit between the scenario tree and the original distribution without considering the optimization problem, would inevitably return an identical scenario tree for both cases, since the original distribution is shared in both cases. Therefore, a tree with three scenarios would generate (at best) two useful scenarios for one case and only one useful scenario for the other case. Obviously, such a scenario tree would perform worse than our tree in both cases, especially in the case where only one scenario is useful.

To put it from a different perspective, one scenario tree used for both cases would need at least twice as many scenarios as our tailor-made trees for each case to reach similar quality of performance<sup>6</sup>. We admit that the problem is artificially set, but it clearly demonstrates our point that scenario trees derived purely from the original distribution, without considering the optimization problem, may lead to some redundant, or little important, scenarios.

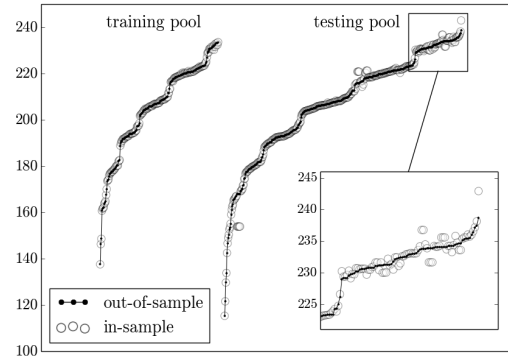
Then, a natural question is how much it actually matters from a computational point of view to keep the number of scenarios small. That will be discussed in Section 4.3, where we discuss potential applications of our approach.

---

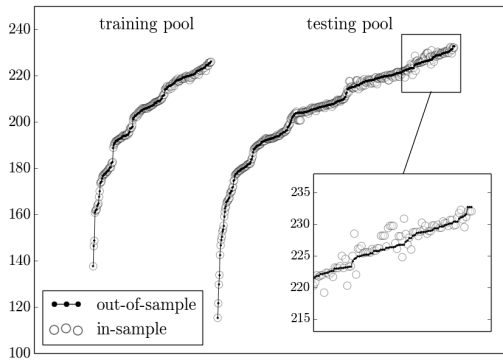
<sup>6</sup>At least according to our definition of the quality of performance, which is given by the loss function.



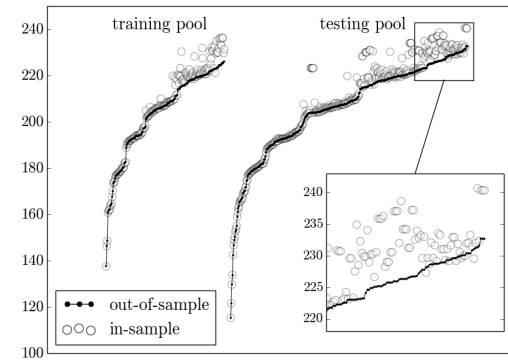
(a) Distribution of sum of items' volume



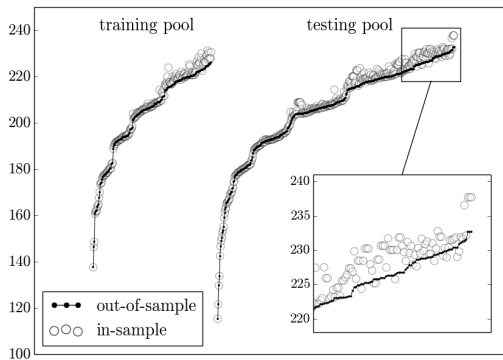
(b) Three scenarios for case 2



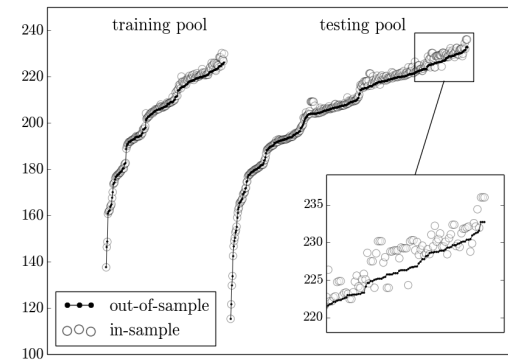
(c) Three scenarios for case 1



(d) 15 sampled scenarios for case 1



(e) 30 sampled scenarios for case 1



(f) 50 sampled scenarios for case 1

Figure 4.2: Numerical results for a stochastic knapsack problem – comparison of three scenarios obtained by our framework and scenarios sampled from the original distribution.

#### 4.2.4 Binary distributions

In this section, we discuss problems where the uncertainty is described by a multivariate Bernoulli distribution (binary distribution in short). This represents a large class of real-world applications: a customer is present or not during delivery services, weather allows a ship to sail a certain edge or not, a machine is broken or not in a scheduling problem,

just to give some examples.

And yet, stochastic programs with binary distributions are rarely studied in the literature and if they are (for example Ball et al. (1995) in the context of network reliability or Bent and Van Hentenryck (2004) in a routing problem), the focus is on the problem as such, not on handling scenarios (generation, reduction etc.). An exception is a paper Prochazka and Wallace (2018), where two useful methods are proposed; one for an efficient out-of-sample evaluation, and the second for reduction of the scenario tree into a minimal number of scenarios needed for an exact solution of the problem. However, to the best of our knowledge, there is no paper offering an efficient method for scenario generation for binary distributions (other than sampling) that would be suitable for solving, approximately, larger instances.

The advantage of our framework is that it does not rely on statistical properties and relationships among distributions. All we require are a heuristic for generating solutions, out-of-sample evaluations and an efficient procedure for minimizing the loss function. The requirements on a good scenario tree and, therefore, the definition of the loss function, do not have to be adjusted.

Although the overall framework remains unchanged, we identify two cases of problems that differ in the procedure of loss function minimization. Let us consider the following example of the knapsack problem, where items have constant value and volume, but it is uncertain whether a particular item will appear or not after the decisions (to pick or not) are made. The optimization program (using the same notation as in 4.2.3) is as follows:

$$\max_{x \in \{0,1\}^n} \sum_{s \in \mathcal{S}} p_s \left( \sum_{i \in \mathcal{I}} r_{si} c_i x_i - d e_s \right) \quad (4.11)$$

$$\text{s.t.} \quad e_s \geq \sum_{i \in \mathcal{I}} r_{si} w_i x_i - W \quad \forall s \in \mathcal{S} \quad (4.12)$$

$$e_s \geq 0 \quad \forall s \in \mathcal{S} \quad (4.13)$$

where  $r_{si}$  is the indicator of appearance of items, taking the value 1 if item  $i$  is present in scenario  $s$ , 0 if not. Assuming we have a pool of solutions, we formulate the loss function (4.3), where scenarios  $r_{si}$  and probabilities  $p_s$  are decision variables. For  $n$  items, we get  $2^n$  possible combinations (scenarios), which results in unsolvable problems for large  $n$ .

Even though the scenarios are inherently binary, there is no reason to follow that restriction when constructing the scenario tree. Relaxation of scenarios (allowing all values for  $r_{si}$ ) decreases the value of the loss function by extending the search space, and therefore, decreases discrepancy between performance of the scenario tree and the true distribution. Moreover, we can find sub-gradients with respect to  $r$  and  $p$  and use the same procedure (Algorithm 7) for minimizing the loss function as in the case of continuous scenarios.

Even though the optimization problem (4.11) – (4.13) gets a different interpretation: suddenly an item can be half-present and half-missing, it remains computationally meaningful. We need to remember that the goal is to find solutions for such a modified problem which are good also in the original problem.

In the following computational test, we consider an example with 8 items, each with a probability of appearance of 50% (independently of each other), so all the scenarios have equal probabilities. In total, there are 256 scenarios. We have a training pool consisting of 50 solutions and a testing pool of 70 solutions (all randomly generated). We want to have 4 scenarios to represent the whole distribution.

In Figure 4.3a), we show the performance of 4 scenarios obtained by our framework when allowing relaxed scenarios. In 4.3b), we choose the best combination (with minimal loss function on the training pool) of 4 binary scenarios. We see that the relaxed scenarios perform much better on the testing pool in preserving ranking among solutions and approximating their true value (they perform better on the training pool by definition). We compare the performance with scenarios that are randomly sampled (random seed 1). Four relaxed scenarios obtained by our framework outperforms 20 sampled scenarios and provide comparable results as 50 sampled scenarios by visual assessment.

There are, however, cases where the relaxation of scenarios is not helpful, albeit possible. An example would be a location-routing problem with uncertain customer appearance. A first-stage decision is the location location of a warehouse, from which customers will be served (second-stage decision). It is uncertain which customers will use the service (contracts are not signed yet). Performing an optimal routing between customers in the second stage requires solving an optimization model that contains a constraint of type

$$\sum_i x_{ij} \geq \hat{d}_{sj} \tag{4.14}$$

where  $x_{ij}$  is a binary decision variable about a choice of traversing an arc from a node  $i$  to  $j$ .  $\hat{d}_{sj}$  indicates whether a customer is present in node  $j$  in scenario  $s$ . In words, if a customer is present, he must be served. Let us assume the constraint is associated with a penalty for violation.

Any entry between 0 and 1 performs as if it was 1 in this constraint (at least one arc to  $j$  must be used). Let us assume that the parameter  $\hat{d}_{sj}$  does not appear anywhere else in the optimization problem, just in (4.14). Then, whether  $\hat{d}_{sj}$  is 1 or any arbitrary number between 0 and 1 does not have any impact on violation/non-violation of the constraint, and hence on the value of the objective function, and therefore on the value of the loss function. As a consequence, there is no change in the value of the loss function if the parameter  $\hat{d}_{sj}$ , which is already greater than zero, is changed by an infinitesimally small number.

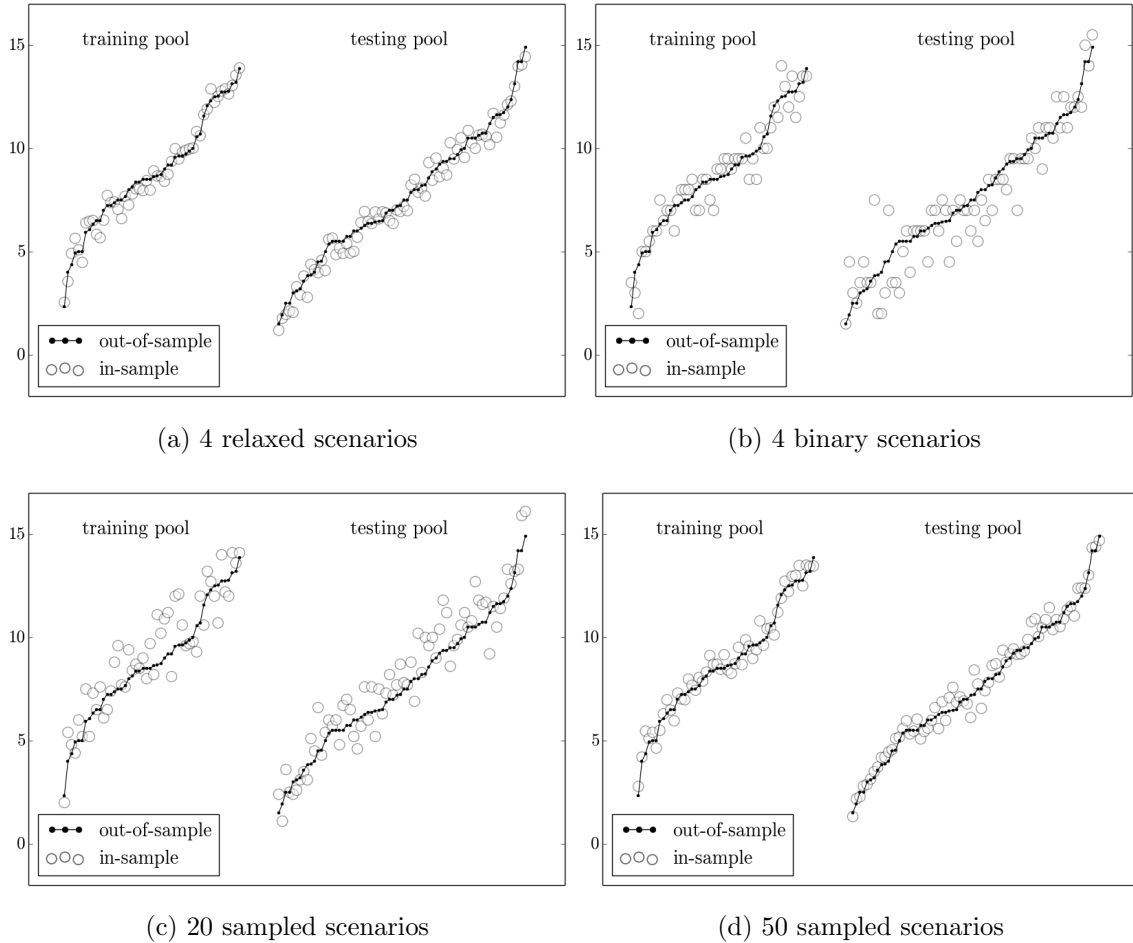


Figure 4.3: Computational test for a problem with binary scenarios.

Therefore, there is no useful change of (sub)gradients in the loss function minimization procedure and the relaxation of the scenarios is pointless.

So even if the scenarios can be principally continuous, there is no advantage in considering them to be so. We would still have to treat them as binaries ( $\hat{d}_{sj}$  is either zero or greater than zero). That means that the minimization of the loss function is more problematic as it leads to solving a non-linear and non-convex problem with binary variables. In such a case, we suggest using meta-heuristics (for example genetic search), which are state-of-art methods for combinatorial optimization problems. But the general scheme – we minimize the discrepancy between in-sample and out-of-sample performance – remains unchanged.

Dealing with a network with  $n$  potential customers, the total number of possible scenarios is exponential in  $n$ . That is, again, often impossible to handle for large instances. Thus we can use only a subset of all possible scenarios. Our framework can help us identify a suitable set.

### 4.2.5 Feasibility

So far in the text, we assumed (relatively) complete recourse. That is, we assumed that all feasible first-stage decisions would lead to feasible second-stage problems for all possible values of the random variables. We also assumed that the heuristic producing our pool of solutions makes sure that they are all first-stage feasible.

In this section, we discuss the case where, for some reason, feasibility of the second-stage problem is an issue. We want to point out that this should be a rather seldom. A constraint saying that (random) demand must always be satisfied leads to a worst-case (and at the same time, most likely, a subjective) model, which hardly makes sense. It may be true that a time window is hard in the sense that outside the time window it is impossible to deliver, but it hardly means that life does not go on if the truck is late. Rather, a penalty is incurred, and the activities continue. So, really hard constraints (violating the ideas behind relatively complete recourse) are extremely rare from an applied perspective. Even so, we shall discuss the issue to some extent here. Note that even in cases where some constraints need to be satisfied in every scenario, it is possible to use a penalty several orders of magnitude larger than the other profits (costs) in the objective function. Then, if there is a solution that can satisfy all the constraints, the optimization model will prefer it. If there is not such a solution, we can see which constraints have been violated. Such information is more valuable for analyzing the problem than a simple report that there is no feasible solution for the problem. For further discussion on feasibility modeling, see King and Wallace (2012a).

In addition, it is also challenging to incorporate a (in)feasibility classification into our framework from a computational point of view as we lose some properties of the loss function, mainly the utilization of sub-gradients. Thus, we recommend to use penalization whenever possible.

However, if feasibility in the second stage really is an issue, for example due to a constraint related to some laws of physics, and which therefore cannot be violated, we introduce a new requirement on a good scenario tree that stands above those formulated in Section 4.2.1. We postulate it as follows:

0. A good scenario tree *classifies feasibility* of solutions “more or less” correctly. That is, true feasible solutions are also feasible in the tree. The same holds for infeasible solutions. But as with outliers, we find it acceptable if occasionally a solution that is feasible in reality, is classified as infeasible by the tree, especially if the solution is weak. On the other hand, we want to avoid cases where very good solutions (high out-of-sample value) that are infeasible in reality, become feasible in the tree.

Let us assume that the heuristic generates both feasible and infeasible solutions. We construct a list  $F$  of feasibility indicators, that is  $F_a = 1$  if  $x_a$  is feasible (out-of-sample),



0 otherwise. Further, let the function  $u(x, \mathcal{T})$  return 1 if  $x$  is feasible in the tree and 0 if the solution is infeasible in the tree.

We keep all the terms from the loss function as they were defined previously to reflect Requirements 1 - 4, and we add some new terms to capture Requirement 0. The loss function is

$$L^F(\mathcal{T}, \mathcal{A}) = L(\mathcal{T}, \mathcal{A}) + \sum_{x_a \in \mathcal{A}} \left( z_3^a u(x_a, \mathcal{T})(1 - F_a) + z_4^a (1 - u(x_a, \mathcal{T}))F_a \right) \quad (4.15)$$

We simply add the weight  $z_3^a$  if an out-of-sample infeasible solution is classified as feasible by the tree, and weight  $z_4^a$  if an out-of-sample feasible solution is classified as infeasible. Since the first case is more critical, penalties  $z_3^a$  are set higher than  $z_4^a$ . Similarly as for  $z_1^a$ , weights  $z_3^a$  and  $z_4^a$  also depend on the quality of the solution  $x_a$ . The higher the out-of-sample value is, the higher weights we set, especially in the case of  $z_3^a$ , so we penalize very good (high value of the objective function) but infeasible solutions that are incorrectly classified as feasible by the tree.

Naturally, it is challenging to set all the weights  $z_1 - z_4$  properly to create a perfect balance between classification of feasibility and approximation of the out-of-sample values. The balance should be derived from the usage of the model and assessment of importance of having feasible solutions.

The main issue is the minimization of the loss function, which is much more difficult in this case due to the added classification terms. The main problem is that we cannot utilize the sub-gradients of the function  $u(x, \mathcal{T})$ . The function is not continuous and returns only two values, that means its derivative is 0 (if it is defined). One needs to either use methods for non-linear optimization that do not utilize derivatives or approximate the function  $u$  with some differentiable function as is demonstrated in the next example.

## Example

Let us consider an alternative version of the stochastic knapsack problem

$$\max_{x \in \{0,1\}^n} \sum_{i \in \mathcal{I}} c_i x_i - d \sum_{s \in \mathcal{S}} p_s e_s \quad (4.16)$$

$$\text{s.t} \quad 0 \geq \sum_{i \in \mathcal{I}} w_{si} x_i - W^{\max} \quad \forall s \in \mathcal{S} \quad (4.17)$$

$$e_s \geq \sum_{i \in \mathcal{I}} w_{si} x_i - W \quad \forall s \in \mathcal{S} \quad (4.18)$$

$$e_s \geq 0 \quad \forall s \in \mathcal{S} \quad (4.19)$$

The total volume of picked items may exceed capacity of the knapsack and we pay the corresponding penalty, but we cannot exceed the total value  $W^{\max}$  ( $> W$ ) in any single scenario.

Classification of feasibility, i.e., the approximation of  $u(x, \mathcal{T})$ , is performed by a simple neural network model with one hidden layer, sigmoid function as the activation function in both layers and the sum of squares as the measurement of the error. As input we use the vector  $e_s$  scaled to  $(0, 1)$ . The neural network is trained on the training pool of heuristically obtained solutions. This simple model works well in our case and correctly classifies (in)feasibility of solutions.

The main advantage of this approach is that the neural network can (back)propagate the sub-gradients of error (miss-classified solutions) via the weights of the neural network to the sub-gradient of the  $e_s$  vector and further to scenarios  $w_{si}$ . Thus, we can, in principal, use Algorithm 7 to find the scenario tree.

We compare the performance of scenarios constructed by our framework with scenarios obtained by pure sampling from the original distribution and by using a copula-based heuristic (Kaut, 2014) in Figure 4.4. The comparison is made on the testing pool as it was not used to construct scenarios by our framework. The testing pool can be perceived as a proxy for the whole search space. Each pool consists of feasible and infeasible solutions sorted in ascending order based on the value of the objective function with highlighted incorrectly classified solutions.

Since no feasible solution leads to a total size exceeding  $W^{\max}$  in any realization (scenario) of the original distribution (by the definition of the feasible solution), the size also does not exceed  $W^{\max}$  in any subset of scenarios drawn from the original distribution. In other words, sampled scenarios will always classify correctly out-of-sample feasible solutions, see Figure 4.4a). However, it may happen that in some scenario from the original distribution, the limit  $W^{\max}$  is exceeded, but such a scenario is not chosen in the sampled subset. Thus, several out-of-sample infeasible solutions are classified as feasible in the tree. Since some of them return high value of the objective function, they have a negative impact on the subsequent optimization as they provide too optimistic and almost always infeasible (out-of-sample) solutions, even if the sample is very large.

We observe a similar phenomena in the case the scenarios are obtained by the copula-based heuristic, Figures 4.4b) and 4.4c). The scenarios are constructed by generating realizations from marginal distributions and then combining them to match the shape of the copula of the original distribution. Thus, even feasible solutions can be occasionally incorrectly classified. This heuristic works extremely well when it comes to matching most of the properties of the original distributions (notice almost perfect estimation of the objective value). However, it is not designed to focus on capturing, in some sense, extreme

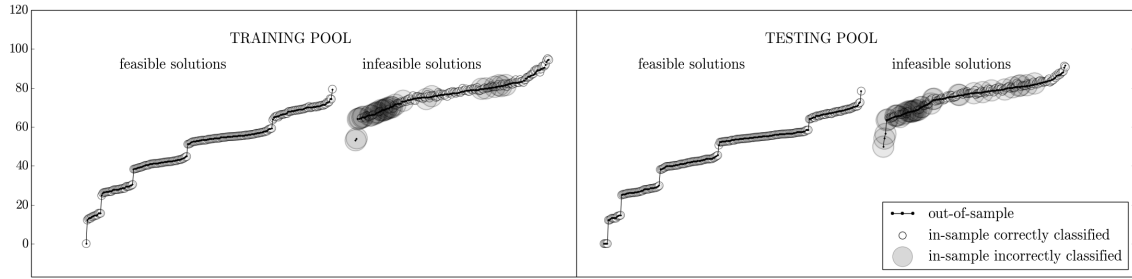
scenarios that cause second-stage infeasibility. Thus, even a large scenario tree consisting of 100 scenarios might cause the occasional appearance of missclassified solutions that look feasible in the tree, but are infeasible in reality (due to one extreme scenario for example).

In contrast to the above described methods, our framework (Figure 4.4d) primarily focuses on correct classification of (in)feasibility, especially on not letting good (high value of the objective function) infeasible solutions be classified as feasible in the tree. To satisfy that requirement, the tree tends to set scenarios more towards their extremes, and thus classify feasible solutions as infeasible more often in favor of correct classification of high-value infeasible solutions. That is a preferable situation for solving the optimization program.

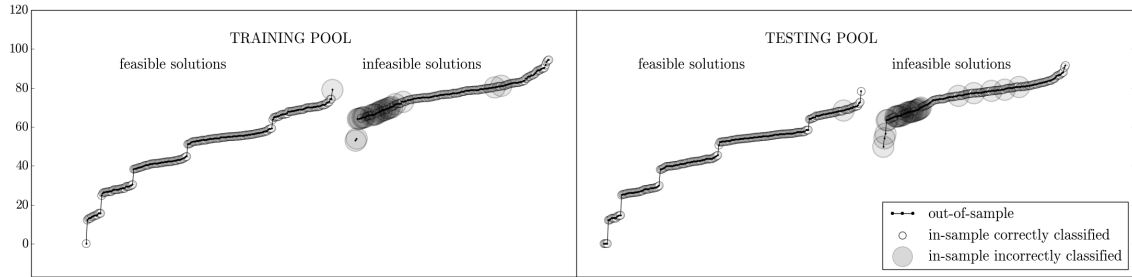
Naturally, there is no guarantee that the presented approach is able to always capture the extreme directions of all scenarios to prevent that a good infeasible solution is incorrectly classified. This risk could be reduced by using a larger training pool.

To further minimize that risk, it would be possible to tighten constraints that cause infeasibility. In our case we could set  $W^{\max}$  smaller, when solving the optimization program with the scenario tree. Then we have more certainty that we are on the “safe” side.

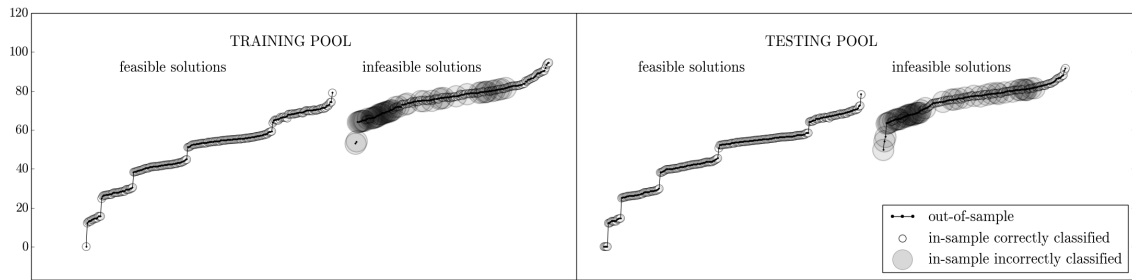
This leads to an idea that it is possible to set other parameters (input data), not only the scenarios, to mimic performance of the original distribution. Our approach provides a framework that can serve that purpose (minimizing discrepancy between in-sample and out-of-sample performance). However, minimization of the loss function becomes more complicated as there are more variables to set.



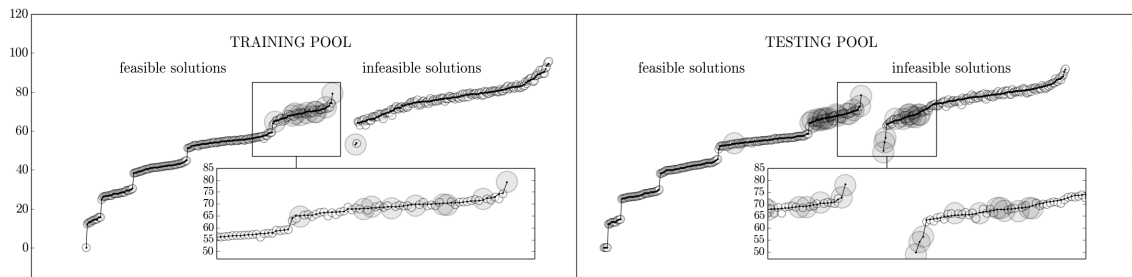
(a) 100 sampled scenarios



(b) 100 scenarios generated by copula-based heuristic



(c) 20 scenarios from copula heuristic



(d) 6 scenarios from our framework

Figure 4.4: Computational test for a problem with infeasible solutions.

### 4.3 Applications

In this section, we summarize the main advantages and drawbacks of the framework. Based on that, we comment for what types of applications our approach could be beneficial, and

where it is better to use a different method.

The main advantage is the number of scenarios needed to represent the underlying distribution. We demonstrate on several examples that just a small number of scenarios can perform as well as a much larger tree since they are “tailor-made” for a particular optimization problem. Our framework enables identifying spots where the scenarios are most useful. Thus, we aim for applications where it is crucial to keep the number of scenarios small.

The main disadvantage is the time to develop two subroutines; a heuristic for generating solutions and the subsequent procedure for minimization of the loss function, which is difficult. Therefore, we do not see any reason to use our framework for optimization problems that are run only once in principal (strategic problems) and/or can handle a large number of scenarios relatively easily (linear programs). In such a case, savings in computational time when solving the program (4.2) with a smaller number of scenarios do not exceed (most likely) time needed for running the heuristic, tuning parameters of the loss function minimization procedure and running it. In addition, we need time to develop these procedures. Thus, for simple linear two-stage models, we recommend using some different method, for which a publicly released code can be found.

A typical example that could utilize our framework would be a stochastic vehicle routing problem<sup>7</sup> (VRP) that needs to be solved repeatedly. Stochasticity in routing problems can be related to uncertain travel conditions (potential congestion), uncertain demand, etc. See an overview of the field in Gendreau et al. (2016). For a dispatching company, this means solving the problem every day (or several times per day) with different input data, but the formulation of the optimization problem is the same. Thus, one can invest into the development of a heuristic, especially if it is likely that one needs the heuristic even for solving the final optimization problem. Typical state-of-art heuristics (genetic search, adaptive neighborhood search, simulated annealing, etc.) for VRPs produce many solutions very quickly during the process. We can imagine that the pool of solutions is obtained by multiple runs of the heuristic with different subset of scenarios (for example randomly sampled), and then, the same heuristic is run with the final scenario tree obtained by minimizing the loss function. We can even imagine, that the two subroutines - solution search and the scenario tree search - can be merged and cleverly implemented into one heuristic with several updates of both subprocesses.

Another application of our framework is on problems with distributions that have no alternative ways to generate scenarios, except for sampling. But sampling often requires too many scenarios to provide a stable solution, and that is sometimes unaffordable from a computational point of view. To this class of problems we can include programs with the

---

<sup>7</sup>Or generally a difficult problem, where the number of scenarios, and thus the number of constraints/variables significantly influence the computational time.

multivariate Bernoulli distribution. An example is VRPs with uncertain appearance of customers. We already made some points that play in favor of our approach when dealing with VRPs. Issues related to binary distributions are analyzed in Section 4.2.4.

There are many real-world application where we need to work with a combination of different distributions (different random variables). Some of them might be empirical, some theoretical, some might be binary, some continuous, etc. It is difficult to handle this issue by other methods for scenario generation. Due to the fact that our framework does not rely on statistical properties of the distribution, but simply looks for a scenario tree that mimics the out-of-sample performance in some (defined) way, it can be, in principal, used in such a case once we are able to evaluate the out-of-sample quality of a solution.

The last field we want to mention is multi-stage optimization (for example Dantzig and Infanger (1993)). Our numerical experiments show that our framework can produce scenario trees with a smaller number of scenarios than other methods with the same level of solution quality. Since the size of the scenario tree growth exponentially with the number of stages in a multi-stage setting, the smaller number of scenarios per stage implies significant reduction in size of the overall tree. Moreover, the scenarios can be generated stage by stage by simply trying to mimic the behavior of the original distribution at that particular stage. Thus, we do not have to control dependencies across the stages as the case when matching statistical properties of the scenario tree and the original distribution. This is not further studied in this paper.

## 4.4 Conclusion

We introduce a new problem-oriented approach for generating scenarios for stochastic optimization. It is not based on matching the scenario tree and the underlying distribution in some probabilistic sense, but it sets the scenario tree in such a way that the tree performs similarly as the original distribution. The performance is evaluated on a pool of solutions that are produced by some heuristics. The similarity of performance is measured by a loss function that we introduce. Its formulation is derived from our postulates on what constitutes a good scenario tree.

Hence, it is the optimization problem that drives the scenario generation by searching for a tree that minimizes the loss function. Thus, the parts of the original distributions that are more crucial for good solutions are approximated with greater emphasis. This approach often leads to a smaller number of scenarios compared with other methods. The main disadvantage is that two main subroutines - the heuristic for generating solutions and the procedure for minimization of the loss function - are not necessarily trivial and need to be developed specially for a given problem.

Thus, the framework is suited for applications where it is critical to use as few scenarios

as possible, for example difficult problems (non-linear, integer) that seriously increase their complexity with the number of used scenarios. We believe it is worthwhile to choose our approach especially in cases where the problem is solved repeatedly (with different input data), so we can utilize the implemented parts several times.

Another usage of our framework is on problems where uncertainty is represented by distributions, for which there is no alternative to pure sampling (even that might be problematic when we have a combination of different distribution types). In this paper, we discuss the case of binary distributions that have many applications in real life, for example a customer that appears randomly in a routing problem, a machine that might not work in a scheduling problem, or a cargo from A to B that is available on a future spot market with a certain probability.

We offer an alternative point of view on the relationship between optimization problems and scenario trees. Even if a different method for scenario generation is used, a simple visual test that compares in-sample and out-of-sample performance of a pool of heuristic solutions, can provide an intuitive way to assess the quality of the tree.

We introduce some ideas for further extension of this research. One is to use our framework on multi-stage optimization problems, where the number of scenarios grows exponentially with the number of stages. Thus, it is desirable to have as few scenarios per stage as possible. Another idea, that is not elaborated in this paper, is to consider other input data (parameters), and not only the scenarios, to be set according to our framework. Thus, we would create a modified problem, whose solution would, hopefully, be similar to the solutions of the original problem using the entire distribution.

## References

- Ball, M. O., Colbourn, C. J., and Provan, J. S. (1995). Network reliability. In Ball, M. O., Magnanti, T. L., Monma, C. L., and Nemhauser, G. L., editors, *Network Models*, volume 7 of *Handbooks in Operation Research & Management Science*, chapter 11. North-Holland, Amsterdam.
- Bent, R. W. and Van Hentenryck, P. (2004). Scenario-based planning for partially dynamic vehicle routing with stochastic customers. *Operations Research*, 52(6):977–987.
- Birge, J. and Louveaux, F. (1997). *Introduction to stochastic programming*. Springer-Verlag, New York.
- Boyd, S. and Vandenberghe, L. (2004). *Convex optimization*. Cambridge university press.
- Cario, M. C. and Nelson, B. (1997). Modeling and generating random vectors with arbitrary marginal distributions and correlation matrix. Technical report, Department of

- Industrial Engineering and Management Sciences, Northwestern University, Evanston, Illinois.
- Dantzig, G. B. and Infanger, G. (1993). Multi-stage stochastic linear programs for portfolio optimization. *Annals of Operations Research*, 45(1):59–76.
- Fairbrother, J., Turner, A., and Wallace, S. (2017). Problem-driven scenario generation: an analytical approach to stochastic programs with tail risk measure. ArXiv e-print 1511.03074.
- Garey, M. and Johnson, D. (1979). *Computers and Intractability, a guide to the theory of NP-completeness*. Freeman, New York.
- Gendreau, M., Jabali, O., and Rei, W. (2016). 50th anniversary invited article—future research directions in stochastic vehicle routing. *Transportation Science*, 50(4):1163–1173.
- Hendrix, E. and Tóth, B. (2010). *Introduction to Nonlinear and Global Optimization*. Springer-Verlag, New York.
- Higle, J. L. and Sen, S. (1991). Stochastic decomposition: An algorithm for two-stage linear programs with recourse. *Mathematics of Operations Research*, 16:650–669.
- Høyland, K. and Wallace, S. W. (2001). Generating scenario trees for multistage decision problems. *Management Science*, 47(2):295–307.
- Kall, P. and Wallace, S. W. (1994). *Stochastic Programming*. John Wiley & Sons, Chichester.
- Kaut, M. (2014). A copula-based heuristic for scenario generation. *Computational Management Science*, 11(4):503–516.
- Kaut, M. and Wallace, S. W. (2007). Evaluation of scenario-generation methods for stochastic programming. *Pacific Journal of Optimization*, 3(2):257–271.
- Kaut, M., Wallace, S. W., Vladimirov, H., and Zenios, S. (2007). Stability analysis of portfolio management with conditional value-at-risk. *Quantitative Finance*, 7(4):397–409.
- King, A. J. and Wallace, S. W. (2012a). *Modeling Feasibility and Dynamics*, chapter 2. In King and Wallace (2012b).
- King, A. J. and Wallace, S. W. (2012b). *Modeling with Stochastic Programming*. Springer Series in Operations Research and Financial Engineering. Springer.



- King, A. J., Wallace, S. W., and Kaut, M. (2012). *Scenario-Tree Generation*, chapter 4. In King and Wallace (2012b).
- Lurie, P. M. and Goldberg, M. S. (1998). An approximate method for sampling correlated random variables from partially-specified distributions. *Management Science*, 44(2):203–218.
- Pflug, G. C. (2001). Scenario tree generation for multiperiod financial optimization by optimal discretization. *Mathematical Programming*, 89(2):251–271.
- Prochazka, V. and Wallace, S. W. (2018). Stochastic programs with binary distributions: structural properties of scenario trees and algorithms. *Computational Management Science*.