Robustness, Complexity, Validation and Risk

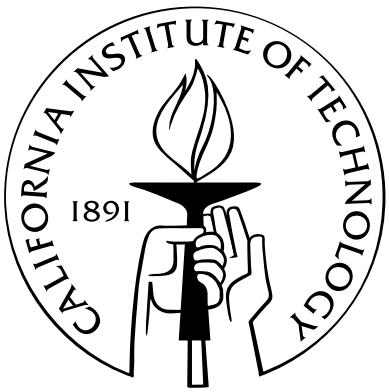
Thesis by

Xin Liu

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy



California Institute of Technology

Pasadena, California

2007

(Defended February 15, 2007)

To Hualin

Acknowledgements

First of all, I am obliged to my advisor, Professor John Doyle. John provided me with the opportunity to pursue my doctoral degree at Caltech, for which I will be forever thankful. His incredible vision, enthusiasm for research and courage to tackle the hardest problems have been and will always be an inspiration to me.

I would like to thank Professor Tau-Mu Yi for initiating the G-protein signaling pathway project and for providing the model and experiment data. I learned a great deal from the discussions we had when working on this project. My gratitude extends to Professors Richard Murray and Mani Chandy for serving on my thesis committee. I also want to thank Professor Hideo Marbuchi for being on my candidacy committee.

I have been very fortunate to work with Professor Ivana Komunjer who, due to unfortunate reasons, is not able to be on my thesis committee. I am grateful for her patience in answering my naive questions when I first started working in econometrics and her encouragement throughout. My appreciation also goes to Professors Jerry Marsden and Yaser Abu-Mustafa for their support during my study at Caltech.

To my friends at Caltech, thank you for all the good times we spent. Lun Li, Jiantao Wang, Zhipu Jin, Ling Shi, Lijun Chen, Huirong Ai, Qiang Yang, Zhengrong Wang, Cheng Jin, Xiaoliang Wei, Ao Tang – thank you all for making this place far away from home feel like home. My first years in the U.S. would not have been the smooth experience that they were, were it not for my buddies. Melvin Flores, Melvin Leok, Stephen Prajna, Domitilla Del Vecchio, Antonis Papachristodoulou, Shreesh Mysore and Harish Bhat, thank you for the new things you taught me and all the fun we had. I thank my friends at CDS, in particular, Maryam Fazel, Tosin

Otitoju, Alfredo Martinez, Aristo Asimakopoulos and Abhishek Tiwari, for those jokes, conversations and drinks at red door that made my days a lot more enjoyable. Especially to Maryam, not only for the collaborations we had, but also for being such a respectable friend.

There are many people who have helped me during my stay at Caltech. Specifically, I thank Marjory Gooding, Jim Endrizzi, Athena Trentin and Parandeh Kia at International Student Programs for advising me through those exhausting immigration procedures and for going out of their way to lend help.

Finally, I would like to thank my parents and my brother for their unconditional love and support throughout the years. They give me strength to pursue my dreams. Last, but in no ways the least, I wish like to thank my dear husband Hualin for loving me, understanding me and always believing in me. I dedicate this thesis to him.

Abstract

A robust design process starts with modeling of the physical system and the uncertainty it faces. Robust design tools are then applied to achieve specified performance criteria. Verification of system properties is crucial as improvements on the modeling and design practices can be made based on results of such verification. In this thesis, we discuss three aspects of this closed-loop process.

First and the most important aspect is the possibility of the feedback from verification to system modeling and design. When verification is hard, what does it tell us about our system? When the system is robust, would it be easy to verify so? We study the relation between robustness of a system property posed as a decision problem and the proof complexity of verifying such property. We examine this relation in two classes of problems: percolation lattices and linear programming problems, and show complexity is upper-bounded by the reciprocal of robustness, i.e., fragility.

The second aspect we study is model validation. More precisely, when given a candidate model and experiment data, how do we rigorously refute the model or gain information about the consistent parameter set? Different methods for model invalidation and parameter inference are demonstrated with the G-protein signaling system in yeast to show the advantages and hurdles in their applications.

While quantification of robustness requirements has been well-studied in engineering, it is just emerging in the field of finance. Robustness specification in finance is closely related to the availability of proper risk measures. We study the estimation of a coherent risk measure, Expected Shortfall (ES). A consistent and asymptotically normal estimator for ES based on empirical likelihood is proposed. Although em-

pirical likelihood based estimators usually involve numerically solving optimization problems that are not necessarily convex, computation of our estimator can be carried out in a sequential manner, avoiding solving non-convex optimization problems.

Contents

List of Figures

List of Tables

List of Acronyms

BCBS	Basel Committee on Banking Supervision
CFP	Cyan Fluorescent Protein
CVaR	Conditional Value-at-Risk
ES	Expected Shortfall
FRET	Fluorescence Resonance Energy Transfer
GDP	Guanosine Diphosphate
G-protein	Guanosine triphosphate-binding protein
GTP	Guanosine Triphosphate
LP	Linear Program
MELE	Maximum Empirical Likelihood Estimator
NP	Non-deterministic Polynomial time
ODE	Ordinary Differential Equation
P	Polynomial time
TCE	Tail Conditional Expectation
VaR	Value-at-Risk
WCE	Worst Conditional Expectation
YFP	Yellow Fluorescent Protein

Chapter 1

Introduction

1.1 Motivation

The omnipresence of uncertainty makes robustness a necessary feature of almost all natural and engineering systems around us. Biological systems exhibit astonishing robustness to uncertainty arising from both their building blocks and the environment. This robustness is crucial for survival and growth of organisms. In engineering design, from aircraft design to communication, building reliable systems robust to uncertainty has always been the goal. Robustness is essential since, for one thing, the mathematical model used to describe the underlying physical system is never perfect. Past experience with financial market crashes as well as sudden fallings of banks also urge individual investors and banks to invest with strategies that are robust to the huge uncertainty in market.

This quest for robustness renders scientists and engineers constantly challenged with the task of understanding and design robust performing systems. Fig 1.1 shows a coarse depiction of the robust design process. Mathematical modeling of the physical system, the uncertainty it is likely to face, and the performance criteria it should achieve is the first step of this process. Robust design is then conducted based on the model of the system with the aim of satisfying performance criteria in the presence of the perturbations specified. Verification is a crucial step to check that the properties designed are actually obtained. A necessary feature of this design process is that it

be closed-looped. In other words, the result of verification gives information about the previous steps and points to directions of improvement.

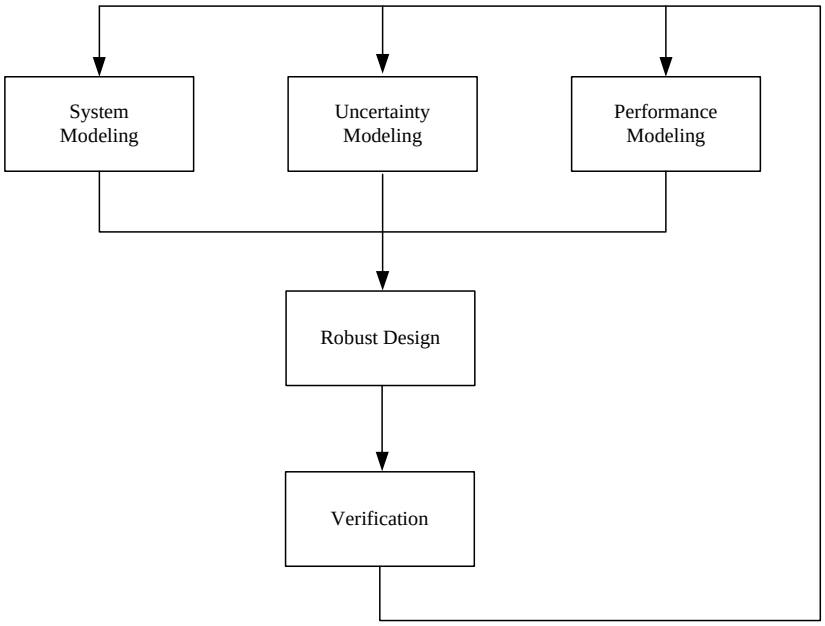


Figure 1.1: Robust design process

In this thesis, we discuss three aspects of the robust analysis and design process. First, and the most important aspect, is the possibility of the feedback from verification to system modeling and design. When verification is very hard, what does it tell us about our designed system? When the system is robust, would it be easy to verify this? Does there exist a fundamental relation between verification and robustness? The second aspect is the construction of mathematical model of the physical system. When we have data from experiments, how do we check the validity of a candidate model structure? When the structure is valid, what can we say about its parameters? The last, but in no means the least important, aspect is the quantification of uncertainties the system is robust to, i.e., quantification of the robustness measure.

The properties or functionalities of a system are often investigated through posing questions about the system. For example, a positive answer to the question "is it safe for a Boeing 777 to fly in windy weather" is a desired property of the Boeing 777

design. A negative answer to the question "does the running of the software cause a computer to crash" would be a necessary requirement of a nice piece of software. Thus a property is robust when its corresponding question is robust, i.e., the answer does not incur qualitative change in the face of uncertainty. Its verification also reduces to proving a solution to the posed question. While some questions are easy to answer, others can be very hard. What does this hardness in proving a property tell us about the underlying system?

In statistical physics, complexity has been associated with phase transition through empirical examination of random systems. Studies of problems such as the satisfiability (SAT) problem, show that the peak of average complexity occurs at phase transition, see, e.g., [15, 55, 56]. The conclusion is then that problems are hard near phase transition and easy away from this critical point. Although this is correct in an average sense and corroborated by empirical data, it yields little information about a particular instance. For example, unsatisfiable 3-SAT instances below phase transition can be significantly harder than a "typical" instance at phase transition [55]. Then what does complexity of an instance tell us, now that it is not bound with phase transition?

Furthermore, biological and engineering systems are hardly ever random. The structure and interactions of biological networks are results of millions of years of natural selection. Engineers deliberately design their systems to be the way they are. So the complexity of designed systems, especially those that are highly optimized to be robust, rather than that of random systems, should be the focus of engineers in the effort to understanding complexity.

As suggested by overwhelming evidences, this proof complexity reveals fragility of the associated property to some uncertainty. Complexity is fundamentally linked to robustness, rather than phase transition. A familiar example in robust control is the structured singular value problem, where lack of short proof of robustness under time invariant perturbations indicates fragility against certain time variant perturbations.

We study the relation between verification complexity and robustness in detail for two classes of problems, namely the percolation lattice problem and Linear Programming (LP). The conclusion for these two case studies is that proof complexity implies fragility. In other words, robust properties are necessarily easy to verify. Our conjecture is that this fundamental relation holds for general systems.

The second aspect we discuss in this thesis is model invalidation and parameter inference. Building a mathematical model of the natural or designed system is the first step of robust design. The model should not be too simple to capture the essence of the original system; nor should it be too complicated to work with. A good model needs to be consistent with known knowledge, able to reproduce observed empirical data, and robust to a reasonable degree of uncertainty. Uncertainties arise from dynamics that are not reflected in the model, fluctuations in environment conditions, as well as experiment errors.

For a specific ODE model, the traditional method of checking model validity consists of carrying out computer simulations to see if experiment data can be reproduced. Uncertainty from model parameters requires exhaustive search over the parameter set, making this approach computationally virtually impossible. Invalidating a model with uncertainty is a co-NP problem, as it requires proving that no parameter combination can result in a model consistent with observed data.

Barrier Method proposed by Prajna [69] is a solution to model validation with uncertainty. Barrier function is conceptually similar to Lyapunov function, and is implemented via SOSTOOLS [70, 71]. While it is theoretically elegant and powerful, its application to real systems, especially complex biological systems, is hindered by computational limitations. We adopt a surface fitting approach [9,57] previously used in chemical engineering to ease the computation load, at the price of introducing a level of relaxation.

When a proposed model is not invalidated, can we infer information about possible parameter combinations from data? Traditional methods concentrate on estimating a

single best parameter and using this estimate to predict system outcome. For a given set of experimental data, there is usually more than one parameter that is capable of producing the observed data. Predictions using a single estimate are therefore limited, even misleading. We study parameter inference with a different approach, where the focus is on the whole set of parameters that are consistent with prior information, e.g., reaction rates lying in certain range, and the data. Surrogate model method and SOS relaxation are used to produce bounds on this consistent parameter set.

In biology, understanding of the mechanism through which complex networks of interacting molecules achieve robust functionalities has to come first, before schemes altering the biological system for pharmaceutical benefits can be designed. Proper mathematical models often help in digging out essences of natural design. Because of the complexity of biological systems, modeling is an important topic in system biology. We study the model validation problem in G-protein signaling system in yeast, a system that is biologically significant yet simple enough for us to see where and how the different methods work and break.

The third aspect we study is the robustness measure. Robustness measure is studied in the field of financial risk management, where quantification of robustness measure is less straightforward than in engineering and is still developing. In problems such as financial investing, or portfolio choice, robustness concern is implicit, but its formulation is not agreed on.

This difficulty is mainly due to people's different perception of risk: a situation considered risky by one person can seem acceptable to another. The general concept of risk in financial risk management is the exposure to uncertainty. Imposing different measures of risk on an investment strategy specifies different uncertainties that the strategy needs to be robust to. A "good" measure of risk should reflect rational decision making under uncertainty. For example, it should explain and encourage diversification that is widely observed in practice. This concept of "goodness" was rather unclear until very recently, when the definition of coherent risk measure was

introduced by Artzner et al. in 1999 [3] and that of convex risk measure by Föllmer and Schied in 2002 [27].

Risk measure is essential in guiding portfolio choice. The aim of portfolio choice is to form a combination of available financial assets, possibly including bonds, stocks, derivatives and other financial instruments, to obtain the maximum anticipated return over a specific period of time while enduring the smallest risk. Here risk should be roughly interpreted as the stochastic counterpart of robustness measure in control system design. For example, in robust control, requiring $\mu(M, \mathbf{\Delta_{TI}}) < 1$, where $\mathbf{\Delta_{TI}}$ is the set of all block-diagonally structured time invariant perturbations within the unit ball, guarantees that the nominal system M is robust against perturbations in $\mathbf{\Delta_{TI}}$. Analogously, an investor imposing an upper bound, say 5%, on the probability of monetary loss from the portfolio wants her investment to be robust to the best 95% of the market movements and considers the worst 5% fluctuations out of control. Whether this is a wise choice of criterion is going to be discussed in Chapter 4.

In portfolio selection, portfolio weights are the decision variable to be determined to achieve the return performance, while keeping undesirable outcome to a tolerable level. It is commonly agreed on that the return performance of a portfolio is measured by the expected return, or mean return. The appropriate measure of risk however is less defined. Variance and quantile have been used as risk measures. However, variance is inappropriate as a measure of risk for asymmetric returns, which is common in financial returns. Quantile-based risk measure, in particular Value at Risk (VaR), focuses on the tail distribution of the return, but it fails to explain diversification and is thus not coherent.

We study a measure of risk called Expected Shortfall (ES), which is essentially the average loss in the unwanted situations. ES is a coherent risk measure that encourages diversification. It is also the candidate measure of credit risk that will potentially be posed to banks by regulators. Determining the value of ES for a portfolio, i.e., estimation of ES, is therefore an important research topic with material practical

impact. A semi-parametric estimation method based on empirical likelihood function is proposed, where no assumptions on the return distribution are made. Asymptotic properties and the computation of the estimator are examined.

1.2 Contributions and Outline

The contributions and outline of the thesis are as follows.

In Chapter 2, we study the relation between robustness of a system property posed as a decision problem and the verification complexity of verifying such property. We examine this relation in two classes of problems: percolation lattices and linear programming (LP) problems. The decision problems are the determination of the existence of a path in a lattice and a solution to a feasibility LP problem. Robustness and verification complexity for both problems are defined. Complexity is shown to be upper-bounded by the reciprocal of robustness, i.e., fragility, grounding the harmony between robustness and verification in system design. We also illustrate that this relation is more relevant in designed systems than the popular attribution of complexity to criticality for percolation lattices. As percolation lattices and LP are proper models for a variety of physical systems, what we study here is a first step in understanding the fundamental link between robustness and complexity, followed by similar explorations for problems that are in NP or are undecidable [63].

Chapter 3 is focused on the problem of model validation. More precisely, when given a candidate model and experiment data, the task is to rigorously refute the model or gain information about the consistent parameter set. Based on the SOS methodology, invalidation is carried out using Barrier function [69] or surrogate approximation. Characterization of parameter correlation based on SOS methods is proposed. These techniques are demonstrated with the G-protein signaling system in yeast to show the advantages and hurdles in their application to system biology.

In Chapter 4, we study estimation of a particular risk measure, the Expected

Shortfall. A maximum empirical likelihood estimator (MELE) for the joint estimation of VaR and ES is proposed. Asymptotic properties of this estimator is developed. In particular, we show that the estimator is consistent and asymptotic normal. In general, a MELE is a solution to an optimization problem that is not necessarily convex. Its implementation thus suffers from difficulties in solving non-convex optimization problems. We show that the computation for our MELE can be carried out in a sequential manner, where the first stage is a comparison of two evaluations and the second stage is an algebraic calculation.

Chapter 5 concludes and gives future directions.

Chapter 2

Robustness and Complexity

Functionalities and their robustness are key properties in biological, economic and engineering systems. Verifying these functionalities has been a long-standing effort in a number of communities. Very often, certain questions are asked about a system in the hope that by answering these questions, we can gain insight in understanding properties of the system concerned. In this chapter, two classes of systems are explored to show that there is a relation between the hardness of answering a simple but meaningful question about the system and the robustness of the system.

The two problems, percolation lattices and linear programming, have been studied extensively in statistical physics and operation research, respectively. These two problems are widely used as models for a variety of physical systems. They are neatly formulated and simply solvable, yet at the same time capturing the essence of the conceptual difficulty in comprehending the relation mentioned above. They provide us with a natural starting point to investigate our conjecture: robust properties have short proofs.

Before looking at these two problems, let us first remind ourselves of the different notions of complexity.

2.1 Notions of Complexity

In common language, a complex system is one with massive number of heterogenous components interwoven together, possibly through complicated hierarchies, to achieve what would not be possible by individual components. As complex systems are studied in almost all research fields, unsurprisingly there are many definitions of complexity. We briefly introduce the concept of computational complexity in contrast with the verification complexity, or proof complexity, that we focus on.

The notion of complexity concerning hardness of problems is computational complexity that came with the introduction of the Turing Machine. Here a problem is a set of related questions. For example, "given an even number, find pairs of prime numbers that sum up to it" is a problem and "find pairs of prime numbers that sum up to 20" is an instance of this problem. Computational complexity describes how the time needed to solve problem instances by the best algorithm scales with the size of problem input.

Problems are categorized into classes by their worst-case complexity. Roughly speaking, a problem is in complexity class P if there exists an efficient program that solves all its instances in polynomial time. The class NP, here NP stands for non-deterministic polynomial, consists of problems for which a candidate solution can be checked in polynomial time. Problems posed as "does there exist an XYZ such that condition ABC is satisfied?" are in NP if conditions ABC can be verified in polynomial time. Problems complement to those in NP are in co-NP. For those problems, a *no* answer can be proved by a counterexample.

The complexity we focus on will be proof complexity, i.e., proof length. For different questions, proofs will often be of different natures. For example, a Lyapunov function is a proof of system stability in dynamical system analysis, whereas a factorization of the integer 15 into factors 3 and 5 is a proof that 15 is not a prime number. A consequence of this is that there is no universal definition of a proof length. The

appropriate meaning of proof complexity will be clear from the system of interest and the form of the proof. In the case of finding polynomial Lyapunov functions, the order of the polynomial is the proof complexity. We know that for stable linear systems, a quadratic Lyapunov function always exists. Therefore, in this sense, proving linear stability is easy.

Proof complexity is our main focus because it is more revealing for engineering design. As we show in the following two problems, percolation lattice and linear programs, a long proof indicates fragility in the system and provides feedback to the robust design process depicted in Figure 1.1.

2.2 Percolation Lattices

A percolation process is a mathematical model of the random spread of a fluid through a medium, where both "fluid" and "medium" can be very broadly interpreted [35]. It was first brought into mathematics by Broadbent and Hammersley in 1957 [11]. In two dimensions, a typical percolation model involves the plane square point lattice $\mathbb{Z}^2$ and a real number p, $0 \leq p \leq 1$. There are two kinds of percolation problems. In site percolation, each site on the lattice, independent of all other sites, is occupied with probability p and vacant with probability $1 - p$, the vacant sites being thought of as junctions that are blocked to the passage of fluid. Thus fluid can only pass through occupied sites. In bond percolation, it is each bond that has a probability p of being occupied and $1 - p$ of being vacant. In fact, every bond percolation may be transformed into a site percolation problem. In this sense, site percolation models are more general. We focus on the two-dimension site percolation model on a square lattice.

Percolation is well-studied in statistical physics, as it is one of the simplest problems that exhibit phase transition. Physicists explore macroscopic properties of lattices as functions of the order parameter p. One popular such property is the existence

of an infinite path of occupied sites in different lattice instances and the critical phenomenon related to it. Phase transition happens at critical probability p_c [81]: when $p < p_c$, infinite horizontal path almost never occurs in a random lattice; but when $p > p_c$, it is almost always observed. In physics, this sudden change is used to model sudden change from one physical phase to another, as when steam turns into water at the condensation temperature.

We are interested in particular lattice instances rather than random lattice realizations as an ensemble. In this case, p does not provide as much information as the fraction of occupied sites in a lattice. This density is denoted by ρ. From this point onward, by a lattice, we mean a site percolation model with a designated occupation configuration, i.e., its occupied and vacant sites are already determined. Figure 2.1 shows a 20×20 lattice, where the occupied sites are colored as black and vacant sites are left as white. In what follows, we sometimes refer to occupied sites as black sites and vacant sites as white sites.

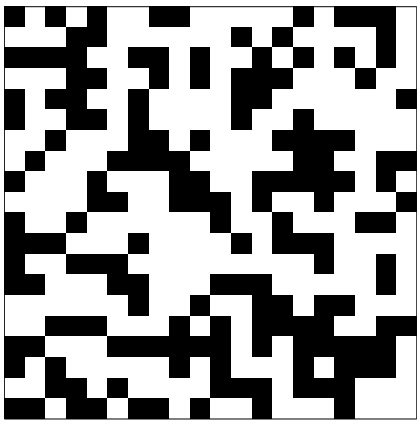
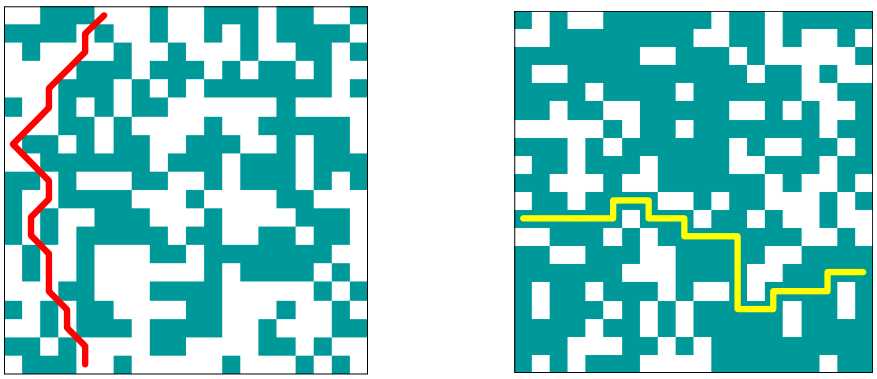


Figure 2.1: A 20×20 site percolation lattice

Two-dimensional percolation lattices can be used as the model for various physical, engineering and social systems. Examples include, among many others, forest fires [12], electrical conductors and spread of epidemic diseases [81]. In the case of forest

fire, we can think of each occupied site of a lattice as a tree planted and vacant sites as non-planted land. Once a tree is on fire, it can ignite its neighbors. Suppose trees on one boundary are ignited at the same time. Then for very small ρ, i.e., trees are very sparsely planted, the fire will likely die out after several steps and the propagation of fire is stopped by a barrier of unplanted land. But for large ρ, i.e., tress are densely planted, it can propagate to the opposite boundary and percolation of fire happens.



(a) Vertical path (b) Horizontal path

Figure 2.2: Vertical and horizontal paths represent cascading failure events and barriers that block these events, respectively

Cascading failure events in dynamical systems can sometimes lead to catastrophical system failure. Therefore a great deal of resources in biological and engineering systems are devoted to building barriers that prevent such cascading from happening or propagating. Percolation lattices can be borrowed as a simplified model to illustrate some of the essence of the challenges involved in the analysis of these systems. The sites represent states of the systems: occupied sites are the states that the system is impossible to be in and vacant sites are the feasible states. Transition between two feasible states can happen only when the two corresponding sites in the lattice are neighbors, i.e., they share an edge or a corner. For simplicity, suppose the first row of the lattice represents initial states of the system and the bottom row of the lattice are the states that cause failure of the system. Then the existence of a series of cascading events leading to system failure is represented by the existence

of a vertical path composed of feasible states. In the system modeled by the lattice in 2.2(a), failure can happen through the vertical path colored in red. Building a barrier to prevent failure is then essentially the same as using occupied sites to form horizontal paths. In 2.2(b), the horizontal path plotted in yellow is a barrier that guarantees that no vertical path exists. Existence of such a barrier means there is no dynamically feasible series of states that leads the initial normal states to a failure.

To make it clear what we mean by neighbors and paths, we describe the problem more formally.

2.2.1 Problem Description

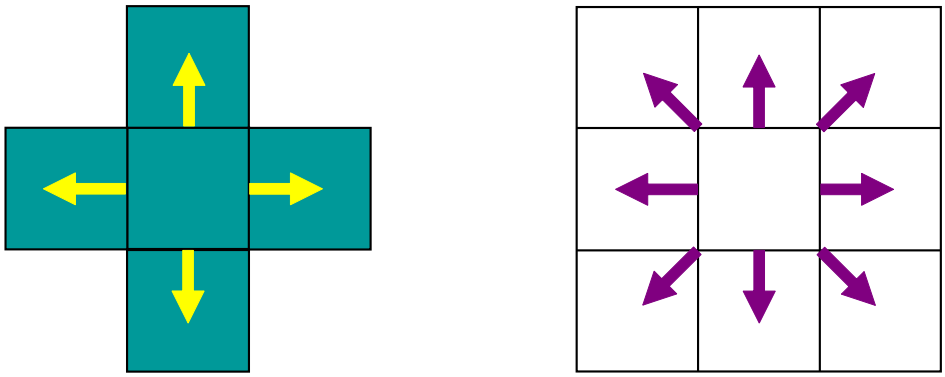
First, some definitions similar to those in [89] are introduced. We place an $N \times N$ lattice on a grid such that each site v is indexed by the two integer coordinates of its center $v = (x_v, y_v) \in \mathbb{Z}^2$, $1 \leq x_v, y_v \leq N$.

Definition 2.1. *Two occupied sites v and w are neighbors if $\|v - w\|_2 = 1$. Two vacant sites v and w are neighbors if $\|v - w\|_2 \leq \sqrt{2}$.*

That is, the neighbors of an occupied site are occupied sites one step horizontal or vertical away from it. An occupied site has at most 4 neighbors. Vacant sites can have neighbors that are one step away from it in a diagonal direction and can have as many as 8 neighbors, as shown in Figure 2.3. Notice that an occupied site can never be a neighbor of a vacant site.

Definition 2.2. *A path in a square percolation lattice is a sequence of sites $v_1, \cdots, v_m$ such that for all $i = 1, \cdots, m-1$, v_i and v_{i+1} are neighbors.*

Thus an occupied path is a sequence of occupied sites $v_1, \cdots, v_m$ such that for all $i = 1, \cdots, m-1$, $\|v_i - v_{i+1}\|_2 = 1$. Similarly, a vacant path is a sequence of vacant sites $v_1, \cdots, v_m$ such that for all $i = 1, \cdots, m-1$, $\|v_i - v_{i+1}\|_2 \leq \sqrt{2}$. We are interested in occupied paths that span horizontally from one end of the lattice



(a) Occupied (b) Vacant

Figure 2.3: Neighborhood rules for occupied and vacant sites

to another and close paths that cross the lattice vertically. Such paths are called crossings.

Definition 2.3. *For a $N \times N$ square percolation lattice that contains all the sites v with coordinate $\{(v_x, v_y) \in \mathbb{Z}^2 : 1 \leq v_x, v_y \leq N\}$, a horizontal cross is a path from site v to w such that $v_x = 1$ and $w_x = N$. A vertical crossing is defined similarly.*

We are only interested in crossings. So hereafter, without ambiguity, an occupied horizontal crossing is referred to as a horizontal path or a black path, corresponding to our coloring of the occupied sites. Similarly, a vacant vertical crossing is called a vertical path or a white path.

Clearly, the absence of vertical path can be shown to be equivalent to the existence of a horizontal path[1]. Thus verifying that there is no white path leading to failure is converted to showing the existence of a horizontal path. Such a horizontal path is usually termed as safety proof in computer science.

[1]This conclusion is a direct application of Proposition 2.2 in [40], where the occupied and vacant paths are paths on a pair of matching graphs.

2.2.2 Robustness and Complexity in Percolation Lattices

Following the previous example of cascading failure, white paths will represent disaster events and black paths will be the designed barrier to block these disasters. Even though mechanisms have been designed to prevent disasters from happening, failures still occur sometimes due to changes in environment or malfunction of these mechanisms. In the lattice setting, these changes or malfunction are associated with perturbations to the lattice by changing the coloring of some lattice sites. Under such perturbations, it is possible that a vertical white path appears in the resulting lattice though no white path was present in the original lattice. Depending on their structures, the minimum number of color changes needed to convert a lattice from having no vertical path to having at least one such path varies greatly even among lattices of the same black cell density. This amount captures how robust the system functionality is towards perturbations modeled by changing in coloring of lattice sites. Suppose the lattice is of size $N \times N$. Let ρ be the density of black sites, i.e., the proportion of black sites. Robustness is defined as follows.

Definition 2.4. *For an $N \times N$ square lattice without vertical path, the robustness of the lattice is*

$$R = B/(\rho N), \tag{2.1}$$

where B is the minimum number of color changes in lattice sites to introduce a vertical path. The fragility is the reciprocal of robustness, $F = 1/R$.

If we view black sites as resources to build barrier from, then the normalization by ρN implies that R is the relative robustness given available resources. Note that if we call paths that do not share common sites as independent paths, then B is the number of independent black paths. It is obvious that the maximum robustness is $R = 1$. This is achieved by devoting all the available resources to building a thick purely horizontal black strip, as the upper left two lattices shown in Figure 2.4. Obviously, for a fixed ρ, it is harder for the system to experience failure when it is more robust.

Another important concept is the verification complexity, i.e., proof length. For systems that function without failure when unperturbed, this fact can be verified through a safety proof, which is a black path in the lattice setting. To show a path exists, there is no simpler way other than to trace out this path and mark all the sites in it. It is then natural to define the verification complexity C as the length l of the shortest horizontal black path, normalized by the size of the lattice N.

Definition 2.5. *For a $N \times N$ square site percolation lattice without vertical path, the complexity for verifying the absence of vertical paths is*

$$C = l/N, \tag{2.2}$$

where l is the length of the shortest horizontal path.

It is obvious that $1 \leq l \leq N^2$ and a tighter upper for l approaches $N^2/2$ when N increases. The smallest complexity occurs when there exists a purely horizontal path consisting occupied sites in the same row. It is intuitive that a shorter path is a less complex verification of the system functionality.

Engineering designs have been aiming at finding systems with shortest proofs. It turns out that this objective does not conflict with designing robustly functional systems, as shown by the following theorem.

Theorem 2.1. *For a given two-dimensional square lattice of size $N \times N$ without vertical paths, let its verification complexity C and robustness R as defined earlier. Then,*

$$RC = \frac{Bl}{\rho N^2} \leq 1. \tag{2.3}$$

Proof. The proof is straightforward from definition, noting that B is the number of independent horizontal paths and ρN^2 is the total number of black sites in the lattice, thus $Bl \leq \rho N^2$. □

This means robustness and proof complexity are to some extent aligned. When the system has high robustness, the proof complexity must be small. Therefore, it is guaranteed that robust systems can be easily verified with a low-complexity proof, as shown in Figure 2.4. On the other hand, when complexity is high, the system is doomed to have low robustness, also illustrated in Figure 2.4. If we can not verify with a simple proof that failures never happen, there is a reason to believe that the system itself is fragile.

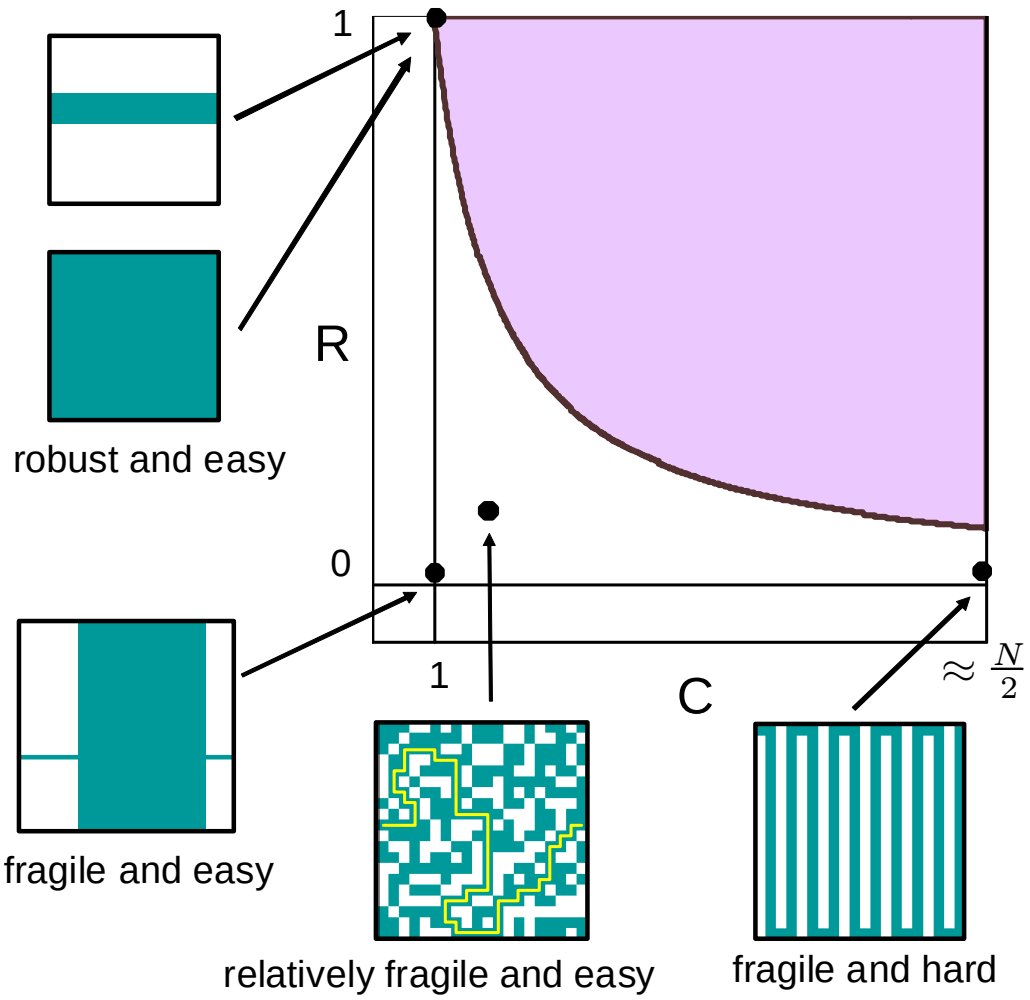


Figure 2.4: The robustness–complexity picture for percolation lattices. Robust lattices are at the upper-left corner of the admissible region, while as hard lattices are necessarily located in the lower-right corner. Random lattices are relatively easy. Fragile lattices can be easy as well.

Various lattice examples with R and C in the admissible region are shown in Figure 2.4. The colored region is excluded due to the $R \cdot C \leq 1$ relation. Four insights can be obtained: (i) Biologically evolved and carefully designed engineering systems are located in the upper-left corner of the region, possessing high robustness and short verification proofs. (ii) The hardest lattices are located in the lower-right corner of the region, being complex with only long verification proofs. (Notice that these instances are almost never observed in randomly generated lattices.) (iii) Fragile lattices can be located in the lower-left corner of the region. This means short verification proofs do not imply robustness. Fragile systems can be easy to verify as well. (iv) Random lattices sit in the middle of the region and are not the hardest instances.

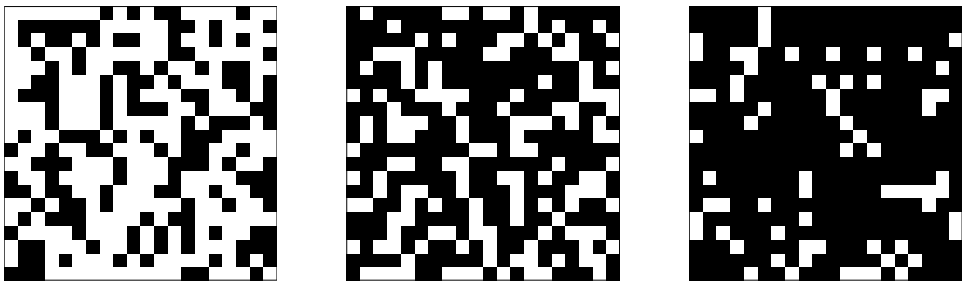
So far, we have studied robustness and complexity for lattice models where vertical paths represent failure events and the aim of robust design is to build horizontal path as thick as possible to block failure from happening. In the cases where events modeled by existence of horizontal paths are to be avoided, such as the forest fire example in the beginning of the chapter, the presence of vertical paths is desirable. Robustness and complexity is then defined accordingly to reflect the maximum number of independent vertical paths and length of the shortest vertical path. It is an simple exercise to show that Theorem 2.1 holds in these situations as well.

2.2.3 Structured and Random Lattices

A random lattice is a realization of the percolation problem introduced at the beginning of Section 2.2.1, where sites are colored randomly according to certain probability p. The density of black sites in the lattice is denoted as ρ as before. In random lattices, p is the expected value of ρ, and ρ approaches p as the size of the lattice increases.

For random lattices, there are three situations regarding the hardness of finding a barrier, as shown in Figure 2.5. When the density of black sites is very low, almost all the random lattices do not have a horizontal black path and verifying this is always easy. Whereas when the density approaches 1, there is almost always a black path

and such a path can usually be found without difficulty. The transition from almost never observing a black path to almost always having a barrier happens the critical density around $\rho = 0.593$ [81, 86]. For lattices around this density, a horizontal path just appears or is just about to appear. It is in general hard to decide whether a barrier exists and, when it does, it tends to be zigzagged and long.

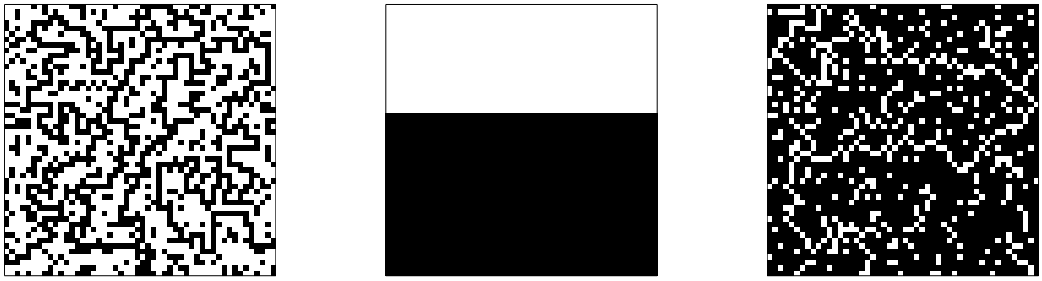


(a) $\rho = 0.4$ (b) $\rho = 0.59$ (c) $\rho = 0.8$

Figure 2.5: Examples of random lattices with different density of black sites. Verification of horizontal barriers in random percolation lattices are the hardest on average when the lattices are placed near the critical density. Phenomenon like this has convinced statistic physicists that the verification complexity is directly linked to criticality.

This phase transition phenomena have been intensively studied in the statistical physics literature. The easy-hard-easy transition around the critical density has lead physicists to believe that the hardness in verifying the existence of a black path is directly associated with phase transition. Consequently, huge amount of effort has been devoted to lattices around phase transition and the systems modeled by these lattices, as they are considered the complex and interesting ones. Failure is considered obvious in systems sufficiently below phase transition and verification of robustness is trivial for systems far above it.

Association of verification complexity with criticality is true in the probability sense for random lattices. However, structured lattices do not observe this relation, as shown by the examples in Figure 2.6. In Figure 2.6(a), a very long horizontal path appears in the region that is below phase transition and is considered easy in the random view. The lattice in Figure 2.6(c) does not have a horizontal barrier.



(a) $\rho = 0.4$ (b) $\rho = 0.59$ (c) $\rho = 0.8$

Figure 2.6: Examples of structured lattices with different density of black sites. Non-random lattices are observed with long horizontal barriers at low density. They also can have short horizontal paths at critical density, refuting the correlation between verification complexity and criticality.

Although it is above phase transition at density $\rho = 0.8$, the absence of a black path is not easy to see. In contrast, at the critical density, the lattice in Figure 2.6(b) is robust and very easy.

Engineering and biological systems are purposeful processes that are very different from being purely random. They are highly optimized to achieve robust resistance to perturbations with limited resources. Lattices that are models of these designed systems have rigorous structure. One such structure is to have black cells carefully positioned to form a solid black strip, despite the low density of black cells. Although the set of these lattices is of measure zero in the space of all lattices and is almost never encountered by the random lattice generating process, they can be built by deterministically choosing the coloring. Structured lattices are more suitable models for engineering and biological systems, though they are mostly neglected from the statistical physics perspective.

Random lattices at phase transition are not always hard. Even when they are, they appear to be easy compared with structured hard lattices, as illustrated in Figure 2.7. In Figure 2.8 we plot random lattices at phase transition on the admissible region in C-R plane. We see that random lattices only occupy a very small area of the

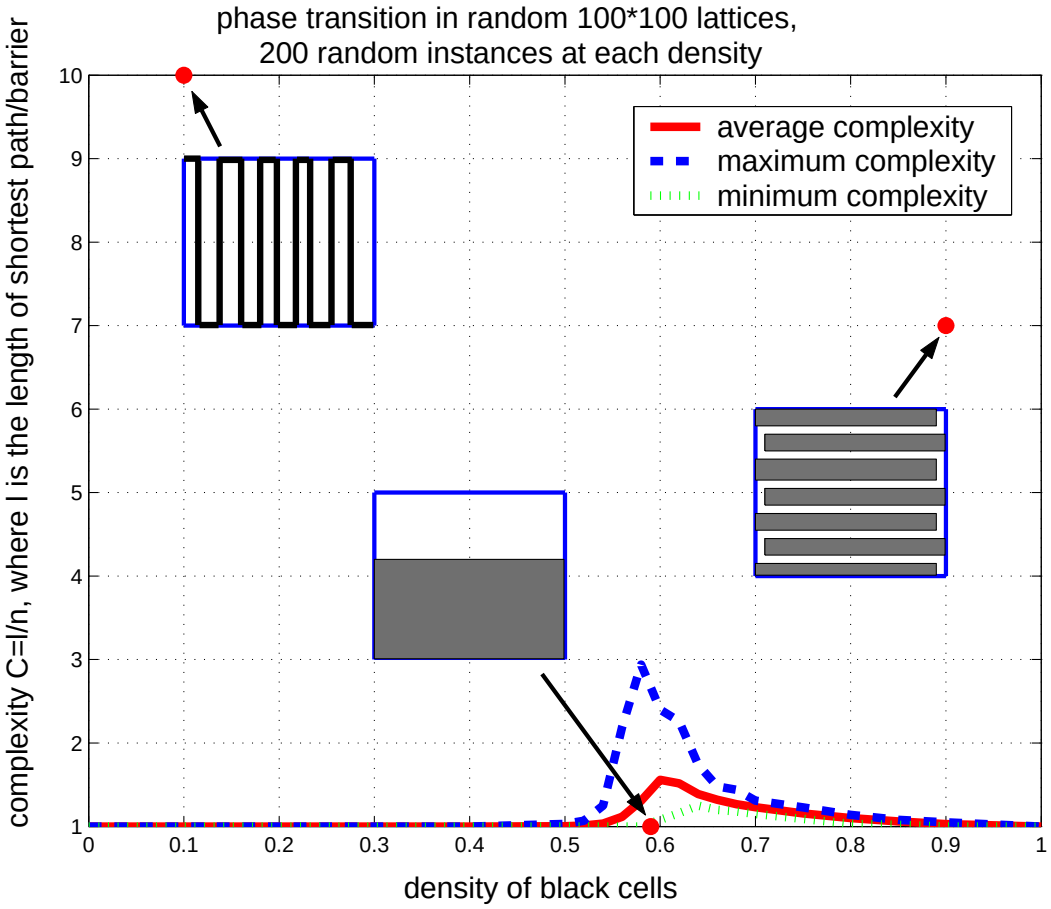


Figure 2.7: Complexity of random percolation lattice instances at different densities. The average complexity, as well as the maximum complexity, of the 200 random lattices peaks at near the critical density $\rho = 0.593$. However, the most complex random lattice is easy when compared with some designed lattices even at noncritical densities. Thus the relation between complexity and criticality is not revealing when designed lattices are under study.

region. The rest of the region is hardly visited by random lattices; however lattices in this unvisited region are often models of structured systems. Therefore, random lattices should not be the focus of engineers. The relationship $C \cdot R \leq 1$ is much more fundamental and revealing than the popular phase transition view when designed systems are of interest.

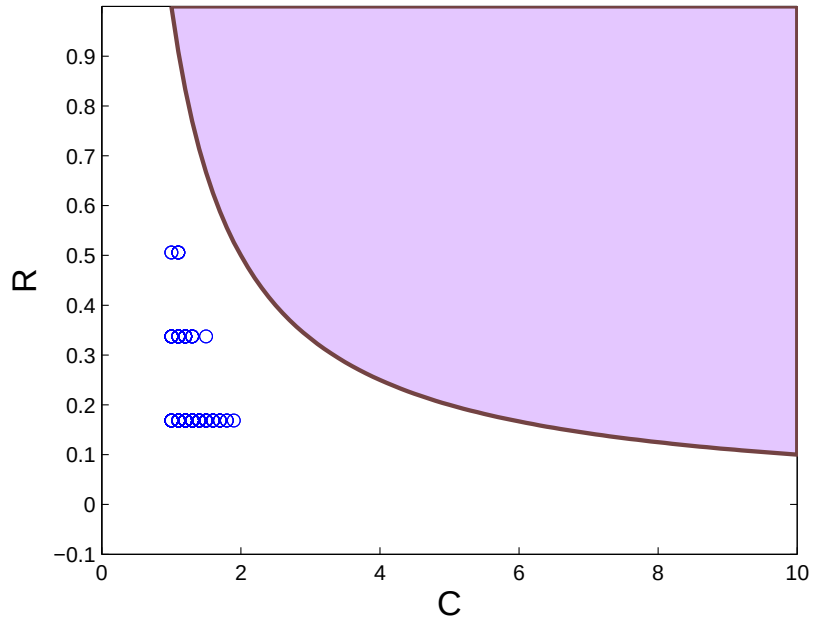


Figure 2.8: The robustness–complexity picture of random lattice at phase transition. 1000 lattices of size 10×10 with density $\rho = 0.593$ are plotted as blue circles.

2.2.4 Percolation Lattice as Linear Programming

The lattice percolation problem can be stated as a feasibility linear programming (LP) problem: a horizontal path exists if, and only if, some linear equalities and inequalities are consistent. Thus, it is natural to ask if the relation between robustness and complexity given in the previous section will hold for any *general* linear program. The answer is positive. Before this is shown in section 2.3 using results on ill-conditioning in linear programming from the numerical analysis literature, the LP formulation of

percolation lattices is derived.

Consider a $N \times N$ lattice. We can assign a vertex to each site and two directed edges in opposite directions to each pair of black neighbor sites. That is, a lattice can be represented by a directed graph $G = (V, T)$, where V is the set of all the sites in the lattice and T is the set of all the connections among the sites allowed by the neighbor rules and the specific coloring, i.e., the data. Then finding a horizontal path in a lattice is equivalent to the problem of finding a path between two subsets of the vertices, one corresponding to one boundary of the lattice and another to the opposite boundary. These two sets of vertices are grouped as V_1 and V_2 in Figure 2.9. For simplification, two auxiliary vertices, along with $2N$ edges are added: a source at left s_l having edges incident to nodes in V_1, and a terminal at right t_r with edges pointing to it from nodes in V_2. The shortest-length path from s_l to t_r corresponds to the shortest horizontal barrier in the lattice.

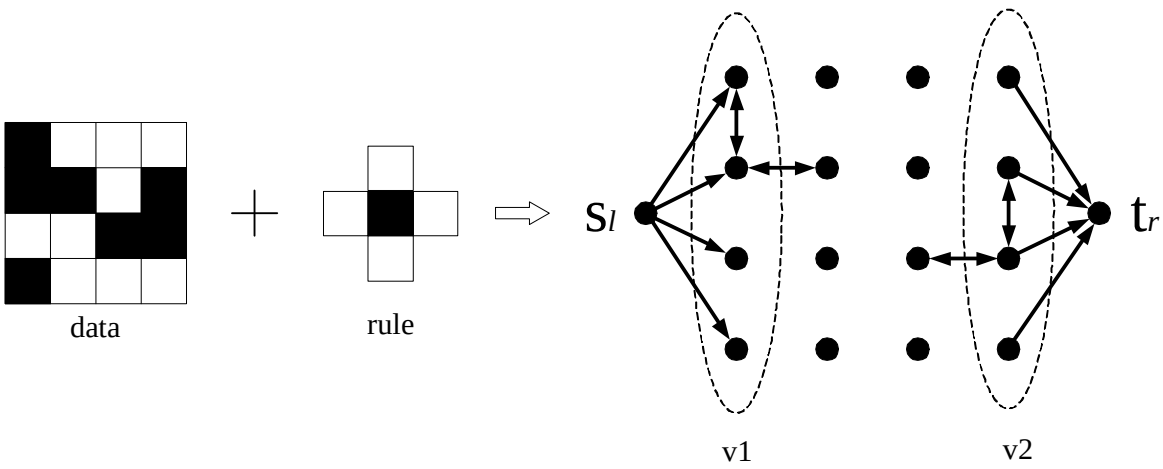


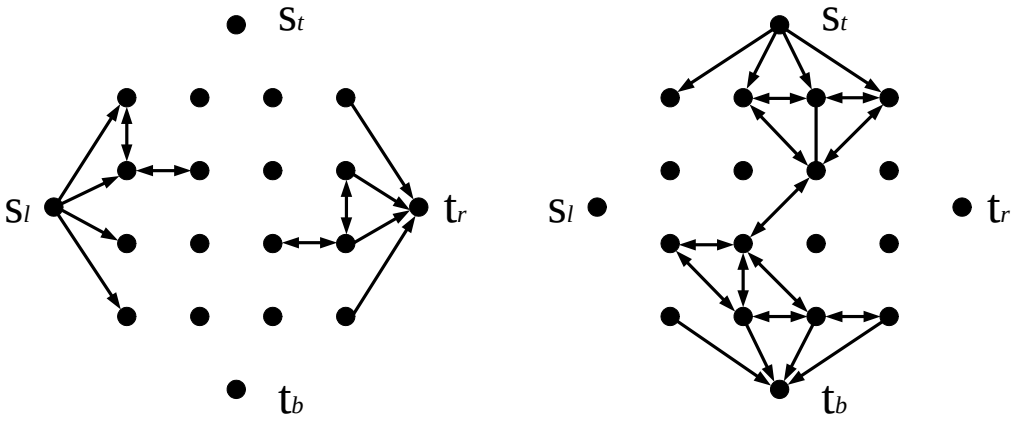
Figure 2.9: The modeling of lattice problems by graphs for an example lattice. An edge with arrows in both directions represents two directed edges in opposite directions. Finding a horizontal path is reduced to finding a path from s_l to t_r in the graph at right.

The path problem can be written as a flow problem. Assign a flow variable f_{ij} to the edge from vertex i to vertex j. All we need to check is whether there exists a legitimate flow from s_l to t_r. A legitimate flow is such that for any node i other

than s_l and t_r, $\sum_j f_{ij} = 0$ where j runs through the neighbors of i. This is a linear constraint, therefore the horizontal path problem is written in the following LP in feasibility form

$$\begin{cases} Af = b \\ Ef \leq 0 \\ f \geq 0, \end{cases}$$

where $A \in \mathbb{R}^{N^2+2\times 2N(2N-1)}$ is the incident matrix of all the edges allowed by the neighbor rule for black sites if all the sites were black; $Ef \leq 0$ together with $f \geq 0$ sets to zero the flows that are blocked by the coloring. b is a vector with -1 for s, 1 for t, and zeros for all other sites.



(a) horizontal (b) vertical

Figure 2.10: The modeling of lattice problems by graphs for the example lattice in Figure 2.9

Figure 2.10 shows the graph models for finding horizontal black path and vertical white path for the lattice in Figure 2.9. Figure 2.10(a) is formed from the four-neighbor rule through occupied (black) sites and is termed the primal graph. Figure 2.10(b) is obtained from the eight-neighbor rule through vacant (white) sites and is called the dual graph. To make the two graphs, and in turn the two LPs, have similar form, two more vertices are added to the graphs — the source on top s_t and the

terminal on bottom t_b.

Determining paths in graphs can be written in LPs in feasibility form. The LP for the primal graph in Figure 2.10(a) is called PH, standing for the primal LP of the horizontal path problem. The Lagrange alternative of PH is written as DH. Similarly, PV and DV come from the graph in Figure 2.10(b).

$$PH \begin{bmatrix} A_h f_h & = & b_h \\ E_h f_h & \leq & 0 \\ f_h & \geq & 0 \end{bmatrix} \qquad PV \begin{bmatrix} A_v f_v & = & b_v \\ E_v f_v & \leq & 0 \\ f_v & \geq & 0 \end{bmatrix}$$

$$DH \begin{bmatrix} A_h^T \lambda_h + E_h^T \nu_h & \geq & 0 \\ b_h^T \lambda_h & = & -1 \\ \nu_h & \geq & 0 \end{bmatrix} \qquad DV \begin{bmatrix} A_v^T \lambda_v + E_v^T \nu_v & \geq & 0 \\ b_v^T \lambda_v & = & -1 \\ \nu_v & \geq & 0 \end{bmatrix}$$

A_h and A_v are the incident matrices determined by the 4-neighbor rule and the eight-neighbor rule, respectively. E_h and E_v reflect the data in the primal and dual lattices respectively. b_h is a vector with -1 for the left source, 1 for the right terminal and 0 for all the other vertices. Similarly, b_v is a vector with -1 for the top source, 1 for the bottom terminal and 0 for all the other vertices. It is obvious that PV is feasible if and only if DH has a solution.

2.3 Linear Programming

Consider the linear programming feasibility problem

$$\mathbf{Fd}(A, b): \quad \begin{array}{l} \text{find } x \in \mathbb{R}^n \text{ such that} \\ Ax \leq b,\ x \geq 0, A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m. \end{array} \tag{2.4}$$

Many engineering or biological systems can be described using this linear programming feasibility problem. A set of design parameters that stabilize the system is the feasible set of stability constraints and physical limitations posed on the system.

The ranges of concentration of critical chemicals that guarantee the functionality of certain biological systems are determined by algebraic inequalities among reaction speed parameter, concentration of reactors and catalysts. Change in emptiness of the feasible set at the presence of small changes in data is undesired in system design or configuration.

2.3.1 Notation

The norm in $\mathbb{R}^n$ and $\mathbb{R}^m$ is the Euclidean 2-norm. We identify the space of linear operators, $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ with the Euclidean space $\mathbb{R}^{m\times n}$. Endow $\mathbb{R}^{m\times n}$ with the operator norm, i.e., if $A \in \mathbb{R}^{m\times n}$, then $\|A\| := \sup\{\|Ax\| : x \in \mathbb{R}^n, \|x\| = 1\} = \bar{\sigma}(A)$, where $\bar{\sigma}(A)$ is the singular value of A. Note that for a linear functional defined on R^n, its induced 2-norm reduces to the usual 2-norm. So, hereafter, the same notation $\|\cdot\|$ is used for all the norms, with the norm it refers to clear from the context. Consider the parameter space of $\mathbb{D} = \mathbb{R}^{m\times n} \times \mathbb{R}^n$ of all possible instances $d = (A, b)$. Following Renegar in [74], define the norm in $\mathbb{D}$ as $\|d\| = \max\{\|A\|, \|b\|\}$.

Denote the set of consistent instances of Fd in $\mathbb{D}$ as

$$\mathcal{F} = \{(A, b) \in \mathbb{D} : \exists\ x \text{ satisfying } Ax \leq b, x \geq 0\}.$$

The complement of $\mathcal{F}$ is denoted by $\mathcal{F}^C$. $\mathcal{F}^C$ consists of (A, b) pairs that render Fd infeasible. The boundary between $\mathcal{F}$ and $\mathcal{F}^C$ is denoted by

$$\mathcal{B} = \partial\mathcal{F} = \partial\mathcal{F}^C = cl(\mathcal{F}) \cap cl(\mathcal{F}^C),$$

where $\partial\mathcal{F}$ and $cl(\mathcal{F})$ represent the boundary and the closure of the set $\mathcal{F}$, respectively. The boundary set $\mathcal{B}$ is the set of Fd instances such that arbitrary small changes in the data can render them feasible or infeasible. Problems on the boundary are called ill-posed since they are not robust to even the smallest change in data. Problems

very close to the boundary are ill-conditioned. The further away a problem is from the boundary, the bigger change the described system is able to endure before its functionalities collapse. Normalization is important in comparing the distances to boundary between systems, since two data sets (A, b) and (kA, kb) capture the same feasible set whenever k is a positive constant.

By the theory of alternatives [10], there exists a linear programming feasibility problem

$$\mathbf{A1}(A,b): \quad \begin{array}{l} \text{find } y \in \mathbb{R}^m \text{ such that} \\ A^T y \geq 0,\ b^T y < 0, y \geq 0, \\ A \in \mathbb{R}^{m\times n},\ b \in \mathbb{R}^m \end{array} \tag{2.5}$$

such that one and only one of Fd (A, b) and A1 (A, b) is feasible given (A, b). This relation between Fd and $A1$ is a variant of the famous Farkas Lemma [64]. Alternatives like Fd and $A1$ are called strong alternatives. Consider another linear programming feasibility problem

$$\mathbf{A2}(A,b): \quad \begin{array}{l} \text{find } y \in \mathbb{R}^m \text{ such that} \\ A^T y \geq \alpha,\ b^T y \leq -\beta, y \geq 0, \\ A \in \mathbb{R}^{m\times n},\ b \in \mathbb{R}^m, \alpha \in \mathbb{R}^n, \alpha > 0, \beta > 0 \end{array} \tag{2.6}$$

A2 and Fd are weak alternatives to each other, for at most one of them can be feasible given (A, b) and it is possible that both of them are infeasible.

Rewrite A2 as

$$\mathbf{A2}(A,b): \quad \begin{array}{rcl} \bar{A}y & \leq & \bar{b} \\ y & \geq & 0 \end{array} \tag{2.7}$$

where

$$\bar{A} = \begin{bmatrix} -A^T \\ b^T \end{bmatrix}, \quad \bar{b} = \begin{bmatrix} -\alpha \\ -\beta \end{bmatrix},$$

and A, b, α, β are as in (2.6). A^T is the adjoin operator of A and b^T is a functional on $\mathbb{R}^n$. Note that $\|A^T\| = \|A\|$, $\|b^T\| = \|b\|$. $\bar{A} \in \mathbb{R}^{(n+1)\times m}$ is a function from $\mathbb{R}^m$ to $\mathbb{R}^{n+1}$ and its norm is defined as operator norm as well. Note that positive scaling of

$\bar{b}$ does not alter the feasibility of $A2$. Without loss of generality, α and β are picked such that

$$\| \begin{bmatrix} -\alpha \\ -\beta \end{bmatrix} \| = \|b\| \tag{2.8}$$

in order for the instance to only depend on the original data $d = (A, b)$.

2.3.2 Robustness and Complexity in LP

Suppose the problem $Fd(A, b)$ is feasible. The *robustness* with respect to errors in the data $d = (A, b)$ can be measured by the smallest relative perturbation in the problem data that renders the problem infeasible. This quantity is interpreted as the distance of the given (feasible or infeasible) instance to the boundary $\mathcal{B}$. Let

$$dist(d, \mathcal{B}) = \inf\{\|d - \tilde{d}\| : \tilde{d} \in \mathcal{B}\}. \tag{2.9}$$

The robustness of an Fd instance with data d is defined as

$$R = \frac{dist(d, \mathcal{B})}{\|d\|}, \tag{2.10}$$

i.e., the distance to ill-posedness normalized by problem size. It is an easy exercise to show that reformulating the system simply by positive scaling does not affect the robustness of an Fd instance.

The *fragility* for this problem is $F = 1/R$; it is exactly the condition measure defined by Renegar [74, 76]. Its properties have been studied in [22, 23, 30]. This is an extension of the usual condition number of a matrix, associated with a system of linear equations. It is known that the computational complexity of several LP algorithms as measured by the number of iterations required for desired solution accuracy is essentially bounded by a logarithmic function of fragility . For example, see [75] for the analysis of an interior point method, and [30] for the ellipsoid method.

Thus, generally speaking, these results indicate that if a large number of iterations are required to find a feasible point, the problem must be ill-conditioned i.e., highly fragile.

A solution x to the inequalities $\{Ax \leq b, x \geq 0\}$ is enough to verify that the LP instance $Fd(A, b)$ is consistent. The size of this solution serves as a measurement of the *verification complexity* of the LP instance. We define the primal verification complexity.

Definition 2.6. *Let the feasibility LP problem $Fd(A, b)$ be given as in (2.4). The primal complexity is the minimum norm of a solution to Fd*

$$C_p = \inf\{\|x\| : \ Ax \leq b, x \geq 0\}. \tag{2.11}$$

We follow the convention and define the infimum of an empty set to be infinity. That is, $C_p = \infty$ when $Fd(A, b)$ is inconsistent. Thus, the feasible point with the smallest norm is taken to be the shortest proof that a solution exists. The relation $R \cdot C \leq 1$ that we advocated in the percolation lattice problem can be shown to hold in this setting using results already known in numerical analysis. Specifically, we use Theorem 1.1(1) in [74] by Renegar. For completeness, a slightly modified version of the theorem fitting our setup and notation is quoted here.

Theorem 2.2 (Renegar)**.** *Assume $d = (A, b) \in \mathbb{D}$. If d is feasible and satisfies $dist(d, \mathcal{B}) > 0$, then there exists a feasible solution x to $Fd\,(A, b)$ such that*

$$\|x\| \leq \frac{\|b\|}{dist(d, \mathcal{B})}. \tag{2.12}$$

When $Fd(A, b)$ is well-posed and consistent, application of Theorem 2.12 yields the same relation as in the percolation lattices: $R \cdot C_p \leq 1$.

Now suppose d is such that $Fd(A, b)$ is infeasible. Then $d \in \mathcal{F}^C$. The robustness with respect to errors in data is still defined the same as in the case when $Fd(A, b)$

is feasible, i.e., $R = dist(d, \mathcal{B})/\|d\|$. If the solutions to the set of inequalities in Fd represent "disaster situations," certainly not only is $Fd(A, b)$ required to be inconsistent in ideal circumstances, but it is desired to be inconsistent when perturbations in (A, b) are present. In this setup, it is preferable that (A, b) is in the interior of $\mathcal{F}^C$ and sits far away from the boundary $\mathcal{B}$.

When $Fd(A, b)$ is infeasible, there must exist at least one solution y to $A1(A, b)$ that can serve as a verification proof to the infeasibility of $Fd(A, b)$. A solution to $A2(A, b)$ also qualifies as a proof. However, there are (A, b) pairs such that $A1(A, b)$ is consistent while $A2(A, b)$ is not. What is interesting is that all the (A, b) instances observing the gap between $A1$ and $A2$ lie on $\mathcal{B}$. In other words, when the infeasibility of $Fd(A, b)$ cannot be verified by providing a solution to $A2(A, b)$, we can actually conclude that the instance is on the boundary, i.e., it is ill-posed.

Lemma 2.1. *If (A, b) is such that $A1(A, b)$ is consistent and $A2(A, b)$ is inconsistent, then for every $\epsilon > 0$, there exists ΔA with $\|\Delta A\| \leq \epsilon$ such that $A1(A + \Delta A, b)$ is infeasible.*

Proof. Let $y_0 \in \{y : b^T y < 0, y \geq 0\}$. It is obvious that $y_0 \neq 0$. Let $A = [a_1 \quad a_2 \quad \cdots \quad a_n]$, i.e., a_i is the ith column of A. Since $A2(A, b)$ is inconsistent, there exists index i, $1 \leq i \leq n$, such that $a_i^T y_0 \leq 0$ for otherwise $A^T y_0 > 0$ and $A1(A, b)$ is feasible.

Let $\Delta A = [\Delta a_1 \quad \Delta a_2 \quad \cdots \quad \Delta a_n] = -\delta \mathbf{1}_{m \times n}$, where $\delta > 0$ and $\mathbf{1}_{m \times n}$ is the $m \times n$ matrix with all elements being 1. Choose δ such that $\|\Delta A\| \leq \epsilon$. Since $y_0 \geq 0, y_0 \neq 0$, $\Delta A y_0 < 0$. Now $(a_i + \Delta a_i)^T y_0 < 0$. Since y_0 is arbitrarily chosen from the set $\{y : b^T y < 0, y \geq 0\}$, the emptiness of $\{(A + \Delta A)^T y \geq 0, b^T y < 0, y \geq 0\}$ follows. □

Lemma 2.1 tells us that we can look at $A2(A, b)$ for verifying inconsistency of $Fd(A, b)$. If a proof does not exist, we know that the problem is fragile to even the smallest perturbation. We define the alternative verification complexity.

Definition 2.7. *Let the feasibility LP problem $Fd(A, b)$ be given as in (2.4). The alternative complexity is the minimum norm of a solution to $A2$.*

$$C_a = \inf\{\|y\| : \bar{A}y \leq \bar{b}, y \geq 0, \|\bar{b}\| = \|b\|\}. \tag{2.13}$$

Again the infimum of an empty set is taken to be infinity. That is, if $A2$ is inconsistent, $C_a = \infty$.

When d is away from the boundary and $A2$ is consistent, applying Theorem 2.12 with some manipulation gives the relation $R \cdot C_a \leq 1$. We have the following lemma.

Lemma 2.2. *When $A2(A, b)$ is consistent and d is away from boundary,*

$$R \cdot C_a \leq 1.^2$$

Consider a well-posed LP instance $Fd(A, b)$, i.e., $d \notin \mathcal{B}$. When it's feasible, a solution x to Fd serves as a proof; otherwise, a solution to $A2$ proves infeasibility. Thus, the verification complexity is defined as follows.

Definition 2.8. *For an LP instance $Fd(A, b)$, its verification complexity is*

$$C = \min\{C_p, C_a\}. \tag{2.14}$$

When $Fd(A, b)$ and $A2(A, b)$ are both inconsistent, the verification complexity of $Fd(A, b)$ is ∞. As Lemma 2.1 shows, these instances are the ill-posed cases. The main result is as follows.

Theorem 2.3. *Let the feasibility LP instance $Fd(A, b)$ as defined in (2.4) be well-posed. Then*

$$R \cdot C \leq 1. \tag{2.15}$$

[2]See Appendix for proof.

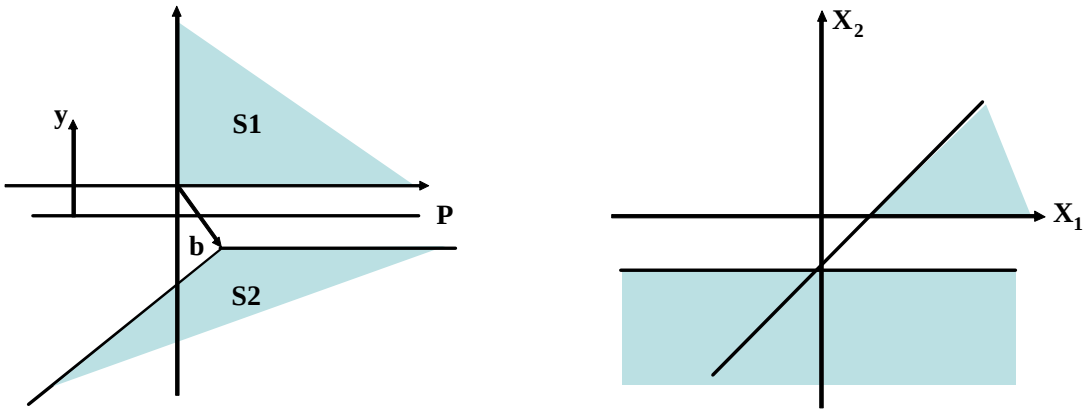
Theorem 2.3 follows directly from the previous discussions.

2.3.3 Some Examples

In this section we show some low-dimensional LP examples to illustrate by simple pictures how the instances on the boundary $\mathcal{B}$ might look and how the complexity of LP instances change as the parameters change.

Let $S1 = \{z \in \mathbb{R}^m : z \geq 0\}$ and $S2 = \{b - Ax : x \in \mathbb{R}^n, x \geq 0\}$. $Fd(A, b)$ has a solution if and only if $S_1 \cap S_2 \neq \emptyset$. Geometrically, finding a solution to $A1(A, b)$ is essentially separating S_1 and S_2 in $\mathbb{R}^m$ by a hyperplane with normal y. In this section, we call $\mathbb{R}^n$ the primal space and $\mathbb{R}^m$ the alternative variable, since the solution to Fd lies in $\mathbb{R}^n$ and the alternative solution to $A1$ lies in $\mathbb{R}^m$.

Example 2.1. $A = \begin{bmatrix} -1 & 1 \\ 0 & 1 \end{bmatrix}, b = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$.



(a) alternative space (b) primal space

Figure 2.11: Finding separating hyperplane for the LP instance,$-x_1 + x_2 \leq -1, x_2 \leq -1, x_1 \geq 0, x_2 \geq 0$

This LP is infeasible since the two shaded areas S_1 and S_2 have no intersection, as shown in Figure 2.11(a). The same instance is depicted in the primal space in Figure

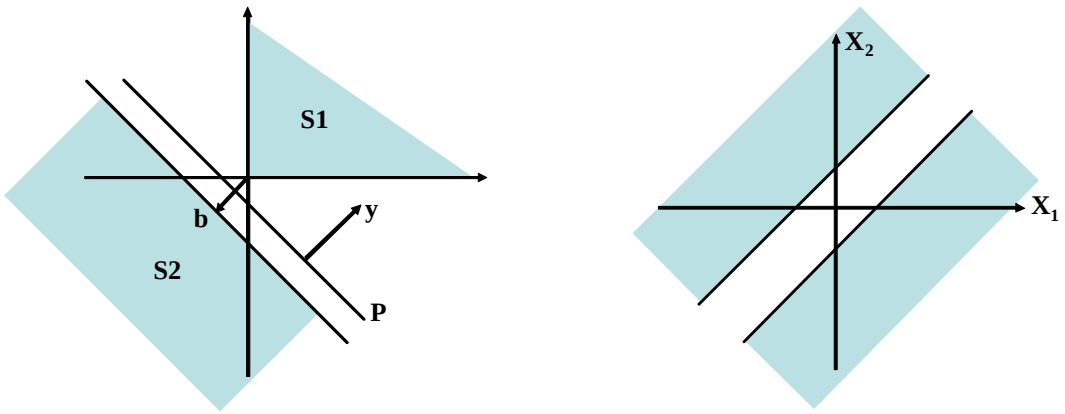
2.11(b). It is easy to see from the plot that an arbitrarily small disturbance

$$\Delta A = \begin{bmatrix} 0 & 0 \\ \epsilon & 0 \end{bmatrix},$$

with $\epsilon > 0$, can make the two shaded areas intersect at $z_1 = \infty$. This perturbation makes the $x_2 \leq -1$ region uplift at right and intersect with the other shaded region at $x_1 \to \infty$. The resulting LP is ill-conditioned and has extremely high complexity.

Example 2.2. $A = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}, b = \begin{bmatrix} -1 \\ -1 \end{bmatrix}.$

Just as Example 2.1, this LP is infeasible. For arbitrarily small $\epsilon > 0$, perturbation $\Delta A = \begin{bmatrix} -\epsilon & 0 \\ 0 & 0 \end{bmatrix}$ leads the LP to feasibility with x_1, x_2 both at infinity. The instance is shown in both alternative and primal space in Figure 2.12.



(a) alternative space (b) primal space

Figure 2.12: LP $x_1 - x_2 \leq -1, -x_1 + x_2 \leq -1, x_1 \geq 0, x_2 \geq 0$ in dual and primal spaces. As seen in the primal space, it is infinitely fragile.

Example 2.3. $A = \begin{bmatrix} a_1 & a_2 \end{bmatrix}, b = -1.$

In this example, we illustrate how complexity of LP instances change in the parameter space. Obviously, $Fd(A, b)$ is always trivially feasible with $x_1 = x_2 = 0$ when

$b \geq 0$, making these cases uninteresting. Therefore only the cases with $b < 0$ are considered and they are normalized so that $b = -1$. The boundary between feasible parameter sets and infeasible ones is the set $\{a_1 = 0, a_2 \geq 0\} \cup \{a_1 \geq 0, a_2 = 0\}$. The LP is infeasible when $a_1 \geq 0, a_2 \geq 0$, and feasible otherwise. As depicted in Figure 2.13, the hard cases only happen around the boundary, confirming that high complexity indicates ill-conditioning. Also note that not all LP instances near the boundary have high complexity, emphasizing that "complexity implies fragility" is a one-way relation.

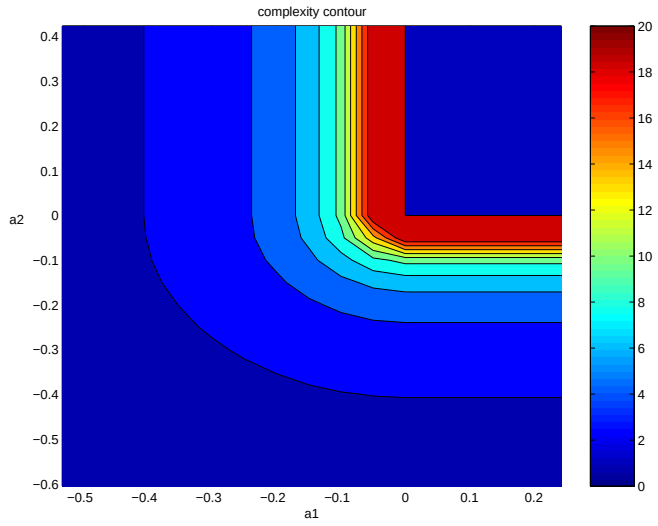


Figure 2.13: Complexity of $a_1x_1 + a_2x_2 \leq -1, x_1 \geq 0, x_2 \geq 0$ changes with a_1 and a_2. Clearly, the set of inequalities is infeasible when $a_1 \geq 0, a_2 \geq 0$ and feasible everywhere else. The figure shows that the complexity of the feasible instances blows up when the instance approaches the boundary. But the complexity of the infeasible ones is constant.

2.4 Conclusions and Discussions

We have looked at two classes of problems and examined their complexity and robustness. It has been shown that the verification complexity is upper bounded by the fragility. This confirms that robust systems necessarily are easily verifiable and systems that are hard to verify are fragile in the presence of certain perturbations.

Note that two-dimensional percolation lattice, as a class of problems, is very special since it observes symmetry in the sense that both failure events and barriers are captured by paths. This symmetry breaks when higher-dimensional lattices or more general LPs are considered. In the higher-dimensional lattice case, the barrier to a path will be a connection of cells that resemble a hyperplane.

In LPs, the barrier to finding a feasible point in the primal space would be creating a hyperplane in the dual space. It is still a feasibility LP problem. However, if we relax the linear restriction and consider feasibility problem

$$\mathbf{Fd}: \quad \begin{array}{l} \text{find } x \in \mathbb{R}^n \text{ such that} \\ f_i(x) \geq 0,\ i = 1, \cdots, m, \end{array} \tag{2.16}$$

where f_i are general polynomials, then proving Fd infeasible is not easy any more (unless all the polynomials are concave, in which case Fd is a convex problem and a solution to its alternative serves as a proof of inconsistency, provided some technical conditions hold). In the general case when f_i may or may not be concave, Positivstellensatz needs to be applied to find a proof that Fd is infeasible. Finding such certificate is a co-NP problem.

As our conclusion $R \cdot C \leq 1$ shows, complexity implies fragility. This fragility can come from a number of sources. Inappropriate design is obviously a possible source. What is normally overlooked is the fact that fragility comes from formulation of the problem, i.e., modeling, as well. For example, an LP instance with a single equation $4x = 12$ can be rewritten as an equivalent instance with two-inequalities $4x \geq 12$, $4x \leq 12$. The first formulation is robust as a feasibility decision problem. At the same time, the second formulation results in a system that is ill-posed: its consistency can be altered by an arbitrarily small perturbation in data [75].

Chapter 3

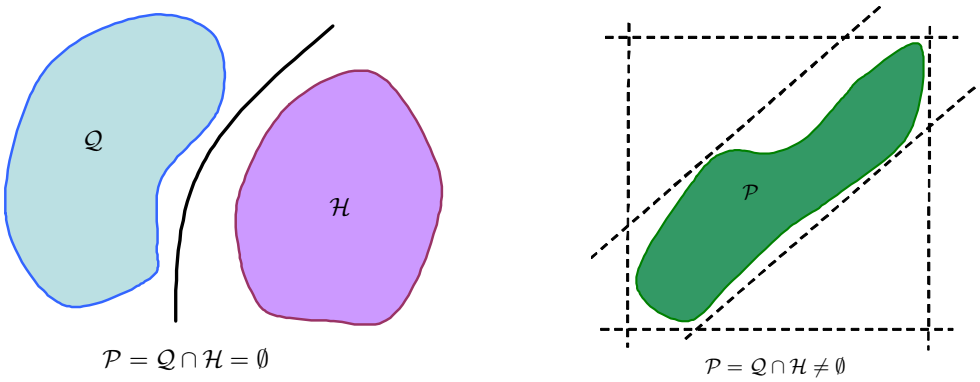
Model Validation of G-protein Signaling System in Yeast

3.1 Introduction

In this chapter, we consider a biological system for which a candidate model and experimental data are available. Some a priori knowledge on the system parameters, in this case reaction rates, has also been established. We then want to explore two problems, as shown in Figure 3.1. The first one is to determine if the model can explain the data at hand, especially when uncertainty exists in experiment observations and the a priori understanding on reaction rates. If the reaction rates determined by the data are not consistent with a priori knowledge, the model is invalidated, as the case in Figure 3.1(a). This invalidation can be rooted in experiment error, discrepancy between the model and real system, or our a priori knowledge.

When the model is actually valid, i.e., there exist parameter combinations that both comply to former knowledge and explain the data at hand, we want to be able to describe these parameter combinations. This can be very hard since the set of such parameters are usually of unknown shape. One way through which we can gain insight is to bound the parameter set as tightly as possible, in order to at least know what parameters are definitely infeasible. This is the case in Figure 3.1(b).

The biological system we study is the heterotrimeric G-protein cycle which is the



(a) Model invalidation (b) Parameter inference

Figure 3.1: Model invalidation and parameter inference from a priori knowledge and experiment data. Q is the set of parameters consistent with data and H is the set of a priori known possible parameters. $P = Q \cap H$ is therefore the set of parameters that are consistent with both prior knowledge and data.

vital component of a G-protein signaling pathway. Guanosine triphosphate-binding protein (G-protein) signaling pathways comprise one of the most important groups of cell-to-cell communication schemes in eukaryotes [49]. Many activities of the human body involve G-protein signaling pathways [13] and diseases such as cancer and diabetes are thought to be related to errors occurring in pathways of this kind. An estimated 50% of current prescription drugs target some factors along the G-protein signaling pathway [13,21,38,50]. Understanding the G-protein signaling transduction pathway quantitatively as well as qualitatively, therefore, can have great impact on new drug discovery and improvements of existing drugs.

In Section 3.2, the general framework and methods for model validation are presented. The example biological system we study is described in Section 3.3, along with the candidate model and experimental data. In Section 3.4, results of applying methods in Section 3.2 to the biological system are reported, and the limitations to these methods are discussed. Section 3.5 concludes.

3.2 General Framework

3.2.1 Problem Setup

Let $u \in \mathbb{R}$, $y \in \mathbb{R}$ denote the input and output vectors of a biological process, and let $p \in \mathbb{R}^k$ be the vector of parameters, usually the reaction rates of interest. A model of the system can often be described in the form of time-invariant ODEs

$$\dot{x} = f(x, u, p), \qquad x(0) = x_0, \tag{3.1}$$

$$y = g(x, u, p), \tag{3.2}$$

where $x \in \mathbb{R}^n$ is the state vector, $u \subset \mathbb{R}^{n_u}$ is the input, $y \subset \mathbb{R}^{n_y}$ is the input and $p \subset \mathbb{R}^k$ is the constant parameter vector. We are interested in the output value at a specific time T, y_T, under constant input. In this setting, the relation between the constant input u and output y_T can be captured by a static mapping $y_T = M(u, p)$, where M comes from numerical integration of equations (3.1)–(3.2) and in general does not possess a closed-form expression. In the rest of the chapter, y is used instead of y_T for simplicity.

Each experiment e on the system provides a data point $(v_e, d_e, \delta_e, \epsilon_e)$. v_e is the measurement of the input u_e and d_e is the measurement of the output y_e. δ_e and ϵ_e are the corresponding measurement uncertainties. They are vectors of appropriate dimension such that

$$\|v_e - u_e\| \leq \delta_e, \tag{3.3}$$

$$\|d_e - y_e\| \leq \epsilon_e, \tag{3.4}$$

where infinity norm is usually employed to bound measurement error on each input and output element.

In many biological processes, some prior knowledge of the parameters is available

either from separate experiments or literature. We assume the prior knowledge on parameters can be expressed by l polynomials. Thus, the prior parameter set is

$$\mathcal{H} = \{p \in \mathbb{R}^k : h_i(p) \leq 0, i = 1, 2, \cdots, l\}. \tag{3.5}$$

In most cases, this knowledge is presented as relative changes along with nominal value for each individual parameter. In these cases, $\mathcal{H}$ can be simplified to

$$\mathcal{H} = \{p \in \mathbb{R}^k : (1 - \beta_i)\bar{p}_i \leq p_i \leq (1 + \beta_i)\bar{p}_i, i = 1, 2, \cdots, k\}, \tag{3.6}$$

where $\bar{p}_i$ is the nominal value of parameter p_i. Writing everything in vectors with $\beta = [\ \beta_1 \quad \cdots \quad \beta_k\ ]^T$, then

$$\mathcal{H} = \{p \in \mathbb{R}^k : -\beta\bar{p} \leq p - \bar{p} \leq \beta\bar{p}\}. \tag{3.7}$$

Such a prior parameter set is a hypercube in the parameter space $\mathbb{R}^k$.

The set of parameters that are consistent with m experiments is

$$\mathcal{P} = \{p \in \mathcal{H} : \forall e = 1, \cdots, m, \exists u_e \text{ with } \|u_e - v_e\| \leq \delta_e \text{ s.t. } \|M(u_e, p) - d_e\| \leq \epsilon_e\}. \tag{3.8}$$

The problem we consider is to either prove $\mathcal{P}$ is empty or describe $\mathcal{P}$ as precisely as we can. If $\mathcal{P}$ is empty, we need to find a certificate such as the separating curve in Figure 3.1(a). This certificate can be based on barrier method [69] or on the use of surrogate models [29, 82]. If not empty, $\mathcal{P}$ can be best characterized by solutions to a series of optimization problems:

$$\begin{aligned} \Phi^* = \quad & \min_p \quad \Phi(p) \\ & \text{s.t.} \quad p \in \mathcal{P}. \end{aligned} \tag{3.9}$$

We choose $\Phi(p)$ depending on the aspects of $\mathcal{P}$ we are interested in. For example, $\Phi(p) = p_j$ or $\Phi(p) = -p_j$ if we are interested in the minimum or maximum values of the j-th parameter.

3.2.2 Model Invalidation

Difficulty of model invalidation arises when the proposed model such as that in (3.1)–(3.2) comes with a set of possible parameter ranges. In this case, there are infinite possible parameter combinations, making simulation methods impossible to carry out. An invalidation method that deals with this difficulty is the barrier method, in which a Lyapunov-like certificate is found [68]. In Section 3.2.2.1, the essence of this method is reviewed. We will see in Section 3.4.1.1 that although this method is theoretically powerful, it is not computationally scalable. Therefore we adopt an approximation method that uses a static mapping, called a surrogate model, to replace the dynamical relationship [93]. This method is discussed in Section 3.2.2.2.

3.2.2.1 Barrier Method

In this subsection, we review the barrier method developed in [68, 69], where ODE models with parameter and state uncertainties are invalidated against experimental data. Barrier function methodology can directly work with system states. The essence of the method is as follows.

For a state space model

$$\dot{x} = f(x(t), u, p, t), \tag{3.10}$$

where $x \in \mathbb{R}^n$ is the state vector, $u \in U \subset \mathbb{R}^{n_u}$ is the constant input vector and $p \in P \subset \mathbb{R}^k$ is the parameter vector. Suppose experiments are carried out which imply that the initial state $x(0) \in \mathcal{X}_0$ and the final state $x(T) \in \mathcal{X}_T$. We also assume that $x(t) \in \mathcal{X}$ for all $t \in [0, T]$. Then the following theorem holds.

Theorem 3.1 (Prajna[1]). *Let the model (3.10) and the sets P, $\mathcal{X}_0, \mathcal{X}_T, \mathcal{X}$ be given. If there exists a function $B : \mathbb{R}^n \times \mathbb{R}^{n_u} \times \mathbb{R}^k \times \mathbb{R} \to \mathbb{R}$ differentiable with respect to x and t such that*

$$B(x_T, u, p, T) - B(x_0, u, p, 0) > 0 \qquad (3.11)$$
$$\forall x_T \in \mathcal{X}_T, x_0 \in \mathcal{X}_0, u \in U, p \in P,$$

$$\frac{\partial B}{\partial x}(x, u, p, t) f(x, u, p, t) + \frac{\partial B}{\partial t}(x, u, p, t) \leq 0 \qquad (3.12)$$
$$\forall x \in \mathcal{X}, u \in U, p \in P, t \in [0, T].$$

Then the model (3.10) and its associated parameter set P are invalidated by $\{\mathcal{X}_0, \mathcal{X}_T, \mathcal{X}\}$.

Such a function B is called a barrier certificate. It is a rigorous proof of inconsistency between model and data. Note that in Theorem 3.1, only two pieces of data concerning states at $t = 0$ and $t = T$ are incorporated. Currently the only way to handle three or more data is to increase the system size, adding n states for each additional data point. See [68] for details and examples. Barrier functions are computed by solving SOS programs [69]. This augmentation of state space quickly drives the size of SOS programs to the extent that it can not be handled by optimization solvers. As such, alternative methods need to be explored.

3.2.2.2 Invalidation with Surrogate Models

The lack of analytical form for the mapping $y = M(u, p)$ corresponding to model (3.1)–(3.2) makes it very hard to prove the consistent parameter set $\mathcal{P}$ defined in (3.8) is empty. An indirect approach is to approximate $M(u, p)$ by a static mapping $S_e(u, p)$ in the region of interest, which is called a surrogate model of $M(u, p)$ [82]. The subscript e in S_e emphasizes the fact that for different region, the surrogate model can be different. S_e is usually picked to be a polynomial. When M is a

[1] This is a variant of the theorem in [68]. See [68, 69] for details.

continuous mapping, the difference between M and S_e can be made arbitrarily small if the order of the polynomial S_e is high enough[2]. Coefficients of the surrogate model are determined through simulations of the dynamical system [28, 29].

Suppose the approximation error is bounded on the region of interest by ζ, i.e.,

$$\|M(u_e, p) - S_e(u_e, p)\| \leq \zeta_e, \quad \forall p \in \mathcal{P}, \|u_e - v_e\| \leq \delta_e, e = 1, \cdots, m. \tag{3.13}$$

Then the set $\mathcal{P}$ can be bounded by two sets,

$$\mathcal{P}_I \subseteq \mathcal{P} \subseteq \mathcal{P}_O, \tag{3.14}$$

where

$$\begin{aligned} \mathcal{P}_I = \{p \in \mathcal{H} : \forall e = 1, \cdots, m, \exists u_e \text{ with } \|u_e - v_e\| \leq \delta_e \\ \text{s.t. } \|S_e(u_e, p) - d_e\| \leq \epsilon_e - \zeta_e\}, \end{aligned} \tag{3.15}$$

$$\begin{aligned} \mathcal{P}_O = \{p \in \mathcal{H} : \forall e = 1, \cdots, m, \exists u_e \text{ with } \|u_e - v_e\| \leq \delta_e \\ \text{s.t. } \|S_e(u_e, p) - d_e\| \leq \epsilon_e + \zeta_e\}. \end{aligned} \tag{3.16}$$

From (3.14), it is easy to see that $\mathcal{P}_O = \emptyset$ is a sufficient condition for inconsistency among model, a priori parameter knowledge and data. Let us consider systems with scalar input and output. That is, $n_u = n_y = 1$. Then the norm conditions in (3.16) reduce to conditions on absolute values. For a set of experimental data $(v_e, d_e, \delta_e, \epsilon_e)$, $e = 1, \cdots, m$, $\mathcal{P}_O$ consists of parameter points p that simultaneously satisfy the following set of polynomial inequalities.

$$\begin{aligned} \mathcal{P}_O = \{p \in \mathbb{R}^k, K_j(p) \geq 0, V_e(u_e, p) \geq 0, W_e(u_e, p) \geq 0, \\ Z_e(u_e) \geq 0, j = 1, \cdots, k, e = 1, \cdots, m\}, \end{aligned} \tag{3.17}$$

[2]This is possible because of the Stone-Weierstrass Theorem.

where

$$K_j(p) = \big((1+\beta_j)\bar{p}_j - p_j\big)\big(p_j - (1-\beta_j)\bar{p}_j\big), \tag{3.18}$$

$$V_e(u_e, p) = d_e - S(u_e, p) + \epsilon_e + \zeta_e, \tag{3.19}$$

$$W_e(u_e, p) = -d_e + S(u_e, p) + \epsilon_e + \zeta_e, \tag{3.20}$$

$$Z_e(u_e) = (u_e - v_e + \delta_e)(v_e + \delta_e - u_e). \tag{3.21}$$

Here K_j and Z_e capture the uncertainty on parameters and input measurements. V_e and W_e are the lower and upper bounds posed implicitly by the model M on the output measurement d_e. $\mathcal{P}_O$ is empty if, and only if, there exists a polynomial $h(u_1, \cdots, u_m, p) : \mathbb{R}^m \times \mathbb{R}^k \to \mathbb{R}$ such that

$$h = h_0 + \sum_{i=1}^{r} q_i h_i < 0, \tag{3.22}$$

where $h_0, h_1, \cdots, h_r$ are polynomials that can be expressed as sum of squares, and $q_1, \cdots, q_r$ are finite products of the members of the set $\{K_j, V_e, W_e, Z_e, j = 1, \cdots, k, e = 1, \cdots, m\}$. This is a direct application of Positivstellensatz (see [65] for details). Searching for h can be done using SOSTOOLS [67], in general with less computational cost than finding barrier certificates.

When quadratic surrogate models are deployed and the order of h is set to 2, finding h is simplified to finding nonnegative multiplier λ's such that

$$\sum_{j=1}^{k} \lambda_{K_j} K_j(p) + \sum_{e=1}^{m} \big(\lambda_{V_e} V_e(u_e, p) + \lambda_{W_e} W_e(u_e, p)\big) + \lambda_{Z_e} Z_e(u_e) < 0 \tag{3.23}$$

for all values of u_e and p. Such λs constitute an invalidation certificate. If these λs can be found, their values provide useful information: constraints that have zero multipliers do not contribute to invalidating the model and those associated with big multipliers tend to have a strong impact to the invalidation. Directional information

can also be obtained. For example, if $\lambda V_e > 0$, then we know that the e-th output measurement is too small to match the prediction from the model.

3.2.3 Robust Parameter Inference

When a model structure is adequate, then an important task is to describe the model as precisely as possible, essentially determining ranges of consistent parameters. Traditionally in chemical engineering and biology, focus has been put on searching for a "best-fitting" parameter combination through local search methods [39, 54]. For a dynamical system described by ODE as in (3.1)–(3.2), a best-fitting parameter p^* is the optimal solution to the following optimization problem.

$$\begin{aligned} \min_{p} \quad & \sum_{e=1}^{m} (y_e - d_e)^2 \\ \text{s.t.} \quad & \dot{x}_e = f(x_e, v_e, p) \\ & y_e = g(x_e, v_e, p), \quad e = 1, \cdots, m \end{aligned} \tag{3.24}$$

where v_e is the e-th input, assumed to be precise.

However, a big disadvantage of this approach is that the best-fitting parameter does not give global information about the set $\mathcal{P}$. It is also not robust to experiments, i.e., the best-fitting parameter shifts around when additional experimental data are available. To obtain information about $\mathcal{P}$ and reveal possible parameter correlation, the optimization problem (3.9) can be studied.

Depending on the property of $\mathcal{P}$ we want to investigate, Φ is set accordingly. For example, when $\Phi = p_j$, the optimal value of (3.9) is the minimum value p_j can take when other uncertainty factors, including parameters and data, vary through their range. This is a measure that ignores parameter correlation and usually gives little information of the feasible set $\mathcal{P}$ [28, 93].

Correlations among parameters are often of interest to scientists. To study such correlations, the ratio between them can be used. For example, for parameters p_i and

p_j, we can find γ_{min} and γ_{max} such that

$$\begin{aligned} \gamma_{min} = \min_{p} \quad & \frac{p_i}{p_j} \\ \text{s.t.} \quad & p \in \mathcal{P} \end{aligned} \qquad\qquad \begin{aligned} \gamma_{max} = \max_{p} \quad & \frac{p_i}{p_j} \\ \text{s.t.} \quad & p \in \mathcal{P} \end{aligned} \tag{3.25}$$

Because $\mathcal{P}$ is described by the model M and can not be explicitly expressed, it is virtually impossible to compute γ_{min} and γ_{max}. Surrogate models can be used to find bounds on these values. That is,

$$\gamma'_{min} \leq \gamma_{min} \leq \gamma_{max} \leq \gamma'_{max}, \tag{3.26}$$

where

$$\begin{aligned} \gamma'_{min} = \min_{p} \quad & \frac{p_i}{p_j} \\ \text{s.t.} \quad & p \in \mathcal{P}_O \end{aligned} \tag{3.27}$$

and γ'_{max} is defined analogously.

(3.27) is a non-convex problem to which, fortunately, SOS relaxation can be applied. Let the surrogate models be polynomials such that $\mathcal{P}_O$ is represented by a set of polynomial inequalities. A lower bound $\underline{\gamma}$ to γ'_{min} can be obtained by solving the SOS relaxation [65] of the following equivalent problem of (3.27)

$$\begin{aligned} \min \quad & -\gamma \\ \text{s.t.} \quad & p_i \geq \gamma p_j, \qquad \forall p \in \mathcal{P}_O. \end{aligned} \tag{3.28}$$

Similarly an upper bound $\overline{\gamma}$ on γ'_{max} can be obtained by SOS relaxation.

Correlation between two parameters p_i and p_j can also be revealed by bounding the feasible region on two-dimensional (p_j, p_i) plane by parallel lines calculated through SOS relaxation of the following optimization problem.

$$\begin{aligned} \min \quad & \gamma \\ \text{s.t.} \quad & \mu p_j + a \leq p_i \leq \mu p_j + a + \gamma, \qquad \forall p \in \mathcal{P}_O. \end{aligned} \tag{3.29}$$

3.3 Model and Data

G-protein signaling pathways can be activated by a variety of ligands and induce different responses, but they all share a similar signal-transducing pathway. The yeast pheromone response pathway in yeast is one of the best studied G-protein signaling pathways [88]. Each type of haploid yeast cells secrete a unique pheromone: **a** cells secrete **a**-factor and α cells secrete α factor. The G-protein coupled receptors on cell surface respond to the pheromone secreted by cells of opposite type and lead to the yeast mating process.

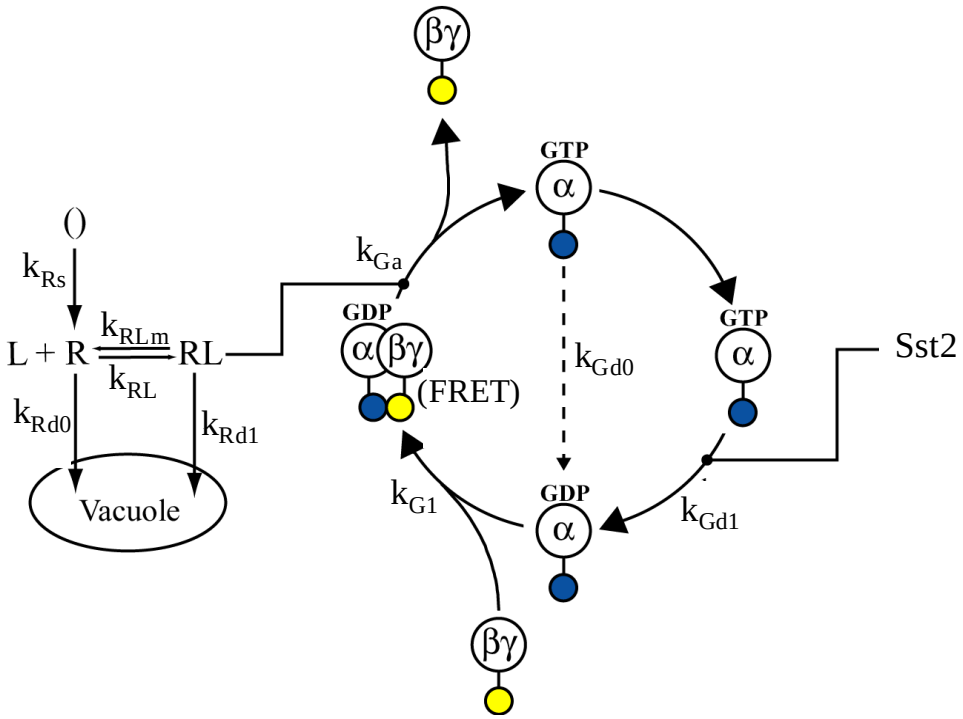


Figure 3.2: Schematic diagram of yeast heterotrimeric G-protein cycle. The blue circle represents CFP and the yellow circle represents YFP.

A central part of this pathway is the heterotrimeric G-protein cycle. A schematic diagram of this cycle in yeast is adapted from [94] and shown here in Figure 3.2. The G-proteins consist of three subunits: G_α, G_β and G_γ. In its inactive form, the G_α subunit is bound to Guanosine diphosphate (GDP) and is associated with

$G_{\beta\gamma}$. Ligand binding to the receptor activates the associated G-protein, leading to substitution of the GDP associated with G_α by GTP followed by the dissociation of G-protein into G_α and $G_{\beta\gamma}$ complexes. The activated G_α or $G_{\beta\gamma}$ in turn activates or inhibits several effector proteins, resulting in different cell responses. The intrinsic GTPase activity of G_α hydrolyzes the bound GTP to GDP, turning G_α back to its inactive state and resulting in the reassociation of G-protein. To quantitatively characterize this process, Fluorescence Resonance Energy Transfer (FRET) was used by Yi et al. [94] to trace the activation/deactivation cycle of G-protein, where Cyan Fluorescent Protein (CFP) and Yellow Fluorescent Protein (YFP) are attached to G_α and $G_{\beta\gamma}$, respectively, as in Figure 3.2.

We follow the model of the heterotrimeric G-protein cycle as in [94] where the following processes are modeled: (1) the binding of ligand (L) to receptor (R); (2) the synthesis and degradation of receptor; (3) the activation of G-protein; (4) the deactivation of G-protein by the catalyst Sst2p and (5) the reassociation of G_α and $G_{\beta\gamma}$. Let x_1 be the concentration of receptor, x_2 of ligand bound to receptor, x_3 of inactive heterotrimeric G-protein and x_4 of activated G_α, i.e., $x_1 = [R]$, $x_2 = [RL]$, $x_3 = [G]$, $x_4 = [G_\alpha]$. Then heterotrimeric G-protein cycle is modeled as

$$\begin{aligned}
\dot{x_1} &= -k_{RL}x_1u + k_{RLm}x_2 - k_{Rd0}x_1 + k_{Rs} \\
\dot{x_2} &= k_{RL}x_1u - k_{RLm}x_2 - k_{Rd1}x_2 \\
\dot{x_3} &= -k_{Ga}x_2x_3 + k_{G1}(G_t - x_3 - x_4)(G_t - x_3) \\
\dot{x_4} &= k_{Ga}x_2x_3 - k_{Gd1}x_4 \\
y &= (G_t - x_3)/G_t
\end{aligned} \tag{3.30}$$

where the input u is the concentration of the pheromone ligand α-factor and the output is the normalized level of $G_{\beta\gamma}$. G_t is the total number of G-protein. Parameter nominal values $\bar{p}_i$ are listed in Table 3.1. Note that $\bar{p}_1$ through $\bar{p}_5$ and $\bar{p}_9$ are measured directly. $\bar{p}_6$ and $\bar{p}_7$ are inferred from experiment data. $\bar{p}_8$ is based on estimates in the

literature.

parameter p_i	physical meaning	nominal value $\bar{p}_i$
p_1	k_{RL}	$10^6 mM^{-1}s^{-1}$
p_2	k_{RLm}	$0.01s^{-1}$
p_3	k_{Rd0}	$0.0004s^{-1}$
p_4	k_{Rd1}	$0.004s^{-1}$
p_5	k_{Rs}	$4mMs^{-1}$
p_6	k_{Ga}	$10^{-5}mM^{-1}s^{-1}$
p_7	k_{Gd1}	$0.1s^{-1}$
p_8	k_{G1}	$1mM^{-1}s^{-1}$
p_9	G_t	$10^4 mM$

Table 3.1: Nominal values of parameters

The *in vivo* dynamics and regulation of the heterotrimeric G-protein cycle is measured in yeast using FRET [94]. The dose-response data at time $t = 60s$ are listed in Table 3.2. In these experiments, the initial states of the system (3.30) are assumed to be known as they are controlled.

experiment e	input $v_e(nM)$	output d_e
1	1	0.083
2	2	0.122
3	5	0.240
4	10	0.352
5	20	0.384
6	50	0.397
7	100	0.400
8	1000	0.397

Table 3.2: Dose response data measured *in vivo* using FRET

3.4 Results and Limitations

3.4.1 Model Invalidation

In this section, model invalidation methods discussed in Section 3.2.2 are used on the ODE model in (3.30) and the dose response data in Table 3.2.

3.4.1.1 Barrier Method

In our application, the output y is directly a function of the state x_3. A single dose-response measurement gives information about the final state of the system. Theorem 3.1 is applied with the first measurement in Table 3.2 using SOSTOOLS [67]. In order to find a barrier function computationally, only 4% uncertainty is allowed for p_6 and p_7 and 5% uncertainty is allowed for the output measurement $y = 0.083$ with all other parameters but p_8 fixed at their nominal values. p_8 is set to 10^{-3}. In addition to these conditions, bounds on the state are also posed by estimates obtained through simulation of the system. In such settings, a barrier certificate is found.

Note that we have to pose very stringent constraints on the system in order for the computation to go through. When constraints are loosened, computational issues arise, e.g., numerical errors are encountered. This makes the application of barrier function to real systems restrictive. Another disadvantage of this approach is that the size of the SOS program grows very fast with the number of measurements, making implementation virtually infeasible with the computational capacity available so far. The advantage of directly incorporating system dynamics becomes a disadvantage computation-wise. We bypass this by using surrogate models.

3.4.1.2 Using Surrogate Models

In the G-protein system, when the measurement errors are assumed to be 10%, i.e., $\epsilon_e = 0.1d_e$, $\delta_e = 0.1v_e$ for all experiments, and uncertainty on parameters are taken to be 30%, i.e., $\beta_i = 0.3$ for all parameters, model (3.30) is invalidated by the 8 measurement points in Table 3.2. Furthermore, we are able to find an invalidation certificate that has only three positive multipliers corresponding to output measurements (measurement 1, 7 and 8): $\lambda_{W_1} = 55.21$, $\lambda_{V_7} = 8.509$, $\lambda_{V_8} = 0.448$.

The multipliers corresponding to measurements 1 and 7 are much bigger than that for measurement 8, indicating the invalidation result is most sensitive to these two data points. We are therefore tempted to invalidate the model using only mea-

surements 1 and 7. Model (3.30) is indeed invalidated by these two measurements with new multiplier values $\lambda_{W_1} = 88.538$, $\lambda_{V_7} = 14.508$. The nonzero multipliers λ_{W_1} and λ_{V_7} give directional information: to be consistent with the model (3.30) and the specified 30% parameter uncertainty around nominal values, the 1st output measure should be lower than recorded and the 7th output should be higher.

When searching for h over lower-degree polynomials fails to render an invalidation certificate, h can be set to higher order. In that case, a certificate is the set of polynomials $h_i, i = 0, \cdots, r$. As long as h is searched over finite degree polynomials, two relaxation steps are taken. The first one is proving emptiness for set $\mathcal{P}_O$ instead of for $\mathcal{P}$ through using surrogate models; the second one is using SOS relaxation to show positiveness of a polynomial. Therefore the result is conservative compared to that obtained from barrier function method. Thus, an invalid model can be deemed as inconclusive.

Higher-order polynomial surrogate models can also be deployed to improve approximation precision. But both approaches usually cause the computation load to grow quite rapidly. There are situations in which a non-polynomial surrogate model might fit the physical model M better than polynomials, however care is needed since proving the emptiness of $\mathcal{P}_O$ using SOS methodology becomes less straightforward, sometimes even impossible.

Independent of our work, Feeley et al. [25] proposed a measure of dataset consistency which is the maximum decrease in absolute data uncertainty before a dataset becomes inconsistent and the use of optimization techniques such as branch and bound to compute the value of this measure. Their method faces the same challenge. That is, examples can be inconclusive unless put through an almost exhaustive search.

3.4.2 Description of Feasible Parameter Set

Given that the model (3.30) is proved to be inconsistent with the dose response data, synthetic data — simulated data contaminated by random error — are used to

study the feasible parameter set $\mathcal{P}$. Simulations are carried out at nominal parameter values, i.e., $d_e = M(u_e, \bar{p}) + \epsilon_e$, where $\epsilon_e \sim \mathcal{N}(0, 0.03y_e)$. The data is shown in Table 3.3. p_6 and p_7 are the G-protein activation and deactivation rates. As they are our main interest, in the following analysis, all but these two parameters are assumed to be fixed at nominal values. We study the (p_6, p_7) set that is consistent with data in Table 3.3.

experiment e	1	2	3	4	5	6	7
input v_e	1	2	5	10	20	50	100
output d_e	0.0354	0.0626	0.1397	0.2245	0.3326	0.4204	0.4359

Table 3.3: Synthetic data obtained by simulating model (3.30) at nominal parameter values

In computing the "best-fitting" parameter, input measurements are considered exact, i.e., $u_e = v_e$. We start with an initial guess, generate a surrogate models S_e of M in its neighborhood N_0 and solve

$$\min \quad \sum_{e=1}^{m} (d_e - S_e(u_e, p))^2. \tag{3.31}$$

If the minimizer p_{N_0} is outside the neighborhood N_0, where the surrogate model is built for, a new surrogate model is built in a neighborhood N_1 of p_{N_0} to again solve (3.31) until for some i, p_{N_i} falls inside N_i [54]. Starting from $p_6 = 2 \times 10^{-5}$, $p_7 = 0.2$, the best-fitting parameter for data points 1–6 in Table 3.3 is $p_6^* = 8.1812 \times 10^{-6}$, $p_7^* = 0.0788$. When all 7 data points are used, $p_6^* = 1.0689 \times 10^{-5}$, $p_7^* = 0.1048$. These are, of course, local minimums.

Note that in this approach, the input uncertainty is not taken into account. Now suppose these are data from experiments and there is a 10% measurement error associated with both input and output values. Suppose the prior knowledge on (p_6, p_7) is that they vary within 50% around nominal values, i.e., $\beta_6 = \beta_7 = 0.5$. As pointed out earlier, bounding individual parameter gives little extra information. This em-

phasizes the importance of capturing parameter correlation. In the G-protein system, the G-protein activation rate p_6 and deactivation rate p_7 are expected to exhibit a positive correlation. Using SOS relaxation of (3.28) and (3.29), the feasible region is bounded from outside as shown in Figure 3.3. Parameter combinations outside the shaded region are not consistent with the simulated data. Using this approach, the shaded region will shrink as more data are available, giving us more understanding of the consistent parameter set.

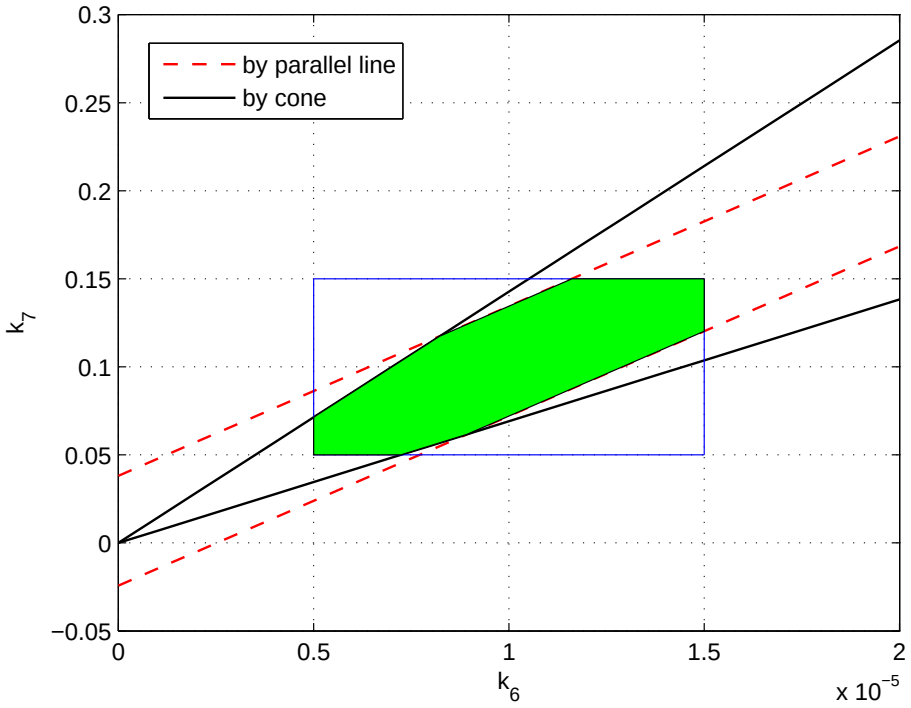


Figure 3.3: (p_6, p_7) combinations consistent with data are inside the shaded region.

3.5 Conclusions

In this Chapter, we studied issues in linking data with model: invalidating candidate model against data or getting information about consistent parameters from data.

Invalidation methods taking uncertainty into account include the barrier function

method and invalidation using surrogate models. Barrier function gives a rigorous and less conservative certificate, but the computation requirement is high and scaling with the number of data points can not be handled efficiently. Invalidation using surrogate models requires less computation effort, but the certificate is conservative. This added conservativeness comes from the approximation error between the surrogate model and the ODE model. When the uncertainty on data and parameter is high, i.e., the range over which surrogate model is built is big, the fitting error tends to grow. Higher-order polynomial or other form of surrogate model can be employed, with increased computation load for invalidation. Furthermore, in practice, this approximation error is obtained by first estimating the fitting error by random sampling and then relaxing it, bringing in another level of conservativeness.

To study the consistent parameter set, a robust approach is proposed, where parameter set is bounded from outside to reveal parameter correlation. In contrast to the single-point fitting approach, where only local information is gained about the feasible set, this robust approach eliminates parameter combinations that are inconsistent with available data. Surrogate models and SOS relaxation are used, therefore the bounds can be conservative: although parameters outside the bounds are inconsistent, there usually are parameters inside the bounds that are also inconsistent.

Much work is needed to improve what we discussed here to handle more data points and the conservativeness of using surrogate models. One possible way is to divide the prior parameter set and work on the subsets individually. A proper division scheme is then needed.

Chapter 4

Estimation of Expected Shortfall – A Risk Measure

4.1 Introduction

Risk measure is central to guiding portfolio choices as well as protecting financial institutions from adverse market movements. In particular, banks are required by regulators to set aside a certain amount of capital to insure solvency. This capital amount is determined according to the risk of the bank's portfolio. Therefore, the definition, estimation and evaluation of a proper risk measure have been central topics in the history of finance theory.

Until the middle of the last century, economists conceived of the financial investments as similar to gambling activities. Risk was discussed more from an intuitive sense and investors were suggested to invest in the individual stock that gives the highest discounted stream of expected future returns [90]. This view failed to explain diversification. Expected Utility theory [6,87] was also proposed to describe preference among uncertain payoffs. Different shape of utility functions represented different attitudes towards risk. For example, concave utility functions correspond to decision making criterion of risk-averse investors. However, risk was still not quantified. The implicit utility functions for different individuals can differ remarkably. Due to these difficulties, while this representation and many later variants, e.g., [32, 36, 73, 91], are

the standard approach in microeconomics, they are difficult to use in practice to guide portfolio choice.

It was first pointed out by Markowitz [51] that investors should care about risk as well as return. He suggested that instead of constructing a portfolio from financial assets with best future expected return, an investor should strike a balance between anticipated portfolio return and risk.

Markowitz proposed using the variance as a measure of risk: a portfolio with less variance is considered less risky. In his framework, risk is quantified by a number, namely the variance of the portfolio. This is the foundation of Modern Portfolio Theory. Starting from then, almost all return-risk models have used variance as the measure of risk. The widely used Capital Asset Pricing Model [83] is one of its descendants. As predominant as the mean-variance portfolio selection methodology, the occurrence of asymmetric return distribution makes variance an unsatisfactory measure of risk, as it treats equally the desirable upside and undesirable downside variations.

The following example illustrates this point by comparing two independently distributed assets whose marginal probability density are shown in Figure 4.1. The return of asset 1, $X1$, follows an Asymmetric Power Distribution[1] with symmetry parameter $\beta = 0.2$ and decay parameter $\lambda = 1$. The return of asset 2, $X2$, is normally distributed with the same mean and variance as that of asset 1. We see that asset 1 performs better than asset 2 in both tails: it is thinner in the lower tail and fatter in the upper tail. But since they have the same mean and variance, they are indistinguishable to investors carrying out the mean-variance selection method.

Motivated by this lack of asymmetric treatment of return shifts by variance, various measures of risk based on the down-side tail distribution of the return was pro-

[1] APD is a broad family of probability distributions indexed by a symmetry parameter $\beta \in (0, 1)$ and a decay parameter $\lambda > 0$. When $\beta = 0.5$, the probability density function is symmetric. The distribution is fat-tailed when $0 < \lambda < 2$ and short-tailed when $\lambda \geq 2$. See [46] for a detailed discussion of the APD family.

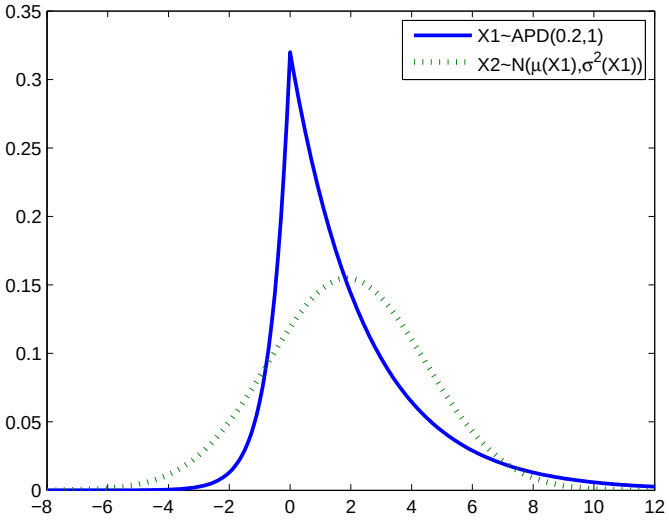


Figure 4.1: Returns from asset 1 and asset 2 have the same mean and variance, yet they differ remarkably at tails. Solid line: return of asset 1 follows APD(0.2,1). Dotted line: return of asset 2 follows $\mathcal{N}(\mu(X1), \sigma^2(X1))$.

posed and studied. Among them, Value-at-Risk (VaR) is one of the most popular. VaR as a risk measure was proposed by the Basel Committee on Banking Supervision (BCBS) in 1996 [4]. Since then, it has become the industry standard for financial risk measuring. α-VaR gives the magnitude of the worst loss in the best $(1-\alpha)100\%$ of the scenarios. It is an attractive measure of risk in many aspects, including conceptual simplicity, computational tractability and availability of various estimation methods [1, 44, 45, 53] and testing procedures [8, 17, 20, 34, 96].

However VaR has several shortcomings. It was pointed out by Artzner et al. [3] that VaR is not a coherent risk measure. In particular, VaR does not satisfy subadditivity, which means that the VaR of a combined portfolio can be bigger than the sum of the VaRs of its component portfolios. Thus using VaR to guide risk management can lead to risky asset allocation [31]. VaR has deficiencies other the lack of subadditivity. By definition, VaR measures the risk of the return by a single point and discards the whole tail distribution. Thus, VaR has a problem in providing

information regarding the size of loss when the loss exceeds VaR. A portfolio with smaller VaR can have bigger losses beyond VaR, a problem called "tail risk" [92]. Using VaR as the risk measure may lead a portfolio manager to make the non-sensible decision of picking the more risky portfolio. Moreover, since VaR concerns only one point on the return distribution, it is easy to manipulate using options [18].

Expected Shortfall (ES) was proposed as an alternative to VaR recently by Artzner et al. [3] as a remedy for VaR's lack of subadditivity. It has gained considerable interest in the financial community since then. In financial terms, ES represents the tail-loss in the market value of a given portfolio over a given time horizon. In mathematical terms, ES at level α is the expected loss of the portfolio's return in the least favorable $100\alpha\%$ situations. It is a coherent risk measure, satisfying subadditivity along with translation invariance, positive homogeneity and monotonicity. Although VaR and ES both have tail risk under certain conditions, ES is less problematic in disregarding distribution of fat tails [92]. Furthermore, because ES is subject to changes in the tail distribution, it is impossible to manipulate the ES of a portfolio by simple trading strategies such as buying and selling options.

There exists a long list of research works on risk assessment in economic models that have emphasized the importance of ES in risk measurement. Examples include axiomatic foundations of ES [3, 27] and theoretical properties of different variants of ES [2, 78, 84]. Triggered by the desirable properties of ES, models of optimal portfolio choice based on ES have been studied by many authors (see [5, 7, 77, 78]). Finally, the link between the Choquet expected utility theory and ES was provided in [5], thus grounding it within the framework of models of choice under uncertainty.

Unsurprisingly, the econometrics literature on ES has been rapidly growing. The approaches to ES estimation offered by the current work can be divided into two categories: fully parametric methods and non-parametric methods [26, 52, 79, 80].

Fully parametric methods are based on parametric assumptions on the return distribution. Parameters of this parametric distribution are then estimated, rendering

estimates of the ES. The most naive approach is to assume normality. However, this assumption often leads to underestimation of the lower-tail based risk measures, including VaR and ES, since financial data are usually fat-tailed. It also fails to capture the skewness of the return data. A partial fix to this is to use a t-distribution instead of a Gaussian. t-distribution addresses the leptokurtosis, but leaves the asymmetry in return unattended. To address this issue, several parametric models have been proposed. Examples include multivariate normal inverse Gaussian distribution [1] and APD family of distributions [46]. Fully parametric models are conceptually easy and computationally straightforward, but are highly fragile to misspecification of the return distribution.

To illustrate this, let us look at the example considered earlier in this section. Suppose the true return $X1$ is as shown in Figure 4.1. Its 5%-ES is 1.4914. First we assume the return is an APD random variable with $\lambda = 1$ and we need to estimate β. With an i.i.d. sample of 10000 observations, we obtain a maximum likelihood estimate $\hat{\beta} = 0.1994$ and compute an estimate of the 5%-ES as 1.4883 based on $\hat{\beta}$. Then suppose we do not know the distribution of $X1$ and naively assume it is Gaussian. From the same sample, we estimate the mean and variance of the Gaussian to be 1.8834 and 6.5443, respectively. The 5%-ES is then computed to be 3.3934 from the estimated Gaussian distribution. While the parametric method is efficient in the case of a correctly-specified model, its estimation error is large (128% in our example) in the case of misspecification.

Nonparametric methods make no parametric assumptions on the return and use kernel functions to estimate the whole distribution. While the result is robust to distributional misspecification, the optimal choice of kernel functions and bandwidth is in general not clear.

In this chapter, we propose an estimator based on empirical likelihood method introduced by Owen [59–61]. The empirical likelihood method makes no assumptions on the distribution of data, thus avoiding the problem of misspecification. The focus

is put on the parameter of interest instead of the whole distribution by specifying certain moment conditions. The parameter value that maximizes this likelihood subject to these moment conditions is defined as the maximum empirical likelihood estimator (MELE) [61, 72]. Since MELE has desirable properties comparable to parametric likelihood methods, empirical likelihood and its variants [16, 42] have gained substantial interest in recent years, both for parameter estimation [43, 95] and for hypothesis testing [24, 85]. A good survey on the theory and application of empirical likelihood methods can be found in [41].

This chapter is structured as follows. In Section 4.2, we give mathematical definition of the risk measures aforementioned, namely VaR and ES, and present the advantages of ES over VaR. An estimator for ES based on empirical likelihood is given in Section 4.3.2. The asymptotic properties of this estimator is presented in Section 4.4. Section 4.5 presents simulation results of the estimator. Section 4.6 briefly concludes.

4.2 Financial Risk Measures

Let us assume that the economic outcomes, i.e., profit and loss, of financial activities can be captured by random variables. Let such a real-valued random variable $X : \Omega \to \mathbb{R}$ that represents the return of certain financial asset over a specific time period be defined on a complete probability space $(\Omega, \mathfrak{U}, P_\Omega)$. Let P denote the probability function of X. Denote by $F_X(\cdot)$ the cumulative distribution function of X such that $F_X(x) = P\{X \leq x\}$ and by $f_X(\cdot)$ the corresponding probability density function. In this chapter we restrict our focus to random variables satisfying the following condition.

Assumption (A0) The cumulative distribution function F_X of the real-valued random variable X is absolutely continuous such that the probability density function f_X w.r.t. the Lebesgue measure exists. f_X is strictly positive and bounded on the

support of X.

Definition 4.1 (Risk measure)**.** *Let $\mathbb{X}$ be a nonempty set of real-valued random variables defined on the probability space $(\Omega, \mathfrak{U}, P_\Omega)$. A mapping $\rho : \mathbb{X} \to \mathbb{R}$ is called a risk measure.*

A risk measure ρ assigns each $X \in \mathbb{X}$ a real number $\rho(X)$ which represents the capital needed for X to guarantee solvency.

4.2.1 Definition of VaR and ES

To define VaR and ES, a careful definition of quantile of a random variable is needed.

Definition 4.2 (Quantile)**.** *Suppose X satisfies **(A0)**. For $\alpha \in (0,1)$, the α-quantile of X is defined as*

$$q_\alpha(X) = F_X^{-1}(\alpha). \tag{4.1}$$

Definition 4.3 (VaR)**.** *Suppose X satisfies **(A0)**. For $\alpha \in (0,1)$, VaR of X at level α, or the α-VaR of X, is defined as*

$$VaR_\alpha(X) = -q_\alpha(X) = q_{1-\alpha}(-X). \tag{4.2}$$

In words, VaR of X at level α is the biggest absolute loss in the best $100(1-\alpha)\%$ situations. Note that $P\{X + \mathrm{VaR}_\alpha(X) < 0\} \leq \alpha$, which means the probability of default is less than α after allocating $\mathrm{VaR}_\alpha(X)$ worth of capital. This is illustrated in Figure 4.2.

Denote by a^- the negative part of a, i.e., $a^- = -a$ for $a < 0$ and $a^- = 0$ for $a \geq 0$. Let $\mathbb{1}_{(A)}$ denote the indicator function: $\mathbb{1}_{(A)} = 1$ when event A occurs and 0 otherwise.

Definition 4.4 (Expected Shortfall)**.** *Assume X satisfies **(A0)** and $E[X^-] < \infty$.*

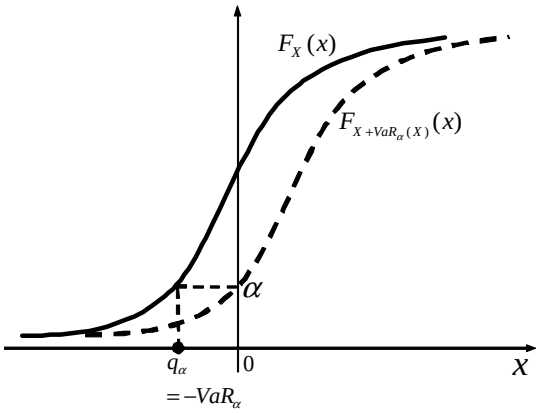


Figure 4.2: $F_{X+VaR_\alpha(X)}(x) = F_X(x - VaR_\alpha(X))$. As a result, the α-level VaR measures the minimum capital required to guarantee default happens with probability no bigger than α: $P\{X + \mathrm{VaR}_\alpha(X) < 0\} = P\{X < -VaR_\alpha(X)\} \leq \alpha$.

Define the Expected Shortfall of X at level $\alpha \in (0,1)$ as

$$ES_\alpha(X) = -\alpha^{-1} E[X \mathbb{1}_{(X \leq q_\alpha(X))}]. \tag{4.3}$$

Note that by the definition of q_α,

$$ES_\alpha(X) = -\frac{E[X \mathbb{1}_{(X \leq q_\alpha)}]}{E[\mathbb{1}_{(X \leq q_\alpha)}]}, \tag{4.4}$$

which is the conditional expected value of $-X$ conditioned on X falling below the α-quantile of X. It is also called Tail Conditional Expectation (TCE or TailVaR [3]). ES is nothing but the expected loss in the worst $100\alpha\%$ of scenarios. A conceptual depiction of VaR and ES is shown in Figure 4.3.

A simple change of variable gives the following representation of ES that associates it with VaR.

$$ES_\alpha(X) = \frac{1}{\alpha} \int_0^\alpha VaR_u(X) du. \tag{4.5}$$

An immediate consequence of expression (4.5) is that ES_α is monotonic decreasing in α: the smaller the level α, the bigger the risk measured by ES_α [2]. As an example,

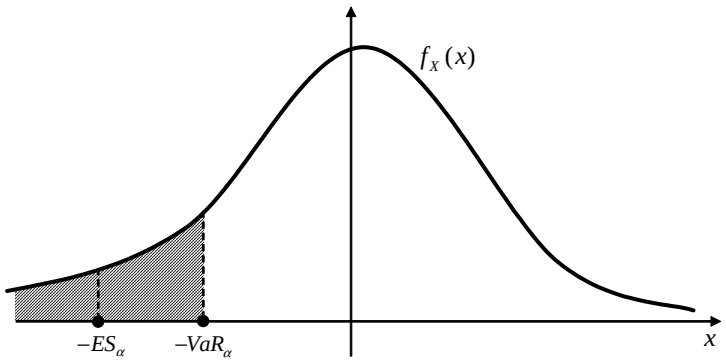


Figure 4.3: A depiction of VaR and ES. The area of the shaded region is α. $-ES_\alpha$ is the conditional mean of the X values in the shaded region.

the manner in which the Expected Shortfall at level α changes as α varies for a standard normally distributed random variable $X \sim \mathcal{N}(0,1)$ is shown in Figure 4.4.

ES is also closely linked to the tick loss function well-studied in the setting of quantile regression [44]. The tick-loss function $L_\alpha(\cdot) : \mathbb{R} \to [0,\infty)$ is defined as $L_\alpha(s) = E[(\mathbb{1}_{(X\leq s)} - \alpha)(s - X)]$. ES can then be represented as

$$ES_\alpha(X) = -E[X] + \frac{1}{\alpha}\min_{s\in\mathbb{R}} L_\alpha(s), \tag{4.6}$$

or equivalently,

$$ES_\alpha(X) = \min_{s\in\mathbb{R}} \alpha^{-1} E\big[\mathbb{1}_{(X\leq s)}(s - X)\big] - s. \tag{4.7}$$

(4.6) enlightens the relation between ES and the regression quantile model proposed by [44] and leads to an easy-to-implement portfolio allocation strategy based on the mean-ES criterion (see [5]) when the distributions of all the assets in the portfolio are available. The right hand side of (4.7) is also called Conditional Value-at-Risk (CVaR) [77, 78]. When ES is represented as the minimum of a convex optimization function, as in (4.7), it is easy to compute and optimize ES under linear portfolio constraints given that all the assets have known distributions.

Although we focus our attention on random variables that meet condition **(A0)**, it is worth mentioning the definitions of VaR and ES for general distributions as some

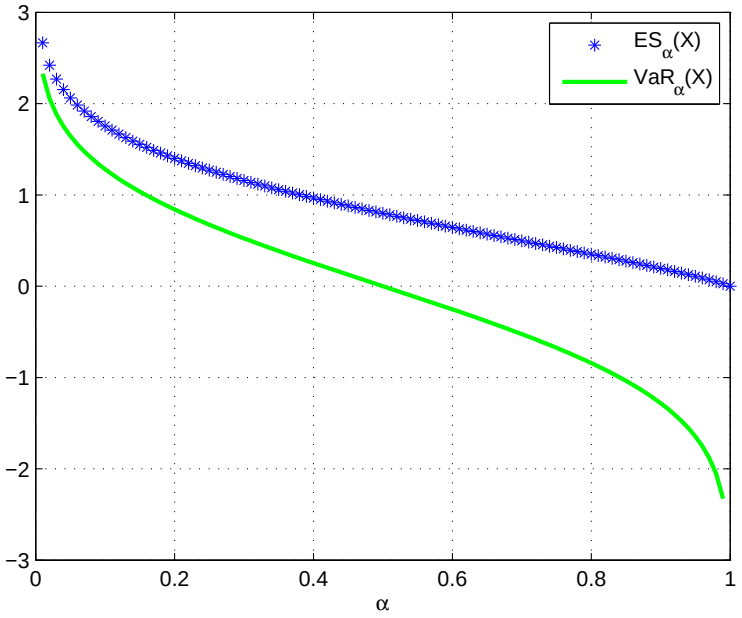


Figure 4.4: $\mathrm{ES}_\alpha(X)$ for $X \sim \mathcal{N}(0,1)$. ES_α is monotonically decreasing as α increases. Note that ES_α reaches 0 when $\alpha = 1$. This is because $\mathrm{ES}_\alpha(X) = E[X]$ when $\alpha = 1$ and $\mathcal{N}(0,1)$ is symmetrically distributed.

of the advantages of ES over VaR manifest themselves for random variables violating **(A0)**.

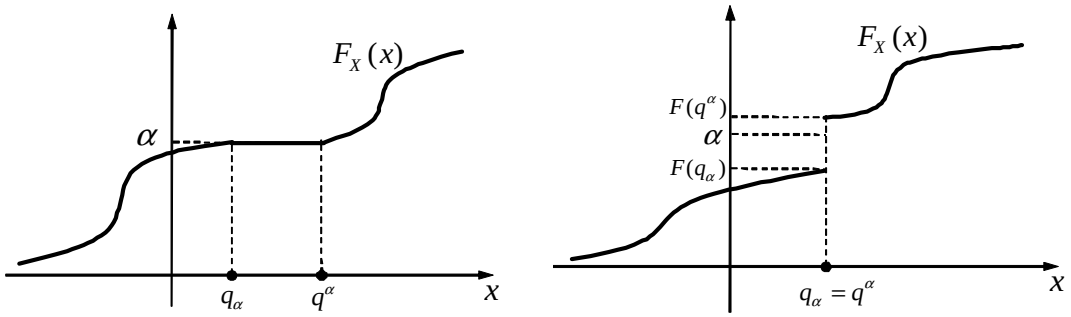
Definition 4.5 (Quantile for general distribution)**.** *For $\alpha \in (0,1)$, the lower α-quantile of X is defined as $q_\alpha(X) = \inf\{x \in \mathbb{R} : F_X(x) \geq \alpha\}$ and the upper α-quantile of X is defined as $q^\alpha(X) = \inf\{x \in \mathbb{R} : F_X(x) > \alpha\}$.*

Definition 4.6 (VaR for general distribution)**.** *For $\alpha \in (0,1)$, VaR of X at level α, or the α-VaR of X, is defined as*

$$VaR_\alpha(X) = -q^\alpha(X) = q_{1-\alpha}(-X). \tag{4.8}$$

Note that when **(A0)** holds, $q_\alpha(X) = q^\alpha(X)$. When **(A0)** does not hold, $q_\alpha(X)$ may or may not equal $q^\alpha(X)$. It is apparent that $q_\alpha(X) \leq q^\alpha(X)$. In Figure 4.5(a) where F_X is continuous but f_X is not strictly positive, $F_X(q_\alpha) = F_X(q^\alpha) = \alpha$ but

$q_\alpha(X) < q^\alpha(X)$; in Figure 4.5(b) where F_X is discontinuous, although $q_\alpha = q^\alpha$, $F_X(q_\alpha) = F_X(q^\alpha) > \alpha$.



(a) (b)

Figure 4.5: Lower and upper quantiles.

Definition 4.7 (Expected Shortfall for general distribution [2])**.** *Assume* $E[X^-] < \infty$. *Define the Expected Shortfall of* X *at level* $\alpha \in (0,1)$ *as*

$$ES_\alpha(X) = -\alpha^{-1}\Big[E[X\mathbb{1}_{\{X\le q_\alpha(X)\}}] + q_\alpha(X)\Big(\alpha - P\{X \le q_\alpha(X)\}\Big)\Big]. \tag{4.9}$$

Expected Shortfall for general distribution is defined as the expected loss in the α-tail. When **(A0)** holds, (4.9) reduces to (4.3). The extra term $\frac{\alpha - P\{X\le q_\alpha(X)\}}{\alpha} q_\alpha(X)$ deals with the case when the probability measure P has an atom at q_α, i.e., $P\{X = q_\alpha\} \neq 0$.

4.2.2 Properties of ES

In this section, we present properties of VaR and ES to demonstrate ES's superiority to VaR. The first set of properties of ES is listed below. For derivation of these properties, refer to [2] .

Proposition 4.1. *Let* ES_α *be defined on the set of random variables* $\mathbb{X}$ *as in (4.3). Fix* $\alpha \in (0,1)$. *It exhibits the following properties:*

(i) translation invariance: $X \in \mathbb{X}$, $a \in \mathbb{R} \Rightarrow ES_\alpha(X+a) = ES_\alpha(X) - a$;

(ii) positive homogeneity: $X \in \mathbb{X}$, $\lambda \geq 0 \Rightarrow ES_\alpha(\lambda X) = \lambda ES_\alpha(X)$;

(iii) monotonicity: $X, Y \in \mathbb{X}$, $X \geq Y \Rightarrow ES_\alpha(X) \leq ES_\alpha(Y)$;

(iv) subadditivity: $X, Y \in \mathbb{X} \Rightarrow ES_\alpha(X+Y) \leq ES_\alpha(X) + ES_\alpha(Y)$;

(v) comonotonic additivity: $X, Y \in \mathbb{X}$ *are comonotonic* $\Rightarrow ES_\alpha(X+Y) = ES_\alpha(X) + ES_\alpha(Y)$;

(vi) law invariance: $X, Y \in \mathbb{X}, P\{X \leq a\} = P\{Y \leq a\}, \forall a \in \mathbb{R} \Rightarrow ES_\alpha(X) = ES_\alpha(Y)$;

(vii) continuity w.r.t. α.

[84] summarized the basic properties of VaR, which we list below.

Proposition 4.2. *Let* VaR_α *be as defined in (4.2) for a fixed* $\alpha \in (0,1)$. *It has the following properties: (i) translation invariance; (ii) positive homogeneity; (iii) monotonicity; (iv) comonotonic additivity; (v) law invariance.*

One important property that is missing for VaR is subadditivity, i.e., $\text{VaR}(X + Y) \leq \text{VaR}(X) + \text{VaR}(Y)$ for $X, Y \in \mathbb{X}$. Subadditivity of a risk measure provides upper bound for the risk of a diversified portfolio and reflects the incentive for diversification in portfolio choice. Several authors have studied the non-subadditivity of VaR. We illustrate it here with a simple example borrowed from [2]. For a more realistic example, see [3] and [31]. Let X_1, X_2 be two random variables defined on $(\Omega, \mathcal{A}, P)$ where there are only three scenarios ω_i, $i = 1, 2, 3$, in Ω with $P(\omega_1) = P(\omega_2) = 0.02$. Suppose $X_i(\omega_j) = -5$ when $i = j$ and 0 otherwise. Then we have $\text{VaR}_{0.05}(X_1) = \text{VaR}_{0.05}(X_2) = 0$ whereas $\text{VaR}_{0.05}(X_1 + X_2) = 5$. If VaR at 5% level is used for monitoring risk, then the fund manager holding X_1 will report no risk, so will the manager holding X_2. However, the bank as a whole holding X_1

and X_2 at the same time does incur positive risk, which creates severe problem for financial institutions.

By the definition introduced by Artzner et al. [3], a risk measure is coherent if it is translation invariant, positively homogeneous, monotonic and subadditive. Therefore, ES is a coherent risk measure, unlike VaR, which fails to be subadditive. Positive homogeneity and subadditivity together imply that ES is convex, i.e., for $X, Y \in \mathbb{X}$, $\alpha \in (0,1)$ and $\lambda \in (0,1)$, $ES_\alpha\big(\lambda X + (1-\lambda)Y\big) \leq \lambda ES_\alpha(X) + (1-\lambda)ES_\alpha(Y)$.

Subadditivity and comonotonic additivity together have a significant practical meaning. Recall that two random variables X and Y defined on $(\Omega, \mathcal{A}, P)$ are comonotonic if there is no pair $\omega_1, \omega_2 \in \Omega$ such that $X(\omega_1) < X(\omega_2)$ and $Y(\omega_1) > Y(\omega_2)$ (see [19]). Suppose X and Y are the random profits from two portfolios. By the argument of subadditivity, the risk of the combined portfolio $X + Y$, if measured by ES_α, is upper bounded by the sum of the individual risks. And from comonotonic additivity, we know that this bound is actually tight when the two portfolios move with the same trend.

It has been pointed out that ES is the smallest coherent and law invariant risk measure that dominates VaR at the same level, see, e.g., [3], [47] and [84]. Essentially, being law invariant means that the risk measure can be estimated from only statistical observations. Recall that VaR has the nice property that $P\{X + VaR_\alpha(X) < 0\} \leq \alpha$. The fact that ES dominates VaR is desirable because it means ES also possess this property: the probability of default is below α after allocating the appropriate capital amount determined by ES. This is apparent since VaR_α is the smallest loss when X falls into the α-tail and ES_α is the average of such loss.

Continuity with respect to α means that a small change in the confidence level will not result in a big change in risk. This insensitivity to α is another advantage over other risk measures including VaR, TCE and WCE [3], which in general can render dramatic change in capital requirement when α undergoes a small change. The fact that VaR is not continuous with α can be seen from Figure 4.5(a).

4.3 Maximum Empirical Likelihood Estimator of ES

4.3.1 Empirical Likelihood Method

Empirical likelihood is a maximum likelihood inference procedure that is based on the empirical likelihood function. To understand the concept, consider a series of independent samples $\{X_i, X_i \in \mathbb{R}^d\}_{i=1}^n$ sharing an unknown distribution function F_0. Denote the corresponding observation of X_i by x_i. Given a series of observations $\{x_i\}_{i=1}^n$, define the empirical likelihood function of a probability distribution function F to be

$$L(F) = \prod_{i=1}^{n} \big(F(x_i) - F(x_i^-)\big), \tag{4.10}$$

where $F(x_i^-) = \lim_{x \nearrow x_i} F(x)$. Clearly, $L(F) = 0$ for continuous distribution functions. The empirical distribution F_n defined as $F_n(x) = \frac{1}{n}\sum_{i=1}^n \mathbb{1}_{(x \geq x_i)}$, where $\mathbb{1}_{(\cdot)}$ is the indicator function as defined before, is the nonparametric maximum likelihood estimate (NPMLE) of F_0 because F_n maximizes $L(F)$.

If we further restrict F to have support on $\{x_i\}_{i=1}^n$, i.e., F assigns probability p_i to x_i with $p_i \geq 0$, $\sum_{i=1}^n p_i = 1$, the empirical log likelihood function can be written as

$$l_{NP}(p_1, \cdots, p_n) = \sum_{i=1}^{n} \ln p_i. \tag{4.11}$$

This restriction is justified since the maximization of $l_{NP}(\cdot)$ automatically pushes towards the support specified by the data.[2] l_{NP} is maximized when $p_i = 1/n$, for $i = 1, \cdots, n$. Then the empirical log likelihood ratio function is

$$l_R(p_1, \cdots, p_n) = \sum_{i=1}^{n} \ln np_i. \tag{4.12}$$

[2]Suppose F places probability p_i on x_i and $L(F) \geq r_0/n^n$. Lemma 2.1 in [62] shows that $1 - \sum_{i=1}^n p_i \leq (1/n) log(1/r_0)$, which means as $L(F)$ approaches the maximum $(1/n)^n$, the probability that F puts on the sample goes to 1.

Empirical likelihood works well for estimation of parameters in moment conditions. Let $g(x_i, \theta) : \mathbb{R}^d \times \Theta \to \mathbb{R}^s$ be a vector of measurable functions, where $\Theta \subset \mathbb{R}^r$. Consider the moment condition expressed as

$$E_0[g(x_i, \theta_0)] = 0, \tag{4.13}$$

where E_0 denotes expectation w.r.t. the true probability distribution F_0 and $\theta_0 \in \Theta$ is the unknown true value of the parameter that we wish to estimate. In the case of estimating the mean of F, $g(x, \theta) = x - \theta$ and θ is the mean parameter. When $g(x, \theta) = \mathbb{1}_{(x \leq \theta)} - \alpha$, θ is the parameter of α-quantile of F. Maximum empirical likelihood estimators (MELE) of θ_0 and F_0 are obtained by maximizing l_R under the moment constraint. That is, $(\hat{\theta}_{EL}, \hat{p}_{1El}, \cdots, \hat{p}_{nEL})$ is the optimal solution to the following constrained optimization problem

$$\sup_{\theta \in \Theta, p_1, \cdots, p_n} \sum_{i=1}^{n} \ln np_i \quad \text{s.t.} \quad p_i \geq 0, i = 1, \cdots, n, \quad \sum_{i=1}^{n} p_i = 1, \quad \sum_{i=1}^{n} p_i g(x_i, \theta) = 0. \tag{4.14}$$

Here and afterwards, the implicit dependence of estimates on the sample size n is omitted from notation for simplicity. Asymptotic properties tell us how the estimates behave when the sample size increases.

The dimension of the optimization variable in (4.14) increases with the data size, posing difficulty to its computation. Fortunately, it can be solved in two stages. First fix $\theta \in \Theta$ and solve (4.14) to get the profile empirical log likelihood ratio function $l(\theta)$ for θ. Since $\{p_i\}_{i=1}^n$ is restricted to be on the standard n-simplex, the supremum in (4.14) is achieved for fixed θ.

$$l(\theta) = \max_{p_1, \cdots, p_n} \sum_{i=1}^{n} \ln np_i \quad \text{s.t.} \quad p_i \geq 0, i = 1, \cdots, n, \quad \sum_{i=1}^{n} p_i = 1, \quad \sum_{i=1}^{n} p_i g(x_i, \theta) = 0. \tag{4.15}$$

$\hat{\theta}_{EL}$ is just the parameter that optimizes this profile ratio function, i.e.,

$$\hat{\theta}_{EL} = \arg \sup_{\theta \in \Theta} l(\theta). \tag{4.16}$$

Here and in the remaining of the chapter, we use "argsup" or "argmax" to mean any maximizer of the objective function and allow a slight abuse of notation when the maximizer is not unique.

The optimization problem (4.15) is convex. By Lagrangian multiplier method, its solution is easily obtained:

$$\hat{p}_i(\theta) = \frac{1}{n} \frac{1}{1 + \lambda(\theta)^T g(x_i, \theta)}, \tag{4.17}$$

where $\lambda(\theta)$ is the Lagrangian multiplier for the moment constraint $\sum_{i=1}^n p_i g(x_i, \theta) = 0$ and satisfies

$$\sum_{i=1}^n \frac{g(x_i, \theta)}{1 + \lambda(\theta)^T g(x_i, \theta)} = 0. \tag{4.18}$$

That is,

$$\lambda(\theta) = \arg\max_{\lambda} \sum_{i=1}^n \ln\Big(1 + \lambda^T g(x_i, \theta)\Big). \tag{4.19}$$

In general, $\lambda(\theta)$ can not be solved analytically. Therefore, from (4.15)–(4.19),

$$\hat{\theta}_{EL} = \arg \sup_{\theta \in \Theta} l(\theta) = \arg \sup_{\theta \in \Theta} \min_{\lambda} -\sum_{i=1}^n \ln\Big(1 + \lambda^T g(x_i, \theta)\Big). \tag{4.20}$$

Note that although the inner optimization with respect to λ is convex, the outer optimization with respect to θ in general is not convex, depending on the format of g. Hence the implementation of MELE suffers from difficulties associated with solving non-convex optimization problems.

Without posing the last constraint, the optimal solution to (4.15) is $\tilde{p}_i = 1/n$, for all $i = 1, \cdots, n$. In the case of mean estimation where $g(x, \theta) = x - \theta$, this maximum is achieved by setting $\theta = \bar{X} = \frac{1}{n}\sum_{i=1}^n x_i$ such that $\sum_{i=1}^n g(x_i, \theta) = \sum_{i=1}^n (x_i - \bar{X}) = 0$.

Therefore the MELE of the mean of F_0 is $\bar{X}$.

4.3.2 Estimator for Expected Shortfall

In this section, the MELE of ES for the marginal distribution F_X of the real-valued random variable X is developed. As before, let X_i, $i = 1, \cdots$, denote the i.i.d. samples of X and let x_i be the observation of X_i. Fix $\alpha \in (0, 1)$. Assume condition **(A0)** holds for X such that its quantile and ES at level α are as defined in (4.1) and (4.3), respectively. Denote by q_0 and ξ_0 the true α-quantile and conditional expectation of X below the α-quantile, i.e., $q_0 = F_X^{-1}(\alpha)$ and $\xi_0 = E[X|X \leq q_0]$. VaR and ES of X at level α are simply $VaR_\alpha(X) = -q_0$, $ES_\alpha(X) = -\xi_0$. The dependence of q_0 and ξ_0 on α is suppressed for simplicity in notation.

To estimate ξ_0, the moment condition is expressed as $E[g(X, \theta_0)] = 0$, where $\theta = (q, \xi)^T \in \Theta \subset \mathbb{R}^2$ and $g : \mathbb{R} \times \Theta \to \mathbb{R}^2$ is a vector of functions $g = (g_1, g_2)^T$ with

$$g_1(x, \theta) = \mathbb{1}_{(x \leq q)} - \alpha, \tag{4.21}$$

$$g_2(x, \theta) = \mathbb{1}_{(x \leq q)}(x - \xi). \tag{4.22}$$

We make the following assumption on Θ.

Assumption (A1) Θ is a compact set and θ_0 is in the interior of Θ.

From (4.14)–(4.16), the MELE for the true parameter $\theta_0 = (q_0, \xi_0)^T$ is

$$\begin{aligned}(\hat{q}_{EL}, \hat{\xi}_{EL}) = \arg \max_{(q,\xi)\in\Theta} \max_{\{p_i\}_{i=1}^n} & \sum_{i=1}^n \ln np_i \\ \text{s.t. (i)} \quad & p_i \geq 0, \quad \text{for } i = 1, \cdots, n \\ \text{(ii)} \quad & \sum_{i=1}^n p_i = 1, \\ \text{(iii)} \quad & \sum_{i=1}^n p_i[\mathbb{1}_{(x_i \leq q)} - \alpha] = 0, \\ \text{(iv)} \quad & \sum_{i=1}^n p_i[\mathbb{1}_{(x_i \leq q)}(x_i - \xi)] = 0.\end{aligned} \tag{4.23}$$

We mentioned in Section 4.3.1 that the implementation of MELE can be less straightforward in general. Fortunately, the estimator for θ_0 defined in (4.23) can be solved in a sequential fashion. To see this, consider the MELE for estimating just q_0. Let $\mathbb{Q} := \{s : (s,t) \in \Theta, t \in \mathbb{R}\}$, $\Xi := \{t : (s,t) \in \Theta, s \in \mathbb{R}\}$. The MELE $\hat{q}_{EL}$ of the α-quantile q_0 of X is defined as the maximizer of the profile empirical likelihood ratio function

$$\begin{aligned} \hat{q}_{EL} = \arg \max_{q \in \mathbb{Q}} \max_{\{p_i\}_{i=1}^n} & \sum_{i=1}^n \ln np_i \\ \text{s.t.} \quad \text{(i)} \quad & p_i \geq 0, \quad \text{for } i = 1, \cdots, n \\ \text{(ii)} \quad & \sum_{i=1}^n p_i = 1, \\ \text{(iii)} \quad & \sum_{i=1}^n p_i[\mathbb{1}_{(x_i \leq q)} - \alpha] = 0. \end{aligned} \tag{4.24}$$

Notice that (4.24) and (4.23) share all but one constraints. The latter has one extra constraint $\sum_{i=1}^n p_i[\mathbb{1}_{(x_i \leq q)}(x_i - \xi)] = 0$. This constraint can be written as

$$\xi = \frac{\sum_{i=1}^n p_i \mathbb{1}_{(x_i \leq q)} x_i}{\sum_{i=1}^n p_i \mathbb{1}_{(x_i \leq q)}}, \tag{4.25}$$

such that for each combination of $q \in \mathbb{Q}$ and $\{p_i\}_{i=1}^n$, there exists a unique $\xi \in \Xi$ satisfying the constraint. Hence the optimizer of (4.24) and (4.23) are linked through the following lemma.

Lemma 4.1. *Fix $\alpha \in (0,1)$. For a sample $\{x_1, \cdots, x_n\}$, let $\mathbb{Q} = \Xi$ be defined as the convex hull of $\{x_1, \cdots, x_n\}$. $(p_1^*, \cdots, p_n^*, q^*)$, $q^* \in \mathbb{Q}$, is an optimal solution to the optimization problem in (4.24) if, and only if, there exists $\xi^* \in \Xi$ such that $(p_1^*, \cdots, p_n^*, q^*, \xi^*)$ is an optimizer of the optimization problem in (4.23).*

Proof. Before we proceed, observe that if $(p_1^*, \cdots, p_n^*, q^*, \xi^*)$ is a feasible solution to (4.23), then so is $(p_1^*, \cdots, p_n^*, q^*)$ to (4.24).

First we prove necessity. Let $(p_1^*, \cdots, p_n^*, q^*)$ be an optimal solution to the optimization problem in (4.24). Let $\xi^* = \sum_{i=1}^n p_i^* \mathbb{1}_{(x_i \leq q^*)} x_i / \alpha$, then $(p_1^*, \cdots, p_n^*, q^*, \xi^*)$

is a feasible solution to (4.23). Now suppose there exists another feasible solution $(\bar{p}_1, \cdots, \bar{p}_n, \bar{q}, \bar{\xi})$ to (4.23) such that $\sum_{i=1}^n \ln \bar{p}_i > \sum_{i=1}^n \ln p_i^*$. Since $(\bar{p}_1, \cdots, \bar{p}_n, \bar{q})$ is a feasible solution to (4.24), this leads to contradiction.

Sufficiency is proved similarly. Let $(p_1^*, \cdots, p_n^*, q^*, \xi^*)$ be an optimal solution to (4.23). Then $(p_1^*, \cdots, p_n^*, q^*)$ is a feasible solution to (4.24). Suppose there exists a feasible solution $(\bar{p}_1, \cdots, \bar{p}_n, \bar{q})$ to (4.24) such that $\sum_{i=1}^n \ln \bar{p}_i > \sum_{i=1}^n \ln p_i^*$. Then $(\bar{p}_1, \cdots, \bar{p}_n, \bar{q}, \bar{\xi})$, where $\bar{\xi} = \sum_{i=1}^n \bar{p}_i \mathbb{1}_{(x_i \le \bar{q})} x_i/\alpha$, is a feasible solution to (4.23) with $\sum_{i=1}^n \ln \bar{p}_i > \sum_{i=1}^n \ln p_i^*$. This contradicts the optimality of $(p_1^*, \cdots, p_n^*, q^*, \xi^*)$ to (4.23). □

A consequence of Lemma 4.1 is that the optimization problem in (4.23) can be solved sequentially. That is, we can first solve (4.24) to obtain $\hat{q}_{EL}$ and then compute $\hat{\xi}_{EL}$ algebraically as

$$\hat{\xi}_{EL} = \frac{1}{\alpha} \sum_{i=1}^n \hat{p}_i(\hat{q}_{EL}) \mathbb{1}_{(x_i \le \hat{q}_{EL})} x_i = \frac{1}{n} \frac{1}{I(\hat{q}_{EL})} \sum_{i=1}^n \mathbb{1}_{(x_i \le \hat{q}_{EL})} x_i. \tag{4.26}$$

Observe that $nI(\hat{q}_{EL})$ is exactly the number of observations that is smaller or equal to the estimated quantile $\hat{q}_{EL}$. Hence (4.26) is nothing but the sample average of the truncated data truncated at the quantile estimate $\hat{q}_{EL}$.

We now look at the computation of $\hat{q}_{EL}$. Note that constraint (iii) in (4.24) requires q to be inside the convex hull generated by observed data $\{x_1, \cdots, x_n\}$. That is, if we rearrange the observations into $\{x_{(i)}\}_{i=1}^n$ such that $x_{(1)} \le x_{(2)} \le \cdots \le x_{(n)}$, $[x_{(1)}, x_{(n)}]$ is the feasible range for q.

Let us first fix a $q \in \mathbb{Q} \cap [x_{(1)}, x_{(n)}]$ and consider the inner optimization with respect to $\{p_i\}_{i=1}^n$ as in Section 4.3.1. Following the general procedure in Section 4.3.1, for each $q \in \mathbb{Q} \cap [x_{(1)}, x_{(n)}]$, the maximizer of (4.24) can be obtained through Lagrange multiplier method as

$$\hat{p}_i(q) = \frac{1}{n} \frac{1}{1 + \lambda(q)[\mathbb{1}_{(x_i \le q)} - \alpha]}, \tag{4.27}$$

where $\lambda(q)$ is the Lagrange multiplier of constraint (iii) in (4.24). In the case of general moment constraints, such as mean, the Lagrangian multiplier can not be expressed analytically and has to be obtained through numerical computation. However, for the case of quantile, the analytical form of $\lambda(q)$ is available as

$$\lambda(q) = \frac{\sum_{i=1}^{n} \frac{1}{n}[\mathbb{1}_{(x_i \le q)} - \alpha]}{\alpha(1-\alpha)} := \frac{I(q) - \alpha}{\alpha(1-\alpha)}, \tag{4.28}$$

where $I(q) := \sum_{i=1}^{n} \frac{1}{n}\mathbb{1}_{(x_i \le q)}$ is the proportion of observations that fall below q [48]. Observe that $I(q) = i/n$ when $x_{(i)} \le q < x_{(i+1)}$ for some $i \in \{1, \cdots, n-1\}$, and $I(x_{(n)}) = 1$.

Taking (4.28) into (4.27) gives

$$\hat{p}_i(q) = \begin{cases} \frac{1}{n}\frac{\alpha}{I(q)} & , \quad \text{for } i : x_i \le q, \\ \frac{1}{n}\frac{1-\alpha}{1-I(q)} & , \quad \text{for } i : x_i > q. \end{cases} \tag{4.29}$$

Therefore the profile empirical likelihood ratio function $l(q)$ resulting from (4.24) does not depend on either $\{p_i\}_{i=1}^{n}$ or $\lambda(q)$:

$$l(q) = -n\Big(I(q)\ln\frac{I(q)}{\alpha} + \big(1 - I(q)\big)\ln\frac{1-I(q)}{1-\alpha}\Big).$$

Note that $l(x_{(n)}) = n\ln\alpha$. The MELE estimator $\hat{q}_{EL}$ of q_0 is

$$\hat{q}_{EL} = \arg\max_{q \in \mathbb{Q} \cap [x_{(1)}, x_{(n)}]} l(q). \tag{4.30}$$

A couple of words on the computation of $\hat{q}_{EL}$ are in order. Difficulty arises when solving (4.30) for $\hat{q}_{EL}$ since the profile ratio function $l(q)$ is discontinuous. Fortunately, an easy computation of $\hat{q}_{EL}$ exists due to the special structure of $l(q)$. It is interesting to observe that in (4.30), q affects $l(q)$ only through $I(q)$. $l(q)$ is strictly concave in $I(q)$ and is maximized at $I(q) = \alpha$. As a result, $\hat{q}_{EL}$ can be computed in the following way:

- When $\alpha < \min\{I(q) : q \in \mathbb{Q} \cap [x_{(1)}, x_{(n)}]\}$, $\hat{q}_{EL} = \min\{q : q \in \mathbb{Q} \cap [x_{(1)}, x_{(n)}]\}$;
- Otherwise, if $\alpha \geq \max\{I(q) : q \in \mathbb{Q} \cap [x_{(1)}, x_{(n)}]\}$, $\hat{q}_{EL} = \max\{q : q \in \mathbb{Q} \cap [x_{(1)}, x_{(n)}]\}$;
- Otherwise, find $j \in \mathbb{N}$, $1 \leq j \leq n-1$, such that $I(x_{(j)}) \leq \alpha < I(x_{(j+1)})$ and $x_{(j)}, x_{(j+1)} \in \mathbb{Q} \cap [x_{(1)}, x_{(n)}]$, then $\hat{q}_{EL} = x_{(k)}$, where $k \in \{j, j+1\}$ is such that $l(x_{(k)}) = \min\{l(x_{(j)}), l(x_{(j+1)})\}$.

4.4 Asymptotic Properties

For each series of sample observations $x_1, \cdots, x_n$, the MELE produces an estimate $(\hat{q}_{EL}, \hat{\xi}_{EL})$. Since $x_1, \cdots, x_n$ are random, the estimate is a random variable for each sample size n. Recall that an estimator $\hat{\beta}$ for some true parameter value β_0 is called consistent if $\hat{\beta}$ converges to β_0 in probability as the sample size n goes to infinity, denoted as $\hat{\beta} \xrightarrow{p} \beta_0$. That is, for every $\varepsilon > 0$, $\delta > 0$, there exists $N_0 \in \mathbb{N}$ such that $P\{\|\hat{\beta}_n - \beta_0\| < \varepsilon\} > 1 - \delta$ for all $n > N_0$. The estimator is asymptotically normal if for some increasing function $v(n)$, usually taken to be $\sqrt{n}$, $v(n)(\hat{\beta} - \beta_0)$ converges in distribution to a zero mean Gaussian distribution as $n \to \infty$, i.e., $v(n)(\hat{\beta} - \beta_0) \xrightarrow{d} \mathcal{N}(0, V)$. V is called the asymptotic variance matrix of the estimator $\hat{\beta}$. Let $\bar{\beta}$ be another consistent and asymptotically normal estimator of β_0 with asymptotic variance matrix U. If $U - V$ is positive semidefinite, we say that $\hat{\beta}$ is more efficient than $\bar{\beta}$. When $\hat{\beta}$ is estimated through some moment conditions, it is called asymptotically efficient if its asymptotic variance matrix is at least as small as that of any other estimator obtained by the same set of moment conditions [37].

MELE for smooth moment functions is shown to be consistent and asymptotically efficient for i.i.d. data under certain regularity conditions [72]. That is, as the sample size goes to infinity, its asymptotic variance matrix achieves the semi-parametric lower bound proved by Chamberlian [14]. Unfortunately, the smoothness condition is not

satisfied by the MELE defined in (4.23). We use the results in [95], which provide asymptotic properties for MELE with possibly discontinuous moment constraint. The following assumption is needed to establish consistency and asymptotic normality of the MELE $(\hat{q}_{EL}, \hat{\xi}_{EL})$ in (4.23).

Assumption (A2) There exists $\gamma_0 > 4$ such that $E[|X|^{\gamma_0}] < \infty$.

Theorem 4.1. *Suppose the conditions in* ***(A0)****–****(A2)*** *hold. Let MELE* $(\hat{q}_{EL}, \hat{\xi}_{EL})$ *be defined as in (4.23), then*

(i) $\hat{\theta} \xrightarrow{p} \theta_0$*;*

(ii) $\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} \mathcal{N}(0, V)$*, where*

$$V = \begin{bmatrix} \dfrac{\alpha(1-\alpha)}{f_X(q_0)^2} & \dfrac{(1-\alpha)(q_0-\xi_0)}{f_X(q_0)} \\ \dfrac{(1-\alpha)(q_0-\xi_0)}{f_X(q_0)} & \dfrac{E[\mathbb{1}_{(X\le q_0)}(X-\xi_0)^2]}{\alpha^2} + \dfrac{1-\alpha}{\alpha}(q_0-\xi_0)^2 \end{bmatrix}.$$

The asymptotic variance for $\hat{q}_{EL}$ is $\frac{\alpha(1-\alpha)}{f_X(q_0)^2}$, which agrees with the optimal variance matrix for quantile estimation proved in [58]. The asymptotic variance for $\hat{\xi}_{EL}$ is

$$\frac{E[\mathbb{1}_{(X\le q_0)}(X-\xi_0)^2]}{\alpha^2} + \frac{1-\alpha}{\alpha}(q_0-\xi_0)^2,$$

which increases as α decreases. This means the more to the tail, the more data are needed to estimate VaR and ES. Note also that the asymptotic variance of $\hat{\xi}_{EL}$ is bigger when the tail of the distribution is flatter.

4.5 Simulation Results

In this section, small sample performance of the MELE $(\hat{q}_{EL}, \hat{\xi}_{EL})$ in estimating $\mathrm{VaR}_\alpha(X)$ and $\mathrm{ES}_\alpha(X)$ of a random variable X is studied by a Monte Carlo experiment. The level α is taken to be 0.01, 0.05 and 0.1 as these are the α levels

of most interest in practice. For each value of α, 1000 replications of i.i.d. sample $\{x_1, \cdots, x_n\}$ are generated. Sample size n is chosen to be 100, 500, 1000, 5000. For each sample, estimate of VaR and ES are obtained from the MELE: $\hat{VaR}_\alpha = -\hat{q}_{EL}$, $\hat{ES}_\alpha = -\hat{\xi}_{EL}$.

The following tables report the mean value of VaR and ES estimates for X with three different distributions. For ease of comparison, the VaR estimates are collected in Table 4.1 and ES estimates are shown in Table 4.2. Distributions of X are taken to be members of Asymmetric Power Distribution (APD) family, i.e., $X \sim APD(\beta, \lambda)$. $APD(0.5, 2)$ is the normal distribution $\mathcal{N}(0, 1/2)$. $\lambda = 1$ corresponds to a Laplace distribution. The probability density function of the three distributions is shown in Figure 4.6.

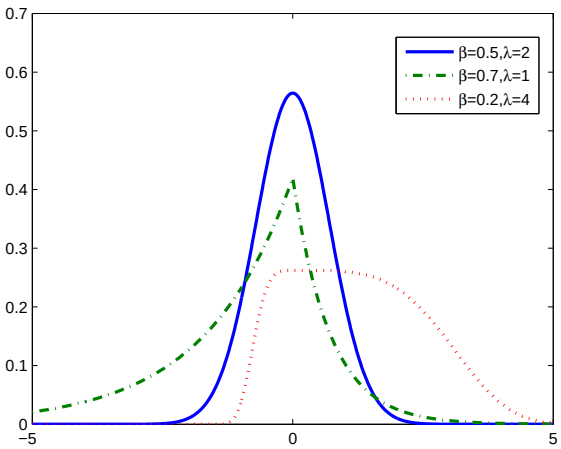


Figure 4.6: Probability density function of APD distributions. Solid line: $\mathcal{N}(0, 1/2)$; Dash-dotted line: asymmetric Laplace distribution

As expected, mean values of MELEs converge to the true value for both VaR and ES estimates at different α levels and for different distributions. Note that for the same distribution, the estimate tends to converge faster for bigger α values. This agrees with our intuition because more data is needed to gain information about lower α tail of the distribution for smaller α.

β	λ	α	VaR_α	$N = 100$	$N = 500$	$N = 1000$	$N = 5000$
0.5	2	0.01	1.6450	1.7801	1.6639	1.6558	1.6469
0.5	2	0.05	1.1631	1.1840	1.1700	1.1684	1.1643
0.5	2	0.1	0.9062	0.9221	0.9103	0.9085	0.9055
0.7	1	0.01	7.0808	8.0969	7.2223	7.1536	7.0949
0.7	1	0.05	4.3984	4.5356	4.4388	4.4276	4.4048
0.7	1	0.1	3.2432	3.3288	3.2641	3.2545	3.2406
0.2	4	0.01	0.8828	0.9343	0.8897	0.8870	0.8836
0.2	4	0.05	0.6014	0.6074	0.6046	0.6043	0.6021
0.2	4	0.1	0.3848	0.3914	0.3869	0.3861	0.3839

Table 4.1: Mean value of MELEs for VaR of APD-distributed random variables

β	λ	α	ES_α	$N = 100$	$N = 500$	$N = 1000$	$N = 5000$
0.5	2	0.01	1.8846	1.7801	1.8563	1.8684	1.8825
0.5	2	0.05	1.4586	1.4184	1.4538	1.4565	1.4574
0.5	2	0.1	1.2410	1.2218	1.2374	1.2398	1.2392
0.7	1	0.01	8.7475	8.0969	8.5704	8.6415	8.7343
0.7	1	0.05	6.0651	5.8676	6.0437	6.0546	6.0578
0.7	1	0.1	4.9099	4.8313	4.8948	4.9054	4.9010
0.2	4	0.01	0.9857	0.9343	0.9721	0.9782	0.9847
0.2	4	0.05	0.7741	0.7453	0.7703	0.7726	0.7734
0.2	4	0.1	0.6317	0.6126	0.6282	0.6304	0.6304

Table 4.2: Mean value of MELEs for ES of APD-distributed random variables

4.6 Conclusions

In this chapter, we have studied the coherent risk measure, Expected Shortfall, as an alternative to the popular measure Value-at-Risk. An estimator based on empirical likelihood is proposed and its asymptotic properties are established. Although in general maximum empirical likelihood estimators are obtained by solving non-convex optimization problems, the proposed estimator for ES can be computed in two steps. The first step involves comparison of two function evaluations and the second step is just an algebraic calculation.

Chapter 5

Conclusions and Future Directions

In this thesis, we have studied three aspects of robust design process depicted in Figure 1.1: relation between robustness and complexity, model validation and risk assessment. By examining percolation lattices and linear programming problems, we show that robust systems necessarily have short proofs and the nonexistence of a short proof indicates fragility in the system. Therefore the existence of the feedback information in Figure 1.1 is established. Model validation methods are examined in the setting of a biological system to illustrate their associated advantages and difficulties. Risk assessment is investigated in the field of finance where the proper mathematical definition of risk is emerging. An estimator of Expected Shortfall is proposed and is shown to have desirable asymptotic properties.

Following the study in this thesis, several research directions can be further pursued. We outline them here.

While together with two other works [33, 63], we have established the existence of feedback in Figure 1.1, it is not yet clear how this feedback can more directly guide system design. When verification is inevitably hard, we know the system property is fragile. As mentioned in Section 2.4, this fragility can come from different sources, including, but not limited to, system specification, modeling and design. It would be extremely enlightening if the complexity of verification could indicate the source and/or the properties of the fragility. One possible approach is to look at different perspectives of the long proof. For example, when the proof is a polynomial, the

order of the polynomial and the magnitude of the coefficients may lead us to different source of fragility. Relaxation is of course useful in exploring fragility with respect to assumptions.

For model validation, it would be interesting to be able to validate a model not just against specific data, but against some overall properties of the system, e.g., the system is stable. This would allow qualitative observations to be included. For linear systems, this is relatively easy because properties such as stability can expressed as a function of its parameters. A starting point can be nonlinear systems with special forms, e.g., systems with polynomial vector fields.

In Chapter 4, a maximum empirical likelihood estimator for ES of a random variable is proposed for i.i.d. data. Several immediate future directions follow. First, it would be interesting to combine the proposed estimator with time series models, such as the GARCH model, to examine how it performs on real financial data. Second, an estimator is needed when the data is not i.i.d., e.g., when data are heterogeneously distributed or when they are dependent or both. Third, we can study parameters in the conditional ES models when the data is still i.i.d. and explanatory variables are present.

Appendix A

Proof of Lemma 2.2

Before proving Lemma 2.2, the following two lemmas are needed. As mentioned earlier, positive scaling of $\bar{b}$ does not affect the feasibility of $A2$. Thus the following lemma is true.

Lemma A.1. *Suppose the (A, b) pair given is such that A2 is consistent for $\alpha > 0, \beta > 0$. Then perturbation on only α and β does not change the feasibility. That is, for any $\Delta\alpha \in \mathbb{R}^n$ and $\Delta\beta \in \mathbb{R}$, there exist $y \in \mathbb{R}^m$, $y \geq 0$ such that $A^T y \geq (\alpha + \Delta\alpha)$, $b^T y \leq -(\beta + \Delta\beta)$.*

Proof. Let y_0 be a solution to $A2(A, b)$, i.e.,

$$\begin{aligned} A^T y_0 &\geq \alpha \\ b^T y_0 &\leq -\beta \\ y_0 &\geq 0 \end{aligned}$$

Denote the i^{th} component of α by α_i and the i^{th} component of $\alpha + \Delta\alpha$ by $(\alpha + \Delta\alpha)_i$. Let $k = \max\{1, \frac{(\alpha+\Delta\alpha)_i}{\alpha_i}, \frac{\beta+\Delta\beta}{\beta}\}$. We have that

$$\begin{aligned} k\alpha &\geq \alpha + \Delta\alpha \\ k\beta &\geq \beta + \Delta\beta \end{aligned}$$

Thus $y_1 = ky_0 \geq 0$ satisfy

$$\begin{aligned} A^T y_1 &\geq k\alpha \geq \alpha + \Delta\alpha \\ b^T y_1 &\leq -k\beta \leq -(\beta + \Delta\beta) \end{aligned}$$

□

Lemma A.2. *$A2(A,b)$ is defined as in (2.7) with $\|\bar{b}\| = \|b\|$. Suppose d is well-posed and $A2(A,b)$ is feasible, then there exists a solution y to $A2(A,b)$ such that*

$$\|y\| \leq \frac{\|b\|}{dist(d, \mathcal{B})}.$$

where $d = (A,b)$ and $dist(d,\mathcal{B})$ is as defined in section 2.3.2.

Proof. Applying Theorem 1.1(1) in [74] gives

$$\|y\| \leq \frac{\|\bar{b}\|}{\xi}$$

where

$$\begin{aligned} \xi &= \inf_{Fd(\bar{A}+\Delta\bar{A},\bar{b}+\Delta\bar{b}) \text{ infeasible}} \{\max\{\|\Delta\bar{A}\|, \|\Delta\bar{b}\|\}\} \\ &= \inf_{Fd(\bar{A}+\Delta\bar{A},\bar{b}) \text{ infeasible}} \{\|\Delta\bar{A}\|\} \\ &\geq \inf_{A2(A+\Delta A,b+\Delta b) \text{ infeasible}} \{\max\{\|\Delta A\|, \|\Delta b\|\}\} \end{aligned}$$

The second equality follows from Lemma A.1 and the inequality is from the norm definition, which gives

$$\|\Delta\bar{A}\| = \left\| \begin{bmatrix} -\Delta A \\ \Delta b \end{bmatrix} \right\| \geq \max\{\|\Delta A\|, \|\Delta b\|\}.$$

By Lemma 2.1, if $A2(A+\Delta A, b+\Delta b)$ is infeasible, then either $Fd(A,b)$ is feasible

or d is on the boundary $\mathcal{B}$. Therefore,

$$\begin{aligned} \xi &\geq \inf_{A2(A+\Delta A, b+\Delta b) \text{ infeasible}} \{\max\{\|\Delta A\|, \|\Delta b\|\}\} \\ &= \inf_{Fd(A+\Delta A, b+\Delta b) \text{ feasible}} \{\max\{\|\Delta A\|, \|\Delta b\|\}\} \\ &= dist(d, \mathcal{B}) \end{aligned}$$

This completes the proof that

$$\|y\| \leq \frac{\|b\|}{dist(d, \mathcal{B})}.$$

□

Proof of Lemma 2.2. By definition, $C_a \leq \|y\|$ for any solution y to $A2(A, b)$. The conclusion then follows from Lemma A.2. □

Appendix B

Proof of Theorem 4.1

Recall that X is the random variable for which we wish to estimate quantile and ES. F_X and f_X are the cumulative and probability distribution function of X, respectively. The moment constraint function $g = (g_1, g_2)^T : \mathbb{R} \times \Theta \to \mathbb{R}$ is

$$\begin{aligned} g_1(x, \theta) &= \mathbb{1}_{(x \leq q)} - \alpha \\ g_2(x, \theta) &= \mathbb{1}_{(x \leq q)}(x - \xi), \end{aligned}$$

where $\theta = (q, \xi)^T$.

Let $r_0 > 0$ be a fixed constant. Before proving Theorem 4.1, the following lemmas are needed.

From assumption **(A2)**, there exists $\gamma_0 > 4$ such that $E[|X|^{\gamma_0}] < \infty$. Then for any $0 \leq \gamma < \gamma_0$, $E[|X|^{\gamma}] < \infty$. This follows from the fact that $|x|^{\gamma} \leq |x|^{\gamma_0} + 1$ for all $x \in \mathbb{R}$. A consequence of this is that $E[(|X| + \eta)^{\gamma}] < \infty$ for bounded constant η. This is used several times in proving the lemmas below.

Lemma B.1. *For $k = 1, 2$,*

$$E\left[\sup_{\|\theta - \theta_0\| \leq r_0} |g_k(x, \theta)|^{\gamma_0}\right] < \infty.$$

Proof.

$$\sup_{\|\theta-\theta_0\|\le r_0} |g_1(x,\theta)|^{\gamma_0} = \sup_{\|\theta-\theta_0\|\le r_0} |\mathbb{1}_{(x\le q)} - \alpha|^{\gamma_0} \le 1.$$

For a fixed x, when $\|\theta - \theta_0\| \le r_0$,

$$|\mathbb{1}_{(x\le q)}(x-\xi)| \le |x-\xi| \le |x-\xi_0| + r_0.$$

Therefore,

$$E\left[\sup_{\|\theta-\theta_0\|\le r_0} |g_2(x,\theta)|^{\gamma_0}\right] \le E\left[(|X-\xi_0| + r_0)^{\gamma_0}\right] \le E\left[(|X| + |\xi_0| + r_0)^{\gamma_0}\right] < \infty.$$

□

Lemma B.2. *For $k = 1, 2$,*

$$\sup_{\|\theta-\theta_0\|\le r_0} E\left[g_k(x,\theta)^4\right] < \infty.$$

Proof.

$$\sup_{\|\theta-\theta_0\|\le r_0} E\left[g_1(X,\theta)^4\right] = \sup_{\|\theta-\theta_0\|\le r_0} E\left[(\mathbb{1}_{(X\le q)} - \alpha)^4\right] \le \sup_{\|\theta-\theta_0\|\le r_0} E[1] = 1.$$

Since

$$|\mathbb{1}_{(x\le q)}(x-\xi)| \le |x| + |\xi_0| + r_0, \qquad \text{for } x \in \mathbb{R}, \|\theta-\theta_0\| \le r_0,$$

$$E\left[(\mathbb{1}_{(X\le q)}(X-\xi))^4\right] \le E\left[(|x| + |\xi_0| + r_0)^4\right] \qquad \text{for } \|\theta-\theta_0\| \le r_0.$$

Therefore,

$$\sup_{\|\theta-\theta_0\|\le r_0} E\left[g_2(X,\theta)^4\right] \le E\left[(|X| + |\xi_0| + r_0)^4\right] < \infty.$$

□

Lemma B.3. *$E\left[g(X,\theta)g^T(X,\theta)\right]$ is finite and its elements are continuous w.r.t. θ. Furthermore, $E\left[g(X,\theta_0)g^T(X,\theta_0)\right]$ is positive definite.*

Proof.

$$E\left[g(X,\theta)g^T(X,\theta)\right] = \begin{bmatrix} (1-2\alpha)E\left[\mathbb{1}_{(X\leq q)}\right]+\alpha^2 & (1-\alpha)E\left[\mathbb{1}_{(X\leq q)}(X-\xi)\right] \\ (1-\alpha)E\left[\mathbb{1}_{(X\leq q)}(X-\xi)\right] & E\left[\mathbb{1}_{(X\leq q)}(X-\xi)^2\right] \end{bmatrix}.$$

Assumption **A(2)** guarantees that the elements are finite and from assumption **A(0)**, it is easy to verify the continuity of $E\left[g(X,\theta)g^T(X,\theta)\right]$ w.r.t. θ.

$$E\left[g(X,\theta_0)g^T(X,\theta_0)\right] = \begin{bmatrix} \alpha(1-\alpha) & 0 \\ 0 & E\left[\mathbb{1}_{(X\leq q_0)}(X-\xi_0)^2\right] \end{bmatrix}.$$

$\alpha(1-\alpha) > 0$ since $\alpha \in (0,1)$. $E\left[\mathbb{1}_{(X\leq q_0)}(X-\xi_0)^2\right] > 0$ since X is absolutely continuous distributed and $f_X(x) > 0$ on its support. Positive definiteness follows. □

Lemma B.4. *For $k = 1, 2$,*

$$\sup_{\|\theta-\theta_0\|\leq r_0} \left\| \frac{\partial E\left[g_k(X,\theta)\right]}{\partial \theta} \right\| < \infty.$$

Proof.

$$\frac{\partial E\left[g_1(X,\theta)\right]}{\partial q} = \frac{\partial}{\partial q}\left(E\left[\mathbb{1}_{(X\leq q)}\right]-\alpha\right) = f_X(q),$$

$$\frac{\partial E\left[g_1(X,\theta)\right]}{\partial \xi} = \frac{\partial}{\partial \xi}\left(E\left[\mathbb{1}_{(X\leq q)}\right]-\alpha\right) = 0.$$

Therefore,

$$\sup_{\|\theta-\theta_0\|\leq r_0} \left\| \frac{\partial E\left[g_1(X,\theta)\right]}{\partial \theta} \right\| = \sup_{\|\theta-\theta_0\|\leq r_0} \left\| \begin{bmatrix} f_X(q) \\ 0 \end{bmatrix} \right\| < \infty,$$

since $f_X(x)$ is bounded.

$$\frac{\partial E\left[g_2(X,\theta)\right]}{\partial q} = \frac{\partial}{\partial q} E\left[\mathbb{1}_{(X\leq q)}(X-\xi)\right] = f_X(q)(q-\xi),$$

$$\frac{\partial E\left[g_2(X,\theta)\right]}{\partial \xi} = \frac{\partial}{\partial \xi} E\left[\mathbb{1}_{(X\leq q)(X-\xi)}\right] = -F_X(q).$$

Therefore,

$$\sup_{\|\theta-\theta_0\|\le r_0} \left\| \frac{\partial E\left[g_2(X,\theta)\right]}{\partial \theta} \right\| = \sup_{\|\theta-\theta_0\|\le r_0} \left\| \begin{bmatrix} f_X(q)(q-\xi) \\ -F_X(q) \end{bmatrix} \right\| < \infty,$$

again because $f_X(x)$ is bounded and $q-\xi$ is bounded when $\|\theta-\theta_0\| \le r_0$. □

The next lemma is on the covering number of the sets of functions

$$\mathscr{F}_{g_k} := \{g_k(\cdot,\theta) : \|\theta-\theta_0\| \le r_0\}, \qquad k = 1, 2.$$

Recall that for $\varepsilon > 0$, the covering number of a set of functions, $\mathscr{F}$, is the smallest number m of functions $g_1, \cdots, g_m$ (not necessarily in $\mathscr{F}$) such that each member of $\mathscr{F}$ is within ε distance from some g_j, $1 \le j \le m$, where the distance is measured by a metric ϱ. Denote this covering number as $N(\varepsilon, \varrho, \mathscr{F})$. For a sample $\{x_1, \cdots, x_n\}$, denote by P_n the empirical distribution that puts mass probability $1/n$ on each x_i.

Lemma B.5. *For $k = 1, 2$, and $\varepsilon > 0$,*

$$N(\varepsilon, \varrho, \mathcal{F}_{g_k}) \le B(P_n)\varepsilon^{-w},$$

where $\limsup_n E[B(P_n)] < \infty$ and w is a positive constant.

Proof. Recall that the envelope of a class of functions $\mathscr{F}$ is a function G such that $|\psi| \le G$ for all $\psi \in \mathscr{F}$. $\mathscr{F}_{g_1}$ has an envelope function $G_1 = 1$. $\mathscr{F}_{g_2}$ has an envelope function $G_2 = |x| + |\xi_0| + r_0$ as

$$|\mathbb{1}_{(x\le x)}(x-\xi)| \le |x-\xi| \le |x| + |\xi| \le |x| + |\xi_0| + r_0,$$

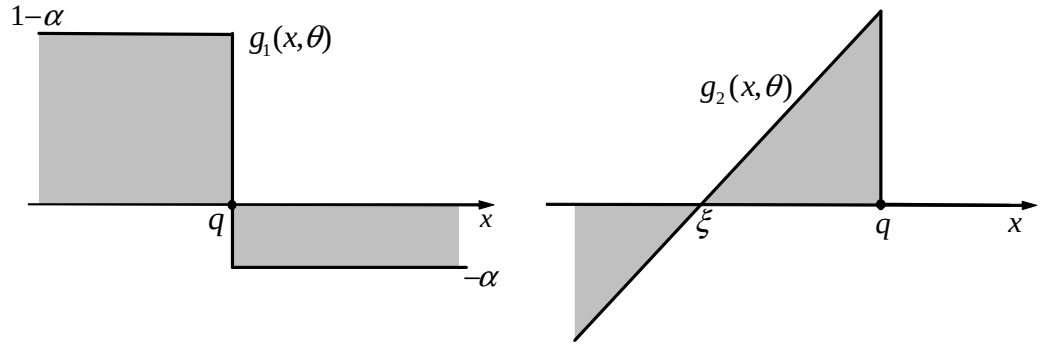
for all $\theta = (q, \xi)^T$, $\|\theta-\theta_0\| \le r_0$.

First we show that the graphs of functions in $\mathscr{F}_{g_1}$ and $\mathscr{F}_{g_2}$ form polynomial classes of sets. Then we use the approximation lemma (P_{27} in [66]) to show the desired result.

Recall that the graph of a real-valued function ψ on a set S is the set

$$G_\psi = \{(s,t) : 0 \le t \le \psi(s) \quad (or) \quad \psi(s) \le t \le 0\}.$$

For a fixed $\theta = (q,\xi)^T$, the graph of $g_1(\cdot,\theta)$ and $g_2(\cdot,\theta)$ are shown as the shaded regions in Figure B.1.



(a) G_{g_1} (b) G_{g_2}

Figure B.1: Graphs of $g_1(\cdot,\theta)$ and $g_2(\cdot,\theta)$ for a fixed $\theta = (q,\xi)^T$

Clearly, graphs of functions in $\mathscr{F}_{g_1}$ form a linear class of sets and graphs of functions in $\mathscr{F}_{g_2}$ form a quadratic class of sets. The approximation lemma (Lemma 25 in [66]) then tells us that for $0 < \varepsilon < 1$,

$$\begin{aligned} N(\varepsilon, L_1(P_n), \mathscr{F}_{g_1}) &\le B_1\varepsilon^{-2} \le B_1\varepsilon^{-4} \\ N(\varepsilon, L_1(P_n), \mathscr{F}_{g_2}) &\le B_2\big(E_{P_n}[G_2]\big)^4\varepsilon^{-4}, \end{aligned}$$

where B_1 and B_2 are positive constants and $E_{P_n}[G_2] = \frac{1}{n}\sum_{i=1}^n G_2(x_i)$.

For a sample $\{x_1, \cdots, x_n\}$, there exist constants c_j, $j = 0, \cdots, 4$ such that

$$\begin{aligned} (E_{P_n}[G_2])^4 &= \left(\frac{1}{n}\sum_{i=1}^n |x_i| + |\xi_0| + r_0\right)^4 \\ &= \sum_{j=0}^4 c_j \left(\frac{1}{n}\sum_{i=1}^n |x_i|\right)^j. \end{aligned}$$

Because for each $j = 0, 1, \cdots, 4$,

$$\limsup_n E\left[\left(\frac{1}{n}\sum_{i=1}^{n}|x_i|\right)^j\right] < \infty,$$

$$\limsup_n E\left[B_2\left(E_{P_n}[G_2]\right)^4\right] < \infty.$$

The desired conclusion follows by letting $B(P_n) = \max\{B_1, B_2 E_{P_n}[G_2]\}$. □

Lemma B.6. *For any sequence of constants $0 < \rho_{n1} < \rho_{n2} \to 0$, there exists a constant c such that*

$$\sup_{\rho_{n1} \le \|\theta - \theta_0\| \le \rho_{n2}} \|E[g(X,\theta)]\| \ge c\rho_{n1}.$$

Proof. For any $0 < \rho_{n1} < \rho_{n2}$, let $\theta' = (q_0, \xi_0 - \rho_{n1})^T$. We have $\|\theta' - \theta_0\| = \rho_{n1} < \rho_{n2}$,

$$E[g_1(X,\theta')] = 0,$$

$$E\big[g_2(X,\theta')\big] = E\big[\mathbb{1}_{(X \le q_0)}(X - \xi_0 + \rho_{n1})\big] = E\big[\mathbb{1}_{(X \le q_0)}\rho_{n1}\big] = \alpha\rho_{n1},$$

and

$$\big\|E\big[g(X,\theta')\big]\big\| = \left\|\begin{bmatrix} 0 \\ -\alpha\rho_{n1} \end{bmatrix}\right\|.$$

Thus,

$$\sup_{\rho_{n1} \le \|\theta - \theta_0\| \le \rho_{n2}} \|E[g(X,\theta)]\| \ge \|E[g(X,\theta')]\| \ge c\rho_{n1}.$$

Let $c = \alpha$ and the conclusion follows. □

Lemma B.7. *For $k = 1, 2$, there exists a constant d such that*

$$E\Big[|g_k(X,\theta) - g_k(X,\theta_0)|^2\Big] \le d\|\theta - \theta_0\|,$$

for all $\|\theta - \theta_0\| \le r_0$.

Proof. First prove for $k = 1$.

$$\begin{aligned} E\Big[|g_1(X,\theta) - g_1(X,\theta_0)|^2\Big] &= E\Big[(\mathbb{1}_{(X\le q)} - \mathbb{1}_{(X\le q_0)})^2\Big] \\ &= F_X(\max\{q,q_0\}) - F_X(\min\{q,q_0\}). \end{aligned}$$

Let $d_1 = \max\{f_X(q) : \theta = (q,\xi)^T, \|\theta - \theta_0\| \le r_0\}$, where f_X is the probability density function of X. The maximum is achieved as the set $\{\theta : \|\theta - \theta_0\| \le r_0\}$ is compact and f_X is bounded. Then,

$$E\Big[|g_1(X,\theta) - g_1(X,\theta_0|^2\Big] \le d_1\|q - q_0\| \le d_1\|\theta - \theta_0\|, \qquad \text{for } \|\theta - \theta_0\| \le r_0.$$

Now look at the case when $k = 2$. Let $\Psi(\theta) = E\Big[|g_2(X,\theta) - g_2(X,\theta_0)|^2\Big]$, then

$$\begin{aligned} \Psi(\theta) &= E\Big[\big(\mathbb{1}_{(X\le q)}(X-\xi) - \mathbb{1}_{(X\le q_0)}(X-\xi_0)\big)^2\Big] \\ &= E\Big[\mathbb{1}_{(X\le q)}(X-\xi)^2 + \mathbb{1}_{(X\le q_0)}(X-\xi_0)^2 - 2\mathbb{1}_{(X\le q)}\mathbb{1}_{(X\le q_0)}(X-\xi)(X-\xi_0)\Big], \end{aligned}$$

with $\Psi(\theta_0) = 0$.

$$\frac{\partial\Psi}{\partial q} = f_X(q)(q-\xi)^2 - 2f_X(q)\mathbb{1}_{(q\le q_0)}(q-\xi_0)(q-\xi),$$

$$\frac{\partial\Psi}{\partial\xi} = \begin{cases} 2E[\mathbb{1}_{(X\le q)}](\xi - \xi_0), & q \le q_0 \\ 2E[\mathbb{1}_{(X\le q)}(\xi - X)], & q > q_0. \end{cases}$$

It is an easy exercise to verify that $\frac{\partial\Psi}{\partial\theta} = \left(\frac{\partial\Psi}{\partial q}, \frac{\partial\Psi}{\partial\xi}\right)^T$ is bounded for all $\|\theta - \theta_0\| \le r_0$. Let

$$d_2 = \sup_{\|\theta-\theta_0\|\le r_0} \left\| \frac{\partial\Psi}{\partial\theta} \right\|,$$

then

$$\|\Psi(\theta) - \Psi(\theta_0)\| \le d_2\|\theta - \theta_0\|, \qquad \text{for } \|\theta - \theta_0\| \le r_0.$$

The conclusion is established by taking $d = \max\{d_1, d_2\}$. □

Proof of Theorem 4.1. We use Theorem 2 and 3 in [95] to show conclusion (i) and (ii), respectively. These two theorems deal with empirical estimator for parameters of conditional distributions. In our case, the marginal distribution is of interest. In the setting of [95], this can be viewed as the case where the explanatory variable is deterministic. Thus the conditional distribution of X with respect to the explanatory variable reduces to the marginal distribution of X. The conditions concerning the smoothness of g when the explanatory variable changes are therefore automatically satisfied.

Our MELE estimator $\hat{\theta}_{EL}$ maximizes $\sum_{i=1}^{n} w_i \ln np_i$ under the moment constraints, where $w_i = 1/n$ for all $i = 1, \cdots, n$. Since the explanatory variable is fixed, this corresponds to the choice of any kernel weights with $K(\|t\|) = 1$ when $\|t\| = 0$. The bandwidth for the kernel weights can also be freely picked to meet the bandwidth condition without altering the weights. Also, for each observation $x \in \mathbb{R}$ of X, g is bounded for θ in the compact set Θ.

Lemma B.1–B.7 verify that the remaining conditions for Theorem 2 and 3 in [95] are satisfied. We thus have that $\hat{\theta}_{EL} \xrightarrow{p} \beta_0$ and $\sqrt{(}\hat{\beta} - \beta_0) \xrightarrow{d} \mathcal{N}(0, V)$ with

$$\begin{aligned} V^{-1} &= \frac{\partial E[g^T(X,\theta_0)]}{\partial \theta}\Big(E[g(X,\theta_0)g^T(X,\theta_0)]\Big)^{-1}\frac{\partial E[g(X,\theta_0)]}{\partial \theta} \\ &= \begin{bmatrix} \dfrac{f_X(q_0)^2}{\alpha(1-\alpha)} + \dfrac{(q_0-\xi_0)^2 f_X(q_0)^2}{E[\mathbb{1}_{(X\le q_0)}(X-\xi_0)^2]} & \dfrac{-\alpha(q_0-\xi_0)f_X(q_0)}{E[\mathbb{1}_{(X\le q_0)}(X-\xi_0)^2]} \\ \dfrac{-\alpha(q_0-\xi_0)f_X(q_0)}{E[\mathbb{1}_{(X\le q_0)}(X-\xi_0)^2]} & \dfrac{\alpha^2}{E[\mathbb{1}_{(X\le q_0)}(X-\xi_0)^2]} \end{bmatrix}. \end{aligned}$$

It is straightforward to check that

$$V = \begin{bmatrix} \dfrac{\alpha(1-\alpha)}{f_X(q_0)^2} & \dfrac{(1-\alpha)(q_0-\xi_0)}{f_X(q_0)} \\ \dfrac{(1-\alpha)(q_0-\xi_0)}{f_X(q_0)} & \dfrac{E[\mathbb{1}_{(X\le q_0)}(X-\xi_0)^2]}{\alpha^2} + \dfrac{1-\alpha}{\alpha}(q_0-\xi_0)^2 \end{bmatrix}.$$

□

Bibliography

[1] K. Aas, I. Haff, and X. Dimakos. Risk estimation using the multivariate normal inverse gaussian distribution. *Journal of Risk*, 8:39–60, 2006.

[2] Carlo Acerbi and Dirk Tasche. On the coherence of expected shortfall. *Journal of Banking and Finance*, 26(7):1487–1503, 2002.

[3] Philippe Artzner, Freddy Delbaen, Jean-Marc Eber, and David Heath. Coherent measure of risk. *Mathematical Finance*, 9(3):203–228, July 1999.

[4] Basel Committee on Banking Supervision. Amendment to the capital accord to incorporate market risks, 1996.

[5] Gilbert W. Bassett, Roger Koenker, and Gregory Kordas. Pessimistic portfolio allocation and choquet expected utility. *Journal of Financial Econometrics*, 2:477–492, 2004.

[6] Daniel Bernoulli. Exposition of a new theory on the measurement of risk. *Econometrica*, 22:23–36, 1954. Translation of the original 1738 paper.

[7] Dimitris Bertsimas, Geoffrey J. Lauprete, and Alexander Samarov. Shortfall as a risk measure: Properties, optimization and applications. *Journal of Economic Dynamics and Control*, 28:1353–1381, 2004.

[8] Herman J. Bierens and Donna K. Ginther. Integrated conditional moment testing of quantile regression models. *Empirical Economics*, 26:307–324, 2001.

[9] George E. P. Box and Norman R. Draper. *Empirical Model-building and Response Surfaces*. Wiley, 1987.

[10] Stephen Boyd and Lieven Vandeberghe. *Convex Optimization*. Cambridge University Press, 2004.

[11] S. R. Broadbent and J. M. Hammersley. Percolation processes. In *Proceedings of the Cambridge Philosophical Society*, volume 53, pages 629–645, 1957.

[12] Jean M. Carlson and John C. Doyle. Highly optimized tolerance: Robustness and design in complex systems. *Physical Review Letters*, 84(11):2529–2532, 2000.

[13] Derek T. Chalmers and Dominic P. Behan. The use of constitutively active GPCRs in drug discovery and functional genomics. *Nat Rev Drug Discov*, 1(8):599–608, August 2002.

[14] Gary Chamberlain. Asymptotic efficiency in estimation with conditional moment restrictions. *Journal of Econometrics*, 34:305–334, 1987.

[15] Peter Cheeseman, Bob Kanefsky, and William M. Taylor. Where the really hard problems are. In *Proceedings of the 11th IJCAI*, pages 331–337, 1991.

[16] Song Xi Chen and Petr Hall. Smoothed empirical likelihood confidence intervals for quantiles.

[17] Peter F. Christoffersen, Jinyong Hahn, and Atsushi Inoue. Testing and comparing value-at-risk measures. *Journal of Empirical Finance*, 8:325–381, 2001.

[18] Jon Danielsson. The emperor has no clothes: Limits to risk modelling. *Journal of Banking and Finance*, 16(7):1273–1296, July 2002.

[19] Dieter Denneberg. *Non-additive Measure and Integral*. Kluwer Academic Publishers, 1994.

[20] Francis X. Diebold and Roberto S. Mariano. Comparing predictive accuracy. *Journal of Business and Ecomonic Statistics*, 13(3):253–263, July 1995.

[21] Jürgen Drews. Drug discovery: a historical perspective. *Science*, 287, March 2000.

[22] John Dunagan, Daniel Spielman, and Shang-Hua Teng. Smoothed analysis of the renegar's condition number for linear programming, 2002.

[23] Marina Epelman and Robert M. Freund. Condition number complexity of an elementary algorithm for resolving a conic linear system. Technical report, Massachusetts Institute of Technology, Operations Research Center, Cambridge, Massachusetts, USA, 1997.

[24] Jianqing Fan and Jian Zhang. Sieve empirical likelihood ratio tests for nonparametric functions. *The Annals of Statistics*, 32(5):1858–1907, 2004.

[25] Ryan Feeley, Pete Seiler, Andrew Packard, and Michael Frenklach. Consistency of a reaction dataset. *Journal of Physical chemistry*, 108:9573–9583, 2004.

[26] Jean-David Fermanian and Olivier Scaillet. Sensitivity analysis of var and expected shortfall for portfolios under netting agreements. *Journal of Banking and Finance*, 29(4):927–958, April 2005.

[27] Hans Föllmer and Alexander Schied. Convex measures of risk and trading constraints. *Finance and Stochastics*, 6:429–447, 2002.

[28] Michael Frenklach, Andrew Packard, Pete Seiler, and Ryan Feeley. Collaborative data processing in developing predictive models of complex reaction systems. *International Journal of Chemical Kinetics*, (1):57–66, 2004.

[29] Michael Frenklach, Andrew Packard, and Pter Seiler. Prediction uncertainty from models and data. In *Proceedings of the American Control Conference*, volume 5, pages 4135–4140, Anchorage, AK, 8-10 May 2002.

[30] Robert Freund and Jorge Vera. Some characterizations and properties of the "distance to ill-posedness" and the condition measure of a conic linear system. *Mathematical Programming*, 86:225–260, 1999.

[31] Rüdiger Frey and Alexander J. McNeil. VaR and expected shortfall in portfolio of dependent credit risks: Conceptual and practical insights. *Journal of Banking and Finance*, 26:1317–1334, 2002.

[32] Milton Friedman and L. J. Savage. The utility analysis of choices involving risk. *The Journal of Political Economy*, 56:279–304, 1948.

[33] Dennice Gayme, John Doyle, Maryam Fazel, Antonis Papachristodoulou, and Stephen Prajna. Optimization based methods for determining basins of attraction in the logistic map and set membership in the mandelbrot set. Under preparation for journal submission.

[34] Raffaella Giacomini and Ivana Komunjer. Evaluation and combination of conditional quantile forcasts. *Journal of Business and Economics Statistics*, 23:416–431, 2005.

[35] Geoffrey Grimmett. *Percolation*. Springer-Verlag, New York, 1989.

[36] G. Hanoch and H. Levy. The efficiecy analysis of choices involving risk. *The Review of Economic Studies*, 36:335–346, 1969.

[37] Lars Peter Hansen. Large sample properties of generalized method of moments estimators. *Econometrica*, 50(4):1029–1054, July 1982.

[38] Andrew L. Hopkins and Colin R. Groom. The druggable genome. *Nat Rev Drug Discov*, 1(9):727–730, September 2002.

[39] Khuloud Jaqaman and Gaudenz Danuser. Linking data to models: Data regression. *Nature Reviews Molecular Cell Biology*, 7:813–819, November 2006.

[40] Harry Kesten. *Percolation Theory for Mathematicians*. Birkhäuser, 1982.

[41] Yuichi Kitamura. Empirical likelihood methods in econometrics: Theory and practice. Cowles Foundation Discussion Paper No. 1569, http://cowles.econ.yale.edu.

[42] Yuichi Kitamura. Empirical likelihood methods with weakly dependent processes. *The Annals of Statistics*, 25(5):2084–2102, 1997.

[43] Yuichi Kitamura, Gautam Tripathi, and Hyungtaik Ahn. Empirical likelihood-based inference in conditional moment restricted models. *Econometrica*, 72(6):1667–1714, 2004.

[44] Roger Koenker and Gilbert Bassett, Jr. Regression quantiles. *Econometrica*, 46(1):33–50, 1978.

[45] Roger Koenker and Quanshui Zhao. Conditional quantile estimation and inference for ARCH models. *Econometric Theory*, 12:793–813, 1996.

[46] Ivana Komunjer. Asymmetric power distribution: Theory and applications to risk measurement. *Journal of Applied Econometrics*, 2006. In press.

[47] Shigeo Kusuoka. On law invariant coherent risk measures. *Advances in Mathematical Economics*, 3:83–95, 2001.

[48] Michael LeBlanc and John Crowley. Semiparametric regression functionals. *Journal of the American Statistical Association*, 90(429):95–105, March 1995.

[49] Harvey Lodish, Arnold Berk, S. Lawrence Zipursky, Paul Matsudaira, David Baltimore, and James Darnell. *Molecular cell biology*. W.H. Freeman & Company, forth edition, 1999.

[50] Philip Ma and Rodney Zemmel. Value of novelty? *Nat Rev Drug Discov*, 1(8):571–572, August 2002.

[51] Harry Markowitz. Portfolio selection. *The Journal of Finance*, 1:77–91, 1952.

[52] Carlos Martins-Filho and Feng Yao. Estimation of value-at-risk and expected shortfall based on nonlinear models of return dynamics and extreme value theory. *Studies in Nonlinear Dynamics and Econometrics*, 10(2), 2006.

[53] Alexander J. McNeil and Rüdiger Frey. Estimation of tail-related risk measures for heteroskedastic financial time series: an extreme value approach. *Journal of Empirical Finance*, 7:271–300, 2000.

[54] David Miller and Michael Frenklach. Sensitivity analysis and parameter estimation in dynamic modeling of chemical kinetics. *International Journal of Chemical Kinetics*, 15:677–696, 1983.

[55] David Mitchell, Bart Selman, and Hector Levesque. Hard and easy distributions of sat problems. In *Proceedings of the Tenth National Conference on Artificial Intelligence (AAAI-92)*, pages 459–465, San Jose, CA, July 1992.

[56] Rémi Monasson, Riccardo Zecchina, Scott Kiekpatrick, Bart Selman, and Lidror Troyansky. Determining computational complexity from characteristic 'phase transition'. *Nature*, 400:133–137, July 1999.

[57] Raymond H. Myers and Douglas C. Montgomery. *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*. Wiley, 2002.

[58] Taisuke Otsu. Conditional empirical likelihood estimation and inference for quantile regression models. University of Wisconsin Madison.

[59] Art B. Owen. Empirical likelihood ratio confidence interval for a single functional. *Biometrika*, 75(2):237–249, 1988.

[60] Art B. Owen. Empirical likelihood ratio confidence regions. *The Annals of Statistics*, 18(1):90–120, 1990.

[61] Art B. Owen. Empirical likelihood for linear models. *The Annals of Statistics*, 19(4):1725–1747, 1991.

[62] Art B. Owen. *Empirical Likelihood*. Chapman and Hall/CRC, 2001.

[63] Antonis Papachristodoulou and Stephen Prajna. Semidefinite relaxations, fragility, and complexity of the number partitioning problem, 2003. Working notes. Control and Dynamical Systems, Caltech, Pasadena, USA.

[64] Christos H. Papadimitriou and Kenneth Steiglitz. *Combinatorial Optimization: Algorithms and Complexity*. Dover, New York, 1998.

[65] Pablo A. Parrilo. *Structured Semidefinite Programs and Semialgebraic Geometry Methods in Robustness and Optimization*. PhD thesis, California Institute of Technology, 2000.

[66] David Pollard. *Convergence of Stochastic Processes*. Springer-Verlag, New York, 1984.

[67] S. Prajna, A. Papachristodoulou, P. Seiler, and P. A. Parrilo. *SOSTOOLS: Sum of squares optimization toolbox for MATLAB*, 2004.

[68] Stephen Prajna. Barrier certificates for nonlinear model validation. In *Proceedings of the IEEE Conference on Decision and Control (CDC)*, Maui, HI, 2003.

[69] Stephen Prajna. *Optimization-Based Methods for Nonlinear and Hybrid Systems Verification*. PhD thesis, California Institute of Technology, 2005.

[70] Stephen Prajna, Antonis Papachristodoulou, and Pablo A. Parrilo. Introducing sostools: A general purpose sum of squares programming solver. In *Proceedings of the IEEE Conference on Decision and Control (CDC)*, Las Vegas, NV, 2002.

[71] Stephen Prajna, Antonis Papachristodoulou, Peter Seiler, and Pablo A. Parrilo. New developments in sum of squares optimization and sostools. In *Proceedings of the American Control Conference (ACC)*, Boston, MA, 2004.

[72] Jing Qin and Jerry Lawless. Empirical likelihood and general estimating equations. *The Annals of Statistics*, 22(1):300–325, 1994.

[73] John Quiggin. A theory of anticipated utility. *Journal of Economic Behavior and Organization*, 3:323–343, 1982.

[74] James Renegar. Some perturbation theory for linear programming. *Mathematical Programming*, 65(1):73–91, 1994.

[75] James Renegar. Incorporating condition measures into the complexity theory of linear programming. *SIAM Journal on Optimization*, 5(3):506–524, 1995.

[76] James Renegar. Linear programming, complexity theory and elementary functional analysis. *Mathematical Programming*, 70(3):279–351, 1995.

[77] R. Tyrrell Rockafellar and Stanislav Uryasev. Optimization of conditional Value-at-Risk. *The Journal of Risk*, 2(3):21–41, 2000.

[78] R. Tyrrell Rockafellar and Stanislav Uryasev. Conditional value-at-risk for general loss distributions. *Journal of Banking and Finance*, 26:1443–1471, 2002.

[79] Olivier Scaillet. Nonparametric estimation and sensitivity analysis of expected shortfall. *Mathematical Finance*, 14(1):115–129, January 2004.

[80] Olivier Scaillet. Nonparametric estimation of conditional expected shortfall. *Revue Assurances et Gestion des Risques/Insurance and Risk Management Journal*, 74:639–660, 2005.

[81] Franz Schwabl. *Statistical Mechanics*. Springer, second edition, 2006.

[82] Pete Seiler, Michael Frenklach, Andrew Packard, and Ryan Feeley. Numerical approaches for collaborative data processing. *Opt. Eng*, 2005.

[83] William F. Sharpe. Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance*, 19(3):425–442, 1964.

[84] Dirk Tasche. Expected shortfall and beyond. *Journal of Banking and Finance*, 27(7):1519–1533, 2002.

[85] Aautam Tripathi and Yuichi Kitamura. Testing conditional moment restrictions. *The Annals of Statistics*, 31(6):2059–2095, 2003.

[86] J. van den Berg and A. Ermakov. A new lower bound for the critical probability of site percolation on the square lattice. *Random Structure and Algorithms*, 8(3):199–212, May 1996.

[87] John von Neumann and Oskar Morgenstern. *Theory of Games and Economic Behavior*. Princeton Univ. Press, Princeton NJ, 1944.

[88] Malcolm S. Whiteway, Cunle Wu, Thomas Leeuw, Karen Clark, Anne Fourest-Lieuvin, David Y. Thomas, and Ekkehard Leberer. Association of the yeast pheromone response G protein $\beta\gamma$ subunits with the MAP kinase scaffold Ste5p. *Science*, 269(5230):1572–1575, 15 September 1995.

[89] John C. Wierman. Duality for k-degree percolation on the square lattice. In Harry Kesten, editor, *Percolation Theory and Ergodic Theory of Infinite Particle Systems*. Springer-Verlag, 1987.

[90] John Burr Williams. *Theory of Investment Value*. Harvard University Press, 1938.

[91] Menahem E. Yaari. The dual theory of choice under risk. *Econometrica*, 55(1):95–115, January 1987.

[92] Yasuhiro Yamai and Toshinao Yoshiba. Comparative analysis of expected shortfall and Value-at-Risk (3): Their validity under market stress. *Monetary And Economic Studies*, pages 181–238, October 2002.

[93] Tau-Mu Yi, Maryam Fazel, Xin Liu, Tosin Otitoju, Jorge Goncalves, Antonis Papachristodoulou, Stephen Prajna, and John Doyle. Application of robust model validation using sostools to the study of g-protein signaling in yeast. In *Proceedings of Foundation of System Biology in Engineering*, 2005.

[94] Tau-Mu Yi, Hiroaki Kitano, and Melvin I. Simon. A quantitative characterization of the yeast heterotrimeric g protein cycle. *Proc. Natl Acad. Sci.*, 16 September 2003.

[95] Jian Zhang and Irène Gijbels. Sieve empirical likelihood and extensions of the generalized least squares. *Board of the Foundation of the Scandinavian Journal of Statistics*, 30:1–24, 2003.

[96] John Xu Zheng. A consistent nonparametric test of parametric regression models under conditional quantile regressions. *Econometric Theory*, 14:123–138, 1998.