Essays in Financial Economics

by

Hyungjune Kang

B.A., Princeton University (2012)

Submitted to the Alfred P. Sloan School of Management in partial fulfillment of the requirements for the degree of Doctor of Philosophy at the MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2018

© Massachusetts Institute of Technology 2018. All rights reserved.

Signature redacted

Author .........................

Alfred P. Sloan School of Management May 1, 2018

Signature redacted

Certified by ..............

Andrey Malenko
Jon D. Gruber Career Development Professor of Finance Thesis Supervisor

Signature redacted

Accepted by ..............

Catherine Tucker
Director, Sloan School of Management PhD Program
Essays in Financial Economics

by

Hyungjune Kang

Submitted to the Alfred P. Sloan School of Management
on May 1, 2018, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

This thesis consists of three chapters.

In the first chapter, I analyze a dynamic game in which a sender of unknown quality persuade a receiver by designing an experiment (model) that transforms signals into recommendations (messages). When the receiver learns the sender's quality by observing the sender's past recommendations and realized events, I show that, due to reputational concern, the sender chooses an experiment that limits the amount of information. A more patient sender is less likely to devise an informative experiment. I also demonstrate that the quality of the sender-optimal experiment—measured by the amount of information—is not monotonic in the sender's quality. The framework can be applied to various settings including financial regulation and analyst forecasting.

In the second chapter, based on joint work with Tetsuya Kaji, we study how Value-at-Risk (VaR) constraint affects the amount of information that price conveys in an economy with asymmetric information. We first show that VaR constraint is different from others (e.g. borrowing and short-sale constraints) in that VaR constraint is relevant only when price is moderate. We find that when some investors follow a VaR rule, realistically high or low prices reveal more information than intermediate prices. We illustrate how the presence of VaR investors affects other investors' incentive to acquire information.

In the third chapter, based on joint work with Tetsuya Kaji, we propose a class of risk measures called the tail risk measures that establish the upper bounds below which the quantities of interest fall with probability at least as much as a pre-specified confidence level. We show that a simple rule based on the Bonferroni inequality can control a tail risk measure at a desired level even when the true risk is unknown and needs to be estimated. Most popular risk measures such as Value-at-Risk and expected shortfall are interpreted as tail risk measures. Empirical applications illustrate how the proposed concept can be applied to practical risk control problems.

Thesis Supervisor: Andrey Malenko
Title: Jon D. Gruber Career Development Professor of Finance
Acknowledgments

I am extremely grateful to my committee chair Andrey Malenko. Since my first year in the Ph.D. program, Andrey has taught me how to develop research ideas and translate them into formal models. I am equally grateful to other committee members Debbie Lucas and Haoxiang Zhu. Their research and career advice has led me to successfully complete the Ph.D. program.

I would also like to thank my friends who stood by me during hard times. I was able to overcome many obstacles with the help from Victor Duarte, Tetsuya Kaji, Joon Ho Kang, So Yoon Kim, Seung Ki Kwak, Hoseok Lee, Louis Lee, Byung Min Nahm, and Jack Beumseok Shim. Special thanks go to Ki Won Ahn who supported me and made me laugh all the time. Any of my accomplishments would not have been possible without her.

Finally, I thank my parents and my brother for their unconditional love and encouragement. I will never forget their unwavering support.
# Contents

1 A Dynamic Model of Reputation in Bayesian Persuasion Games 15

1.1 Introduction ........................................ 15

  1.1.1 Related Literature .............................. 17

  1.1.2 Outline ........................................ 20

1.2 Model .............................................. 20

  1.2.1 Model Setup .................................... 21

  1.2.2 Model Discussion .............................. 23

  1.2.3 Assumptions .................................... 25

  1.2.4 Relationship to Standard Bayesian Persuasion 28

1.3 Benchmark Equilibrium ............................ 29

  1.3.1 Static model ................................... 30

  1.3.2 Dynamic model ................................. 34

1.4 Extensions ......................................... 42

  1.4.1 Perfect ex-post learning ....................... 42

  1.4.2 Full commitment ................................ 43

1.5 Conclusion .......................................... 45

1.6 Appendix ........................................... 49

2 Value-at-Risk Constraint and Price Informativeness 59

2.1 Introduction ...................................... 59

2.2 Model ............................................. 63

  2.2.1 Asset payoff, information structure, and preference 64

  2.2.2 Optimal demand of the informed ............... 67
2.2.3 Equilibrium Concept ........................................... 71
2.3 Equilibrium Price Function ...................................... 71
  2.3.1 Benchmark: GS economy without VaR investors .......... 71
  2.3.2 VaR economy .................................................. 72
2.4 Price Informativeness ............................................. 74
  2.4.1 Overview of VaR equilibrium ................................ 74
  2.4.2 Uninformed Demand ......................................... 74
  2.4.3 Price informativeness in the VaR economy ............... 77
  2.4.4 Price informativeness in other constrained economies ... 83
2.5 Price and Return Distribution .................................. 84
  2.5.1 Price volatility .............................................. 84
  2.5.2 Price and return distribution ................................ 86
2.6 Endogenous Information Acquisition ............................ 88
2.7 Conclusion .......................................................... 91
2.8 Appendix ............................................................ 95

3 Controlling Tail Risk Measures with Estimation Error ............ 101
  3.1 Introduction ..................................................... 101
  3.2 Tail Risk Measures ............................................. 105
  3.3 Controlling Tail Risk Measures ................................. 110
  3.4 Empirical Application to Expected Shortfall ................. 111
  3.5 Conclusion ....................................................... 115
  3.6 Appendix ........................................................ 121
List of Figures

1-1 Example of ex-post updating rule ........................................ 23

1-2 Timeline of the benchmark model ........................................ 23

1-3 The value function and optimal policy in the single period. Left: Gaussian distribution with $\rho = 0.25$, and $\omega_c = 0.62$, $y = 0.47$. Middle: “Trigger” model with $\omega_c = 0.68$, $y = 0.75$. Right: Clayton copula $C = (u^{-\eta} + v^{-\eta} - 1)^{-1/\eta}$ with $\eta = 2$, and $\omega_c = 0.23$, $y = 0.9$. ........................ 32

1-4 Continuation value for various distributional assumptions and $\mu = 0.2, 0.5, 0.8$ (from bottom to top). The distributions and parameters are the same as the ones in Figure 1-3. ................................................ 37

1-5 $(\theta, \theta')$ is normal with $\rho = 0.25$, $\omega_c = 0.62$, $y = 0.47$ in the two-period model. Left: Plot of $\hat{b}(\mu)$. Middle: Value function with various $b_1$ values ($b_2$ is kept fixed at 1). Right: Expected payoffs under the optimal and full disclosure scheme for $b_1 = 0.03$. ....................... 41

1-6 Continuation values in the two-period model under the two formulations for $\mu = 0.2, 0.5, 0.8$. .................................................. 43

1-7 The Sender’s value function and optimal policy under the benchmark and full commitment case. $C$ is Gaussian with correlation $\rho = 0.25$, $\omega_c = 0.62$, $y = 0.47$, $b_1 = 0.03$, $b_2 = 1$. ......................... 44
2-1 VaR investor optimal demand. The unconstrained demand is the optimal demand of the informed traders without the VaR rule. ‘VaR when go long’ (short) illustrates the upper (lower) bound of the trading position. The thick line green line is the VaR investor’s optimal demand when VaR constraint is approximated by linear lines. 68

2-2 Equilibrium GS and VaR Prices. The left (right) panel corresponds to GS (VaR) economy. The top figures plot the uninformed demand against price. Varying m, the relationship between the signal and equilibrium price is illustrated in the middle. The bottom figures plot the equilibrium price against m, by varying the signal. The parameter values used here are \( \rho_I = \rho_U = 1, \alpha = 0.3, q = 0.01, \tau_d = 5, \tau_c = 10, \tau_m = 0.5, \bar{m} = 0, w_{IF} = 5\%, w_{IV} = 25\%, w_U = 70\% \). 75

2-3 Uninformed investor optimal demand. The dashed line and solid curve correspond to the uninformed optimal demands in the GS economy and VaR economy, respectively. The parameter values used here are \( \rho_I = \rho_U = 1, \alpha = 0.3, q = 0.01, \tau_d = 5, \tau_c = 10, \tau_m = 0.5, \bar{m} = 0, w_{IF} = 5\%, w_{IV} = 25\%, w_U = 70\% \). 76

2-4 Price informativeness. The dashed and solid lines represent price informativeness (measured by \( Var[d|s]/Var[d|P] \)) of GS and VaR economy, respectively. The parameter values used here are the same as in Figure 2-3. 77

2-5 Iso-price curves and density contour map. The solid lines represent iso-price curves in the \((m, s)\) plane with corresponding price levels \( P = 6, 4, 2, 0, -2, -4, -6 \), respectively. The starred and circled dot points are the states of nature in which the informed demands 0 in the VaR and GS economies, respectively. The x-points plot the conditional mean of \((m, s)\) given each price level of the GS economy. The parameters used here are the same as in Figure 3. 79

2-6 Constrained and unconstrained regions along iso-price curves. 80
Price informativeness in leverage (borrowing) and short-sale constrained economies. In (a), the constraint is modeled by \( x \leq a\tilde{P} + b \) and in (b), \( x \geq -\xi \), where \( x \) is the number of units that constrained agents can buy. The parameters values are \( \rho_I = \rho_U = 1, \alpha = 0.3, q = 0.01, \tau_d = 5, \tau_e = 25, \tau_m = 0.5, \bar{m} = 0, w_{IF} = 5\%, \ w_{bc} = w_s = 25\%, \ w_U = 70\%, \ a = 2, b = -1, \xi = 0.1 \), where \( w_{bc} \) and \( w_s \) denote the population of borrowing and short-sale constrained but informed investors.

Equilibrium price and noisy supply for various signals. This figure plots the equilibrium price against noisy supply \( m \) under various realizations of the signal. The dashed lines represent the GS equilibrium prices for \( s = 4.7, 2.35, 0, -2.35, -4.7 \) (top to bottom). The solid lines are the equilibrium prices in the VaR economy with the same signals as in the GS economy. The \( s \) values correspond to \( 0, \pm 5 \text{std}(s), \pm 10 \text{std}(s) \), respectively. The underlying parameters are \( \rho_I = \rho_U = 1, \alpha = 0.3, q = 0.01, \tau_d = 5, \tau_e = 50, \tau_m = 0.5, \bar{m} = 0, w_{IF} = 5\%, \ w_{IV} = 25\%, \ w_U = 70\% \).

Population of informed and uninformed agents. Solid: with VaR investors. Dashed: without VaR investors. The cost of acquiring information is \( c = 2 \).

Simulations of \( n = 100 \) normal returns with \( S = 10,000 \) iterations illustrate distortion of risk probability, \( \Pr(X < -\text{VaR}_\alpha) \neq \alpha \). Draws outside the range are shown as clusters at the boundaries.
List of Tables

2.1 Parameter Choice. ................................................. 87
2.2 Price and Return Moments and Tail Loss Probabilities when $\tau_e = 50$. 87
2.3 Price and Return Moments and Tail Loss Probabilities when $\tau_e = 10$. 88

3.1 To control ES at 90%, we estimate ES of the daily stock returns at 95% from Jan 4, 2016 to Dec 31, 2017, and the upper bounds of the one-sided confidence bounds (CB) of level 95%. The CIs are based on bootstrap with 10,000 simulations. ................................. 114
Chapter 1

A Dynamic Model of Reputation in Bayesian Persuasion Games

1.1 Introduction

When can an interested party (sender) of uncertain ability persuade a decision maker (receiver) towards some action the sender prefers? How much information is revealed in a persuasion? How does the content of information change over time if the sender were to persuade the receiver repeatedly? These questions are worthwhile investigating because repeated persuasion is common in the real world. For instance, a bank or company may want to convince a regulator that risk is being controlled, or a pharmaceutical company may seek an FDA approval, or a prosecutor would like to elicit conviction from a judge by presenting evidence. In such repeated games, the receiver observes whether the sender’s recommendation turns out to be correct, and therefore infers the sender’s quality over time. Consequently, the sender’s strategy can deviate much from that of a simple one-shot game.

For example, suppose that the sender is the Federal Reserve (Fed) who develops a model to project the next year’s inflation rate. If the model’s outcomes are given only in wide intervals (e.g. between 1% and 2%) the model is not very informative. It is unlikely that the realized rate will lie outside the projected interval, so the rational public does not learn much about the Fed’s ability in making correct predictions.
If the Fed instead makes narrow projections (e.g. between 1% and 1.1%), there will be both hits and misses, and consequently the public can make a meaningful inference about the Fed's ability. If the ex-post public observation conforms to the Fed’s previous projection, the Fed will be perceived as competent and can therefore persuade the public more easily in the future. Otherwise, the Fed loses its reputation and will find it harder to elicit its preferred action from the public.

The Fed’s reputation management can be particularly important since the value of persuasion may not be constant across time and the states of the world. For instance, the Fed may put more value weights in being able to elicit certain actions from the public in bad times. Or, there may be periods when the Fed makes important decisions (e.g. interest rate hike) for which competence is more valuable. How would the Fed “substitute” its reputation over time? How does the Fed optimally choose the amount of information disclosed to the public? How can the Fed be incentivized to transmit more information? These questions can be of importance when social welfare depends positively on the amount of information, because agents can make better and more efficient decisions.

The interest in the role of reputation is not confined to academic research. Blinder (2000) summarizes a survey result in which central bankers and academics were asked to answer a series of questions including the importance of credibility to a central bank and ways to “establish or maintain credibility.” In the former question, the central bankers’ average response recorded 4.83 on a scale of 1 to 5 with 5 being the most important. In the latter question, transparency and “history of honesty” were among the top choices. The survey result strongly suggests that central bankers do care much about their public perception and that their policy choices are affected by such concern.

In this paper, I first propose a tractable framework of dynamic Bayesian persuasion game in which the sender’s ability is unknown. Second, under fairly general conditions, I prove that the sender’s static value function is concave in her reputation parameter (perceived ability). Third, I demonstrate that in a dynamic model the sender dislikes to gamble for her reputation, i.e., discloses as little information as
possible in order to minimize the variation of the receiver’s posterior belief about her competence. Finally, I show how the sender’s optimal choice of experiment hinges on the value of persuasion of the current period relative to that of future periods. Interestingly, extreme types (either competent or incompetent) are more likely to carry out informative experiments than intermediate types. The main contribution of my paper is to 1) introduce a new reputation-based channel that explains limited disclosure of information, 2) build a framework that is able to uniquely quantify the amount of information revealed in equilibrium, and 3) analyze how the sender manages her reputation over time. The required assumptions on the state-signal joint distribution are strikingly simple and can be used in many applications.

1.1.1 Related Literature

Early works have examined various forms of strategic communication. Akerlof (1970) and Spence (1973) consider models of costly signaling, and Crawford and Sobel (1982) explore a cheap talk model in which information revelation is costless. The more recent strand of literature called Bayesian persuasion studies design of information structure when the designer has commitment power and information is verifiable. This paper builds on this framework.

In Bayesian persuasion — introduced by Kamenica and Gentzkow (2011) — the sender designs an experiment, or information structure, that maps each payoff-relevant state to a message, or recommendation. The sender’s incentive is not perfectly aligned with the receiver’s, so the sender’s objective is to construct an experiment so as to maximize the probability of persuading the receiver to take an action the sender prefers. The crucial assumptions of the framework are that the sender can commit to an experiment ex-ante and cannot conceal or falsify any experiment outcomes. By characterizing the space of posteriors induced by some experiment, Kamenica and Gentzkow (2011) give necessary and sufficient conditions under which the sender can persuade the receiver.

Since Kamenica and Gentzkow (2011), there has been a growing literature on Bayesian persuasion. Goldstein and Leitner (2017) apply the framework in a bank
stress test setting, in which a regulator communicates with depositors through test results. They argue that disclosing too much information can be suboptimal for the regulator due to high probability of bank runs. In a similar vein, Faria-e-Castro et al. (2017) study the optimal decision of a regulator, noting that high level of transparency can alleviate adverse selection problem on banks but at the same time increase the likelihood of runs. In Leitner and Yilmaz (2017), a regulator is the receiver, and the sender is a corporation constructing a risk model for both its internal risk control and proposal submitted to the regulator. Full transparency might not be optimal for the firm protected by limited liability, because it allows the regulator to interfere more often.

Most existing works in Bayesian persuasion posit that an experiment simply transforms a true state into a message. Their premise is that the sender is able to carry out a perfectly revealing experiment. In this sense, the sender is assumed to be omnipotent in her ability to figure out the true state exactly. As a result, in repeated games the sender does not incur any cost even if her recommendation is not in line with ex-post realized events. However, as Blinder (2000) suggests, in many applications the sender’s capacity is uncertain and the receiver constantly evaluates the sender’s ability based on realized outcomes.

The main departure of my model from the literature is twofold. First, the sender has limited ability in identifying the true state. In my model, an experiment, or an information transmission mechanism, maps signals to messages. A signal contains information about the true state but can be imperfect. The precision of the signal represents the sender’s type (competence), and the sender’s perceived competence level is her reputation. Second, the model is dynamic in the sense that the sender persuades the receiver in a repeated game. In each period, an outcome of an experiment is revealed to the receiver who takes an action based on the result. Afterwards, the receiver learns (perfectly or imperfectly) the true state. By comparing the sender’s previous projection and the ex-post observation, the receiver reassesses the sender’s ability.

This paper is not the first one to study the interaction between reputation and
information disclosure. Gentzkow and Shapiro (2006) analyze a problem in which a sender (media firm) of unknown type decides whether to reveal information truthfully to the public. They demonstrate that generally the firm strategically distorts its signal and reports the one that is most consistent with the receiver's prior, but that if the public is able to observe the truth ex-post, the bias goes away due to the firm's reputational concern. There are a few key differences between their model and the model to be introduced in this paper. First, under their framework the sender's choice is whether to tell the truth or to lie. Absent concern on reputation, the low quality type finds it optimal to mimic the high quality type and therefore misreports in equilibrium. However, when reputation is at stake — the receiver learns the true state ex-post and updates its belief about the sender — the low quality type no longer does so. In my model, the sender cannot falsify its information and rather chooses the information accuracy that a message contains. It is not immediately clear how much information the sender would decide to reveal. Second, in Gentzkow and Shapiro (2006), the firm chooses its strategy after learning its type, so it does not internalize the impact of its strategy on other types due to lack of commitment. In my model, information is symmetric and the sender commits to a mechanism ex-ante. Third, my framework allows flexibility in the receiver's updating rule and in the joint distribution of the state and signal. For instance, the receiver's observation can be either exogenous or endogenous to the sender's policy choice depending on the formulation (only exogenous in Gentzkow and Shapiro (2006)). In addition, my results are derived under any joint distribution that satisfies a notion of positive dependence.

Few recent papers are also worth mentioning. The first set of papers explores Bayesian persuasion in a dynamic setting. In a model of real options, Orlov et al. (2017) investigate how the timing of information revelation depends on the level of disagreement of two parties. In Ely et al. (2015), the sender (mystery novel writer) persuades the receiver (reader) who derives utility from suspense and surprise. It is shown that at the optimum the sender allocates “budget of variance” evenly over time. The second set of papers considers a static persuasion game when the sender is
privately informed. Perez-Richet (2014), Hedlund (2017), and Kosenko (2017) assume that the sender observes private signal before designing an experiment. The signal can be the sender's type (Perez-Richet (2014)) which directly affects the sender's payoff, or some information about the state of the world (Hedlund (2017) and Kosenko (2017)). Despite some similarities between their models and mine, dynamic implications are very different. In my model, the sender’s type is the precision of the signal she is able to utilize, not the content of the signal itself (e.g. positive/negative). The informativeness of today’s experiment has a one-to-one relation with the variation of the posterior distribution of the sender’s ability. As a result, the sender’s choice can be linked to her preference over the dispersion of her future reputation.

1.1.2 Outline

Section 1.2 formally introduces the model with an example. Section 1.3 characterizes the equilibrium of the static and dynamic model. Section 1.4 provides some extensions of the model and Section 1.5 concludes.

1.2 Model

The model is based on the Bayesian persuasion game of Kamenica and Gentzkow (2011) extended to multiple periods. In most previous models, an experiment is a mapping from the set of true states to a set of messages. In my model, however, it is a function from the set of signals to a set of messages.

\[ \Theta \rightarrow \hat{\Theta} \rightarrow M \]

The above roughly summarizes the model. The Sender designs an experiment which is a mapping \( G : \hat{\Theta} \rightarrow M \), while it is defined as \( G : \Theta \rightarrow M \) in standard models. A signal \( \hat{\theta} \) is informative about the true state \( \theta \), but may not be perfectly correlated with \( \theta \). The function \( H : \Theta \rightarrow \hat{\Theta} \) depends on the Sender's type (competence), and therefore is not a choice variable. Although the Sender is not assumed to observe \( \hat{\theta} \)
when committing to an experiment, the competence mapping \( \mathcal{H} \) proxies the Sender’s ability. For example, if \( \mathcal{H} \) is a constant mapping, i.e., the Sender is totally incompetent, any experiment will be completely uninformative about the true state. If \( \mathcal{H} \) is an identity mapping, the Sender is perfectly competent, in which case developing a completely revealing experiment is possible.

In the dynamic model of this paper, choosing an experiment \( \mathcal{G} \) is tantamount to adjusting the speed of learning the Sender’s type. A detailed explanation is provided below after the model is formally introduced.

### 1.2.1 Model Setup

Time period runs through \( t = 1, 2, \cdots T < \infty \). At the beginning of the game, the Sender’s type is drawn according to \( \Pr[\text{Competent}] = \mu \). The Sender does not know her own type, so information is symmetric and there is no adverse selection.\(^1\) \( \theta_t \) is independent from the Sender’s type.

**States**

\( \theta_t \sim F \) with support \( \Theta = [\bar{\theta}, \tilde{\theta}] \) is drawn, where \( -\infty \leq \theta < \bar{\theta} \leq \tilde{\theta} < \infty \). \( \{\theta_t\}_{t=1}^{\infty} \) are independently and identically distributed (i.i.d.) over time. Assume that \( F \) has a strictly positive continuous density \( f \).

**Information Design**

The Sender (she) designs an experiment (model; mechanism; information structure) \( \mathcal{G}_t : \hat{\Theta}_t \to M_t \), where \( \hat{\Theta}_t \) and \( M_t \) are the signal and message space at time \( t \), respectively. As is standard in Bayesian persuasion games, the Sender commits to \( \mathcal{G}_t \) without observing \( \theta_t \) or \( \hat{\theta}_t \).

---

\(^1\)If adverse selection is present, in a dynamic model the competent type will generically design a fully transparent experiment in the first period in order to convince the Receiver of her competence ex-post. Therefore, only a pooling equilibrium is sustained in which both types disclose information fully and then the Receiver perfectly learns the Sender’s type onwards.
A signal $\hat{\theta}_t$ informative about $\theta_t$ is drawn according to

$$\hat{\theta}_t = \begin{cases} \theta_t & \text{Sender is Competent} \\ \theta'_t & \text{Sender is Incompetent,} \end{cases}$$

where $\theta'_t$ is a random variable with the same marginal as $\theta_t$ (more on $\theta'$ below). Neither $\theta_t$ nor $\hat{\theta}_t$ is observable to the Sender or the Receiver.

An experiment $G_t$ maps signals to messages. Since $\hat{\theta}_t$ and $\theta_t$ are assumed to have the same marginal, $\hat{\Theta}_t = \Theta_t$. To simplify the model, I further assume that the Sender designs a partition $P_t$ of $\hat{\Theta}_t$, and that $G_t(\hat{\theta}_t) = P_t(\hat{\theta}_t)$, where $P_t(\hat{\theta}_t)$ is the element of $P_t$ that contains $\hat{\theta}_t$. It follows that a message can be identified with an element in the partition $P_t$. The partition structure becomes a common knowledge.

**Payoffs**

Short-lived Receiver (he) can either act ($a_t = 1$) or do not act ($a_t = 0$). Assume that the Receiver’s utility is

$$u(a_t = 1, \theta_t) = \begin{cases} 0 & \theta_t < \theta_c \\ 1 & \theta_t \geq \theta_c \end{cases}$$

$$u(a = 0, \theta_t) = y \in (0, 1),$$

where $\theta_c \in (\theta, \hat{\theta})$ and $0 < y < 1$ are given. $y$ is the Receiver’s outside option when no action is taken. Similar to Kamenica and Gentzkow (2011), I assume that when the Receiver is indifferent between the two actions, he chooses the Sender-preferred-action $a = 1$. It then follows that the Receiver will choose $a_t = 1$ if and only if

$$Pr[\theta_t \geq \theta_c | \mathcal{I}^t] \geq y,$$

where $\mathcal{I}^t$ denotes the information set generated by the past history.

The Sender’s period-$t$ utility is given by

$$v(a_t, \theta_t) = b_t \cdot 1_{\{a_t = 1\}},$$

---

2 The assumption that $\theta_t$ and $\theta'_t$ are identically distributed is not important. See the Assumptions section.

3 This restriction implies that the sender cannot adopt a mixed strategy.
where \( b_t \geq 0 \) represents the value of persuasion in period \( t \), and \( 1_{\{\cdot\}} \) is the indicator function. In the benchmark model, \( b \) is assumed to be deterministic and therefore depends only on time.\(^4\) However, in real world applications, \( b \) might be stochastic. For example, \( b \) might depend on the state of the economy. I later discuss how the results change when \( b \) is not deterministic.

### Learning

At the end of the period, the event \( P_t(\theta_t) \in \mathcal{P}_t \) is publicly observed. The Receiver updates the Sender’s type based on the realization.

\[ \begin{align*}
\text{Sender is competent: } & \hat{\theta} = \theta \\
\text{Sender is incompetent: } & \hat{\theta} = \theta' \\
\text{Observe } & P(\hat{\theta}) \text{ before action} \\
\text{Observe } & P(\hat{\theta}) \text{ after action} \\
\text{Positively update} & \\
\text{Negatively update} &
\end{align*} \]

**Figure 1-1:** Example of ex-post updating rule

\[ \begin{align*}
\text{Period } t \\
\text{Nature draws } S's \text{ type} & \\
S \text{ designs experiment} & \\
\text{Nature draws } \theta, \theta' & \\
\text{Message sent} & \\
S, R \text{ observe } P(\hat{\theta}) & \\
\text{Update on } \mu &
\end{align*} \]

**Figure 1-2:** Timeline of the benchmark model

### 1.2.2 Model Discussion

\( \theta_t \) is assumed to be independent across time, so any information on \( \theta_t \) is irrelevant to future predictions of \((\theta_s)_{s \geq t}\). Therefore, the Sender’s choice of current period’s

\(^4b_t \) can be thought as the Sender’s relative Pareto weight in welfare analysis.
experiment is affected only by her incentive to persuade the Receiver in the current period and to manage the future perception of her ability. Devising a very informative mechanism can risk the Sender’s future reputation; if the ex-post observation conforms to the received message, the Receiver will positively update the Sender’s quality. Otherwise, the Receiver will negatively update.

Provided that the Receiver’s action is kept fixed, choosing the level of informiveness of an experiment is tantamount to selecting the dispersion of posterior distributions of Sender’s type, and consequently adjusting the speed of learning the Sender’s type. If the Sender likes (dislikes) to gamble for her perceived ability in equilibrium, she prefers to reveal as much (little) information as possible and expedite (delay) the learning process of her true type. One extreme is a case where the experiment does not convey any information, i.e., $\mathcal{P} = \{[\underline{\theta}, \bar{\theta}]\}$. In this case, regardless of ex-post realization, no learning occurs. In another extreme where the experiment is fully transparent, i.e., $\mathcal{P} = \cup_{\hat{\theta} \in \Theta}\{\{\hat{\theta}\}\}$. Under this scheme, complete learning takes place immediately. It is a priori unclear which option is more preferable by the Sender, and one of the main objectives of this paper is to determine conditions under which the Sender prefers one over the other.

The Receiver’s utility takes a rather simple form; he only assesses the probability that the true state is above the threshold $\theta_c$, not, for example, the average value of the true state. This formulation is used for analytical simplicity, as probabilities are invariant under the inverse transformation and all relevant quantities can be expressed in terms of the underlying copula. If there are multiple receivers and their actions have strategic complementarity (e.g. bank depositors deciding to run on banks), the usual global games approach à la Morris and Shin (1998) yields that the outcome depends on whether the fundamental is above or below a threshold, so the model can be interpreted as a reduced-form model of a coordination game with strategic complementarity (see Goldstein and Huang (2016)).

Ex-post learning depends heavily on the chosen experiment. For example, if the mechanism sends out a binary message ‘Pass’ or ‘Fail,’ the Receiver can only observe whether the entity in question indeed passed or failed ex-post. This formulation is
suitable when the fundamental (state) is hard to observe accurately. For example, in a bank stress setting (Goldstein and Leitner (2017) and Faria-e-Castro et al. (2017)), since factors driving a bank to fail are complex, depositors are likely to observe only whether ex-post results are consistent with the authority’s prediction (whether bank failed). However, in other settings such as projecting the next year’s inflation rate, the receiver can observe the realized outcome with high precision. I consider this variation of the model in which the receiver learns \( \theta_t \) perfectly ex-post, and show that many of the results still hold under the alternative formulation.

1.2.3 Assumptions

Denote the copula of \((\theta, \theta')\) by \(C(\cdot, \cdot)\). That is, \(C\) is a bivariate function such that for all \(u, v \in \mathbb{R}\),

\[
Pr[\theta \leq u, \theta' \leq v] = C(F(u), F(v)).
\]

Throughout this paper, I assume that \(F\) has a strictly positive and continuous density \(f\), and that \(C\) is twice continuously differentiable almost everywhere. The partial derivatives of \(C\) are denoted by subscripts, e.g. \(C_1\) and \(C_2\). Also, let \(\delta : [0, 1] \to [0, 1]\) defined by \(\delta(x) = C(x, x)\) be the diagonal section of the copula \(C\).

\(A1.\) \(1 - F(\theta_c) < y.\)

\(A2.\) Each set in \(P_t\) is a Borel set \(\subset [\theta, \bar{\theta}]\).

\(A3.\) (Positive Stochastic Monotonicity Dependence; \(PSMD^6\)):

\(Pr[\theta \leq x | \theta' = x']\) is decreasing in \(x'\) for all \(x\). In other words, for \(x'_1 > x'_2\), the posterior distribution \(G(\cdot | x'_1)\) can be ranked according to the first order stochastic dominance:

\[
G(\cdot | x'_1) \succeq G(\cdot | x'_2)
\]

\(^5\)It is known that \(0 \leq \delta'(x) \leq 2\) for all \(x \in [0, 1]\). There is no copula such that its diagonal section is strictly concave, because \(\delta(0) = 0, \delta(1) = 1, \delta(x) \leq x\). For a Gaussian copula with positive correlation, \(\delta\) is convex. See Nelsen (2006) for details.

\(^6\)This terminology is taken from Fang and Wu (2018).
It will be useful for future purpose to rewrite the assumptions in terms of the inversely transformed variables.

Defining \( \omega = F(\theta) \), and \( \omega' = F(\theta') \), we see that each of \( \omega \) and \( \omega' \) is uniformly distributed. If we let \( \bar{F}_{\omega|\omega'}(\omega|\omega') = F_{\theta|\theta'}(F^{-1}(\omega)|F^{-1}(\omega')) \), all of the above assumptions still hold. \( \omega \)-denoted variables should always be construed as the inversely transformed \( \theta \)-variables.

The assumptions can be rewritten as:

\[ A1. \ 1 - \omega_c < y. \]

\[ A2. \EE \text{Each set in } \mathcal{P}_t^\omega \EE{\text{is a Borel set } \subset [0,1].} \]

\[ A3. \ (PSMD) \ C_{22}(u,v) \leq 0 \EE{\text{for all } u,v \in [0,1]^2.} \]

\[ A4. \ P[\omega \geq \omega_c | \omega' \geq \omega_c] > y. \]

In the language of copula, \( A4 \) can be reformulated as

\[ 1 - \frac{\omega_c - C(\omega_c, \omega_c)}{1 - \omega_c} > y. \]

**Interpretation**

\( A1 \) assumes that the Receiver’s default action absent any new information is \( a_t = 0 \). Otherwise, the Sender always can achieve \( a_t = 1 \) by choosing the coarsest partition \( \mathcal{P}_t = \{ [\theta, \bar{\theta}] \} \), so the first best is trivially attained for the Sender.

\( A3 \) assumes that the distribution of \( \theta \) conditional on \( \theta' \) is monotonic in the stochastic dominance sense (see Fang and Wu (2018)). Although \( \theta' \) is a noisy proxy for \( \theta \), it is required that \( \theta' \) is positively correlated with \( \theta \) and therefore \( \theta' \) contains information about \( \theta \). Without this assumption, the message “\( \theta \) is likely to be high” can in fact convey more negative information than the message “\( \theta \) is likely to be low.” Other

\footnote{A necessary condition for both \( A1 \) and \( A4 \) to hold is that \( C(\theta_c, \theta_c) \geq \theta_c^2 \). In fact, \( PSMD \) implies \textit{positive quadrant dependence} property, so if \( (X, Y) \) is \( PSMD \), then \( C(x, y) \geq xy \forall 0 \leq x, y \leq 1 \). See Nelsen (2006) for details.}
than PSMD, for my baseline results no other assumptions are made on the dependence structure between the states and signals. In most information models, only few class of distributions are used (e.g. binary or Gaussian) simply because of their tractability. It is not clear whether their results hold under general distributions, and they can even be reversed (see the “anything goes” result in Breon-Drish (2015)).

A4 ensures that even the worst type Sender ($\mu = 0$) can persuade the Receiver with positive probability when a message is positive enough. Suppose this were not true. Then incompetent types (say $\mu < \bar{\mu}$) are never able to persuade the Receiver unless they alter the Receiver’s perception. In my model, the Sender’s utility only depends on whether the Receiver acts, so all types below $\bar{\mu}$ have the same static payoff of 0. Consequently, as in exercising a call option, an incompetent Sender will prefer to gamble for her reputation, i.e., make a narrow prediction and hope it turns out to be correct. In real life applications, the Sender’s ability may have a direct effect on her payoff as well as an indirect effect through persuasion power. Instead of including the direct effect in the utility function, I assume that even the worst type is able to persuade the Receiver by pooling extremely good states; a more incompetent type has lower likelihood of successful persuasion, because she needs to pool better states in order for the Receiver to act. For the sake of simplicity, I maintain the assumption that the Sender’s payoff only depends on the binary outcome.

In fact, imposing the existence of $x < \bar{\theta}$ such that $Pr[\theta \geq \theta_c | \theta' \geq x] > y$ suffices for the above points to hold. Instead, I assume A4 for algebraic simplicity; otherwise some competent types are able to persuade the Receiver with a message “[x, $\bar{\theta}$]” where $x < \theta_c$, but others require $x \geq \theta_c$. Since the Receiver only calculates the conditional probability that $\theta \geq \theta_c$, the weaker assumption leads to a kink in the Sender’s value function.

The assumption that $\theta$ and $\theta'$ have the same marginal can be relaxed with some adjustments. If they have different marginals (e.g. different supports), apparently a partition of $[\theta, \bar{\theta}]$ cannot be treated as an information transmission mechanism. Alternatively, I can assume that an experiment is a partition of $[0, 1]$ and the message corresponds to whether $\omega$ falls into one of the partition interval. With this modifica-
tion, all of my results still hold.

Examples of PSMD

Since \textit{PSMD} is a key assumption of this paper, it is useful to see whether \textit{PSMD} is satisfied under some of widely used settings in economics.

1. Let \( Y = X + \epsilon \), where \( X \) and \( \epsilon \) are independent Gaussian random variables. Then, \((X, Y)\) satisfies \textit{PSMD}.

2. Let \( X \) be any random variable, and define \( Y = 1_{\{X \geq c\}} \) for some given constant \( c \). Then, \((X, Y)\) satisfies \textit{PSMD}.

3. Suppose \( \{z_i\}_{i=1}^n \) are \( i.i.d. \) and \( X = \min\{z_1, z_2, \cdots, z_n\} \) and \( Y = \frac{\sum_{i=1}^n z_i}{n} \). Then, \((X, Y)\) satisfies \textit{PSMD}.

1.2.4 Relationship to Standard Bayesian Persuasion

Before characterizing the equilibrium, it is constructive to see an example that illustrates how my model differs from the that of standard Bayesian persuasion. Below is a discrete state version of my model, built on the baseline example of Kamenica and Gentzkow (2011).

Example. There are two states \( \Theta = \{G, B\} \) with \( Pr[\theta = G] = p \). The set of signals is \( \Theta' = \{g, b\} \) and \( \theta' \) is positively correlated with \( \theta \) by

\[
Pr[\theta' = s|\theta = S] = (1 + \rho)/2,
\]

with \( \rho \in [0, 1] \) for \( S = G, B \) (Kamenica and Gentzkow (2011) corresponds to \( \rho = 1 \)). The Receiver’s action set is \( \{0, 1\} \), and the Receiver acts if and only if the probability of the good state is at least \( y \in (p, 1) \). The Sender derives utility of \( 1_\{1\}_1 \) and designs an experiment \( h : \hat{\Theta} \to \Delta(M) \). Appealing to the Revelation principle in Kamenica and Gentzkow (2011), without loss of generality assume \( M = \{0, 1\} \). The Sender can be one of the two types: \textit{Competent} or \textit{Incompetent}. If Competent, \( \hat{\theta} = \theta \), and if Incompetent, \( \hat{\theta} = \theta' \). The Sender’s type is independent from other states,
and \( Pr[\text{Competent}] = \mu \). A simple algebra exercise yields that the set of feasible posteriors \( \alpha_m = Pr[G|m] \) is characterized by

\[
\frac{p(1-\mu)(1-\rho)}{p(1-\mu)(1-\rho) + (1-p)\left(\mu + \frac{(1-\mu)(1+\rho)}{2}\right)} \leq \alpha_m \leq \frac{p\left(\mu + \frac{(1-\mu)(1+\rho)}{2}\right)}{p\left(\mu + \frac{(1-\mu)(1+\rho)}{2}\right) + \frac{(1-\rho)(1-\mu)(1-\rho)}{2}}
\]

subject to the usual Bayesian plausibility constraint (the average of posteriors should be the prior). Under the following experiment, the upper bound is attained for \( \alpha_1 \) and the lower bound for \( \alpha_0 \):

\[
h(m=1|g) = 1, \ h(m=1|b) = 0.
\]

It can be understood that the uncertainty of signal accuracy puts a restriction on the Receiver's posterior. The Sender's ability to use a precise input to an experiment translates to her capacity to alter the Receiver's belief freely. Intuitively, if the Sender is known to be incompetent and \( \theta' \) is pure noise (\( \rho = 0 \)) there is no way for the Sender to modify the Receiver's belief. If there are multiple states, the feasible set of posteriors is a convex set governed by multiple inequalities.

Similar to the usual approach, the equilibrium can be characterized by optimizing over all feasible posteriors, which corresponds to concavifying the Sender's value function. The difference is that the problem is a constrained optimization. With this in mind, we can now proceed to solving for equilibrium of the benchmark model.

### 1.3 Benchmark Equilibrium

First, let us establish a technical result. All omitted proofs are provided in Appendix.

**Lemma 1.1** \( Pr[\theta \geq \theta_c | x \leq \theta' \leq y] \) is strictly increasing in \( x \) and \( y \), respectively.

**Proof.** See Appendix. 

Intuitively, by PSMD, \( \theta \) is likely to be high if \( \theta' \) is. Lemma 1 implies that a message \( \hat{\theta} \in [a, b] \) is more positive than a message \( \hat{\theta} \in [c, d] \) for \( a \geq d \).
1.3.1 Static model

Let us characterize the equilibrium information structure in the single period model, i.e. when $T = 1$. In this case, we can assume $b_1 = 1$. For any partition $\mathcal{P}$ of $[\theta, \bar{\theta}]$, write $\mathcal{P} = \mathcal{P}^{a=0} \cup \mathcal{P}^{a=1}$, where

$$\mathcal{P}^{a=0} = \bigcup_{P \in \mathcal{P}} \{P : Pr[\theta \geq \theta_c | \hat{\theta} \in P] < y\}$$
$$\mathcal{P}^{a=1} = \bigcup_{P \in \mathcal{P}} \{P : Pr[\theta \geq \theta_c | \hat{\theta} \in P] \geq y\}.$$

The Sender’s problem is to solve

$$\sup_{\mathcal{P}} \left[ \sum_{P \in \mathcal{P}^{a=1}} Pr[\hat{\theta} \in P] \right].$$

Since the Receiver’s action is binary, by the Revelation principle in information design (see Kamenica and Gentzkow (2011)) analogous to the one in Myerson (1979), it is without loss of generality to assume that in the static model the message space is binary. The intuition is that the message space can be divided into two groups (say $M_0$ and $M_1$), each of which being the set of messages that lead to the same action in equilibrium. Then, all messages in $M_i$ can be relabeled as a single message $m_i$. In my model it turns out that the binary partition can be characterized by a single threshold.

**Proposition 1.1** In the static model, for any given $\mu$, the equilibrium outcome can be implemented by an experiment that exhibits a binary partition $\mathcal{P} = \{[\theta, z^*_\theta(\mu)),[z^*_\theta(\mu),\bar{\theta}]\}$. The threshold $z^*_\theta(\mu)$ is uniquely defined by the implicit equation

$$Pr\left[\theta \geq \theta_c | \hat{\theta} \geq z^*_\theta(\mu)\right] = y.$$

**Proof.** See Appendix. □

There are several observations to make. The Sender’s choice of information partition follows a cutoff strategy. Intuitively, pooling the highest states is least costly for the Sender to persuade the Receiver, i.e., the set $A$ that maximizes $Pr[\theta \geq \theta_c | \hat{\theta} \in A]$ with fixed $Pr[\hat{\theta} \in A]$ is an upper-half interval.

In equilibrium, the Receiver acts on the positive message ($\hat{\theta} \in [z^*_\theta, \bar{\theta}]$) and does
not on the negative message. When he acts, he is indifferent between acting and not acting. The reason is that if he strictly prefers to act, the Sender can strictly increase the probability of successful persuasion by lowering the threshold while maintaining the Receiver to act on the positive message. Consequently, the Receiver’s equilibrium payoff is equal to his outside option \( y \).

Although the threshold is unique, there are many experiments that lead to the same equilibrium outcome in the static model. In fact, any partition \( \mathcal{P} = \mathcal{P}' \cup \{[\theta, z^*_0(\mu)], \theta] \} \) works, where \( \mathcal{P}' \) is an arbitrary partition of \([\theta, z^*_0(\mu)]\). The reason is that for any event in \( \mathcal{P}' \) the Receiver does not act; for any \( P' \in \mathcal{P}' \), \( Pr[\theta \geq \theta_c | \tilde{\theta} \in P'] < y \). The Sender only cares whether the Receiver acts in an one-shot game so any partition of \([\theta, z^*_0(\mu)]\) leads to the same outcome.

The multiplicity of equilibrium information structures is a generic feature of static Bayesian persuasion games. However, as is shown in the next section, the equilibrium information structure is unique in the dynamic model with reputational concern.

**Theorem 1.1** In the static model, the threshold \( z^*_0(\mu) \) is strictly decreasing in \( \mu \), and the Sender’s value function \( V(\mu) \) is strictly increasing in \( \mu \). Moreover, \( V(\mu) \) is strictly concave in \( \mu \).

**Proof.** See Appendix. ■

Intuitively, the Sender is better off when she is able to use more precise signal. Note that this is also the case in the example from the previous section; higher \( \mu \) corresponds to the Sender’s latitude in selecting the Receiver’s posteriors. It is not, however, clear whether the Sender would prefer to gamble for her reputation by designing an informative structure and hope the truth is ruled in her favor. Theorem 1.1 establishes that she prefers not to disclose information more than necessary; her utility is concave in her competence, and therefore chooses to reveal as little information as possible subject to the persuasion constraint.

For illustration, I consider three joint distributions that can be applied to various settings. First, \((\theta, \theta')\) is assumed to be jointly normal. This formulation is undeniably the most frequent one in information based models. Second, I suppose that
\[ \theta = \min\{A_1, A_2\}, \text{ and } \theta' = (A_1 + A_2)/2, \] where \( A_1, A_2 \) are independent random variables. In some cases, an economic outcome is determined by the worst realization of the fundamentals. For instance, a regulator may care about the failure of the banking system which is triggered by the worst performance of few banks. However, the regulator may not be able to run an experiment based on these trigger signals but rather use the average one such as overall credit quality. Another instance is when a credit rating agency (CRA) develops a credit risk model that evaluates the default probability of a firm (see Goldstein and Huang (2017)). For the exposition, I assume both \( A_1 \) and \( A_2 \) are standard Gaussian random variables. Third, I use an Archimedean copula defined by \( C(u, v) = \phi^{-1}(\phi(u) + \phi(v)) \), where \( \phi \) is a well-behaved generator function. This class of copulas is widely used in practice because it is easily
generalizable to multivariate distributions with a small number of parameters that characterize dependence structure.

The next two lemmas will be useful later in characterizing the equilibrium in the dynamic model when additional assumptions are imposed on the joint distribution.

**Lemma 1.2** Suppose $C_{22}(u, v) \leq 0$ for $u \geq v$. Then, the single period value function $V$ satisfies $V'' \geq 0$.

**Proof.** See Appendix. $lacksquare$

Below is an example in which Lemma 1.2 is true.

**Example.** Suppose the copula is given by the Farlie-Gumbel-Morgenstern copula:

$$C(u, v) = uv + \eta uv(1 - u)(1 - v), \; 0 < \eta \leq 1/3.$$ 

If $A1, A2$, and $A4$ are satisfied, then the results of Lemma 1.2 hold.

**Lemma 1.3** Suppose $C$ is the Gaussian copula with correlation $\rho > 0$. That is,

$$C(u, v) = \Phi_\rho \left( \Phi^{-1}(u), \Phi^{-1}(v) \right),$$

where $\Phi_\rho$ is the bivariate Gaussian cumulative distribution function with correlation $\rho$ and $\Phi^{-1}$ is the inverse of the standard Gaussian cumulative distribution function. If any of the following conditions hold, $V'' \geq 0$.$^8$

- **Condition 1:** $1 - (1 - \omega_c)/\gamma \geq 1/2$.
- **Condition 2:** $\rho \geq 1/\sqrt{2}$ and $\omega_c \geq 1/2$.
- **Condition 3:** $\rho \leq 1/\sqrt{2}$, $\omega_c \leq 1/2$, and $1 - (1 - \omega_c)/\gamma \geq \zeta(y, \rho)$, where $\zeta(y, \rho)$ is the unique solution of

$$\sqrt{2\pi e^{\frac{\Phi^{-1}(\rho)^2}{2}}} \Phi^{-1}(x) = \frac{9y(1 + \rho)}{1 + 2\rho}$$

**Proof.** See Appendix. $lacksquare$

---

$^8$From numerical simulations, the result seems to hold for any $\rho > 0$, but I provide the conditions on the primitives under which the observation can be proven analytically.
1.3.2 Dynamic model

In the dynamic model, the Sender’s payoff is given by

$$\sum_{t=1}^{T} b_t \cdot \mathbb{1}_{\{a_t=1\}}$$

where $T < \infty$, and $b$ is a non-negative deterministic function.

The Sender cares not only about her current payoff but also her future payoffs. The latter depends on her reputation in future periods, and the current choice of experiment affects the posterior of her reputation; a highly informative experiment enables the Receiver to learn her ability accurately, while an uninformative one does not.

The Revelation principle does not hold in the dynamic setting. The same logic applied before — grouping messages by the action they induce — cannot be appealed because relabeling messages alters the distribution of posteriors about the Sender’s ability; although refining the negative message has no effect on current utility, it changes the way the Receiver evaluates the Sender’s capacity, and therefore leads to a different utility outcome. However, the next proposition and its corollary establish that even in the dynamic model the equilibrium partition is binary and, more importantly, unique unlike in the static model.

To simplify the analysis for the dynamic model, I assume a stronger version of $A_2$, similar to the one in Leitner and Yilmaz (2017): $A_2'$. Each set in $P_t$ is convex.

That is, I impose each partition element to be an interval. Coupled with $PSMD$, it is clear that an experiment assigns more positive message to higher states in equilibrium. Under $A_2'$, the Sender is prohibited to develop an experiment that contains results such as “$	heta$ is likely to be either very high or very low.” Although $A_2'$ is imposed for simplicity, it is a reasonable one; in a bank stress test setting, the regulator may find it politically hard to devise such a model, pooling the two extreme set of states into the same message. In addition, if the state of the nature is an outcome of another agent’s action (e.g. financial health of an individual bank as in Goldstein and
Leitner (2017)), a non-monotonic recommendation rule can result in moral hazard in the sense that the agent may prefer to discard its value (e.g. bank assets) in order to obtain a better recommendation.

Hereafter, $A2'$ is assumed instead of $A2$.

**Theorem 1.2** If $V : [0, 1] \rightarrow \mathbb{R}$ is strictly increasing and concave, for given $\mu$ the optimal information partition $\mathcal{P}(\mu)$ that solves

$$
\sup_{\mathcal{P}} \left[ Pr[a = 1; (\mu, \mathcal{P})] + \int_0^1 V(\mu')dQ_{\mathcal{P}}(\mu') \right]
$$

(1.1)

is a binary partition $\{[\theta, z^*_1), [z^*_1, \bar{\theta}]\}$, where $Q_{\mathcal{P}}$ is the measure of the posterior of $\mu$ that $\mathcal{P}$ induces. The optimal information partition is unique.

**Proof.** See Appendix.

In general, solving (1.1) is complex, as the argument is infinitely-dimensional. However, if $V$ is strictly concave, the solution is a binary partition with a single cutoff. Thanks to this feature, the optimization becomes a simple one-dimensional problem. Applying Theorem 1.1 and Theorem 1.2, the next corollaries follow.

**Corollary 1.1** In the two period model, the optimal information partition in the first period is binary and unique: $\mathcal{P}_1 = \{[\theta, z^*_1), [z^*_1, \bar{\theta}]\}$.

**Corollary 1.2** In the $T$ period model, if $b_1 = b_2 = \cdots = b_{T-1} = 0$ and $b_T > 0$, the optimal information partition in periods $t = 1, 2, \cdots, T - 1$ is the coarsest partition: $\mathcal{P}_t = \{[\theta, \bar{\theta}]\}$.

The uniqueness result is in sharp contrast to that of a standard static Bayesian persuasion model. To illustrate this point, consider the leading example of Kamenica and Gentzkow (2011) (see the example in section 2.4). If two experiments induce the Receiver to act with equal probability, the Sender is indifferent between the two. For example, if in equilibrium the message $m = 0$ leads to $a = 0$, the Sender might as well choose to reveal the true state instead of sending $m = 0$. Although the alternative yields the same expected payoff to the Sender, the information content
of the mechanism is very different. That is, with a static model it is generically impossible to quantify the amount of information that is revealed in equilibrium.

In the dynamic model with reputation, if the next period value function is strictly concave in \( \mu \), the Sender dislikes to gamble for her reputation. That is, if two information structures lead to the same payoff in the current period, she prefers the one that minimizes the dispersion of the posterior of \( \mu \). Since any event of a given partition \( \mathcal{P}' \) of \([\theta, z^*_i]\) induces no-action, the equilibrium partition of the lower interval is the lower interval itself.

If the Sender chooses a binary partition with threshold \( z_{\theta t} \), her expected utility at time \( t \) can be written by

\[
b_t I(z_{\theta t} \geq z^*_0(\mu_t)) \cdot (1 - F(z_{\theta t})) + CV_{t+1}(\mu_t, z_{\theta t})
\]

\[
= b_t I(z_{\omega t} \geq z^*_0(\mu_t)) \cdot (1 - z_{\omega t}) + CV_{t+1}(\mu_t, z_{\omega t}) := J_t(\mu_t, z_{\omega t}),
\]

where

\[
CV_{t+1}(\mu_t, z_t) = \lambda_t(V_{t+1}(\mu_t) - V_{t+1}(0)) + \lambda_h(V_{t+1}(\mu_h) - V_{t+1}(0)) + V_{t+1}(0),
\]

\[
\lambda_t = \mu_t z_t + (1 - \mu_t) C(z_t, z_t)
\]

\[
\mu_t = \frac{\mu_t z_t}{\lambda_t}
\]

\[
\lambda_h = \mu_t (1 - z_t) + (1 - \mu_t) (1 - 2 z_t + C(z_t, z_t))
\]

\[
\mu_h = \frac{\mu_t (1 - z_t)}{\lambda_h},
\]

and \( V_{t+1} \) and \( z^*_o \) are the Sender's value function at \( t + 1 \) and the single period policy threshold, respectively.

\( \lambda_h \ (\lambda_t) \) is the probability that the Sender's high (low) message conforms to the Receiver's ex-post observation. \( \mu_h \) and \( \mu_l \) are the corresponding posterior reputation based on a positive update.

If \( V_{t+1} \) is strictly concave, because the average of the posteriors must be the prior, \( CV_{t+1}(\mu_t, \cdot) \) attains the unique maximum at \( z_{\omega t} = 0 \) and \( z_{\omega t} = 1 \). These policies correspond to the coarsest information partition from which no information is revealed to the Receiver. Although by choosing these policies the Sender can claim the highest
continuation value next period, she cannot persuade the Receiver in period \( t \), because the Receiver's default action is \( a_t = 0 \).

(1.2) shows the trade-off that the Sender faces. The first term represents the probability of successful persuasion in period \( t \). By choosing a high enough threshold \( z_{w} \), she can guarantee that the Receiver chooses \( a_t = 1 \) upon getting the good message. She would not like to choose too high \( z_{w} \), as the probability of persuasion diminishes. However, choosing an interior \( z_{w} \) widens the distribution of the posterior of \( \mu \). If she is risk-averse to a reputation gamble, she might want to choose a higher threshold. The overall effect is not clear. Intuitively, if the former effect is stronger, she would choose the myopic policy \( z_{w} = z_{w0}(\mu_t) \). If the latter is stronger, \( z_{w} = 0 \) or \( z_{w} = 1 \) would be optimal.

![Figure 1-4: Continuation value for various distributional assumptions and \( \mu = 0.2, 0.5, 0.8 \) (from bottom to top). The distributions and parameters are the same as the ones in Figure 1-3.](image)

In Figure 1-4 the continuation value is plotted for the two-period model under various distributional structures as in Figure 1-3. Since the single period value function
is strictly concave (Proposition 1.1), \( CV \) attains its maximum at the endpoints. One could also conjecture that \( CV(\mu, \cdot) \) is convex. The reason is that when information partition is binary, more information is conveyed in the entropic sense when the cutoff is closer to the middle (median for symmetric distributions). Since the Sender wishes not to give more information than necessary, \( CV \) will have a unique global minimum in the interior.

Indeed, the conjecture seems to hold under the assumptions of my model. However, actually proving the convexity for general distributions is quite involved, so instead I give few sufficient conditions under which the conjecture is true. For some cases I establish the result analytically. The next lemma is used to prove the convexity of \( CV(\mu, \cdot) \) when \( \theta \) and \( \theta' \) are jointly Gaussian.

**Lemma 1.4** Suppose that a copula \( C \) satisfies PSMD and \( \delta \) is its diagonal section. If \( V_t < 0, V_t'' \geq 0 \), and
\[
\frac{\delta(z)\delta''(z)}{2} \geq \left( \delta'(z) - \frac{\delta(z)}{z} \right)^2 \quad \text{and} \quad \frac{(1 - 2z + \delta(z))\delta''(z)}{2} \geq \left( 1 - \frac{\delta(z)}{1 - z} - \delta'(z) \right)^2
\]
for all \( z \in (0, 1) \), then \( CV_{t+1}(\mu, \cdot) \) defined in (1.3) is convex in the second argument. In particular, the inequalities hold when \( C \) is a Gaussian copula with correlation \( \rho > 0 \).

**Proof.** See Appendix. 

If convexity of \( CV(\mu, \cdot) \) is assumed, the Sender's optimal choice of experiment can be describe formally as follows:

**Theorem 1.3** Suppose that \( CV_{t+1} \) defined in (1.3) is convex in the second argument, and that \( V_t'' < 0 \). Then, there exists \( \bar{b}_t(\mu) \) such that:

1. If \( b_t \leq \bar{b}_t(\mu) \), the Sender chooses the trivial information structure in period \( t \), i.e., \( \mathcal{P}_t = \{[\theta, \emptyset]\} \).

2. If \( b_t > \bar{b}_t(\mu) \), the Sender chooses the myopic policy in period \( t \), i.e., \( \mathcal{P}_t = \{[\theta, z^*_0(\mu)], [z_0^*(\mu), \emptyset]\} \), where \( z^*_0 \) is the single period threshold.

**Proof.** See Appendix. 

38
Corollary 1.3 Assume that $(\theta, \theta')$ is normally distributed with correlation $\rho > 0$ and $\omega_c$ and $y$ satisfy any of the conditions in Lemma 1.3. Then, in the two period model, for any given $\mu$, there exists $\bar{b}(\mu)$ such that:

1. If $b_1 / b_2 \leq \bar{b}(\mu)$, the Sender chooses the coarsest information structure in the first period, i.e., $P_1 = \{[\theta, \bar{\theta}]\}$.

2. If $b_1 / b_2 > \bar{b}(\mu)$, the Sender chooses the myopic policy in the first period, i.e., $P_1 = \{[\theta, z^*_\theta(\mu)], [z^*_\theta(\mu), \bar{\theta}]\}$, where $z^*_\theta$ is the single period threshold.

Proof. Follows from Lemma 1.3, Lemma 1.4 and Theorem 1.3.

Theorem 1.3 establishes that if the continuation value is convex in policy threshold, the Sender’s optimal information structure is either the coarsest or myopic one, and the choice depends on the current value of persuasion relative to the future value. If it is valuable for the Sender to persuade the Receiver today (impatient), the Sender adopts the myopic policy and maximizes the probability of persuasion today. If the Sender puts more weight on the future (patient), she chooses the most conservative policy - not disclosing any information today due to reputational concern. I call this policy “dynamic” as opposed to “myopic.” The decision cutoff $\bar{b}(\mu)$ is not monotonic in $\mu$. The reason is that for extreme types ($\mu \approx 0$ or $\mu \approx 1$) the posterior distribution of reputation is not sensitive to policy choice. Consequently, they opt to persuade today even the value is low. For intermediate types, however, designing an informative experiment is risky as the posterior of their ability varies a lot depending on their policy. The value of persuasion today needs to be high enough to incentivize these types to transmit information today.

An (counter-intuitive) implication is that the quality of experiments (amount of information) does not necessarily correspond to the Sender’s quality (ability to exploit precise signal). For example, consider a media company with bad reputation that attempts to predict an election outcome. According to my model, this company might be more willing to devise a more useful prediction model than other firms with better skills. The reason is that any new ex-post observation (either right or wrong) does not have a huge impact on its perceived ability due to a strong prior; the firm optimally
chooses to derive utility in the current period rather than waiting until future periods. Another example can be found in analyst forecasts of stock returns. Medium-rated analysts are most likely to suffer from reputational loss due to mistakes, and thus are most likely to place conservative bets. As a result, their predictions contain least amount of information, and clients might be better off subscribing to less renowned analysts.

Figure 1-5 illustrates the above points. In the two period model, the decision threshold \( \tilde{b}(\mu) \) has an inverse U-shape (left panel). In the middle panel, the Sender’s value function is plotted for various values of \( b_1 \), fixing \( b_2 = 1 \). Although the value function is increasing in \( b_1 \), it is not strictly increasing. That is, for intermediate types, higher value of persuasion today \( (b_1) \) may not make them better off, because they decide not to disclose any information anyways (and therefore fail to persuade the Receiver in the current period). The right panel illustrates how the Sender’s value function differs from her expected payoff when she chooses the fully transparent experiment.

If more information leads to higher social welfare, the Sender’s reputational concern can exacerbate loss of welfare.\(^9\) Consider a variation of my model in which there are some receivers who make decisions based upon the information generated by the Sender’s experiment but whose action does not affect the Sender’s utility. For example, 1) a company or bank develops a risk model addressed to only a subset of regulators, or 2) a group of shareholders needs to persuade only the majority in a proxy fight. For illustration, assume that there are continuum of receivers and their outside options are distributed according to \( p(\cdot) \), and that the Sender needs to persuade a certain fraction of the receivers, say the ones with \( y < \bar{y} \). In equilibrium the non-targeted receivers \( (y > \bar{y}) \) will claim their outside option. Absent any information, the targeted receivers’ aggregate welfare is given by \( W_0 = \int_{\bar{y}}^{\bar{y}} yp(y)dy \). Under the Sender’s myopic policy, it is given by \( q \int_{0}^{\bar{y}} (\bar{y} - y)p(y)dy + W_0 \), where \( q \) is the probability that the receivers

\(^9\)Of course, in my benchmark model of the Sender and Receiver, in equilibrium the Receiver’s expected utility is always his outside option \( y \); the experiment optimally chosen by the Sender is such that the Receiver is indifferent between acting and not acting conditional on the positive message. As a result, social welfare can be measured by the Sender’s expected payoff.
positive message is sent (note that under this policy the positive message is tailored such that “θ ≥ θ_c” occurs with probability $\hat{y}$). If $b_1 \ll b_2 \ll 1$, due to her reputational concern, the Sender runs an uninformative experiment in the first period, although social welfare is higher under the myopic policy. In this case, destroying the long-term value ($b_2$) or boosting the short-term value ($b_1$) of persuasion can increase social welfare. An example of such a scheme is to enforce a contract (if available) with the Sender to be renewable based on short-run performance.
1.4 Extensions

1.4.1 Perfect ex-post learning

It has been assumed that the Receiver’s time-t ex-post observation is restricted within the Sender’s information partition at time t, i.e., \( P_t(\theta_t) \) is observed. In this subsection, I consider a (exogenous) perfect learning mechanism, i.e. the actual \( \theta_t \) is observed ex-post. This mechanism is suitable for situations in which the Sender predicts an outcome that can be accurately verified ex-post (e.g. predicting inflation rate or the outcome of a sports match). All the assumptions (A1-A4) remain the same. There is no learning in the static model, so Proposition 1.1 and Theorem 1.1 hold. The next proposition asserts that Corollary 1.1 remains the same under the alternative formulation.

**Proposition 1.2** In the two period model with perfect ex-post learning, the optimal information partition in the first period is binary and unique.

**Proof.** See Appendix. □

In the two-period model, the value function is given by

\[
V(\mu) = \sup_{z_\omega \in [0,1]} \left[ b_1 \mathbb{1}_{z_\omega \geq z^*_{\omega 0}(\mu)} \cdot (1 - z_\omega) + b_2 \int_{0}^{z_\omega} (\mu + (1 - \mu)C_1(x, z_\omega)) V_0 \left( \frac{\mu}{\mu + (1 - \mu)C_1(x, z_\omega)} \right) dx \right. \\
+ b_2 \int_{z_\omega}^{1} (\mu + (1 - \mu)(1 - C_1(x, z_\omega))) V_0 \left( \frac{\mu}{\mu + (1 - \mu)(1 - C_1(x, z_\omega))} \right) dx \\
+ 2b_2 (1 - \mu)(z_\omega - C(z_\omega, z_\omega))V_0(0) \right],
\]

where \( V_0 \) is the single period value function, and \( z^*_{\omega 0} \) is the static model cutoff. The first term represents the value of current period’s persuasion. The second and third expressions correspond to the expected payoff when the negative and positive messages turn out to be correct, respectively. The Sender gets the last term when her message does not conform to the realized event.

A numerical simulation shows that the value function under this alternative formulation is qualitatively the same as the one in the benchmark case, so all of the
established results are expected to hold (see Figure 1-6).

1.4.2 Full commitment

In this subsection, I consider how the Sender's policy changes when she must fully commit to an experiment once and for all at the beginning of the game. That is, in all subsequent periods she must adhere to the experiment she chose in the first period. This setting is applicable when 1) experiments are conducted fairly often and therefore are required to follow a standardized procedure, or 2) it is costly to set up a different experiment every period.

Let us consider the two-period model, i.e. $T = 2$. Define

$$\gamma(z, \mu) = \mathbb{1}_{\{z \geq z^*(\mu)\}} (1 - z),$$

where $z^*$ is the persuasion cutoff in the static model. Similar to the benchmark model,
the sender optimally follows a threshold strategy. The sender’s problem is

\[ V^{FC}(\mu) = \sup_{z \in [0,1]} \left[ b_1 \gamma(z, \mu) + b_2 (\lambda_l(z) \gamma(z, \mu_l(z)) + \lambda_h(z) \gamma(z, \mu_h(z)) + (1 - \lambda_l(z) - \lambda_h(z)) \gamma(z, 0) \right] , \]

where \( \lambda_l, \lambda_h, \mu_l, \mu_h \) are defined in (1.3).

Figure 1-7: The Sender’s value function and optimal policy under the benchmark and full commitment case. \( C \) is Gaussian with correlation \( \rho = 0.25, \omega_c = 0.62, y = 0.47, b_1 = 0.03, b_2 = 1. \)

Clearly, if the Sender is required to adhere to a fixed policy, the “dynamic” policy cannot be adopted; some information needs to be disclosed, because otherwise the Sender cannot persuade the Receiver in any period. Figure 1-7 compares the Sender’s value function and optimal policy under the two formulations. Under full commitment, very low types select the same strategy as the worst type, because they are likely to be the worst type in the next period, in which case the message leading to \( a = 1 \) needs to be sufficiently positive. For higher types, the myopic policy is chosen. In general, more information is transmitted under full commitment than under
the benchmark case, and the discrepancy is bigger for lower $b_1$, because the dynamic policy is more adopted in the benchmark.

1.5 Conclusion

I have developed a tractable model of dynamic Bayesian persuasion in which the receiver ex-post learns the true state (perfectly or imperfectly) and updates his belief on the sender’s competence level. If the sender designs a highly informative experiment, the receiver can learn the sender’s type accurately, so the sender’s choice not only affects the receiver’s action but also the way the receiver perceives the sender in the future. Under mild conditions I showed that the sender prefers to minimize the variation of the receiver’s posterior belief about the sender’s type (binary information partition), and adopts an extreme strategy of not disclosing any information if the value of future persuasion is high enough. The equilibrium information structure is unique, so I was able to exactly quantify the amount of information that the optimal experiment transmits, a feature that standard models do not deliver. Another new finding is the non-monotonic relationship between the quality of experiment and Sender’s ability. Both extremely competent and extremely incompetent types are more likely to run informative experiments than intermediate ones since the Receiver’s ex-post observation is less consequential to the posterior of the Sender’s type. This observation sheds a new light on interpreting both quality and quantity of information that is communicated between two parties. The framework of this paper allowed a general state-signal dependence structure and distinct forms of learning in order to accommodate various real-world settings.

This paper suggests potentially interesting venues for further research. In particular, understanding the dynamic policy of a sender with state-dependent persuasion value can be important. For example, a central bank may weigh more on its ability to prevent bank failures in bad times (countercyclical persuasion value). Or, a company may derive higher utility from persuading a regulator in good times when potential NPV is higher (procyclical persuasion value). Another possible extension of this pa-
per is to allow the receiver’s observation to be action-dependent. For instance, the receiver may learn the realization only on good outcomes or only on bad outcomes. The sender’s optimal strategy can be very different under the two schemes.
Bibliography


1.6 Appendix

Proof of Lemma 1.1:

It suffices to show the result for \( \omega \) variables. Define

\[
G(x, y) := \Pr[\omega \geq \omega_c \mid x \leq \omega' \leq y] = 1 - \frac{C(\omega_c, y) - C(\omega_c, x)}{y - x}.
\]

Also, write

\[
G(x) := \Pr[\omega \geq \omega_c \mid \omega' \geq x] = 1 - \frac{\omega_c - C(\omega_c, x)}{1 - x}.
\]

Note that

\[
\frac{\partial}{\partial x} G(x, y) = \frac{C_2(\omega_c, x)(y - x) - (C(\omega_c, y) - C(\omega_c, x))}{(y - x)^2} > 0
\]

by concavity of \( C \) in the second argument (A3). Similarly,

\[
\frac{\partial}{\partial y} G(x, y) = \frac{-C_2(\omega_c, y)(y - x) + (C(\omega_c, y) - C(\omega_c, x))}{(y - x)^2} > 0
\]

Proof of Proposition 1.1: Let us work with the \( \omega \)-space. With abuse of notation, I use \( \mathcal{P} \) instead of \( \mathcal{P}^\omega \) when no confusion arises.

Let \( \mathcal{P} = \mathcal{P}^{a=0} \cup \mathcal{P}^{a=1} \) be an optimal information structure (partition of \([0, 1]\)), where \( \mathcal{P}^{a=i} \) denotes the set of partition sets in \( \mathcal{P} \) that lead to action \( a = i \). By the Revelation principle in information design (see Kamenica and Gentzkow (2011)), there exists an equilibrium experiment such that each \( \mathcal{P}^{a=i} \) is a singleton. Let \( P^i \) be the element. I claim that \( P^1 \) must be an upper-half interval almost surely.

Clearly, it must be that \( \sup P^1 = 1 \), because otherwise adding the region \([\sup P^1, 1]\) to \( P^1 \) makes the sender strictly better off. Now define

\[
\bar{\omega}^1 = \inf_{\omega} \{ m([\omega, 1] \cap P^1) = 1 - \omega \},
\]

where \( m \) is the Lebesgue-measure. If \( m(P^1 \setminus [\omega^1, 1]) = \alpha > 0 \), by Lemma 1,

\[
\Pr[\omega \geq \omega_c \mid \hat{\omega} \in [1 - \bar{\omega}^1 - \alpha, 1]] > \Pr[\omega \geq \omega_c \mid \hat{\omega} \in P^1] \geq y.
\]
Hence, there exists \( \epsilon > 0 \) such that (note that \( \alpha < 1 - \bar{\omega}^1 \) by \( A1 \))

\[
Pr \left[ \omega \geq \omega_c \left| \hat{\omega} \in [1 - \bar{\omega}^1 - \alpha - \epsilon, 1] \right. \right] \geq y.
\]

The sender is strictly better off under the binary partition with threshold \( 1 - \bar{\omega}^1 - \alpha - \epsilon \), so a contradiction. Then, it must be that \( \alpha = 0 \), establishing Proposition 1.

**Proof of Theorem 1.1:** Let us work with the \( \omega \)-space. \( z_\omega \) denotes \( F(z_0) \).

Note that

\[
Pr \left[ \omega \geq \omega_c \left| \hat{\omega} \geq x \right. \right] = \begin{cases} 
\mu + (1 - \mu) \cdot G(x) & x > \omega_c \\
\mu \cdot \frac{1 - \omega_c}{1 - x} + (1 - \mu) \cdot G(x) & x \leq \omega_c
\end{cases}
\]

If \( P = \{[0, x), [x, 1]\} \) with \( x \geq \omega_c \), \( [x, 1] \) is an action region by \( A4 \). Therefore, \( \omega^* \leq \omega_c \).

In equilibrium, \( Pr[\omega \geq \omega_c | \hat{\omega} \geq z^*_\omega] \) must be equal to \( y \), because otherwise the Sender can lower \( z^*_\omega \) and increase the probability of successful persuasion. In other words, the optimal \( z^*_\omega(\mu) \) is the solution to

\[
\mu \cdot \frac{1 - \omega_c}{1 - z} + (1 - \mu) \cdot G(z) = y. \tag{1.4}
\]

By Lemma 1, the left hand side is strictly increasing in \( z \), so there exists a unique \( z^*_\omega(\mu) \) that solves (1.4).

For algebraic simplicity, define

\[
\alpha_1(x) = \frac{1 - \omega_c}{1 - x} = Pr[\omega \geq \omega_c | \omega \geq x]
\]

\[
\alpha_2(x) = \frac{x - C(\omega_c, x)}{1 - x} = Pr[\omega \geq \omega_c | \omega \geq x] - Pr[\omega \geq \omega_c | \omega' \geq x].
\]

(1.4) can be rewritten as (drop the subscript \( \omega \) for simplicity)

\[
\alpha_1(z^*(\mu)) - (1 - \mu)\alpha_2(z^*(\mu)) = y.
\]

Taking the first and second derivatives, respectively, after rearranging terms

\[
z' \cdot (\alpha'_1 - (1 - \mu)\alpha'_2) = -\alpha_2 \tag{1.5}
\]

\[
z'' \cdot (\alpha'_1 - (1 - \mu)\alpha'_2) = -z' \cdot (\alpha''_1 - (1 - \mu)\alpha''_2) \cdot z' + 2\alpha'_2 \tag{1.6}
\]

50
After simplifying (1.5),

\[
\frac{dz^*}{d\mu} \left[ \frac{1 - \omega_c}{1 - z^*} - (1 - \mu) \left( \frac{1 - C(\omega_c, z^*)}{1 - z^*} - C'_2(\omega_c, z^*) \right) \right] > 0
\]

\[\text{(1.7)}\]

The inequality inside the bracket is positive because \( \frac{1 - C(\omega_c, z^*)}{1 - z^*} \geq 1 \geq C_2(\omega_c, z^*) \) and

\[
\frac{1 - \omega_c}{1 - z^*} - \left( \frac{1 - C(\omega_c, z^*)}{1 - z^*} - C_2(\omega_c, z^*) \right) = C_2(\omega_c, z^*) - \frac{\omega_c - C(\omega_c, z^*)}{1 - z^*} > 0,
\]

where \( A3 \) is used in the last inequality. Therefore, \( z^*_\omega \) is strictly decreasing in \( \mu \) and \( V(\mu) = 1 - z^*_\omega(\mu) \) is strictly increasing in \( \mu \), establishing the first part of the proposition.

After simplifying (1.6),

\[
\frac{d^2 z^*}{d\mu^2} \left[ \frac{1 - \omega_c}{1 - z^*} - (1 - \mu) \left( \frac{1 - C(\omega_c, z^*)}{1 - z^*} - C'_2(\omega_c, z^*) \right) \right] = - \frac{dz^*}{d\mu} \cdot \left[ 2(1 - C_2(\omega_c, z^*)) + (1 - \mu) \cdot \frac{dz^*}{d\mu} \cdot C_{22}(\omega_c, z^*) \right] > 0,
\]

where the inequality follows from \( C_2 \leq 1 \), and that \( \frac{dz^*}{d\mu}, C_{22}(\omega_c, z^*_\omega) < 0 \). Therefore,

\[\frac{d^2 z^*_\omega}{d\mu^2} > 0, \text{ and } V(\mu) = 1 - z^*_\omega(\mu) \text{ is strictly concave in } \mu. \]

**Proof of Lemma 1.2 and 1.3.** For algebraic tractability, \( z \) and \( z' \) represent \( z^*_\omega(\mu) \) and \( \frac{dz^*_\omega}{d\mu} \), respectively. It is implicitly assumed that \( C \) and its derivatives are evaluated at \( (\omega_c, z^*_\omega(\mu)) \).

Differentiating (1.8) with respect to \( \mu \),

\[
z'''(\alpha_1' - (1 - \mu)\alpha_2')
\]

\[= - 3z'' \left[ (\alpha'''_1 - (1 - \mu)\alpha''_2)z' + \alpha_2' \right] - (z')^2 \left[ (\alpha'''_1 - (1 - \mu)\alpha''_2)z' + 3\alpha_2'' \right]
\]

\[= - (z')^2 \left[ \frac{3}{\alpha_2} \left( [\alpha'''_1 - (1 - \mu)\alpha''_2]z' + 2\alpha_2' \right) \cdot \left( [\alpha''_1 - (1 - \mu)\alpha''_2]z' + \alpha_2' \right) \right]
\]

\[= A + \left[ \frac{[\alpha''_1 - (1 - \mu)\alpha''_2]z' + 3\alpha_2''}{\alpha_2} \right] = B \]

\[\text{(1.10)}\]
A, B can be further reduced to

\[
A = \frac{3}{\alpha_2(1 - z)^2} \left[ (1 - \mu)C_{22}z' + 2(1 - C_2) \right] \cdot \left[ (1 - \mu)C_{22}z' + (1 - C_2) - \alpha_2 \right]
\]

\[
B = \frac{1}{\alpha_2(1 - z)^2} \left[ 3C_{22}z' \left( C_2 - \frac{\omega_c - z}{1 - z} \right) + 6(1 - C_2)\alpha_2 + (1 - \mu)(1 - z)z' C_{22}\alpha \right]
\]

and

\[
A + B = \frac{3C_{22}z'}{\alpha_2(1 - z)^2} \left\{ \left( C_2 - \frac{\omega_c - z}{1 - z} \right) - (1 - \mu)\frac{z - C}{1 - z} + (1 - \mu)^2 C_{22}z' \right.
\]

\[
+ 3(1 - \mu)(1 - C_2) + \frac{2(1 - C_2)^2}{C_{22}z'} + \frac{(1 - \mu)(1 - z)C_{22}}{3C_{22}} \right\}
\]

\[
\geq \frac{3C_{22}z'}{\alpha_2(1 - z)^2} \left\{ \left( C_2 - \frac{\omega_c - C}{1 - z} \right) + (1 - \mu)^2 C_{22}z' + 3(1 - \mu)(1 - C_2) \right.
\]

\[
+ \frac{2(1 - C_2)^2}{C_{22}z'} + \frac{(1 - \mu)(1 - z)C_{22}}{3C_{22}} \right\}, \tag{1.11}
\]

where the inequality follows from \(C_{22}z' \geq 0\) and the fact that

\[
C_2 - \frac{\omega_c - z}{1 - z} - (1 - \mu)\frac{z - C}{1 - z} \geq C_2 - \frac{\omega_c - z}{1 - z} \geq 0.
\]

If \(C_{22}(\omega_c, z^*) \leq 0\), then every term in the curly bracket is positive, so \(A + B \geq 0\). Consequently, from (1.10) and 1.11, \(V'' \geq 0\), proving Lemma 1.2.

Now, assume that \(C\) is the Gaussian copula with correlation \(\rho > 0\), it can be shown that

\[
\text{sign}(C_{22}(\omega_c, z^*)) = -\text{sign} \left( \rho \Phi^{-1}(\omega_c) + (1 - 2\rho^2)\Phi^{-1}(z^*) \right).
\]

If \(z^*(\mu = 1) = 1 - \frac{1 - \omega_c}{y} \geq 1/2\), then for every \(\mu\), \(\Phi^{-1}(\omega_c) \geq \Phi^{-1}(z^*(\mu)) \geq 0\), so \(C_{22}(\omega_c, z^*) \leq 0\). Therefore \(V'' \geq 0\).

If \(\rho \geq \frac{1}{\sqrt{2}}\) and \(\omega_c \geq 1/2\),

\[
\rho \Phi^{-1}(\omega_c) + (1 - 2\rho^2)\Phi^{-1}(z^*) \geq (\rho + 1 - 2\rho^2)\Phi^{-1}(\omega_c) \geq 0,
\]

so \(V'' \geq 0\).

Now, let \(\rho \leq \frac{1}{\sqrt{2}}\), \(\omega_c \leq 1/2\) and \(1 - \frac{1 - \omega_c}{y} \geq \zeta(y, \rho)\). In order for \(V'' \geq 0\), it suffices
to show that
\[
\left( C_2 - \frac{\omega_c - C}{1 - z} \right) + 3(1 - \mu)(1 - C_2) + \frac{2(1 - C_2)^2}{C_{22} z'} \geq \frac{(1 - \mu)(1 - z)C_{222}}{3C_{22}}. \tag{1.12}
\]

Note that
\[
\frac{2(1 - C_2)^2}{C_{22} z'*'} = \frac{2(1 - C_2)^2}{z' - C(\omega_c, z')} \frac{\sqrt{1 - \rho^2} e^{\frac{w^2}{2} \frac{-\Phi^{-1}(z^*)^2}{2}}}{\rho} \times \left( \frac{1 - \omega_c}{1 - z'} - (1 - \mu) \left( \frac{1 - C(\omega_c, z')}{1 - z'} - C_2(\omega_c, z^*) \right) \right),
\]
where
\[
W = \frac{\Phi^{-1}(\omega_c) - \rho \Phi^{-1}(z^*)}{\sqrt{1 - \rho^2}}
\]
Note that
\[
\Phi(-W) = (1 - C_2) \geq \frac{z^* - C(\omega_c, z^*)}{z^*}
\]
so
\[
\frac{2(1 - C_2)^2}{C_{22} z'*'} \geq \frac{2\sqrt{1 - \rho^2}}{\rho} \frac{\Phi(-W) e^{\frac{w^2}{2} \frac{-\Phi^{-1}(z^*)^2}{2}}}{\Phi(\Phi^{-1}(z^*)) e^{\frac{-\Phi^{-1}(z^*)^2}{2}}} \left( \frac{1 - \omega_c}{1 - z^*} - (1 - \mu) \left( \frac{1 - C(\omega_c, z^*)}{1 - z^*} - C_2(\omega_c, z^*) \right) \right)
\]
\[
\geq \frac{2\sqrt{1 - \rho^2}}{\rho} \left( \frac{1 - \omega_c}{1 - z^*} - (1 - \mu) \left( \frac{1 - C(\omega_c, z^*)}{1 - z^*} - C_2(\omega_c, z^*) \right) \right),
\]
where the fact that \( \Phi(x) e^{\frac{x^2}{2}} \) is increasing and that \(-W \geq \Phi^{-1}(z^*)\) if \( \sqrt{1 - \rho^2} \geq \rho \), \( \omega_c \leq 1/2 \), are used.

Both sides of (1.12) are linear in \( \mu \) (ignoring dependence between \( \mu \) and \( z^* \)), and the inequality holds for \( \mu = 1 \), so it suffices to show when \( \mu = 0 \). This corresponds to proving
\[
\left( 1 + \frac{2\sqrt{1 - \rho^2}}{\rho} \right) \left( C_2 - \frac{\omega_c - C}{1 - z^*} \right) + 3(1 - C_2) \geq \frac{(1 - z^*)C_{222}}{3C_{22}} \tag{1.13}
\]
The RHS can be computed as
\[
(1 - z^*) \frac{\sqrt{2\pi}}{3} e^{\Phi^{-1}(z^*)^2/2} \left( -\frac{\rho}{1 - \rho^2} \Phi^{-1}(\omega_c) - \frac{1 - 2\rho^2}{1 - \rho^2} \Phi^{-1}(z^*) \right)
\leq (1 - z^*) \frac{\sqrt{2\pi}}{3} e^{\Phi^{-1}(z^*)^2/2} \left( -\frac{1 + 2\rho}{1 + \rho} \Phi^{-1}(z^*) \right).
\]

Using \( \frac{\sqrt{1-\rho^2}}{\rho} \geq 2 \), the LHS of (1.13) is at least
\[
3 \left( 1 - \frac{\omega_c - C}{1 - z^*} \right) \geq 3(1 - \omega_c),
\]
where the fact that \( C(\omega_c, z^*) \geq \omega_c z^* \) is used in the last inequality. Noting that
\[
\frac{1 - \omega_c}{1 - z^*} \geq \frac{1 - \omega_c}{1 - z_{\mu=1}^*} = y,
\]
the result of Lemma 1.3 is proven. ■

**Proof of Theorem 1.2:** Again let us work with the \( \omega \)-space. Consider any partition \( P = P_{a=0}^1 \cup P_{a=1}^1 \), where \( P_{a=i}^1 \) is the region in which the Receiver chooses \( a = i \). If there are multiple partition elements in \( P_{a=0}^1 \) choose two partition elements \( Q_k, Q_{k+1} \) and collapse to one partition \( \bar{Q} = Q_k \cup Q_{k+1} \). Define
\[
\lambda_j = Pr [\omega \in Q_j, \hat{\omega} \in Q_j] = \mu m(Q_k) + (1 - \mu) Pr [\omega \in Q_j, \omega' \in Q_j]
\]
\[
\mu_j = Pr [\epsilon \omega \in Q_j, \hat{\omega} \in Q_j]
\]
By the law of iterated expectation
\[
\lambda_k \mu_k + \lambda_{k+1} \mu_{k+1} = \bar{\lambda} \bar{\mu}.
\] (1.14)

The incremental gain of collapsing \( Q_k \) and \( Q_{k+1} \) is given by
\[
\left[ \bar{\lambda} V(\bar{\mu}) + (m(\bar{Q}) - \bar{\lambda}) V(0) \right]
- \left[ \lambda_k V(\mu_k) + (m(Q_k) - \lambda_k) V(0) \right]
- \left[ \lambda_{k+1} V(\mu_{k+1}) + (m(Q_{k+1}) - \lambda_{k+1}) V(0) \right]
= \bar{\lambda} V(\bar{\mu}) - \left[ \lambda_k V(\mu_k) + \lambda_{k+1} V(\mu_{k+1}) + (\bar{\lambda} - \lambda_k - \lambda_{k+1}) V(0) \right].
\] (1.15)
Note that

\[
\frac{\lambda - \lambda_k - \lambda_{k+1}}{1 - \mu} = Pr[\omega \in \bar{Q}, \omega' \in \bar{Q}] - Pr[\omega \in Q_k, \omega' \in Q_k] - Pr[\omega \in Q_{k+1}, \omega' \in Q_{k+1}] > 0.
\]

Hence, by strict concavity of \( V \) and Jensen’s inequality, it follows from (1.15) that collapsing two partition elements yields better payoff to the sender. In other words, under the optimal scheme, \( P_1^{a=0} \) must be a singleton.

The same logic applies to \( P_{a=1} \); collapsing any two partition elements of \( P_{a=1} \) yields higher continuation payoff to the sender. In addition, if the receiver were to act for each partition element, he would do so under the concatenated partition element as well. Therefore under the optimal partition rule \( P_1^{a=1} \) should contain no more than one element.

**Proof of Lemma 1.4.** For notational simplicity, I drop subscripts.

Differentiating (1.3) twice with respect to \( z \) and rearranging terms, we have

\[
\frac{\partial^2}{\partial z^2} (CV) = \frac{\partial^2 \lambda_i}{\partial z^2} \cdot [V(\mu_i) - V(0) - \mu_i V'(\mu_i)] + \frac{\partial^2 \lambda_h}{\partial z^2} \cdot [V(\mu_h) - V(0) - \mu_h V'(\mu_h)]
+ \lambda_i \left( \frac{\partial \mu_i}{\partial z} \right)^2 \cdot V''(\mu_i) + \lambda_h \left( \frac{\partial \mu_h}{\partial z} \right)^2 \cdot V''(\mu_h).
\]

Since \( V'' \geq 0 \), the function \( V(x) - V(0) - xV'(x) + \frac{x^2}{2} V''(x) \) is increasing and is 0 for \( x = 0 \),

\[
V(x) - V(0) - xV'(x) + \frac{x^2}{2} V''(x) \geq 0.
\]

So

\[
\frac{\partial^2}{\partial z^2} (CV) \geq (V(\mu_i) - V(0) - \mu_i V'(\mu_i)) \cdot \left( \frac{\partial^2 \lambda_i}{\partial z^2} - 2 \lambda_i \left( \frac{\partial \mu_i}{\partial z} \right)^2 / \mu_i^2 \right)
+ (V(\mu_h) - V(0) - \mu_h V'(\mu_h)) \cdot \left( \frac{\partial^2 \lambda_h}{\partial z^2} - 2 \lambda_h \left( \frac{\partial \mu_h}{\partial z} \right)^2 / \mu_h^2 \right).
\]

Using \( \frac{\partial \mu_i}{\partial z} / \mu_i = \frac{1 - \mu}{\lambda_i} \cdot \left( \frac{\delta(z)}{z} - \delta'(z) \right) \) and \( \frac{\partial \mu_h}{\partial z} / \mu_h = \frac{1 - \mu}{\lambda_h} \cdot \left( \frac{1 - \delta(z)}{1 - z} - \delta'(z) \right) \).
we see that for \( \frac{\partial^2}{\partial z^2} (CV) > 0 \) it suffices to show that
\[
\frac{\delta''(z) \lambda_l(\mu, z)}{2} \geq (1 - \mu) \left( \frac{\delta'(z) - \delta(z)}{z} \right)^2
\]
\[
\frac{\delta''(z) \lambda_h(\mu, z)}{2} \geq (1 - \mu) \left( \frac{1 - \delta(z)}{1 - z} - \delta'(z) \right)^2.
\]
Both \( \lambda_l \) and \( \lambda_h \) are linear in \( \mu \), so the above inequalities hold for all \( \mu \) if they hold for \( \mu = 0 \) and \( \mu = 1 \). The latter is trivial, and for \( \mu = 0 \) the inequalities follow from the given assumptions.

**Proof of Theorem 1.3.** Recall that
\[
J_t(\mu_t, z_{wt}) = b_t \{ z_{wt} \geq z^*_{w0}(\mu_t) \} \cdot (1 - z_{wt}) + CV_{t+1}(\mu_t, z_{wt})
\]
If a policy \( z_{wt} < z^*_{w0}(\mu_t) \) is adopted, clearly \( z_{wt} = 0 \) is optimal, because \( CV_{t+1}(\mu_t, \cdot) \) has its global maximum at 0 and 1. \( z_{wt} = 1 \) yields the same functional value as \( z_{wt} = 0 \), so the optimal policy can be assumed to be in \( I = [z^*_{w0}(\mu_t), 1] \), i.e.,
\[
V_t(\mu_t) = \sup_{z_{wt} \in I} J_t(\mu_t, z_{wt}) = \sup_{z_{wt} \in I} J_t(\mu_t, z_{wt})
\]
\( J_t(\mu_t, \cdot) \) is convex in \( I \), so the global maximum is attained at either of the two endpoints: one corresponds to the myopic policy and the other to the coarsest policy (no information). The result of Theorem 1.3 then follows.

**Proof of Proposition 1.2.**

Let \( V \) be the strictly concave single period value function.

Similar to the proof of Proposition 3, let us work with \( \omega \)-space.

Suppose there are two partition elements \( Q_k, Q_{k+1} \) that belong to \( \mathcal{P}^{a=0} \). Let us collapse them to one partition element \( \bar{Q} = Q_k \cup Q_{k+1} \). The incremental gain of collapsing
is
\[
\int_Q \tilde{\lambda}(x) \left[ V \left( \frac{\mu}{\lambda(x)} \right) - V(0) \right] dx \\
- \int_{Q_k} \lambda_k(x) \left[ V \left( \frac{\mu}{\lambda_k(x)} \right) - V(0) \right] dx \\
- \int_{Q_{k+1}} \lambda_{k+1}(x) \left[ V \left( \frac{\mu}{\lambda_{k+1}(x)} \right) - V(0) \right] dx \\
= \int_Q \left[ \tilde{\lambda}(x) V \left( \frac{\mu}{\lambda(x)} \right) - \lambda_k(x) V \left( \frac{\mu}{\lambda_k(x)} \right) - (\tilde{\lambda}(x) - \lambda_k(x)) V(0) \right] dx \\
+ \int_{Q_{k+1}} \left[ \tilde{\lambda}(x) V \left( \frac{\mu}{\lambda(x)} \right) - \lambda_{k+1}(x) V \left( \frac{\mu}{\lambda_{k+1}(x)} \right) - (\tilde{\lambda}(x) - \lambda_{k+1}(x)) V(0) \right] dx
\]
where
\[
\tilde{\lambda}(x) = \mu + (1 - \mu) Pr \{ \omega' \in \tilde{Q} \mid \omega = x \} \\
\lambda_k(x) = \mu + (1 - \mu) Pr \{ \omega' \in Q_k \mid \omega = x \} \\
\lambda_{k+1}(x) = \mu + (1 - \mu) Pr \{ \omega' \in Q_{k+1} \mid \omega = x \}.
\]
From the fact that \(\tilde{\lambda}(x) > \lambda_k(x), \lambda_{k+1}(x)\) and Jensen’s inequality, the incremental gain is strictly positive. Hence, \(\mathcal{P}^{a=0}\) must be singleton. A similar logic applies to \(\mathcal{P}^{a=1}\).
Chapter 2

Value-at-Risk Constraint and Price Informativeness

2.1 Introduction

Many commercial banks and other financial institutions are subject to regulations that restrict their risk exposure. In particular, various regulatory agencies including the Federal Deposit Insurance Corporation (FDIC) require banks to report and control their Value-at-Risk (VaR) measure, which quantifies the threshold loss value that the probability of realizing this loss is equal to a pre-specified level such as 1% or 5%. For example, the Basel Accord allows banks to use VaR measures, if necessary, when calculating the riskiness of their position. Their “eligible capital” must exceed certain proportion of their “risk weighted asset holdings” (Brunnermeier and Pedersen (2009)) which can be partially based on the VaR measure.

There have been extensive studies that investigate how VaR constraint affects equilibrium through changes in agents’ optimal asset holdings. For instance, Basak and Shapiro (2001) examine the impact of a static VaR constraint on an agent’s portfolio choice problem, and show that the constraint may make the agent take excessive risk in the worst state of the economy, because it is expensive to insure in those states and VaR is only pertinent to probability of big losses. Cuoco et al. (2008) demonstrate that under a more reasonable dynamic VaR constraint, the agent holds
less risky position than she would absent the constraint. In a general equilibrium framework, Danielsson et al. (2004) illustrate that when regulated banks form VaR measures from “backward-looking belief revision rules,” price volatility increases while price level falls.

However, little attention is given to the informational role of VaR constraint. Banks that are subject to this constraint usually possess superior information than others. In general, equilibrium price reflects private information about asset fundamentals, so agents with less or no information learn from the price. Since such uninformed agents do not know whether the constraint is binding for banks, the VaR rule may deter them from making precise inference. Of course, their behavior has a huge impact on the equilibrium, since market has to clear in every state of the nature. It is not clear at all to imagine how the equilibrium would look like with the constraint.

In this paper, we study how VaR driven risk management affects the equilibrium price in the presence of information asymmetry. In particular, using a Rational Expectations Equilibrium (REE) model, we explore how price informativeness depends on the price level and characterize the equilibrium price and return distributions when some informed investors use a VaR rule to control the probability of extreme tail loss events. We find that 1) price contains less information when price is around its unconditional mean, 2) price volatility may decrease or increase depending on the precision of private information, 3) risk is transferred to uninformed investors and they expect higher return at the expense of increased return volatility, and 4) less investors choose to stay informed when VaR investors are present. In particular, when price is reasonably high or low, price is much more informative than that of a constraint-free economy and can almost fully reveal the signal at some level.

The inference problem by the uninformed induces a non-monotonic relationship between price informativeness and price level. Prices very close to the mean level do not contain much information, since uninformed agents cannot tell whether informed traders are buying or selling. In turn, the uninformed face much uncertainty about asset fundamentals. When price is moderately away from its mean, the uninformed estimate that with high probability the constraint is binding in a specific region, and
possible signal variation is very small. For larger price deviation, the uninformed cannot project whether the constraint is binding or not, because the VaR constraint is slack when price is very far from the asset fundamental. Finally, if price is at an extreme level (high or low), it is very likely that the VaR investors are submitting their unconstrained demand. Therefore, price becomes as informative as in a constraint-free economy.

Even though the above result suggests that the price informativeness is not monotonic in the deviation of price from its mean, the probability of realizing a very high or low price is extremely small in our model. Simulating with various parameter values, we find that the such probability is less than the order of $10^{-8}$. In this respect, for a reasonable price range, price contains more information when it deviates more from its mean. Interestingly, at such moderate levels, price is more informative than it would be in an economy without constraints.

The price volatility depends on the conflicting forces of the VaR constraint: allocation and information effects (see Bai et al. (2006)). In general, in an economy with information asymmetry, a constraint has two roles in determining equilibrium. First, it restricts how assets are allocated. When informed agents face a constraint and have to change their asset holdings, such change must be met by the demand by the uninformed and noise traders, since the market has to clear. This mechanism alters how price reacts to noise supply. Second, the constraint affects the inference problem of the uninformed. Not only are they unable to differentiate signal from noise, but they also do not know whether the constrained investors are hitting the constraint. In Section 5, we provide conditions under which effect is stronger than the other. Qualitatively, when private signal is very precise, the informational effect dominates, because the constraint is meaningful in only limited price regions.

Our model is not the first to look at a constrained economy with asymmetric information. For example, in a REE framework, Yuan (2005) studies the equilibrium dynamics when some set of informed investors are borrowing (leverage) constrained. The constraint is modeled by imposing an upper bound of long position $x$ by an increasing linear function of price. In a similar context, Bai et al. (2006) study the
characteristics of equilibrium when some informed investors cannot short more than some exogenous amount. Yuan (2006) incorporates both borrowing and short-sale constraints. In such models, constraints are generally associated with “backward-bending” demand curve of the uninformed and with price crash; at a critical price level, a small increase in price lowers the variance of the uninformed inference problem, and the uninformed may demand increasingly more at a higher price. While our model produces a similar upward sloping uninformed demand curve in some regions, the likelihood of such an event is essentially zero.

In all of the above models, the amount of information reflected in price is globally lower than that of the constraint-free economy. Our model, however, produces wide regions in which price is more informative than it would have been without the constraint. Moreover, in leverage or short sale constrained economy, price volatility generally increases, whereas in our economy it depends on the signal precision.

A key feature of our model that contributes to such contrasting equilibrium dynamics from the existing models is that, unlike others, VaR constraint does not bind in far-out regions. In other words, when price deviates much from the conditional mean of the payoff given private signal, VaR is vacuous, because the probability of realizing negative return is bounded by above, regardless of the magnitude of trade. Therefore, only when price is around the fundamental value does the VaR constraint matter.

To fix ideas, consider an informed but VaR-constrained investor who believes a stock is worth $10 on average and is worth at least $8 with 95% probability. Further assume that the investor wants to have 95% confidence in not losing more than a specified amount, say 20% of her wealth. Being risk-averse, she would hold a finite position at any price even without the constraint. Now, assume that the current stock price is $8. Since the stock is thought to be undervalued, she would buy the stock. Regardless of the position size, the probability of incurring loss is within 5%, so the VaR constraint has no effect on her portfolio choice. If, instead, the stock price is $9, the constraint matters; although the investor still goes long the asset, her position must be limited in order to contain the probability of extreme loss. At even higher
price, say $9.9, the VaR rule may become irrelevant again, because she dislikes to place a big bet even absent any constraint, because her expected profit is small.

As the above example illustrates, VaR-based risk management creates a subtle relationship between security price and optimal portfolio holding even in a simple portfolio choice problem. When fully taken to a model of information asymmetry, its effects on the equilibrium outcome are far from trivial. The objective of this paper is to understand its consequences. Section 2.2 fully describes the model, and derive informed agents' optimal demand. In Section 2.3, we provide analytical equations that the uninformed demand and equilibrium price must satisfy. The analytical expression of the GS equilibrium price is also provided. Section 2.4 studies the uninformed inference problem and price informativeness. A comparison between the VaR economy and other (leverage and short-sale) constrained economies is also made. In Section 2.5, we discuss the characteristics of VaR equilibrium and how it deviates from the GS equilibrium and the price/return distribution. Section 2.6 explores endogenous choice of information acquisition and Section 2.7 concludes.

2.2 Model

The model is similar to that of Yuan (2005) who extends Grossman and Stiglitz (1980) to a model with borrowing-constrained investors. There is one risky asset and one risk-free asset. The risky asset pays a random payoff \( d \) at \( t = 1 \), and the risk-free asset pays \( 1 \) at \( t = 1 \). The risk-free asset is the numeraire. In other words, the risk free rate is normalized to 0.\(^1\) There are two periods \( t = 0 \) and \( t = 1 \) and four types of agents; unconstrained informed, VaR-informed, uninformed, and noisy traders. Therefore, the economy is identical to Grossman and Stiglitz (1980) except that some set of informed traders follow a VaR rule. In this paper, VaR-informed agents can be viewed as banks that are subject to regulations. Hedge funds can be loosely labeled as unconstrained informed agents who maximize expected utility.

\(^1\)The model can also allow a positive risk-free rate \( r \), and the results remain the same.
without any constraint.² The uninformed traders are retail investors, although they can be institutional investors with informational disadvantage than others. Finally, noise traders buy or sell the asset due to exogenous liquidity reasons.³

All agents trade the risky asset at $t = 0$. The random noisy traders’ supply is denoted by $\tilde{m}$.

### 2.2.1 Asset payoff, information structure, and preference

Informed traders observe a common signal $\bar{s}$ that conveys information of the asset fundamental:

$$\bar{s} = \tilde{d} + \tilde{\epsilon}_s,$$

where $\tilde{d}$, $\tilde{\epsilon}_s$, and $\tilde{m}$ are jointly independent normal variables with mean and variance $(0, \tau_{d}^{-1})$, $(0, \tau_{\epsilon}^{-1})$, and $(\tilde{m}, \tau_{m}^{-1})$, respectively. Every informed trader observes the same signal $\bar{s}$, so the common signal is a sufficient statistics of the asset fundamental.

Assume that both informed and uninformed agents have mean-variance preference:⁴ for $j \in \{U, I\}$

$$E[U_j(\tilde{W}^j | \mathcal{F}_j)] = E[\tilde{W}^j | \mathcal{F}_j] - \frac{1}{2\rho_j} V[\tilde{W}^j | \mathcal{F}_j],$$

where $\mathcal{F}_j$ is agent $j$’s information set, $\rho_j$ is agent $j$’s risk tolerance, and $\tilde{W}^j$ is agent $j$’s random wealth at $t = 1$. Note that for the informed $\mathcal{F} = \{\bar{s}, \bar{P}\}$ and for the uninformed $\mathcal{F} = \{\bar{P}\}$.

Unconstrained-informed investors (hedge funds) choose to maximize their expected utility conditional on their private signal and given price. That is, for given $(s, P)$,

---

²In reality, these traders do face some constraints such as short-sale constraint. We discuss the implication of short-sale constraint in a later section.

³When there are no noise traders, price becomes fully revealing, and in equilibrium all agents submit a price inelastic demand, the Grossman-Stiglitz paradox. See also Hellwig (1980) and Verrecchia (1982).

⁴In a noisy REE static model, it is conventional to assume that investors have CARA utility preference. When the payoff and noise terms are jointly normal, CARA paramounts to mean-variance specification. However, in our model price is not globally linear in $(s, m)$, so the uninformed inference problem has to do with distribution of doubly truncated normal variables. See Yuan (2005) for example.
they solve

$$\max_x \mathbb{E}[\tilde{W}_1|s, P] - \frac{1}{2\rho_I} \mathbb{V}[\tilde{W}_1|s, P]$$

(2.1)

where

$$\tilde{W}_1 = W_0 + x(\tilde{d} - P).$$

VaR-informed investors (banks) do the same but subject to a VaR constraint. That is, they solve

$$\max_x \mathbb{E}[\tilde{W}_1|s, P] - \frac{1}{2\rho_I} \mathbb{V}[\tilde{W}_1|s, P]$$

(2.2)

subject to

$$\tilde{W}_1 = W_0 + x(\tilde{d} - P)$$

$$Pr[\Delta \tilde{W}_1 \leq -\alpha|s, P] \leq q,$$

(2.3)

where $\alpha, q > 0.5$

In our model, VaR investors are assumed to follow the exogenously given constraint (2.3). That is, it is assumed that the banks control the true tail risk in the sense that the VaR is calculated with regard to their information set, not to a regulator’s. There are two interpretations. First, even absent any regulation, most banks voluntarily attempt to contain tail risk. Since its formal introduction in J.P. Morgan’s RiskMetrics (1995), VaR has been one of the most popular risk management tools for banks and institutional investors. Subsequently, advanced implementation methods such as those in Barone-Adesi et al. (1999) have been proposed for practical use. Second, while a regulator may not know the banks’ signal contemporaneously, banks may face punishment from the regulator ex-post if tail events occur more often than what VaR measure prescribes. Basel III, for example, stipulates that banks are subject to assessments based on “a periodic comparison of the bank’s daily risk measures (expected shortfall or value at risk) with the subsequent daily profit or loss” (BCBS (2016)). It is, therefore, in banks’ best interest to control the “true” VaR.

---

5 We are normalizing $W_0 = 1$, so $\Delta \tilde{W}_1$ is the investors’ rate of return.
A few papers further explore this idea. Cuoco and Liu (2006) show that, when the capital charge of a bank depends on its previous self-reported VaR measure in the form of multiples and the bank is penalized when realized returns involve more frequent extreme losses than the VaR prediction, VaR regulation induces the bank to truthfully report its risk exposure. Berkowitz and O’Brien (2002) and Péignon et al. (2008) provide empirical evidence that banks overstate their VaR estimates rather than understate, using the dataset of large U.S. and Canadian commercial banks, respectively.

Adrian and Shin (2014) endogenously derive the constraint in a model with moral hazard. They show that when a bank (agent) funds its fund through collateralized borrowing from creditors (principal) but can decide which project to invest in, under specific distributions supported in extreme value theory, the optimal contract entails a VaR rule, which keeps the probability of default at a constant level. In this risk-shifting moral hazard problem, the VaR probability is determined endogenously as a solution to optimal contracting.

Uninformed agents (retail investors) do not receive any signal, but infer the signal from the time-0 price. That is, they do not observe anything but the price at \( t = 0 \). Given price \( P \), they solve

\[
\max_x \mathbb{E}[\hat{W}_1|P] - \frac{1}{2\rho_U} \mathbb{V}[\hat{W}_1|P]
\]

(2.4)

subject to

\[
\hat{W}_1 = W_0 + x(d - P).
\]

Assume that there are \( w_{IF}, w_{IV} \) and \( w_u \) mass of unconstrained informed (Informed and free of constraints), VaR-informed (Informed and VaR constrained) and uninformed investors (Uninformed), respectively.
2.2.2 Optimal demand of the informed

Unconstrained-informed demand

Unconstrained informed agents choose the quantity that maximizes their expected payoff (2.1) given \((s, P)\). Since \(s\) is a sufficient statistics of the asset payoff, it is easy to see that the optimal demand is given by

\[
x_{IF}(s, P) = \rho I \cdot \frac{\mathbb{E}[d|s] - P}{\mathbb{V}[d|s]} = \rho I \left( \tau_e s - (\tau_e + \tau_d)P \right).
\]

(2.5)

VaR-informed demand

Since agents’ expected utility is quadratic and concave in risky asset demand \(x\), the VaR investors’ optimal demand is either equal to the unconstrained demand or is at the boundary. The demand is depicted in Figure 2-1.

(2.3) can be rewritten as

\[
Pr[x(d - P) < -\alpha|s, P] \leq q
\]

A mean-variance informed investor goes long (shorts) the asset when the price is lower (higher) than the conditional mean of the payoff \(\mathbb{E}[d|s] = s\tau_e / (\tau_e + \tau_d)\). First, consider the case \(P < \mathbb{E}[d|s]\) in which \(x^* > 0\). Using the fact that \(\mathbb{V}[d|s] = 1 / (\tau_e + \tau_d)\), the constraint can be written as:

\[
\frac{\alpha}{x} \geq P - \frac{\tau_e}{\tau_e + \tau_d} \cdot s - \frac{\Phi^{-1}(q)}{\sqrt{\tau_e + \tau_d}},
\]

where \(\Phi(\cdot)\) is the standard normal cumulative distribution function. When the price is very low (\(RHS < 0\)), the constraint is not binding as the investor will optimally choose \(x > 0\). If the \(RHS > 0\), then the constraint imposes an upper bound on the long position:

\[
x \leq \frac{\alpha}{P - \frac{\tau_e}{\tau_e + \tau_d} s - \frac{\Phi^{-1}(q)}{\sqrt{\tau_e + \tau_d}}}
\]

\(\Phi^{-1}(q)\) will be a small number in our calibration (1%-5%), so \(\Phi^{-1}(q)\) is a negative number.
Figure 2-1: VaR investor optimal demand. The unconstrained demand is the optimal demand of the informed traders without the VaR rule. ‘VaR when go long’ (short) illustrates the upper (lower) bound of the trading position. The thick line green line is the VaR investor’s optimal demand when VaR constraint is approximated by linear lines.

Similarly, when price is high (\( P > \mathbb{E}[d|s] \)) but not very high (\( P - \frac{\tau_e}{\tau_e + \tau_d} s + \frac{\Phi^{-1}(q)}{\sqrt{\tau_e + \tau_d}} < 0 \)), the VaR constraint restricts short position:

\[
x \geq \frac{\alpha}{P - \frac{\tau_e}{\tau_e + \tau_d} s + \frac{\Phi^{-1}(q)}{\sqrt{\tau_e + \tau_d}}}
\]

Intuitively, the constraint is most severe around the conditional mean of the payoff, and becomes loose as price moves away from the mean. In particular, when price is very far from the mean in either direction, the VaR constraint is always vacuous. This is because if price is very low (high) compared to the average payoff, investors would want to buy (short) assets, and regardless of the size of their position, the probability of realizing negative return is bounded above by a small number. For extreme prices,
this number is smaller than \( q \), so the VaR constraint is always satisfied.

Also note that when price is very close to the mean, although the constraint is tight, a mean-variance investor optimally chooses a small quantity (long or short) even without the constraint. This implies that the VaR constraint does not bind in a neighborhood of the conditional expectation of the payoff. Therefore, the only binding region is where price is a little distant from the expectation.

Figure 2-1 illustrates this point. When the signal is fixed at the mean level 0, informed investors go long (short) when price is lower (higher) than their expected asset payoff. When price is very close to 0, the VaR rule allows investors to go long or short some amount, but being risk averse, they choose not to buy or sell up to that amount. When price moves a little away from 0, the VaR rule limits their position. However, when price is very far from 0, the VaR rule becomes slack again.

In this respect, the VaR constraint has fundamentally different implications from leverage or short-sale constraints. Under the latter constraints, binding regions are always the ones in which price is far away from the signal in either direction. The fact that the VaR constraint only binds in intermediate regions is crucial in analyzing price informativeness in later sections. The observation made so far remains true under any dependence structure of \( \tilde{s} \) and \( \tilde{d} \), which is formally stated as follows.

**Proposition 2.1** Suppose \( \tilde{s} \) is any signal informative of \( \tilde{d} \) (\( \tilde{s} \) and \( \tilde{d} \) are not necessarily Gaussian). Then, for a fixed signal realization \( s \), there exist \( P(s) < \tilde{P}(s) \) such that the VaR-constrained informed demand \( x_{IV}(s, P) \) equals the unconstrained informed demand \( x_{IF}(s, P) \) for both \( P < \tilde{P} \) and \( P \geq \tilde{P} \).

Hereafter, the joint Gaussian structure is maintained for tractability.

If investors are too risk averse, in an unconstrained world they always demand less in absolute terms than what the VaR rule allows. Since we are interested in the effect of the VaR rule, assume that this is not the case.\(^7\) Then, there will be intermediate price regions in which VaR binds. According to the constraint (2.3), in

\[ \Phi^{-1}(q) > \frac{4\alpha}{\rho_I}. \]

\(^7\) This means that in Figure 2-1, the VaR curve and unconstrained demand line intersect. In Appendix A, it is shown that this is mathematically equivalent to \( [\Phi^{-1}(q)]^2 > \frac{4\alpha}{\rho_I} \).
these regions the demand is a hyperbolic function of price. In order to get closed expressions for conditional expectation and variance for the uninformed, we instead approximate the demand of these regions by piecewise linear functions, as in Figure 2-1. In particular, we use tangent lines at two intersection points of the VaR bound curve and the unconstrained demand.\footnote{Under various piecewise linear approximation methods, the results of this paper remain qualitatively the same.} Formally, throughout this paper, the VaR constraint is modeled by

\[
x^+(s, P) \leq \max \left[ \beta_h (\mu_{d/s} - P - u_h) + g_h, \beta_l (\mu_{d/s} - P - u_l) + g_l \right]
\]

\[
x^-(s, P) \leq \min \left[ \beta_h (E[d/s] - P + u_h) - g_h, \beta_l (E[d/s] - P + u_l) - g_l \right],
\]

where \(\beta_h, u_h, g_h, \beta_l, u_l, g_l\) are defined in Appendix, and \(x^+ = \max(x, 0), x^- = -\min(x, 0)\). Note that \(\mu_{d/s} := E[d/s] = \frac{\tau_e}{\tau_e + \tau_d}s\). The above remark is summarized in the following lemma.

**Lemma 2.1** The VaR investors' optimal demand that solves (2.2) under (2.6) is given by

\[
x^*_F(s, P) = \begin{cases} 
  x^*_F(s, P) & P - \mu_{d/s} \in (-\infty, -u_h) \cup (-u_l, u_l) \cup (u_h, \infty) \\
  \beta_h (\mu_{d/s} - P - u_h) + g_h & P - \mu_{d/s} \in [-u_h, -u_m] \\
  \beta_l (\mu_{d/s} - P - u_l) + g_l & P - \mu_{d/s} \in (-u_m, -u_l] \\
  \beta_l (\mu_{d/s} - P + u_l) - g_l & P - \mu_{d/s} \in [u_l, u_m) \\
  \beta_h (\mu_{d/s} - P + u_h) - g_h & P - \mu_{d/s} \in [u_m, u_h], 
\end{cases}
\]

where \(x^*_F(s, P) = \rho_l (\mu_{d/s} - P)\) is the unconstrained optimal demand. In particular, the demand is piecewise linear in \((s, P)\).
2.2.3 Equilibrium Concept

The equilibrium concept in this paper is Rational Expectations Equilibrium (REE). The equilibrium price $P(s, m)$ is a measurable function in $(s, m)$ under which 1) agents maximize their expected utility conditional on their information set, and 2) the market clears for every $(s, m)$:

$$w_{IV} x_{IV}^*(s, P) + w_{IF} x_{IF}^*(s, P) + w_{U} x_{U}^*(P) = m,$$

where $x_{IF}^*$ and $x_{IV}^*$ are given in (2.5) and (2.7), and $x_{U}^*$ is the solution to (2.4) under an equilibrium price. To make a meaningful comparison between economies with and without VaR investors, throughout this paper we fix the density of the informed traders $w_{IF} + w_{IV} := w_I$ constant.

2.3 Equilibrium Price Function

2.3.1 Benchmark: GS economy without VaR investors

In the economy without VaR constrained investors ($w_{IV} = 0$ so $w_{IF} = w_I$), the equilibrium is characterized by Grossman and Stiglitz (1980). Pávölgyi and Venter (2015) show that in the space of continuous price functions in $(s, m)$, the linear GS price function is the unique equilibrium. They also prove that there are non-linear equilibria outside the space of continuous functions. In general, in a GS economy, any equilibrium price contains information about a linear relationship between $s$ and $m$. Therefore, observing a particular price, uninformed agents can learn that $s + Cm = l$ for some constants $C$ and $l$. Pávölgyi and Venter (2015) illustrate that when price is restricted to be continuous, this relationship is the only information reflected in price. However, if price is not necessarily continuous, it is possible to learn more from price alone. Pávölgyi and Venter (2015) construct “linear cuts” that allow uninformed investors to know which region of the $(s, m)$ line (left or right cut) that the true $(s, m)$ lies. In their example, there are uncountably many discontinuous equilibrium price functions that exhibit price jumps and crashes. In a similar context
but extending to more general exponential family of distributions, Breon-Drish (2015) also demonstrates that there are many discontinuous equilibria.

In order to have a meaningful comparison between GS and VaR economy, we restrict our attention to the continuous equilibrium. The unique (continuous) equilibrium of the GS economy is given by

$$P_{GS}(s, \tilde{m}) = a_{s}^{GS}s - a_{m}^{GS}m,$$

where

$$a_{s}^{GS} = \frac{\frac{\tau_{e}}{\tau_{dp}} + \frac{w_{I}}{\rho_{I}} \frac{w_{U}}{w_{I}} \left( \frac{T_{d}}{\tau_{e}} + 1 + \frac{T_{d}}{\tau_{m}} (w_{I} \rho_{I} \tau_{e})^{-2} \right)^{-1}}{\frac{\tau_{e} + T_{d}}{\tau_{dp}} + \frac{w_{I} \rho_{I}}{w_{I} \rho_{I}}}$$

$$a_{m}^{GS} = \frac{1}{w_{I} \rho_{I}} a_{s}^{GS}$$

$$\tau_{dp} = \frac{\frac{\tau_{e} + T_{d}}{\tau_{e}} + \frac{T_{d}}{\tau_{m}} (w_{I} \rho_{I} \tau_{e})^{-2}}{\frac{\tau_{e}}{\tau_{e} + \frac{1}{\tau_{m}} (w_{I} \rho_{I} \tau_{e})^{-2}}.$$ 

2.3.2 VaR economy

Now, consider the economy with density $w_{IV} > 0$ of VaR investors. The methodology closely follows Yuan (2005) who considers a fictional economy populated only with informed traders with supply $\tilde{m}^{fic} = \tilde{m} - w_{U}x_{U}(\tilde{P})$. Again, we restrict our attention to price functions $P(s, m)$ that are continuous in both arguments. The basic idea is that the strategy of the informed is fixed in the sense that they do not learn any information from price. In other words, in any equilibrium for a given price $P$ and signal $s$, the informed demand is the same, because private signal is a sufficient statistics. For detailed arguments, see Yuan (2005), Pávolgyi and Venter (2015), or Breon-Drish (2015). As noted previously, throughout the paper, we maintain the following parameter assumption.

Assumption: $\rho_{I} [\Phi^{-1}(q)]^{2} > 4\alpha$.

Proposition 2.2 In the fictitious economy with only informed traders and noisy sup-
ply \( \hat{m}^{fic} \), the equilibrium \( P^{fic}(\hat{s}, \hat{m}^{fic}) \) is the unique solution to the following equation.

\[
\sum_{\text{region } r} w_{IV} \cdot \mathbb{1}_{\{\hat{s} \in S^r(P)\}} x^r_{IV} (\hat{s}, P) + w_{IF} \rho_F (\tau_e + \tau_d) \left( \frac{\tau_e}{\tau_e + \tau_d} \hat{s} - P \right) = \hat{m}^{fic},
\]

where \( x^r_{IV} \) is defined in (2.9) in Appendix.

Solving out the above equation, \( P^{fic}(\hat{s}, \hat{m}^{fic}) \) can be written as a piecewise-linear function of \( (\hat{s}, \hat{m}^{fic}) \).

\[
P^{fic} = \sum_{\text{region } r} \mathbb{1}_{\{\hat{s} \in S^r(P^{fic})\}} \left( a_0^r + a_s^r \hat{s} - a_m^r \hat{m}^{fic} \right).
\]

All the constants \( a_0^r, a_s, \) and \( a_m^r \) are defined in Appendix.

**Proposition 2.3** For a given price \( \hat{P}^{fic} = P \) of the fictional economy and an estimate of the aggregate uninformed demand \( \hat{x}_U = x_U \), the optimal uninformed demand is given by

\[
f(P, x_U) = \rho_U \left( \frac{E_U - P}{V_U} \right),
\]

where

\[
E_U = \sum_{\text{region } r} \left[ \pi_r \cdot E_r \right],
\]

\[
V_U = \sum_{\text{region } r} \left[ \pi_r \cdot \left( V_r + (E_r)^2 \right) \right] - (E_U)^2,
\]

\[
\pi_r \propto \sigma_r \exp \left( \frac{\mu_r^2}{2\sigma_r^2} \left[ 1 - \frac{(\sigma_r^2)^{-1}}{a_s^2} \left( \frac{a_m^r}{a_s^r} \right)^2 \right] \right) \left( \Phi(\theta_r) - \Phi(\theta_r^r) \right),
\]

\[
\sum_{\text{region } r} \pi_r = 1.
\]

The expressions for \( E_r, V_r, \sigma_r, \mu_r, \theta_r \) are provided in the Appendix as functions of \( P \) and \( x_U \).

**Proposition 2.4** For given price \( \hat{P}^{fic} = P \), the equilibrium uninformed demand is a solution to

\[
x^*_U = f(P, x^*_U).
\]

For any \( P \), there always exists a solution \( x^*_U \) that satisfies the above equation.
If the above equation has a unique solution \( x^*_U \) for every \( P \), \( x^*_U \) can be understood as a function of \( P \). While we were unable to determine the necessary and sufficient condition on the parameters that results in uniqueness of equilibrium, under reasonable parameters, numerical simulations suggest that in general the solution is unique. Hereafter, we assume that each \( P \) corresponds to only one value of \( x^*_U \).

**Theorem 2.1** The equilibrium price \( P = P(\bar{s}, \bar{m}) \) of the economy with VaR investors is the solution to the following equation.

\[
\bar{m} = w_{IV} \sum_{\text{region } r} \mathbb{1}_{\{\bar{s} \in S_r(P)\}} x_r(\bar{s}, P) + w_{IF} \rho_f(\tau_e + \tau_d)(\frac{\tau_e}{\tau_e + \tau_d} \bar{s} - P) + w_U x^*_U(P).
\]

**Proof** It follows from Proposition 2.4 and the market clearing condition.

## 2.4 Price Informativeness

### 2.4.1 Overview of VaR equilibrium

Figure 2-2 gives a brief overview of the equilibrium of VaR economy. The left and right figures correspond to GS and VaR equilibria, respectively. To get a clear picture of the equilibrium characteristics, it is important to first understand how the uninformed demand depends on price.

### 2.4.2 Uninformed Demand

Figure 2-3 plots uninformed optimal demand as a function of observed price. Consider when price is above the mean level 0 (i.e. the right half of Figure 2-3). In both economies the uninformed sell the asset, because of the likelihood that high price is driven by large noisy demand. However, in the constrained economy, the uninformed short more at moderately high price. While we provide a detailed explanation in the subsection 2.4.3, below is a short description of Figure 2-3. Consider an uninformed investor in the GS economy when observed equilibrium price is slightly above its average level. Then, she would short the risky asset, because conditional on the
Figure 2-2: Equilibrium GS and VaR Prices. The left (right) panel corresponds to GS (VaR) economy. The top figures plot the uninformed demand against price. Varying \( m \), the relationship between the signal and equilibrium price is illustrated in the middle. The bottom figures plot the equilibrium price against \( m \), by varying the signal. The parameter values used here are \( \rho_I = \rho_U = 1 \), \( \alpha = 0.3 \), \( q = 0.01 \), \( \tau_d = 5 \), \( \tau_e = 10 \), \( \tau_m = 0.5 \), \( \bar{m} = 0 \), \( w_{IF} = 5\% \), \( w_{IV} = 25\% \), \( w_U = 70\% \).

observed price she expects that signal is not as high to justify such price. She estimates that on average informed traders are also shorting the asset. Now, consider a situation in which some of informed agents become VaR constrained. If she observes the same price, then she would like to short more assets. This is because if the signal were the same, the VaR constraint forces the demand of the constrained investors upward and the price would have been higher. The uninformed investor rationally thinks that the signal must have been lower and consequently sells more assets at the same price. The exact opposite is true when price is moderately low in which case the agent buys more.
Figure 2-3: Uninformed investor optimal demand. The dashed line and solid curve correspond to the uninformed optimal demands in the GS economy and VaR economy, respectively. The parameter values used here are $\rho_I = \rho_U = 1$, $\alpha = 0.3$, $q = 0.01$, $\tau_d = 5$, $\tau_e = 10$, $\tau_m = 0.5$, $\bar{m} = 0$, $w_{IF} = 5\%$, $w_{IV} = 25\%$, $w_U = 70\%$.

At even higher but not too extreme price, the uninformed agent starts to sell less than the amount she would in the constraint-free economy. As price goes up, the likelihood that the informed are unconstrained increases, because the VaR rule is effective only when price is not too far away from the conditional mean of the payoff given signal, i.e., $P - \frac{\tau_e}{\tau_e + \tau_d}s$ is not too high. At this price level, although she estimates with high probability that the informed agents must be shorting, she cannot tell whether they are submitting their constrained ($P - \frac{\tau_e}{\tau_e + \tau_d}s$ is moderately high) or unconstrained ($P - \frac{\tau_e}{\tau_e + \tau_d}s$ is extremely high) demand. The estimated signals in the two scenarios are substantially different. At a certain price level, these events are equally likely, and the uncertainty of her inference problem becomes extremely noisy. As a risk-averse investor she reduces her short position, i.e., demands more.

When price is extremely high, the VaR rule is unlikely to bind. The uninformed agent projects that the informed submit their unconstrained GS demand at this price, and her optimal demand converges to her demand in the GS economy.
2.4.3 Price informativeness in the VaR economy

Figure 2-4: Price informativeness. The dashed and solid lines represent price informativeness (measured by $\text{Var}[d|s]/\text{Var}[d|P]$) of GS and VaR economy, respectively. The parameter values used here are the same as in Figure 2-3.

Figure 2-4 depicts the price informativeness of the VaR economy. Following the convention of the literature, price informativeness is measured by $\frac{\text{Var}[d|s]}{\text{Var}[d|P]}$. Since $s$ is a sufficient statistics for the asset payoff, this quantity is always bounded above by 1. In the GS economy, it is equal to 1 when investors are risk neutral or signal is perfect ($\tilde{\varepsilon}_s = 0$). More importantly, it is independent on price, because price is a linear function of $s$ and $m$. However, Figure 2-4 shows that in the VaR model it highly depends on the price level.

Unlike in the GS economy, the equilibrium price, signal and payoff are not jointly normal. An alternative measure that quantifies price informativeness is the entropy defined by $h(p) = -\int_{-\infty}^{\infty} f_{dp}(x|p) \log f_{dp}(x|p) \, dx$, where $f_{dp}(\cdot|\cdot)$ is the conditional density of payoff given price. When this quantity is higher, payoff is more uncertain given particular realization of $p$. Using this measure, we find the same result as Figure 2-4; as price increases from its mean, the entropy decreases and then increases until a certain point after which decreases and converges to the GS value. Also note that $\text{Var}[d|s]$ is independent on the signal by the joint normality assumption.

77
Due to the symmetry of the model, it is sufficient to consider price regions above the mean level (i.e. \( P > 0 \), the right half of Figure 2-4). Figure 2-4 suggests that the precision of the inference problem of the uninformed is very volatile in the price movement. An intuitive description of the mechanism is as follows, while a rigorous explanation follows after Figure 2-5.

For a slightly positive price \((P \in (0, 0.55])\), uninformed traders predict with high probability that the VaR constraint is binding, but are uncertain as to whether it is the long or short constraint. This uncertainty creates a relatively large variance in the inference problem. At a higher price \((P \in (0.55, 2.6])\), uninformed traders know with very high precision that the true state belongs to a narrow region. Hence, the variance becomes very small, or price is very informative. At even higher price \((P \in (2.6, 3.8])\), they are unable to tell whether the VaR constraint binds or not, because price is fairly high but not too high. Furthermore, the anticipated signals in each scenario are far from each other, which results in a larger variance in the inference problem. Finally, for an extremely high price \((P \in (3.8, \infty))\), it becomes very unlikely for the constraint to bind, and price conveys just as much information as it does in the constraint-free world. The price informativeness, therefore, eventually converges to that of the GS economy.

The above observations are formalized in Figure 2-5. The six solid curves are iso-price curves in the \((m, s)\) plane corresponding to \(P = 6, 4, 2, 0, -2, -4, -6\), respectively, from top to bottom. The contour map of the joint density of \((m, s)\) is also drawn. The unconditional density of \((m, s)\) is given by

\[
    f(m, s) = \left( \frac{\tau_m}{2\pi(\tau_e + \tau_d)} \right)^{1/2} e^{-\frac{\tau_c\tau_d}{2(\tau_e + \tau_d)} s^2} e^{-\frac{\tau_m}{2}(m-m)^2},
\]

so the level sets of the joint density are a collection of ellipses. Given price \(P\), the maximum likelihood point of \((m, s)\) is the one that touches the innermost ellipse. In the starred states of the nature along the iso-price curves, VaR-constrained agents hold a square position (zero net position). The circled points are the states of the GS economy in which the informed hold the same position at various price levels. Finally, the \(x\)-points represent the GS conditional mean of \((m, s)\) given \(P = 6, 4, 2, 0, -2, -4, -6\),
Figure 2-5: Iso-price curves and density contour map. The solid lines represent iso-price curves in the \((m, s)\) plane with corresponding price levels \(P = 6, 4, 2, 0, -2, -4, -6\), respectively. The starred and circled dot points are the states of nature in which the informed demands \(0\) in the VaR and GS economies, respectively. The x-points plot the conditional mean of \((m, s)\) given each price level of the GS economy. The parameters used here are the same as in Figure 3.

respectively. Since \((P, s, m)\) are jointly normal in GS, these points are the most likely states given a price level, or tangency points of the GS iso-price lines to the innermost ellipse. Note that the straight lines connecting the circled and x-points are the GS iso-price lines.

Just as in Figure 2-1, each iso-price curve has 7 regions; from left to right along the iso-price curve, label the regions by Region 1 to Region 7 (see Figure 2-6). Region 1 is generated by low signal and noisy supply (or large noisy demand). In this region, the VaR-informed short and submit their unconstrained demand. In Regions 2 and 3, they short but are constrained. They submit price-elastic and price-inelastic demands in Region 2 and 3, respectively (see Figure 2-1). Region 4 is again an unconstrained region in which price and expected mean of the payoff are close. In Regions 5 and
Figure 2-6: Constrained and unconstrained regions along iso-price curves.

6, they go long the asset price-inelastically and price-elastically, respectively. Finally they enter unconstrained long position in Region 7. For each given price $P$ and $1 \leq i \leq 6$, denote by $A_{i,i+1}(P)$ the intersection point of Region $i$ and Region $i+1$. The dependence on $P$ is suppressed when no confusion arises.

Define the Marginal Rate of Informational Substitution as

$$MRIS := -\frac{\partial p}{\partial m} \frac{\partial s}{\partial p}.$$ 

The $MRIS$ at a point $(m, s)$ measures the sensitivity of signal on noisy supply fixing the price corresponding to $(m, s)$. That is, it quantifies how much signal must respond to a unit change in noisy supply in order to keep the equilibrium price at the same level. Note that in $R_2$ and $R_5$, $MRIS$ is low because VaR-informed investors demand price-elastically; from (2.7), their demand is a function of $\frac{\tau_e}{\tau_e + \tau_d} s - P$. A change in $s$ makes them demand as if they have the same signal but different price. Therefore, in $R_2$ and $R_5$, a small change in signal leads to a big change in their demand, and in order for the market to clear, the corresponding noisy supply $m$ must move a lot. To put it another way, in these regions, a unit change in $m$ causes a little
variation in $s$, or the $MRIS = \frac{\partial s}{\partial m}$ is small. Using the reverse logic, the $MRIS$ is high in $R_3$ and $R_6$.

Interpolating from Figure 2-5, it can be seen that for prices close to but higher than 0, all of the regions $R_3$, $R_4$, and $R_5$ are likely to be the true states of the world. Intuitively, at these prices, it is hard to tell with certainty whether the informed are buying or selling. The signal must be low when the state is in $R_3$ and high when in $R_5$. Therefore, the projection of the signal is imprecise, and price conveys less informational content than it would in the GS economy.

Now, consider a higher price $P$, say $P = 2$ in Figure 2-5 (the third solid line from the top). The uninformed rationally estimate that the informed are likely to be shorting and constrained. Suppose that they project that the true state of the world lies in $R_2$, in which $MRIS$ is low. We claim that any state $(m, s)$ that is in the interior of $R_2$ is not likely to be the true state. To see this, consider a point $(m + \Delta_m, s + \Delta_s)$ slightly right to $(m, s)$ along the iso-price curve at $P = 2$. The $MRIS$ is low, so $\Delta_m \gg \Delta_s$. Combined with the fact that $m$ is hugely negative in this region ($P$ is high and the signal is low, so the corresponding noisy supply $m$ has to be considerably negative), it follows that $f_{m,s}(m + \Delta_m, s + \Delta_s) \gg f_{m,s}(m, s)$. In other words, any interior points of $R_2$ is very unlikely to be realized, so conditional on the true state lying in $R_2$, the rightmost point $A_{2,3}$ has extremely high density. If the true state were in $R_3$, instead, $A_{2,3}$ is again the state that has the maximum density. The reason is that if $(m, s)$ and $(m - \Delta_m, s - \Delta_s)$ were both in $R_3$, because the $MRIS$ is high, $\Delta_s \gg \Delta_m$, and $f_{m,s}(m - \Delta_m, s - \Delta_s) > f_{m,s}(m, s)$. In any case, only a small neighborhood of $A_{2,3}$ constitutes the "likely states." As a result, the price conveys much information about the signal. Graphically, Figure 2-5 suggests that a neighborhood around $A_{2,3}$ absorbs much of the density when $P = 2$.

The above argument is not an artifact of the linear approximation of the constraint. Under the actual form (2.3), it is still true that in states where VaR investors short-sell but are constrained, the $MRIS$ is monotonic along their respective iso-price curves. By a similar logic used above, the state that is most likely to occur is when $MRIS$ is not too high nor too low. Therefore, the likely states are in a small neighborhood of
the state with an intermediate MRIS value, and, in turn, price is very informative at a moderately high level.

Suppose that price becomes even higher (between the second and third solid lines from the top in Figure 2-5). Before, the uninformed estimated that the true state is extremely unlikely to lie in $R_1$. But now, such likelihood becomes substantial. Recall that when signal is very low relative to price, VaR investors are unconstrained. Observing a higher price, the uninformed therefore infer that the state is more likely to be in $R_1$ than it would be at a previously lower price. As such likelihood increases, their inference problem becomes noisier; they cannot tell whether the state is in $R_1$ or $R_2$, and the implied signals by the two regions differ by a lot. As risk-averse agents, the uninformed would like to reduce their short position, or demand more than before. Graphically in Figure 2-5, the starred point on the $P = 4$ line moves to the right of that on the $P = 2$ line, because the increase in the uninformed demand must be met by an increase in noisy supply. The point $A_{1,2}$ also moves to the right, because it is in a neighborhood of the starred point, the state in which the informed agents demand 0. Recall that $A_{1,2}$ is the density-absorbing state conditional on the constraint binding. As $A_{1,2}$ moves to the right, the likelihood of the state belonging to $R_2$ relative to $R_1$ decreases. This again, worsens the uninformed inference problem, which positively feeds back to creating even more noise. Eventually, at a price level, the two regions become equally likely to contain the true state of the nature. At that level, price reveals very little information about the signal. In Figure 2-5, such level corresponds to approximately $P = 4$, at which the innermost ellipse touches $A_{1,2}$ and $R_1$ at the same time.

The above mechanism explains why uninformed demand increases rapidly at an intermediate price level in Figure 2-3. When the uninformed scale down their position, their inference problem becomes less precise, and then they reduce their position even further, and so on. In particular, around this critical price, the uninformed demand more at a higher price. Such a “backward-bending” of uninformed demand curve also documented in Yuan (2005) eventually creates a jump or crash of equilibrium price.

When price is extremely high, say $P = 6$ (the top solid line in Figure 2-5), the
constraint almost surely does not bind. Even though the expected signal is high, it is not as high as to justify such high price. The uninformed estimate with high precision that the state of the nature is in \( R_1 \), in which the VaR investors trade without being constrained. Therefore, in this price region, the price informativeness converges to that of the GS economy.

The price informativeness curve (Figure 2-4) is now well understood. As price moves away from 0 in the positive direction, it first increases and then decreases and finally increases. However, the two latter regions are extremely unlikely to be realized. Simulations using various parameters estimate that such price occurs with probability less than \( 10^{-8} \). So realistically, a neighborhood around \( P = 0 \) should be the region of interest, in which the price informativeness increases in the deviation of price from its mean.

### 2.4.4 Price informativeness in other constrained economies

As noted earlier, VaR constraint is distinctive in that the constraint only binds for moderate price levels. For illustration, we contrast our results to the ones of an economy under other popular constraints. In particular, we show how price informativeness varies across different price levels in leverage (Yuan (2005)) and short-sale (Yuan (2006) and Bai et al. (2006)) constrained economies.

Figure 2-7 demonstrates that price informativeness is almost globally lower than that of the GS economy. In a leverage (borrowing) constrained economy, price reveals little information about the asset payoff when price is low, and just as much as in the GS world when high. The exact opposite is true for short-sale constrained economy. The reason is that in the former (latter), the constraint binds when price is lower (higher) than the conditional payoff of the asset given signal. As price deviates much, the likelihood that the constraint is binding becomes negligible, so the price informativeness converges to that of the GS economy.
Figure 2-7: Price informativeness in leverage (borrowing) and short-sale constrained economies. In (a), the constraint is modeled by \( x \leq aP + b \) and in (b), \( x \geq -\xi \), where \( x \) is the number of units that constrained agents can buy. The parameters values are \( \rho_l = \rho_U = 1, \alpha = 0.3, q = 0.01, \tau_d = 5, \tau_e = 25, \tau_m = 0.5, \bar{m} = 0, w_{IF} = 5\%, w_{bc} = w_s = 25\%, w_U = 70\%, a = 2, b = -1, \xi = 0.1 \), where \( w_{bc} \) and \( w_s \) denote the population of borrowing and short-sale constrained but informed investors.

### 2.5 Price and Return Distribution

In this section, we study how VaR constraint affects price and return distributions.

#### 2.5.1 Price volatility

To get a better grasp of the price dynamics, we compare the equilibrium price of the VaR and GS economies in Figure 2-8, as a function of \( m \) for various values of \( s \). This figure juxtaposes the VaR prices with the GS prices. Several observations can be made from this figure.

First, similar to Figure 2-5, each VaR price line (holding signal \( s \) fixed and varying \( m \)) has seven regions. For price above the mean 0, the VaR price is below the GS price; when price is moderately high, by the mechanism discussed in subsections 2.4.2 and 2.4.3, the uninformed short more than they would in the GS economy given the same price \( P \) and signal \( s \).\textsuperscript{10} By the market clearing condition, the noisy supply

\textsuperscript{10}As discussed in previous sections, this is only true when price is moderately high; when price is higher, the uninformed demand is actually lower than that of the GS model. The reverse holds for
Figure 2-8: Equilibrium price and noisy supply for various signals. This figure plots the equilibrium price against noisy supply $m$ under various realizations of the signal. The dashed lines represent the GS equilibrium prices for $s = 4.7, 2.35, 0, -2.35, -4.7$ (top to bottom). The solid lines are the equilibrium prices in the VaR economy with the same signals as in the GS economy. The $s$ values correspond to $0, \pm 5 \text{std}(s), \pm 10 \text{std}(s)$, respectively. The underlying parameters are $\rho_I = \rho_U = 1, \alpha = 0.3, q = 0.01, \tau_d = 5, \tau_e = 50, \tau_m = 0.5, \bar{m} = 0, w_{IF} = 5\%, w_{IV} = 25\%, w_U = 70\%$.

that corresponds to $(P, s)$ must be lower in the VaR model. In other words, if $P = P_{GS}(s, m) = P_{IV}(s, m')$, then $m > m'$. Therefore, at the same $(s, m)$, $P_{IV}(s, m) < P_{GS}(s, m)$. The opposite is true when price is below its mean; the uninformed buy more and the VaR price is higher than the GS price for each $(s, m)$.

Second, unlike the GS model, market liquidity depends on the state of the world.\textsuperscript{11} a very low price. However, such events happen with virtually zero probability, so here we restrict our attention to a reasonable price range.

\textsuperscript{11}As a convention in the literature, liquidity is measured by $\partial p/\partial m$, the sensitivity of price on noisy supply. When this quantity is large in absolute term, price responds more to noise, and the
The liquid regions are the states in which the VaR-informed traders submit price-elastic-constrained demand. In these regions, $MRIS$ is low — signal is insensitive to noise — and consequently price is insensitive to noise. By the reverse logic, in states in which they submit price-inelastic-constrained demand, market is illiquid. Note that if the exact VaR constraint (2.3) were used, the equilibrium would exhibit smooth curves.

The allocation effect of VaR constraint gives rise to liquid and illiquid regions, which are the states of the world in which the constraint binds and VaR investors demand less in absolute value. In these regions, the noisy demand must react accordingly in order for the market to clear. In contrast, in non-binding regions it is the information effect of VaR constraint that makes the VaR and GS prices differ from each other. Even though the constraint is slack, uninformed traders do not know that and only infer from price. Since there is a probability that the constraint binds, they demand more when price is lower and less when higher.

The price volatility is determined by the relative strength of the above effects. If liquid regions are wide and uninformed demand deviates much from the benchmark, price becomes less volatile. Otherwise, price is more volatile. In the next subsection, we describe under which circumstance one effect dominates the other.

2.5.2 Price and return distribution

We compare the equilibrium price and return of the uninformed of the benchmark GS economy and VaR economy. The mass of informed traders is kept at 30%, while the proportion of VaR investors varies from 0% (GS) to 30%.

Table 2.2 shows unconditional moments of price and return, using the parameter values described in Table 2.1 when signal is precise (high $\tau_c$). As more informed traders follow the VaR rule (increase in $w_{IV}$), price becomes less volatile. In particular, price volatility is lower in the VaR model than that in the GS model. When there is less

market is considered illiquid.
Table 2.1: Parameter Choice.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\rho_I$</td>
<td>Risk tolerance of the informed</td>
<td>1</td>
</tr>
<tr>
<td>$\rho_U$</td>
<td>Risk tolerance of the uninformed</td>
<td>1</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>VaR critical value</td>
<td>30%</td>
</tr>
<tr>
<td>$q$</td>
<td>VaR probability</td>
<td>1%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Asset/signal</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$\tau_d$</td>
<td>Precision of asset payoff</td>
<td>5</td>
</tr>
<tr>
<td>$\tau_e$</td>
<td>Precision of noise</td>
<td>10 or 50</td>
</tr>
<tr>
<td>$\tau_m$</td>
<td>Precision of the noisy supply</td>
<td>0.5</td>
</tr>
<tr>
<td>$\bar{m}$</td>
<td>Mean of noisy supply</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Population</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$w_{IF}$</td>
<td>Mass of the unconstrained informed</td>
<td>0 - 30%</td>
</tr>
<tr>
<td>$w_{IV}$</td>
<td>Mass of the VaR informed</td>
<td>30 - 0%</td>
</tr>
<tr>
<td>$w_U$</td>
<td>Mass of the uninformed</td>
<td>70%</td>
</tr>
</tbody>
</table>

Table 2.2: Price and Return Moments and Tail Loss Probabilities when $\tau_e = 50$.

<table>
<thead>
<tr>
<th>VaR $w_{IV}(%)$</th>
<th>GS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
</tr>
<tr>
<td>Price mean</td>
<td>0</td>
</tr>
<tr>
<td>std</td>
<td>0.424</td>
</tr>
<tr>
<td>Uninformed mean (%)</td>
<td>0.16</td>
</tr>
<tr>
<td>std (%)</td>
<td>3.97</td>
</tr>
<tr>
<td>$\mathbb{P}_U[\Delta W &lt; -\alpha] (%)$</td>
<td>0.005</td>
</tr>
</tbody>
</table>

Table 2.2: Price and Return Moments and Tail Loss Probabilities when $\tau_e = 50$.

noise in the signal, the VaR constraint is binding only for a small price region for each $s$, because VaR investors know the random payoff $d$ with high precision. Table 2.2 suggests that when $\epsilon_s$ has less noise, information effect dominates the allocation effect.

The mean of the price is 0, due to the symmetry of our model. However, uninformed investors' return becomes much more volatile (3.97 vs. 21.63). This, in turn, increases the probability of big losses. Since the price distribution changes only by little, the increase in return volatility is due to a change in uninformed investors' asset holdings, which is documented in Figure 2-3.

If information is imprecise (low $\tau_e$), price volatility increases in the number of VaR-


<table>
<thead>
<tr>
<th></th>
<th>VaR $w_{IV}$ (%)</th>
<th>GS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>Price mean</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>std</td>
<td>0.367</td>
<td>0.369</td>
</tr>
<tr>
<td>Uninformed mean (%)</td>
<td>6.68</td>
<td>10.72</td>
</tr>
<tr>
<td>return std (%)</td>
<td>27.52</td>
<td>34.84</td>
</tr>
<tr>
<td>$\mathbb{P}_U[\Delta \tilde{W} &lt; -\alpha]$ (%)</td>
<td>4.802</td>
<td>6.045</td>
</tr>
</tbody>
</table>

Table 2.3: Price and Return Moments and Tail Loss Probabilities when $\tau_c = 10$.

informed traders, holding the total informed population fixed, although the difference is not big (0.367 vs. 0.371 in Table 2.3). Intuitively, when signal is noisy, the VaR constraint is binding for a wide price region. Therefore, the allocation effect of the constraint dominates its information effect and causes price to be more volatile. But even when signal is inaccurate, the uninformed still bear much increased risk and expect higher payoff. In other words, risk is transferred from the informed to the uninformed regardless of signal precision.

### 2.6 Endogenous Information Acquisition

We now allow the uninformed investors to acquire information at a constant cost $c > 0$ similar to Verrecchia (1982). The decision depends on the current price $P$. Since we assume that any agent without the VaR constraint can choose to be informed, let us denote the risk tolerance of non VaR-constrained agents by $\rho$. Let $U$ be their corresponding utility function. VaR investors are assumed to be informed at all times. An equilibrium consists of a price function $P = P(s, m)$ and populations of informed and uninformed agents $w_{IF}(P), w_U(P)$ such that

1. VaR investors choose their optimal demand:

   $$ x_{IV}^*(s, P) \in \arg \max_x \mathbb{E}[U_{IV}(\tilde{W}_{IV}; x_{IV}) | s, P] $$

   subject to the VaR constraint (2.3).
2. Unconstrained agents submit their optimal demand:

\[ x^*_IF(s, P) \in \arg \max_x \mathbb{E}[U(W; x) \mid s, P] \]

\[ x^*_U(P) \in \arg \max_x \mathbb{E}[U(W; x) \mid P] \]

3. For any \( P \),

\[ w^*_IF(P) > 0 \Rightarrow \mathbb{E}[\mathbb{E}[U(W; x^*_IF(s, P)) \mid s, P] \mid P] - c \geq \mathbb{E}[U(W; x^*_U(P)) \mid P] \]

\[ w^*_U(P) > 0 \Rightarrow \mathbb{E}[\mathbb{E}[U(W; x^*_IF(s, P)) \mid s, P] \mid P] - c \leq \mathbb{E}[U(W; x^*_U(P)) \mid P] \]

4. The market clears: For every \((s, m)\)

\[ w^*_IV x^*_IV(s, P) + w^*_IF(P) x^*_IF(s, P) + w^*_U(P) x^*_U(P) = m. \]

The third requirement is the optimality condition of information acquisition; when the current price is \( P \), and if there is a strictly positive mass of informed traders \((w^*_IF(P) > 0)\), it must be that the expected utility gain from becoming informed

\[ \mathbb{E}[U(W; x^*_IF(s, P)) \mid s, P] \mid P] - \mathbb{E}[U(W; x^*_U(P)) \mid P] - c \]

is non-negative. It follows that for \( w^*_IF(P), w^*_U(P) > 0 \) to hold, unconstrained traders are indifferent between being informed and uninformed:

\[ \mathbb{E}[U(W; x^*_IF(s, P)) \mid s, P] \mid P] - c = \mathbb{E}[U(W; x^*_U(P)) \mid P]. \]

Analytical solution is hard to obtain, so we solve for the equilibrium numerically.

Figure 2-9 shows the dependence between price and population composition with respect to information acquisition. The population of VaR-informed agents are assumed to be kept at 10%. The area bounded by solid lines represents the result in the VaR economy. Note that no agent (other than fixed VaR investors) acquires information when price is around its mean. This is because conditional on moderate price deviation from its average level, payoff \( d \) is expected to be close to the current price \( P \) even after conditioning private signal \( s \) (if acquired). Since expected utility of the informed is an increasing function in the absolute difference between expected payoff
Figure 2-9: Population of informed and uninformed agents. Solid: with VaR investors. Dashed: without VaR investors. The cost of acquiring information is $c = 2$.

and price $(\mathbb{E}[d|s] - P)$, the expected utility gain of being informed is not big enough to justify the cost $c$. As price deviates more from its mean, more agents choose to be informed. Consequently, price becomes more informative as price is further away.

To better understand the effect of VaR constraint on information acquisition, we compare the VaR economy to the economy without VaR investors. To make a fair comparison, we assume that in the latter there is a fixed proportion (10%) of unconstrained investors who acquire information at any price level. That is, some investors must stay informed (e.g. banks) but are not subject to regulations or do not employ VaR-based risk management. The population composition in the VaR-free-economy is given in Figure 2-9 with a dashed line. It can be seen that VaR constraint generally discourages information acquisition; less agents become informed at any price level. Moreover, as price moves away from its mean, the discouragement is more severe. This is consistent with our earlier observation that when some informed investors are subject to VaR constraint, price reveals more information when it is not near the mean, thereby decreasing the value of buying additional private information.
2.7 Conclusion

In this paper, we showed that, in an information-asymmetric economy in which some informed agents have Value-at-Risk (VaR) constraint, the extent to which price conveys information depends on the price level. In particular, moderately high or low prices reveal more information in the constrained model than in the constraint-free model. We also illustrated that price volatility depends on the precision of private information; price can be less volatile than in the benchmark when signal is precise. The constraint transfers much of the risk to the uninformed agents. As a result, they realize big losses with increased probability. When information acquisition is an endogenous choice, we demonstrated that VaR constraint disincentivizes investors from being informed.

All of the above distinctive effects arise through the information channel of the constraint. When some informed traders become constrained, the inference problem of the uninformed is altered, which potentially has a huge effect on the equilibrium outcome. The main results of this paper stem from the unique characteristics of the VaR constraint; the fact that the constraint binds only when price and signal are not far away is essential in characterizing price informativeness.
Bibliography


2.8 Appendix

Proof of Proposition 2.1

Due to symmetry, let us only consider the case when \( P < \mathbb{E}[\tilde{d} | s] \). The VaR-constrained informed (henceforth the bank) optimally chooses \( x^* > 0 \). Since the bank enters a long position, for \( x > 0 \) the VaR constraint can be written as

\[
Pr \left[ \tilde{d} \leq P - \frac{\alpha}{x} \big| s, P \right] \leq q
\]

\[
\Leftrightarrow F_{\tilde{d}|s} \left( P - \frac{\alpha}{x} \right) \leq q,
\]

where \( F_{\tilde{d}|s} \) is the cdf of \( \tilde{d} \) conditional on \( s \). For low enough \( P \), the constraint is trivially satisfied for any \( x \) (say \( P \leq F_{\tilde{d}|s}^{-1}(q) \)), so the constrained demand must equal to the unconstrained one. ■

Proof of Proposition 2.3

Define the following constants.

\[
u_h = \frac{1}{2\sqrt{\tau_c + \tau_d}} \left( -\Phi^{-1}(q) + \sqrt{[\Phi^{-1}(q)]^2 - \frac{4\alpha}{\rho_l}} \right)
\]

\[
u_l = \frac{1}{2\sqrt{\tau_c + \tau_d}} \left( -\Phi^{-1}(q) - \sqrt{[\Phi^{-1}(q)]^2 - \frac{4\alpha}{\rho_l}} \right)
\]

\[g_h = \rho_l(\tau_c + \tau_d)u_h, \quad g_l = \rho_l(\tau_c + \tau_d)u_l\]

\[\beta_h = \rho_l(\tau_c + \tau_d)\frac{u_h}{u_l}, \quad \beta_l = \rho_l(\tau_c + \tau_d)\frac{u_l}{u_h}\]

\[u_m = \frac{[\beta_h u_h - g_h] - [\beta_l u_l - g_l]}{\beta_h - \beta_l}\]

Also, define constants

\[\kappa_p = \frac{\tau_c + \tau_d}{\tau_c}\]

\[\chi_h = \kappa_p u_h, \quad \chi_m = \kappa_p u_m, \quad \chi_l = \kappa_p u_l,\]

95
and regions in $\mathbb{R}$

\[
S^{h,+}(P) = (\kappa_P P + \chi_h, \infty) \\
S^{m,+}(P) = (\kappa_P P + \chi_m, \kappa_P P + \chi_h] \\
S^{l,+}(P) = (\kappa_P P + \chi_l, \kappa_P P + \chi_m] \\
S^o(P) = [\kappa_P P - \chi_l, \kappa_P P + \chi_l] \\
S^{h,-}(P) = (-\infty, \kappa_P P - \chi_h) \\
S^{m,-}(P) = [\kappa_P P - \chi_h, \kappa_P P - \chi_m) \\
S^{l,-}(P) = (\kappa_P P - \chi_m, \kappa_P P - \chi_l).
\]

Finally, define the region-dependent coefficients

\[
a_0^0 = a_0^{h,+} = a_0^{h,-} = 0, \quad a_0^o = a_m^{h,+} = a_m^{h,-} = \frac{1}{\rho_I(\tau_c + \tau_d)(w_{IF} + w_{IV})} \\
a_0^{m,+} = \frac{w_{IV}(g_l - \beta_l u_l)}{w_{IF} \rho_I(\tau_c + \tau_d) + \beta_l w_{IV}}, \quad a_0^{m,-} = \frac{1}{w_{IF} \rho_I(\tau_c + \tau_d) + \beta_l w_{IV}} \\
a_0^{l,+} = \frac{w_{IV}(g_l - \beta_l u_l)}{w_{IF} \rho_I(\tau_c + \tau_d) + \beta_l w_{IV}}, \quad a_0^{l,-} = \frac{1}{w_{IF} \rho_I(\tau_c + \tau_d) + \beta_l w_{IV}} \\
a_0^{m,-} = -a_0^{m,+}, \quad a_m^{m,-} = a_m^{m,+} \\
a_0^{l,-} = -a_0^{l,+}, \quad a_m^{l,-} = a_m^{l,+},
\]

and the region-common $s$-coefficient

\[
a_s = \frac{\tau_c}{\tau_c + \tau_d}.
\]

Then, rewriting (2.7), the VaR investors' optimal demand depends on the region:

\[
x^*_IF(\tilde{s}, \tilde{P}) \\
x^*_IV(\tilde{s}, \tilde{P}) = \left\{
\begin{array}{ll}
x^*_IF(\tilde{s}, \tilde{P}) & \tilde{s} \in S^{h,+}(\tilde{P}) \cup S^o(\tilde{P}) \cup S^{h,-}(\tilde{P}) \\
-\beta_h(\tilde{P} - \frac{\tau_c}{\tau_c + \tau_d} \tilde{s} + u_h) + g_h & \tilde{s} \in S^{m,+}(\tilde{P}) \\
-\beta_l(\tilde{P} - \frac{\tau_c}{\tau_c + \tau_d} \tilde{s} + u_l) + g_l & \tilde{s} \in S^{l,+}(\tilde{P}) \\
-\beta_l(\tilde{P} - \frac{\tau_c}{\tau_c + \tau_d} \tilde{s} - u_l) - g_l & \tilde{s} \in S^{l,-}(\tilde{P}) \\
-\beta_h(\tilde{P} - \frac{\tau_c}{\tau_c + \tau_d} \tilde{s} - u_h) - g_h & \tilde{s} \in S^{m,-}(\tilde{P})
\end{array}
\right.
\]

(2.9)
where \( x^*_F(\hat{s}, \hat{P}) = -\rho F(\tau_e + \tau_d) \left( \hat{P} - \frac{\tau_e}{\tau_e + \tau_d} \hat{s} \right) \) is the unconstrained optimal demand. Denote by \( x^*_I \) the above optimal VaR demand in each region.

Define
\[
\mu_r(P, x_U) = \frac{1}{a_s} \cdot \frac{1}{1 + \left( \frac{a_s}{a_m} \right)^2 \frac{\tau_e \tau_d}{\tau_m (\tau_e + \tau_d)}} (P - a_0^r + a_m^r (\hat{m} - w_U x_U))
\]
\[
\sigma^2_r = \frac{1}{\frac{\tau_e \tau_d}{\tau_e + \tau_d} + \left( \frac{a_s}{a_m} \right)^2 \tau_m}
\]

Also, let the endpoints of \( S^r(P) \) be \( c^r(P) \) and \( \bar{c}^r(P) \), allowing \( \pm \infty \). \(^{12}\) Again, for each region \( r \), let
\[
\bar{\theta}_r(P, x_U) = \frac{\bar{c}^r(P) - \mu_r(P, x_U)}{\sigma_r}, \quad \theta_r(P, x_U) = \frac{c^r(P) - \mu_r(P, x_U)}{\sigma_r}
\]
\[
\lambda_r(P, x_U) = -\phi(\bar{\theta}_r) - \phi(\theta_r)
\]
\[
\delta_r(P, x_U) = \lambda_r \left( \frac{\theta_r \phi(\bar{\theta}_r) - \bar{\theta}_r \phi(\theta_r)}{\phi(\bar{\theta}_r) - \phi(\theta_r)} \right),
\]
with the understanding that \( \phi(+\infty) = \phi(-\infty) = 0 \) and \( \Phi(+\infty) = 1, \Phi(-\infty) = 0 \).

Denote by \( E_r(P, x_U) \) and \( V_r(P, x_U) \) the expectation and variance of \( d \) conditional on \( P \) in the fictional economy, the estimate of the aggregate uninformed demand \( x_U \), and \( s \in S^r(P) \). That is,
\[
E_r(P, x_U) : = \mathbb{E} \left[ d | P = P^{fic}(\hat{s}, \hat{m}^{fic}), \hat{x}_U = x_U, \hat{s} \in S^r(P) \right]
\]
\[
V_r(P, x_U) : = \mathbb{V} \left[ d | P = P^{fic}(\hat{s}, \hat{m}^{fic}), \hat{x}_U = x_U, \hat{s} \in S^r(P) \right].
\]

Simple algebra yields that
\[
E_r(P, x_U) = \frac{\tau_e}{\tau_e + \tau_d} \left[ \mu_r(P, x_U) + \lambda_r(P, x_U) \sigma_r \right]
\]
\[
V_r(P, x_U) = \frac{1}{\tau_e + \tau_d} + \left( \frac{\tau_e}{\tau_e + \tau_d} \right)^2 (1 - \delta_r(P, x_U)) \sigma^2_r.
\]

Finally, the conditional probability \( \pi_r(P, x_U) : = \mathbb{E} [\hat{s} \in S^r(P) | P = P^{fic}(\hat{s}, \hat{m}^{fic}), \hat{x}_U = \hat{P}] \)

\(^{12}\)For example, \( \epsilon^{h,+}(P) = \kappa_p P + \chi_h, \bar{\epsilon}^{h,+}(P) = \infty \) and \( \epsilon^{m,+}(P) = \kappa_p P + \chi_m, \bar{\epsilon}^{m,+}(P) = \kappa_p P + \chi_m. \)
is given by
\[
\pi_r \propto \int_{s(P)}^{e^r(P)} \exp \left( -\frac{\tau_c r_d}{2(\tau_c + \tau_d)} s^2 \right) \exp \left( -\frac{\tau_m}{2(a_m^r)^2} \left( P - a_0^r - a_s s + a_m^r (\bar{m} - w_U x_U) \right)^2 \right) ds \\
\propto \sigma_r \exp \left( \frac{\mu_r^2}{2\sigma_r^2} \left[ 1 - (\sigma_r^2 \tau_m)^{-1} \left( \frac{a_m^r}{a_s} \right)^2 \right] \right) (\Phi(\bar{\theta}^r) - \Phi(\bar{\theta}_r)).
\]

**Proof of Proposition 2.4**

Using the fact that for any \( a \), \( \lim_{t \to \pm \infty} t [\Phi(t + a) - \Phi(t)] = 0 \) and \( \phi'(x) = -x\phi(x) \), by L'Hôpital rule,

\[
\lim_{t \to \pm \infty} \frac{\phi(t + a) - \phi(t)}{\Phi(t + a) - \Phi(t)} + t \\
= \lim_{t \to \pm \infty} \frac{(t + a)\phi'(t + a) + t\Phi(t + a) - \Phi(t)}{\Phi(t + a) - \Phi(t)} \\
= \lim_{t \to \pm \infty} \frac{-(t + a)\phi'(t + a) + t\Phi(t + a) - \Phi(t)}{\phi(t + a) - \phi(t)} + t
\]

\[
= \begin{cases} 
0 & t \to \infty \\
-a & t \to -\infty.
\end{cases}
\]

Hence, \( [\phi(t + a) - \phi(t)]/[\Phi(t + a) - \Phi(t)] = -t + O(1) \) as \( t \to \pm \infty \).

Using the definition of \( \theta_r \) and \( \lambda_r \), it is immediate that for except \( l+ \) region \( r \)

\[
\lambda_r(P, x) = -\frac{\mu_r(P, x)}{\sigma_r} + O(1), \quad x \to \infty
\]

and

\[
\lambda_{l+}(P, x) \to 0, \quad x \to \infty.
\]

As a result,

\[
E_r(P, x_U) = \frac{\tau_c}{\tau_c + \tau_d} [\mu_r(P, x_U) + \lambda_r(P, x_U)\sigma_r]
\]

is bounded above as \( x_U \to +\infty \). Similarly, as \( x_U \to -\infty \), \( E_r(P, x_U) \) is bounded below.
(note that in this case $\lambda_{h,+}(P, x_U) \to 0$). Moreover, $V_r(P, x_U) = O(1)$, so the function

$$x_U - f(P, x_U)$$

goes to $+\infty$ as $x_U \to +\infty$ and $-\infty$ as $x_U \to -\infty$. Since the above function is continuous in $x_U$, by the Intermediate Value Theorem, there exists a solution for $x_U$.

$\blacksquare$
Chapter 3

Controlling Tail Risk Measures with Estimation Error

3.1 Introduction

Assessment of risk and its control play an important role in finance, management, actuarial science, and operations research. In the context of investment decision making, Markowitz (1952) and Roy (1952) developed theory of optimal investment decisions that take into account risk. Value-at-Risk (VaR) has been adapted in Wall Street since the 1970’s. Artzner et al. (1999) proposed the concept of coherence as a desirable property of risk measures. Rockafellar and Uryasev (2000) proposed expected shortfall (ES), a coherent risk measure that is now widely used in banking regulations Chen (2014).\(^1\) ES is generalized to distortion risk measures, a class of risk measures that are coherent Acerbi (2002). Ahmadi-Javid (2012) propose another coherent risk measure that gives an upper bound on VaR and expected shortfall and has computational advantages. Various other risk measures are succinctly summarized in Dowd and Blake (2006).

Risk measures are often used as one of the key components in large systems. In

\(^1\)This quantity is originally named conditional Value-at-Risk. We use expected shortfall since conditional Value-at-Risk sometimes refers to a different concept in the literature Chernozhukov and Umantsev (2001); Wang and Zhao (2016).
investment decision making, there is vast literature on portfolio optimization with VaR or ES constraints (e.g. see Pflug (2000), Campbell et al. (2001), Favre and Galeano (2002), Pflug (2000), Rockafellar and Uryasev (2000), Ciliberti et al. (2007)). The equilibrium consequences of VaR constraints are also studied, both from partial equilibrium standpoint (Basak and Shapiro (2001), Cuoco and Liu (2006), and Cuoco et al. (2008)) and from general equilibrium standpoint (Danielsson et al. (2004)). In banking regulations, the required capital for each bank is calculated based on VaR measure (Chen (2014)). However, since risk measures depend on the underlying unknown return distributions, we need to estimate them to be used in practice.


How crucial is the fact that risk measures are estimated? Concerns about accuracy of practically used VaR estimates have been raised from the early stages of the literature. (Jorion (1996), Hendricks (1996), Pritsker (1997b), Pritsker (1997a), Pritsker (2006), Barone-Adesi and Giannopoulos (2001), Berkowitz and O’Brien (2002), and Aussenegg and Miazhynskaia (2006)). Also, Caccioli et al. (2017) note that due to high dimensionality of institutional portfolios and the lack of stationarity in the long run, portfolio optimization is plagued by (relatively) small sample problems. Correspondingly, they derive “lower bounds” to estimation errors in the context of ES optimization. Estimation errors can thus be a non-negligible aspect in risk control.

McNeil et al. (2005) also warn the danger of direct interpretation of VaR using estimated quantities, noting that the estimate of the loss distribution is subject to
estimation error and model risk as well as market liquidity. In the practice of banking regulations, banks are usually required to hold about 3 times their Value-at-Risk estimate in order to cope with such uncertainty (Chen (2014)). This number 3 is called the multiplier, and Stahl (1997) justifies the choice of 3 retrospectively by showing that it approximately accounts for model uncertainty for VaR. The Basel III also uses ES as a risk measure, and Leippold and Vanini (2002) argue that the multiplier of 1.5 is appropriate for dealing with model uncertainty with ES. See also Hendricks and Hirtle (1997), Lopez (1998), and Novak (2010) regarding this matter. However, there has not been much discussion on how to integrate potential estimation errors into risk management, beyond quantifying the distributions (standard errors) of the estimates.

The aim of this paper is to develop a method to reflect the precision assessment on the control of risk measures that are given by probability bounds. For example, consider VaR of confidence level $1 - \alpha$: $\Pr(X \leq -\text{VaR}_\alpha) \leq \alpha$. Controlling VaR at $1 - \alpha$ thus means bounding the probability of a portfolio loss exceeding the VaR by $\alpha$. True VaR is unobservable, however, so an estimate $\widehat{\text{VaR}}_\alpha$ is used to the control risk in practice. The problem with this estimate is that we are not ensured to have $\Pr(X \leq -\widehat{\text{VaR}}_\alpha) \leq \alpha$; thus, the probability of a large loss is no longer controlled in the original sense. Suppose that a one-sided $(1 - \alpha)$-confidence interval for VaR is available, satisfying $\Pr(\text{VaR}_\alpha \geq \widehat{\text{VaR}}_\alpha) \leq \alpha$. Then, by the Bonferroni inequality, we have

$$\Pr(X \leq -\widehat{\text{VaR}}_\alpha) \leq \Pr(X \leq -\text{VaR}_\alpha) + \Pr(-\text{VaR}_\alpha \leq -\widehat{\text{VaR}}_\alpha) = 2\alpha.$$ 

Thus, $\widehat{\text{VaR}}_\alpha$ "controls" true VaR with confidence level $1 - 2\alpha$. Retrospectively, if we want to control true VaR at $1 - \alpha$, we may have chosen $\alpha/2$ for each probability on the right-hand side.

This paper generalizes this observation and defines a class of risk measures to which the above argument is applicable, which we call tail risk measures. Intuitively, if a risk measure prescribes the probability of some bad events to be lower than some threshold, it is susceptible to the use of the Bonferroni inequality as above. More
precisely, we let $\chi(F_X)$ be the measure of "badness" of a random variable $X$, where $F_X$ is the cumulative distribution function of $X$. A tail risk measure associated with $\chi$ gives the number $c_\alpha$ such that the probability of an event in which $\chi(F)$ exceeds $c_\alpha$ is bounded by $\alpha$. Thus, if we have a valid $(1 - \alpha)$ confidence interval for this tail risk measure, we may use the upper bound of the confidence interval to control the tail risk measure with confidence level $(1 - 2\alpha)$.

We then show that two most popular risk measures, VaR and ES, have interpretations as tail risk measures.

Our method assumes that we have a valid (one-sided) confidence interval for the tail risk measure of interest. Many papers in the literature derive at least asymptotically valid inference methods in conjunction with estimation methods; in addition to aforementioned papers such as Chen (2008), Linton and Xiao (2013), and Hill (2015), Belomestny and Krätschmer (2012) derive asymptotic normality for plugin estimators of law-invariant coherent risk measures; Christoffersen and Gonçalves (2005) provide a resampling method that can correctly estimate the precision of VaR and ES estimates under mild conditions; Gao and Song (2008) establish asymptotic distribution of these estimates.

We demonstrate that our framework can be easily applied to portfolio optimization problems with tail risk measure constraints. For subadditive risk measures such as expected shortfall, one can employ our method to obtain the upper bound at the individual asset level and approximate the portfolio constraints by a linear combination of the upper bounds. The method does not require any knowledge on the joint distribution and therefore avoids the computational burden related to the estimation of the correlation structure. Using historical data, we show that our method does not suggest extremely conservative portfolio management—the upper bound estimates are only about 25% larger than the point estimates. We deduce that a moderately conservative portfolio choice helps one control the true tail risk.

This paper is organized as follows. Section 3.2 introduces the notion of a tail

---

2Note that there are two distinct concepts referred to as "confidence levels." One is associated with the concept of the risk measure (as in Value-at-Risk with confidence level $1 - \alpha$), and the other with the statistical inference of the risk measure.

3VaR is also known to be subadditive in some special cases Ibragimov (2005).
risk measure and explains the concept using examples. Section 3.3 shows that the Bonferroni argument works for tail risk measures, providing a method to control the true probabilities. Section 3.4 illustrates the use of our method in investment decision making, using the stock return data obtained from Yahoo! Finance.

3.2 Tail Risk Measures

This section defines the tail risk measure and introduces its examples. We show that the popular risk measures, Value-at-Risk and expected shortfall, are examples of the tail risk measure.

We first introduce our mathematical setup. Let \((\Omega, \mathcal{F}, P)\) be a probability space that is rich enough to contain all random variables and events to be introduced below. Let \(X : \Omega \to \mathbb{R}\) be a random variable representing the return of a portfolio. Denote by \(F_X\) and \(Q_X\) the distribution function and quantile function of \(X\). We set \(U\) to be the uniformly distributed random variable on \((0, 1)\) such that \(Q_X(U) = X\) almost surely. For a random variable \(Y : \Omega \to \mathbb{R}\), denote by \(\sigma(Y) \subset \mathcal{F}\) the \(\sigma\)-algebra generated by \(Y\). For event \(E \in \mathcal{F}\), denote by \(F_{X,E}\) the conditional distribution function of \(X\) conditional on the occurrence of event \(E\).

The tail risk measure is a quantity such that the probability that some characteristic of the portfolio return \(X\) exceeding it is bounded by some pre-specified number. The characteristic is given by a function \(\chi\) that maps the return distribution \(F_X\) to a real number. The bound on the probability is denoted by \(\alpha\) and we call \(1 - \alpha\) the confidence level.

**Definition 3.1** For \(\alpha \in (0, 1)\) and a map \(X_E \mapsto \chi(X_E) \in \mathbb{R}\), the tail risk measure of confidence level \(1 - \alpha\) is the infimum of \(c_\alpha\) that satisfies

\[
\sup_{E \in \sigma_0} \{ \Pr(\omega \in E) : \chi(F_{X,E}) \geq c_\alpha \} \leq \alpha
\]

for some fixed \(\sigma_0 \subset \sigma(U)\). If the supremum is attained at \(E = \emptyset\), the infimum of \(c_\alpha\) is defined to be \(\infty\).
Intuitively, the tail risk measure controls the probability of the occurrence of some bad event. The supremum with respect to $E$ represents the search for the worst event over a set of events $\sigma_0$; $\chi$ represents the measure of badness of the return $F_{X,E}$ given the event $E$; $c_\alpha$ gives the threshold above which $\chi$ will not exceed with probability $1 - \alpha$.

As a regulator of the banking system, one may want to bound the probability of financial crisis at some level. As a portfolio manager, one may want to minimize the probability of the portfolio incurring a significant loss. This notion of bounding the probability is the key idea that the tail risk measure captures.

Note that the function $\chi$ does not depend on $E$ or $\alpha$, and the supremum is taken over a subset of the $\sigma$-algebra generated by $U$, not $\sigma(X)$ nor any other subset of $\mathcal{F}$. In general, risk measures are defined as functions that map random variables to real numbers; dependence on $E$ (the $\sigma$-algebra) gives too much freedom to qualify as a risk measure. Also, we want $\alpha$ to represent some probability of a fixed concept, so not allowing $\chi$ to depend on $\alpha$ is also natural. The third point will be revisited at the end of this section.

First, we show that risk measures popularly used in practice are tail risk measures. In particular, we explain how Value-at-Risk, expected shortfall, and some distortion risk measures can be written as tail risk measures. Then, we motivate the above definition by explaining why variance can not (and should not) be interpreted as a tail risk measure.

**Example 1 (Value-at-Risk)** For $X : \Omega \to \mathbb{R}$, *Value-at-Risk of confidence level* $1 - \alpha$ is defined as

$$\text{VaR}_\alpha(X) := \inf\{x \in \mathbb{R} : \Pr(X < -x) \leq \alpha\}.$$ 

In words, Value-at-Risk gives the upper bound on the loss of the portfolio that occurs with probability of at least $1 - \alpha$, or equivalently, the worst loss one can expect with probability $1 - \alpha$ or higher.
This quantity can be cast as the infimum of $c_\alpha$ that satisfies
\[
\sup_{E \in \sigma(U)} \left\{ \Pr(\omega \in E) : -\sup \text{supp}(F_{X,E}) \geq c_\alpha \right\} \leq \alpha,
\]
where $\text{supp}(F_{X,E})$ denotes the support of distribution $F_{X,E}$. Therefore, Value-at-Risk at $1 - \alpha$ is a tail risk measure of confidence level $1 - \alpha$, holding $\sigma_0 = \sigma(U)$ and $\chi(F_{X,E}) = -\sup \text{supp}(F_{X,E})$.

**Example 2 (Expected shortfall)** Value-at-Risk has some undesirable features as a risk measure; it ignores the fatness of the tail loss and does not satisfy the so-called *coherence*. In response to these problems, a new risk measure is proposed and gradually adapted in practice (Acerbi and Tasche (2002)).

For $X : \Omega \rightarrow \mathbb{R}$, expected shortfall of confidence level $1 - \alpha$ is defined in Acerbi et al. (2008) as
\[
ES_\alpha(X) := \frac{1}{\alpha} \int_0^\alpha \text{VaR}_\gamma(X)d\gamma.
\]
Note that Value-at-Risk is (the negative of) the $\alpha$-quantile of $X$. Thus, for continuous returns, expected shortfall is equal to $\mathbb{E}[-X \mid X \leq -\text{VaR}_\alpha]$, the *expected* loss in the worst $\alpha$ events. Expected shortfall is known to be coherent (Acerbi and Tasche (2002)).

Expected shortfall can be cast as the infimum of $c_\alpha$ that satisfies
\[
\sup_{E \in \sigma(U)} \left\{ \Pr(\omega \in E) : \mathbb{E}[-X \mid \omega \in E] \geq c_\alpha \right\} \leq \alpha.
\]
Therefore, expected shortfall at $1 - \alpha$ is a tail risk measure of confidence level $1 - \alpha$, holding $\sigma_0 = \sigma(U)$ and $\chi(F_{X,E}) = \mathbb{E}[-X \mid E] = \int -xdF_{X,E}$.

**Remark 3.1** If the distribution of $X$ is discontinuous at $\alpha$, i.e., there does not exist $c$ that solves $\Pr(X \leq c) = \alpha$, then the event $U \in (0, \alpha)$ is not in $\sigma(X)$. For this reason, we use $\sigma(U)$ instead of $\sigma(X)$ in the definition.

**Example 3 (Tail Value-at-Risk)** A quantity that is closely related to expected shortfall is tail Value-at-Risk. For $X : \Omega \rightarrow \mathbb{R}$, tail Value-at-Risk of confidence level
$1 - \alpha$ is given by

$$\text{TVaR}_\alpha(X) := \mathbb{E}[X | -X \geq \text{VaR}_\alpha(X)].$$

This coincides with expected shortfall if the distribution of $X$ is continuous at $\text{VaR}_\alpha(X)$. Tail Value-at-Risk is known not to be coherent (Acerbi (2004)).

Tail Value-at-Risk can be cast as the infimum of $c_\alpha$ that satisfies

$$\sup_{E \in \sigma(X)} \left\{ \Pr(\omega \in E) : \mathbb{E}[X | \omega \in E] \geq c_\alpha \right\} \leq \alpha.$$

Therefore, tail Value-at-Risk at $1 - \alpha$ is a tail risk measure of confidence level $1 - \alpha$, holding $\sigma_0 = \sigma(X) \subset \sigma(U)$ and $\chi(F_{X,E}) = \int -x dF_{X,E}$.

Example 4 (Distortion risk measure) Let $K : [0, 1] \rightarrow [0, 1]$ be increasing, $K(0) = 0$, and $K(1) = 1$. The distortion risk measure for the distortion function $K$ is defined as

$$\rho_K(X) := \int_0^1 \text{VaR}_\gamma(X) dK(\gamma) = \int_0^1 -Q_X dK.$$

In this paper we consider distortion functions that satisfy an additional assumption that there exists $\kappa \in (0, 1)$ such that $K(u) < 1$ for $u < \kappa$ and $K(u) = 1$ for $u \geq \kappa$. Value-at-Risk is a distortion risk measure with $K(u) = 1\{u \geq \alpha\}$; expected shortfall is a distortion risk measure with $K(u) = \max\{u/\alpha, 1\}$. Therefore, both are distortion risk measures that are considered in this paper. With this assumption, we may write $\rho_K(X)$ as the infimum of $c_\alpha$ such that $\alpha = \kappa$ and

$$\sup_{E \in \sigma(U)} \left\{ \Pr(\omega \in E) : \int_0^1 -Q_{X,E}(u) dK\left(\frac{u}{\kappa}\right) \geq c_\alpha \right\} \leq \alpha,$$

where $Q_{X,E}$ denotes the generalized inverse function of $F_{X,E}$.

To motivate our definition of the supremum taken only over a subset of the $\sigma$-algebra generated by $U$, it is useful to look at examples of risk measures in the literature that are not tail risk measures.

Example 5 (Variance) For $X : \Omega \rightarrow \mathbb{R}$ with a second moment, variance is defined
by

\[ \text{Var}(X) := \mathbb{E}[(X - \mathbb{E}[X])^2]. \]

Variance is not a feature of \( X \) conditional on some event, but a feature that captures how disperse the entire distribution is; for example, variance penalizes positive and negative returns equally. As variance depends on the entirety of the distribution, it cannot be represented as a quantity of \( X \) conditional on an event that restricts a part of the distribution; hence, it is not a tail risk measure.

On the other hand, if we allow the supremum to be taken over a more general sub-\( \sigma \)-algebra, then one can represent variance in a similar form. Assume that \( \mathcal{F} \) is rich enough so it contains some random variable \( V \) that is independent of \( U \). Then, variance of \( X \) is the infimum of \( c_\alpha \) that satisfies

\[
\sup_{E \in \sigma(V)} \{ \Pr(\omega \in E) : \mathbb{E}[(X_E - \mathbb{E}[X_E])^2] \geq c_\alpha \} \leq \alpha
\]

for arbitrary \( \alpha \in [0, 1] \), since \( \text{Var}(X | V) = \text{Var}(X) \).

However, it is not sensible to regard a decision maker who aims to control variance as one who wants to control the conditional variance on any events at confidence level 0. Put differently, the event that attains the supremum is the one in which \( \Pr(V \in V(E)) = \alpha \) and hence the conditional distribution of \( X \) is kept equal to its marginal. Therefore, considering such an event as the “worst” case goes astray from what it aims to mean.

**Example 6 (Entropic Value-at-Risk)** For \( X : \Omega \to \mathbb{R} \) that has a finite moment-generating function \( M_X(z) := \mathbb{E}[e^{zX}] \) for non-positive values \( z \leq 0 \), Ahmadi-Javid (2012) considers *entropic Value-at-Risk* with confidence level \( 1 - \alpha \) by

\[
\text{EVaR}_{\alpha}(X) := \inf_{z < 0} \frac{\log(M_X(z)/\alpha)}{-z}.
\]

Since \( M_X \) depends on the entirety of the distribution of \( X \), it is not a tail risk measure.

Similarly as the previous example, however, if we allow a random variable \( V \) in-

---

\(^4\)Note that we measure the risk of the negative tail of \( X \) while Ahmadi-Javid (2012) measures that of the positive tail, hence the sign differences.
dependent of \( U \), entropic Value-at-Risk can be written as the infimum of \( c_\alpha \) that satisfies
\[
\sup_{E \in \sigma(V)} \left\{ \Pr(\omega \in E) : \inf_{z < 0} \frac{\log(M_{X_E}(z))/\alpha}{-z} \geq c_\alpha \right\}.
\]
However, again, this does not capture what we want tail risk measures to capture.

### 3.3 Controlling Tail Risk Measures

This section provides a simple rule to control tail risk measures with observable quantities. The problem of traditional approach is that the true risk is unobservable, and upon decision making, the estimated value has been used without taking into account the estimation error. We show that tail risk measures allow a very simple rule based on the Bonferroni inequality to control the true risk.

The intuition is best explained by Value-at-Risk. The Value-at-Risk of confidence level \( 1 - \alpha \) satisfies
\[
\Pr(X \leq -\text{VaR}_\alpha) \leq \alpha.
\]
However, if one replaces \( \text{VaR}_\alpha \) with an estimated quantity \( \tilde{\text{VaR}}_\alpha \), the inequality no longer holds. Suppose that one can construct a one-sided \( (1 - \alpha) \)-confidence interval such that
\[
\Pr(\text{VaR} \in (-\infty, \tilde{\text{VaR}}_\alpha]) \geq 1 - \alpha.
\]
Then, by the Bonferroni inequality,
\[
\Pr(X \leq -\tilde{\text{VaR}}_\alpha) \leq \Pr(X \leq -\text{VaR}) + \Pr(-\text{VaR} \leq -\tilde{\text{VaR}}) = 2\alpha.
\]
Thus, we may control Value-at-Risk of confidence level \( 1 - 2\alpha \) with \( \tilde{\text{VaR}} \).

The following theorem shows that the same logic holds for tail risk measures defined in the previous section. If one wants to control a tail risk measure of confidence level \( 1 - \beta \), then one may pick \( \alpha > 0 \) and \( s > 0 \) such that \( \alpha + s \leq \beta \), and compute the upper bound of a one-sided \( (1 - s) \)-confidence interval of an estimator for the tail risk measure of confidence level \( 1 - \alpha \). This allows us to control the true risk with the
knowledge of a valid confidence interval, taking into account the estimation error.

**Theorem 3.1** Let $c_\alpha$ be the true tail risk measure of confidence level $1 - \alpha$, and $\bar{c}_{\alpha,s}$ the $(1 - s)$-confidence bound of $c_\alpha$, that is, $\Pr(\bar{c}_{\alpha,s} \leq c_\alpha) \leq s$. Then for any $\alpha, s \in (0, 1)$, we have

$$\sup_{E \in \sigma_0} \Pr(\omega \in E \cap \chi(F_{X,E}) \geq \bar{c}_{\alpha,s}) \leq \alpha + s.$$

In other words, controlling $\bar{c}_{\alpha,s}$ controls $c_{\alpha+s}$ ex ante.

**Proof:** Observe that

$$\sup_{E \in \sigma_0} \Pr(\omega \in E \cap \chi(F_{X,E}) \geq \bar{c}_{\alpha,s}) \leq \sup_{E \in \sigma_0} \{\Pr(\omega \in E) : \chi(F_{X,E}) \geq c_\alpha\} + s \leq \alpha + s.$$

The first inequality follows from the Bonferroni inequality and the second from the property of $c_\alpha$.

**Remark 3.2** Although for any fixed choice of $\alpha$ and $s$ such that $\alpha + s \leq \beta$ the theorem is valid, one cannot "hunt" for the pair $(\alpha, s)$ that gives the lowest upper bound $\bar{c}_{\alpha,s}$. This will change the distribution of $\bar{c}_{\alpha,s}$ (due to randomness introduced by the confidence level hunting), and the probability bound may no longer be valid.

One key assumption of our method is the availability of a valid (one-sided) confidence interval for the tail risk measure we aim to control. To this end, we utilize results on properties of risk measure estimators examined in: Christoffersen and Gonçalves (2005), Gao and Song (2008), Chen (2008), Linton and Xiao (2013), Hill (2015), Belomestny and Krätschmer (2012).

### 3.4 Empirical Application to Expected Shortfall

An investor is considering to invest in three securities, namely, the stock of the Bank of America Corp., the stock of Morgan Stanley, and the index fund of the Dow Jones Industrial Averages. She wants to control the expected shortfall of level 90% of her
portfolio below some threshold $c$. Since expected shortfall is a coherent risk measure, it is subadditive, meaning that the expected shortfall of the sum of two portfolios does not exceed the sum of expected shortfalls of each portfolio. Therefore, she only needs to calculate expected shortfall for each security before she makes an investment decision, and does not need to recalculate the expected shortfall of the composite portfolio every time.

The investor is to control the "true" expected shortfall of her final portfolio by the combination of the estimated expected shortfalls and their confidence intervals. Our statistical objective is thus to estimate the expected shortfalls of level $1 - \tau_1$ of the three securities and their joint confidence intervals of level $1 - \tau_2$ so that $\tau_1 + \tau_2 \leq 0.1 = 100\% - 90\%$. We choose $\tau_1 = \tau_2 = 0.05$. Denoting $Y_t$ the daily price, we compute the daily return by

$$X_t := \frac{Y_t - Y_{t-1}}{Y_{t-1}}.$$

We use two popular methods to determine the point estimates and their confidence intervals. The first method is *Historical Simulation* (HS), which is a non-parametric estimation method that treats the returns as independent and identically distributed. Under this scheme, we take the lower 5\% average of the returns to estimate expected shortfall. Although HS is simple and robust to the underlying distributional assumptions, it does not take into account the serial dependence structure of asset returns. Moreover, forecasts with HS does not reflect the market conditions at the time when the forecasts are made (see Barone-Adesi et al. (1999)). To account for such shortcomings, we employ another method now widely called *Filtered Historical Simulation* (FHS) by Barone-Adesi et al. (1999). Unlike HS, FHS accommodates many stylized facts of financial return series such as volatility clustering. Since time-varying volatilities are estimated, forecasts under FHS reflect the then market conditions.

Following Hansen and Lunde (2005), Brownlees et al. (2012), and Barone-Adesi et al. (1999), we model daily stock returns as GARCH(1,1) process. Formally, daily
returns \{r_t\} are modeled according to:

\[ r_t = \mu + \epsilon_t = \mu + \sigma_t z_t \quad (3.1) \]

\[ \sigma_t^2 = \omega + \alpha \epsilon_{t-1}^2 + \beta \sigma_{t-1}^2, \quad (3.2) \]

where \{z_t\} are i.i.d. standard normal random variables. To ensure non-negativity of the variance process, parameters are restricted to \(\omega, \alpha, \beta \geq 0\). For stationarity, it is further assumed that \(\alpha + \beta < 1\).

While the original method of Barone-Adesi et al. (1999) can be used to determine point estimates of VaR and ES measures under a GARCH structure, if does not take into account the estimation error in parameter estimates. Since it is critical for us to form a valid (one-sided) confidence intervals of tail risk measures, we follow the method suggested in Christoffersen and Gonçalves (2005), which improves the one in Barone-Adesi et al. (1999). For concreteness, we restate their procedure.

**Step 1.** Use a sample of historical returns \(\{r_t\}_{t=1}^T\) to estimate \((\hat{\mu}, \hat{\omega}, \hat{\alpha}, \hat{\beta})\) in (3.1) and (3.2) using QMLE. Construct the standardized residuals \(\hat{z}_t = \frac{r_t - \hat{\mu}}{\hat{\sigma}_t}\).

**Step 2.** From \(\{\hat{z}_t\}_{t=1}^T\), sample with replacement \(\{z_t^*\}_{t=1}^T\). Use these bootstrapped residuals to reconstruct a bootstrap return sample \(\{r_t^*\}_{t=1}^T\).

**Step 3.** Use the sample \(\{r_t^*\}_{t=1}^T\), to re-estimate GARCH(1,1) in (3.1) and (3.2), and obtain \((\hat{\mu}^*, \hat{\omega}^*, \hat{\alpha}^*, \hat{\beta}^*)\).

**Step 4.** With \((\hat{\mu}^*, \hat{\omega}^*, \hat{\alpha}^*, \hat{\beta}^*)\) and the historical returns \(\{r_t\}_{t=1}^T\), determine the one-period-ahead volatility \(\hat{\sigma}_{T+1}^*\) by applying (3.1) and (3.2) iteratively.

**Step 5:** Use the bootstrapped standardized residuals \(\hat{z}_t^* = \frac{r_t^* - \hat{\mu}^*}{\hat{\sigma}_t^*}\) to estimate expected shortfall of residuals and from (3.1) derive the bootstrap estimate of expected shortfall.

**Step 6:** Repeat Step 2-5 for a large number of simulations.

We use the daily returns from January 4, 2016 to December 31, 2017, consisting of 502 business days in total. The data are retrieved from Yahoo! Finance.
<table>
<thead>
<tr>
<th></th>
<th>Estimate (%)</th>
<th>Marginal CB (%)</th>
<th>Joint CB (%)</th>
<th>Conservative CB (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>HS</td>
<td>Bank of America Corp.</td>
<td>3.96</td>
<td>4.65</td>
<td>4.85</td>
</tr>
<tr>
<td></td>
<td>Morgan Stanley</td>
<td>3.98</td>
<td>4.75</td>
<td>4.95</td>
</tr>
<tr>
<td></td>
<td>Dow Jones Industrial Average</td>
<td>1.55</td>
<td>1.81</td>
<td>1.89</td>
</tr>
<tr>
<td>FHS</td>
<td>Bank of America Corp.</td>
<td>3.12</td>
<td>3.60</td>
<td>3.75</td>
</tr>
<tr>
<td></td>
<td>Morgan Stanley</td>
<td>2.55</td>
<td>3.04</td>
<td>3.19</td>
</tr>
<tr>
<td></td>
<td>Dow Jones Industrial Average</td>
<td>1.10</td>
<td>1.27</td>
<td>1.32</td>
</tr>
</tbody>
</table>

Table 3.1: To control ES at 90%, we estimate ES of the daily stock returns at 95% from Jan 4, 2016 to Dec 31, 2017, and the upper bounds of the one-sided confidence bounds (CB) of level 95%. The CIs are based on bootstrap with 10,000 simulations.

Table 3.1 gives the estimates of the expected shortfalls of level 95% of the three securities, together with the 95% confidence bounds. Here, the confidence bound means $M$ such that the ray $(-\infty, M]$ contains the true parameter with the specified probability. The values represent the negative daily returns in the worst 5% cases in percentage points; for instance, under HS, the expected return of the Bank of America Corp. stock is $-3.96\%$ in a day in the worst 5% cases. The marginal confidence bound is the bound below which each estimate would ex ante fall individually for the 95% of all times, while the joint confidence bound is the bound of the hyperray below which the entire vector of estimates would ex ante fall for the 95% of all times. Needless to say, there are many ways to construct a joint confidence set, e.g., by a hyperellipse or hypercube. In this example we use a hyperray since we want the smallest lower bound on each axis. The joint bounds are taken proportional to the marginal ones, so the joint one is always greater than the marginal one. It can be seen that the inflation of the bounds due to the joint coverage requirement is not very big. This tight joint confidence bounds are made possible because we have the joint distribution. If we assume no knowledge on the joint distribution, we might simply split the probability equally among assets, 5%/3 in this case, to ensure the joint coverage even in the worst case; we call it conservative confidence bounds.\(^5\) These confidence bounds are calculated by the bootstrap with 10,000 iterations.

With these estimates, the investor can control her expected shortfall, for its subad-

\(^5\)This joint coverage is ensured by the Bonferroni inequality. In this paper, we avoid the use of “Bonferroni confidence bounds,” lest it incurs unnecessary confusion.
ditivity, by meeting the following criterion McNeil et al. (2005, p. 240): for example, using FHS,

$$3.75\alpha + 3.19\beta + 1.32(1 - \alpha - \beta) \leq c$$

for $0 \leq \alpha \leq 1$ and $0 \leq \beta \leq 1 - \alpha$, where $\alpha$ and $\beta$ denote the shares of the first two securities in her portfolio. For example, she can maximize the expected return of her portfolio subject to this constraint. To extend the setup to allow for short positions, observe that the expected shortfall of a short position is the highest 100$r\%$ expected return of each security. Thus, simply add the new security $-Y_i$ to the set of securities and carry out the same exercise.

### 3.5 Conclusion

In this paper, we addressed the issue of estimation errors in financial risk control. When we use estimated risk measures in the practice of risk management—be it portfolio optimization or banking regulation—we incur additional uncertainty due to estimation errors. Observing that many risk measures are designed to control the probabilities of bad events, we proposed a method to control the “true but unobservable” risk probability under the assumption that a valid confidence interval is available.

Let $\chi(F)$ be a quantity that measures “badness” associated with the return distribution $F$. We defined a class of risk measures, called *tail risk measures*, which bound the maximum probabilities of events that entail large values of $\chi$. For example, if $\chi(F)$ is the negative of the expectation of $F$, ES can be given as a quantity that bounds the probability with which the negative expectation of the return exceeds ES. We showed, most notably, that VaR and ES, arguably the two most popular risk measures in the present, are given as tail risk measures. We also gave examples of risk measures, variance and EVaR, that are not considered as tail risk measures.

Next, we established a method to control the true risk probability using the Bonferroni inequality. The idea is to construct a valid (one-sided) confidence interval of
the risk measure, and use the upper bound of the confidence interval as the risk estimate. Because of the properties of the risk measure and the confidence interval, we showed that this upper bound manages to control the true probability that the risk measure was designed to control. The key to the proof was the Bonferroni inequality.

Our empirical application showed how our method can be used to control ES of a portfolio that consists of three assets. We used both historical simulations and filtered historical simulation to compute ES of daily stock returns of the Bank of America Corp., Morgan Stanley, and Dow Jones Industrial Average, and applied the bootstrap method of Christoffersen and Gonçalves (2005) to construct confidence intervals on ES.

In addition to the risk control literature, this paper contributes to the work on microfounding the multiplier that is used in modern banking regulations (Stahl (1997) and Leippold and Vanini (2002)). Practical risk control suffers from estimation error and model risk. The previous literature gave justification on the use of multiplier 3 on VaR control from the perspective of accounting for model risk. This paper provides new insights on how much conservativeness is required in order to guard against possible estimation error.
Bibliography


3.6 Appendix

The defining feature of VaR is that the probability of the loss of return variable $X$ exceeding $\text{VaR}$, $\Pr(X < -\text{VaR}_\alpha)$, is bounded by, often equal to, $\alpha$. We call this probability the risk probability. Once we substitute $\text{VaR}_\alpha$ with its estimator $\text{VaR}_\alpha$, however, we do not know whether $\Pr(X < -\text{VaR}_\alpha)$ is bounded by $\alpha$.

This section carries out simulation studies to illustrate how the substituted risk probability compares to $\alpha$. Our simulation consists of two iterations. In the first iteration, we draw $n$ return observations $X_1, \ldots, X_n$ and estimate $\text{VaR}_\alpha$. For this estimate, there is a corresponding probability by which the new draw falls below it, i.e., $\Pr(X_{n+1} < -\text{VaR}_\alpha | \text{VaR}_\alpha)$. Since $\text{VaR}_\alpha$ may realize at any value, this probability can itself be larger or smaller than $\alpha$. In the next iteration, we repeat this exercise for $S$ times and compute the unconditional probability by which the new draw falls below the estimator, i.e., $\Pr(X_{n+1} < -\text{VaR}_\alpha)$. For simplicity, we hereafter drop the subscript $n + 1$. Note that by the law of iterated expectations,

$$\Pr(X < -\text{VaR}_\alpha) = \mathbb{E}[\mathbb{E}[\mathbb{1}\{X < -\text{VaR}_\alpha\} | \text{VaR}_\alpha]] = \mathbb{E}[\Pr(X < -\text{VaR}_\alpha | \text{VaR}_\alpha)].$$

This probability naturally depends on the estimation method we employ for VaR. Thus, we take three methods and compute the unconditional probability for each. We will see, somewhat surprisingly, that even when an estimation method does a great job in estimating VaR, it does not necessarily make the unconditional probability close to $\alpha$.

To make clear that the issue is not a mere byproduct of the intrinsically complicated financial data structure, we keep the simulation setup as simple as possible. First, we assume that the return variables $X_1, \ldots, X_n$ follow independent standard normal distribution, where $X_{(1)} \leq \cdots \leq X_{(n)}$ denote their order statistics. To reflect the small sample nature of financial data (Caccioli et al. (2017)), we let $n = 100$ and $\alpha = 0.01$. To accurately assess unconditional probabilities, we carry out $S = 10,000$ iterations of VaR estimation. Denoting by $\Phi$ the cdf of the standard normal distribution, the true VaR is $-\Phi^{-1}(0.01) \approx 2.33$. 
The following three methods are employed to estimate VaR: maximum likelihood estimation (MLE), a method of Weissman (1978), and historical simulation. Roughly speaking, these three methods correspond to parametric, semi-parametric, and non-parametric methods, respectively. All of them are "correctly specified;" in MLE, we maximize the independent normal likelihood; in Weissman, we assume that the return distribution is in the domain of attraction of Gumbel distributions, which indeed is so for normal returns; in historical simulation, we take the empirical quantile, which is a valid estimator for i.i.d. random variables.

MLE maximizes the normal likelihood of the data. In particular, we first estimate the two parameters of the normal distribution by 

\[ \hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} X_i \quad \text{and} \quad \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \hat{\mu})^2, \]

and then estimate VaR by

\[ \hat{\text{VaR}}_{\alpha,1} := -\hat{\mu} - \hat{\sigma} \Phi^{-1}(0.01). \]

Figure 3-1a shows the histogram of VaR estimated by MLE around the true VaR represented by the red line. Being a correctly specified parametric model, MLE estimates VaR more accurately than other methods, compared to histograms in Figure 3-1c and 3-1e. Mean squared error (MSE) of MLE estimator is 0.0378, the smallest of the three. However, in Figure 3-1b we see the sign of risk probability distortion. It shows the histogram of conditional risk probabilities \( \Pr(X < -\hat{\text{VaR}}_{\alpha,1} \mid \hat{\text{VaR}}_{\alpha,1}) \), together with the unconditional risk probability depicted by the orange line. The unconditional probability 1.21% is much larger than \( \alpha = 1\% \). This means that, by repeatedly using MLE estimates, the chances that the loss exceeds the estimate are larger than intended. That is, we underestimate the risk.

Weissman’s estimator uses the smallest \( k \) observations to estimate a tail quantile, relying on extreme value theory. We set \( k = n/10 = 10 \) and estimate VaR by\(^6\)

\[ \hat{\text{VaR}}_{\alpha,2} := -\left( \frac{1}{k} \sum_{i=1}^{k} X_{(i)} - X_{(k)} \right) \log \left( \frac{k}{\alpha n} \right) - X_{(k)}. \]

Figure 3-1c gives the histogram of Weissman’s VaR estimates. The performance of

\[^6\text{Weissman (1978) proposes two estimators: one based on MLE and the other on minimum variance unbiased estimation (MVUE). We use the first one.}\]
this method (as an estimator of VaR) is in the middle of the three methods, giving MSE of 0.0943. However, we see in Figure 3-1d that there is a large upward distortion of the risk probability entailed by this method; Pr(\(X < -\widehat{\text{VaR}_{\alpha,2}}\)) = 0.0143 > Pr(\(X < -\text{VaR}_\alpha\)) = 0.01. Again, the expected frequency that the loss exceeds Weissman’s VaR is larger than the targeted probability, which results in underestimation of risk.

Historical simulation in this context is simply implemented by the empirical \(\alpha\)-quantile of historical data. Thus, the estimate is the maximum loss of the historical return, i.e.,

\[
\widehat{\text{VaR}}_{\alpha,3} := -X_{(1)}.
\]

In Figure 3-1e we see that the histogram of historical estimates is more dispersed than the previous estimates, with MSE equal to 0.2197. Figure 3-1f is the histogram of conditional risk probabilities for historical simulation. Despite the VaR estimates being “worse” than the first two methods, the unconditional risk probability 0.0099 is very close to the intended probability 0.01. However, we cannot take this result at face value since the simulation assumes a very restrictive case of i.i.d. returns and it is known in the literature that historical simulation does not usually work well with financial time series data (Pritsker (2006)).

Finally, the source of distortion may be either the bias or variance of the estimator, or both of them. de Haan et al. (2016) discuss bias that comes from applying extreme value theory. It is noteworthy to acknowledge that our method works whenever we have a valid confidence interval, regardless of the original estimator suffering from large bias or large variance. Moreover, our method does not require that an estimator is available; if we can obtain the (one-sided) confidence interval of the true VaR, we can control the true risk probability applying the results of this paper.
Figure 3-1: Simulations of $n = 100$ normal returns with $S = 10,000$ iterations illustrate distortion of risk probability, $\Pr(X < -\overline{VaR}) \neq \alpha$. Draws outside the range are shown as clusters at the boundaries.