Financial Portfolio Risk Management: Model Risk, Robustness and Rebalancing Error

Xingbo Xu

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
in the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2013

ABSTRACT

Financial Portfolio Risk Management:

Model Risk, Robustness and Rebalancing Error

Xingbo Xu

Risk management has always been in key component of portfolio management. While more and more complicated models are proposed and implemented as research advances, they all inevitably rely on imperfect assumptions and estimates. This dissertation aims to investigate the gap between complicated theoretical modelling and practice. We mainly focus on two directions: model risk and reblancing error.

In the first part of the thesis, we develop a framework for quantifying the impact of model error and for measuring and minimizing risk in a way that is robust to model error. This robust approach starts from a baseline model and finds the worst-case error in risk measurement that would be incurred through a deviation from the baseline model, given a precise constraint on the plausibility of the deviation. Using relative entropy to constrain model distance leads to an explicit characterization of worst-case model errors; this characterization lends itself to Monte Carlo simulation, allowing straightforward calculation of bounds on model error with very little computational effort beyond that required to evaluate performance under the baseline nominal model. This approach goes well beyond the effect of errors in parameter estimates to consider errors in the underlying stochastic assumptions of the model and to characterize the greatest vulnerabilities to error in a model. We apply this approach to problems of portfolio risk measurement, credit risk, delta hedging, and counterparty risk measured through credit valuation adjustment.

In the second part, we apply this robust approach to a dynamic portfolio control problem. The sources of model error include the evolution of market factors and the influence of these factors on asset returns. We analyze both finite- and infinite-horizon problems in a model in which returns are driven by factors that evolve stochastically. The model incorporates transaction costs and leads to simple and tractable optimal robust controls for multiple assets. We illustrate the performance of the controls on historical data. Robustness does improve performance in out-of-sample tests in which the model is estimated on a rolling window of data and then applied over a subsequent time period. By acknowledging uncertainty in the estimated model, the robust rules lead to less aggressive trading and are less sensitive to sharp moves in underlying prices.

In the last part, we analyze the error between a discretely rebalanced portfolio and its continuously rebalanced counterpart in the presence of jumps or mean-reversion in the underlying asset dynamics. With discrete rebalancing, the portfolio's composition is restored to a set of fixed target weights at discrete intervals; with continuous rebalancing, the target weights are maintained at all times. We examine the difference between the two portfolios as the number of discrete rebalancing dates increases. We derive the limiting variance of the relative error between the two portfolios for both the mean-reverting and jump-diffusion cases. For both cases, we derive "volatility adjustments" to improve the approximation of the discretely rebalanced portfolio by the continuously rebalanced portfolio, based on on the limiting covariance between the relative rebalancing error and the level of the continuously rebalanced portfolio. These results are based on strong approximation results for jump-diffusion processes.

Contents

List of Figures

List of Tables

Acknowledgements

First and foremost, I am deeply grateful to my advisor Professor Paul Glasserman. It is my honor to work with him, and my first debt of gratitude must go to him for his continuous guidance and encouragement. Without his support and guidance, I would not have accomplished this.

Special thanks to my committee, Professor Mark Broadie and Professor Xuedong He, Professor Garud Iyengar and Professor Steven Kou, for their support, guidance and helpful suggestions. I have attended many of their classes and seminars. Their guidance has served me well and I owe them my heartfelt appreciation.

My sincere thanks to many Ph.D fellows: Dr. Zhiwei (Tony) Qin, Haowen Zhong for being there all these five years; Dr. Shiqian Ma, Dr. Xianhua Peng, Dr. Zongjian Liu, Dr. Yunan Liu, Dr. Xiaowei Liu, Dr. Guodong Pang for their invaluable advice from the day I joined IEOR; Junyi Zhang, Dr. Pengfei Zang, Xinyun Chen, Yan Liu, Chun Wang, Yupeng Chen, Juna Li, Yunru Han, Hua Zheng, and many others for all the cheerful moments. Of course, I shall never forget my wonderful officemates in MUDD 313A: Rodrigo Carrasco, Tulia Humphr, Matthieu Plumettaz, Arseniy Kukanov, Dr. Yixi Shi, Dr. A. Cecilia Zenteno, Andrew Ahn and Jinbeom Kim.

Last but not the least, I am deeply indebted to my parents, Qineng Xu and Qingju Chen. It is them who make my accomplishment meaningful. I would like to give my special thank to my wife, Yina Lu, who is the greatest gift of my life. Without her love and support throughout these years, I could have not overcome the hard times.

To my family:

Father Qineng Xu

Mother Qingju Chen

Wife Yina Lu

Chapter 1

Introduction

Risk measurement has always been a vital component of financial portfolio management. Research in this area using advanced mathematical models has experienced rapid development over the past two decades. However, simplified assumptions used in traditional academic research models are usually inadequate to characterize the complicated practical situations. The implementation of these simplified models could lead to significant unexpected risks in financial activities. This dissertation aims to bridge this gap between the existing academic research and practice from the following two directions:

- Model risk, referring to the unexpected uncertainty in model assumption and estimation;
- Portfolio rebalancing error, referring to the discrepancy between theoretical continuously rebalanced portfolio and discretely rebalanced one.

1.1 Robust Risk Measurement and Model Risk

Risk measurement relies on modeling assumptions. Errors in these assumptions introduce errors in risk measurement. This makes risk measurement vulnerable to *model risk*, which refers to the uncertainty of the model being used.

In practice, model risk is sometimes addressed by comparing the results of different models — see Morini (2011) for an extensive treatment of this idea with applications to many different markets. An alternative approach to model uncertainty is to mix multiple models. This idea is developed from a Bayesian perspective in, for example, Draper (1995) and Raftery et al. (1997) and applied to portfolio selection in Pesaran et al. (2009). More often, if it is considered at all, model risk is investigated by varying model parameters. Importantly, we consider model errors that go beyond parameter sensitivity to consider the effect of changes in the probability law that defines an underlying model. This includes those model errors that are not reflected in parameter perturbations. For example, the main source of model risk might result from an error in a joint distribution of returns that cannot be described through a change in a covariance matrix.

In Chapter 2, we develop tools for quantifying model risk and making risk measurement robust to modeling errors. Our goals are as follows:

- to measure model error that is more general than parameter uncertainty, given a baseline nominal model;
- to identify the sources of model error to which a measure of risk is most vulnerable and to identify which changes in the underlying model have the greatest impact; and
- to propose decisions that are robust in terms of model error.

To work with model errors described by changes in probability laws, we need a way to quantify such changes, and for this we use *relative entropy*. Relative entropy offers a non-parametric way to describe the difference between two probability distributions. In Bayesian statistics, the relative entropy between posterior and prior distributions measures the information gained through additional data. In characterizing model error, we interpret relative entropy as a measure of the additional information required to make a perturbed model preferable to a baseline model. Thus, relative entropy becomes a measure of the plausibility of an alternative model. Indeed, relative entropy has been applied for model calibration and estimation in numerous sources, including Avellaneda (1998), Avellaneda et al. (2001, 1997), Buchen and Kelly (1996), Cont and Deguest (2010), Cont and Tankov (2004, 2006), Gulko (1999, 2002), and Basurto and Goodhart (2009). In working with heavy-tailed distributions, for which relative entropy may be undefined, we use a related notion of α-divergence, as do Dey and Juneja (2010) in a portfolio selection problem.

With the relative entropy, we are able to identify the worst-case scenario of a particular model when the model risk is bounded by some pre-determined level. It is a convenient choice because the worst-case alternative within a relative entropy constraint is typically given by an exponential change of measure. Monte Carlo simulation combines conveniently with our approach to identify the details of the worst-case scenario. At the same time that we simulate a nominal model and estimate a nominal risk measure, we can estimate a bound or bounds on model risk with virtually no additional computational effort: we simply multiply the nominal risk measure on each path by a factor (a likelihood ratio or Radon-Nikodym derivative) that captures the worst-case change of probability measure. To simulate under the worst-case model is again straightforward because simulating under the original model and then multiplying any output by the worst-case likelihood

ratio is equivalent to simulating the output from the worst-case model. This is similar to importance sampling, except that the usual goal of importance sampling is to reduce estimation variance without changing the mean of the estimated quantity; here, the objective is to understand how the change in probability measure changes the means and other model properties. This simulation-based approach also allows us to limit which stochastic inputs to a model are subject to model error.

Our method can potentially be generalized to a much higher level as follows. If we can find a measurement or premetric, whose worst-case likelihood ratio can be derived easily, then most of our analysis can be carried out without much difficulty. A possible choice of such measurement or premetric has the form $E[\phi(m)]$, where the function $\phi(m) \geq 0$ for any likelihood ratio m. Relative entropy and $\alpha-$divergence are special cases.

We further develop tools following this approach that are robust in a sense similar to the way the term is used in the optimization and control literature. Robust optimization seeks to optimize against worst-case errors in problem data — see Ben-Tal et al. (2000), Bertsimas and Pachamanova (2008), and Goldfarb and Iyengar (2003), for example. The errors in problem data considered in this setting are generally limited to uncertainty about parameters, though distributional robustness is considered in, e.g., Nilim and El Ghaoui (2005) and Natarajan et al. (2008). Our approach builds on the robust control ideas developed in Hansen and Sargent (2007), Hansen et al. (2006), and Petersen et al. (2000), and applied to dynamic portfolio selection in Glasserman and Xu (2013). In this line of work, it is useful to imagine an adversary that changes the probability law in the model dynamics; the robust control objective is to optimize performance against the worst-case change of probability imposed by the adversary. Similarly, here we may imagine an adversary

changing the probability law of the inputs to a risk calculation; we want to describe this worst-case change in law and quantify its potential impact on risk measurement. In both settings, the degree of robustness is determined through either a constraint or a penalty on relative entropy that limits the adversary's ability to make the worst case arbitrarily bad.

In the second half of Chapter 2, we examine model errors in models for mean-variance portfolio optimization, conditional value-at-risk, Gaussian copula of portfolio credit risk, delta hedging error and credit valuation adjustment. Using simulation-based tools, we are able to identify the worst-case scenario to model error with relative entropy constraints. The results show that our approach based on stochastic robustness goes well beyond parameter sensitivity in exploring model error.

1.2 Application in Dynamic Portfolio Control

In Chapter 3, the robust approach introduced in Chapter 2 is applied to a dynamic portfolio control problem in details.

Classic mean-variance portfolio optimization (see Markowitz (1952)), like most optimization problems that rely on estimated quantities, is vulnerable to the error-amplifying effects of combining optimization with estimation. Any reasonable estimation procedure applied to multiple assets will overestimate the expected returns of some assets or underestimate their risk, and an optimization procedure that ignores this fact will drive a portfolio to overinvest in precisely these assets. Dynamic portfolio control introduces a further complication by requiring a model of the evolution of asset prices. Any practical model is likely to be misspecified, in addition to being subject to estimation error, and optimization will again amplify the effects of error, in this case model error.

The approach of robustness in parameters has been developed extensively in the portfolio optimization. For example, Ben-Tal et al. (2000), Bertsimas and Pachamanova (2008), Goldfarb and Iyengar (2003) focus on parameter uncertainty, and, Lim et al. (2011) approaches it from a different perspective in . Robustness to uncertainty over a set of distributions in portfolio optimization is analyzed in, e.g., Nilim and El Ghaoui (2005), Natarajan et al. (2008), Natarajan et al. (2010), Delage and Ye (2010) and Goh and Sim (2010), primarily in single-period formulations.

Here we use a stochastic notion of robustness introduced in Chapter 2 that allows model uncertainty in the law of evolution of the stochastic inputs to a model. This robustness approach is combined with the following features: we study multi-period (finite- and infinite-horizon) portfolio control problems; returns are driven by factors that evolve stochastically; transaction costs are incorporated. We assume both the relationship between returns and factors and the evolution of the factors are subject to model error and thus treated robustly. Simple optimal controls are developed that remain tractable for multiple assets. Performance is demonstrated through both in-sample and out-of-sample on historical data.

For the factor model and factor dynamics, we start from the (non-robust) model of Garleanu and Pedersen (2012). Their model uses linear dynamics and a quadratic objective to achieve tractability with considerable flexibility and generality that lends itself to further study. Their analysis, motivated by realistic trading strategies, focuses on the impact of the speed of mean-reversion in factor dynamics and how this affects portfolio control and, ultimately, equilibrium asset prices. By building on their framework, we retain a high degree of tractability, and we can study the effect of robustness in a current and independent model, rather than in a model introduced specifically for the comparison. As a byproduct, we can also see the effect of model uncertainty on factor dy-

namics and the factor model of returns: the adversary in the robust formulation can perturb both, and the adversary's optimal choice points to the ways in which the investor is most vulnerable to model error. We test our portfolio rules on the same commodity futures as Garleanu and Pedersen (2012). Briefly, we find that robustness leads to better performance in out-of-sample tests in which the model is re-estimated on a rolling window; robust rules guard against model and estimation error by trading less aggressively on signals from the factors.

1.3 Porfolio Relabancing Error

The analysis of a portfolio's dynamics is often simplified by assuming that the constituent assets can be traded continuously. For a trading strategy defined by portfolio weights, meaning the fraction of the portfolio held in each asset, continuous trading leads to an idealized model in which the actual weights match the target weights at each instant. For highly liquid stocks bought and sold on electronic exchanges, continuous trading is often a close approximation of reality. But for many other asset classes the practical reality of discrete trading cannot be entirely ignored. A portfolio manager may not be able to maintain an ideal set of portfolio weights continuously in time; transactions costs and liquidity constraints may limit the portfolio manager to rebalancing the portfolio to target weights at discrete intervals.

We analyze the error in approximating a discretely rebalanced portfolio with one that is continuously rebalanced and thus more convenient to model. Our focus is on the effect of jumps and mean reversion in the dynamics of the underlying assets. For both features, we examine the limiting difference between the continuous and discrete portfolios as the rebalancing frequency increases.

Our main results are as follows. With either mean reversion or jumps, we derive the limiting variance of the relative error between the two portfolios. With mean reversion and no jumps, we show that the limiting error, scaled by the square root of the number of rebalancing dates, is asymptotically normal and independent of the level of the continuously rebalanced portfolio; moreover, the limiting distribution is identical to the one achieved without mean reversion. In the presence of jumps, we show that the scaled relative error converges to the sum of a normal random variable and a compound Poisson random variable. For both the mean-reverting and jump-diffusion cases, we derive "volatility adjustments" to improve the approximation of the discretely rebalanced portfolio by the continuously rebalanced portfolio. These adjustments are based on the limiting covariance between the relative rebalancing error and the level of the continuously rebalanced portfolio.

The simpler case in which the underlying assets are modeled as a multivariate geometric Brownian motion is analyzed in Glasserman (2012). The analysis there is motivated by the *incremental risk charge* (IRC) introduced by the Basel Committee on Banking Supervision (2007, 2009). The IRC is intended to capture the effect of potential illiquidity of assets in a bank's trading portfolio. It models illiquidity by imposing a fixed rebalancing frequency for each asset class: some bonds, for example, might have a liquidity interval of two weeks, and tranches of asset backed securities might have liquidity intervals of a month or even a quarter. The IRC is thus based on the difference between discrete and continuous rebalancing.

The possibility of jumps in asset prices is clearly relevant to portfolio risk and to the modeling of less liquid assets. One would also expect jumps to have a qualitatively different effect on rebalancing error than pure diffusion — adding jumps should cause the discretely rebalanced portfolio to stray farther from the target weights — and this is confirmed in our results. The potential impact

of mean reversion is less evident: one might expect mean reversion to offset part of the effect of discrete rebalancing if it helps restore a portfolio's weights to their targets. We will see that this is the case, but only for the volatility adjustment that comes from the covariance between the rebalancing error and the portfolio level. The distribution of the relative rebalancing error itself is, in the limit, unaffected by the presence of mean reversion.

Discretely rebalanced portfolios arise in models of transaction costs and discrete hedging, including Bertsimas et al. (2000), Boyle and Emanuel (1980), Duffie and Sun (1990), Leland (1985), and Morton and Pliska (1995). Sepp (2012) examines the asymptotic error of delta hedging with proportional transaction costs under a jump-diffusion model with lognormal jump sizes. Guasoni et al. (2011) analyze the effect of discrete rebalancing on the measurement of tracking error and portfolio alpha. In their analysis of leveraged ETFs, Avellaneda and Zhang (2010) examine the impact of discrete rebalancing and derive an asymptotic relation between the behavior of the fund and the underlying asset as the rebalancing frequency increases. Jessen (2010) studies the discretization error for CPPI portfolio strategies using simulation. Haugh (2011) studied constant proportion trading strategy, which is closely related to leveraged ETF. Although these applications do not fit precisely within the specifics of our setting, we nevertheless view our analysis as potentially relevant to extending work on these applications. In Glasserman and Xu (2010), we use a continuously rebalanced portfolio to design an importance sampling procedure to estimate the tail of a discretely rebalanced portfolio in a pure-diffusion setting, and the results we develop here suggest potential extensions to models with jumps.

The distribution of the difference between a diffusion process and its discrete-time approximation has received extensive study motivated by simulation methods, as in Kurtz and Protter (1991).

Jacod and Protter (1998) study this error for more general processes, including processes with jumps. Tankov and Voltchkova (2009) apply the results of Jacod and Protter (1998) to analyze the error in discrete delta-hedging, thus extending the results of Bertsimas et al. (2000) to models with jumps. In their analysis of discretization methods, Kloeden and Platen Kloeden and Platen (1992) develop strong approximation results for stochastic Taylor expansions; Bruti-Liberati and Platen (2005, 2007) derive corresponding expansions for jump-diffusion processes. These results provide very useful tools for our investigation of rebalancing error.

1.4 Outline

In Chapter 2, the definition of the robust risk measurement is proposed, and is further generalized to heavy-tail distribution. The approach is extended to incorporate constraints in the form of expectations, and to cope with the case where model uncertainty is restricted within certain sources. In the second half of the chapter, several models are examined to illustrate that the robust risk measurement captures more general model uncertainty that goes beyond parameter uncertainty.

In Chapter 3, the robustness approach is applied to a dynamic portfolio control problem with practical features. We find a closed-form value function iteration, and study the effect of robustness through empirical experiments. Stability results are also derived for both finite and infinite-horizon cases.

In Chapter 4, the limiting behavior of constant-weight portfolio rebalancing error is investigated. Closed-form limits of the errors are derived for the models with jump and mean-reversion effects. Using the results, volatility estimators of the discretely rebalanced portfolios are obtained.

In later half of the chapter, the main tool, strong approximation for stochastic differential equation, is provided, together with detailed proof of the main results.

Chapter 2

Robust Risk Measurement and Model Risk

2.1 Introduction

Risk measurement relies on modeling assumptions, which themselves can be a source of risk, called *model risk*. A variety of reasons contribute to model uncertainty. For example, misspecification among alternative models, insufficient or polluted data, sampling variability, imperfect models structure, or even the evolution of the market dynamics (or say the "true" model), can all lead to the inevitable result, i.e., model error.

In this chapter, a non-parametric robust risk measurement is introduced to measure potential model uncertainty. The measurement uses relative entropy to capture the discrepancy between the baseline model and the unknown "true" model. In Bayesian statistics, the relative entropy between posterior and prior distributions measures the information gained through additional data. In characterizing model error, we interpret relative entropy as a measure of the additional information

required to make a perturbed model preferable to a baseline model. Thus, relative entropy becomes a measure of the plausibility of an alternative model.

We illustrate through numerical examples that this robustness approach enables us to consider more general uncertainty compared with parameter uncertainty. This approach is incorporated with Monte Carlo simulation, which lends us the power to investigate the details of the worst possible model with given uncertainty level. We extend the robustness approach to account for constraints in the form of expectation, and to restrict the uncertainty within certain random source. For heavy-tailed distribution, a generalization of relative entropy, $\alpha-$divergence, is used.

The rest of this chapter is organized as follows. Section 2.2 provides an overview of our approach and develops the main supporting theoretical tools. In Section 2.3, we discuss the implementation of the approach through a set of techniques we call robust Monte Carlo. The remainder of this chapter is devoted to illustrative applications: Section 2.4 considers portfolio variance; Section 2.5 considers conditional value-at-risk; Section 2.6 examines the Gaussian copula model of portfolio credit risk; Section 2.7 investigates delta hedging, comparing the worst-case hedging error with various specific sources of model error; and Section 2.8 studies model risk in the dependence between exposures and default times in credit valuation adjustment (CVA).

2.2 Overview of the Approach

We begin by introducing the main ideas in a simple setting. Let X denote the stochastic elements of a model — this could be a scalar random variable, a random vector, or a stochastic process. Let $V(X)$ denote some measure of risk associated with the outcome X. We will introduce conditions

on V later, but for now we keep the discussion informal. If the law of X is correctly specified, then the expectation $E[V(X)]$ is the true value of the risk measure of interest.

We incorporate model uncertainty by acknowledging that the law of X may be misspecified. We consider alternative probability laws that are not too far from the nominal law in a sense quantified by relative entropy. For probability densities g and $\tilde{g}$ with a well-defined likelihood ratio $m = \tilde{g}/g$, we define the relative entropy of $\tilde{g}$ with respect to g to be

$$\mathcal{R}(g, \tilde{g}) = E[m \log m] = \int \frac{\tilde{g}(x)}{g(x)} \log \frac{\tilde{g}(x)}{g(x)} \, g(x) \, dx.$$

In Bayesian statistics, relative entropy measures the information gain in moving from a prior distribution to a posterior distribution. In our setting, it measures the additional information that would be needed to make an alternative model $\tilde{g}$ preferable to a nominal model g.

It is easy to see that $\mathcal{R} \geq 0$, and $\mathcal{R}(g, \tilde{g}) = 0$ only if $\tilde{g}$ and g coincide almost everywhere (with respect to g). Relative entropy is not symmetric in g and $\tilde{g}$ and does not define a distance in the usual sense, but $\mathcal{R}(g, \tilde{g})$ is nevertheless interpreted as a measure of how much the alternative $\tilde{g}$ deviates from g. (Our views of g and $\tilde{g}$ are generally not symmetric either: we favor the nominal model g but wish to consider the possibility that $\tilde{g}$ is correct.) The expression $E[m \log m]$, defining relative entropy through a likelihood ratio, is applicable on general probability spaces and is thus convenient. Indeed, we will usually refer to alternative models through the likelihood ratio that connects an alternative probability law to a nominal law, defining $\tilde{g}(x)$ to be $m(x)g(x)$. With the nominal model g fixed, we write $\mathcal{R}(m)$ instead of $\mathcal{R}(g, \tilde{g})$.

To quantify model risk, we consider alternative models described by a set $\mathcal{P}_\eta$ of likelihood

ratios m for which $E[m \log m] < \eta$. In other words, we consider alternatives within a relative entropy "distance" η of the original model. We then seek to evaluate, in addition to the nominal risk measure $E[V(X)]$, the bounds

$$\inf_{m \in \mathcal{P}_\eta} E[m(X)V(X)] \text{ and } \sup_{m \in \mathcal{P}_\eta} E[m(X)V(X)]. \tag{2.2.1}$$

The expression $E[m(X)V(X)]$ is the expectation under the alternative model defined by m. For example, in the scalar case $m = \tilde{g}/g$,

$$E[m(X)V(X)] = \int \frac{\tilde{g}(x)}{g(x)} V(x) g(x)\, dx = \int V(x)\tilde{g}(x)\, dx.$$

The bounds in (2.2.1) thus bound the range of possible values for the risk measure consistent with a degree of model error bounded by η.

The standard approach to the maximization problem in (2.2.1) is to form the dual problem

$$\inf_{\theta > 0} \sup_{m} E[mV(X) - \frac{1}{\theta}(m \log m - \eta)].$$

(We will often suppress the argument of m to simplify notation, as we have here.) For given $\theta > 0$, the inner supremum problem has as solution of the form

$$m_\theta^* = \frac{\exp(\theta V(X))}{E[\exp(\theta V(X))]}, \tag{2.2.2}$$

provided the expectation in the denominator is finite. In other words, the worst-case model error is

characterized by an exponential change of measure defined through the function V and a parameter $\theta > 0$. The lower bound in (2.2.1) is solved the same way but with $\theta < 0$. The explicit solution we get in (2.2.2) is the greatest advantage of working with relative entropy to quantify model error. In Section 2.3, we will apply (2.2.2) at multiple values of θ to trace out bounds at multiple levels of relative entropy.

2.2.1 A First Example: Portfolio Variance

To help fix ideas, we introduce a simple example. Let X denote a vector of asset returns and suppose, for simplicity, that X is modeled by a multivariate normal distribution $N(\mu, \Sigma)$, $\Sigma > 0$, on $\mathbb{R}^n$. We consider a portfolio with weights $a = (a_1, \ldots, a_n)^\top$ summing to 1, and we use portfolio variance as our risk measure

$$E[V(X)] = E[a^\top (X - \mu)(X - \mu)^\top a].$$

We are interested in the worst-case variance

$$\sup_{m \in \mathcal{P}_\eta} E[mV(X)] = \sup_{m \in \mathcal{P}_\eta} E[ma^\top (X - \mu)(X - \mu)^\top a].$$

In formulating the problem this way, we are taking μ as known but otherwise allowing an arbitrary change in distribution, subject to the relative entropy budget of η.

From (2.2.2), we know that the worst-case change of measure has the form

$$m_\theta^* \propto \exp\left(\theta\left[a^\top (X-\mu)(X-\mu)^\top a\right]\right).$$

We find the worst-case density of X by multiplying the original $N(\mu, \Sigma)$ density by the likelihood ratio; the result is a density proportional to

$$\exp\left(\theta\left[a^\top (x-\mu)(x-\mu)^\top a\right]\right) \times \exp\left(-\frac{1}{2}(x-\mu)^\top \Sigma^{-1}(x-\mu)\right).$$

In other words, the worst-case density is itself multivariate normal $N(\mu, \tilde{\Sigma})$,

$$\tilde{\Sigma} = (\Sigma^{-1} - 2\theta a a^\top)^{-1},$$

with $\theta > 0$ sufficiently small that the matrix inverse exists. For small θ,

$$\tilde{\Sigma} = \Sigma + 2\theta \Sigma a a^\top \Sigma + o(\theta^2),$$

and the worst-case portfolio variance becomes

$$\begin{aligned} a^\top \tilde{\Sigma} a &= a^\top \Sigma a + 2\theta a^\top \Sigma a a^\top \Sigma a + o(\theta^2) \\ &= a^\top \Sigma a + 2\theta \left(a^\top \Sigma a\right)^2 + o(\theta^2). \end{aligned}$$

That is, the resulting worst-case variance of the portfolio is increased by approximately 2θ times the square of the original variance.

This simple examples illustrates ideas that recur throughout this chapter. We are interested in finding the worst-case error in the risk measure — here given by portfolio variance — but we are just as interested in understanding the change in the probability law that produces the worst-case change. In this example, the worst-case change in law turns out to stay within the family of multivariate normal distributions: we did not impose this as a constraint; it was a result of the optimization. So, in this example, the worst-case change in law reduces to a parametric change — a change in Σ. In this respect, this example is atypical, and, indeed, we will repeatedly stress that the approach to robustness we use goes beyond merely examining the effect of parameter changes to gauge the impact of far more general types of model error.

The worst-case change in distribution we found in this example depends on the portfolio vector a. Here and throughout, it is convenient to interpret model error as the work of a malicious adversary. The adversary perturbs our original model, but the error introduced by the adversary is not arbitrary — it is tailored to have the most severe impact possible, subject to a relative entropy budget constraint. The bounds in (2.2.1) measure the greatest error the adversary can introduce, subject to this constraint.

The portfolio variance example generalizes to any quadratic function $V(x) = x^\top A_q x$, $A_q > 0$. A similar calculation shows that under the worst-case change of measure X remains normally distributed with

$$X \sim N(\tilde{\mu}, \tilde{\Sigma}), \quad \tilde{\Sigma} = (\Sigma^{-1} - 2\theta A_q)^{-1}, \quad \tilde{\mu} = \tilde{\Sigma}\Sigma^{-1}\mu.$$

The relative entropy associated with this change of measure evaluates to

$$\eta(\theta) = \frac{1}{2}\left(\log(\det(\Sigma\tilde{\Sigma}^{-1}) + \mathrm{tr}(\Sigma^{-1}\tilde{\Sigma} - I) + (\mu - \tilde{\mu})^{\top}\Sigma^{-1}(\mu - \tilde{\mu})\right).$$

By inverting the mapping $\theta \mapsto \eta(\theta)$, we can find the worst-case θ associated with any relative entropy budget η. In most of our examples, it is easier to evaluate model error at various values of θ and calculate the corresponding value for relative entropy, rather than to specify the level of relative entropy in advance; we return to this point in Section 2.3.

2.2.2 Optimization Problems and Precise Conditions

As the portfolio variance example illustrates, risk measurement often takes place in the context of an investment or related decision. We therefore extend the basic problem of robust evaluation of $E[V(X)]$ to optimization problems of the form

$$\inf_{a\in\mathcal{A}} E[V_a(X)], \tag{2.2.3}$$

for some parameter a ranging over a parameter set $\mathcal{A}$. For example, a could be a vector of portfolio weights or a parameter of a hedging strategy. We will introduce conditions on V_a and the law of X.

Assumption 2.2.1. *For the minimization problem (2.2.3)*

1. *The decision parameter set $\mathcal{A}$ is compact, $V_a(x)$ is convex in a for any x. Thus,* $\inf_a E[V_a(X)] < \infty$.

2. *For all $a \in \mathcal{A}$, the moment generating function $\Psi_g(\theta, a) = E[\exp(\theta V_a(X))]$ exists for θ in some open set containing the origin. If $P(V_a(X) > 0) > 0$, then $\Psi_g(\theta, a) \uparrow \infty$ as $\theta \uparrow \theta_{\max}(a)$, where $\theta_{\max}(a) := \sup\{\theta : \Psi_g(\theta, a) < \infty\}$; if $P(V_a(X) < 0) > 0$, then $\Psi_g(\theta, a) \uparrow \infty$ as $\theta \downarrow \theta_{\min}(a)$ where $\theta_{\min}(a) := \inf\{\theta : \Psi_g(\theta, a) < \infty\}$.*

Part (1) of the assumption ensures feasibility of the optimization problem. (For a maximization problem, we would require that $V_a(x)$ be concave in a.) Part (2) ensures the finiteness of $\Psi_g(\theta, a)$ and its derivative, so that the corresponding exponential change of measure is well-defined. We denote by $(\theta_{\min}(a), \theta_{\max}(a))$ the interval (possibly infinite) in which $\Psi_g(\theta, a)$ is finite and thus an exponential change of measure defined by $\exp(\theta V_a(X))$ is well-defined.

We formulate a robust version of the optimization problem (2.2.3) as

$$\inf_a \sup_{m \in \mathcal{P}_\eta} E[mV_a(X)]. \tag{2.2.4}$$

Here, we seek to optimize against the worst-case model error imposed by a hypothetical adversary. The dual to the inner maximization problem is

$$\inf_a \inf_{\theta > 0} \sup_m E[mV_a(X) - \frac{1}{\theta}(m \log m - \eta)]. \tag{2.2.5}$$

For any $\theta > 0$ and decision parameter a, if part (2) of Assumption 2.2.1 is satisfied, the optimal change of measure for the adversary is described by the likelihood ratio

$$m^*_{\theta,a} = \exp(\theta V_a(X))/E[\exp(\theta V_a(X))], \tag{2.2.6}$$

where we need $\theta \in (0, \theta_{\max}(a))$. By substituting (2.2.6) into (2.2.5), we get

$$\inf_a \inf_{\theta>0} \frac{1}{\theta} \log E[\exp(\theta V_a(X))] + \frac{\eta}{\theta}. \tag{2.2.7}$$

If $\theta_{\max}(a) < \infty$, then as $\theta \uparrow \theta_{\max}(a)$, the objective function in (2.2.7) goes to infinity, so the infimum over θ will automatically make the optimal θ smaller than $\theta_{\max}$. That is, we can safely consider $\theta < \infty$ instead of $\theta \in (0, \theta_{\max}(a))$. This allows us to change the order of $\inf_a$ and $\inf_\theta$ in (2.2.5), whereas $\theta_{\max}(a)$ depends on the decision a. Now we can relax the constraints for θ in both (2.2.5) and (2.2.7) to $\theta > 0$. Assumption 2.2.1 is relevant to the $\inf_a$ and $\inf_\theta$ ordered as (2.2.7). To swap the order, we need the following assumption.

Assumption 2.2.2. *1. If (θ^*, a^*, m^*) solve (2.2.5), then $\theta^* \in [0, \theta^*_{\max})$ for some $\theta^*_{\max} \in [0, \infty]$ such that for any $\theta \in [0, \theta^*_{\max})$, the set $\{a \in \mathcal{A} : E[\exp(\theta V_a(X))] < \infty\}$ is compact.*

2. For $\theta \in [0, \theta^*_{\max})$, $E[\exp(\theta V_a(X))]$ is lower semi-continuous in a.

Proposition 2.2.3. *Under Assumptions 2.2.1–2.2.2, problem (2.2.5) is equivalent to*

$$\inf_{\theta>0} \inf_a \sup_m E[mV_a(X) - \frac{1}{\theta}(m \log m - \eta)]. \tag{2.2.8}$$

For fixed $\theta \in (0, \theta^*_{\max})$, the corresponding optimal objective function of inner $\inf_a \sup_m$ in (2.2.8) becomes

$$\begin{aligned} H(\theta) + \frac{\eta}{\theta} &:= \inf_a \sup_m E[mV_a(X) - \frac{1}{\theta} m \log m + \frac{\eta}{\theta}] \\ &= \frac{1}{\theta} \log E[\exp(\theta V_{a^*(\theta)}(X))] + \frac{\eta}{\theta}, \end{aligned} \tag{2.2.9}$$

where the optimal decision is

$$a^*(\theta) = \arg\inf_a \frac{1}{\theta} \log E[\exp(\theta V_a(X))], \tag{2.2.10}$$

and the worst-case change of measure is

$$m_\theta^* = \exp(\theta V_{a^*(\theta)}(X))/E[\exp(\theta V_{a^*(\theta)}(X))]. \tag{2.2.11}$$

Because $E[\exp(\theta V_a(X))]$ is not necessarily continuous in a, the lower semi-contintuity condition in Assumption 2.2.2 is needed to guarantees that the infimum in (2.2.10) can be attained.

For a fixed value of a,

$$\lim_{\theta \to 0+} \frac{1}{\theta} \log E[\exp(\theta V_a(X))] = E[V_a(X)],$$

corresponding to the nominal case without model uncertainty. To avoid too much technical complication, we only consider a simple case. When $\log E[\exp(\theta V_a(X))]/\theta$ is continuous both in a and , we can define the optimal decision and objective function when θ approaches 0 as follows:

$$\begin{aligned} a^*(0) &= \lim_{\theta \to 0+} \arg\inf_a \frac{1}{\theta} \log E[\exp(\theta V_a(X))] = \arg\inf_a E[V_a(X)], \\ H(0) &= \lim_{\theta \to 0+} \frac{1}{\theta} \log E[\exp(\theta V_{a^*(0)}(X))] = E[V_{a^*(0)}(X)]. \end{aligned}$$

The constrained problem (2.2.4) is equivalent to

$$\inf_{\theta>0} H(\theta) + \frac{\eta}{\theta}. \tag{2.2.12}$$

As a consequence of (Petersen et al. 2000, Theorem 3.1), when the set of $\theta > 0$ leading to finite $H(\theta)$, is non-empty, (2.2.12) has a solution $\theta > 0$ and the optimal value and solution solve the original constraint problem (2.2.4).

Proposition 2.2.4. *With assumption 2.2.2, the objective function in (2.2.10) is convex in a.*

Proof. Because $V_a(x)$ is convex in a for any x, the objective function $E[V_a(X)]$ is convex in a. Because $\theta > 0$, the objective function in (2.2.10) is convex as well.

□

For given $\eta > 0$, we can find an optimal θ^*_η, with $m^*(\theta^*_\eta, a^*(\theta^*_\eta))$ and $a^*(\theta^*_\eta)$ as optimal solutions, and

$$\eta = E[m^*(\theta^*_\eta, a^*(\theta^*_\eta)) \log m^*(\theta^*_\eta, a^*(\theta^*_\eta))], \tag{2.2.13}$$

i.e., the uncertainty upper bound is reached at the optimal perturbation. So with $\theta^*_\eta > 0$, and the adversary's optimal choice as (2.2.11), the original constraint problem (2.2.4) has the optimal objective

$$E[m^*(\theta^*_\eta, a^*(\theta^*_\eta))V_{a^*(\theta^*_\eta)}(X)] = \frac{E[V_{a^*(\theta^*_\eta)}(X)\exp(\theta^*_\eta V_{a^*(\theta^*_\eta)}(X))]}{E[\exp(\theta^*_\eta V_{a^*(\theta^*_\eta)}(X))]}, \tag{2.2.14}$$

which differs from the objective function of the penalty version (2.2.9) through the constant term.

In practice, we may be interested in seeing the relation between the level of uncertainty and the worst-case error, which involves comparing different values of η. In this case, rather than repeat

the procedure above multiple times to solve (2.2.12), we can work directly with multiple values of $\theta > 0$ and evaluate $\eta(\theta)$ with each, as in (2.2.13). Working with a range of values of θ, this allows us to explore the relationship between η and the worst-case error (and this is the approach we use in our numerical examples). This method requires that η be an increasing function of θ, a property we have observed numerically in all of our examples.

2.2.3 Robustness with Heavy Tails: Extension to α-Divergence

In order to use relative entropy to describe model uncertainty, we need the tails of the distribution of $V(X)$ to be exponentially bounded, as in Assumption 2.2.1. To deal with heavy-tailed distribution, we can use an extension of relative entropy called α-divergence and defined as (see also Rényi (1961) and Tsallis (1988))

$$\mathcal{D}_\alpha(m) = \mathcal{D}_\alpha(g, \tilde{g}) = \frac{1 - \int \tilde{g}^\alpha(x) g^{1-\alpha}(x) dx}{\alpha(1-\alpha)} = \frac{1 - E[m^\alpha]}{\alpha(1-\alpha)},$$

with m the likelihood ratio $\tilde{g}/g$, as before, and the expectation on the right taken with respect to g. Relative entropy can be considered a special case of α-divergence, in the sense that $\mathcal{R}(m) = E[m \log m] = \lim_{\alpha \to 1+} D_\alpha(m)$.

With relative entropy replaced by α-divergence, the constraint problem (2.2.4) becomes

$$\inf_a \sup_{m:\mathcal{D}_\alpha(m)<\eta} E[mV_a(X)].$$

The corresponding penalty problem is

$$\begin{aligned} & \inf_a \inf_{\theta>0} \sup_m \quad E[mV_a(X) - \frac{1}{\theta}(\mathcal{D}_\alpha(m) - \eta)] \\ = \quad & \inf_{\theta>0} \inf_a \sup_m \quad E[mV_a(X) - \frac{1}{\theta}(\mathcal{D}_\alpha(m) - \eta)]. \end{aligned} \tag{2.2.15}$$

The supremum is taken over valid likelihood ratios — nonnegative random variables with mean 1. Dey and Juneja (2010) apply an equivalent polynomial divergence and minimize it subject to linear constraints through a duality argument. We use a similar approach. We need the following condition to ensure that the proposed likelihood ratio is non-negative.

Assumption 2.2.5. *For any a, $V_a(X) > 0$ almost surely under the nominal measure, and $E[V_a(X)^{\frac{\alpha}{\alpha-1}}] < \infty$.*

Proposition 2.2.6. *Suppose Assumption 2.2.5 holds. For any $a \in \mathcal{A}$, $\theta > 0$ and $\alpha > 1$, the pair $(m^*(\theta, \alpha, a), c(\theta, \alpha, a))$ that solves the following equations with probability 1 is an optimal solution to (2.2.15):*

$$m^*(\theta, \alpha, a) = (\theta(\alpha - 1)V_a(X) + c(\theta, \alpha, a))^{\frac{1}{\alpha-1}}, \tag{2.2.16}$$

$$\text{for some constant } c(\theta, \alpha, a), \text{ such that } \theta(\alpha - 1)V_a(X) + c(\theta, \alpha, a) \geq 0, \tag{2.2.17}$$

$$\text{and } E\left[(\theta(\alpha - 1)V_a(X) + c(\theta, \alpha, a))^{\frac{1}{\alpha-1}}\right] = 1. \tag{2.2.18}$$

Proof. The objective of (2.2.15) is concave in m. Proceeding as in (Dey and Juneja 2010, Proof of Theorem 2), we can construct a new likelihood ratio $(1 - t)m^* + tm$ using an arbitrary m; the

objective becomes

$$\begin{aligned} K(t) &:= E\left[((1-t)m^* + tm)V_a + \frac{1}{\theta\alpha(1-\alpha)}((1-t)m^* + tm)^\alpha\right] + \frac{\eta}{\theta}, \\ K'(0) &= E\left[\left\{V_a + \frac{1}{\theta(1-\alpha)}(m^*)^{\alpha-1}\right\}(m-m^*)\right]. \end{aligned} \tag{2.2.19}$$

In order to have $K'(0) = 0$ for any m, we need the term inside braces in (2.2.19) to be constant. By the definition of m^*, $K'(0) = 0$ holds, so m^* is optimal. □

If $V_a(X)$ is not bounded from below, then when $\theta > 0$ and $\alpha \geq 0$, (2.2.17) cannot be satisfied. For the case in which the adversary seeks to minimize the objective function (that is, to get the lower bound of the error interval), we need $\alpha < 0$ to satisfy (2.2.17).

A feasible likelihood ratio exists in a neighborhood of $\theta = 0$, by the following argument. In the nominal case $\theta = 0$, we have $m^*(0, \alpha, a) = c(0, a)^{\frac{1}{\alpha-1}}$, so we can always choose $c(0, \alpha, a) = 1$. By continuity, we can find a set $[0, \theta_0)$ such that $c(\theta, \alpha, a)$ satisfying (2.2.17) and (2.2.18) exists for any $\theta \in [0, \theta_0)$. Once $c(\theta, \alpha, a)$, (2.2.16) gives an optimal change of measure (not necessarily unique). The optimal decision becomes

$$a^*(\theta) = \arg\min_a \frac{\alpha-1}{\alpha} E[(\theta(\alpha-1)V_a(X) + c(\theta,\alpha,a))^{\frac{1}{\alpha-1}} V_a(X)] + \frac{c(\theta,\alpha,a)}{\theta\alpha(1-\alpha)}. \tag{2.2.20}$$

In contrast to the relative entropy case, it is not clear whether the objective in (2.2.20) is convex in a.

Measuring potential model error through α-divergence focuses uncertainty on the tail decay of the nominal probability density. For example, in the simple scalar case $V_a(x) = x^k$, taking $\alpha > 1$

leads to a worst-case density function

$$\tilde{g}_X(x) \approx c x^{k/(\alpha-1)} g_X(x), \tag{2.2.21}$$

for $x \gg 0$, where g_X is the density function of X under the nominal measure. Incorporating model uncertainty makes the tail heavier, asymptotically, by a factor of $x^{k/(\alpha-1)}$.

Remark 2.2.1. As illustrated using relative entropy and $\alpha-$divergence, our method can potentially be generalized to a much higher level as follows. If we can find a measurement or premetric, with which the worst-case likelihood ratio can be derived easily, then most of our analysis can be carried out without much difficulty. A possible choice of such measurement or premetric has the form $E[\phi(m)]$, where the function $\phi(m) \geq 0$ for any likelihood ratio m and is convex in m. Relative entropy and $\alpha-$divergence are special cases.

2.3 Implementation: Robust Monte Carlo

In this section, we present methods for estimating the model error bounds in practice through what we call *robust Monte Carlo*. In addition to calculating bounds, we present ways of examining the worst-case model perturbation to identify the greatest model vulnerabilities, and we also show how to constrain the possible sources of model error.

2.3.1 Estimating the Bounds on Model Error

We assume the ability to generate independent replications $X_1, X_2, \ldots$ of the stochastic input X, recalling that X may be a random variable, a random vector, or a path of a stochastic process. A standard Monte Carlo estimator of $E[V(X)]$ is

$$\frac{1}{N}\sum_{i=1}^{n} V(X_i).$$

For any fixed θ and likelihood ratio $m_\theta \propto \exp(\theta V(X))$, we can estimate the expectation of $V(X)$ under the change of measure defined by m_θ by generating the X_i from the original (nominal) measure and forming the estimator

$$\frac{\sum_{i=1}^{N} V(X_i)\exp(\theta V(X_i))}{\sum_{i=1}^{N}\exp(\theta V(X_i))}, \tag{2.3.1}$$

which converges to $E[m_\theta V(X)]$ as $N \to \infty$. Assuming $V \geq 0$, we have $E[V(X)] \leq E[m_\theta V(X)]$ if $\theta > 0$ and $E[V(X)] \geq E[m_\theta V(X)]$ if $\theta < 0$. Our estimator of these bounds requires virtually no additional computational effort beyond that required to estimate the nominal value $E[V(X)]$.

From the same replications $X_1, \ldots, X_N$, we can estimate the likelihood ratio by setting

$$\hat{m}_{\theta,i} = \frac{\exp(\theta V(X_i))}{\sum_{j=1}^{N}\exp(\theta V(X_j))/N}, \quad i = 1, \ldots, N.$$

This in turn allows us to estimate the relative entropy at θ as

$$\hat{\eta}(\theta) = \frac{1}{N}\sum_{i=1}^{N} \hat{m}_{\theta,i} \log \hat{m}_{\theta,i}. \tag{2.3.2}$$

Thus, we can easily estimate $(\eta(\theta), E[m_\theta V(X)])$ across multiple values of θ. Given a relative entropy budget η, we then lookup the smallest and largest values of $E[m_\theta V(X)]$ estimated with $\hat{\eta}(\theta) \leq \eta$ to get the model error bounds at that level of η. We will illustrate this procedure through several examples.

Just as importantly, we can use the same simulation to analyze and interpret the worst-case model error. We do this by estimating expectations $E[m_\theta h(X)]$ of auxiliary functions $h(X)$ under the change of measure by evaluating estimators of the form

$$\frac{1}{N}\sum_{i=1}^{N} \hat{m}_{\theta,i} h(X_i). \tag{2.3.3}$$

Through appropriate choice of h, this allows us to estimate probabilities, means, and variances of quantities of interest, for example, that provide insight into the effect of the worst-case change in probability law.

In some case, we may want to sample from the worst-case law, and not evaluate expectations under the change of measure. If V is bounded, we can achieve this through acceptance-rejection: to simulate under the law defined by θ, we generate candidates X from the original nominal law and accept them with probability $\exp(\theta V(X))/M$, with M chosen so that this ratio is between 0

and 1. If V is unbounded, we need to truncate it at some large value and the sampling procedure then incurs some bias as a result of the truncation.

These techniques extend to problems of optimization over a decision parameter a, introduced in Section 2.2.2, for which a standard estimator is

$$\min_a \frac{1}{N} \sum_{i=1}^{N} V_a(X_i),$$

For $\theta > 0$, a worst-case objective function estimator is

$$\frac{\sum_{i=1}^{N} V_{\hat{a}^*}(X_i) \exp(\theta V_{\hat{a}^*}(X_i))}{\sum_{i=1}^{N} \exp(\theta V_{\hat{a}^*}(X_i))}, \tag{2.3.4}$$

where the estimated optimal decision parameter is

$$\hat{a}_N^* = \arg\inf_a \frac{1}{\theta} \log \sum_{i=1}^{N} \frac{\exp(\theta V_a(X))}{N},$$

and the estimated optimal likelihood ratio is

$$\hat{m}_{\theta,i}^* = \frac{\exp(\theta V_{\hat{a}_N^*}(X_i))}{\sum_{j=1}^{N} \exp(\theta V_{\hat{a}_N^*}(X_j))/N}, \quad i = 1, \ldots, N.$$

By continuous mapping theorem, for given $\hat{a}_N^*$ and any $\theta \in [0, \theta_{\max})$ the averages of both numera-

tor and denominator of (2.3.4) are consistent estimators. That is,

$$\frac{1}{N}\sum_{i=1}^{N} V_{\hat{a}_N^*}(X_i)\exp(\theta V_{\hat{a}_N^*}(X_i)) \to E[V_{\hat{a}_N^*}(X)\exp(\theta V_{\hat{a}_N^*}(X))],$$
$$\frac{1}{N}\sum_{i=1}^{N} \exp(\theta V_{\hat{a}_N^*}(X_i)) \to E[\exp(\theta V_{\hat{a}_N^*}(X))].$$

Hence, (2.3.4) is a consistent estimator for (2.2.14) with $\hat{a}^*$.

In the case where $E[\exp(\theta V_a(X))]$ is continuous in a and the optimal decision a^* is unique, it is easy to show that $\hat{a}^*$ converges to a^* in distribution. More generalized results can be found in Sample Average Approximate literature, e.g., Shapiro et al. (2009).

Similar estimators are available in the α-divergence framework. For given $\theta > 0$, $\alpha > 1$ and a, we estimate the worst-case likelihood ratio as

$$\hat{m}^*_{\theta,\alpha,a,i} = (\theta(\alpha-1)V_a(X_i) + \hat{c}(\theta,\alpha,a))^{\frac{1}{\alpha-1}},$$

$$\text{for some constant } \hat{c}(\theta,a), \text{ s.t. } \theta(\alpha-1)V_a(X_i) + \hat{c}(\theta,\alpha,a) > 0, \text{ for each } i$$

$$\text{with } \frac{1}{N}\sum_{i=1}^{N}\left[(\theta(\alpha-1)V_a(X_i) + \hat{c}(\theta,\alpha,a))^{\frac{1}{\alpha-1}}\right] = 1.$$

For given $\theta > 0$ and $\alpha > 1$, we solve for an optimal a as

$$\hat{a}^*(\theta) \;=\; \arg\min_a \frac{\alpha-1}{\alpha}\sum_{i=1}^{N}[(\theta(\alpha-1)V_a(X_i) + \hat{c}(\theta,\alpha,a))^{\frac{1}{\alpha-1}} V_a(X_i)] + \frac{\hat{c}(\theta,\alpha,a)}{\theta\alpha(1-\alpha)}.$$

The robust estimator for the objective becomes

$$\frac{1}{N}\sum_{i=1}^{N} V_a(X_i)\hat{m}_{\theta,\alpha,a^*(\theta),i}.$$

2.3.2 Incorporating Expectation Constraints

When additional information is available about the "true" model, we can use it to constrain the worst-case change of measure. Suppose the information available takes the form of constraints on certain expectations. For example, we may want to constrain the mean (or some higher moment) of some variable of a model. We formulate this generically through constraints of the form $E[mh_i(X)] \leq \eta_i$ or $E[mh_i(X)] = \eta_i$ for some function h_i and scalars η_i.

Such constraints can be imposed as part of an iterative evaluation of model risk. In (2.3.3), we showed how a change of measure selected by an adversary can be analyzed through its implications for auxiliary functions. If we find that the change of measure attaches an implausible value to the expectation of some $h_i(X)$, we can further constrain the adversary not just through the relative entropy constraint but through additional constraints on these expectations. This helps ensure the plausibility of the estimated model error and implicitly steers the adversary to allocate the relative entropy budget to other sources of model uncertainty. The adversary's problem becomes

$$\sup_{m\in\mathcal{P}_{\mathcal{M}}} E[mV(X)], \tag{2.3.5}$$

where $\mathcal{P}_{\mathcal{M}} = \{m : \mathcal{R}(m) \leq \eta, E[mh_i(X)] \leq \eta_i, i = 1, ..., n_M\}$ for some $\eta_i, \eta \in [0, \infty)$.

Here we have added n_M constraints on the expectations of $h_i(X)$ under the new measure.

We can move the constraints into the objective with Lagrange multipliers θ_i and transform (2.3.5) into a penalty problem; the argument in Petersen et al. (2000) still holds as the terms of $h_i(X)$ can be combined with that of V:

$$\inf_{\theta>0,\theta_i>0} \sup_m E\left[mV(X) - \frac{1}{\theta}(m\log m - \eta) - \sum_{i=1}^{n_M} \frac{mh_i(X) - \eta_i}{\theta_i}\right].$$

When θ and the θ_i are fixed, the problem can be treated as before in (2.2.8).

Proposition 2.3.1. *For fixed $\theta > 0$ and $\theta_i > 0$, $i = 1, ..., n_M$, such that*

$$E\left[\exp\left(\theta\left[V(X) - \sum_{i=1}^{n_M} \frac{h_i(X)}{\theta_i}\right]\right)\right] < \infty.$$

The worst change of measure is

$$m_\theta^* \propto \exp\left(\theta\left[V(X) - \sum_{i=1}^{n_M} \frac{h_i(X)}{\theta_i}\right]\right).$$

The optimization over (θ, θ_i) becomes

$$\inf_{\theta>0,\theta_i>0} \frac{1}{\theta}\log E\left[\exp\left(\theta\left[V(X) - \sum_{i=1}^{n_M} \frac{h_i(X)}{\theta_i}\right]\right)\right] + \frac{\eta}{\theta} + \sum_{i=1}^{n_M} \frac{\eta_i}{\theta_i}.$$

For equality constraints, the optimization is over $\theta_i \in \mathbb{R}$.

This is a standard result on constraints in exponential families of probability measures. It is used in Avellaneda et al. (2001) and Cont and Tankov (2004), for example, where the constraints calibrate a base model to market prices. Glasserman and Yu (2005) and Szechtman and Glynn

(2001) analyze the convergence of Monte Carlo estimators in which constraints are imposed by applying weights to the replications.

For an optimization problem as in (2.2.3), adding constraints entails solving another layer of optimization. For example, if the original problem is a minimization problem as in (2.2.3), then for given (θ, θ_i), the optimal decision becomes

$$a^*(\theta, \theta_i) = \arg\inf_a \frac{1}{\theta} \log E\left[\exp\left(\theta\left[V_a(X) - \sum_{i=1}^{n_M} \frac{h_i(a,X)}{\theta_i}\right]\right)\right] + \sum_{i=1}^{n_M} \frac{\eta_i}{\theta_i}.$$

2.3.3 Restricting Sources of Model Uncertainty

In some cases, we want to go beyond imposing constraints on expectations to leave entire distributions unchanged by concerns about model error. We can use this device to focus robustness on parts of the model of particular concern. We will see a different application in Section 2.8 where we use an exponential random variable to define a default time in a model with a stochastic default intensity. In that setting, we want to allow the default intensity to be subject to model uncertainty, but we want to leave the exponential clock unchanged as part of the definition of the default time.

Suppose, then, that the stochastic input has a representation as (X, Y), for a pair of random variables or vectors X and Y. We want to introduce robustness to model error in the law of X, but we have no uncertainty about the law of Y. For a given $\theta > 0$, we require that

$E[\exp(\theta V_a(X,Y))|Y=y] < \infty$ for any y, and formulate the penalty problem

$$\inf_a \sup_m \quad E[m(X,Y)V_a(X,Y) - \frac{1}{\theta}(m(X,Y)\log m(X,Y) - \eta)] \tag{2.3.6}$$

$$\text{s.t.} \quad E[m(X,Y)|Y=y] = 1,\ \forall y \tag{2.3.7}$$

$$m(x,y) \geq 0\, \forall x, y.$$

We have written $m(X,Y)$ to emphasize that the likelihood ratio may be a function of both inputs even if we want to leave the law of Y unchanged.

Proposition 2.3.2. *For problem (2.3.6) with $\theta > 0$ and $E[\exp(\theta V_a(X,Y))|Y=y] < \infty$ for all y:*

1. *Any likelihood ratio that satisfies (2.3.7) preserves the law of Y.*

2. *For any a, the likelihood ratio*

$$m^*(x,y) = \frac{\exp(\theta V_a(x,y))}{E[\exp(\theta V_a(X,Y))|Y=y]}, \tag{2.3.8}$$

   is an optimal solution to the maximization part of problem (2.3.6).

3. *The corresponding optimal decision becomes*

$$a^*(\theta) = \arg\inf_a \frac{1}{\theta} E\left[\log E[\exp(\theta V_a(X,Y))|Y]\right].$$

Proof. The feasible set of likelihood ratios m is convex, and the objective function is concave in

m, so it suffices to check first-order conditions for optimality. Define

$$\bar{K}(t) = E\left[(tm^* + (1-t)m)\, V_a(X,Y) - \frac{1}{\theta}((tm^* + (1-t)m)\log(tm^* + (1-t)m) - \eta)\right],$$

where m is an arbitrary likelihood ratio satisfying (2.3.7). Obviously, m^* satisfies (2.3.7). Taking the derivative of $\bar{K}$ at zero and substituting for m^*, we get

$$\begin{aligned}
\bar{K}'(0) &= E\left[\left(V_a(X,Y) - \frac{1}{\theta}\log m^* - \frac{1}{\theta}\right)(m^* - m)\right] \\
&= E\left[E\left[\left(V_a(X,Y) - \frac{1}{\theta}\log m^* - \frac{1}{\theta}\right)(m^* - m)|Y\right]\right] \\
&= E\left[\frac{1}{\theta}\left(\log E[\exp(\theta V_a(X,Y))|Y] - 1\right) E\left[(m^* - m)|Y\right]\right]. \qquad (2.3.9)
\end{aligned}$$

By constraint (2.3.7), for any $Y = y$, the conditional expectation $E[(m^* - m)|Y]$ in (2.3.9) equals zero, so $\bar{K}'(0) = 0$. Hence m^* is an optimal solution satisfying constraint (2.3.7).

Next, we show that any likelihood ratio satisfying (2.3.7) preserves the distribution of Y. Let a tilde indicate the distribution following the change of measure.

$$\begin{aligned}
\tilde{P}(Y \in D) &= E[m^*(\theta, X, Y) I_{Y \in D}] \\
&= E\left[E[m^*(\theta, X, Y) I_{Y \in D}|Y]\right] \\
&= E\left[I_{Y \in D} E[m^*(\theta, X, Y)|Y]\right] \\
&= E[I_{Y \in D}] = P(Y \in D) \\
&\quad \text{for any } Y\text{-measurable set } D.
\end{aligned}$$

Thus, the likelihood ratio for the marginal law of Y is identically equal to 1, indicating that the distribution of Y is unchanged. □

To implement (2.3.8), we need to generate multiple copies $X_1, \ldots, X_N$ for each outcome of y and then form the Monte Carlo counterpart of (2.3.8),

$$\hat{m}^*(x, y) = \frac{\exp(\theta V_a(x, y))}{\sum_{i=1}^{N} \exp(\theta V_a(X_i, y))/N} \tag{2.3.10}$$

Robust Monte Carlo Recap: We conclude this section with a brief summary of the implementation tools of this section.

- By simulating under the nominal model and weighting the results as in (2.3.1), we can estimate the worst-case error at each level of θ. We can do this across multiple values of θ at minimal computational cost. By also estimating $\eta(\theta)$ as in (2.3.2), we can plot the worst-case error as a function of relative entropy.
- To examine the effect of the change of measure define by θ, we can estimate moments and the expectations of other auxiliary functions using (2.3.3). We can also sample directly from the measure defined by θ using acceptance-rejection — exactly if V is bounded and approximately if not.
- We can constrain the worst-case change of measure through constraints on moments or other auxiliary functions using Proposition 2.3.1. This technique can be used iteratively to constrain the potential model error if the values estimated through (2.3.3) appear implausible.
- Using Proposition 2.3.2 and (2.3.10), we can constrain the worst-case model to leave certain

marginal distributions unchanged. This too can be used iteratively to focus robustness on the most uncertain features of model.

2.4 Portfolio Variance

The rest of the chapter deals with applications of the ideas developed in the previous sections. In Section 2.2.1, we illustrated the key ideas of robust risk measurement through an application to portfolio variance. Here we expand on this example.

2.4.1 Mean-Variance Optimal Portfolio

We extend our earlier discussion of portfolio variance to cover the selection of mean-variance optimal portfolios under model uncertainty. For the mean-variance objective, let $\gamma > 0$ be a risk-aversion parameter and consider the optimization problem

$$\inf_a -E[a^\top X - \frac{\gamma}{2} a^\top (X - E[X])(X - E[X])^\top a]. \tag{2.4.1}$$

As before, a denotes a vector of portfolio weights. To illustrate the method of Section 2.3.2, we constrain the mean vector and limit uncertainty to the covariance matrix, which leads to the robust problem

$$\begin{aligned} \inf_a \sup_m E[mV_a(X)] &= \inf_a \sup_m -E[m(a^\top X - \frac{\gamma}{2} a^\top (X-\mu)(X-\mu)^\top a)] \\ &\text{s.t.} \qquad E[mX] = \mu. \end{aligned}$$

Following the argument in Section 2.3.2, for some a, $\theta > 0$ and θ_μ (θ_μ corresponds to the vector $(1/\theta_i)$), the worst-case likelihood ratio is

$$\begin{aligned} m^* &\propto \exp\left(\theta(V_a(X) - \theta_\mu^\top(X-\mu))\right) \qquad (2.4.2) \\ \text{where } \theta_\mu \text{ solves} &\qquad \inf_{\theta_\mu} \frac{1}{\theta}\log E\left[\exp\left(\theta\left[V(X) - \theta_\mu^\top X\right]\right)\right] - \theta_\mu^\top \mu. \end{aligned}$$

Proceeding as in Section 2.2.1, we find that the worst-case change of measure preserves the normality of X. The term with θ_μ is linear in X and therefore affects only the mean of X. Because we have constrained the mean, m^* satisfies

$$m^* \propto \exp(\frac{\theta\gamma}{2} a^\top (X-\mu)(X-\mu)^\top a), \qquad (2.4.3)$$

Matching (2.4.2) and (2.4.3), we find that $\theta_\mu = -a$.

For given $\theta > 0$, let $\mathcal{A}(\theta) = \{a : \Sigma^{-1} - \theta\gamma a a^\top > 0\}$ denote the set of portfolio vectors a that ensure that the resulting covariance matrix is positive definite. Then for given (a, θ) such that $\theta > 0$ and $a \in \mathcal{A}(\theta)$, the worst-case change of measure has $X \sim N(\mu, \tilde{\Sigma})$, where $\tilde{\Sigma}^{-1} = \Sigma^{-1} - \theta\gamma a a^\top$. We can find the optimal a by numerically solving

$$\begin{aligned} a^*(\theta) &= \arg\inf_{a\in\mathcal{A}(\theta)} \frac{1}{\theta}\log E[\exp\left(\theta\left[V(X) - \theta_\mu^\top X\right]\right)] + \theta_\mu^\top \mu \\ &= \arg\inf_{a\in\mathcal{A}(\theta)} \frac{1}{\sqrt{\det(I - \theta\gamma a a^\top \Sigma)}} - a^\top \mu. \qquad (2.4.4) \end{aligned}$$

The corresponding relative entropy is

$$\eta(\theta) = \frac{1}{2}\left(\log(\det(\Sigma\tilde{\Sigma}^{-1})) + \mathrm{tr}(\Sigma^{-1}\tilde{\Sigma} - I)\right).$$

To illustrate, we consider an example with 10 assets, where $\mu_i = 0.1$, $\sigma_{ii} = 0.3$ and $\rho_{ij} = 0.25$ for $i \neq j$, $i, j = 1, ..., 10$, and $\gamma = 1$. We refer to the optimal portfolio at these parameter values as the nominal portfolio (NP). At each θ value, we compute the robust portfolio (RP), meaning the one that is optimal under the change of measure defined by θ. In the left panel of Figure 2.1, we plot the performance of the two portfolios (as measured by the mean-variance objective — recall that we are minimizing) against relative entropy (which we also compute at each θ). The performance of the NP portfolio under the nominal model is simply a horizontal line. The performance of the RP portfolio under the nominal model is always inferior, as it must be since NP is optimal in the nominal model. However, under the worst-case model, the RP values are better than the NP values, as indicated by the upper portion of the figure. In the lower portion of the figure we see the performance of the nominal portfolio under the best-case model perturbation possible at each level of relative entropy. The vertical gap between the two portions of the NP curve indicate the model risk at each level of relative entropy.

One of the themes of this chapter is that model error as gauged by relative entropy does not necessarily correspond to a straightforward error in parameters. To illustrate, in the right panel we examine the performance of the nominal portfolio under specific parameter perturbations. We vary the common correlation parameter from $\rho = 0.05$ (which produces the best performance) to $\rho = 0.45$ (which produces the worst); the relative entropy first decreases and then increases as

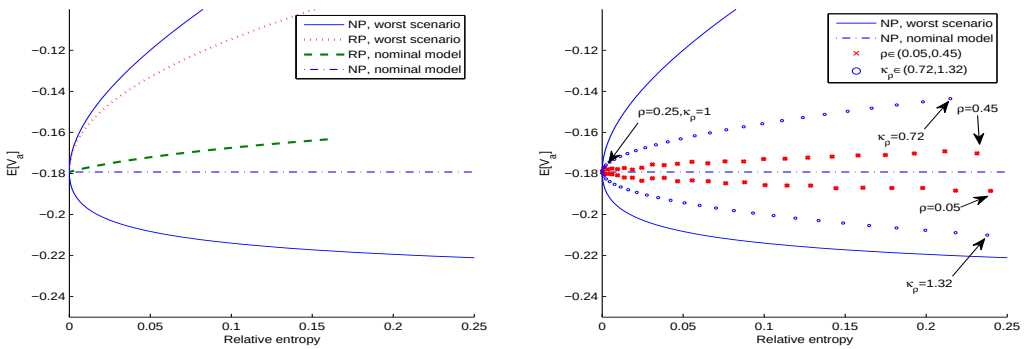


Figure 2.1: Expected performance vs relative entropy. The left panels shows the performance of the nominal portfolio (NP) and the robust portfolio (RP) under the nominal and worst-case models. The right panel shows the performance of the nominal portfolio under perturbations in model parameters. Higher values on the vertical scale indicate worse performance.

ρ moves through this range. We also examine the effect of multiplying the covariance matrix of the assets by $\kappa_\rho \in (0.72, 1.32)$. The key point — and one to which we return often — is that the worst-case change of measure results in significantly worse performance than any of these parameter perturbations.

Glasserman and Xu (2013) study a dynamic version of the mean-variance problem with stochastic factors and transaction costs. The analysis results in closed-form solutions for both the investor and adversary. For general multi-period problems, Iyengar (2005) develops a robust version of dynamic programming.

2.4.2 Empirical Example

To apply these ideas to data, we use daily returns from the CRSP database on the 126 stocks that were members of the S&P500 index from January 1, 1990, to December 31, 2011.

We first estimate the mean μ and covariance Σ of daily return using the first 12 years of data, through the end of 2001. For the covariance matrix we use the shrinkage method in Ledoit and

Wolf (2003). Based on the estimated mean and covariance matrix, we construct the mean-variance optimal portfolio

$$a = (\gamma\Sigma)^{-1}(\mu - \theta_\mu I) \tag{2.4.5}$$

$$\text{where } \theta_\mu = (I^\top(\gamma\Sigma)^{-1}\mu - 1)/(I^\top(\gamma\Sigma)^{-1}I)$$

and $\gamma = 10$. We assume a static portfolio with total capital of 1. We take the portfolio variance from the initial time period as a forecast of the future variance for the same portfolio. We compare this forecast with the realized variance in 2002, when the dot-com bubble burst.

In the first column of Table 2.1, we see that the realized variance in 2002 is quite large compared to the forecast using the previous 12 years of data. Confidence intervals equal to two times the standard error of the realized variance and forecast have no overlap. The sampling variability in the initial period is not large enough to explain the realized variance.

	2002	2008
Realized variance	0.35×10^{-3}	0.65×10^{-3}
$\pm 2\times$Std. Err.	$(0.29, 0.42) \times 10^{-3}$	$(0.53, 0.77) \times 10^{-3}$
Forecast variance	0.21×10^{-3}	0.21×10^{-3}
$\pm(2\times$Std. Err.+Model Err.)	$(0.20, 0.22) \times 10^{-3}$	$(0.20, 0.22) \times 10^{-3}$
$\theta = 100$	$(0.21, 0.25) \times 10^{-3}$	$(0.17, 0.22) \times 10^{-3}$
$\theta = 500$	$(0.18, 0.32) \times 10^{-3}$	$(0.14, 0.32) \times 10^{-3}$
$\theta = 900$	$(0.16, 0.47) \times 10^{-3}$	$(0.12, 0.58) \times 10^{-3}$

Table 2.1: Realized and forecast variance with model uncertainty.

Next we introduce error intervals based on relative entropy. We use the portfolio variance as

the objective and obtain the worst-case variance at different levels of θ. Let

$$\text{Model Error} = |\text{nominal variance-worst variance}|.$$

Now we can form a new interval by combining both standard error and model error. In the lower part of Table 2.1, the new interval almost reaches the realized variance in 2002 when $\theta = 500$, and it covers the confidence interval of realized variance when $\theta = 900$. By considering both sampling variability and model error, we can cover the 2002 scenario.

This gives us a rough sense of the level of robustness needed to capture a sharp change like that in 2002. We now position ourselves at the end of 2007 and undertake a similar analysis. Again, we use the previous 12 years of data to form a forecast, which is 0.21×10^{-3}. We choose $\theta = 900$ as the robustness level, based on the study of 2002, so that the whole confidence interval of 2002 is contained.

The model errors for the forecast of 2002 were 0.10×10^{-3} and 0.25×10^{-3} for $\theta = 500$ and 900 respectively, and they change to 0.10×10^{-3} and 0.36×10^{-3} in the forecast of 2008. The forecast with both standard error and model error forms a pretty wide interval, which has a slight overlap with the confidence interval of the realized variance in 2008. Although the crisis in 2008 was more severe than the drop in 2002, the market change in 2002 provides a rough guide of potential model risk. The particular combination we have used of sampling error and model error is somewhat heuristic, but it nevertheless shows one way these ideas can be applied to historical data.

2.4.3 The Heavy-Tailed Case

To illustrate the use of α-divergence in the heavy-tailed setting, we now suppose that the vector of asset returns is given by $X \sim \mu + Z_T$, where $Z_T \sim t_\nu(\Sigma, \nu)$ has a multivariate t distribution with $\nu > 2$ degrees of freedom and covariance matrix $\nu\Sigma/(\nu-2)$. Because neither the t-distribution nor a quadratic function of X has a moment generating function, we use α-divergence as an uncertainty measure. With a fixed portfolio weight vector a, Proposition 2.2.6 yields the worst-case likelihood ratio

$$\begin{aligned} m^*(\theta,\alpha) &= (\theta(\alpha-1)V_a(X) + c(\theta,\alpha))^{\frac{1}{\alpha-1}} \\ \text{with } c(\theta,\alpha) \text{ s.t.} &\quad E[m^*(\theta,\alpha)] = 1 \\ \text{where } V_a(X) &= a^\top (X-\mu)(X-\mu)^\top a. \end{aligned} \tag{2.4.6}$$

To illustrate, we consider an portfolio with $n = 10$ assets, $\nu = 4$, $\mu_i = 0.1$, $\Sigma_{ii} = 0.28 + 0.02 \times i$ and $\rho_{ij} = 0.25$ for $i, j = 1, ..., n$ and $i \neq j$. We use a randomly generated portfolio weight vector

$$a = [0.0785,\ 0.1067,\ 0.1085,\ 0.1376,\ 0.0127,\ 0.2204,\ 0.0287,\ 0.1541,\ 0.1486,\ 0.0042],$$

and simulate $N = 10^7$ samples to examine the worst-case scenario. Table 2.2 shows the portfolio variance across various values of θ and α, with $\theta = 0$ corresponding to the baseline nominal model. For fixed α, increasing θ increases the uncertainty level and increases the worst-case variance. The middle column of the table shows results using estimated parameters at $\alpha = 2.5$; we return to these at the end of this section.

θ	$\alpha = 2$	$\alpha = 2.5$	$\alpha = 2.5$, worst parameters (DOF)	$\alpha = 3$	$\alpha = 3.5$
0	0.109	0.109	0.109	0.109	0.109
0.1	0.159	0.131	0.130 (3.15,3.65)	0.125	0.122
0.4	0.308	0.174	0.174 (2.84,3.18)	0.152	0.143
0.7	0.458	0.210	0.209 (2.77,2.93)	0.173	0.159
1	0.607	0.241	0.238 (2.74,2.84)	0.190	0.171

Table 2.2: Worst-case portfolio variance at different levels of θ and α. The middle column reports estimates using parameters estimated at $\alpha = 2.5$, showing first the portfolio variance and then the degrees of freedom parameter (in parentheses) estimated using $\nu_\alpha = \nu + k_{no} - k_{\theta,\alpha}$ and maximum likelihood.

We saw in (2.2.21) that the choice of α influences the tail of $V(X)$ under the worst-case change of measure. A smaller α in Table 2.2 yields a heavier tail, but this does not necessarily imply a larger portfolio variance. To contrast the role of α with θ, we can think of choosing α based on an assessment of how heavy the tail might be and then varying θ to get a range of levels of uncertainty. In both cases, some calibration to the context is necessary, as in the empirical example of the previous section and in the discussion below.

To understand the influence of the α parameter, we examine the tail of the portfolio excess return, $r = a^\top X - \mu$. Figure 2.2 plots the tail probability of $|r|$ on a log-log scale. Because r has a t distribution, the log of the density of $|r|$, denoted by $g_{|r|}(x)$, is asymptotically linear

$$\log g_{|r|}(x) \approx -(\nu + 1) \log x, \text{ for } x \gg 0.$$

Using the fact that

$$\log\left(m^*(\theta, \alpha)\right) = \frac{1}{\alpha - 1} \log(\theta(\alpha - 1)r^2 + c(\theta, \alpha)) \approx \frac{2}{\alpha - 1} \log |r|,$$

we find (as in (2.2.21)) that

$$\log\left(\tilde{g}_{|r|}(x)\right) - \left(\log g_{|r|}(x)\right) \approx \frac{2}{\alpha - 1} \log x, \text{ for } x \gg 0 \tag{2.4.7}$$

where $\tilde{g}_{|r|}$ is the density of $|r|$ under the change of measure. This suggests that the difference of the slopes in Figure 2.2 between the nominal and worst scenario should be roughly $2/(\alpha - 1)$. Asymptotically, the tail under the worst scenario is similar to a t distribution with degrees of freedom $\nu - 2/(\alpha - 1)$.

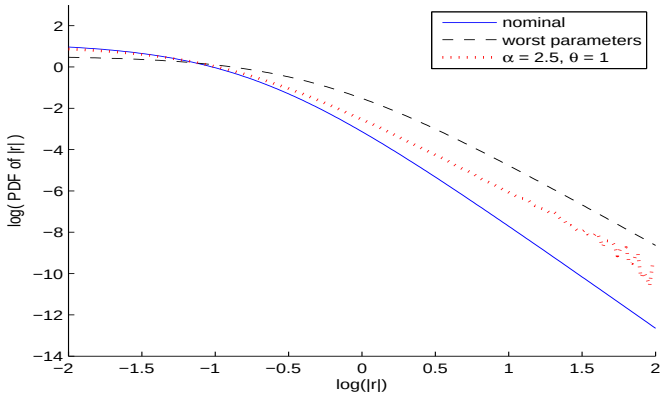


Figure 2.2: Tail density of absolute returns $|r|$.

We fit linear functions to the curves in Figure 2.2 in the region $\log(|r|) \in (0.5, 2)$ and compare the slopes of nominal k_{no} and worst scenario $k_{\theta,\alpha}$. Table 2.3 lists the differences $k_{\theta,\alpha} - k_{no}$; as we increase θ, the difference of slopes gets closer to the limit $2/(\alpha - 1)$ in (2.4.7).

By re-weighting the sample under the nominal model using $m^*(\theta, \alpha)$, we can estimate model parameters as though the worst-case model were a multivariate t. We estimate the degrees of freedom parameter using

$$\nu_{\alpha,\theta} = \nu + k_{no} - k_{\theta,\alpha} \tag{2.4.8}$$

	$\alpha = 2$	$\alpha = 2.5$	$\alpha = 3$	$\alpha = 3.5$
$2/(\alpha - 1)$	2	1.333	1	0.8
$\theta = 0.1$	1.090	0.846	0.694	0.590
$\theta = 0.4$	1.631	1.159	0.899	0.735
$\theta = 0.7$	1.773	1.231	0.943	0.764
$\theta = 1$	1.840	1.263	0.962	0.777

Table 2.3: Difference of slopes $k_{\theta,\alpha} - k_{no}$ of the worst-case and nominal densities, as in Figure 2.2.

and estimate the covariance matrix as

$$\text{worst covariance} = E[ma^\top (X-\mu)(X-\mu)^\top a] \approx \frac{1}{N}\sum_{i=1}^{N} m(X_i) a^\top (X_i - \mu)(X_i - \mu)^\top a.$$

We can then generate a second set of samples from the t distribution with these parameters to see how this compares with the actual change of measure.

In the middle of Table 2.2, we show the estimated $\nu_{\alpha,\theta}$ using (2.4.8) as the first number in parentheses. The second value is a maximum likelihood estimate using $m^*_{\theta,\alpha}$ to weight the nominal samples. The two values are relatively close; we use only (2.4.8) in sampling under the worst-case parameter values and in Figure 2.2. The variance results under the parameters estimated at $\alpha = 2.5$ are very close to those estimated under the worst-case model at $\alpha = 2.5$, suggesting that the worst case might indeed be close to a t distribution. Interestingly, Figure 2.2 shows that using the parameters from the worst case actually produces a heavier tail; the worst-case change of measure magnifies the variance through relatively more small returns than does the approximating t distribution. In Table 2.4, we see that the α-divergence under the approximating t is much larger. Thus, the adversary has economized the use of α-divergence to magnify the portfolio variance without making the tail heavier than necessary.

θ	$\alpha = 2.5$	approximating t-dist.
0.1	0.001	0.086
0.4	0.012	0.230
0.7	0.031	0.287
1	0.058	0.323

Table 2.4: Comparison of α-divergence using the worst-case change of measure and the approximating t distribution from the worst case.

2.5 Conditional Value at Risk (CVaR)

The next risk measure we consider is conditional value at risk (CVaR), also called expected shortfall. The CVaR at quantile δ for a random variable X representing the loss on a portfolio is defined by

$$\begin{aligned} CVaR_\delta &= E[X|X > VaR_\delta], \\ \text{where } VaR_\delta \text{ satisfies } 1-\delta &= P(X > VaR_\delta). \end{aligned}$$

As in Rockafellar and Uryasev (2002), CVaR also equals to the optimal value of the minimization problem

$$\min_a \frac{1}{1-\delta} E[(X-a)_+] + a, \tag{2.5.1}$$

for which the optimal a is VaR_δ.

To put this problem in our general framework, we set $V_a(X) = (1-\delta)^{-1}(X-a)_+ + a$. The main source of model error in measuring CVaR is the distribution of X. As in previous sections, we can introduce robustness to model uncertainty by considering a hypothetical adversary who changes the distribution of X. Of particular concern is the worst-case CVaR subject to a plausibility constraint

formulated through relative entropy or α-divergence. Jabbour et al. (2008) and Zhu and Pykhtin (2007) consider robust portfolio optimization problems using CVaR but different types of model uncertainty.

To illustrate the general approach, we introduce two specific examples that offer some analytic tractability, one in the relative entropy setting and one using α-divergence.

2.5.1 Relative Entropy Uncertainty

Suppose X follows a double exponential distribution $DE(\mu_{de}, b_{de})$ with location parameter μ_{de} and scale parameter b_{de}, meaning that its density function is

$$g(x) \propto \exp\left(-\frac{|x-\mu_{de}|}{b_{de}}\right).$$

Then for given a and $\theta > 0$, the density function of X under the worst-case change of measure becomes

$$\tilde{g}(x) = m^*_{\theta,a}(x)g(x) \propto \exp\left(-\frac{|x-\mu_{de}|}{b_{de}} + \frac{\theta}{1-\delta}(x-a)_+\right).$$

The values of a and δ are connected by $P(X > a) = 1-\delta$ under the nominal distribution. Because $\theta/(1-\delta) > 0$, we need $1/b_{de} > \theta/(1-\delta)$ to ensure this density function is well-defined. The exponent is a piecewise linear function of the argument x, so $\tilde{g}$ can be considered a generalization of the double exponential distribution.

We can find the VaR and CVaR explicitly in this example. First, we evaluate the normalization

constant (2.2.10):

$$E[\exp(\theta V_a(X))] = \begin{cases} \exp(\theta a)\left[1+\frac{1}{2}\left(\frac{1}{1-\frac{\theta b_{de}}{1-\delta}}-1\right)\exp(\frac{\mu_{de}-a}{b_{de}})\right], & \text{if } a > \mu_{de} \\ \frac{1}{2}\exp(\theta a)\left[\left(\frac{1}{1-\frac{\theta b_{de}}{1-\delta}}+\frac{1}{1+\frac{\theta b_{de}}{1-\delta}}\right)\exp(\frac{\theta}{1-\delta}(\mu_{de}-a))\right. & \\ \qquad \left. +\left(1-\frac{1}{1+\frac{\theta b_{de}}{1-\delta}}\right)\exp(\frac{a-\mu_{de}}{b_{de}})\right], & \text{else.} \end{cases} \tag{2.5.2}$$

Denote the cumulant generating function of $V_a(X)$ by $\Upsilon_a(\theta) = \log E[\exp(\theta V_a(X))]$; then

$$a^*(\theta) \quad = \quad \arg\min_a \frac{1}{\theta}\Upsilon_a(\theta),$$

To find a^*, we observe that the function $E[\exp(\theta V_a(X))]$ is convex in a and its derivative at $a = \mu_{de}$ is

$$\frac{d}{da}E[\exp(\theta V_a(X))]\Big|_{a=\mu_{de}} = \frac{\theta}{2}(2+\frac{\theta b_{de}-1}{1-\delta-\theta b_{de}})\exp(\theta\mu_{de}).$$

This is positive provided $\delta > 1/2$, so we can solve the first order condition for $a > \mu_{de}$ to get

$$a^*(\theta) = \mu_{de} - b_{de}\log\left(\frac{2(1-\delta-\theta b_{de})}{1-\theta b_{de}}\right),$$

which is the VaR under the worst-case change of measure.

The VaR for the nominal model is

$$VaR_\delta = \mu_{de} - b_{de}\log(2(1-\delta)).$$

and the nominal CVaR is

$$CVaR_\delta = VaR_\delta + b_{de} = \mu_{de} - b_{de}\log(2(1-\delta)) + b_{de}.$$

Under the worst-case change of measure at parameter θ, the CVaR becomes

$$CVaR_{\delta,\theta} = a^*(\theta) + \frac{1}{\frac{1}{b_{de}} + \frac{\theta}{1-\delta}}.$$

So, here we can see explicitly how the worst-case CVaR increases compared to the nominal CVaR. The corresponding relative entropy is

$$\begin{aligned}
\eta(\theta) &= E[m^*_{a^*(\theta),\theta}\log m^*_{a^*(\theta),\theta}] \\
&= \theta\frac{E[V_{a^*(\theta)}\exp(\theta V_{a^*(\theta)}(X))]}{E[\exp(\theta V_{a^*(\theta)}(X))]} - \log E[\exp(\theta V_{a^*(\theta)}(X))] \\
&= \theta\Upsilon'_{a^*(\theta)}(\theta) - \Upsilon_{a^*(\theta)}(\theta).
\end{aligned}$$

Figure 2.3 shows the nominal and worst-case densities starting from a nominal density that is $DE(0,1)$, using $\delta = 95\%$ and $\theta = 0.03$. The nominal 95% VaR is $a = 2.30$; the worst-case model error (for CVaR) at $\theta = 0.03$ shifts more mass to the right tail and increases the VaR to 3.19. The CVaR increases from 3.30 to 3.81. The increase in VaR and the corresponding increase in CVaR reflect the magnitude of underestimation of risk consistent with this level of the uncertainty parameter θ.

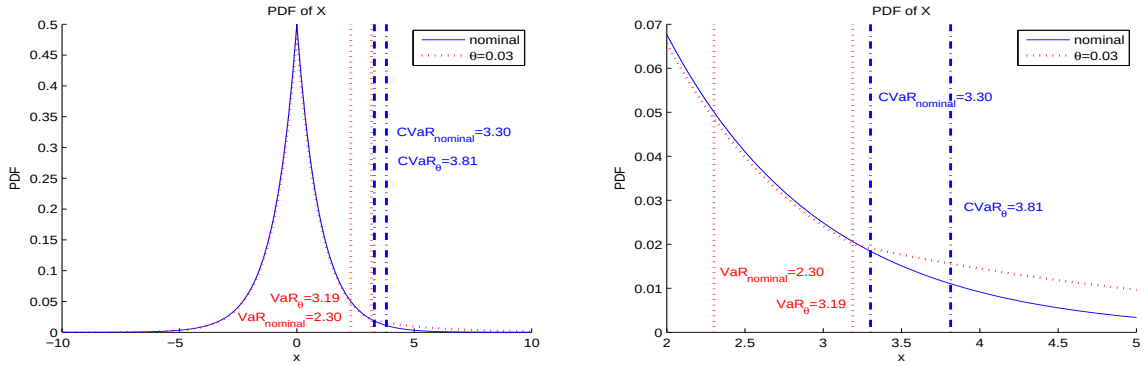


Figure 2.3: The dotted red line shows the worst-case density, with $\delta = 95\%$ and $\theta = 0.03$, relative to a $DE(0, 1)$ nominal density (the solid blue line). The right panel gives a magnified view of the right tail.

2.5.2 The Heavy-Tailed Case

If the nominal distribution of the loss random variable X is heavy-tailed, then $E[\exp(\theta V_a(X))]$ is infinite and the calculations in (2.5.2) and following do not apply. In this case, we need to use α-divergence as the uncertainty measure. With $\alpha > 1$, $\theta > 0$ and a fixed, the worst case likelihood ratio now becomes

$$m^*_{\theta,a}(X) = (\theta(\alpha - 1)V_a(X) + c(\theta, \alpha, a))^{\frac{1}{\alpha-1}}, \tag{2.5.3}$$

for some constant $c(\theta, \alpha, a)$ satisfying (2.2.17) and (2.2.18).

If the density function of X under the nominal distribution is regularly varying with index ξ, i.e. $\lim_{x\to\infty} g(tx)/g(x) = t^\xi$ for any $t > 0$ and some index $\xi < 0$, then under the worst-case change of measure it is regularly varying with index $\xi + 1/(\alpha - 1)$, as suggested by (2.2.21). We require $\xi + 1/(\alpha - 1) < 0$ to guarantee the new density function is well-defined. Because $\alpha > 1$, the worst index is smaller than the nominal one, meaning that the worst-case distribution has a heavier tail.

For purposes of illustration, it is convenient to choose as nominal model a generalized Pareto distribution with density function

$$g(x) = \frac{1}{b_{gp}}(1 + \frac{\xi_{gp}}{b_{gp}}x)^{-\frac{1}{\xi_{gp}}-1}, \text{ for } x \geq 0, \text{ some } b_{gp} > 0 \text{ and } \xi_{gp} > 0,$$

or a generalized extreme value distribution with density

$$g(x) = \frac{1}{\xi_{gev}}(1 + \xi_{gev}x)^{-\frac{1}{\xi_{gev}}-1} \exp\left(-(1 + \xi_{gev}x)^{-\frac{1}{\xi_{gev}}}\right), \text{ for } x \geq 0 \text{ and } \xi_{gev} > 0.$$

These are regularly varying with index $-(1 + 1/\xi)$, with $\xi = \xi_{gp}$ or $\xi = \xi_{gev}$, accordingly.

Figure 2.4 shows two examples — a generalized Pareto density on the left, and a generalized extreme value distribution on the right, each shown on a log scale. In each case, the figure compares the nominal distribution and the worst-case distribution with $\alpha = 4$. As in Figure 2.3, the worst-case model error shifts the VaR to the right and increases the weight of the tail beyond the shifted VaR, increasing the CVaR.

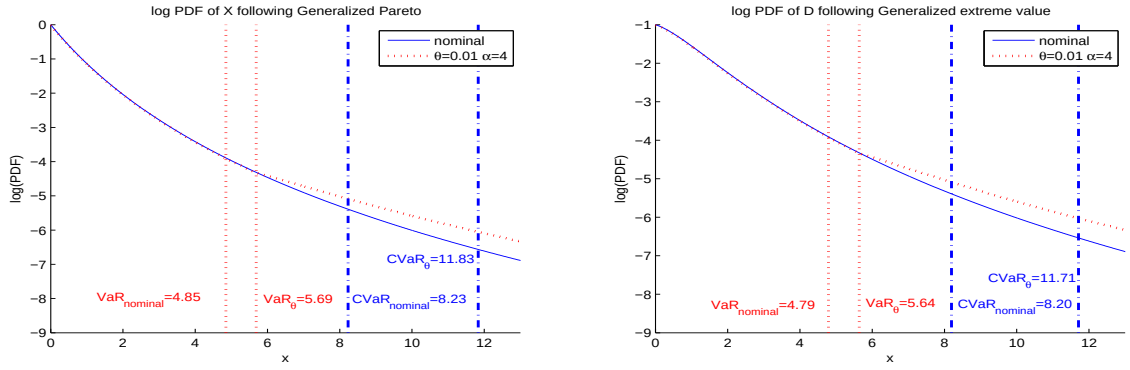


Figure 2.4: Density of X. The nominal distribution is generalized Pareto (left) or generalized extreme value (right), with parameters $b_{gp} = 1$ (scale), $\xi_{gp} = 0.3$ (shape), and $\xi_{gev} = 0.3$ (shape). Other parameters are $\theta = 0.01$, $\alpha = 4$, and $\delta = 95\%$.

A recurring and inevitable question in incorporating robustness into risk measurement is how much uncertainty to allow — in other words, where to set θ or α. If the distribution of X is estimated from historical data, then the precision with which the tail decay of X is estimated (the exponential decay in the light-tailed setting and power decay in the heavy-tailed setting) can provide some guidance on how much uncertainty should be incorporated, as we saw in Section 2.4.2. Also, the Monte Carlo approach presented in Section 2.3 illustrates how auxiliary quantities (for example, moments of X) can be calculated under the worst-case change of measure to gauge its plausibility.

2.6 Portfolio Credit Risk

In this section, we apply robustness to the problem of portfolio credit risk measurement. We develop the application within the framework of the standard Gaussian copula model; the same techniques are applicable in other models as well.

2.6.1 The Gaussian Copula Model

We consider a portfolio exposed to n obligors, and we focus on the distribution of losses at a fixed horizon. Let D_i denote the default indicator for ith obligor, meaning that

$$D_i = \begin{cases} 1, & \text{if the } i\text{th obligor defaults within the horizon;} \\ 0, & \text{otherwise.} \end{cases}$$

A default of obligor i produces a loss of c_i, so the total loss from defaults is

$$L = \sum_{i=1}^{n} c_i D_i.$$

We are interested in robust measurement of tail probabilities $P(L > l)$ for loss thresholds l.

In the Gaussian copula model, each default indicator D_i is represented through the indicator of an event $\{X_i^D > l_i\}$, where X_i^D has a standard normal distribution, and the threshold l_i is chosen so that $P(D_i = 1) = P(X_i^D > l_i) = p_i$, for a given default probability p_i. Dependence between default indicators is introduced through correlations between the X_i^D. For simplicity, we focus on a single-factor homogeneous model in which the X_i^D are given by

$$X_i^D = \rho Z + \sqrt{1 - \rho^2}\epsilon_i,$$

where Z, $\epsilon_1, \ldots, \epsilon_n$ are independent standard normal random variables. We interpret Z as a broad risk factor that affects all obligors, whereas ϵ_i is an idiosyncratic risk associated with the ith obligor only. We have $n = 100$ obligors, each with a 1% default probability p_i, so $l_i = 2.33$. The loss given default is $c_i \equiv 1$ for all $i = 1, \ldots, n$.

2.6.2 Robustness and Model Error

The Gaussian copula model offers an interesting application because it is both widely used and widely criticized for its shortcomings. Taking the Gaussian copula as a reference model, our interest lies in examining its greatest vulnerabilities to model error — in other words, finding which

perturbations of the model (in the sense of relative entropy) produce the greatest error in measuring tail loss probabilities $P(L > l)$. Importantly, we are interested in going beyond parameter sensitivities to understand how the worst-case error changes the structure of the model.

Taking our risk measure as $P(L > l)$ means taking $V(Z, \epsilon_1, \ldots, \epsilon_n) = I_{L>l}$, so the worst-case change of measure at parameter θ is

$$m_\theta^* \propto \exp(\theta I_{L>l}).$$

$$\Rightarrow \tilde{P}(L \in dy) = \begin{cases} \frac{\exp(\theta)}{C} P(L \in dy) & \text{if } l > y; \\ \frac{1}{C} P(L \in dl) & \text{otherwise.} \end{cases} \tag{2.6.1}$$

Here, $C > 1$ is a normalization constant. This change of measure lifts the probabilities of losses greater than l and lowers the probability of all other scenarios. Equivalently, we can say that the probability of any outcome of the default indicators $(D_1, \ldots, D_n)$ is increased by $\exp(\theta)/C$ if it yields a loss greater than l and is lowered by a factor of C otherwise.

We investigate the implications of this transformation to the model through numerical experiments. We take $l = 5$, which yields $P(L > l) = 3.8\%$. Our results are based on simulation with $N = 10^6$ samples.

Figure 2.5 shows how the loss probability varies with relative entropy. The solid blue line shows results under the worst-case change of measure defined by (2.6.1). The dotted red line shows results under parameter changes only; these are determined as follows. At each relative entropy level, we simulate results under the worst-case change of measure (2.6.1); we estimate all model

parameters (the means, standard deviations, and correlations for the normal random variables Z, $\epsilon_1, \ldots, \epsilon_n$); we then simulate the Gaussian copula model with these modified parameters.

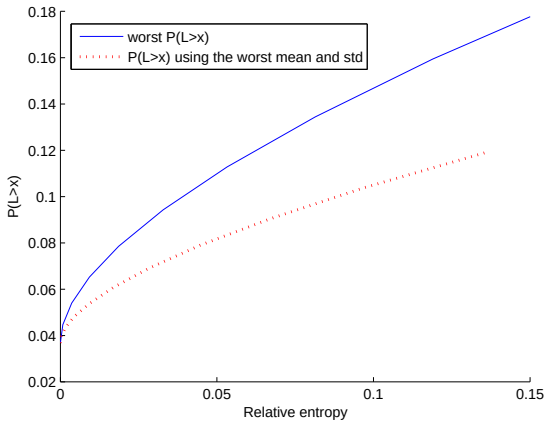


Figure 2.5: Loss probability as a function relative entropy. The solid blue line shows results under the worst-case change of measure. The dotted red line shows results using parameter values estimated from the worst-case change of measure. The comparison shows that the vulnerability to model error goes well beyond errors in parameters.

A comparison of the lines in Figure 2.5 confirms that the worst-case change of measure has an impact that goes well beyond a change in parameter values. If we compare the two curves at the same relative entropy, the worst-case model continues to show a higher loss probability. In other words, focusing on parameter changes only does not fully utilize the relative entropy budget. The changes in parameter values do not maximize the model error at a given relative entropy budget.

Table 2.5 reports parameter estimates obtained under the worst-case model at two values of θ. They indicate, in particular, that the parameters of the ϵ_i are affected very little by the change in distribution. Indeed, with 95% confidence, Jarque-Bera and Anderson-Darling tests reject normality of Z at $\theta \geq 1$ but fail to reject normality of the ϵ_i even at $\theta = 2$. The model is more vulnerable to errors in the dependence structure introduced by Z than to errors in the distribution of the idiosyncratic terms.

	$\theta = 0.5$	$\theta = 2$
$\max(\rho_{\epsilon_i\epsilon_j}, \rho_{\epsilon_i,Z})$	4.3×10^{-3}	0.013
$\min(\rho_{\epsilon_i\epsilon_j}, \rho_{\epsilon_i,Z})$	-3.4×10^{-3}	-4.7×10^{-3}
$average(\lvert\rho_{\epsilon_i\epsilon_j}\rvert, \lvert\rho_{\epsilon_i,Z}\rvert)$	5.6×10^{-3}	6.4×10^{-3}
$average(\mu_{\epsilon_j})$	7.6×10^{-4}	6.8×10^{-3}
$average(\sigma_{\epsilon_j})$	1.00	1.01
$average(skew_{\epsilon_j})$	1.7×10^{-3}	0.013
$average(excess\ kurtosis_{\epsilon_j})$	8.2×10^{-4}	0.017
mean of Z	0.047	0.39
standard deviation of Z	1.04	1.23

Table 2.5: Statistics of ϵ_j and Z under the worst-case change of measure.

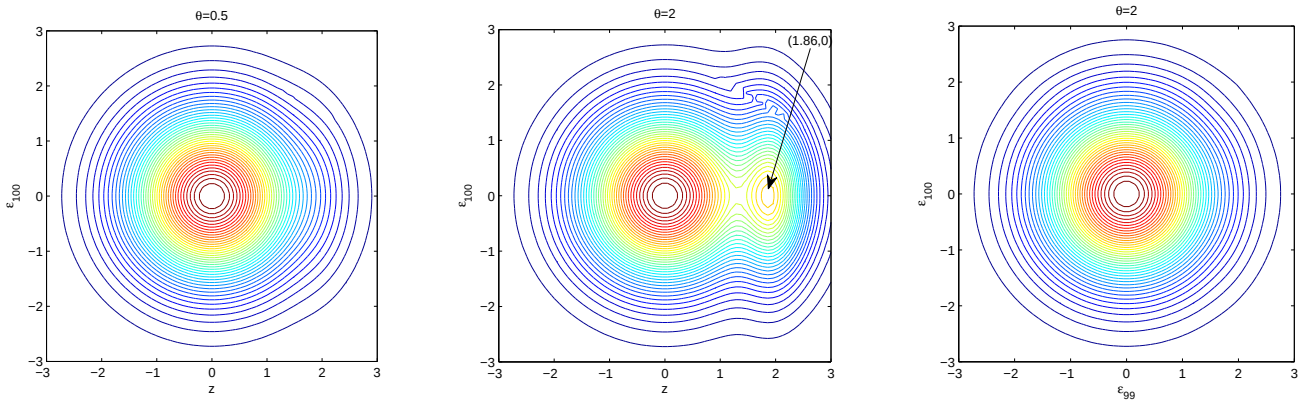


Figure 2.6: Contours of joint densities of (Z, ϵ_{100}) with $\theta = 0.5$ (left) and $\theta = 2$ (middle), and joint density of $(\epsilon_{99}, \epsilon_{100})$ at $\theta = 2$ (right)

To gain further insight into the worst-case change of distribution, we examine contour plots in Figure 2.6 of the joint density function of ϵ_{100} and Z. The joint density function is derived by using the original joint density function and the likelihood ratio m^*_θ. The leftmost figure shows $\theta = 0.5$, and the next two correspond to $\theta = 2$. The increase in θ shifts probability mass of Z to the right but leaves the joint distribution of the ϵ_i essentially unchanged. This shift in Z changes the dependence structure in the copula and produces the lift in the probability mass function of L described by (2.6.1). In the middle panel of Figure 2.6, we see a slight asymmetry in the upper

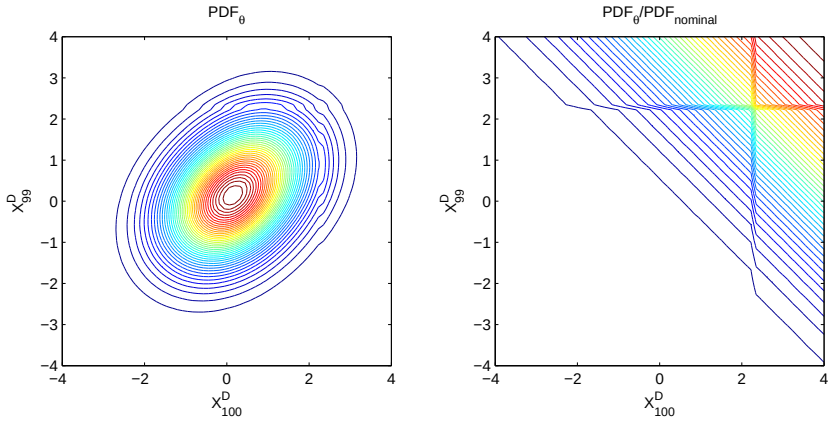


Figure 2.7: Contours of the joint density of (X_{99}^D, X_{100}^D) under the worst scenario $\theta = 2$ (left), and the ratio of the worst-case joint density to the nominal density (right).

right corner, reflecting the fact that defaults are more likely when both the ϵ_i and Z are increased. The left panel of Figure 2.7 shows contours of the joint density of (X_{99}^D, X_{100}^D) under the worst-case change of measure, which distorts the upper-right corner, reflecting the increased probability of joint defaults. The right panel shows the ratio of the worst-case density to the nominal density.

Figure 2.8 shows the nominal and worst-case marginal distributions of Z and L. The worst case makes Z bimodal and inflates the distribution of L beyond the threshold of 5. In particular, the greatest vulnerability to model error takes us outside the Gaussian copula model, creating greater dependence between obligors in the direction of more likely defaults, rather than just through a change of parameters within the Gaussian copula framework.

Next, we illustrate the effect of imposing constraints on Z, using the method of Section 2.3.2. We constrain the first moment to equal 0 or the first two moments to equal 0 and 1; one might take these values to be part of the definition of Z. To match relative entropy values, we find that an unconstrained value of $\theta = 2$ corresponds to constrained values $\theta = 2.7$ (with one constraint) and $\theta = 3.7$ (with two constraints); see Table 2.6. Figure 2.9 compares the marginal distribution of Z

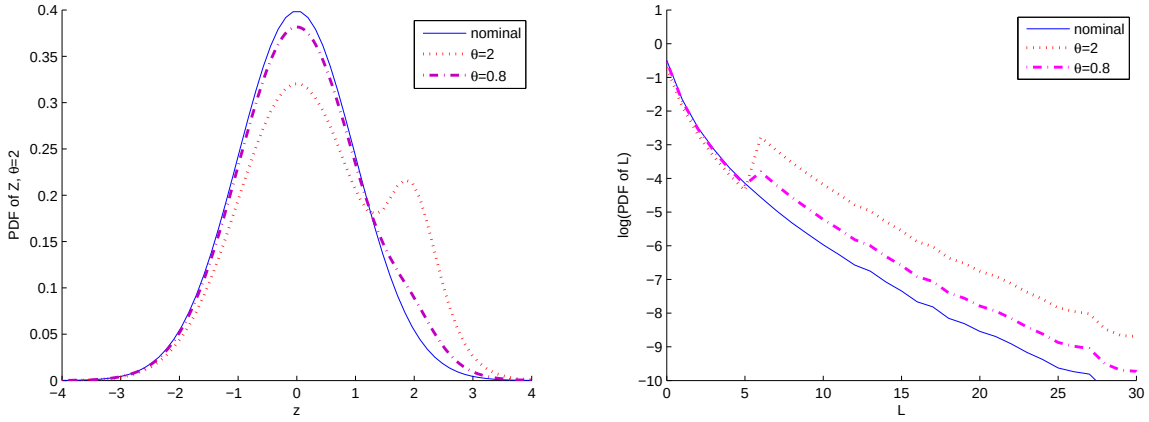


Figure 2.8: Marginal density of Z and L under worst scenario with $\theta = 0.8$ and $\theta = 2$ vs nominal model.

under the constrained and unconstrained worst-case changes of measure. The constraints lower the height of the second peak in the bimodal distribution of Z. Not surprisingly, the worst-case value of $P(L > l)$ decreases as we add constraints.

	$P(L > l)$
nominal, $\theta = 0$	0.037
unconstrained, $\theta = 2$	0.221
constraint on 1st moment of Z, $\theta = 2.7$	0.186
constraint on 1st and 2nd moments of Z, $\theta = 3.7$	0.153
constraint on marginal distribution of Z, $\theta = 4$	0.152

Table 2.6: Default probability for unconstrained and constrained cases. The values of θ for the constrained cases are chosen to keep the relative entropy fixed across all three cases.

We can further restrict the marginal distribution of Z through the method of Section 2.3.3. Such a restriction is important if one indeed takes Z as an overall market risk factor and not simply a tool for constructing a copula. Using 10^3 samples for Z and 10^4 samples of ϵ for each realization of Z, we report the resulting probability in the last row of Table 2.6, taking $\theta = 4$ to make the

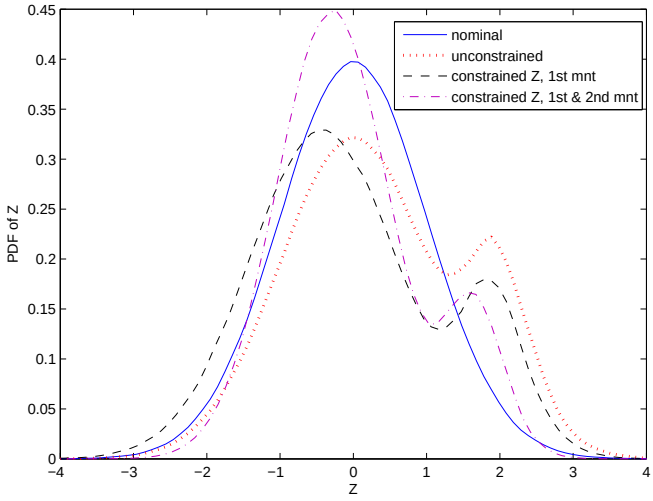


Figure 2.9: Density of Z under the nominal, unconstrained worst-case and constrained worst-case measures.

relative entropy roughly equal to that in the unconstrained case with $\theta = 2$. The default probability is slightly smaller than the case with constraints on 1st and 2nd moments.

Figure 2.10 shows the distribution of ϵ under the worst scenario, taking $\theta = 9$ to make the effect more pronounced. With the marginal distribution of Z held fixed, the potential model error moves to the idiosyncratic terms. The worst-case joint density of $(\epsilon_{99}, \epsilon_{100})$ puts greater weight on large values of either ϵ_{99} or ϵ_{100}. The worst-case marginal density of ϵ_{100} changes in a way similar to the marginal density of Z in Figures 2.8 and 2.9.

2.7 Delta Hedging Error

In our next application, we take hedging error as our measure of risk. This application goes beyond our previous examples by adding model dynamics to the robust risk measurement framework. The nominal model specifies dynamics for the evolution of an underlying asset, which leads to a

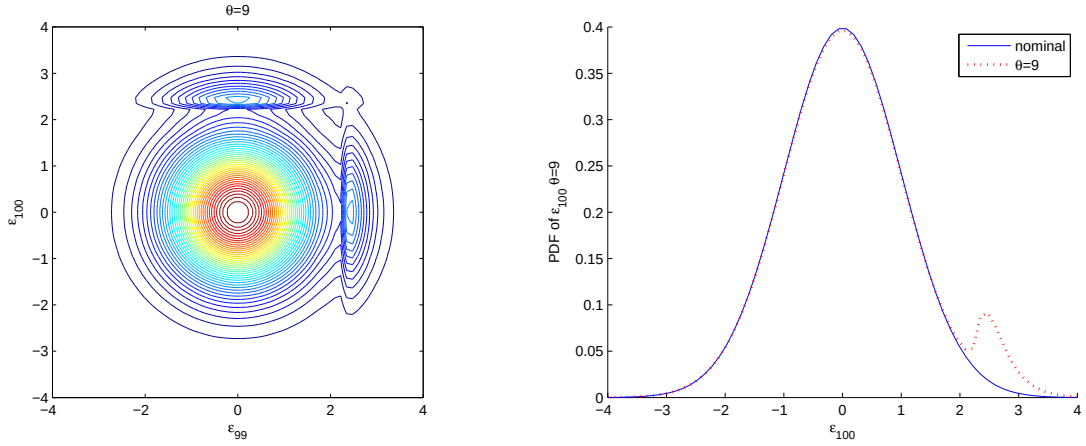


Figure 2.10: The marginal distribution of Z is fixed. The left figure is the joint density of $(\epsilon_{99}, \epsilon_{100})$ under the worst scenario, and the right figure is the marginal density of ϵ_{100} under the worst scenario. Both figures have $\theta = 9$.

hedging strategy for options written on the underlying asset. Model risk in this context can take the form of misspecification of the dynamics of the underlying asset, rather than just a marginal distribution at a fixed point in time. A hypothetical adversary can change the dynamics of the underlying asset and will do so in a way that maximizes hedging error subject to a relative entropy constraint. Our objectives are to quantify the potential hedging error, develop a hedging strategy that is robust to model error, and to identify the greatest sources of vulnerability to model error in the nominal model.

2.7.1 Delta Hedging: Nominal Model

For simplicity, we take the nominal model to be the Black-Scholes framework. The risk-neutral dynamics of the underlying asset are given by

$$\frac{dS_t}{S_t} = r_n dt + \sigma_n dW_t,$$

and the drift under the physical measure is μ_n. The risk-neutral drift enters in the option delta, but hedging error is generated under the physical measure so the physical drift is also relevant. The subscript n indicates that these parameters apply to the nominal model.

We consider the problem of discrete hedging of a European call option with strike K and maturity T: the interval $[0, T]$ is divided into N_T equal periods, and the hedging portfolio is rebalanced at the start of each period. With discrete rebalancing, we introduce hedging error even under the nominal model.

We consider a discrete-time implementation of a self-financing delta hedging strategy. At time $t = 0$, the proceeds of the sale of the option (at price $C(0, T, S_0)$) are used to form a portfolio of stock and cash, with r_n the interest rate for holding or borrowing cash. We denote by $\delta_{\sigma_n}(t, S_t)$ the number of shares of stock held at time t. It equals to the delta of the call option, which is the first derivative of the value of the option with respect to the underlying price under Black-Scholes model. At time 0, the portfolio's cash and stock positions are given by

$$\begin{aligned} \texttt{cash}(0) &= C(0, T, S_0) - S_0 \delta_{\sigma_n}(0, S_0), \\ \texttt{stock}(0) &= S_0 \delta_{\sigma_n}(0, S_0). \end{aligned}$$

After the rebalancing at time $kT/N_T = k\Delta t$, they are given by

$$\begin{aligned} \texttt{cash}(k) &= e^{r_n \Delta t} \texttt{cash}(k-1) - S_{k\Delta t}(\delta_{\sigma_n}(k\Delta t, S_{k\Delta t}) - \delta_{\sigma_n}((k-1)\Delta t, S_{(k-1)\Delta t})), \\ \texttt{stock}(k) &= S_{k\Delta t} \delta_{\sigma_n}(k\Delta t, S_{k\Delta t}). \end{aligned}$$

At maturity, the option pays $(S_T - K)_+$ with strike K, resulting in a hedging error is

$$H_e = (S_T - K)_+ - \texttt{cash}(N_T) - \texttt{stock}(N_T).$$

If adjusted continuously, the hedging portfolio should replicated the price of the option under the Black-Scholes model, i.e., $H_e = 0$. But the discrete hedging gives a non-zero H_e.

For our measure of hedging performance, we use $E[|H_e|]$, the expected absolute hedging error. A hypothetical adversary seeks to perturb the dynamics of S to magnify this hedging error. In our general formulation, we would take X to be the discrete path of the underlying asset and $V(X) = |H_e|$.

Alternative approaches to related problems include the uncertain volatility formulation of Avellaneda et al. (1995), where the volatility is assumed to lie within a closed interval but is otherwise unknown. In Mykland (2000), uncertainty is defined more generally through bounds on integrals of coefficients. Tankov and Voltchkova (2009) study the best volatility parameter to use for delta hedging to minimize expected squared hedging error under a jump-diffusion model for the underlying asset. Bertsimas et al. (2000) analyze asymptotics of the delta hedging error as $N_T \to \infty$.

In delta hedging, the volatility is unknown and is typically extracted from option prices. If the nominal model holds, then the minimizer of hedging error is indeed the nominal volatility σ_n. Under our formulation of robustness with discrete delta hedging, we can calculate a robust value of this input σ_n in the sense of minimizing the maximum value of the hedging error $E[|H_e|]$ at given value of θ. The result is illustrated in Figure 2.11 for an example with an initial stock price of $S_0 = 100$, strike $K = 100$, maturity $T = 1$, nominal volatility $\sigma_n = 0.2$, risk-free rate $r_n = 0.05$,

resulting in a Black-Scholes call price of 10.45 at $t = 0$. The drift under the physical measure is $\mu_n = 0.1$ The figure shows the nominal and robust values of delta as functions of the underlying asset; the robust σ_n is optimized against the worst-case change of measure at $\theta = 0.5$. The robust delta is slightly larger out-of-the-money and smaller in-the-money. Figure 2.11 suggests that if we are restricted to delta-hedging but are allowed to use different values for volatility, then the nominal value is almost the best we can do. Branger et al. (2012), among others, also find that Black-Scholes delta hedging performs surprsingly well, even when its underlying assumptions are not satisfied.

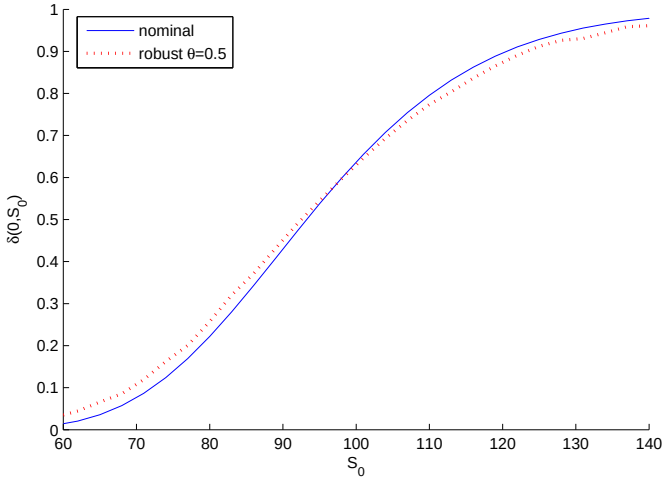


Figure 2.11: Optimal delta versus S_0 with $\theta = 0.5$.

2.7.2 Model Error and Hedging Error

Now we take a dynamic perspective on hedging error. We use simulation to investigate the vulnerability of discrete delta hedging to model error and to examine the worst-case change of measure

that leads to hedging errors. We continue to use the Black-Scholes model as the nominal model with the same parameters as before. Our simulation results use 10^8 paths.

From simulated sample paths and (2.2.11), we can estimate the optimal likelihood ratio m_θ^* for each path (we use $\theta = 0.5$ for most results), which enables us to estimate the density function of $|H_e|$ under the worst-case change of measure. The density is illustrated in Figure 2.12, where we can see that the change of measure makes the right tail heavier. In Figure 2.12, the tail is fit through a non-parametric method, using the "ksdensity" command in MATLAB with a normal kernel and bandwidth 0.1.

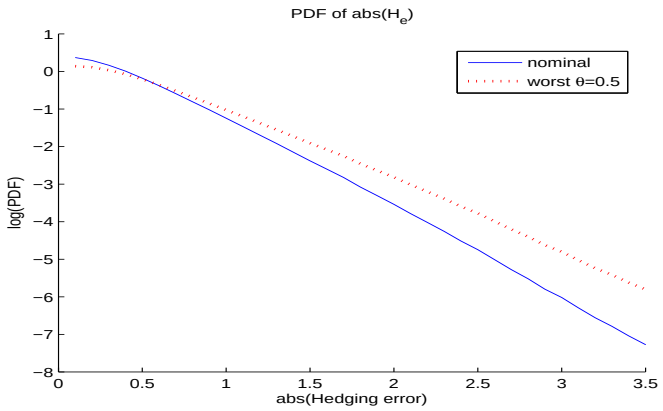


Figure 2.12: Density of absolute hedging error under nominal and worst scenario, with $\theta = 0.5$.

To investigate the dynamics of the underlying asset under the worst-case change of measure — in other words, to investigate the adversary's strategy — we generate paths conditional on reaching points (t, S_t). For every $t = T(2 + 8k/N_T)$, for $k = 1, ..., 12$, and every $S_t = 70 + 6l$ for $l = 1, ..., 10$, we simulate N sample paths conditioned to pass through (t, S_t) by using Brownian bridge sampling. If we use $\texttt{path}_i$ to denote the ith simulated path, then the conditional likelihood

given $S_t = x$ is

$$\begin{aligned} m^*(\texttt{path}_i | S_t = x) &= \frac{\tilde{g}(\texttt{path}_i | S_t = x)}{g(\texttt{path}_i | S_t = x)} \\ &= \frac{\tilde{g}(\texttt{path}_i)\, g(S_t \in (x, x+dx))}{g(\texttt{path}_i)\, \tilde{g}(S_t \in (x, x+dx))} \\ &\propto \frac{\tilde{g}(\texttt{path}_i)}{g(\texttt{path}_i)} = m^*(\texttt{path}_i) \end{aligned}$$

Because the expectation of the conditional likelihood ratio should be 1, we apply the normalization

$$m^*(\texttt{path}_i | S_t = x) = \frac{m^*(\texttt{path}_i)}{\sum_{j=1}^{N} m^*(\texttt{path}_j)/N}$$

across the N simulated paths.

As a point of comparison for the simulation results, it is useful to consider potential sources of hedging error. With discrete rebalancing, we would expect a large move in the underlying asset to produce a large hedging error. Figure 2.13 plots the option gamma and the time-decay theta, and these suggest that the hedging error is particularly vulnerable close to maturity when the underlying is near the strike. (Time in the figure runs from left to right, with time 1 indicating option maturity.) Indeed, the gamma at the strike becomes infinite at maturity.

In Figure 2.14, we use the simulation results to plot contours of the worst drift (upper left) and worst volatility (lower left) of the Brownian increment in the step immediately following the conditional value at (t, S_t). The conditional worst drift is highest close to maturity and just below the strike and it is lowest close to maturity and just above the strike, as if the adversary were trying

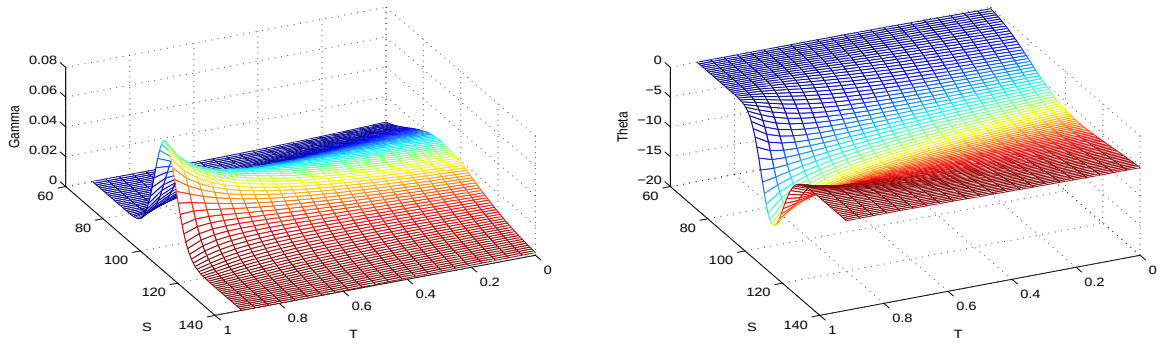


Figure 2.13: Gamma and Theta for European call option

to push the underlying toward the strike near maturity to magnify the hedging error. In fact, at every step t, the worst-case drift has an $S-$shape centered near the strike.

The worst-case volatility is also largest near the strike and near maturity, consistent with the view that this is where the model is most vulnerable. If the underlying is far from the strike, large hedging errors are likely to have been generated already, so the adversary does not need to consume relative entropy to generate further hedging errors. The contours of relative entropy show that the adversary expends the greatest effort near the strike and maturity. There is a slight asymmetry in the relative entropy and worst-case volatility below the strike near inception. This may reflect the asymmetry in gamma around the strike, which is greater far from maturity.

It should also be noted that the adversary's strategy is path-dependent, so Figure 2.14 does not provide a complete description. In particular, at any (t, S_t), we would expect the adversary to expend greater relative entropy — applying a greater distortion to the dynamics of the underlying — if the accumulated hedging error thus far is small than if it is large. The contours in the figure implicitly average over these cases in conditioning only on (t, S_t).

To generate Figure 2.14, we used kernel smoothing. The smoothed value at (s, t) is a weighted

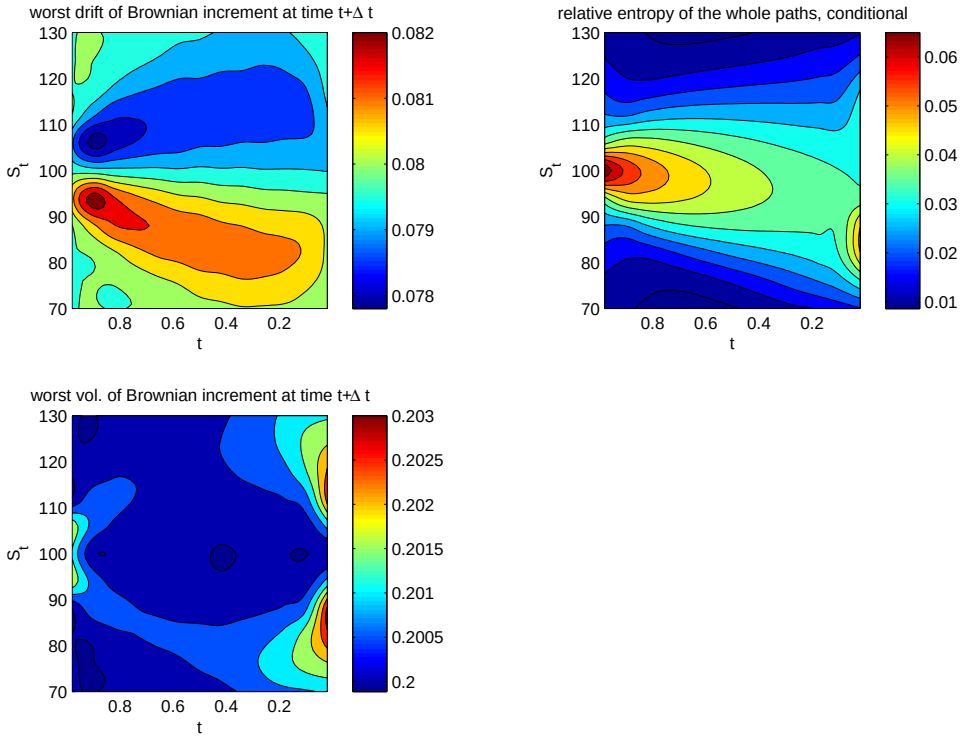


Figure 2.14: Conditional on S_t, the worst-case drift (upper left), relative entropy (upper right), and worst-case volatility (lower left), all using $\theta = 0.5$.

average of results at (s_i, t_i), $i = 1, ..., n$, using a kernel $K(.,.) > 0$,

$$g_{smooth}(s,t) = \frac{\sum_{i=1}^{n} g(s_i,t_i)K((s_i,t_i),(s,t))}{\sum_{i=1}^{n} K((s_i,t_i),(s,t))}.$$

In particular, we used $K((s',t'),(s,t)) = \phi(\|(s',t') - (s,t)\|/a)$, with ϕ the density of the standard normal distribution and $\|\ \|$ a scaled Euclidean normal under which the distance between adjacent corners in the grid is 1. That is, $\|(60,1) - (60,0)\| = 1$, $\|(60,1) - (140,1)\| = 1$, and so on. The constant a is chosen so that for any neighboring nodes (s,t) and (s',t') on the grid, $\|(s,t) - (s',t')\|/a = 1$.

2.7.3 Comparison With Specific Model Errors

In this section, we examine specific types of model errors and compare them against the worst case. In each example, we replace the Black-Scholes dynamics of the underlying asset with an alternative model. For each alternative, we evaluate the hedging error and the relative entropy relative to the nominal model. By controlling for the level of relative entropy, we are able to compare different types of model error, including the worst case, on a consistent basis.

In each plot in Figure 2.15, the horizontal axis shows the relative entropy of the perturbed model (with respect to the nominal model), and the vertical axis is the absolute hedging error estimated from simulation. We take values of θ in $[0, 0.23]$.

In panel (a) of Figure 2.15, we perturb the nominal model through serial correlation: we replace the i.i.d. Brownian increments with AR(1) dynamics. The perturbed model thus has $\Delta\tilde{W}_t = \rho\Delta\tilde{W}_{t-1} + \sqrt{1-\rho^2}\epsilon_t$ and $\Delta\tilde{W}_1 = \epsilon_1$, where ϵ_t are independent and normally distributed with mean 0 and variance Δt. With $\rho \in (-0.15, 0.15)$, the relative entropy reaches a minimum near $\rho = 0$. The expected hedging error seems to be robust with respect to serial dependence, never getting close to the worst case error except near the origin. The second plot in (a) suggests that a larger ρ leads to smaller hedging error. For larger $\rho > 0$, $\Delta\tilde{W}$ is more mean reverting, which may explain the smaller hedging error.

In panel (b) of Figure 2.15, we use Merton's jump-diffusion model,

$$\frac{dS_t}{S_{t-}} = (r_n - \lambda_J E[\exp(Y_i) - 1]))dt + \sigma_n dW_t + dJ_t$$

where J is a compound Poisson, $J_t = \sum_{i=1}^{N_t} \exp(Y_i)$, with N_t a Poisson process with intensity λ_J,

and Y_i i.i.d. $N(0, \sigma_J)$, with $\sigma_J = 1$. When increasing σ_J from 0 to 1, or the jump intensity λ_J from 0 to 0.2, both the relative entropy and the expected hedging error increase almost linearly, with similar slope.

Panels (c) and (d) of Figure 2.15 test the Heston stochastic volatility model, in which the square of volatility $v_t = \sigma_t^2$ follows the dynamics

$$dv_t = \kappa_\sigma(\mu_\sigma - v_t)dt + \sigma_\sigma\sqrt{v_t}dW_t^v, \tag{2.7.1}$$

where W_t^v is a Brownian motion, $\rho = corr(W_t^v, W_t)$. We pick $\kappa_\sigma = 5$, $\mu_\sigma = \sigma_n^2 = 0.04$, $\rho = -0.2$ and $\sigma_\sigma = 0.05$.

When discretized to dates $t_i = i\Delta t$, $i = 1, ..., N_T$, the likelihood ratio for the price process becomes

$$\begin{aligned} m(s_{t_1}, ..., s_{t_{N_T}}) &= \frac{\tilde{g}(s_{t_1}, ..., s_{t_{N_T}})}{g(s_{t_1}, ..., s_{t_{N_T}})} \\ &= \frac{E_v[\tilde{g}(s_{t_1}, ..., s_{t_{N_T}}|v_{t_1}, ..., v_{t_N})]}{g(s_{t_1}, ..., s_{t_{N_T}})} \end{aligned}$$

where g and $\tilde{g}$ are the joint density functions of prices under the nominal and Heston models, respectively. In the second equality, $\tilde{g}(.|.)$ denotes the conditional density of prices given the variance process, and the expectation is taken over the variance process. The conditional expectation is approximated using 1000 sample paths of v.

As the speed of mean-reversion κ_σ changes from 3 to 20, the relative entropy and the expected hedging error decrease. As κ_σ becomes larger, the expected hedging error gets closer to the nomi-

nal value, while relative entropy appears to converge to some positive value. With a large κ_σ, any deviation from the nominal variance decays quickly, leaving only a short-term deviation introduced by the diffusion term of (2.7.1).

As the long-run limit μ_σ varies from 0.036 to 0.044, relative entropy and expected hedging error attain their lowest values near 0.04, which is the nominal value of squared volatility. Holding fixed the level of relative entropy, the expected hedging error is very similar when $\mu_\sigma < 0.04$ and $\mu_\sigma > 0.04$. As the volatility of volatility σ_σ varies from 0 to 0.13, both relative entropy and expected hedging error increase. As σ_σ gets closer to zero, the volatility behaves more like a constant, which is the nominal model. And as the correlation ρ between the two Brownian motions varies from -0.5 to 0.7, the change in hedging error is very small, with the maximum hedging error obtained when ρ is close to nominal value -0.2. The relative entropy reaches the minimum value when ρ equals the nominal value -0.2.

For our last comparison, we use the variance-gamma model of Madan et al. (1998),

$$\begin{aligned} S_t &= S_0 \exp((\mu_{vg} + \omega_{vg})t + X_t) \\ \text{where} \quad X_t &= \mu_{vg}\Gamma(t; 1, \nu_{vg}) + \sigma W_{\Gamma(t;1,\nu_{vg})} \\ \omega_{vg} &= \frac{1}{\nu_{vg}} \log(1 - \mu_{vg}\nu_{vg} - \frac{1}{2}\sigma^2\nu_{vg}) \end{aligned}$$

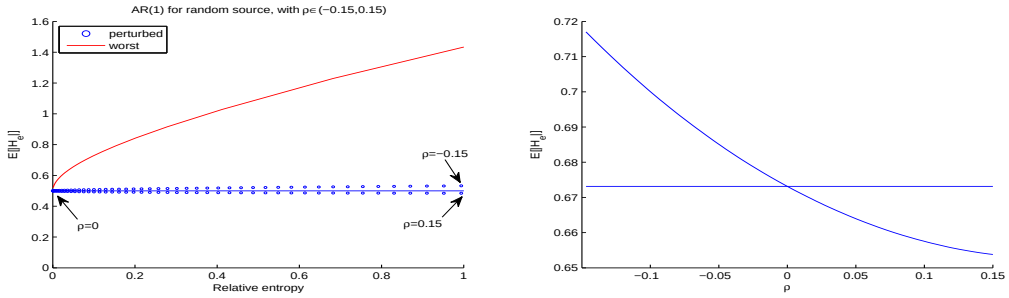
where $\Gamma(t; 1, \nu_{vg})$ is the gamma process with unit mean rate. Parameter μ_{vg} controls the skewness of return and ν_{vg} controls the kurtosis; see panels (b) and (c) of Figure 2.16. The figure suggests that skewness and kurtosis have limited impact on hedging error.

It is noteworthy that in most of the examples in Figures 2.15 and 2.16 the observed hedging

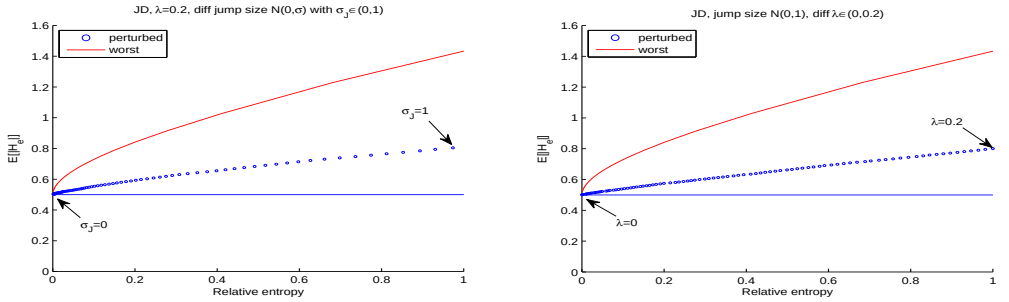
error is significantly smaller than that of the worst-case achievable at the same level of relative entropy. As our final test, we add constraints on the evolution of the underlying asset, thus limiting the adversary's potential impact.

First we constrain the moments of the realized mean and realized variance of the returns of the underlying asset. Let $\overline{\Delta W} = \sum_{i=1}^{N_T} \Delta W_i / N_T$ be the average of the Brownian increments ΔW_i along a path. We constrain the mean $E[m\overline{\Delta W}] = 0$ and the realized variance $E[m \sum_{i=1}^{N_T} (\Delta W_i - \overline{\Delta W})^2/(N_T - 1)] = \Delta t$. Figure 2.17(a) shows that this has only a minor effect on the worst-case hedging error. In Figure 2.17(b), we constrain the mean and variance of the realized variance as a way of constraining total volatility. Here, the reduction in the worst-case hedging error is more pronounced.

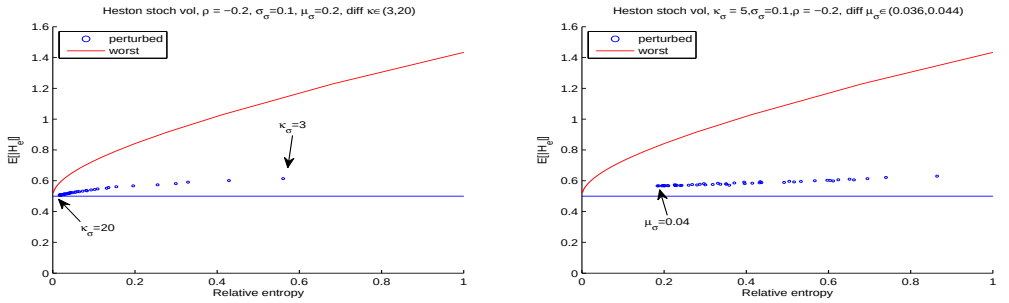
The overall conclusion from Figure 2.17 is that even with constraints on the first two moments of the underlying asset returns, the worst-case hedging error generally remains larger than the hedging errors we see in Figures 2.15 and 2.16 under specific alternatives. To put it another way, the hypothetical adversary shows much more creativity in undermining the Black-Scholes delta-hedging strategy than is reflected in these models. Indeed, the alternatives are all time-homogeneous, whereas a key feature of Figure 2.14 is that the greatest vulnerabilities occur close to maturity and, to a lesser extent, at inception.



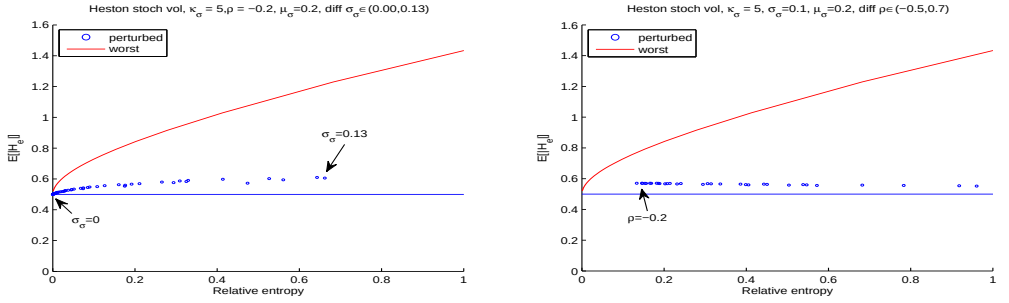
(a) AR(1) perturbation (left) and the effect of varying ρ (right).



(b) Jump-diffusion perturbation: the effect of different jump sizes (left) and jump intensity (right).

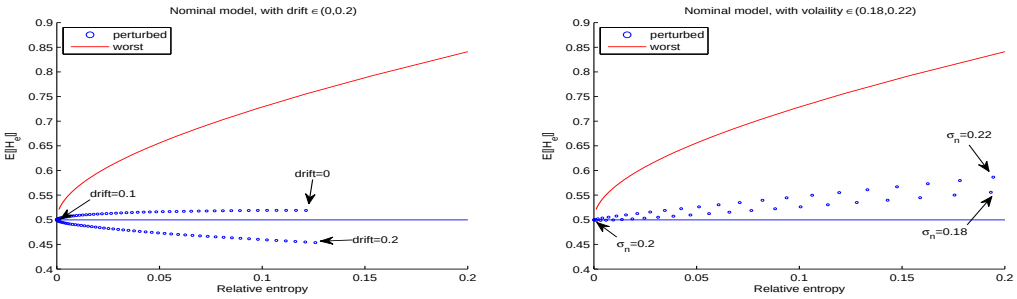


(c) Heston stochastic volatility model with different mean reversion speed κ_σ (left) and different long term limit μ_σ (right).

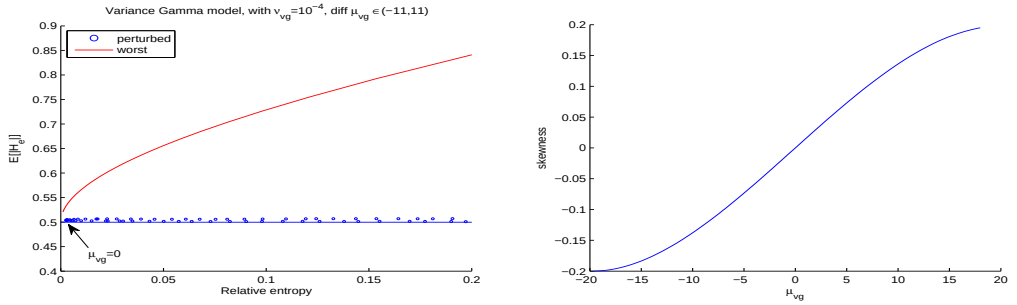


(d) Heston stochastic volatility perturbation with different volatility of volatility σ_σ (left) and correlation of diffusion ρ (right).

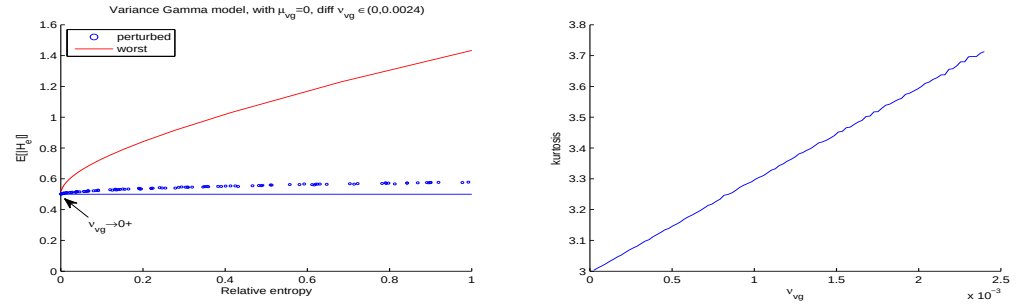
Figure 2.15: Hedging error under various changes in the underlying dynamics.



(a) Nominal model with changed drift (left) or volatility (right).



(b) Variance-Gamma model with varying θ_{vg} (left). Skewness vs θ_{vg} (right).

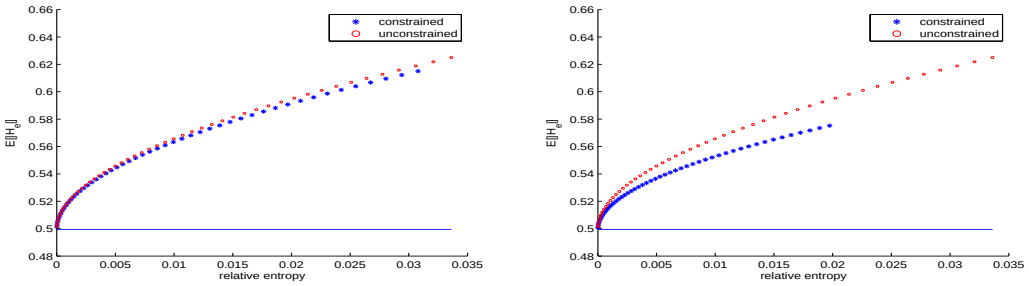


(c) Variance-Gamma model with varying ν_{vg}. Kurtosis vs ν_{vg} (right).

Figure 2.16: Hedging errors under various changes in the underlying dynamics.

2.8 Credit Valuation Adjustment (CVA)

Our final application of the robust risk measurement framework examines credit valuation adjustment (CVA), which has emerged as a key tool for quantifying counterparty risk among both market participants and regulators.



(a) Constraints on means of realized mean and realized variance

(b) Constraints on mean and variance of realized variance.

Figure 2.17: The blue dots are for constraint cases, and the red dots are for the unconstrained case.

2.8.1 Background on CVA

CVA measures the cost of a counterparty's default on a portfolio of derivatives. Rather than model each derivative individually, we will work with a simplified model of the aggregated exposure between two parties. We model this aggregated exposure as an Ornstein-Uhlenbeck process X_t,

$$dX_t = \kappa_x(\mu_x - X_t)dt + \sigma_x dW_t^x. \tag{2.8.1}$$

This allows the aggregated exposure to be positive for either party (and thus negative for the other); we can think of the two parties as having an ongoing trading relationship so that new swaps are added to their portfolio as old swaps mature, keeping the dynamics stationary. Alternatively, we can take X as a model of the exposure for a forward contract on a commodity or FX product where the underlying asset price is mean-reverting.

The time-to-default for the counterparty is modeled through a stochastic default intensity λ_t,

which follows a CIR-jump process

$$d\lambda_t = \kappa_\lambda(\mu_\lambda - \lambda_t)dt + \sigma_\lambda\sqrt{\lambda_t}dW_t^\lambda + dJ_t,$$

where W^x and W^λ are Brownian motions with correlation ρ, and J_t is a compound Poisson process with jump intensity ν_j and jump sizes following an exponential distribution with mean $1/\gamma$. The long-run limit of X matches the initial value, $X_0 = \mu_x$, and similarly $\lambda_0 = \mu_\lambda$. As in Zhu and Pykhtin (2007), the CIR-jump model guarantees that $\lambda_t \geq 0$.

Given the default intensity process, the time of default τ is

$$\tau = \Lambda^{-1}(\xi), \quad \text{where } \Lambda(t) = \int_0^t \lambda_s\, ds \quad \text{and} \quad \xi \sim Exp(1), \tag{2.8.2}$$

meaning that ξ has a unit-mean exponential distribution and is independent of everything else. The CVA for a time horizon T is then given by

$$CVA = (1-R)E[e^{-r\tau}I_{\tau<T}\max(X_\tau, 0)],$$

where R is the recovery rate. In other words, the loss at default of the counterparty is $(1 - R)\max(X_\tau, 0)$, and we take the expected present value of this loss on the event $\{\tau < T\}$ that the default occurs within the horizon. (This is a unilateral CVA, because we have included the default time of only one of the two parties.) We will study how model uncertainty affects the CVA.

In the following example, we set parameters at $T = 2$ years and divide the time horizon evenly into $N_T = 200$ steps, corresponding to around two periods per week. The risk-free rate is $r = 0.02$,

the recovery rate is $R = 0.3$, the long-run limit of the exposure is $\mu_x = 0$, the long-run limit of the default intensity is $\mu_\lambda = 0.02$, the exposure volatility is $\sigma_x = 0.2$, the default intensity has volatility $\sigma_\lambda = 0.2$, and the mean reversion coefficient $\kappa_x = \kappa_\lambda = 1$, which corresponds to a half-life of about 1.4 years. For simplicity, we initially omit jumps in the intensity.

We have the freedom to choose the units of X to fit the context. For example, if the volatility is 0.1 million dollars, we can measure X in multiple of a half million dollars to get $\sigma_x = 0.2$. Alternatively, suppose the underlying exposure is that of a netted portfolio of swaps with notional value 0.111 billion dollars, 10-year maturity and quarterly payments. If the interest rate is roughly constant, then the change in the value in the early years is roughly proportional to the change in the swap rate, or about $0.111 \sum_{i=1}^{40} e^{-ri\Delta t} \Delta S \Delta t = \Delta S$ billion dollars with swap rate S. Then we can model the change in swap rate using dynamics similar to (2.8.1) with $\kappa_s = \kappa_x$, $S_0 \geq 0$ and $\sigma_s = \sigma_x$, which corresponds to 20% volatility for the swap rate.

We apply our robust Monte Carlo approach to measure model risk. In this application, it is essential that the distribution of ξ in (2.8.2) remain unchanged: the adversary can change the dynamics of the default intensity (as well as the exposure), but having ξ be a unit-mean exponential in (2.8.2) is part of what it means for λ to be the default intensity, so this element is not subject to model error.

We enforce this condition through the method in Section 2.3.3. We simulate $N = 10^4$ sample paths for X and λ, and use $N_\xi = 10^4$ samples of ξ. For each realization of ξ, all N paths of λ are generated using (2.8.2), yielding a total of $N \times N_\xi$ paths. (Paths of X and λ are generated using

an Euler approximation.) For path (X^i, λ^i) and given ξ,

$$\begin{aligned} \hat{m}^*(X^i, \lambda^i, \xi) &= \frac{\exp(\theta V(X^i, \lambda^i, \xi))}{\sum_{j=1}^N \exp(\theta V(X^j, \lambda^j, \xi))/N} \\ \text{where } V(X^i, \Lambda^i, \xi) &= (1-R)e^{-r\tau} I_{\tau<T} \max(X^i_\tau, 0) \\ \text{and } \tau &= \Lambda_i^{-1}(\xi),\ \Lambda_i(t) = \int_0^t \lambda_i(s)ds. \end{aligned} \tag{2.8.3}$$

We call $V(X^i, \Lambda^i, \xi)$ the realized CVA for sample (X^i, Λ^i, ξ).

2.8.2 Analysis of the Worst-Case Model Error

We use the simulation results to examine the worst-case model. As a first step, we estimate values for some parameters to see how these parameters are affected by the change of measure.

We consider three cases for our experiments: $\rho = 0.3$, $\rho = 0$, and $\rho = -0.3$, the first of these corresponding to wrong-way risk because it makes default more likely when the exposure is large. These values of ρ are parameters to the nominal model; the nominal model is then distorted by the change of measure, and we re-estimate the correlations and other parameters. For example, to estimate the kth moment of the increments of W_x, we use

$$\hat{\mu}^x_k = \frac{\sum_{i=1}^N \sum_{j=1}^{N_\xi} \sum_{t=1}^{N_{ij}} \left(\Delta W_t^{x,i}\right)^k m^*(X^i, \lambda^i, \xi_j)}{\sum_{i=1}^N \sum_{j=1}^{N_\xi} N_{ij}},$$

where N_{ij} is the number of steps until the default, with $N_{ij} = N$ if no default occurs within the horizon.

The results are summarized in Table 2.7. The columns show estimates for different values of

θ, with $\theta = 0$ corresponding to the nominal value. Positive larger θ corresponds to possible larger CVA in the worst-case scenario, while negative smaller θ corresponds to possible smaller CVA. We first report estimates for CVA and, just below each value, CVA as a percentage of the notional 0.111. The impact of model error is illustrated through the range of values across different values of θ. The impact is asymmetric, with positive θ values having a greater effect than smaller θ values, particularly at larger values of ρ. This is at least partly explained by controlling for differences in relative entropy $\mathcal{R}(m_\theta)$.

		θ								
nominal ρ		-12	-9	-6	-3	0	3	6	9	12
-0.3	CVA	3.98	4.80	5.86	7.26	9.14	11.67	15.14	19.90	26.46
	$\times 10^4$	(0.36%)	(0.43%)	(0.53%)	(0.65%)	(0.82%)	(1.05%)	(1.36%)	(1.79%)	(2.38%)
0		5.13	6.27	7.78	9.82	12.60	16.45	21.84	29.45	40.23
		(0.46%)	(0.56%)	(0.70%)	(0.88%)	(1.14%)	(1.48%)	(1.97%)	(2.65%)	(3.62%)
0.3		6.34	7.82	9.81	12.49	16.17	21.25	28.31	38.07	51.36
		(0.75%)	(0.70%)	(0.88%)	(1.13%)	(1.46%)	(1.91%)	(2.55%)	(3.42%)	(4.63%)
-0.3	$\mathcal{R}(m_\theta)$	2.53	1.68	0.89	0.27	0.00	0.40	1.99	5.60	12.53
0	$\times 10^3$	3.61	2.42	1.30	0.39	0.00	0.61	3.08	8.86	20.27
0.3		4.73	3.18	1.71	0.52	0.00	0.80	4.04	11.44	25.49
-0.3	ρ	-0.299	-0.299	-0.299	-0.299	-0.299	-0.299	-0.299	-0.299	-0.299
0		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.3		0.299	0.299	0.299	0.299	0.299	0.299	0.299	0.299	0.299
-0.3	σ_x	0.200	0.200	0.200	0.200	0.200	0.200	0.200	0.200	0.201
0		0.200	0.200	0.200	0.200	0.200	0.200	0.200	0.201	0.201
0.3		0.200	0.200	0.200	0.200	0.200	0.200	0.200	0.201	0.201
-0.3	σ_λ	0.201	0.201	0.201	0.201	0.201	0.201	0.201	0.201	0.201
0		0.201	0.201	0.201	0.201	0.201	0.201	0.201	0.201	0.202
0.3		0.200	0.201	0.201	0.201	0.201	0.201	0.201	0.202	0.202
-0.3	drift of $\sigma_x W^x$	-1.27	-1.06	-0.79	-0.43	0.05	0.70	1.59	2.83	4.56
0	$\times 10^4$	-1.27	-1.06	-0.79	-0.43	0.05	0.70	1.59	2.83	4.56
0.3		-1.73	-1.47	-1.13	-0.67	-0.06	0.76	1.89	3.43	5.49
-0.3	drift of $\sigma_\lambda W^\lambda$	-0.48	-0.39	-0.27	-0.12	0.07	0.32	0.66	1.13	1.78
0	$\times 10^4$	-0.48	-0.39	-0.27	-0.12	0.07	0.32	0.66	1.13	1.78
0.3		-0.96	-0.80	-0.60	-0.34	0.014	0.49	1.12	1.99	3.12

Table 2.7: Worst-case results and parameters for CVA example.

The estimates of ρ, σ_x, and σ_λ are consistently close to their nominal values and thus unaffected by the change of measure. The means of the scaled Brownian motions $\sigma_x W^x$ and $\sigma_\lambda W^\lambda$, both of which are zero under the nominal measure, are increasing in θ, though the magnitude of change is small. In short, the wide range of CVA values are all consistent with nearly the same parameters values; changes in parameter values are not the primary source of model risk.

Next, we consider changes in the marginal distributions of X_τ and τ, considering only outcomes in which $\tau < T$. The upper panels of Figure 2.18 show the marginal density of X_τ. At $\theta = 12$, the adversary is trying to increase the CVA, so the density is shifted to the right; setting $\theta = -12$ has the opposite effect. The lower panels show the cumulative distribution of τ. Here, a larger θ value makes default more likely within the horizon (thus increasing the CVA), whereas a smaller θ makes default less likely.

The most interesting aspect of the worst-case change of measure is the effect on the dependence between τ and X_τ. We can get a first indication of this dependence from the correlations estimated at different θ values reported in Table 2.8. The correlations consistently increase with θ.

θ	-12	-9	-6	-3	0	3	6	9	12
$\rho_{X_\tau,\tau}$	-0.075	-0.045	-0.012	0.025	0.068	0.116	0.178	0.263	0.390
$\rho_{X_\tau,\xi}$	0.039	0.064	0.093	0.125	0.159	0.200	0.248	0.313	0.413

Table 2.8: Correlations between X_τ and τ, conditional on $\tau < T$.

To further examine the dependence, in Figure 2.19 we plot contours of the joint density of τ and X_τ for different values of θ, taking $\rho = 0.3$. Despite this correlation in the driving Brownian motions, we do not observe much dependence between X_τ and τ in the nominal case $\theta = 0$ (upper right). At $\theta = 12$, we see a marked increase in dependence. We also see that the most likely way to get a large realized CVA is to have a default toward the end of the horizon, after the exposure

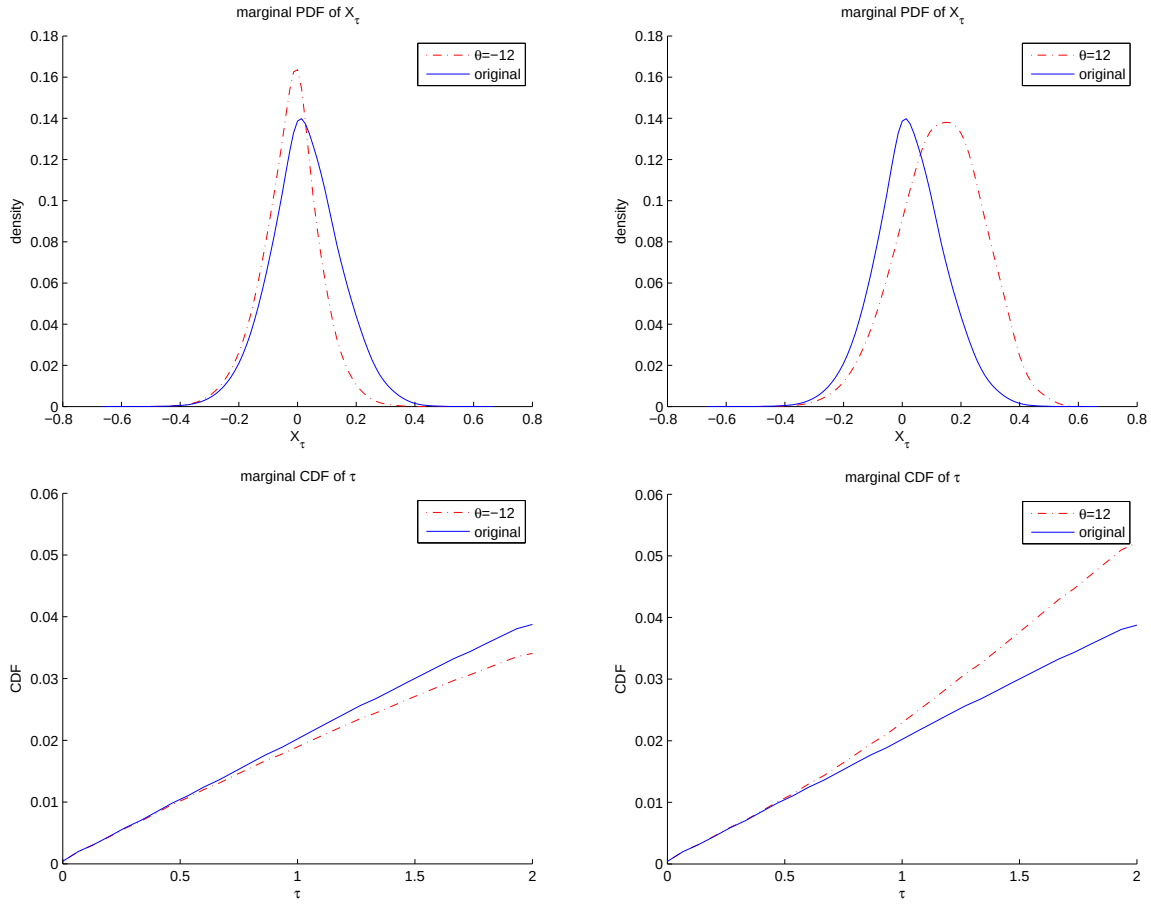


Figure 2.18: Marginal distribution of X_τ and τ for $\tau < T$.

has a chance to accumulate. In other words, the least costly way for the adversary to generate a large CVA is to push τ toward T and push X upward. In the lower right corner, we show the joint distribution obtained using the parameter values estimated at $\theta = 12$. This once again supports the view that the change in parameter values does not capture the most important features of the worst-case model. The case $\theta = -12$ in the upper left shows some negative dependence between τ and X_τ; here the adversary tries to generate a small CVA with a quick default near $X = 0$ or no default at all.

In Figure 2.20, we plot contours of the joint density of (ξ, X_τ) for the same cases that appear

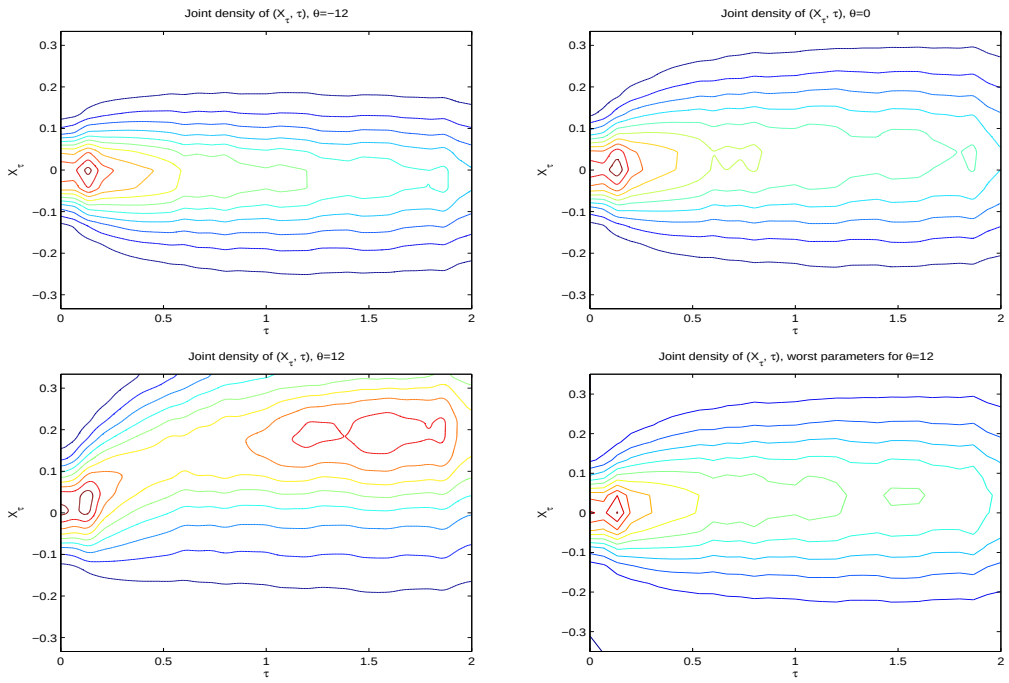


Figure 2.19: Joint density of X_τ and τ for $\tau < T$, $T = 2$.

in Figure 2.19. These are consistent with the pattern in Figure 2.19, but recall that the marginal distribution of ξ does not change so the pattern here is more purely determined by the change in dependence. This is further illustrated in Figure 2.21, which shows contours of the copula for X_τ and ξ.

In Figure 2.22, we revisit the comparisons of Figures 2.18–2.20, except now we constrain the change of measure to leave the marginal law of X unchanged. The effect is to force a much greater change in the dependence structure since the adversary has less flexibility to change the marginals.

As another perspective on the worst-case change of measure, in Figure 2.23 we plot some statistics of the Brownian increments $\sigma_x W^x$ and $\sigma_\lambda W^\lambda$ on paths with defaults. Each plot starts up to 100 steps (1 year) before the default. The horizontal axis is the time remaining until default, so the origin corresponds to the time of default and 1 corresponds to 1 year before default.

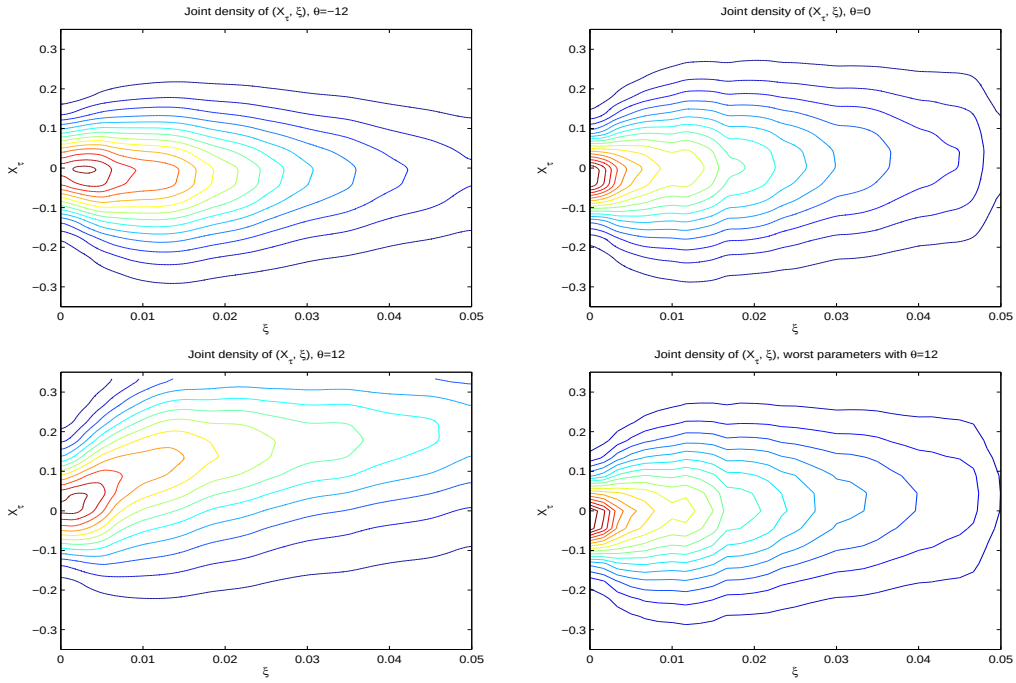


Figure 2.20: Joint density of X_τ and ξ for $\tau < T$, $T = 2$.

With $\theta = 12$, the worst-case means and standard deviations of the increments are significantly higher than their original values within 100 steps. Further from default, the abnormality of means decreases. This may explain why parameters estimated using all the increments under the worst-case scenario are not so different from the original parameter values. In estimating worst-case parameters using all the increments, the effect of the abnormal increments is diluted by other increments whose distributions are much less perturbed.

Interestingly, the standard deviations increase as we move further away from default. This seems to be a consequence of the fact that closer to default there is a strong upward trend with reduced volatility for both λ and X.

In Table 2.9, we estimate CVA at different values of θ and then at different parameter values. For example, to match the CVA at $\theta = 6$, we would need to make dramatic changes in the input

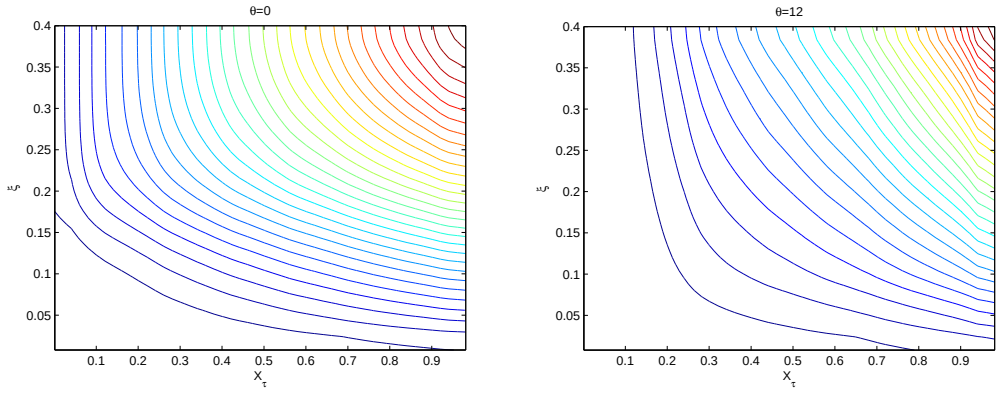


Figure 2.21: Copula of (X_τ, ξ).

parameters — increasing ρ to 0.95 or increasing ρ together with μ_λ. Once again we find that the worst-case model error is not simply described by a change in parameters.

	$T = 2$	
	CVA	$\rho_{X_\tau,\tau}$
worst parameters $\theta = 12$	0.0016	0.0695
$\theta = 0$	0.0016	0.0675
$\theta = 3$	0.0021	0.1164
$\theta = 6$	**0.0028**	**0.1777**
$\theta = 9$	0.0038	0.2634
$\theta = 12$	**0.0051**	**0.3903**
$\rho = 0.4$	0.0018	0.0884
$\rho = 0.95$	0.0027	0.2083
$\mu_\lambda = 0.04$	0.0027	0.0271
$\mu_\lambda = 0.06$	0.0031	-0.0169
$\mu_\lambda = 0.025$, $\rho = 0.65$	0.0025	0.1231
$\mu_\lambda = 0.025$, $\rho = 0.9$	**0.0030**	**0.1722**
$\mu_\lambda = 0.05$, $\rho = 0.9$	0.0037	-0.0010

Table 2.9: CVA and $\rho_{X_\tau,\tau}$ using parameters estimated from the worst-case scenario, the worst-case scenario, and scenarios with perturbed parameters.

By simply increasing ρ, both CVA and $\rho_{X_\tau,\tau}$ increase. However, the increase in $\rho_{X_\tau,\tau}$ is greater than the increase in CVA compared to the worst scenario; moreover, even with a very extreme value like $\rho = 0.95$, the changes in CVA and $\rho_{X_\tau,\tau}$ are limited. When only μ_λ is increased, the

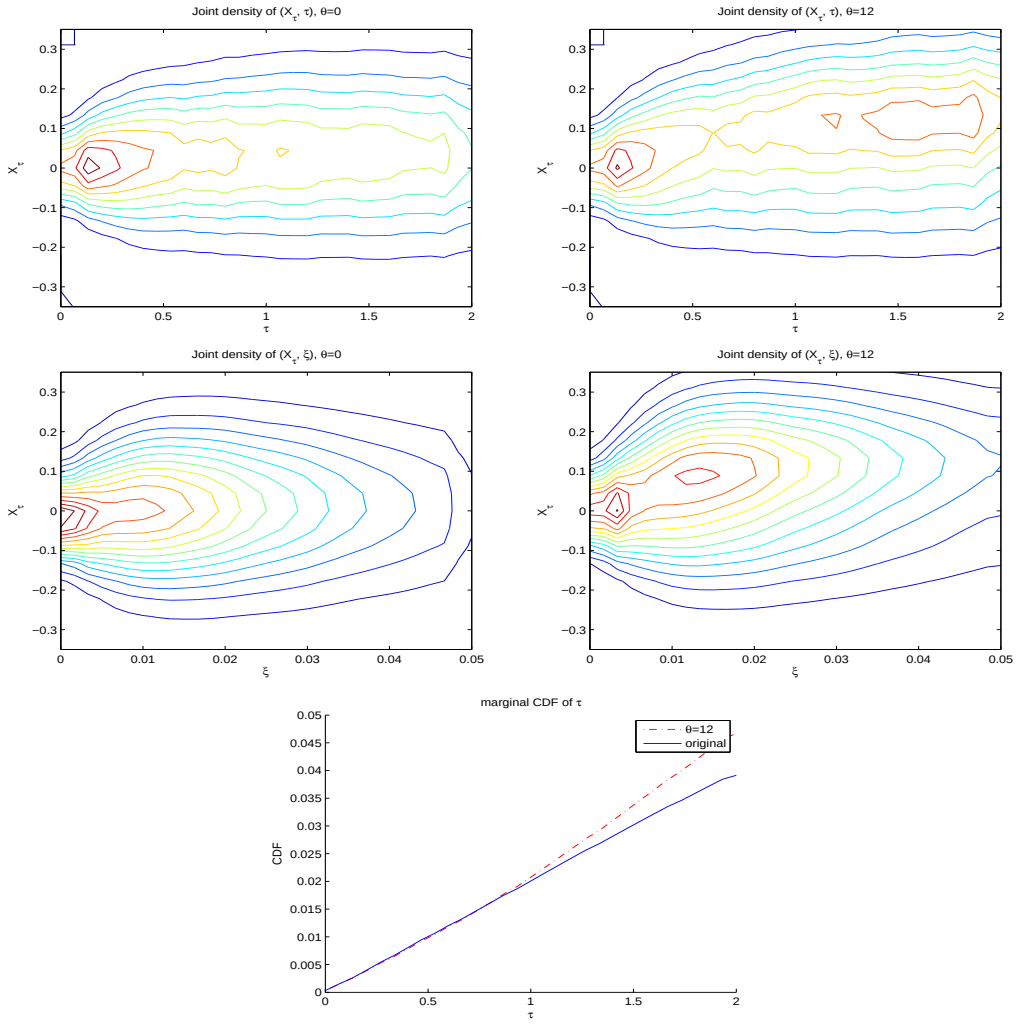


Figure 2.22: Marginal densities of τ and X_τ(first row), joint density of (X_τ, τ) (second row) and joint density of (X_τ, ξ) (third row). The dynamics of X are fixed.

CVA increases, but when $\mu_\lambda \geq 0.06$, $\rho_{X_\tau,\tau}$ turns negative. A possible explanation is that as μ_λ becomes very large, those paths with small realizations of W^λ also default before T, contributing small values of X_τ, hence a smaller $\rho_{X_\tau,\tau}$.

In order to reach the level of the worst-case CVA and $\rho_{X_\tau,\tau}$, we need to have $\mu_\lambda \approx 0.025$ and $\rho = 0.9$ to reach the level at $\theta = 6$. For the level at $\theta = 12$, we can set $\mu_\lambda \approx 0.07$ to reach the level of CVA, but we cannot reach a similar level for $\rho_{X_\tau,\tau}$ even with very high correlation $\rho = 0.8$.

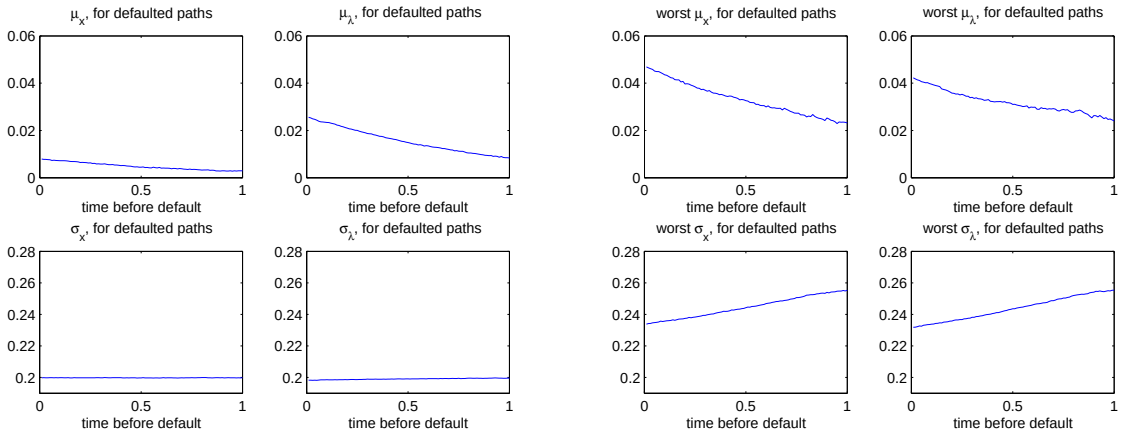


Figure 2.23: Statistics for increments before defaults using $\theta = 0$ (left) and $\theta = 12$ (right).

Compared to Figure 2.19, Figure 2.22 shows much less distortion in the joint density of (X_τ, τ) and (X_τ, ξ), and the change of dependence shows up later in the horizon. With the dynamics of X fixed, the adversary's only control is through the default intensity λ, trying to make the default occur when X has a larger value. Early in the horizon, X typically has small values, so the adversary chooses not to expend relative entropy early. Hence, the perturbed distribution of τ is similar to what it was before early in the horizon.

We have also tested the case with jumps in the dynamics of λ, with parameters $\nu_j = 1.5$ and $\gamma = 0.01$ from El Bachir and Brigo (2008) and other parameters are unchanged. The results are very similar to what we had before, except that in Figure 2.23, the dynamics of W_x and W_λ have very minor changes even before defaults.

In Figure 2.24, we plot the worst-case jump intensity and the worst-case mean jump sizes. Both increase with θ, but the magnitudes of the changes are small.

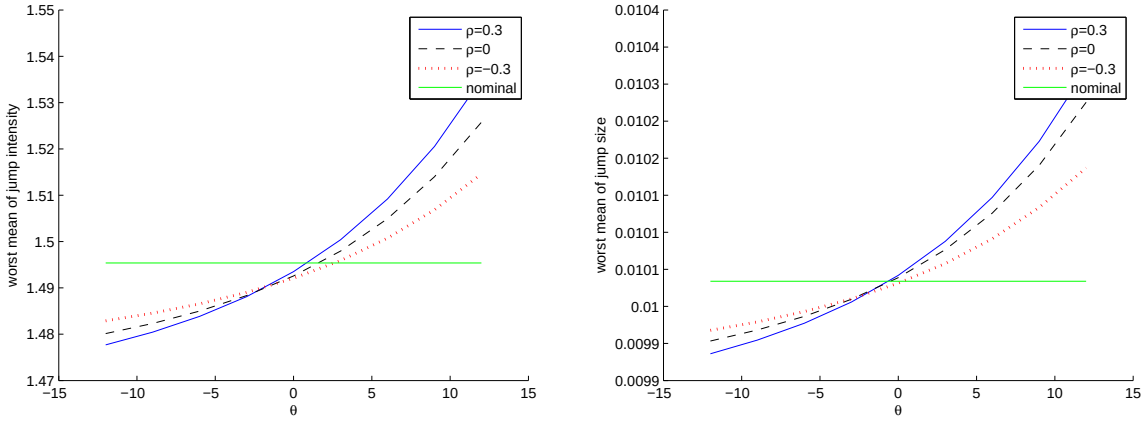


Figure 2.24: Worst-case jump intensity and worst-case mean jump size.

2.9 Concluding Remarks

This chapter develops a general approach and specific tools for quantifying model risk and bounding the impact of model error. By taking relative entropy as a measure of "distance" between stochastic models, we get a simple representation of the worst-case deviation from a baseline nominal model — this worst-case deviation is characterized by an exponential change of measure. Applying this representation with simulation allows us to bound the effect of model error across multiple values of relative entropy with minimal computational effort beyond that required to simulate the baseline nominal model alone. We have also shown how to incorporate additional information into the analysis to impose constraints on moments and other auxiliary functions of the underlying model or to leave certain marginal distributions of the underlying model unchanged; and we have extended these ideas to heavy-tailed distributions with α-divergence replacing relative entropy.

Using these tools, we have examined model error in mean-variance portfolio optimization, conditional value-at-risk, the Gaussian copula model of portfolio credit risk, delta hedging, and

credit valuation adjustment. A recurring theme in these examples is that the worst-case model deviation generally looks very different from a change of parameters within the baseline nominal model. Thus, our approach based on stochastic robustness goes well beyond parameter sensitivity in exploring model error to identify the greatest vulnerabilities in the stochastic structure of a model.

Chapter 3

Robust Portfolio Control with Stochastic Factor Dynamics

3.1 Introduction

Mean-variance portfolio optimization is known for its sensitivity to small perturbation. This is one of the main difficulty faced by its practical application. The natural extension to multiple horizons not only inherits this unpleasant pitfall, it also faces the challenge of evolution of the future market. In this chapter, tools introduced in Chapter 2 are used to construct a portfolio control that is robust towards model risk.

Our model builds on a practical setting, where the returns of assets depend on factors following some times-series model. The objective follows mean-variance framework while the transaction costs are subtracted from gross returns. We assume both the relationship between returns and factors and the evolution of the factors are subject to model error, and derive optimal controls that

optimize the worst-case objective provided that the true model is relatively close to the nominal one, measured by relative entropy. Thus, it is robust with respect to moderate model risk. The original problem (without robustness) can be solved via Linear-Quadratic-Gaussian programming. With the extra consideration of robustness, a closed-form value function iteration is derived and the stability for infinite-horizon problem is studied. Numerical results shows improvement from the consideration of robustness, which comes from the reduced risk exposure relative to the baseline model.

The rest of this chapter is organized as follows. Section 3.2 formulates the basic portfolio control problem and its robust extension. Section 3.3 solves the finite-horizon problem. Section 3.4 examines the effect of varying the degree of robustness and compares robust and non-robust solutions. Section 3.5 solves the infinite-horizon control problem, and Section 3.6 presents numerical results. Most proofs are collected in Appendix A.

3.2 Problem Formulation

3.2.1 Dynamics and Objective

We consider a portfolio optimization problem in which asset returns are driven by factors with stochastic dynamics. Examples of portfolio control problems with factor models of returns include, among many others, the work of Bielecki and Pliska (1999), Campbell and Viceira (2002), and Pesaran and Timmermann (2012). The formulation in Garleanu and Pedersen (2012), which we now review, leads to particularly explicit solutions in both its original and robust form.

The investor has access to n_x underlying assets evolving in discrete time. The changes in prices

of the assets from time t to time $t+1$ are indicated by a vector $r_{t+1} \in \mathbb{R}^{n_x}$, specified by

$$r_{t+1} = \mu + B_f f_t + u_{t+1}, \tag{3.2.1}$$

where $\mu \in \mathbb{R}^{n_x}$ represents an expected or "fair" return, $f_t \in \mathbb{R}^{n_f}$ is a vector of factors influencing price changes and known to the investor at time t, $B_f \in \mathbb{R}^{n_x \times n_f}$ is a factor loading matrix, and $u_1, u_2, \ldots$, are i.i.d. random vectors in $\mathbb{R}^{n_x}$ following a multivariate normal distribution with mean zero and covariance matrix Σ_u. The factors are mean-reverting, evolving according to the equation

$$f_{t+1} = C_f f_t + v_{t+1}, \tag{3.2.2}$$

with coefficient matrix $C_f \in \mathbb{R}^{n_f \times n_f}$ and i.i.d. noise vectors $v_1, v_2, \ldots$ in $\mathbb{R}^{n_f}$ following a multivariate normal distribution with mean zero and covariance matrix Σ_v. We assume that the v_t are independent of the u_t. To make the factors stable, we assume throughout that $\sigma(C_f) < 1$, where $\sigma(\cdot)$ gives the spectral radius of a square matrix. Equations (3.2.1) and (3.2.2) allow the possibility that prices become negative. However, we measure performance based on price changes, so this does not present a problem. The probability of this occurring can also be made very small through parameter choices.

Denote by $x_t \in \mathbb{R}^{n_x}$ the vector of shares of underlying assets held in the portfolio just after any transactions made at time t; in other words, at time t the portfolio's holdings are rebalanced from

x_{t-1} to x_t. Rebalancing the portfolio imposes transactions costs modeled as

$$\frac{1}{2}\Delta x_t^\top \Lambda \Delta x_t, \text{ where } \Delta x_t = x_t - x_{t-1}, \tag{3.2.3}$$

with cost matrix Λ symmetric and positive definite. For a square matrix A, we write $A > 0$ (or $\geq$, $<$, $\leq$) if A is positive definite (or positive semi-definite, negative definite, negative semi-definite, respectively). If a small transaction of dx shares temporarily moves the market price unfavorably by the amount Λdx, then a transaction of size Δx_t results in a total cost of $\Delta x_t^\top \Lambda \Delta x_t/2$, compared to executing the transaction at the original price. The simple model penalizes large trades and provides tractability.

In a mild abuse of notation, we use Δx to denote an investment *policy* — that is, a rule for determining transactions given the information available. With this convention, we write the objective introduced by Garleanu and Pedersen (2012), which models an investor seeking to maximize the present value of risk-adjusted excess gains, net of transaction costs, as follows:

$$\begin{aligned}
&\sup_{\Delta x} E\left[\sum_{t=0}^{T} \beta^t \left((x_{t-1} + \Delta x_t)^\top (r_{t+1} - \mu) - \frac{\gamma}{2} Var_t[(x_{t-1} + \Delta x_t)^\top r_{t+1}] - \frac{1}{2}\Delta x_t^\top \Lambda \Delta x_t\right)\right] \\
&= \sup_{\Delta x} E\left[\sum_{t=0}^{T} \beta^t \left(x_t^\top (B_f f_t + u_{t+1}) - \frac{\gamma}{2} x_t^\top \Sigma_u x_t - \frac{1}{2}\Delta x_t^\top \Lambda \Delta x_t\right)\right].
\end{aligned} \tag{3.2.4}$$

The objective (3.2.4) consists of three terms. The first term is the discounted sum of future excess returns with discount factor $\beta \in (0, 1)$. The third term measures discounted transaction costs. The difference between these two terms measures the discounted net cash flow to the investor. The middle term is a risk penalty, in which $\gamma > 0$ measures the investor's risk aversion. The notation

Var_t denotes conditional variance of excess return, given information up to time t, including position x_t, and we use E_t analogously. Measuring risk through the expectation of the discounted sum of conditional variances is a compromise made for tractability, as is the case with the quadratic measure of transaction costs. This also makes the objective time consistent, which is not typically true for dynamic mean-variance problems (e.g., Basak and Chabakauri (2010)). Interestingly, even if we drop the risk penalty (setting $\gamma = 0$) a term of exactly this form appears in the solution to a robust formulation. We view (3.2.4) as a guide to selecting sensible strategies rather than as a precise representation of an investor's preference. In our numerical tests, we therefore evaluate performance through a Sharpe ratio as well as directly through (3.2.4).

Given the Markov structure of the problem, it suffices for the investor to consider policies under which Δx_t is a deterministic function of (x_{t-1}, f_t), and the supremum in (3.2.4) is taken over such policies. In choosing an optimal policy, the investor must, as usual, balance risk and reward. In addition, the combination of the factor structure in (3.2.1) and the mean-reversion in (3.2.2) requires the investor to balance the benefits of acting on a signal before the factors decay against the costs of large transactions.

Remark 3.2.1. Our model of transaction costs can be generalized to incorporate more features while preserving tractability. One generalization is to incorporate permanent price impact by adding $x_t^\top \tilde{\Lambda} \Delta x_t$ to the transaction costs, where $\tilde{\Lambda} \Delta x_t$ is the permanent price impact caused by transaction Δx_t. Because this term is linear in Δx_t, we still get an explicit iteration similar to Proposition 3.3.6. Moallemi and Saglam (2012) consider more general models with transaction costs and more general performance objectives than (3.2.4); they forego explicit solutions and instead optimize numerically within the class of linear rebalancing rules.

3.2.2 Robust Formulation

We now introduce model uncertainty by allowing perturbations in the stochastic dynamics of the model. The stochastic input to the model is the sequence $\{(u_t, v_t), t = 1, 2, \dots\}$ of noise terms, so we will translate uncertainty about the model into uncertainty about the law of this sequence. As is usually the case in discussions of robustness, it is convenient to describe uncertainty through the possible actions of a hypothetical adversary who changes the model to maximize harm to the original agent — in our setting, the investor. We will constrain the actions of the adversary shortly, but first we briefly illustrate the effect the adversary can have by changing the law of the noise sequence.

The noise vectors all have mean zero under the original model. If the adversary changes the conditional mean of v_{t+1} to $-Df_t$, for some $n_f \times n_f$ matrix D, then (3.2.2) becomes

$$f_{t+1} = (C_f - D)f_t + \tilde{v}_{t+1},$$

where $\tilde{v}_{t+1}$ has the law of the original v_{t+1}. Thus, the adversary can change the dynamics of the factors and, for example, accelerate the speed of mean-reversion and potentially reduce the value of the factors to the investor. By instead setting the conditional mean of v_{t+1} to $(I - C_f)\bar{f}$, for some fixed $\bar{f}$, the adversary moves the long-run mean of the factors from the origin to $\bar{f}$. Changing the conditional mean of u_{t+1} to $-\bar{B}_f f_t$ changes (3.2.1) to

$$r_{t+1} = \mu + (B_f - \bar{B})f_t + \tilde{u}_{t+1},$$

$\tilde{u}_{t+1}$ having the original distribution of u_{t+1}, and thus allows the adversary to change the factor loadings. By changing the covariance of u_{t+1}, the adversary changes the covariance of the price changes r_{t+1}. [1] The adversary can also introduce correlation between u_{t+1} and v_{t+1}. We will see that the adversary's optimal controls take advantage of dynamic information about both portfolio holdings and factor levels and thus go beyond robustness to uncertainty about static parameters like C_f and B_f.

These examples also serve to illustrate that alternative but equivalent formulations of the original control problem (3.2.4) can lead to distinct formulations when we consider robustness. For instance, replacing $x_t^\top(r_{t+1}-\mu)$ on the left side of (3.2.4) with its conditional expectation $E_t[x_t^\top(r_{t+1}-\mu)] = x_t^\top B_f f_t$ clearly has no effect on the investor's portfolio choice or its performance as measured by (3.2.4). However, by including u_{t+1} in the objective we allow the adversary to influence performance by changing the distribution of this term. Put differently, including this term leads the investor to a strategy that is robust to errors in both the return model (3.2.1) and the factor dynamics (3.2.2), whereas omitting u_{t+1} focuses robustness exclusively on the factor dynamics. We solve and test both formulations.

We now formulate the adversary's actions more precisely. Let g_u and g_v denote the probability densities of u_t and v_t. The adversary may choose a new joint density $\tilde{g}_t$ for (u_t, v_t), which could, in the most general formulation, depend on past values (u_s, v_s), $s < t$, of the noise sequence. However, we will restrict our analysis to the Markovian case in which any dependence of the

[1] The uncertainty introduced by u_{t+1} can potentially cause the factor loading model to deviate from Arbitrage Pricing Theory (APT). We can enforce APT constraint by imposing linear constraints on the conditional expectation of u_{t+1}, i.e., $E_t[u_{t+1}] = \Omega\lambda_t + \Omega f_t$ for given f_t, where λ_t is the vector of risk premium of factors f_t, and Ω represents the degree of possible deviation of conditional expectation of u_{t+1}. As a special case, we can also assume the uncertainty only comes from v_{t+1}. This alternative approach will be tested in the numerical section.

density function $\tilde{g}_t$ on the past is captured through dependence on the state (x_{t-1}, f_t). If we set

$$m_t = \frac{\tilde{g}_t(v_t, u_t | x_{t-1}, f_t)}{g_v(v_t) g_u(u_t)}, \quad M_t = \prod_{s=1}^{t} m_s, \tag{3.2.5}$$

then M_t is the likelihood ratio relating the distribution of $(u_1, v_1, \ldots, u_t, v_t)$ selected by the adversary to the original distribution. Because g_u and g_v are multivariate normal densities, so the denominator of m_t is supported on all of $\mathbb{R}^{n_x} \times \mathbb{R}^{n_x}$ and is never zero.

As in Hansen and Sargent (2007), Hansen et al. (2006) and Petersen et al. (2000), we limit the adversary by constraining or penalizing the relative entropy of the change of measure. How tightly we constrain or penalize the relative entropy determines the degree of model uncertainty by limiting how far the adversary can change the stochastic evolution of the data away from the investor's model. The relative entropy at time t is $E[M_t \log M_t]$, which is always positive and is equal to zero only when the adversary leaves the original measure unchanged by taking $M_t \equiv 1$. Given M_0 and a sequence of one-period likelihood ratios $m = \{m_t, t = 1, 2, \ldots\}$ as in (3.2.5), let

$$\mathcal{R}_\beta(m) = (1-\beta) \sum_{t=0}^{\infty} \beta^t E[M_t \log M_t] = \sum_{t=0}^{\infty} \beta^{t+1} E\left[M_t E_t[m_{t+1} \log m_{t+1}]\right] \tag{3.2.6}$$

denote the infinite-horizon discounted sum of relative entropy, where the term $1 - \beta$ is introduced to simplify the final expression. We can give the adversary a budget $\eta > 0$ and constrain the measure change to satisfy $\mathcal{R}_\beta(m) \leq \eta$. When truncated at a finite upper limit T, the two sums in (3.2.6) no longer coincide — the discount factor on the right would need to be replaced with $(\beta^{t+1} - \beta^{T+1})$, leading to a control problem that depends on both t and T, and not just on the

time-to-go $T-t$. To avoid this feature and to preserve consistency with the infinite-horizon case, we use the rightmost sum in (3.2.6), truncated at T, as our finite-horizon measure of discounted entropy.

Constraining the adversary's measure change to satisfy $\mathcal{R}_\beta(m) \leq \eta$ results in the robust control problem (for either a finite or infinite horizon)

$$\sup_{\Delta x} \inf_{m:\mathcal{R}_\beta(m)<\eta} E\left[\sum_t \beta^t M_t \left(x_t^\top (B_f f_t + u_{t+1}) - \frac{\gamma}{2} x_t^\top \Sigma_v x_t - \frac{1}{2}\Delta x_t^\top \Lambda \Delta x_t\right)\right]. \tag{3.2.7}$$

Here, the investor seeks to optimize performance in the face of model uncertainty by maximizing performance against the worst case stochastic perturbation to the original model, considering only perturbations that are sufficiently close to the original model to satisfy the relative entropy constraint. As before, the supremum is taken over policies under which each Δx_t is a deterministic function of (x_{t-1}, f_t); the infimum is taken over measure changes satisfying the relative entropy constraint and having the form in (3.2.5) in which each new density $\tilde{g}_t$ is determined by (x_{t-1}, f_t). Thus, (x_{t-1}, f_t) remains Markovian under any policy pair $(\Delta x, m)$.

The Lagrangian of the constrained problem (3.2.7) is a penalty problem with parameter $\theta > 0$

$$\begin{aligned}\sup_{\Delta x} \inf_{m} E\Bigg[\sum_{t=0}^{T} \beta^t M_t \Big(&x_t^\top (B_f f_t + E_t[m_{t+1} u_{t+1}]) - \frac{\gamma}{2} x_t^\top \Sigma_v x_t \\ &-\frac{1}{2}\Delta x_t^\top \Lambda \Delta x_t + \theta\beta E_t[m_{t+1}\log m_{t+1}]\Big)\Bigg],\end{aligned} \tag{3.2.8}$$

with the constraint $\mathcal{R}_\beta(m) < \eta$ replaced by an admissibility condition $\mathcal{R}_\beta(m) < \infty$. Hansen et al. (2006), Claim 5.4, establish the equivalence of constrained and penalized formulations through

convex duality under mild conditions, complementing a similar result in Petersen et al. (2000). We work directly with (3.2.8) because it is more amenable to explicit solution. This formulation also has an interpretation in terms of dynamic risk measures, in which case θ measures the investor's aversion to ambiguity; see Ruszczyński (2010).

Our restriction to investment policies and measure changes that are Markovian in the sense that their dependence on the past is fully captured by dependence on the state (x_{t-1}, f_t) does not change the value of (3.2.8) in the finite-horizon case, provided the measure changes satisfy a more general rectangularity condition; this is shown in Theorems 2.1–2.2 of Iyengar (2005) for discrete state spaces, but his argument applies here as well. The infinite-horizon problem raises stability issues, but Section 7.6 of Hansen and Sargent (2007) shows an analogous reduction to Markovian strategies, under modest technical conditions, for problems of the type we consider. We avoid a digression into these issues by limiting ourselves to Markovian strategies throughout.

3.3 Finite-Horizon Robust Problem

3.3.1 Robust Bellman Equation for U

To lighten notation, we define

$$Q(x, \Delta x, f) = x^\top B_f f - \frac{\gamma}{2} x^\top \Sigma_u x - \frac{1}{2} \Delta x^\top \Lambda \Delta x.$$

For a fixed horizon $T < \infty$, and any $0 \leq t \leq T$, the finite-horizon value function $U_{t,T}$ for the penalty problem (3.2.8) is given by

$$U_{t,T}(M_t, x_{t-1}, f_t) = \sup_{\Delta x} \inf_{m} E_t \left[\sum_{s=t}^{T} \beta^s M_s \left(Q(x_s, \Delta x_s, f_s) \right.\right.$$
$$\left.\left. + E_s[m_{s+1} u_{s+1}^\top] x_s + \theta \beta E_s[m_{s+1} \log m_{s+1}] \right) \right], \tag{3.3.1}$$

with (x_{-1}, f_0) fixed and $U_{T+1,T} = 0$. That $U_{t,T}$ is indeed a function of only (M_t, x_{t-1}, f_t) follows from our restriction to Markovian strategies $(\Delta x, m)$ for the investor and the adversary. In particular, under a fixed pair of policies, the conditional expectations inside the summation reduce to functions of (x_s, f_s), and x_s is a function of (x_{s-1}, f_s).

Define the one-step robust dynamic programming operator $\mathcal{T}$ acting on functions $h : \mathbb{R} \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_f} \to \mathbb{R}$ by

$$\mathcal{T}(h)(M, x, f) = \sup_{\Delta x_+} \inf_{m_+} M \left(Q(x_+, \Delta x_+, f) \right.$$
$$\left. + E[m_+ u_+^\top] x_+ + \beta E[\theta m_+ \log m_+] + \beta E[h(M m_+, x_+, f_+)] \right), \tag{3.3.2}$$

where $x_+ = x + \Delta x_+$, $f_+ = C_f f + v_+$, $v_+ \sim N(0, \Sigma_v)$, and $u_+ \sim N(0, \Sigma_u)$. The supremum is over $\Delta x_+ \in \mathbb{R}^{n_x}$ and the infimum is over m_+ of the form in (3.2.5). It is always feasible (and optimal) for the adversary to choose m_+ with finite relative entropy.

Proposition 3.3.1. *$U_{t,T}$ satisfies, for $0 \leq t \leq T$, the robust Bellman equation*

$$U_{t,T} = \mathcal{T}(U_{t+1,T}). \tag{3.3.3}$$

Proof. The proof is the same as that of Theorem 2.2 of Iyengar (2005), even though the setting there is a discrete state space. In particular, the rectangularity condition required there holds in our setting. (We detail the argument for the infinite-horizon case in the proof of Proposition 3.5.1.) □

We have taken $U_{T+1,T} = 0$ as our terminal condition for simplicity. If the underlying assets are futures contracts (the focus of Section 3.6), then we can interpret this condition as having all contracts mature at $T + 1$. Alternatively, we could assign $U_{T+1,T}$ a liquidation value for the portfolio, considering both asset prices at time $T + 1$ and the transactions costs incurred in selling off the portfolio's holdings. This formulation would require recording price levels (the cumulative sum of the price differences r_t) in the state vector, which could be done quite easily. The final portfolio is just the scalar product of x_{T+1} and the price vector, so this formulation remains within the linear-quadratic framework. We omit this extension for simplicity, particularly since it does not apply to the infinite-horizon problem.

3.3.2 Bellman Equation for V

To solve the Bellman equation (3.3.3), we will follow the approach in Hansen and Sargent (2007) and prove that $U_{t,T}$ can be decomposed as a product of M_t and a function of (x_{t-1}, f_t), which will

simplify (3.3.3) so that we can solve it analytically. First, we write U in the form

$$U_{t,T}(M_t, x_{t-1}, f_t) = M_t V_{t,T}(M_t, x_{t-1}, f_t), \tag{3.3.4}$$

taking this as the definition of $V_{t,T}$, since $M_t \in (0, \infty)$. Then (3.3.3) becomes

$$\mathcal{T}(M_t m_{t+1} V_{t+1,T})(M_t, x_{t-1}, f_t) = M_t V_{t,T}(M_t, x_{t-1}, f_t). \tag{3.3.5}$$

Recalling the definition of $\mathcal{T}$, we can divide both sides of (3.3.5) by M_t to get

$$\begin{aligned} V_{t,T}(M_t, x_{t-1}, f_t) = \sup_{\Delta x_t} \inf_{m_{t+1}} \{ & Q(x_t, \Delta x_t, f_t) \\ & + E_t[m_{t+1} u_{t+1}^\top x_t + \beta\theta m_{t+1} \log m_{t+1} + \beta m_{t+1} V_{t+1,T}(M_{t+1}, x_t, f_{t+1})] \} . \end{aligned} \tag{3.3.6}$$

Set $V_{T+1,T} = 0$. It now follows by induction that $V_{t,T}$ does not depend on M_t: Suppose this is true of $V_{t+1,T}$; then, under our Markovian restriction on strategies, the conditional expectations in (3.3.6) are functions of (x_{t-1}, f_t), and thus so is $V_{t,T}$.

Since $\theta > 0$, the term in the conditional expectation of (3.3.6) is convex in m_{t+1}, so we can solve the minimization problem through the first-order conditions, which leads to the optimal choice

$$m_{t+1}^* = \frac{\exp\left\{-\frac{1}{\theta}\left(V_{t+1,T}(x_t, f_{t+1}) + \frac{1}{\beta} x_t^\top u_{t+1}\right)\right\}}{E_t\left[\exp\left\{-\frac{1}{\theta}\left(V_{t+1,T}(x_t, f_{t+1}) + \frac{1}{\beta} x_t^\top u_{t+1}\right)\right\}\right]}. \tag{3.3.7}$$

This is a positive function of x_t, f_{t+1}, u_{t+1} and v_{t+1}, normalized to integrate to 1, so it has the form

required by (3.2.5). (We verify that the normalization in the denominator is finite in the case of interest as part of Theorem 3.3.8.) Substituting expression (3.3.7) into (3.3.6), we get the following recursion for V:

$$V_{t,T}(x_{t-1}, f_t) = \sup_{\Delta x_t \in \mathbb{R}^d} \left\{ Q(x_t, \Delta x_t, f_t) - \beta\theta \log E_t \left[\exp\{ -\frac{V_{t+1,T}(x_t, f_{t+1}) + \frac{1}{\beta} x_t^\top u_{t+1}}{\theta} \} \right] \right\}, \tag{3.3.8}$$

which we abbreviate as $V_{t,T} = \mathcal{T}^V(V_{t+1,T})$ by defining $\mathcal{T}^V$ as the operator on the right. We summarize this transformation as follows:

Proposition 3.3.2. *For any $T < \infty$, any solution to (3.3.8) gives a solution $U_{t,T}(M_t, x_{t-1}, f_t) = M_t V_{t,T}(x_{t-1}, f_t)$ to (3.3.3), where the adversary's choice is given by (3.3.7).*

The recursion in (3.3.8) has the form of a risk-sensitive optimal control problem. With $\beta = 1$, the recursion can be unwound and V expressed as the value function of a control problem; this case is treated extensively in Whittle (1981), Whittle and Whittle (1990). With $\beta < 1$, there is no non-recursive expression for V: V cannot be expressed as the value function for a control problem with a time-separable objective, nor is it equivalent to specifying an exponential utility function or any other standard utility function. This discounted case is treated in Hansen and Sargent (1995), though their convexity condition does not hold in our setting; see also Skiadas (2003). Portfolio optimization problems with risk-sensitive criteria are solved in, e.g., Bielecki and Pliska (1999), Bielecki et al. (2005), Fleming and Sheu (2001), with the risk-sensitive objective posited from the outset. It should be stressed that in, our setting, the risk-sensitive problem (3.3.8) emerges only as an intermediate step in solving the robust control problem, in response to the adversary's optimal strategy, and not as the primary objective.

3.3.3 Saddlepoint Condition and Solution to the Bellman Equation

Building on Whittle (1981), Whittle and Whittle (1990), we will give conditions leading to a quadratic solution for $V_{t,T}$. To motivate the argument, we first observe that starting with $V_{T+1,T} = 0$ and taking one step backward in (3.3.8) we find that $V_{T,T}$ is quadratic in (x_{T-1}, f_T). If, for some $t \leq T$, $V_{t+1,T}$ is quadratic in (x_t, f_{t+1}) with no dependence on M_t, then the last term in (3.3.8) becomes

$$-\theta\beta \log \left(E_t[\exp\{-\frac{1}{\theta}G\}] E_t[\exp\{-\frac{1}{\theta\beta}x_t^\top u_{t+1}\}] \right) = -\theta\beta \log E_t[\exp\{-\frac{1}{\theta}G\}] - \frac{1}{2\beta\theta}x_t^\top \Sigma_u x_t, \tag{3.3.9}$$

where G is a quadratic function of (x_t, f_{t+1}) (and thus of v_{t+1}), and the conditional expectation factors because of the independence of u_{t+1} under the original probability measure. Under the saddlepoint conditions given below, (3.3.9) then reduces to a quadratic function of (x_t, f_t). So, the right side of (3.3.8) is a quadratic function of x_t, f_t, and x_{t-1}. Maximizing over x_t under the saddlepoint conditions, the right side of (3.3.8) becomes a quadratic function of (x_{t-1}, f_t). Thus, $V_{t,T}$ is quadratic in (x_{t-1}, f_t) and has no functional dependence on M_t.

A consequence of these properties of $V_{t,T}$ is that it is quadratic in v_{t+1} — implying that v_{t+1} continues to be normally distributed under the change of measure, though with a different mean and covariance — and linear in u_{t+1} — implying that u_{t+1} continues to be normally distributed but with a different mean. The absence of a cross term multiplying v_{t+1} and u_{t+1} in m^*_{t+1} preserves the independence of the two vectors.

In light of the foregoing discussion, we posit the representation

$$V_{t,T}(x_{t-1}, f_t) = x_{t-1}^\top A_{xx}^{(t,T)} x_{t-1} + x_{t-1}^\top A_{xf}^{(t,T)} f_t + f_t^\top A_{ff}^{(t,T)} f_t + A_0^{(t,T)},$$

and set

$$A^{(t,T)} = \begin{bmatrix} A_{xx}^{(t,T)} & \frac{1}{2} A_{xf}^{(t,T)} \\ \frac{1}{2}(A_{xf}^{(t,T)})^\top & A_{ff}^{(t,T)} \end{bmatrix}.$$

Without loss of generality, we take $A^{(t,T)}$ to be symmetric. We introduce two conditions to ensure that this structure is preserved by the recursion (3.3.8). To state the conditions generically, we drop the superscript (t, T).

Condition 3.3.3. $\Sigma_v^{-1} + \frac{2}{\theta} A_{ff} > 0$.

Condition 3.3.4. $J_1 > 0$, *with* $\gamma_\theta = \gamma + (1/\theta\beta)$ *and*

$$J_1 = \gamma_\theta \Sigma_u + \Lambda - 2\beta A_{xx} + \frac{\beta}{\theta} A_{xf} (\Sigma_v^{-1} + \frac{2}{\theta} A_{ff})^{-1} A_{xf}^\top > 0. \tag{3.3.10}$$

By analogy with Whittle and Whittle (1990), pp.81–83, we call these saddlepoint conditions. The following lemma provides sufficient conditions for the required properties.

Lemma 3.3.5. *(i) If* $A_{xx} \leq 0$ *and Condition 3.3.3 holds, then Condition 3.3.4 holds. (ii) If* $A_{ff} \geq 0$*, then Condition 3.3.3 holds.*

The lemma helps explain the name "saddlepoint" and shows that the conditions we have are weaker than concavity in x and convexity in f. We now apply our conditions to the Bellman

equation (3.3.8). The result is similar to that of Hansen and Sargent (1995); however, a separate argument is needed because we do not have joint concavity. Given a matrix A, define

$$J_2 = B_f f_t + \Lambda x_{t-1} + \beta A_{xf} C_f f_t - 2\frac{\beta}{\theta} A_{xf}(\Sigma_v^{-1} + \frac{2}{\theta} A_{ff})^{-1} A_{ff} C_f f_t, \tag{3.3.11}$$

$$J_3 = B_f + \beta A_{xf} C_f - 2\frac{\beta}{\theta} A_{xf}(\Sigma_v^{-1} + \frac{2}{\theta} A_{ff})^{-1} A_{ff} C_f. \tag{3.3.12}$$

The matrix J_3 appears when we substitute a quadratic function into (3.3.8); the vector J_2 depends on the state (x_{t-1}, f_t) and will be used to describe the investor's optimal portfolio. The proof of the following result, and that of most results throughout this chapter, appears in Appendix A.

Proposition 3.3.6. *If a symmetric matrix A satisfies Conditions 3.3.3 and 3.3.4, then $\mathcal{T}^V$ maps a quadratic function with coefficients (A, A_0) to a quadratic function with coefficients $(\mathcal{S}(A), \mathcal{U}(A, A_0))$, where*

$$\mathcal{S}(A)_{xx} = -\frac{1}{2}\Lambda + \frac{1}{2}\Lambda J_1^{-1}\Lambda^\top, \tag{3.3.13}$$

$$\mathcal{S}(A)_{xf} = \Lambda J_1^{-1}\left[B_f + \beta A_{xf} C_f - 2\frac{\beta}{\theta} A_{xf}(\Sigma_v^{-1} + \frac{2}{\theta} A_{ff})^{-1} A_{ff} C_f\right], \tag{3.3.14}$$

$$\mathcal{S}(A)_{ff} = \beta C_f^\top A_{ff} C_f - 2\frac{\beta}{\theta} C_f^\top A_{ff}(\Sigma_v^{-1} + \frac{2}{\theta} A_{ff})^{-1} A_{ff} C_f + \frac{1}{2} J_3^\top J_1^{-1} J_3, \tag{3.3.15}$$

$$\mathcal{U}(A, A_0) = \frac{\beta\theta}{2}\log|I + \frac{2}{\theta}\Sigma_v A_{ff}| + \beta A_0, \tag{3.3.16}$$

and $|\cdot|$ denotes the determinant of a matrix.

We now show that the properties we need for a quadratic representation of $V_{t,T}$ are indeed preserved by (3.3.8). We write $\mathcal{S}^n$ for the n-fold iteration of the mapping $\mathcal{S}$.

Proposition 3.3.7. *(i) If A is symmetric then $\mathcal{S}^n(A)$ is symmetric for all $n > 0$. (ii) If A is symmetric and satisfies $\Lambda^{1/2} J_1^{-1} \Lambda^{1/2} < I$, $A_{ff} \geq 0$, and Condition 3.3.4, then*

$$-\Lambda/2 < \mathcal{S}^n(A)_{xx} < 0 \quad \textit{and} \quad \mathcal{S}^n(A)_{ff} \geq 0, \quad \textit{for } n > 0.$$

Hence, Conditions 3.3.4 and 3.3.3 hold for all $n > 0$.

3.3.4 Optimal Controls

We can now summarize the optimal controls for the adversary and the investor. We use $\tilde{E}$ to denote expectation under the change of measure selected by the adversary. The conditions on $A^{(T,T)}$ in the following result hold, in particular, for the terminal condition $A^{(T+1,T)} = 0$ corresponding to $U_{T+1,T} = 0$, but they hold more generally as well.

Theorem 3.3.8. *Suppose $A^{(T,T)}$ satisfies the conditions in Proposition 3.3.7 so that Conditions 3.3.4 and 3.3.3 hold for all $A^{(t,T)}$, $t \leq T$. (i) Under the adversary's optimal change of measure, the conditional distribution of (u_{t+1}, v_{t+1}) given $(u_1, v_1), \ldots, (u_t, v_t)$ is normal with conditional covariance $\tilde{\Sigma}_{t+1}$,*

$$\tilde{\Sigma}_{t+1} = \begin{bmatrix} \Sigma_u & 0 \\ 0 & (\tilde{\Sigma}_v)_{t+1} \end{bmatrix}, \quad \textit{where} \quad (\tilde{\Sigma}_v)_{t+1} = \left(\Sigma_v^{-1} + \frac{2}{\theta} A_{ff}^{(t+1,T)} \right)^{-1},$$

and conditional mean $(\mu_{u,t+1}, \mu_{v,t+1})$*, with* $\mu_{u,t+1} = -\Sigma_u x_t/(\theta\beta)$ *and*

$$\begin{aligned}\mu_{v,t+1} &= -\frac{1}{\theta}(\tilde{\Sigma}_v)_{t+1}\left((A_{xf}^{(t+1,T)})^\top x_t + 2A_{ff}^{(t+1,T)} C_f f_t\right) \\ &= -\frac{1}{\theta}(\tilde{\Sigma}_v)_t \frac{\partial V_{t+1,T}(x_t, \tilde{E}_t[f_{t+1}])}{\partial f_t}.\end{aligned} \tag{3.3.17}$$

(ii) The investor's optimal choice is

$$\Delta x_t^* = 2\Lambda^{-1} A_{xx}^{(t,T)}(x_{t-1} + \frac{1}{2}(A_{xx}^{(t,T)})^{-1} A_{xf}^{(t,T)} f_t) = \Lambda^{-1}\frac{\partial V_{t,T}}{\partial x}, \tag{3.3.18}$$

so

$$x_t = J_1^{-1} J_2 = (I + 2\Lambda^{-1} A_{xx}^{(t,T)}) x_{t-1} + \Lambda^{-1} A_{xf}^{(t,T)} f_t. \tag{3.3.19}$$

The effect of the adversary's control is to change the evolution of the factors from (3.2.2) to

$$f_{t+1} = C_f f_t + \mu_{v,t+1} + \tilde{v}_{t+1}, \quad \tilde{v}_{t+1} \sim N(0, (\tilde{\Sigma}_v)_{t+1});$$

in particular, this makes

$$\begin{aligned}\tilde{E}_t[f_{t+1}] &= C_f f_t + \mu_{v,t+1} \\ &= (I - \frac{2}{\theta}(\tilde{\Sigma}_v)_{t+1} A_{ff}^{(t+1,T)}) C_f f_t - \frac{1}{\theta}(\tilde{\Sigma}_v)_{t+1}(A_{xf}^{(t+1,T)})^\top x_t \\ &= (\tilde{\Sigma}_v)_{t+1}(\Sigma_v^{-1} C_f f_t - \frac{1}{\theta}(A_{xf}^{(t+1,T)})^\top x_t),\end{aligned} \tag{3.3.20}$$

where $\tilde{E}$ denotes expectation under the change of measure selected by the adversary. Because $(\tilde{\Sigma}_v)_{t+1} \leq \Sigma_v$, we interpret the first term in (3.3.20) as shrinking the persistence of the factors and thus potentially reducing their value to the investor; the second term in (3.3.20) indicates that the adversary also exploits the investor's current portfolio in setting the conditional mean of the factors, as suggested by the expression following (3.3.17).

Corollary 3.3.9. *We can write the investor's optimal choice (3.3.19) as*

$$x_t = (I + 2\Lambda^{-1}A_{xx}^{(t,T)})x_{t-1} - 2\Lambda^{-1}A_{xx}^{(t,T)}(-\frac{1}{2}(A_{xx}^{(t,T)})^{-1}A_{xf}^{(t,T)}f_t) \tag{3.3.21}$$

$$= (\gamma_\theta\Sigma_u + \Lambda - 2\beta A_{xx}^{(t+1,T)})^{-1}$$

$$\times (\Lambda x_{t-1} + \gamma_\theta\Sigma_u(\gamma_\theta\Sigma_u)^{-1}B_f f_t - 2\beta A_{xx}^{(t+1,T)}(-\frac{1}{2}(A_{xx}^{(t+1,T)})^{-1}A_{xf}^{(t+1,T)}\tilde{E}_t[f_{t+1}]). \tag{3.3.22}$$

For (3.3.21), as in Garleanu and Pedersen (2012), we can interpret this choice as a weighted average of the current portfolio x_{t-1} and a target portfolio given by

$$target = -\frac{1}{2}(A_{xx}^{(t,T)})^{-1}A_{xf}^{(t,T)}f_t. \tag{3.3.23}$$

The target portfolio maximizes the quadratic function $V_{t,T}$, given the factor level f_t. If transaction costs were waived for one period, the investor would move immediately to the target; otherwise, the investor's optimal trade (3.3.18) is proportional to the difference between the current portfolio and the target, the proportion depending on the cost matrix Λ. Recall that the A matrix depends on the robustness parameter θ through the recursions in Proposition 3.3.6.

In (3.3.22), the term $(\gamma_\theta\Sigma_u)^{-1}B_f f_t$, which we call the *myopic* portfolio, maximizes the single-

period mean-variance objective

$$x_t^\top B_f f_t - \frac{\gamma_\theta}{2} x_t^\top \Sigma_u x_t.$$

Thus, (3.3.22) represents x_t as a weighted average of the current portfolio x_{t-1}, the myopic portfolio, and the conditional expectation of the target portfolio one step ahead. Comparing this expression to equation (16) of Garleanu and Pedersen (2012), we can interpret the effect of the robust solution as replacing the original conditional expectation of the factors with their conditional expectation under the adversary's change of measure and implicitly increasing the investor's risk aversion parameter from γ to γ_θ. Interestingly, if we omitted the variance penalty $\gamma x_t^\top \Sigma_u x_t/2$ from the original objective (3.2.4), it would still appear in the robust formulation, because $\gamma_\theta > 0$ even if $\gamma = 0$. Uncertainty in the linear term $u_{t+1}^\top x_t$ has the effect of increasing risk aversion.

3.4 Comparison with the Non-Robust Case

In this section, we examine the effect of varying the robustness parameter θ, including the non-robust formulation $\theta = \infty$ as a limiting case. We affix θ as a subscript or superscript to indicate functions and quantities tied to a specific value of the parameter. The non-robust version is indicated by a subscript or superscript ∞.

We denote by $U_{t,T}^\infty$ and U^∞, respectively, the finite-horizon and infinite-horizon value functions for the (non-robust) objective (3.2.4) and define a dynamic programming operator acting on functions $h : \mathbb{R}^{n_x} \times \mathbb{R}^{n_f} \to \mathbb{R}$ by

$$\mathcal{T}^\infty(h)(x,f) = \sup_{\Delta x_+ \in \mathbb{R}^{n_x}} \{x_+^\top B_f f_+ - \frac{\gamma}{2} x_+^\top \Sigma_v x_+ - \frac{1}{2} \Delta x_+^\top \Lambda \Delta x_+ + E\left[h(x_+, f_+)\right]\},$$

with $x_+ = x + \Delta x_+$, and the expectation taken over $f_+ = C_f f + v$, $v \sim N(0, \Sigma_v)$. Then $U^\infty_{t,T}$ satisfies the recursion $\mathcal{T}^\infty(U^\infty_{t+1,T}) = U^\infty_{t,T}$. Garleanu and Pedersen (2012) show that this dynamic programming equation maps a quadratic function backward to another quadratic function. We can therefore write

$$U^\infty_{t,T}(x_{t-1}, f_t) = x_{t-1}^\top \mathcal{S}_\infty^{T-t}(A)_{xx} x_{t-1} + x_{t-1}^\top \mathcal{S}_\infty^{T-t}(A)_{xf} f_t + f_t^\top \mathcal{S}_\infty^{T-t}(A)_{ff} f_t + \mathcal{U}_\infty^{T-t}(A, A_0), \tag{3.4.1}$$

with $\mathcal{S}_\infty^{T-t}(A)$ and $\mathcal{U}_\infty^{T-t}(A, A_0)$ the coefficients of $U^\infty_{t,T}$ at time t when $U^\infty_{T,T}$ is quadratic with coefficient matrix A. Here, $\mathcal{S}^n_\infty$ is the n-fold iteration of $\mathcal{S}_\infty$, but $\mathcal{U}^n_\infty$ is defined recursively by setting $\mathcal{U}^1_\infty = \mathcal{U}_\infty$ and $\mathcal{U}^n_\infty(A, A_0) = \mathcal{U}_\infty(\mathcal{S}^{n-1}_\infty(A), \mathcal{U}^{n-1}_\infty(A, A_0))$; see the analogous dependence on A and A_0 in (3.3.16).

The non-robust case can be considered a special case of the robust formulation. Condition 3.3.3 holds automatically when $\theta = \infty$, and if Condition 3.3.4 holds for some matrix A for $\theta = \infty$, then it also holds for any $\theta \in (0, \infty)$. This is because the last term in (3.3.10) is positive definite for $\theta \in (0, \infty)$ but vanishes when $\theta = \infty$, and γ_θ is decreasing in θ, so $J_1^\theta \geq J_1^\infty$. It is also easy to verify that Proposition 3.3.7 holds at $\theta = \infty$. As we vary θ (smaller θ indicating greater robustness), the coefficient matrices are ordered as follows:

Lemma 3.4.1. *If (A, A_0) satisfies $A_{ff} \geq 0$, $\Lambda^{1/2} J_1^{-1} \Lambda^{1/2} < I$, and Condition 3.3.4 for some $0 < \theta_1 < \theta_2 \leq \infty$, then for any $n \geq 0$, $\mathcal{S}^n_{\theta_1}(A) \leq \mathcal{S}^n_{\theta_2}(A)$ and $\mathcal{U}^n_{\theta_1}(A, A_0) \leq \mathcal{U}^n_{\theta_2}(A, A_0)$.*

To illustrate, suppose we start the recursions for two parameter levels $0 < \theta_1 < \theta_2 \leq \infty$ from

the same terminal condition with coefficients (A, A_0) (including $A = A_0 = 0$ as a special case). Suppose the conditions of Lemma 3.4.1 hold. We make the following observations:

(a) In the portfolio decomposition (3.3.21), the weight on the previous position satisfies $1 + 2\Lambda^{-1}\mathcal{S}^n_{\theta_1}(A)_{xx} < 1 + 2\Lambda^{-1}\mathcal{S}^n_{\theta_2}(A)_{xx}$, so more robustness (smaller θ) leads to less weight on the previous position x_{t-1} and more weight on the target portfolio (3.3.23). The less-robust investor puts greater trust in the persistence of the factors described by the model and thus attaches greater value to the previous portfolio. However, the coefficient of f_t in (3.3.21) can either increase or decrease with θ because $\mathcal{S}^n_\theta(A)_{xf}$ can increase or decrease or change in a more complicated way.

(b) From the decomposition (3.3.22), we find similarly that increasing robustness decreases weight on the myopic portfolio $(\gamma_\theta \Sigma_u)^{-1} B_f f_t$, and it also decreases the size of the myopic portfolio because γ_θ increases with θ. If we remove u_{t+1} from (3.2.4) and limit robustness to the factor dynamics (3.2.2) only, then $\gamma_\theta \equiv \gamma$ and the myopic portfolio does not vary with θ.

(c) Also from (3.3.22), we see that increasing robustness puts more weight on the conditional expectation of the target portfolio while decreasing the coefficient on the conditional expectation of the factors. In numerical examples, we find that the conditional expectation of the target portfolio is very sensitive to θ.

We can also interpret the effect of robustness from the optimal controls

$$\Delta x^*_t = \Lambda^{-1} \frac{\partial V_{t+1,T}}{\partial x} \text{ and } \mu_{v,t+1} = -\frac{1}{\theta}(\tilde{\Sigma}_v)_t \frac{\partial V_{t+1,T}(x_t, \tilde{E}_t[f_{t+1}])}{\partial f_t}.$$

These expressions are already very suggestive, as they show the investor and the adversary using their controls to increase and decrease V, respectively. Also, the quadratic function $V_{t,T}$ is concave in x_{t-1} and convex in f_t, making it a hyperbolic paraboloid. The cross term $x_t^\top B_f f_t$ in the objective function leads to the cross term coefficient $A_{xf} \neq 0$ in the value function. The presence of this term means that the axes of the hyperbolic paraboloid are twisted and not orthogonal to each other. As a result, the minimum point for f is linear in x, and maximum point for x is linear in f, properties exploited by both players. This can also be seen in (3.3.17) and (3.3.19).

If there were no cross term in the objective and we had $A_{xf}^{t+1,T} = 0$, then the coefficient of $C_f f_t$ in (3.3.17) would be a negative definite matrix, and the effect of the adversary's choice of $\mu_{v,t+1}$ would thus be to accelerate the mean reversion of the factors in (3.2.2) and reduce their value to the investor. In fact, in the limit as θ approaches zero, the coefficient of f_t in (3.3.17) becomes $-C_f$ which eliminates any persistence in the factor dynamics (3.2.2). With a nonzero cross term, the adversary can do further harm by moving the factors in a direction that depends on the investor's current portfolio.

We conclude this section by verifying that value iteration for the non-robust problem converges; this is needed to confirm that the solution to the Bellman equation found in Garleanu and Pedersen (2012) is in fact the value function for the infinite-horizon problem and that the corresponding control is optimal. In the following, J_1 is evaluated with $\theta = \infty$.

Proposition 3.4.2. *If A is such that $J_1 > 0$, $A_{ff} \geq 0$, and*

$$A - \begin{bmatrix} 0 & 0 \\ 0 & \frac{1}{2\gamma} B_f^\top \Sigma_u^{-1} B_f \end{bmatrix} \leq 0,$$

then the iteration of (3.4.1) converges; i.e., $\lim_{n\to\infty}(\mathcal{S}_\infty^n(A), \mathcal{U}_\infty^n(A, A_0))$ *exists. The control (3.3.18) obtained from the limit is optimal, and the quadratic function defined by the limit solves the Bellman equation and is the value function for the infinite-horizon problem.*

3.5 Infinite-Horizon Robust Problem

3.5.1 Formulation and Bellman equation

For the robust infinite-horizon problem, define U by setting $t = 0$ and $T = \infty$ on the right side of (3.3.1). This robust value function is bounded above by the non-robust value function (corresponding to $\theta = \infty$ in Lemma 3.4.1), and it is bounded below because the investor can choose $x_t \equiv 0$.

Proposition 3.5.1. *With* $\mathcal{T}$ *the operator defined in (3.3.2),* U *satisfies*

$$U = \mathcal{T}(U). \tag{3.5.1}$$

Similarly, by arguing as in Proposition 3.3.2 and the subsequent discussion, we arrive at the following result.

Proposition 3.5.2. *Suppose* $V(x_{t-1}, f_t)$ *satisfies* $V = \mathcal{T}^V(V)$*, with* $\mathcal{T}^V$ *as defined by (3.3.8). Then* $U(M_t, x_{t-1}, f_t) = M_t V(x_{t-1}, f_t)$ *satisfies (3.5.1) with*

$$m_{t+1}^* = \frac{\exp\left\{-\frac{1}{\theta}\left(V(x_t, f_{t+1}) + \frac{1}{\beta}x_t^\top u_{t+1}\right)\right\}}{E_t\left[\exp\left\{-\frac{1}{\theta}\left(V(x_t, f_{t+1}) + \frac{1}{\beta}x_t^\top u_{t+1}\right)\right\}\right]}, \tag{3.5.2}$$

provided the normalization in (3.5.2) is finite.

This reduces the problem of finding a solution to the robust Bellman equation (3.5.1) to one of solving $V = \mathcal{T}^V(V)$. In solving for the infinite-horizon V, the finite-horizon recursions for $V_{t,T}$ in Proposition 3.3.6 become simultaneous equations. Given coefficients (A, A_0) of a quadratic function, define J_1, J_2, and J_3 as in (3.3.10)–(3.3.12).

Proposition 3.5.3. *If A is symmetric and (A, A_0) satisfy $A_{ff} \geq 0$, $J_1 > 0$, $\Lambda^{1/2} J_1^{-1} \Lambda^{1/2} < I$, and*

$$\begin{aligned}
A_{xx} &= -\frac{1}{2}\Lambda + \frac{1}{2}\Lambda J_1^{-1}\Lambda^\top, \\
A_{xf} &= \Lambda J_1^{-1}\left[B_f + \beta A_{xf} C_f - 2\frac{\beta}{\theta} A_{xf}(\Sigma_v^{-1} + \frac{2}{\theta}A_{ff})^{-1} A_{ff} C_f\right], \\
A_{ff} &= \beta C_f^\top A_{ff} C_f - 2\frac{\beta}{\theta} C_f^\top A_{ff}(\Sigma_v^{-1} + \frac{2}{\theta}A_{ff})^{-1} A_{ff} C_f + \frac{1}{2} J_3^\top J_1^{-1} J_3, \\
(1-\beta)A_0 &= -\frac{\beta\theta}{2}\log|I + \frac{2}{\theta}\Sigma_v A_{ff}|,
\end{aligned}$$

then the quadratic function V defined by (A, A_0) is a fixed point of $\mathcal{T}^V$.

This is a direct consequence of Proposition 3.3.7. With this result, we can solve the equations for (A, A_0) and check the conditions in the statement of the proposition (which ensure the saddle-point conditions we need for optimality). If these are satisfied, then we have a fixed point V, from which we get a solution U to the Bellman equation $\mathcal{T}(U) = U$ by setting $U(M, x, f) = MV(x, f)$. Such a solution provides candidate optimal controls for both the investor and the adversary — controls that attain the supremum and the infimum in the one-step operator $\mathcal{T}$. The calculation of these controls is similar to that in Theorem 3.3.8, but simpler because of that stationarity implicit in the infinite-horizon setting. We summarize the calculation as follows:

Lemma 3.5.4. *Suppose the conditions of Proposition 3.5.3 hold and define V from (A, A_0) accordingly. (i) Under the change of measure (3.5.2), the conditional distribution of (u_{t+1}, v_{t+1}) given $(u_1, v_1), \ldots, (u_t, v_t)$ is normal with conditional covariance $\tilde{\Sigma}$,*

$$\tilde{\Sigma} = \begin{bmatrix} \Sigma_u & 0 \\ 0 & \tilde{\Sigma}_v \end{bmatrix}, \quad \textit{where} \quad \tilde{\Sigma}_v = \left(\Sigma_v^{-1} + \frac{2}{\theta} A_{ff}\right)^{-1},$$

and conditional mean $(\mu_{u,t+1}, \mu_{v,t+1})$, with $\mu_{u,t+1} = -\Sigma_u x_t/(\theta\beta)$ and

$$\begin{aligned} \mu_{v,t+1} &= -\frac{1}{\theta}\tilde{\Sigma}_v \left(A_{xf}{}^{\top} x_t + 2A_{ff} C_f f_t\right) \\ &= -\frac{1}{\theta}\tilde{\Sigma}_v \frac{\partial V(x_t, \tilde{E}_t[f_{t+1}])}{\partial f_t}. \end{aligned}$$

(ii) The supremum over Δx in the Bellman equation $V = \mathcal{T}^V(V)$ is given by the investment choice

$$\Delta x_t^* = 2\Lambda^{-1} A_{xx}(x_{t-1} + \frac{1}{2}(A_{xx})^{-1} A_{xf} f_t) = \Lambda^{-1} \frac{\partial V(x_{t-1}, f_t)}{\partial x},$$

under which

$$x_t = J_1^{-1} J_2 = (I + 2\Lambda^{-1} A_{xx}) x_{t-1} + \Lambda^{-1} A_{xf} f_t.$$

3.5.2 Stability

Lemma 3.5.4 provides explicit expressions for the controls obtained by solving the robust Bellman equation. As is often the case in infinite-horizon problems, we need additional conditions to verify

that a solution to the Bellman equation is in fact the value function (3.3.1) (with $T = \infty$) and that the corresponding controls are optimal. For these properties, we need to impose stability properties on the evolution of the controlled system. The key property is the admissibility condition $\mathcal{R}_\beta(m) < \infty$, with $\mathcal{R}_\beta$ as defined in (3.2.6). Although it refers only to the adversary's control, this property is best viewed as a condition on the controls of both players because the adversary's choice of m_t may depend on the investor's choice of portfolio. Define the state y_t and the extended state y_t^e by setting

$$y_t = \begin{bmatrix} x_{t-1} \\ f_t \end{bmatrix} \text{ and } y_t^e = \begin{bmatrix} u_t \\ x_{t-1} \\ f_t \end{bmatrix}.$$

The state evolution depends on the chosen pair of policies $(\Delta x, m)$. We use $\|\cdot\|$ to denote the usual Euclidean vector norm. The full stability condition we use is as follows.

Definition 3.5.5. *(β-Stability.) We call a policy pair $(\Delta x, m)$ and the resulting extended state evolution β-stable if $\mathcal{R}_\beta(m) < \infty$ and if $\alpha^t \tilde{E}[\|y_t^e\|^2] \Rightarrow 0$ for some $\alpha \in (\beta, 1)$, for all y_0^e.*

The mean square convergence to zero of $\alpha^{t/2} y_t^e$ under the change of measure is sufficient to ensure that the infinite discounted sum (with discount factor β) of a quadratic function of the extended state is finite. We thus interpret β-stability as ensuring that the adversary cannot drive the investor reward to $-\infty$ and that the investor cannot drive the relative entropy penalty to $+\infty$; in particular, the condition avoids the possibility of getting $\infty - \infty$ in the robust value function. For the controls obtained from the Bellman equation (the controls in Lemma 3.5.4), a simpler condition

characterizes β-stability. We use the condition

$$\mathcal{R}_\alpha(m) < \infty \text{ for some } \alpha \in (\beta, 1). \tag{3.5.3}$$

Lemma 3.5.6. *Suppose U is a solution to the robust Bellman equation (3.5.1) for which $U(M, x, f)$ is the product of M and a quadratic function of (x, f). If the quadratic function satisfies Conditions 3.3.4 and 3.3.3, then the resulting controls $(\Delta x, m)$ are β-stable if and only if (3.5.3) holds.*

3.5.3 Optimality

We now verify that that the policies provided by the robust Bellman equation through Lemma 3.5.4 do indeed solve the robust control problem in a suitable sense. Suppose both the investor and the adversary choose their policies (Markov, as we assume throughout), and let x_t^* and m_t^* be the resulting portfolios and likelihood ratios. Then the value attained by this pair of policies, starting from (M_0, x_{-1}, f_0), whenever this expression is well-defined, is given by

$$\begin{aligned} U^*(M_0, x_{-1}, f_0) &= E\left[\sum_{t=0}^{\infty} \beta^t M_t \left(Q(x_t^*, \Delta x_t^*, f_t) + \theta\beta m_{t+1}^*(\log m_{t+1}^* + u_{t+1}^\top x_t^*)\right)\right] & (3.5.4)\\ &= \tilde{E}\left[\sum_{t=0}^{\infty} \beta^t \left(Q(x_t^*, \Delta x_t^*, f_t) + \theta\beta m_{t+1}^*(\log m_{t+1}^* + u_{t+1}^\top x_t^*)\right)\right], & (3.5.5) \end{aligned}$$

where, as before, $\tilde{E}$ denotes expectation under the adversary's change of measure. We show that a solution to the robust Bellman equation (3.5.1) is indeed the value attained under the corresponding policies, and the policy forms an equilibrium. Once one player has selected a policy, we call a policy selected by the other player a β-stable response if the resulting policy pair is β-stable.

Theorem 3.5.7. *Suppose U^* is a solution to the robust Bellman equation (3.5.1), and suppose $U^*(M, x, f)$ is the product of M and a quadratic function of (x, f) satisfying Conditions 3.3.3–3.3.4. Suppose the corresponding policy pair $(\Delta x^*, m^*)$ satisfies $\mathcal{R}_\alpha(m^*) < \infty$ for some $\alpha \in (\beta, 1)$. Then*

(i) U^* is the value attained under the policy pair $(\Delta x^*, m^*)$.

(ii) The investor's best β-stable response to m^* is Δx^*. The adversary's best β-stable response to Δx^* is m^*.

This result justifies the controls that come out of Lemma 3.5.4. It is worth noting that a violation of β-stability entails either a portfolio size that grows exponentially or an infinite relative entropy penalty. The restriction to β-stable policy pairs is therefore sensible, and it is appropriate to view the policy Δx^* derived from the robust Bellman equation as the investor's optimal choice in the face of the model uncertainty captured by the robust formulation.

3.5.4 Convergence of Value Iteration

From Lemma 3.5.4, we see that the key step in solving the infinite-horizon robust control problem is solving the equations in Proposition 3.5.3, which restate the condition $V = \mathcal{T}^V(V)$ for quadratic V. A natural approach is to start from some initial (A, A_0) and apply the equations iteratively. Each application of the equations is an application of the operator $\mathcal{T}^V$, so the question of convergence of this iterative approach is equivalent to the question of convergence of the finite-horizon function $V_{t,T}$ as $t \to -\infty$ with $V_{T,T}$ the quadratic function determined by the starting point (A, A_0). Hansen and Sargent (1995) consider the case where the objective function is concave in the state variable,

which allows a simple proof through a monotone convergence argument, but our setting is beyond the scope of their result.

Over a finite horizon T, each $(A^{t+1,T}, A_0^{t+1,T})$ determines candidate controls through the prescription in Theorem 3.3.8. Under these controls, the state y_t evolves as in (A.2) but with a time-dependent transition matrix

$$\bar{\Psi}_{t,T} = \begin{bmatrix} I & 0 \\ -\frac{1}{\theta}(\tilde{\Sigma}_v)_{t+1} A_{xf}^{(t+1,T)} & (\tilde{\Sigma}_v)_{t+1} \Sigma_v^{-1} C_f \end{bmatrix} \begin{bmatrix} I + 2\Lambda^{-1} A_{xx}^{(t+1,T)} & \Lambda^{-1} A_{xf}^{(t+1,T)} \\ 0 & I \end{bmatrix}.$$

If both factors in this representation have norm less than $\beta^{-1/2}$, then we have convergence of $\mathcal{S}^n(A)$ $\mathcal{U}^n(A, A_0)$ for any initial (A, A_0) that satisfies $\Lambda^{1/2} J_1^{-1} \Lambda^{1/2} < I$ and $A_{ff} \geq 0$. The norm here can be any matrix norm for which $\|M^n\|^{1/n} \to \sigma(M)$, such as any p-norm.

We have not found simple sufficient conditions that ensure this uniform stability condition. The condition can easily be checked for each $\bar{\Psi}_{t,T}$ at each iteration as part of an iterative algorithm, but given the difficulty of verifying the condition in advance we omit the details of the result. In our numerical experiments, we have never observed a failure to converge starting either from zero or the solution of the non-robust case and, indeed, the convergence appears to be quite fast. An alternative to iteration is the decomposition method covered in Hansen and Sargent (2007). This approach leads to conditions that guarantee a solution, but it requires a lengthy and technical digression so we omit it.

3.6 Numerical Results

3.6.1 Data Description and Model Estimation

In order to test the effect of the robust formulation, we work with the application to commodity futures in Garleanu and Pedersen (2012), using futures prices on the following commodities: aluminun, copper, nickel, zinc, lead, and tin from the London Metal Exchange; gas oil from the Intercontinental Exchange; WTI crude, RBOB unleaded gasoline, and natural gas from the New York Mercantile Exchange; gold and silver from the New York Commodities Exchange; and coffee, cocoa, and sugar from the New York Board of Trade. For consistency with Garleanu and Pedersen (2012), we use daily prices for the period 01/01/1996 – 01/23/2009 for our in-sample tests; we use data through 04/09/2010 for out-of-sample tests. As discussed in Garleanu and Pedersen (2012), extracting price changes from futures prices requires some assumptions on how contracts are rolled, and this makes it difficult to reproduce exactly the same time series of price changes. We choose the contract with the largest volume on each day. In some early samples when volumes for some commodities are not available, we choose the contract closest to maturity that does not expire in the current month. Our estimates and results are quite close and adequate for the purpose of examining the effect of robustness.

For each commodity, Garleanu and Pedersen (2012) introduce factors f^{5D}, f^{1Y} and f^{5Y}, which are the moving averages of price changes over the previous five days, one year and five years, normalized by their respective standard deviations. Using these factors, we estimate the following

model of price changes for each commodity:

$$r^s_{t+1} = \underset{(0.54)}{0.004} + \underset{(2.41)}{11.43}\, f^{5D,s}_t + \underset{(2.42)}{107.55}\, f^{1Y,s}_t - \underset{(-1.65)}{218.76}\, f^{5Y,s}_t + u^s_{t+1},$$

the superscript s indexing the 15 commodities. This is a pooled panel regression — the coefficients are the same across all commodities — with parameters estimated using feasible generalized least squares. The numbers reported under the coefficients are t-statistics. Similarly, for the factor dynamics we get the following estimates:

$$\Delta f^{5D,s}_{t+1} = \underset{(-0.67)}{-0.2510} f^{5D,s}_t + v^{5D,s}_{t+1},$$

$$\Delta f^{1Y,s}_{t+1} = \underset{(-0.64)}{-0.0039} f^{1Y,s}_t + v^{1Y,s}_{t+1},$$

$$\Delta f^{5Y,s}_{t+1} = \underset{(-0.78)}{-0.0010} f^{5Y,s}_t + v^{5Y,s}_{t+1}.$$

The matrix C_f is thus diagonal and, in light of the t-statistics, a potential source of model error to be captured in the v_{t+1} terms. With $f = (f^{5D,1}, f^{1Y,1}, f^{5Y,1}, ..., f^{5D,15}, f^{1Y,15}, f^{5Y,15})^\top$, the form of the loading matrix B_f follows from the regression equation for r^s_t. Erb and Harvey (2006) documented the 1-year momentum factor in commodity futures prices. Asness et al. (2009), Moskowitz et al. (2012) documented 1-year and 5-year many asset classes.

We adopt the choices in Garleanu and Pedersen (2012) in estimating Σ_v and Σ_u, and in setting the risk aversion parameter to $\gamma = 10^{-9}$, the one-day discount factor $\beta = \exp(-0.02/260)$ corresponding to a 2% annual rate, and the transaction cost matrix to $\Lambda = \lambda\Sigma_u$, with $\lambda = 3 \times 10^{-7}$.

3.6.2 In-Sample Tests

This section reports results of in-sample tests in which we evaluate portfolio performance on the same price data used to estimate the model. We compare performance at various levels of the robustness parameter θ, including the non-robust case $\theta = \infty$. The "No TC" case is a strategy that ignores transaction costs and thus reduces to the mean-variance optimal portfolio $x_{t+1} = (\gamma\Sigma_u)^{-1}B_f f_t$. With $\Lambda = \lambda\Sigma_u$, the myopic portfolio corresponds to taking $\beta = 0$, and it evolves as $x_{t+1} = \frac{\lambda}{\lambda+\gamma}x_t + \frac{1}{\lambda+\gamma}\Sigma_u^{-1}B_f f_t$.

Tables 3.1 and 3.2 summarize performance results. The robust results in Table 3.1 are based on allowing the changes in both returns (through u_t) and factor dynamics (through v_t); in Table 3.2, robustness is limited to v_t by omitting u_{t+1} from the original problem (3.2.4). As we discussed in Section 3.2.2, alternative but equivalent non-robust objectives can lead to different robust problems.

The columns labeled "mean/std" report annualized performance ratios computed as

$$\sqrt{260} \times \text{Mean(daily \$ profit)}/\text{Standard deviation(daily \$ profit)}.$$

We refer to these loosely as Sharpe ratios though they are calculated from differences rather than percentage changes because the assets are futures contracts — each contract has zero initial value, and total portfolio value can become negative. The columns labeled "Obj" report the objective function value

$$\text{Mean(daily \$ profit)} - \frac{\gamma}{2}\text{Variance(daily \$ profit)}.$$

The difference between gross and net performance is the effect of transaction costs.

To provide a rough indication of the statistical significance of our comparisons, we group the data into consecutive batches and calculate standard errors across batches. For each level of the robustness parameter θ, we calculate an approximate t-statistic (using the first estimator in Theorem 1 of Muñoz and Glynn (1997)) for the difference in performance between the robust and non-robust strategies. With sufficient stationarity and mixing in the underlying data, these statistics are indeed asymptotically t-distributed, though the conditions required are not guaranteed to hold in practice. We report results based on 40 batches and have obtained very similar results with 30 batches and with 50 batches.

Not surprisingly, ignoring transactions costs leads to good gross performance and terrible net performance in the first row of Table 3.1. The myopic rule produces less extreme differences but overall poor results. These portfolios help illustrate the value of dynamic control rules. The non-robust rule is optimized to the net objective function, so there is no reason to expect to see any benefit to robustness by this criterion. In Table 3.1, we see some deterioration in the net objective as we increase robustness (decrease θ); in Table 3.2, the net objective is relatively insensitive over a wide range of θ values. Interestingly, when we compare performance based on Sharpe ratios, for which none of the rules has been optimized, adding robustness appears to improve performance in both cases, though the differences are not significant as measured by our t-statistics. The in-sample improvement in the Sharpe ratio for the robust portfolios is primarily due to a reduction in the standard deviation in the denominator. At high robustness levels, the net excess return of the robust portfolio can be lower than that of the non-robust portfolio but with smaller standard deviation.

	θ	Obj $\times 10^{-6}$				mean/std			
		Gross	t-stat	Net	t-stat	Gross	t-stat	Net	t-stat
No TC		1.22		-159.68		0.82		-11.67	
Myopic		-0.21		-0.22		0.08		0.08	
Non-robust		0.70		0.62		0.60		0.56	
Robust	10^{10}	0.70	-0.04	0.62	0.12	0.61	-0.08	0.57	-0.13
	10^{9}	0.55	-0.65	0.51	-0.49	0.61	-0.29	0.57	-0.48
	10^{8}	0.15	-1.24	0.13	-1.09	0.63	0.09	0.57	-0.35
	10^{7}	0.02	-1.37	0.02	-1.21	0.72	0.89	0.66	0.70
	10^{6}	0.00	-1.40	0.00	-1.23	0.79	1.04	0.77	1.18
	10^{5}	0.00	-1.40	0.00	-1.23	0.82	1.13	0.82	1.40

Table 3.1: In-sample performance comparisons using the full data series with robustness in returns (u_t) and factor dynamics (v_t). For each θ, the t-stats compare performance of the robust rule at that θ with the non-robust case, based on grouping the data into 40 batches.

	θ	Obj $\times 10^{-6}$				mean/std			
		Gross	t-stat	Net	t-stat	Gross	t-stat	Net	t-stat
Non-robust		0.70		0.62		0.60		0.56	
Robust	10^{10}	0.70	0.40	0.62	0.53	0.60	0.80	0.56	0.89
	10^{9}	0.71	0.38	0.62	0.52	0.60	0.81	0.57	0.89
	10^{8}	0.72	0.35	0.65	0.48	0.61	0.98	0.57	1.05
	10^{7}	0.72	0.07	0.66	0.19	0.65	1.15	0.62	1.19
	10^{6}	0.51	-0.50	0.48	-0.37	0.72	0.97	0.68	0.96
	10^{5}	0.23	-1.02	0.22	-0.87	0.75	0.84	0.71	0.77

Table 3.2: In-sample performance comparisons using the full data series with robustness in factor dynamics (v_t) only. For each θ, the t-stats compare performance of the robust rule at that θ with the non-robust case, based on grouping the data into 40 batches.

3.6.3 Out-of-Sample Tests

To compare out-of-sample performance, we re-estimate the model parameters each week from 01/01/1996 through 04/09/2010 using the previous 6 months of data. Each time the parameters are estimated, the investment control rule remains fixed for one week until the parameters are next updated. Thus, at each point in time, the investment policy is based solely on prior market data. Updating the parameter estimates based on a rolling 6-month window is also more reflective of how such a model would be used in practice.

Table 3.3 (with robustness to both u_t and v_t) and Table 3.4 (with robustness to v_t only) summarize the results. Over a wide range of θ values, the robust control rules show improved net performance as measured by either the objective function value or the Sharpe ratio. In effect, the robust rules acknowledge the uncertainty in the estimated model and thus trade less aggressively than the non-robust rule, and this improves out-of-sample performance. Allowing robustness to both the model of returns and the model of factor dynamics (Table 3.3) results in somewhat better results overall than focusing robustness on the factor dynamics.

As with the in-sample tests, the improvement mainly comes from the reduction in risk. In Table 3.5, t-statistics for the difference in net returns between the robust and non-robust portfolios are estimated using batch means with 40 batches. None of the robust portfolios has a significantly better net return than that of the non-robust portfolio.

Our subsequent analysis focuses on the less favorable case in which robustness is limited to the factor dynamics.

To illustrate the effect of robustness, Figure 3.1 shows the evolution of the positions in gold and crude oil under various strategies. Ignoring transaction costs leads to wild swings on a much wider scale, so we omit this case from the graph. Positions under the robust rules (shown at $\theta = 10^7$ and $\theta = 10^4$ with robustness to v_t only) fluctuate less than those chosen by the non-robust rule. At the same time, by anticipating the evolution of the factors, the robust rules are quicker to respond than the myopic portfolio. The figures and numerical results suggest that $\theta = 10^7$ provides a reasonable level of robustness and $\theta = 10^4$ is overly conservative. The third panel scales the non-robust positions to facilitate comparison. We discuss scaling strategies in Section 3.6.4.

Figure 3.2 compares net returns over the full time period and provides further insight into

	θ	Obj $\times 10^{-6}$				mean/std			
		Gross	t-stat	Net	t-stat	Gross	t-stat	Net	t-stat
No TC		-59.76		-8257.60		0.53		-5.74	
Myopic		-11.80		-11.90		-0.57		-0.59	
non-robust		-36.78		-42.92		0.35		0.04	
robust	10^9	-8.17	2.24	-10.50	2.32	0.39	0.93	0.15	3.07
	10^8	0.29	2.06	0.02	2.18	0.40	0.21	0.26	1.76
	10^7	0.10	1.99	0.08	2.13	0.46	0.38	0.37	1.73
	10^6	0.01	1.99	0.01	2.12	0.52	0.66	0.49	2.21
	10^5	0.00	1.99	0.00	2.12	0.53	0.69	0.52	2.32
	10^4	0.00	1.99	0.00	2.12	0.53	0.69	0.53	2.34

Table 3.3: Out-of-sample performance comparisons using a rolling 6-month estimation window with robustness in returns (u_t) and factor dynamics (v_t). For each θ, the t-stats compare performance of the robust rule at that θ with the non-robust case, based on grouping the data into 40 batches.

	θ	Obj $\times 10^{-6}$				mean/std			
		Gross	t-stat	Net	t-stat	Gross	t-stat	Net	t-stat
non-robust		-36.78		-42.92		0.35		0.04	
robust	10^9	-31.16	1.69	-36.38	1.72	0.37	1.21	0.08	2.67
	10^8	-18.70	1.91	-22.16	1.95	0.40	1.06	0.16	2.65
	10^7	-5.10	2.05	-6.70	2.12	0.44	0.58	0.25	1.99
	10^6	-0.28	2.03	-0.85	2.13	0.42	-0.39	0.27	0.81
	10^5	0.27	2.01	0.11	2.13	0.37	-0.73	0.26	0.38
	10^4	0.18	2.00	0.14	2.13	0.41	-0.26	0.33	0.86

Table 3.4: Out-of-sample performance comparisons using a rolling 6-month estimation window with robustness in factor dynamics (v_t) only. For each θ, the t-stats compare performance of the robust rule at that θ with the non-robust case, based on grouping the data into 40 batches.

differences across strategies. Ignoring transaction costs results in disastrously poor performance, so this case is omitted from the figure. The performance of the myopic portfolio degrades over the time. Interestingly, much of the benefit of the robust rule, compared with the non-robust rule, appears to be due to a small number of days. The non-robust rule can outperform the robust rule over long periods of time; adding robustness reduces the impact of a small number of bad bets by trading less aggressively on the signals from the factors. Consistent with what we see in Figure 3.2,

θ	10^9	10^8	10^7	10^6	10^5	10^4
robust in v only	0.96	0.69	0.33	0.03	-0.09	-0.12
robust in v and u	0.24	-0.07	-0.14	-0.15	-0.15	-0.15

Table 3.5: t-statistics of the difference of net returns between robust and non-robust portfolios for out-of-sample tests, based on grouping the data into 40 batches.

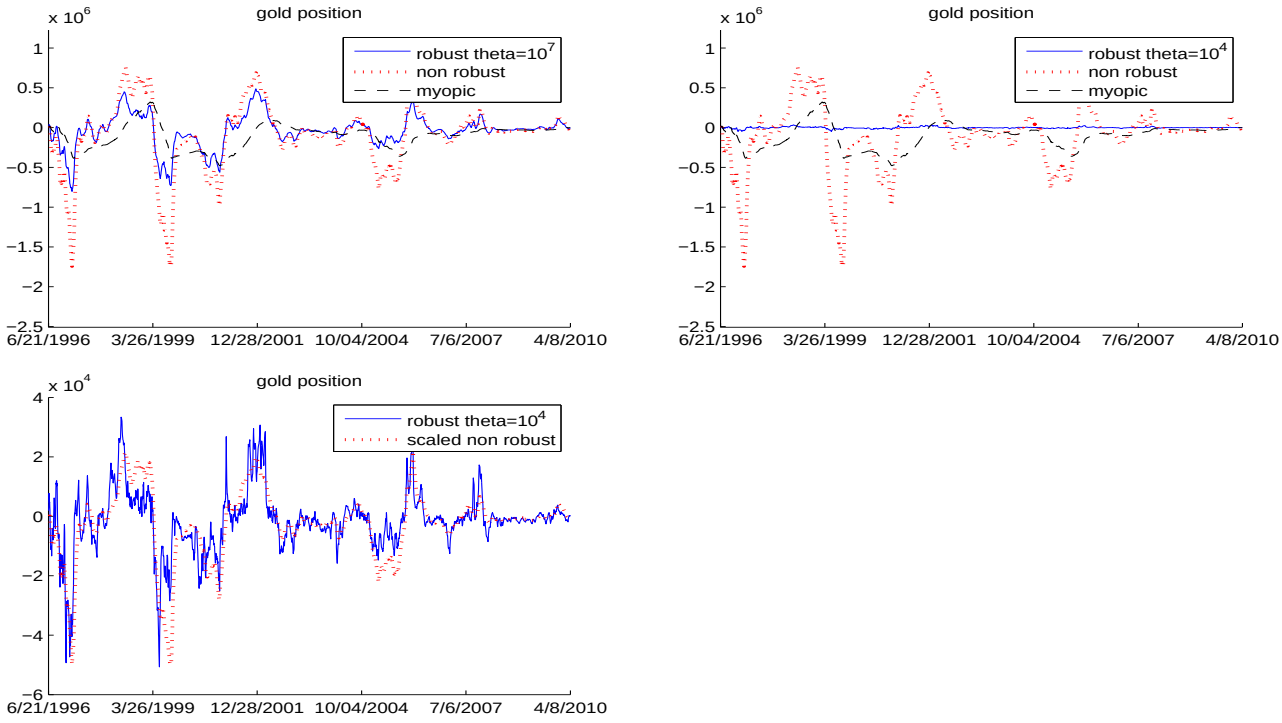


Figure 3.1: Positions in gold in out-of-sample tests under various control rules. The lower-left figure shows the positions for risk scaled non-robust portfolio and corresponding robust portfolio.

the improved Sharpe ratio under the robust rules results mainly from a smaller denominator rather than a larger numerator. We have also observed in QQ-plots (not included) that the tails of the out-of-sample distributions of daily returns of the robust portfolio are lighter than those of the non-robust portfolio.

The largest losses in Figure 3.2 occur near September 27, 1999, and February 2, 2006, so we examine events around these days in greater detail. Leading up to this date, the loading matrix (B_f) and the mean reversion matrix (C_f) were relatively slow moving. As shown in Figure 3.3,

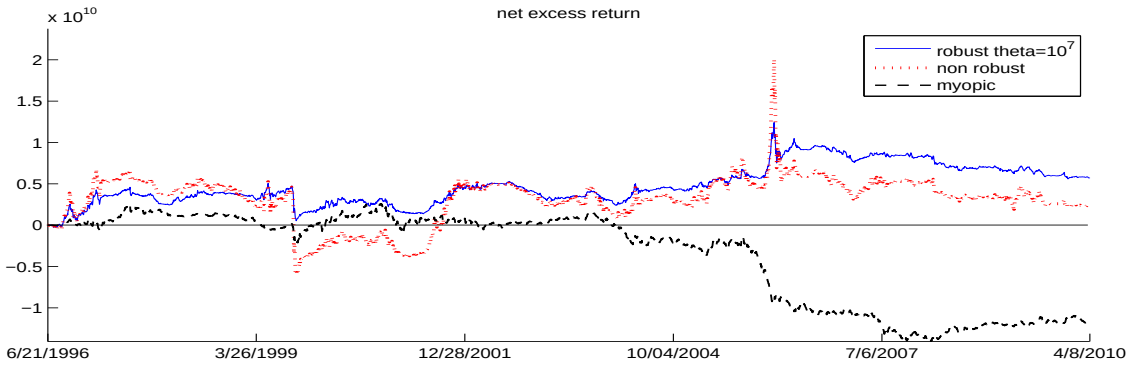


Figure 3.2: Net portfolio returns in out-of-sample tests.

both portfolios had short positions in gold, though more aggressively under the non-robust rule. On September 26, 15 European central banks signed an agreement to limit gold sales (Weber (2003)); the price of gold rose 6% the next day and 11% the day after. This spike results in large losses for the short positions in our test portfolios, but the loss is tempered under the robust rule.

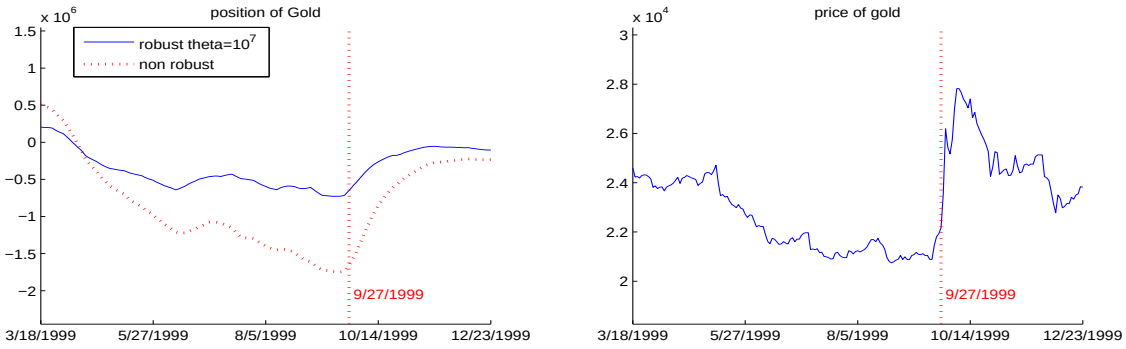


Figure 3.3: Parameter estimates near September 29, 1999

To ensure that our results are not overly influenced by a single day, we repeat the comparison removing days September 27-28, 1999, from the data. Table 3.6 shows that the robust portfolio still outperforms the non-robust portfolio.

Whereas the large price change on September 27, 1999, was limited to gold, changes around February 2, 2006, were spread across multiple commodities, and the portfolio losses resulted from

	θ	Obj $\times 10^{-6}$				mean/std			
		Gross	t-stat	Net	t-stat	Gross	t-stat	Net	t-stat
non-robust		-32.11		-37.98		0.49		0.18	
robust	10^7	-3.89	1.89	-5.45	1.97	0.56	1.05	0.37	2.58
	10^6	-0.02	1.86	-0.58	1.98	0.48	-0.03	0.33	1.24

Table 3.6: Out-of-sample results with two extreme days (9/27-28/1999) removed from the data.

large positions rather than large price changes. The largest positions for both the robust and non-robust portfolios on that date are in aluminum, zinc, gold, and sugar. The prices for these commodities are shown in Figure 3.4.

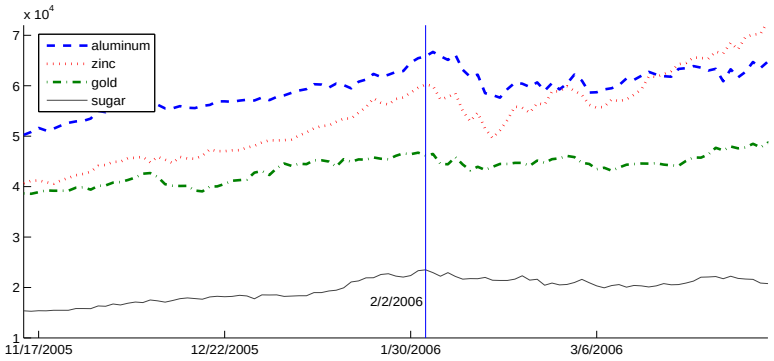


Figure 3.4: Prices for aluminum, gold, zinc, and sugar before and after February 2, 2006. These are the commodities in which the portfolios hold the largest positions on that date.

The position sizes for these commodities are shown in Figure 3.5. The steady price increases in the first half of Figure 3.4 lead to growing positions, particularly for the non-robust portfolio. The positions change smoothly; this is consistent with the representation in (3.3.21) — more precisely, the infinite-horizon version without the superscripts (t, T) — of the portfolio as a weighted average of the previous position and a target position, together with the observation that the factors are moving smoothly as a consequence of the pattern of price changes. The large positions produce large losses on February 2. The price drop in sugar, for example, is barely perceptible, yet it pro-

duces the largest losses of any of the commodities because of the large position accumulated. The robust portfolio suffers smaller losses because it is less aggressive in building up large positions in response to the increasing factor levels. Interestingly, the two portfolios hold fairly similar positions in zinc and gold, despite the large difference in their sugar positions. The non-robust portfolio positions continue to grow quickly following the price drop. We attribute this, informally, to the non-robust portfolio ascribing greater persistence to the factors.

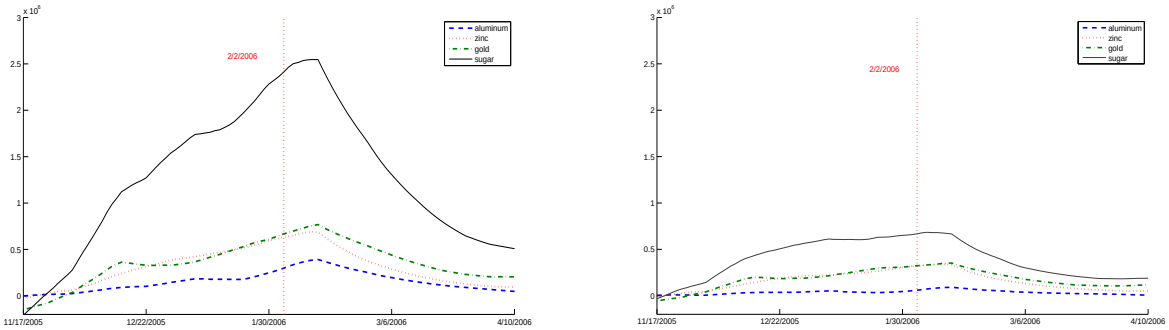


Figure 3.5: Four largest positions for the non-robust (top) and robust (bottom) portfolios around February 2, 2006.

Table 3.7 lists relative entropy values for in-sample tests with standard errors reported in parentheses. Using the results in Theorem 3.3.8, 10^4 sample paths with the same length as the history for in-sample tests are simulated using the estimated model, and the relative entropy is estimated using (3.2.6). For each M_t, the conditional relative entropy $E_t[m_{t+1} \log m_{t+1}]$ is calculated using the closed form (A.3). To achieve the similar level of relative entropy for the robustness only in v with $\theta = 10^9$, one need to set $\theta = 10^7$ when robustness in both v and u are considered. Both objective function and Sharpe ratio with $\theta = 10^7$ in Table 3.3 are better than those with $\theta = 10^9$ in Table 3.4. The difference of these performance measurements suggests that the improvement is brought by considering the extra source of uncertainty from u.

For out-of-sample tests, there is no exact way to capture the relative entropy budget at each time, so we simply use the relative entropy for the same value of θ in in-sample tests.

θ	Robustness in v		Robustness in v and u	
10^{10}	0.020	(1.4×10^{-4})	0.017	(1.2×10^{-4})
10^{9}	0.30	(1.2×10^{-3})	0.17	(6.4×10^{-4})
10^{8}	1.1	(5.7×10^{-3})	0.27	(1.0×10^{-3})
10^{7}	2.4	(0.016)	0.31	(9.7×10^{-4})
10^{6}	5.0	(0.048)	0.40	(8.2×10^{-4})
10^{5}	12	(0.16)	0.44	(8.1×10^{-4})

Table 3.7: Relative entropy for in-sample tests. Standard errors are reported in parentheses.

3.6.4 Scaling and Trimming

In this section, we will compare the robust portfolios with simple heuristics to make the non-robust portfolios less aggressive by adding constraints. We consider three alternatives.

Risk scaled portfolio: For the out-of-sample test, at each time the model is updated, the position of the non-robust portfolio is scaled by a factor. For given robustness level $\theta > 0$, the scaling factor is computed using the previous 6 months' realized return so that the variance of the net return of the scaled portfolio equals that of the robust portfolio. The scaling factor is applied to those positions in the subsequent week. For the first 6 months, we still use their own performance for scaling.

Capital scaled portfolio: First, define the total exposure to be the sum of absolute exposures, $\sum_i |x_{i,t}|p_{i,t}$, with $p_{i,t}$ being the price of the i^{th} asset at time t. Whenever the total exposure of the non-robust portfolio exceeds a predetermined threshold, it will be scaled down proportionately so

that the total exposure of the scaled portfolio equals the threshold. We choose the maximum total exposure of a robust portfolio as the threshold.

Trimmed portfolio: Here we trim the non-robust portfolio so that at any time t the position for each asset will be bounded by some upper and lower bounds. We set the bounds to be the maximum and minimum positions of the robust portfolio for each asset.

Tables 3.9 and 3.10 report out-of-sample performance for these three portfolios. Columns labeled with "RS", "CS" and "T" refer to risk scaled portfolio, capital scaled portfolio and trimmed portfolio, respectively. The $t-$statistics in parentheses, which the compare performance of the robust portfolio with these three portfolios, suggest that none of these alternatives performs consistently as well as the robust method. Actually, the robust portfolio perform significantly better than these three portfolios in terms of objective function when robustness is not too extreme, i.e., when θ is not too small.

Among the three constrained portfolio, the risk scaled portfolios have relatively closer performance to the corresponding robust portfolios. Interestingly, there is a heuristic reason for this. Suppose that we scale down the non-robust portfolio by a factor $s \in (0, 1)$, such that the positions of the resulting portfolio becomes $x_t^s = sx_t^\infty$, where x_t^∞ is the position of the non-robust portfolio. Then

$$x_t^s = (I + 2\Lambda^{-1}A_{xx}^{(t,T)})x_{t-1}^s - 2\Lambda^{-1}A_{xx}^{(t,T)} \times s \times target,$$

where the matrix A is computed under the case $\theta = \infty$. So the scaled portfolio follows original non-robust policy but with scaled target. (Garleanu and Pedersen 2012, Prop. 3) show that under

the specification $\Lambda = \lambda\Sigma_u$ for some $\lambda > 0$, the target portfolio of the non-robust portfolio can be written as a discounted sum of expected myopic portfolios, $(\gamma\Sigma_u)^{-1}B_f f_t$, at all future times. So scaling the target portfolio is very close to scaling up the risk aversion parameter γ to $\gamma^s = \gamma/s$, though the discounting factor for myopic portfolios will change slightly when γ changes.

On the other hand, from (3.3.8), (3.3.9) and Appendix A.1, robustness in the mean of u with $\theta > 0$ is equivalent to increasing γ to γ_θ. Thus, scaling down the non-robust portfolio by s is close to considering robustness in the mean of u with

$$\theta = \frac{s}{((1-s)\beta\gamma)}. \tag{3.6.1}$$

The performance of the risk scaled portfolios in Table 3.10 is close to that of the corresponding robust portfolios. This suggests that most of the improvement is explained by the robustness in u, since the gap between the risk scaled portfolio and the robust portfolio can be considered as the extra benefit brought by considering robustness in v. This is consistent with the observation in Table 3.7, where the performance of the robust portfolio considering only the uncertainty in v is much less than the portfolio with robustness in both u and v at the same relative entropy level.

In Figure 3.1, the lower-left figure shows the position of gold for risk scaled portfolio and the corresponding robust portfolio with $\theta = 10^4$. The robust portfolio is different from the risk scaled portfolio, especially when it has some extreme positions.

Table 3.8 reports average scaling parameters over time. For the cases with robustness in both v and u, the scaling parameters are very close to γ/γ_θ, which supports our observation on the effect of scaling.

θ	10^9	10^8	10^7	10^6	10^5	10^4
robust in v only	0.96	0.81	0.53	0.27	0.11	0.037
robust in v and u	0.53	0.11	0.013	1.5×10^{-3}	1.5×10^{-4}	1.5×10^{-5}
γ/γ_θ	0.5	0.091	9.9×10^{-3}	1.0×10^{-3}	1.0×10^{-4}	1.0×10^{-5}

Table 3.8: Average of scaling parameters for out-of-sample tests, so that the realized variance of the scaled non-robust portfolio equals to that of the robust portfolio.

3.7 Concluding Remarks

We have developed robust portfolio control rules using a stochastic and dynamic notion of robustness to model error. Our analysis covers both finite- and infinite-horizon multi-period problems. We work with a factor model of returns, in which factors evolve stochastically. The relationship between returns and factors and the evolution of the factors are subject to model error and are treated robustly. We incorporate transaction costs and develop simple optimal controls that remain tractable for multiple assets. Robustness significantly improves performance in out-of-sample tests on historical data.

Using this approach requires choosing a value for the parameter θ, which controls the degree of robustness or pessimism. In principle, one would want to select this parameter to reflect the reliability of a model based on the data available to support it. Conveniently, we find that our results are consistent over a wide range of θ values, so the exact choice of this parameter does not dominate our empirical results. Methods for selecting this parameter nevertheless remain an interesting topic for further investigation.

		Obj $\times 10^6$							
	θ	Gross	RS	CS	T	Net	RS	CS	T
non-robust		-36.78				-42.92			
robust	10^9	-31.16	-31.59	-35.43	-34.67	-36.38	-36.97	-41.29	-40.16
			(-0.93)	(-2.13)	(-2.43)		(-0.96)	(-2.22)	(-2.79)
	10^8	-18.70	-19.70	-30.45	-27.31	-22.16	-23.35	-35.70	-32.27
			(-1.22)	(-3.24)	(-2.99)		(-1.19)	(-3.41)	(-3.17)
	10^7	-5.10	-6.70	-21.07	-14.63	-6.70	-8.18	-25.21	-18.37
			(-1.35)	(-3.80)	(-3.03)		(-1.32)	(-4.19)	(-3.49)
	10^6	-0.28	-0.80	-7.92	-6.59	-0.85	-1.33	-11.20	-9.41
			(-0.79)	(-3.52)	(-2.71)		(-0.70)	(-4.19)	(-3.54)
	10^5	0.27	0.15	-1.1	-2.78	0.11	0.05	-1.93	-4.54
			(-0.39)	(-1.18)	(-2.05)		(-0.19)	(-1.77)	(-3.03)
	10^4	0.18	0.14	0.36	-0.50	0.14	0.12	0.18	-1.48
			(-0.41)	(0.45)	(-0.93)		(-0.17)	(0.12)	(-1.96)
		mean/std							
	θ	Gross	RS	CS	T	Net	RS	CS	T
non-robust		0.35				0.04			
robust	10^9	0.37	0.35	0.37	0.38	0.08	0.06	0.06	0.08
			(-1.24)	(-1.05)	(-0.35)		(-1.48)	(-2.64)	(-1.68)
	10^8	0.40	0.35	0.42	0.43	0.16	0.10	0.12	0.13
			(-1.14)	(-0.88)	(0.17)		(-1.19)	(-2.47)	(-1.32)
	10^7	0.44	0.33	0.41	0.53	0.25	0.14	0.13	0.24
			(-0.63)	(-0.33)	(1.14)		(-0.49)	(-1.56)	(0.09)
	10^6	0.42	0.31	0.34	0.42	0.27	0.18	0.10	0.10
			(0.29)	(1.01)	(0.75)		(0.56)	(0.44)	(-0.30)
	10^5	0.37	0.30	0.39	0.17	0.26	0.23	0.21	-0.16
			(0.62)	(1.87)	(0.08)		(1.03)	(1.75)	(-0.80)
	10^4	0.41	0.31	0.44	0.23	0.33	0.29	0.34	-0.10
			(0.19)	(1.41)	(0.11)		(0.64)	(1.53)	(-0.77)

Table 3.9: Out-of-sample performance comparisons using a rolling 6-month estimation window with robustness factor dynamics (v_t) only. Scaled portfolio is derived from scaling the non-robust portfolio so that the realized variance of its net return equals to that of the corresponding robust portfolio. ((RS) is for risk scaled portfolio, and (CS) is for capital scaled portfolio. (T) indicates trimmed portfolio is derived by trimming the position of the non-robust portfolio so that the positions of trimmed portfolio is bounded by the positions of corresponding robust portfolio. The t-statistics in parentheses compare performance of the robust rule with scaling or trimming, based on grouping the data into 40 batches.

		Obj $\times 10^6$							
	θ	Gross	RS	CS	T	Net	RS	CS	T
non-robust		-36.78				-42.92			
robust	10^9	-8.17	-9.25	-26.98	-18.18	-10.5	-11.05	-31.85	-22.37
			(-1.83)	(-3.76)	(-3.74)		(-1.07)	(-4.11)	(-4.39)
	10^8	0.29	0.16	-4.54	-4.11	0.02	0.07	-6.18	-6.24
			(-1.06)	(-2.59)	(-2.44)		(0.38)	(-3.25)	(-3.41)
	10^7	0.10	0.08	0.31	-0.10	0.08	0.08	0.25	-0.62
			(-0.59)	(0.98)	(-0.51)		(0.10)	(0.80)	(-1.31)
	10^6	0.01	0.01	0.04	0.05	0.01	0.01	0.04	-0.01
			(-0.73)	(1.37)	(0.18)		(-0.52)	(1.38)	(-0.39)
	10^5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
			(-0.71)	(1.41)	(0.05)		(-0.68)	(1.41)	(-0.15)
	10^4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
			(-0.70)	(1.41)	(-0.01)		(-0.69)	(1.41)	(-0.04)
		mean/std							
	θ	Gross	RS	CS	T	Net	RS	CS	T
non-robust		0.35				0.04			
robust	10^9	0.39	0.35	0.42	0.53	0.15	0.18	0.13	0.23
			(-0.75)	(-0.72)	(1.25)		(1.75)	(-2.41)	(0.01)
	10^8	0.40	0.35	0.34	0.19	0.26	0.31	0.11	0.15
			(-0.07)	(0.97)	(-0.75)		(1.22)	(0.47)	(-1.72)
	10^7	0.46	0.35	0.44	0.17	0.37	0.34	0.38	-0.14
			(-0.21)	(1.10)	(-0.72)		(0.49)	(1.47)	(-1.45)
	10^6	0.52	0.34	0.43	0.22	0.49	0.34	0.42	-0.00
			(-0.46)	(0.62)	(-0.60)		(-0.24)	(0.82)	(-1.14)
	10^5	0.53	0.35	0.43	0.12	0.52	0.35	0.43	0.05
			(-0.49)	(0.53)	(-0.77)		(-0.46)	(0.56)	(-0.95)
	10^4	0.53	0.35	0.43	0.10	0.53	0.35	0.43	0.09
			(-0.49)	(0.52)	(-0.82)		(-0.49)	(0.53)	(-0.84)

Table 3.10: Out-of-sample performance comparisons using a rolling 6-month estimation window with robustness in returns (u_t) factor dynamics (v_t). Scaled portfolio is derived from scaling the non-robust portfolio so that the realized variance of its net return equals to that of the corresponding robust portfolio. (RS) is for risk scaled portfolio, and (CS) is for capital scaled portfolio. (T) indicates trimmed portfolio is derived by trimming the position of the non-robust portfolio so that the positions of trimmed portfolio is bounded by the positions of corresponding robust portfolio.The t-statistics in parentheses compare performance of the robust rule with scaling or trimming, based on grouping the data into 40 batches.

Chapter 4

Portfolio Rebalancing Error with Jumps and Mean Reversion in Asset Prices

4.1 Introduction

The gap between continuous- and discrete-time modeling for portfolio dynamic is a very practical issue. Continuous modeling is convenient in terms of tractability, while discrete trading is an important realistic feature of trading activity and portfolio management.

We investigate the error in approximating a discretely rebalanced portfolio with one that is continuously rebalanced. We focus on the effects of jumps and mean reversion in the dynamics of the underlying assets, with constant target weights. By increasing the rebalancing frequency to infinity, results are derived for the speed of convergence and the limit of the error, scaled by the square root of the number of rebalancing dates. Specifically, when it involves jumps, the limit follows a compound Poisson distribution; when it only has the mean reversion feature, the

limit follows a normal distribution. Because of the convergence, the closed-form volatility of the continuously rebalanced portfolio can serve as an approximation of that of the discrete portfolio. Thanks to the limiting results, we are able to derive an improved volatility estimator for the discrete portfolio.

The main tools needed for the proof of the main results are the strong approximation results for stochastic differential equations. They are stochastic Taylor expansion, which are developed by Kloeden and Platen (1992) for diffusion processes, and Bruti-Liberati and Platen (2005, 2007) for jump-diffusion processes.

As a direct application, the results of this chapter can be used to guide the construction of efficient simulation algorithm for tail risk measurement of discretely rebalanced portfolio (see Glasserman and Xu (2010)).

The rest of this chapter is organized as follows. Section 4.2 introduces the mean-reverting and jump-diffusion models and states our main results on the limiting rebalancing error. Section 4.3 derives our volatility adjustments for discretely rebalanced portfolios. Numerical examples are given in Section 4.4. The rest of the chapter is then devoted to proving our main results. In Section B.1, we provide background on strong approximation and then apply these tools to our results for the jump-diffusion model. Section B.6 covers the mean-reverting case. Proofs for the volatility adjustments are given in Section B.7. Section 4.5 addresses complications that arise from the possibility of portfolio values becoming negative, which we interpret as a default.

4.2 Model Dynamics and Main Results

We begin by introducing two models of the dynamics of the d underlying assets in the portfolio, one with mean reversion and one with jumps. The first model is as follows:

Exponential Ornstein-Uhlenbeck (EOU) model:

$$\frac{d\underline{S}_i(t)}{\underline{S}_i} = \mu_i dt + dL_i(t), \quad i = 1, \ldots, d,$$

$$dL_i(t) = \kappa(\mu_i^l - L_i)dt + \sigma_i^\top dW(t), \quad L_i(0) = 0.$$

For each $i = 1, \ldots, d$, the drift μ_i and volatility vector $\sigma_i = (\sigma_{i1}, \ldots, \sigma_{id})$ are constants. The model is driven by $W = (W_1, \ldots, W_d)^\top$, a d-dimensional standard Brownian motion, and each L_i is a Ornstein-Uhlenbeck process. We recover geometric Brownian motion as a special case by taking $\kappa = 0$.

We also investigate portfolios under the following dynamics for asset prices:

Jump-Diffusion (JD) model:

$$\frac{dS_i(t)}{S_i(t-)} = \mu_i dt + \sum_{j=1}^{d} \sigma_{ij} dW_j(t) + d(\sum_{j=1}^{N(t)} (Y_j^i - 1)), \quad i = 1, \ldots, d.$$

Here, N is a Poisson process with intensity $0 < \lambda < \infty$, and $Y_j^i > 0$ is the jump size associated with the i^{th} asset at the j^{th} jump of N. The $\{Y_j^i\}_i$ are i.i.d. across different values of j. All of W, N and $\{Y_j^i\}$ are mutually independent. Each S_i is right-continuous, so the left limit $S_i(t-)$ is the value of S_i just prior to a possible jump at t.

The two models could be combined to introduce both mean reversion and jumps in the asset

dynamics. However, our interest lies in analyzing the impact of each of these features, so we keep them separate. To avoid confusion between the two models, we underline variables that are specific to the EOU case.

Given a model of asset dynamics, we consider portfolios defined by a fixed vector of weights $w = (w_1, \ldots, w_d)^\top$, such that $\sum_{i=1}^d w_i = 1$. Interpret w_i as the fraction of value invested in the i^{th} asset. The weights could be the result of a portfolio optimization, but we do not model the portfolio selection problem. In considering only fixed weights, we exclude portfolios in which the weights themselves change with asset prices, and this is a restriction on the scope of our results. Kallsen (2000) showed that under an exponential Levy model such as our JD model, constant weights are in fact optimal for investors with power and logarithmic utilities. There is a sizeable literature that argues the merits of rebalancing to fixed weights. Kim and Omberg (1996) studied portfolio optimization with mean reversion, but their framework does not fit our setting. See, e.g., Chapters 4–6 of Dempster (2009) and the many references cited there.

With continuous rebalancing to target weights $w_1, \ldots, w_d$, the value of the portfolio in the EOU model evolves as

$$\frac{d\underline{V}(t)}{\underline{V}(t)} = \sum_{i=1}^d w_i \mu_i dt + \sum_{i=1}^d w_i \, dL_i(t),$$

and thus

$$\underline{V}(t) = \underline{V}(0) \exp\{(\mu_w - \frac{1}{2}\sigma_w^2)t + \sum_{i=1}^d w_i \sigma_i^\top \int_0^t e^{-\kappa(t-s)} dW_s + (1 - e^{-\kappa t})\bar{\mu}^l)\}, \qquad (4.2.1)$$

where $\bar{\mu}^l = \sum_i w_i \mu_i^l$, $\mu_w = \sum_i w_i \mu_i$, $\Sigma = (\Sigma_{ij})$ with $\Sigma_{ij} = \sum_{k=1}^d \sigma_{ik}\sigma_{jk}$ and $\sigma_w = \sqrt{w^\top \Sigma w}$.

In the jump-diffusion model, portfolio value evolves as

$$\begin{aligned}\frac{d\underset{\sim}{V}(t)}{\underset{\sim}{V}(t-)} = \sum_{i=1}^d w_i \frac{dS_i(t)}{S_i(t-)} &= \mu_w dt + \sum_{i=1}^d w_i \sigma_i^\top dW(t) + \sum_{i=1}^d w_i d(\sum_{j=1}^{N(t)} Y_j^i - 1) \\ &= \mu_w dt + \sigma_w d\tilde{W}(t) + d(\sum_{j=1}^{N(t)} \sum_{i=1}^d w_i (Y_j^i - 1)),\end{aligned}$$

where $\tilde{W}$ is a scalar Brownian motion, $\tilde{W}(t) = \sum_{i,j} w_i \sigma_{ij} W_j(t)/\sigma_w$. This expression assumes that $\underset{\sim}{V}$ remains strictly positive, a requirement we will return to shortly. The solution to this equation is then given by

$$\underset{\sim}{V}(t) = \exp\{(\mu_w - \frac{1}{2}\sigma_w^2)t + \sigma_w \tilde{W}(t)\} \prod_{j=1}^{N(t)} [\sum_{i=1}^d w_i Y_j^i]. \tag{4.2.2}$$

We fix a horizon T over which we analyze the evolution of the portfolio. For the discretely rebalanced case, we fix a rebalancing interval $\Delta t = T/N$, corresponding to a fixed number N of rebalancing dates in $(0, T]$. Denote the value of the discretely rebalanced portfolio by $\hat{\underset{\sim}{V}}$ (or $\underline{\hat{V}}$ in the EOU case). With discrete rebalancing, the portfolio composition is restored to the target weights at each rebalancing opportunity. Thus, the portfolio value evolves as

$$\hat{\underset{\sim}{V}}((n+1)\Delta t) = \hat{\underset{\sim}{V}}(n\Delta t) \sum_{i=1}^d w_i \frac{S_i((n+1)\Delta t)}{S_i(n\Delta t-)}, \quad n = 1, \ldots, N-1,$$

and similarly for $\underline{\hat{V}}$. We normalize the initial portfolio value to $\underset{\sim}{V}(0) = \hat{\underset{\sim}{V}}(0) = \underline{\hat{V}}(0) = 1$.

To ensure that the continuously rebalanced portfolio preserves strictly positive value (i.e., to

rule out bankruptcy), we impose the requirement that, almost surely,

$$\sum_{i=1}^{d} w_i Y^i > 0, \tag{4.2.3}$$

where $Y^1, \ldots, Y^d$ have the distribution of the jump sizes associated with the d assets. That this condition is sufficient can be seen from (4.2.2), and differentiating (4.2.2) reproduces the stochastic differential equation that precedes it. This condition still allows jumps to decrease portfolio value to levels arbitrarily close to zero. It holds automatically if all portfolio weights are positive. The condition is crucial for our analysis because we work with the relative error between the discrete and continuous portfolios, and the denominator in the relative error is the value of the continuous-time portfolio. We also make the following technical assumption on the jump sizes:

$$\left\| \frac{Y^k}{\sum_i w_i Y^i} \right\|_3 < \infty \text{ and } \|Y^k\| < \infty \text{ for } k = 1, \ldots, d; \tag{4.2.4}$$

and later,

$$\left\| \log\left(\frac{Y^k}{\sum_i w_i Y^i}\right) \right\| < \infty. \tag{4.2.5}$$

Here, $\|.\|_3$ indicates the L_3-norm of a random variable, and $\|.\|$ indicates the L_2-norm. Assumptions (4.2.3)–(4.2.4) will be in force whenever we consider the jump-diffusion model; we use (4.2.5) in Section 4.3.

Even under these assumptions, we cannot rule out the possibility that the discretely rebalanced portfolio value drops to zero and lower. We therefore adopt the convention that the portfolio value

is absorbed at zero if it would otherwise become less than or equal to zero; we refer to this event as bankruptcy. We will show (in Section 4.5) that we can ignore the possibility of bankruptcy for our limiting results because the effect becomes negligible asymptotically. Thus, in most of our discussion, we treat the discretely rebalanced portfolio as a positive process.

We now proceed to state our main results for the EOU model. Our first result approximates the relative error between the discrete and continuous portfolios with a sum of independent random variables and identifies the limiting variance of the relative error.

Theorem 4.2.1. *For the EOU model, there exist random variables* $\{\epsilon_{n,N}, n = 1, \dots, N, N = 1, 2, \dots\}$*, with* $\{\epsilon_{1,N}, \dots, \epsilon_{N,N}\}$ *i.i.d. for each* N*, such that*

$$E[(\frac{\hat{\underline{V}}(T) - \underline{V}(T)}{\underline{V}(T)} - \sum_{n=1}^{N} \epsilon_{n,N})^2] = O(\Delta t^2); \tag{4.2.6}$$

in particular, with $\bar{\sigma} = \sum_{i=1}^{d} w_i \sigma_i$

$$\epsilon_{n,N} = \sum_{i=1}^{d} w_i \int_{(n-1)\Delta t}^{n\Delta t} \int_{(n-1)\Delta t}^{s} (\sigma_i - \bar{\sigma})^\top dW(r)(\sigma_i - \bar{\sigma})^\top dW(s),$$

and

$$Var[\epsilon_{n,N}] = \sigma_L^2 \Delta t^2 := [\frac{1}{2}(w^\top(\Sigma \circ \Sigma)w - 2w^\top \Sigma \Omega \Sigma w + (w^\top \Sigma w)^2)]\Delta t^2, \tag{4.2.7}$$

where "$\circ$*" denotes elementwise multiplication of matrices,* Ω *is a diagonal matrix with* $\Omega_{ii} = w_i$*.*

Thus,

$$N\ Var\left[\frac{\hat{\underline{V}}(T)-\underline{V}(T)}{\underline{V}(T)}\right]\to\sigma_L^2T^2.$$

The variance parameter in this result can be understood as

$$\sigma_L^2=Var[\frac{1}{2}(\sum_{i=1}^{d}w_i(\sigma_i^\top Z)^2-(\sum_{i=1}^{d}w_i\sigma_i^\top Z)^2)],$$

where $Z\sim N(0,I)$ in $\mathbb{R}^d$. We now supplement this characterization of the limiting variance with the limiting distribution of the error:

Theorem 4.2.2. *As $N\to\infty$,*

$$\sqrt{N}(\hat{\underline{V}}(T)-\underline{V}(T),\frac{\hat{\underline{V}}(T)-\underline{V}(T)}{\underline{V}(T)})\Rightarrow(\underline{V}(T)\underline{X},\underline{X}),$$

where $\underline{X}\sim N(0,\sigma_L^2T^2)$ is independent of $\underline{V}(T)$, and $\Rightarrow$ denotes convergence in distribution.

The limits in Theorems 4.2.1 and 4.2.2 coincide with those proved in Glasserman (2012) for asset prices modeled by geometric Brownian motion. Thus, we may paraphrase these results as stating that the presence of mean-reversion does not change the relative rebalancing error, as measured by its limiting distribution. The absolute error $\hat{\underline{V}}(T)-\underline{V}(T)$ does change. In both cases, its limiting distribution is that of the independent product of the continuous portfolio ($\underline{V}(T)$ or $\underline{V}(T)$) and X, but the distribution of the continuous portfolio is itself changed by the presence of mean-reversion.

A key feature of Theorem 4.2.2 is the asymptotic independence between the portfolio value and the relative error. We will see, however, that with appropriate scaling there is a non-trivial

covariance between these terms, and the strength of the limiting covariance depends on the speed of mean-reversion. We take up this issue when we consider volatility adjustments in the next section.

We proceed to the limiting variance of the relative error in the jump-diffusion model. For each asset $i = 1, \ldots, d$, introduce the compound Poisson process

$$J_t^i = \sum_{j=1}^{N(t)} \left(\frac{Y_j^i}{\sum_k w_k Y_j^k} - 1 \right).$$

To simplify notation, we define

$$\bar{Y}_j^i = \frac{Y_j^i}{\sum_k w_k Y_j^k} - 1,$$

and then the compensated version of J_t^i becomes $\tilde{J}_t^i = J_t^i - \lambda \mu_i^y t$, where $\mu_i^y = E[\bar{Y}^i]$. Let $\Delta \tilde{J}_n^i = \tilde{J}^i(n\Delta t) - \tilde{J}^i((n-1)\Delta t)$ and $\Delta W_n = W(n\Delta t) - W((n-1)\Delta t)$. Denote $X_N := (\hat{\underset{\sim}{V}}(T) - \underset{\sim}{V}(T))/\underset{\sim}{V}(T)$.

Theorem 4.2.3. *For the JD model, under assumptions (4.2.3) and (4.2.4),*

$$E[(\frac{\hat{\underset{\sim}{V}}(T) - \underset{\sim}{V}(T)}{\underset{\sim}{V}(T)} - \sum_{n=1}^{N} \tilde{\epsilon}_{n,N})^2] = O(\Delta t^2), \tag{4.2.8}$$

where

$$\tilde{\epsilon}_{n,N} = \epsilon_{n,N} + \sum_{i=1}^{d} w_i [b_i^\top \Delta W_n \Delta \tilde{J}_n^i + \int_{(n-1)\Delta t}^{n\Delta t} \int_{(n-1)\Delta t}^{s-} d\tilde{J}^i(r) d\tilde{J}^i(s)], \tag{4.2.9}$$

and $b_i = \sigma_i - \bar{\sigma}$, $i = 1, \ldots, d$. *And*

$$\begin{aligned} Var[\tilde{\epsilon}_{n,N}] &= \tilde{\sigma}_L^2 \Delta t^2 \\ &= Var[\epsilon_{n,N}] + \Delta t^2 (w^\top (b^\top b \circ M) w) + \frac{\Delta t^2}{2} w^\top M \circ M w, \end{aligned}$$

where $Var[\epsilon_{n,N}]$ *is as in (4.2.7),* $b = [b_1, b_2, ..., b_d]$, *and* M *is the* $d \times d$ *matrix with entries*

$$m_{ij} := \lambda E[\bar{Y}^i \bar{Y}^j]. \tag{4.2.10}$$

Thus

$$\begin{aligned} &Var(X_N) \to \tilde{\sigma}_L^2 T^2, \\ &\textit{where } X_N = \sqrt{N} \frac{\hat{\underset{\sim}{V}}(T) - \underset{\sim}{V}(T)}{\underset{\sim}{V}(T)}. \end{aligned}$$

In (4.2.9), the $\epsilon_{n,N}$ are the error terms that arise in the case of geometric Brownian motion (i.e., with $\lambda = 0$ in the JD model and, equivalently, with $\kappa = 0$ in the EOU model). As in the EOU model, the relative error has a limit distribution.

Theorem 4.2.4. *Under assumptions (4.2.3) and (4.2.4), if the jump part is not degenerate, i.e.* $\lambda \neq 0$ *and* $P(Y^i = 1, i = 1, \ldots, d) \neq 1$, *then*

$$\sqrt{N} \frac{\hat{\underline{V}}(T) - \underline{V}(T)}{\underline{V}(T)} \Rightarrow X,$$

where $X \stackrel{d}{=} \underline{X} + \sqrt{T}\sum_{j=1}^{N(t)}\sum_{i=1}^{d} w_i b_i^\top \xi_j \bar{Y}_j^i$ *and* $\xi_j \sim N(0, I)$ *are i.i.d.* d*-dimensional standard normal vectors for* $j \geq 1$*, independent of everything else. The limit does not hold in the* L_2 *sense.*

The jump-diffusion model produces a heavier-tailed distribution for the relative error, resulting in the failure to converge to a limiting normal distribution. One can get some intuition from the asymptotics of $\tilde{\epsilon}_{n,N}$ in (4.2.9), where the third term is nonzero only when there are at least two jumps in the period. Though the third term in (4.2.9) converges to zero in probability, it does contribute to the limiting variance as well as the third absolute moment, both of which are of order $\Theta(\Delta t^2)$.

Because of the presence of $\bar{Y}$ in the limit distribution, we do not have an asymptotic independence result for the JD case, but $\log \underset{\sim}{V}(T)$ and X_N are asymptotically uncorrelated, as shown later in Proposition 4.3.2.

4.3 Volatility Adjustments

We now apply and extend the limiting results of the previous section to develop volatility adjustments that approximate the effect of discrete rebalancing. To motivate this idea, consider the continuous-time dynamics of the portfolio value in (4.2.2), and consider first the case without mean reversion, $\kappa = 0$. In this setting, $\underline{V}$ is a geometric Brownian motion with volatility σ_w, with $\sigma_w^2 = w^\top \Sigma w$, as defined following (4.2.2). The parameter σ_w is a useful measure of portfolio risk under continuous rebalancing. The corresponding parameter for horizon T in the EOU model is (the square root of)

$$\sigma_{w,\kappa}^2 := \frac{1}{T} Var[\log \underline{V}(T)] = \sigma_w^2 \frac{1 - \exp(-2\kappa T)}{2\kappa T}, \tag{4.3.1}$$

and, in the jump-diffusion model, under assumption (4.2.5)

$$\sigma_{w,J}^2 := \frac{1}{T} Var[\log \underset{\sim}{V}(T)] = \sigma_w^2 + \lambda E\left[\left(\log \sum_{i=1}^{d} w_i Y^i\right)^2\right]. \tag{4.3.2}$$

In practice, $\sigma_{w,\kappa}$ and $\sigma_{w,J}$ serve reasonably well for large N as an approximation for discretely rebalanced portfolio. Our objective is to correct these parameters to capture the impact of discrete rebalancing.

4.3.1 Volatility Adjustment with Mean Reversion

From the definition of X_N, we can write value of the discretely rebalanced portfolio as

$$\hat{\underline{V}}(T) = \underline{V}(T)(1 + X_N/\sqrt{N}),$$

which shows that $\hat{\underline{V}}(T)$ is the product of the continuously rebalanced portfolio value and a correction factor that is asymptotically normal and independent of $\underline{V}(T)$. We would like to calculate the "volatility" of $\hat{\underline{V}}(T)$ — the standard deviation of its logarithm, normalized by $\sqrt{T}$ — but because $\hat{\underline{V}}(T)$ is potentially negative, we cannot do this directly. Instead, we note that

$$\bar{\underline{V}}(T) := \underline{V}(T)\exp(X_N/\sqrt{N}) = \hat{\underline{V}}(T) + O_p(1/N),$$

which yields a strictly positive approximation. The $O_p(1/N)$ error in this approximation is negligible compared to the $O_p(1/\sqrt{N})$ difference between the discrete and continuous portfolios, and we will confirm that making this approximation does not change the limiting variance.

For $\underline{\bar{V}}(T)$ we have

$$\begin{aligned}\frac{Var[\log \underline{\bar{V}}(T)]}{T} &= \frac{1}{T}Var[\log \underline{V}(T) + \frac{X_N}{\sqrt{N}}] \\ &= \sigma_w^2 \frac{1-e^{-2\kappa T}}{2\kappa T} + \frac{Var[X_N]}{TN} + \frac{2Cov[\log \underline{V}(T), X_N]}{T\sqrt{N}} \\ &= \sigma_{w,\kappa}^2 + \sigma_L^2 T\Delta t + o(\Delta t) + \frac{2Cov[\log \underline{V}(T), X_N]}{T\sqrt{N}}, \end{aligned} \tag{4.3.3}$$

with $\sigma_{w,\kappa}$ as in (4.3.1) and σ_L^2 the variance parameter in (4.2.7). Although X_N is asymptotically independent of $\underline{V}(T)$, the covariance term does not vanish fast enough to be negligible. In the following proposition, we find the limit of the third term, and verify the validity of replacing $\underline{\hat{V}}$ with $\underline{\bar{V}}$:

Proposition 4.3.1. *(i) The limiting covariance is given by*

$$\sqrt{N}Cov[\log \underline{V}(T), X_N] \to \underline{\gamma}_L T^2,$$

where

$$\underline{\gamma}_L = e^{-\kappa}(\gamma_L + \sum_i w_i(\bar{\sigma}^\top \sigma_i)\kappa(\mu_i^l - \bar{\mu}^l)),$$

with

$$\gamma_L = \mu^\top \Omega\Sigma w - \mu_w \sigma_w^2 + \sigma_w^4 - w^\top \Sigma\Omega\Sigma w. \tag{4.3.4}$$

(ii) Moreover, $E[(\underline{\bar{V}}(T) - \underline{\hat{V}}(T))^2] = O(N^{-2})$*, and*

$$N(Var[\log \underline{\bar{V}}(T)] - Var[\log \underline{V}(T)]) \to (\sigma_L^2 + 2\underline{\gamma}_L)T^2.$$

This result applied to (4.3.3) suggests the following adjustment to the volatility for the discretely rebalanced portfolio:

$$\sigma_{adj}^2 = \sigma_{w,\kappa}^2 + (\sigma_L^2 + 2\underline{\gamma}_L)\Delta t. \tag{4.3.5}$$

At $\Delta t = 0$, we recover the volatility for the continuously rebalanced portfolio, but for small $\Delta t > 0$, the adjusted volatility includes a correction for discrete rebalancing. The parameter γ_L in (4.3.4) is the limiting covariance derived in Glasserman (2012) for assets modeled by multivariate geometric Brownian motion; thus, at $\kappa = 0$ we recover the volatility adjustment derived there in the absence of mean reversion, as expected. The second part of the proposition confirms that the difference between $\underline{\bar{V}}(T)$ and $\underline{\hat{V}}(T)$ is negligible. In Section 4.4.2, we present numerical results illustrating the performance of the volatility adjustment (4.3.5) in approximating the effect of discrete rebalancing.

4.3.2 Volatility Adjustment in the Jump-Diffusion Model

We follow similar steps in the jump-diffusion model. We set $\underset{\sim}{\bar{V}}(T) := \underset{\sim}{V}(T)\exp(X_N/\sqrt{N})$ with

$$X_N = \sqrt{N}\sum_{n=0}^{N-1}\left(\frac{\underset{\sim}{\hat{V}}((n+1)\Delta t)}{\underset{\sim}{V}((n+1)\Delta t)} - \frac{\underset{\sim}{\hat{V}}(b\Delta t)}{\underset{\sim}{V}(n\Delta t)}\right),$$

and then

$$\frac{Var[\log \bar{\underset{\sim}{V}}(T)]}{T} = \sigma_{w,J}^2 + \frac{Var[X_N]}{TN} + \frac{2Cov[\log \underset{\sim}{V}(T), X_N]}{T\sqrt{N}}, \tag{4.3.6}$$

with $\sigma_{w,J}$ as defined in (4.3.2).

Proposition 4.3.2. *(i) The limiting covariance is given by*

$$\sqrt{N}Cov[\log \underset{\sim}{V}(T), X_N] \to \tilde{\gamma}_L T^2,$$

where

$$\tilde{\gamma}_L := \gamma_L + \lambda[\sum_i w_i \bar{\sigma}^\top \sigma_i \mu_i^y] + \lambda \sum_i w_i(\mu_i - \sigma_i^\top \bar{\sigma} + \lambda \mu_i^y) E[\bar{Y}^i(\log \sum_l w_l Y^l - \mu_J)]$$

and

$$\mu_J = E[\log \sum_i w_i Y_j^i].$$

(ii) Moreover, $E[(\hat{\underset{\sim}{V}}(T) - \bar{\underset{\sim}{V}}(T))^2] = O(N^{-2})$ *and*

$$N(Var[\log \bar{\underset{\sim}{V}}(T)] - Var[\log \underset{\sim}{V}(T)]) \to (\tilde{\sigma}_L + \tilde{\gamma}_L)T^2.$$

The resulting volatility adjustment is

$$\tilde{\sigma}_{adj}^2 = \sigma_{w,J}^2 + (\tilde{\sigma}_L^2 + 2\tilde{\gamma}_L)\Delta t. \tag{4.3.7}$$

The asymptotic variance parameters for the relative error (σ_L^2 and $\tilde{\sigma}_L^2$) do not depend on the drift parameters μ_i, but, interestingly, the drifts do appear in the asymptotic covariance γ_L (and $\underline{\gamma}_L$ and $\tilde{\gamma}_L$). We will see that in a stochastic Taylor expansion of the relative error, the μ_i appear only in those terms with norms of order $O(\Delta t^{3/2})$. For the variance, it turns out that only terms with norms up to order $O(\Delta t)$ are relevant, but the covariance involves terms of norm $O(\Delta t^{3/2})$, and these involve the μ_i.

Since the volatility adjustments are explicitly related to the weights, one could reverse the approximation as a guideline for adjusting portfolio weights to control the portfolio volatility σ with discrete rebalancing.

4.4 Numerical Experiments and Further Discussion

4.4.1 Example for the Jump-Diffusion Model

We begin with the JD model model and examine the approximation for the relative error provided by Theorem 4.2.4.

We calibrated the JD model from the daily returns of global equity indices based on the method introduced in Das and Uppal (2004). The weights are computed as the optimal weights for power utility with risk aversion parameter $\gamma = 2$ following the results of Das and Uppal (2004)[1]. The data used is from March 2009 to March 2011, and the calibrated results are as in Table 4.1. Jump sizes are modeled by Merton's jump model with $\log(Y^i) \sim N(\mu_J^i, \sigma_J^i)$. We calibrate the parameters by

[1]The negative weights could cause defaults, even in the continuous portfolio, though this occurs very rarely with out estimated value of σ_J. In our numerical examples, we exclude paths with defaults. We address this issue in Section 4.5.

assuming the jump sizes are perfectly correlated as in Das and Uppal (2004). However, perfectly correlated jumps would have the same effect as constant jump sizes because we are considering relative error. To make the example more interesting, we simulate independent jumps sizes instead.

	SP500	FTSE	NIK	DAX	SSMI	CAC	STI	HSI	MXX	SET50	MERV
λ	3.0142										
w	-1.22	-0.22	0.22	0.87	-3.30	0.82	0.44	-0.47	1.32	1.17	1.38
μ	0.15	0.13	0.12	0.25	0.09	0.12	0.17	0.21	0.25	0.35	0.40
$\mu_J \times 10^{-2}$	-0.74	0.24	-1.71	-0.10	0.22	1.28	0.00	0.18	-0.85	-0.01	0.46
$\sigma_J \times 10^{-2}$	2.91	2.65	1.47	2.92	2.24	4.68	2.46	2.87	2.58	3.56	4.69
$\Sigma \times 10^{-2}$	3.14	2.00	0.27	2.35	1.52	2.56	0.50	0.44	2.14	0.41	3.17
	2.00	2.84	0.75	2.94	1.93	3.26	0.92	1.00	1.72	1.05	2.48
	0.27	0.75	4.53	0.60	0.77	1.03	1.76	2.74	0.33	1.81	0.30
	2.35	2.94	0.60	3.83	2.29	3.87	0.92	0.98	2.00	1.12	2.93
	1.52	1.93	0.77	2.29	2.02	2.54	0.65	0.68	1.14	0.72	1.78
	2.56	3.26	1.03	3.87	2.54	4.37	1.05	1.14	2.08	1.16	3.02
	0.50	0.92	1.76	0.92	0.65	1.05	2.54	2.47	0.68	1.86	0.78
	0.44	1.00	2.74	0.98	0.68	1.14	2.47	4.65	0.88	2.58	0.68
	2.14	1.72	0.33	2.00	1.14	2.08	0.68	0.88	2.74	0.73	2.65
	0.41	1.05	1.81	1.12	0.72	1.16	1.86	2.58	0.73	4.88	0.91
	3.17	2.48	0.30	2.93	1.78	3.02	0.78	0.68	2.65	0.91	6.23

Table 4.1: Parameters estimated from S&P 500, FTSE 100, Nikkei 225, DAX, Swiss Market Index, CAC 40, FTSE Straits Times Index for Singapore, Hang Seng, Mexico IPC, Thai Set 50 and Argentina Merval.

Figure 4.1 shows QQ plots of the value of discrete portfolios versus the limit as described in Theorem 4.2.4, both simulated over 2500 replications. We choose N to be 4, 12 and 360 to represent quarterly, monthly and daily rebalancings. As the number of steps N gets larger, the figure indicates convergence to the theoretical limit, though relatively slower than in the EOU model.

Since the limiting distribution is not normal, we do not have an asymptotic independence result of the type in Theorem 4.2.2. But the numerical results in Table 4.2 still show the correlation

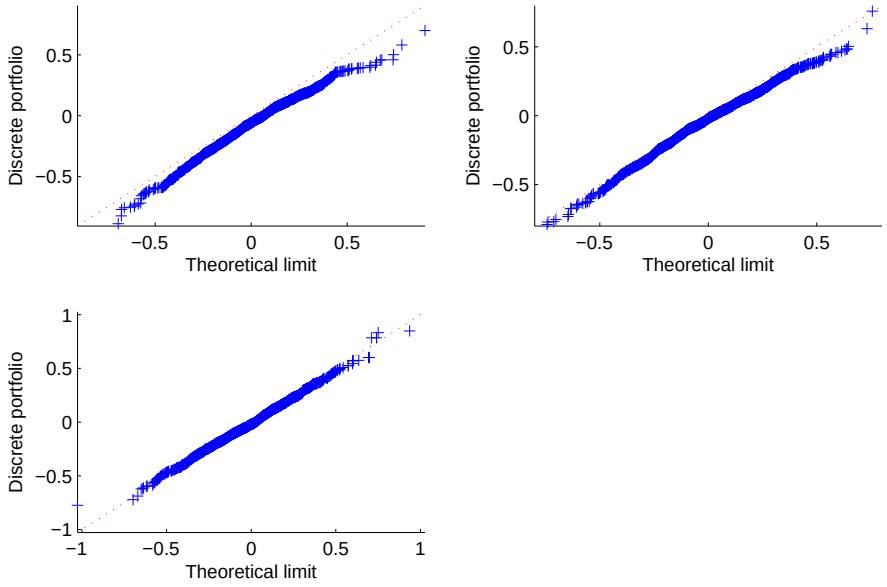


Figure 4.1: Jump-diffusion model: QQ plots of X_N versus X at $N = 4$ (upper left), $N = 12$ (upper right), $N = 360$ (lower left).

between $\log \underset{\sim}{V}(T)$ and X_N decreasing toward zero as N increases. This is to be expected because part (i) of Proposition 4.3.2 shows the covariance of $\log \underset{\sim}{V}(T)$ and X_N converging to zero at rate $O(1/\sqrt{N})$, and X_N has a non-degenerate limiting variance. In separate experiments, we have found large discrepancies in the QQ plots when σ_J^i are doubled. Estimation of m_{ij} in (4.2.10) becomes unstable, and condition (4.2.4) may be violated. Table 4.3 shows the error reduction of

N	4	12	360
JD	-12%	-13%	-4%
EOU	-85%	-61%	-13%

Table 4.2: Correlations for JD model and EOU model, between $\log \underset{\sim}{V}(T)$ (or $\log \underline{V}(T)$) and X_N, with 2500 replicates.

volatility as

$$1 - |\frac{\tilde{\sigma}_{adj} - \hat{\sigma}_N}{\sigma_{w,J} - \hat{\sigma}_N}|, \tag{4.4.1}$$

where $\tilde{\sigma}_{adj}$ is defined in (4.3.7). This measure shows the relative improvement achieved in approximating the volatility using the adjustment; a small value indicates small improvement, and a value close to 1 indicates good improvement. These estimates are based on 50,000 replications. When the correlation between $\underset{\sim}{V}(T)$ and X_N is small, the error reduction tends to be unstable. As suggested by (4.3.6), when N is small and the covariance term in (4.3.6) is negative, the error reduction can be small, or even negative. In this situation, numerical errors, especially from computing the required expectation of the $\bar{Y}^i$, can contaminate the results.

N	4	12	360
JD	87%	46%	2%
EOU	69%	55%	18%

Table 4.3: Volatility error reductions for JD model and EOU model, with 50,000 replications. Formula (4.4.1) and (4.4.2) are used for JD model and EOU model, respectively.

4.4.2 Example for the EOU model

For the purpose of illustration, we use the same parameters w, μ and Σ from Section 4.4.1. We use the mean-reversion rate $\kappa = 1$ and long-run levels $\mu_i^l = 0.1 \times i/d$, $i = 1, \ldots, d$. Figure 4.2 illustrates the convergence to normality as N increases, using 2500 replicates.

Table 4.2 reports estimated correlations between $\log \underset{\sim}{V}(T)$ and X_N using the same parameters as Figure 4.2. As expected, the correlation decreases toward zero as N increases.

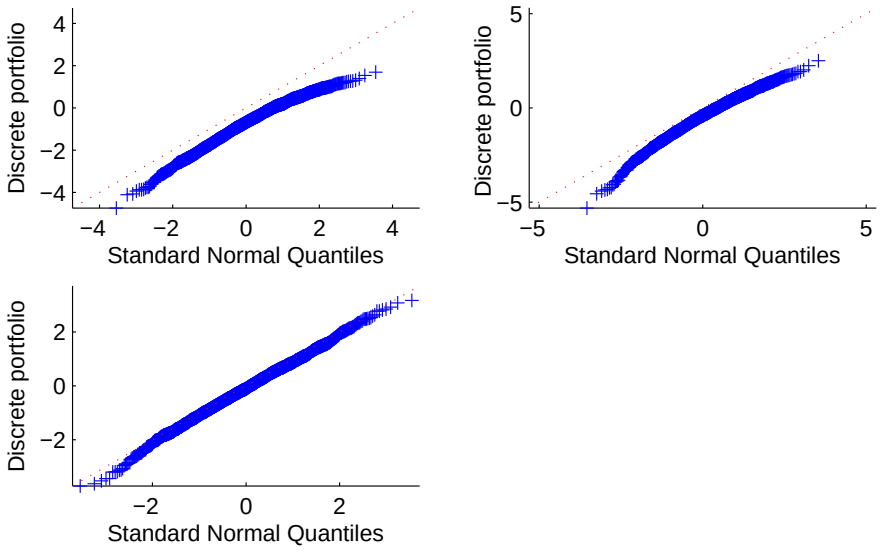


Figure 4.2: EOU model: QQ plots of $X_N/\underline{\sigma}_L T$ versus standard normal at $N = 4$ (upper left), $N = 12$ (upper right) and $N = 360$ (lower left).

Table 4.3 evaluates the volatility adjustment by reporting the estimated error reduction using the adjustment, calculated as

$$1 - |\frac{\sigma_{adj} - \hat{\sigma}_N}{\sigma_{w,\kappa} - \hat{\sigma}_N}|, \tag{4.4.2}$$

where σ_{adj} is defined in (4.3.5) and $\hat{\sigma}_N$ is the volatility of the discretely rebalanced portfolio as estimated by simulation. The results in Table 4.3 show appreciable error reduction, especially when the number of rebalancing dates N is small. When N becomes large, the denominator $\sigma_{w,\kappa} - \hat{\sigma}$ will become very small. The magnitude of the reduction is not necessarily monotone in N. More examples for the diffusion case without mean reversion can be found in Glasserman (2012).

4.5 Dealing with Defaults

As explained in Section 4.2, jumps in asset values can produce negative portfolio values, even under continuous rebalancing. Here we address this issue in greater detail.

Assume that once a portfolio defaults (i.e., drops to zero or below), it is absorbed at zero forever. It follows from (4.2.2) that such a default occurs in a continuously rebalanced portfolio if and only if there is a jump before time T with $\sum_i w_i Y^i \leq 0$. Under assumption (4.2.3), the continuously rebalanced portfolio will therefore never default.

The discretely rebalanced portfolio will default at time t in the n^{th} time interval if and only if t is the first time that $t \in [(n-1)\Delta t, n\Delta t]$ with $\tilde{t} = t - \Delta t \lfloor \frac{t}{\Delta t} \rfloor$ and

$$\hat{R}_{n,N}(t) = \left(\frac{\hat{\underset{\sim}{V}}(t)}{\hat{\underset{\sim}{V}}((n-1)\Delta t)} \right) = \sum_{i=1}^{d} w_i \exp\{(\mu_i - \frac{1}{2}\sum_{j=1}^{d} \sigma_{ij}^2)\tilde{t} + \sigma_i^\top W(\tilde{t})\} \prod_{j=1}^{N(\tilde{t})} Y_j^i \leq 0. \quad (4.5.1)$$

Let I_d^n denote the indicator of default for the discrete portfolio, where $I_d^n = 1$ means that the portfolio defaults in n^{th} time interval, while $I_d^n = 0$ if not.

Proposition 4.5.1. *Under conditions (4.2.3) and (4.2.4), we have*

$$\| \frac{\hat{\underset{\sim}{V}}(T) - \underset{\sim}{V}(T)}{\underset{\sim}{V}(T)} \| - \| \frac{\hat{\underset{\sim}{V}}(T) - \underset{\sim}{V}(T)}{\underset{\sim}{V}(T)} I_{\{I_d^n = 0 \textit{ for all } n=1,\ldots,N\}} \| = O(\Delta t).$$

Proposition 4.5.1 confirms that we can ignore possible defaults in the discretely rebalanced portfolio, because the limits in Theorem 4.2.2 and 4.2.4 are scaled by $\sqrt{N} = \Delta t^{-1/2}$, while the errors introduced by ignoring defaults are of order $O(\Delta t)$. In fact, we can even weaken our as-

sumptions to allow $\sum_i w_i Y^i \leq 0$, replacing (4.2.3) with the condition

$$E[\left(\frac{w_j Y^j}{\sum_i w_i Y^i}\right)^2 | \sum_i w_i Y^i < 0] < \infty, \text{ for all } j = 1, ..., d,$$

This suffices to show that defaults have a negligible effect on the relative error using a similar argument.

4.6 Concluding Remarks

In this chapter, we have analyzed the error between a discretely rebalanced portfolio and its continuously rebalanced counterpart in the presence of jumps or mean-reversion in the underlying asset dynamics. With discrete rebalancing, the portfolios composition is restored to a set of fixed target weights at discrete intervals; with continuous rebalancing, the target weights are maintained at all times. We examined the difference between the two portfolios as the number of discrete rebalancing dates increases. With either mean reversion or jumps, we derive the limiting variance of the relative error between the two portfolios. With mean reversion and no jumps, we show that the scaled limiting error is asymptotically normal and independent of the level of the continuously rebalanced portfolio. With jumps, the scaled relative error converges in distribution to the sum of a normal random variable and a compound Poisson random variable.

For both the mean-reverting and jump-diffusion cases, we derive volatility adjustments to improve the approximation of the discretely rebalanced portfolio by the continuously rebalanced portfolio, based on on the limiting covariance between the relative rebalancing error and the level

of the continuously rebalanced portfolio. These results are based on strong approximation results for jump-diffusion processes.

Appendix A

Additional Proofs for Chapter 3

A.1 Proof of Proposition 3.3.6

Proof. Proof Set $x_+ = x + \Delta x_+$ and $f_+ = C_f f + v_+$ with $v_+ \sim N(0, \Sigma_v)$. By substituting a quadratic expression for $V_{t+1,T}$ into (3.3.8), applying (3.3.9), and then substituting for f_+, the maximization problem becomes

$$\begin{aligned}
&\max_{\Delta x_+ \in \mathbb{R}^d} \Bigg\{ x_+^\top B_f f - \frac{\gamma_\theta}{2} x_+^\top \Sigma_u x_+ - \frac{1}{2} \Delta x_+^\top \Lambda \Delta x_+ \\
&\qquad -\theta\beta \log E[\exp\{-\frac{1}{\theta}(x_+^\top A_{xx} x_+ + x_+^\top A_{xf} f_+ + f_+^\top A_{ff} f_+ + A_0)\}] \Bigg\} \\
= &\max_{\Delta x_+ \in \mathbb{R}^d} \Bigg\{ x_+^\top B_f f - \frac{\gamma_\theta}{2} x_+^\top \Sigma_u x_+ - \frac{1}{2} \Delta x_+^\top \Lambda \Delta x_+ + \beta x_+^\top A_{xx} x_+ + \beta A_0 \\
&\qquad - \theta\beta \log E[\exp\{-\frac{1}{\theta}[x_+^\top A_{xf}(C_f f + v_+) + (C_f f + v_+)^\top A_{ff}(C_f f + v_+)]\}] \Bigg\},
\end{aligned}$$

where the expectation is over v_+. Evaluating this expectation, the problem becomes

$$\max_{\Delta x_+ \in \mathbb{R}^d} \Big\{ x_+^\top B_f f - \frac{\gamma_\theta}{2} x_+^\top \Sigma_u x_+ - \frac{1}{2}\Delta x_+^\top \Lambda \Delta x_+ + \beta x_+^\top A_{xx} x_+ + \beta A_0 \\ - \frac{\beta}{2\theta}(2f^\top C_f^\top A_{ff} + x_+^\top A_{xf})(\Sigma_v^{-1} + \frac{2}{\theta}A_{ff})^{-1}(2f^\top C_f^\top A_{ff} + x_+^\top A_{xf})^\top \\ + \beta x_+^\top A_{xf} C_f f + \beta (C_f f)^\top A_{ff}(C_f f) - \frac{\theta\beta}{2}\log(|\Sigma_v^{-1} + \frac{2}{\theta}A_{ff}|^{-1}/|\Sigma_v|) \Big\}. \tag{A.1}$$

The terms involving x_+ in (A.1) are contained in a quadratic function of the form $-\frac{1}{2}x_+^\top J_1 x_+ + x_+^\top J_2$ with J_1 and J_2 defined in (3.3.10) and (3.3.11), respectively (taking $x_{t-1} = x$ and $f_t = f$ in J_2). Condition 3.3.4 guarantees that the maximum is achieved at $x_+ = J_1^{-1}J_2$.

Substituting the optimal control into (A.1) and collecting terms, the maximum value becomes

$$\frac{1}{2}J_2^\top J_1^{-1} J_2 - \frac{1}{2}x^\top \Lambda x + \beta A_0 + \beta (C_f f)^\top A_{ff} C_f f \\ -2\frac{\beta}{\theta}(C_f f)^\top A_{ff}(\Sigma_v^{-1} + \frac{2}{\theta}A_{ff})^{-1} A_{ff} C_f f + \frac{\beta\theta}{2}\log|I + \frac{2}{\theta}\Sigma_v A_{ff}|.$$

This expression can be written as $x^\top \mathcal{S}(A)_{xx} x + x^\top \mathcal{S}(A)_{xf} f + f^\top \mathcal{S}(A)_{ff} f + \mathcal{U}(A, A_0)$ with $\mathcal{S}(A)$ and $\mathcal{U}(A, A_0)$ as defined in (3.3.13)–(3.3.16). □

A.2 Proof of Proposition 3.3.7

Proof. Proof (i) Symmetry is immediate from (3.3.13) and (3.3.15). (ii) Let $J_{(n),1}$ denote the value of J_1 obtained by replacing A with $\mathcal{S}^n(A)$, with $J_{(0),1} = J_1$ and $\mathcal{S}^0(A) = A$. Take as induction hypothesis that $J_{(k),1} > 0$, $\Lambda^{1/2} J_{(k),1}^{-1} \Lambda^{1/2} < I$, and $\mathcal{S}^k(A)_{ff} \geq 0$, for all $k = 0, 1, \ldots, n$; these

conditions hold at $n = 0$. From (3.3.13) we have $\mathcal{S}^n(A)_{xx} = (-\Lambda + \Lambda J^{-1}_{(n-1),1}\Lambda)/2$, so the induction hypothesis implies that $-\Lambda/2 < \mathcal{S}^n(A)_{xx} < 0$. We need to show that the properties asserted in the induction hypothesis are preserved at $k = n + 1$. We get $J_{(n+1),1} > 0$ from (3.3.10) using $\gamma_\theta \Sigma_u > 0$, $\Lambda > 0$, $\mathcal{S}^n(A)_{xx} < 0$ and $\mathcal{S}^n(A)_{ff} \geq 0$. Next, using (3.3.10), $\Lambda^{1/2} J^{-1}_{(n+1),1} \Lambda^{1/2}$ evaluates to

$$\left(I + \Lambda^{-1/2} \left[\gamma_\theta \Sigma_u - 2\beta \mathcal{S}^n(A)_{xx} + \frac{\beta}{\theta} \mathcal{S}^n(A)_{xf} (\Sigma_v^{-1} + \frac{2}{\theta} \mathcal{S}^n(A)_{ff})^{-1} \mathcal{S}^n(A)_{xf}^\top \right] \Lambda^{-1/2} \right)^{-1}.$$

By the induction hypothesis, the term in square brackets is positive definite, from which we get $\Lambda^{1/2} J^{-1}_{(n+1),1} \Lambda^{1/2} < I$.

It remains to show that $\mathcal{S}^{(n+1)}(A)_{ff} \geq 0$. To simplify notation, we detail the case $n = 0$, which allows us to write $\mathcal{S}(A)$ as A; the same argument applies for all n. Because A_{ff} is symmetric and positive semi-definite and $\Sigma_v > 0$, we have

$$\begin{pmatrix} A_{ff} & A_{ff} \\ A_{ff} & A_{ff} + \frac{\theta}{2} \Sigma_v^{-1} \end{pmatrix} \geq 0.$$

By the second property at the top of p.651 of Boyd and Vandenberghe (2004), the Schur complement

$$A_{ff} - A_{ff} (\frac{\theta}{2} \Sigma_v^{-1} + A_{ff})^{-1} A_{ff}$$

is positive semi-definite, and so then is $\beta C_f^\top (A_{ff} - A_{ff} (\frac{\theta}{2} \Sigma_v^{-1} + A_{ff})^{-1} A_{ff}) C_f$, which combines

the first two terms in (3.3.15). The last term in (3.3.15) is $J_3^\top J_1^{-1} J_3/2 \geq 0$, so we conclude that $\mathcal{S}(A)_{ff}$ is positive semi-definite. □

A.3 Proof of Theorem 3.3.8

Proof. Proof The investor's optimal choice (3.3.19) is established in the proof of Proposition 3.3.6. For the adversary, we know from the fact that the exponent in the numerator on the right side of (3.3.7) is quadratic in v_{t+1} and linear in u_{t+1} that m_{t+1}^* is well-defined and is a ratio of multivariate normal densities that takes the form

$$\begin{aligned}\log m_{t+1}^* =& Constant - \frac{1}{2}[u_{t+1}^\top - \mu_{u,t+1}^\top, v_{t+1}^\top - \mu_{v,t+1}^\top](\tilde{\Sigma})_{t+1}^{-1} \begin{bmatrix} u_{t+1} - \mu_{u,t+1} \\ \\ v_{t+1} - \mu_{v,t+1} \end{bmatrix} \\ &+ \frac{1}{2}[u_{t+1}^\top, v_{t+1}^\top] \begin{bmatrix} \Sigma_u^{-1} & 0 \\ \\ 0 & \Sigma_v^{-1} \end{bmatrix} \begin{bmatrix} u_{t+1} \\ \\ v_{t+1} \end{bmatrix}.\end{aligned}$$

On the other hand, by substituting the quadratic form of $V_{t+1,T}$ into the right side of (3.3.7) we get

$$\begin{aligned}\log m_{t+1}^* = &- \frac{1}{\theta}\Big(x_t^\top A_{xx}^{(t+1,T)} x_t + x_t^\top A_{xf}^{(t+1,T)}[C_f f_t + v_{t+1}] \\ &+ [v_{t+1} + C_f f_t]^\top A_{ff}^{(t+1,T)}[v_{t+1} + C_f f_t] + A_0^{(t+1,T)}\Big) \\ &- \frac{1}{\beta\theta} x_t^\top u_{t+1} - \log E_t[\exp\{-\frac{1}{\theta} V_{t+1,T}\}] - \frac{1}{2\beta\theta} x_t^\top \Sigma_u x_t.\end{aligned}$$

Matching the coefficients of u_{t+1} and v_{t+1} yields the expressions for $\mu_{v,t+1}$, $\mu_{u,t+1}$ and $\tilde{\Sigma}_t$. □

A.4 Proof of Corollary 3.3.9

Equation (3.3.21) is immediate. For (3.3.22), using (3.3.19) and substituting (3.3.11) and (3.3.20), we get

$$x_t = J_1^{-1}\left(B_f f_t + \Lambda x_{t-1} + \beta A_{xf}^{(t+1,T)}(\tilde{E}_t[f_{t+1}] - \mu_{v,t+1}) - \frac{2\beta}{\theta}A_{xf}^{(t+1,T)}(\tilde{\Sigma}_v)_{t+1}A_{ff}^{(t+1,T)}C_f f_t\right).$$

Moving J_1 to the left-hand side, and substituting (3.3.10) and (3.3.17), we have

$$\begin{aligned}
&\left(\gamma_\theta \Sigma_u + \Lambda - 2\beta A_{xx} + \frac{\beta}{\theta}A_{xf}(\Sigma_v^{-1} + \frac{2}{\theta}A_{ff})^{-1}A_{xf}^\top\right)x_t \\
&= \left(B_f f_t + \Lambda x_{t-1} + \beta A_{xf}^{(t+1,T)}\tilde{E}_t[f_{t+1}] + \frac{\beta}{\theta}A_{xf}^{(t+1,T)}\tilde{\Sigma}_{v,t+1}((A_{xf}^{(t+1,T)})^\top x_t + 2A_{ff}^{(t+1,T)}C_f f_t)\right. \\
&\left.-\frac{2\beta}{\theta}A_{xf}^{(t+1,T)}(\tilde{\Sigma}_v)_{t+1}A_{ff}^{(t+1,T)}C_f f_t\right).
\end{aligned}$$

Rearranging terms and, as before, writing γ_θ for $\gamma + (1/\theta\beta)$, we get equation (3.3.22).

A.5 Proof of Lemma 3.4.1

Proof. Proof We use induction on $n = T - t$. The statement of the lemma holds at $n = 0$. For some $t \leq T$, suppose $U_{t,T}^{\theta_1}(1, x, f) \leq U_{t,T}^{\theta_2}(1, x, f)$, for all (x, f). Then for any (x_{t-1}, f_t) and any m_t,

$$\begin{aligned}
&x_{t-1}^\top E[m_t u_t] + \beta E[m_t U_{t,T}^{\theta_1}(1, x_{t-1}, f_t) + \theta_1 \beta m_t \log m_t] \\
&\leq x_{t-1}^\top E[m_t u_t] + \beta E[m_t U_{t,T}^{\theta_2}(1, x_{t-1}, f_t) + \theta_2 \beta m_t \log m_t],
\end{aligned}$$

because the relative entropy is nonnegative. It follows that the infima over m_t of the two sides are ordered the same way, and then also the suprema over x_t in evaluating $\mathcal{T}(U^{\theta_i}_{t-1,T})$ using (3.3.2) Thus, $U^{\theta_1}_{t-1,T}(1, x_{t-1}, f_t) \leq U^{\theta_2}_{t-1,T}(1, x_{t-1}, f_t)$. Since (x_{t-1}, f_t) is arbitrary and the conditions for Proposition 3.3.7 are in force, this entails $\mathcal{S}^{n+1}_{\theta_1}(A) \leq \mathcal{S}^{n+1}_{\theta_2}(A)$ and $\mathcal{U}^{n+1}_{\theta_1}(A, A_0) \leq \mathcal{U}^{n+1}_{\theta_2}(A, A_0)$.

□

A.6 Proof of Proposition 3.4.2

Proof. Proof Consider the modified problem

$$\tilde{U}(x_0, f_1) = \sup_{\Delta x} \sum_{t=1}^{\infty} E[\beta^t(-\frac{1}{2\gamma} f_t^\top B_f^\top \Sigma_u^{-1} B_f f_t + x_t^\top B_f f_t - \frac{\gamma}{2} x_t^\top \Sigma_u x_t - \frac{1}{2}\Delta x_t^\top \Lambda \Delta x_t)]. \quad \text{(A.1)}$$

The term we have added (quadratic in f_t) completes the square in (x_t, f_t) and makes the one-step reward in the modified problem non-positive for all t under any policy. It does so without changing the optimal control because the evolution of f_t is unaffected by the investor's decisions.

The added terms make (A.1) a standard LQG problem to which we can apply standard results. The corresponding dynamic programming operator maps the set of negative semi-definite quadratic functions into itself; thus, we may abbreviate it as mapping coefficients $(\tilde{A}, \tilde{A}_0)$, $\tilde{A} \leq 0$, to coefficients $(\tilde{\mathcal{S}}(\tilde{A}), \tilde{\mathcal{U}}(\tilde{A}, \tilde{A}_0))$. We can invoke Proposition 4.4.1 of Bertsekas (2007) by setting $\bar{x}_t = \beta^{t/2} x_t$ and $\bar{f}_t = \beta^{t/2} f_t$ and recalling that $\sigma(C_f) < 1$. Starting from any $(\tilde{A}, \tilde{A}_0)$ with $\tilde{A} \leq 0$, iterative application of the dynamic programming operator for the modified problem produces a

convergent sequence; that is, $\lim_{n\to\infty}(\tilde{\mathcal{S}}^n(\tilde{A}), \tilde{\mathcal{U}}^n(\tilde{A}, \tilde{A}_0))$ exists, and the limit is the same for all $(\tilde{A}, \tilde{A}_0)$ with $\tilde{A} \leq 0$.

If (A, A_0) satisfies the conditions of the proposition, then the original and modified dynamic programming operators are related by $\mathcal{U}_\infty = \tilde{\mathcal{U}}$ and

$$\mathcal{S}_\infty(A) = \tilde{\mathcal{S}}(A) + \begin{bmatrix} 0 & 0 \\ 0 & \frac{1}{2\gamma} B_f^\top \Sigma_u^{-1} B_f \end{bmatrix}.$$

With $\theta = \infty$, $\mathcal{S}_\infty(A)_{xx}$ and $\mathcal{S}_\infty(A)_{xf}$ depend only on A_{xx} and A_{xf}, and $\mathcal{S}_\infty(A)_{ff}$ depends linearly on A_{ff}, so

$$\begin{aligned}\mathcal{S}_\infty^2(A) &= \tilde{\mathcal{S}}\left(\tilde{\mathcal{S}}(A) + \begin{bmatrix} 0 & 0 \\ 0 & \frac{1}{2\gamma} B_f^\top \Sigma_u^{-1} B_f \end{bmatrix}\right) + \begin{bmatrix} 0 & 0 \\ 0 & \frac{1}{2\gamma} B_f^\top \Sigma_u^{-1} B_f \end{bmatrix} \\ &= \tilde{\mathcal{S}}^2(A) + \beta \begin{bmatrix} 0 & 0 \\ 0 & \frac{1}{2\gamma} C_f^\top B_f^\top \Sigma_u^{-1} B_f C_f \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & \frac{1}{2\gamma} B_f^\top \Sigma_u^{-1} B_f \end{bmatrix}.\end{aligned}$$

Applying this step repeatedly, we get

$$\mathcal{S}_\infty^n(A) = \tilde{\mathcal{S}}^n(A) + \sum_{t=0}^{n-1} \beta^t \begin{bmatrix} 0 & 0 \\ 0 & \frac{1}{2\gamma} (C_f^\top)^t B_f^\top \Sigma_u^{-1} B_f C_f^t \end{bmatrix}. \tag{A.2}$$

With $\sigma(C_f) < 1 < \beta^{-1/2}$, the sum converges, so $\lim_{n\to\infty} \mathcal{S}_\infty^n(A)$ exists.

For $\mathcal{U}_\infty$, the expression $\mathcal{U}_\infty(A, A_0) = \beta A_0 + \beta tr(\Sigma_v A_{ff})$ yields

$$\mathcal{U}_\infty^n(A, A_0) = \beta^n A_0 + \sum_{i=1}^{n} \beta^i trace(\Sigma_v \mathcal{S}^i(A)_{ff}).$$

Thus, $\lim_{n\to\infty} \tilde{\mathcal{U}}$ exists. The remaining assertions in the proposition follow from corresponding statements for the modified problem. □

A.7 Proof of Proposition 3.5.1

Proof. Proof By collecting the one-period reward and cost in a function G, we can write

$$U(M_0, x_{-1}, f_0) = \sup_{\Delta x} \inf_{m} E\left[\sum_{t=0}^{\infty} \beta^t G(M_t m_{t+1}, x_t, \Delta x_t, f_t)\right].$$

Writing Δx_s and m_s for the policies applied at time s, we have

$$\begin{aligned}
&U(M_0, x_{-1}, f_0) \\
&\leq \sup_{\Delta x_0} \inf_{m_1} \sup_{\{\Delta x_t, t\geq 1\}} \inf_{\{m_t, t\geq 2\}} E\left[\sum_{t=0}^{\infty} \beta^t G(M_t m_{t+1}, x_t, f_t)\right] \\
&= \sup_{\Delta x_0} \inf_{m_1} \left(E[G(M_0 m_1, x_1, \Delta x_1, f_1)] + \sup_{\{\Delta x_t, t\geq 1\}} \inf_{\{m_t, t\geq 2\}} E\left[\sum_{t=1}^{\infty} \beta^t G(M_t m_{t+1}, x_t, \Delta x_t, f_t)\right]\right) \\
&= \sup_{\Delta x_0} \inf_{m_1} \left(E[G(M_0 m_1, x_1, \Delta x_1, f_1)] + \beta E[U(M_1, x_0, f_1)]\right);
\end{aligned}$$

i.e., $U \leq \mathcal{T}(U)$. Now fix a policy π for the investor and let

$$U^{\pi}(M_0, x_{-1}, f_0) = \inf_{\{m_t, t\geq 1\}} E\left[\sum_{t=0}^{\infty} \beta^t G(M_t m_{t+1}, x_t^{\pi}, \Delta x_t^{\pi}, f_t)\right]$$
$$= \inf_{m_1}\left(E[G(M_0 m_1, x_0^{\pi}, \Delta x_0^{\pi}, f_0)] + \inf_{\{m_t, t\geq 2\}} E\left[\sum_{t=1}^{\infty} \beta^t G(M_t m_{t+1}, x_t^{\pi}, \Delta x_t^{\pi}, f_t)\right]\right),$$

the notation x_t^{π} indicating the portfolio process under policy π. For any $\epsilon > 0$, the policy π can be chosen so that

$$\inf_{\{m_t, t\geq 2\}} E\left[\sum_{t=1}^{\infty} \beta^t G(M_t m_{t+1}, x_t^{\pi}, \Delta x_t^{\pi}, f_t)\right] \geq \beta E[U(M_0 m_1, x_0^{\pi}, f_1)] - \epsilon,$$

and since $U \geq U^{\pi}$, we get

$$U(M_0, x_{-1}, f_0) \geq \inf_{m_1}\left(E[G(M_0 m_1, x_0^{\pi}, \Delta x_0^{\pi}, f_0)] + \beta E[U(M_0 m_1, x_0^{\pi}, f_0)] - \epsilon\right),$$

and since this holds for any x_0^{π},

$$U(M_0, x_{-1}, f_0) \geq \sup_{\Delta x_0} \inf_{m_1}\left(E[G(M_0 m_1, x_0, \Delta x_0, f_0)] + \beta E[U(M_0 m_1, x_0, f_1)] - \epsilon\right).$$

We conclude that $U \geq \mathcal{T}(U)$ because $\epsilon > 0$ is arbitrary. □

A.8 Proof of Lemma 3.5.6

The key property of the controls provided by the solution to the Bellman equation is that they lead to linear dynamics for the state evolution. Using the form of the controls in Lemma 3.5.4, we can write the state evolution under the resulting change of measure as

$$y_{t+1} = \Psi y_t + \begin{bmatrix} 0 \\ \\ \mu_{v,t+1} \end{bmatrix} + \begin{bmatrix} 0 \\ \\ \tilde{v}_{t+1} \end{bmatrix}, \tag{A.1}$$

with $\tilde{v}_{t+1} \sim N(0, \tilde{\Sigma}_v)$ and

$$\Psi = \begin{bmatrix} I + 2\Lambda^{-1}A_{xx} & \Lambda^{-1}A_{xf} \\ \\ 0 & C_f \end{bmatrix}.$$

Because $\mu_{v,t+1}$ is itself a linear function of x_t and f_t, we can also write this as

$$y_{t+1} = \bar{\Psi} y_t + \begin{bmatrix} 0 \\ \\ \tilde{v}_{t+1} \end{bmatrix}, \tag{A.2}$$

with

$$\bar{\Psi} = \begin{bmatrix} I + 2\Lambda^{-1}A_{xx} & \Lambda^{-1}A_{xf} \\ \\ -\frac{1}{\theta}\tilde{\Sigma}_v A_{xf}^\top (I + 2\Lambda^{-1}A_{xx}) & -\frac{1}{\theta}\tilde{\Sigma}_v A_{xf}^\top \Lambda^{-1} A_{xf} + (I - \frac{2}{\theta}\tilde{\Sigma}_v A_{ff}) C_f \end{bmatrix}.$$

Before proving Lemma 3.5.6, we show that in this setting β-stability reduces to a condition on $\bar{\Psi}$.

Lemma A.8.1. *Under the conditions of Lemma 3.5.6, β-stability holds if and only if $\sigma(\bar{\Psi}) < \beta^{-1/2}$.*

Proof. Proof (*Lemma A.8.1.*) Suppose first that $\sigma(\bar{\Psi}) < \beta^{-1/2}$. From (A.2) we have

$$y_t = \bar{\Psi}^t y_0 + \sum_{i=0}^{t-1} \bar{\Psi}^i \begin{bmatrix} 0 \\ \\ \tilde{v}_{t-i} \end{bmatrix},$$

where the $\tilde{v}_t$ are i.i.d. $N(0, \tilde{\Sigma}_v)$ random vectors under the adversary's change of measure. Thus, $\tilde{E}[y_t] = \bar{\Psi}^t y_0$. The matrix norm induced by the usual Euclidean norm has the property that $\|\bar{\Psi}^n\|^{1/n} \to \sigma(\bar{\Psi})$ (see, e.g., (7.10.12) in Meyer (2001)). Thus, there exists a $c_1 > 0$ and an $\alpha_1 \in (0, 1)$ such that $\beta^{t/2}\|\bar{\Psi}^t\| < c_1\alpha_1^t$, and then

$$\beta^{t/2}\|\bar{\Psi}^t y_0\| \leq \beta^{t/2}\|\bar{\Psi}^t\|\|y_0\| \leq c_1\alpha_1^t\|y_0\|.$$

In particular, $\beta^{t/2}\tilde{E}[y_0]$ converges to zero exponentially fast.

Next we consider the covariance matrix of y_t,

$$Cov(y_t) = \sum_{i=0}^{t-1} (\bar{\Psi}^i)^\top \tilde{\Sigma}_v \bar{\Psi}^i$$

We may choose $\alpha_1 > \beta^{1/2}$ and set $\bar{\beta} = \beta/\alpha_1^2 < 1$, so $\bar{\beta}^{t/2}\|\bar{\Psi}^t\| < c_1$. Then $\bar{\beta}^{t/2}\|(\bar{\Psi}^\top)^t\| < c_2$, for some $c_2 > 0$ — the norm of a matrix and the norm of its transpose are equivalent (in the sense of norms) because both are equivalent to the max norm defined as the as maximum of the absolute values of all entries of the matrix. This allows us to bound the norm of the covariance matrix of

$\beta^{t/2} y_t$ as follows:

$$\begin{aligned}
\beta^t \| \sum_{i=0}^{t-1} (\bar{\Psi}^i)^\top \tilde{\Sigma}_v \bar{\Psi}^i \| &\leq \beta^t \sum_{i=0}^{t-1} \|(\bar{\Psi}^i)^\top\| \|\bar{\Psi}^i\| \|\tilde{\Sigma}_v\| \\
&\leq \alpha_1^{2t} \|\tilde{\Sigma}_v\| \sum_{i=0}^{t-1} \bar{\beta}^i \|\bar{\Psi}^i\| \|(\bar{\Psi}^i)^\top\| \\
&\leq c_1 c_2 \alpha_1^{2t} \|\tilde{\Sigma}_v\| (t-1).
\end{aligned}$$

Thus, the covariance of $\beta^{t/2} y_t$ also converges to zero, from which we conclude that $\beta^{t/2} y_t$ converges to zero in mean square under the change of measure. Furthermore, $u_{t+1} \sim N(\mu_{u,t+1}, \Sigma_u)$ under the adversary's change of measure, with $\mu_{u,t+1}$ linear in y_t with constant coefficients. It follows that $\beta^{t/2} u_{t+1}$ and thus $\beta^{t/2} y_t^e$ converges to zero in mean square. Retracing the steps above, we see that we can replace β with some $\alpha \in (\beta, 1)$ and preserve this convergence because $\sigma(\bar{\Psi}) < \alpha^{-1/2}$ for some such α. Thus, $\alpha^{t/2} \tilde{E}[\|y_t^e\|] \to 0$ for some $\alpha \in (\beta, 1)$.

Next we evaluate the discounted sum of relative entropy. The summands are given by

$$E_t[m_{t+1} \log m_{t+1}] = \frac{1}{2} \left(\mu_{u,t+1}^\top \Sigma_v^{-1} \mu_{u,t+1} + \mu_{v,t+1}^\top \Sigma_v^{-1} \mu_{v,t+1} + tr(\Sigma_v^{-1} \tilde{\Sigma}_v - I) + \log \frac{|\Sigma_v|}{|\tilde{\Sigma}_v|} \right). \tag{A.3}$$

Since $\tilde{\Sigma}_v$ does not depend on t, it suffices to show that

$$\tilde{E} \left[\sum_{t=0}^{\infty} \beta^t \left(\mu_{v,t+1}^\top \Sigma_v^{-1} \mu_{v,t+1} + \mu_{u,t+1} \Sigma_u^{-1} \mu_{u,t+1} \right) \right] < \infty. \tag{A.4}$$

Because both $\alpha^{t/2} \mu_{u,t+1}$ and $\alpha^{t/2} \mu_{v,t+1}$ are linear in $\alpha^{t/2} y_t$, both converge to zero in mean square, and then because $\beta < \alpha$, (A.4) holds.

Finally, for the converse, if $\beta^{t/2}\tilde{E}[\|y_t^e\|] \to 0$, then $\beta^{t/2}\tilde{E}[y_t] \to 0$, which means $\beta^{t/2}\bar{\Psi}^t y_0 \to 0$. Because this holds for all y_0, we must (by (7.10.5) in Meyer (2001)) have $\sigma(\bar{\Psi}) < \beta^{-1/2}$. □

Proof. Proof (*Lemma 3.5.6.*) We know from the proof of Lemma A.8.1 that β-stability implies $\mathbf{R}_\alpha(m) < \infty$ for some $\alpha \in (\beta, 1)$, so it remains to establish the converse. If $\mathbf{R}_\alpha(m) < \infty$, then, in particular,

$$\tilde{E}\left[\sum_{t=0}^{\infty} \alpha^t \mu_{v,t+1}^\top \Sigma_v^{-1} \mu_{v,t+1}\right] < \infty.$$

Each $\mu_{v,t+1}$ is normally distributed, say $\mu_{v,t+1} \sim N(\mu_{z,t+1}, \Sigma_{z,t+1})$, so we may write

$$\tilde{E}[\alpha^t \mu_{v,t+1}^\top \Sigma_v^{-1} \mu_{v,t+1}] = \alpha^t \left(\mu_{z,t+1}^\top \Sigma_v^{-1} \mu_{z,t+1} + \alpha^t \tilde{E}[(\mu_{v,t+1} - \mu_{z,t+1})^\top \Sigma_v^{-1} (\mu_{v,t+1} - \mu_{z,t+1})]\right).$$

Both terms on the right side are non-negative, hence convergence of the series entails $\alpha^{t/2}\mu_{z,t+1} \to 0$. We can choose $c' \in (0, \infty)$ such that $\alpha^{t/2}\|\mu_{z,t+1}\| \leq c'$. Taking the expectation of both sides of (A.1) and (A.2) we get

$$\bar{\Psi}^t y_0 = \tilde{E}[y_t] = \Psi^t y_0 + \sum_{i=0}^{t-1} \Psi^i \begin{bmatrix} 0 \\ \\ \mu_{z,t-i} \end{bmatrix}.$$

Thus, choosing $\|y_0\| = 1$,

$$\alpha^{t/2}\|\bar{\Psi}^t\| \leq \alpha^{t/2}\|\Psi^t\| + \alpha^{t/2} \sum_{i=0}^{t-1} \|\Psi^i\| \|\mu_{z,t-i}\|. \tag{A.5}$$

Because of the upper triangular structure of Ψ, Proposition 3.3.7, and the condition that $\sigma(C_f) < 1$,

we have that $\sigma(\Psi) = \alpha_2 \in (0,1)$, and we can find a constant $c \in (0,\infty)$ such that for all $t > 0$, $\|\Psi^t\| \leq c\alpha_2^t$.

Thus,

$$\begin{aligned}\alpha^{t/2}\sum_{i=0}^{t-1}\|\Psi^i\|\|\mu_{z,t-i}\| &\leq c'\sum_{i=0}^{t-1}\|\Psi^i\|\alpha^{(1+i)/2} \\ &\leq cc'\sum_{i=0}^{t-1}\alpha_2^i\alpha^{(1+i)/2} \leq cc't\alpha^{1/2}\frac{1-(\alpha_2\alpha^{1/2})^{t-1}}{1-\alpha_2\alpha^{1/2}},\end{aligned} \tag{A.6}$$

which is bounded by some constant for all $t > 0$. We can then choose a slightly smaller $\bar{\alpha} \in (\beta,\alpha)$ for which $\bar{\alpha}^{t/2}\sum_{i=0}^{t-1}\|\Psi^i\|\|\mu_{z,t-i}\| \to 0$ geometrically fast. We conclude from (A.5) that $\bar{\alpha}^{t/2}\|\bar{\Psi}\| \to 0$ geometrically fast and so by (7.10.5) in Meyer (2001), $\sigma(\bar{\Psi}) < \bar{\alpha}^{1/2}$. Lemma A.8.1 now yields β-stability. □

A.9 Proof of Theorem 3.5.7

Proof. Proof (i). Let V^* be the quadratic function in the statement of the theorem. Let Δx_t^* and x_t^* denote the transactions and positions under the corresponding policy. Because U^* satisfies the robust Bellman equation, we have

$$\begin{aligned}U^*(1,x_{-1},f_0) &= V^*(x_{-1},f_0) \\ &= Q(x_0^*,\Delta x_0^*,f_0) + \theta\beta\tilde{E}[\log m_1^* + u_1^\top x_0^*] + \beta\tilde{E}[V^*(x_0^*,f_1)] \\ &= \tilde{E}\left[\sum_{t=0}^{n}\beta^t\left(Q(x_t^*,\Delta x_t^*,f_t) + \theta\beta\tilde{E}[\log m_{t+1}^* + u_{t+1}^\top x_t^*]\right) + \beta^{n+1}V^*(x_n^*,f_{n+1})\right],\end{aligned}$$

for any finite n. The β-stability property allows us to write this as

$$V^*(x_{-1}, f_0) = \tilde{E}\left[\sum_{t=0}^{\infty} \beta^t \left(Q(x_t^*, \Delta x_t^*, f_t) + \theta\beta \tilde{E}[\log m_{t+1}^* + u_{t+1}^\top x_t^*]\right)\right.$$
$$\left. - \sum_{t=n+1}^{\infty} \beta^t \left(Q(x_t^*, \Delta x_t^*, f_t) + \theta\beta(\tilde{E}[\log m_{t+1}^* + u_{t+1}^\top x_t^*])\right) + \beta^{n+1} V^*(x_n^*, f_{n+1})\right].$$

By β-stability, the infinite sum of terms $\beta^{t+1}\tilde{E}[\log m_{t+1}^*]$ is finite, so the tail sum from $t = n+1$ vanishes as $n \to \infty$. From the change of measure in Lemma 3.5.4, we see that $\tilde{E}_t[u_{t+1}^\top x_t^*]$ is a quadratic function of the extended state y_t^e, as are V^* and Q. Again by β-stability, the tail sum of discounted quadratic terms vanishes as $n \to \infty$, and $\tilde{E}[\beta^{n+1} V^*(x_n^*, f_{n+1})] \to 0$ as well. Thus, letting $n \to \infty$ we see that U^* is indeed the value attained under the policy pair $(\Delta x^*, m^*)$.

(ii) Now suppose the adversary has chosen policy m^*. By selecting policy Δx^*, the investor can attain the value U^*. Consider an alternative policy indicated by trades $\Delta\bar{x}_t$ and positions $\bar{x}_t$. Because U^* satisfies the robust Bellman equation, we have

$$U^*(1, x_{-1}, f_0) \geq Q(\bar{x}_0, \Delta\bar{x}_0, f_0) + \theta\beta\tilde{E}[\log m_1^* + u_1^\top \bar{x}_0] + \beta\tilde{E}[U^*(1, \bar{x}_0, f_1)].$$

The expression on the right is the value to the investor of applying $\Delta\bar{x}_0$ and subsequently following the policy Δx^* while the adversary consistently follows m^*. By applying this inequality iteratively, we can compare U^* with the value attained by applying $\Delta\bar{x}_0, \ldots, \Delta\bar{x}_n$ and subsequently following

policy Δx^*:

$$U^*(1, x_{-1}, f_0) \geq \tilde{E}\left[\sum_{t=0}^{n} \beta^t \left\{Q(\bar{x}_t, \Delta\bar{x}_t, f_t) + \theta\beta\tilde{E}[\log m^*_{t+1} + u^\top_{t+1}\bar{x}_t]\right\}\right] + \beta^{n+1}\tilde{E}[U^*(1, \bar{x}_n, f_{n+1})].$$

If $(\Delta\bar{x}, m^*)$ is β-stable, then the value attained under this pair of policies is well-defined, finite, and given by

$$\bar{U}(1, x_{-1}, f_0) = \tilde{E}\left[\sum_{t=0}^{\infty} \beta^t \left\{Q(\bar{x}_t, \Delta\bar{x}_t, f_t) + \theta\beta\tilde{E}[\log m^*_{t+1} + u^\top_{t+1}\bar{x}_t]\right\}\right].$$

We thus have

$$\begin{aligned} &U^*(1, x_{-1}, f_0) - \bar{U}(1, x_{-1}, f_0) \\ &\quad \geq \quad \beta^{n+1}\tilde{E}[U^*(1, \bar{x}_n, f_{n+1})] - \tilde{E}\left[\sum_{t=n+1}^{\infty} \beta^t \left\{Q(\bar{x}_t, \Delta\bar{x}_t, f_t) + \theta\beta\tilde{E}[\log m^*_{t+1} + u^\top_{t+1}\bar{x}_t]\right\}\right]. \end{aligned}$$

The last term on the right vanishes as $n \to \infty$ because $\bar{U}(1, x_{-1}, f_0) < \infty$ under the assumed β-stability of $(\Delta\bar{x}, m^*)$. For the first term on the right, $U^*(1, \bar{x}_n, f_{n+1}) = V^*(\bar{x}_n, f_{n+1})$ is quadratic, so, again, β-stability of $(\bar{x}, m^*)$ ensures that this term vanishes as $n \to 0$.

Using the saddle point condition, i.e. Conditions 3.3.3 and 3.3.4, we can interchange the order of minimum and maximum in the robust Bellman's equation. Then a similar argument applies for the adversary's response to the investor's choice of policy. □

Appendix B

Additional Proofs for Chapter 4

B.1 Asymptotic Error via Strong Approximation

In this section, we develop tools for the strong approximation of jump-diffusion models which we will need to prove our results for that case.

If X solves $dX_t = \tilde{a}(X_t)dt + \tilde{b}(X_t)dW_t + \tilde{c}(X_t)dJ_t$, and $\|X_N - X\|_2 = O(\Delta t^k)$, then we call X_N a *strong approximation* of order k. In the absence of jumps, Kloeden and Platen (1992) show the same order then applies to almost sure convergence. Bruti-Liberati and Platen (2005) and Bruti-Liberati and Platen (2007) treat strong approximation for the jump-diffusion case. In following their approach it is convenient to think of dt as having order 1, and dW and dJ as each having order $1/2$, in terms of their L_2-norm. Approximations of order k then involve keeping all terms of order k or lower.

We use the following representations of the continuous and discrete portfolios. We set

$$V(1) = \prod_{n=1}^{N} \frac{V(n\Delta t)}{V((n+1)\Delta t)} = \prod_{n=1}^{N} R_{n,N},$$

and

$$\hat{V}(1) = \prod_{n=1}^{N} \frac{\hat{V}(n\Delta t)}{\hat{V}((n+1)\Delta t)} = \prod_{n=1}^{N} \hat{R}_{n,N},$$

where

$$\begin{aligned}\hat{R}_{n,N} &:= \frac{\hat{V}(n\Delta t)}{\hat{V}((n-1)\Delta t)} \\ &= \sum_{i=1}^{d} w_i \exp\{(\mu_i - \frac{1}{2}\sum_{j=1}^{d} \sigma_{ij}^2)\Delta t + \sigma_i^\top \Delta W_n\} \prod_{j=N((n-1)\Delta t)+1}^{N(n\Delta t)} Y_j^i\end{aligned}$$

and

$$\begin{aligned}R_{n,N} &:= \frac{V(n\Delta t)}{V((n-1)\Delta t)} \\ &= \exp\{(\mu_w - \frac{1}{2}\sigma_w^2)\Delta t + \bar{\sigma}^\top \Delta W_n\} \prod_{j=N((n-1)\Delta t)+1}^{N(n\Delta t)} (\sum_{i=1}^{d} w_i Y_j^i).\end{aligned}$$

Then

$$\frac{\hat{R}_{n,N}}{R_{n,N}} = \sum_{i=1}^{d} w_i \exp\{(\mu_i - \mu_w - \frac{1}{2}\|\sigma_i\|^2 + \frac{1}{2}\sigma_w^2)\Delta t + (\sigma_i - \bar{\sigma})^\top \Delta W_n\}$$
$$\times \prod_{j=N((n-1)\Delta t)+1}^{N(n\Delta t)} \frac{Y_j^i}{\sum_{i=1}^d w_i Y_j^i}$$
$$= \frac{\hat{R}^c_{n,N}}{R^c_{n,N}} \frac{\sum_{i=1}^d w_i \exp\{(\mu_i - \mu_w - \frac{1}{2}\|\sigma_i\|^2 + \frac{1}{2}\sigma_w^2)\Delta t + (\sigma_i - \bar{\sigma})^\top \Delta W_n\} \prod_j Y_j^i}{\sum_i w_i \exp\{(\mu_i - \mu_w - \frac{1}{2}\|\sigma_i\|^2 + \frac{1}{2}\sigma_w^2)\Delta t + (\sigma_i - \bar{\sigma})^\top \Delta W_n\} \prod_j \sum_i w_i Y_j^i},$$

where $\hat{R}^c_{n,N}/R^c_{n,N}$ is the ratio of returns in the absence of jumps, as in Glasserman (2012),

$$\frac{\hat{R}^c_{n,N}}{R^c_{n,N}} = \sum_{i=1}^{d} w_i \exp\{(\mu_i - \mu_w - \frac{1}{2}\|\sigma_i\|^2 + \frac{1}{2}\sigma_w^2)\Delta t + (\sigma_i - \bar{\sigma})^\top \Delta W_n\}.$$

B.2 Background on Strong Approximations

As in Kloeden and Platen (1992) and Platen (1982), we use the following notation. For a string $\omega = (i_1, \ldots, i_{k-1}, i_k)$ of indices, let $\omega- := (i_1, \ldots, i_{k-1})$ and $-\omega := (i_2, \ldots, i_k)$, for $k > 0$. The length of the string is given by $l(\omega) = k$, and $n(\omega)$ denotes the number of zeros in the string ω. Define the hierarchical sets $\mathcal{A}_l = \{\omega | l(\omega) + n(\omega) \leq 2l\}$, and the corresponding remainder sets $\mathcal{B}(\mathcal{A}_l) = \{\omega \notin \mathcal{A}_l, -\omega \in \mathcal{A}_l\}$, for $l = \frac{1}{2}, 1, \frac{3}{2}, 2, \ldots$.

For a predictable g satisfying certain regularity and integrability conditions in the main theorem

of Platen (1982), an iterated integral I_ω is defined as follows:

$$I_\omega[g]_t = \begin{cases} g(t) & \text{if } l(\omega) = 0; \\ \int_0^t I_{\omega-}[g]_z dz & \text{if } i_{l(\omega)} = 0 \text{ and } l(\omega) > 0; \\ \int_0^t I_{\omega-}[g]_z dW_z^i & \text{if } i_{l(\omega)} = i > 0 \text{ and } l(\omega) > 0; \\ \int_0^t I_{\omega-}[g]_z d\tilde{J}_z^i & \text{if } i_{l(\omega)} = i < 0 \text{ and } l(\omega) > 0. \end{cases}$$

To have a better understanding of the notation, one can interpret the string $\omega = (i_1, ..., i_k)$ as the order for iterated integration, with the direction from left to right corresponding to the order of integration from innermost to outermost integral. Each entry i_k indicates the process against which the integral is taken. For example, $i_k > 0$ indicates an integral against the i_k^{th} component of the Brownian Motion, while $i_k < 0$ indicates an integral against $\tilde{J}^{i_k}$.

The main result of Platen (1982) shows that under our particular setting where all coefficient functions are linear, we have the Ito-Taylor expansion

$$f(t, X_t) = \sum_{\omega \in \mathcal{A}_l} I_\omega[f_\omega(0, X_0)]_t + \sum_{\omega \in \mathcal{B}(\mathcal{A}_l)} I_\omega[f_\omega(\cdot\,, X_\cdot)]_t.$$

Here we choose $f(x) = x$ and coefficients are defined by

$$f_\omega(t,x) = \begin{cases} x & \text{if } l(\omega) = 0; \\ \tilde{a}(x) & \text{if } l(\omega) = 1, i_1 = 0\,; \\ \tilde{b}_{i_1}(x) & \text{if } l(\omega) = 1, i_1 > 0\,; \\ \tilde{c}(x) & \text{if } l(\omega) = 1, i_1 < 0\,; \\ L^{i_1} f_{-\omega} & \text{if } l(\omega) > 1; \end{cases}$$

where

$$L^i f(t,x) = \begin{cases} \frac{\partial f}{\partial t} + \tilde{a}\frac{\partial f}{\partial x} + \frac{1}{2}\sum_j \tilde{b}_j^2 \frac{\partial^2 f}{\partial x} & \text{if } i = 0; \\ \tilde{b}_i \frac{\partial f}{\partial x} & \text{if } i > 0; \\ f(t, x + \tilde{c}(x)) - f(t,x) & \text{if } i < 0. \end{cases}$$

A more detailed treatment of strong approximations and this notation can be found in Platen (1982).

For our application, we need to approximate $\sum w_i X_i(\Delta t) := \hat{R}_{n,N}/R_{n,N}$, where

$$X_{i,N}(t) = \exp\{(\mu_i - \mu_w - \frac{1}{2}\|\sigma_i\|^2 + \frac{1}{2}\sigma_w^2)t + (\sigma_i - \bar{\sigma})^\top W(t)\} \prod_{j=1}^{N(t)} \frac{Y_j^i}{\sum w_k Y_j^k}.$$

Each $X_{i,N}$ satisfies the following SDE:

$$\begin{aligned} \frac{dX_{i,N}(t)}{X_{i,N}(t-)} &= (\mu_i - \mu_w - \frac{1}{2}\|\sigma_i\|^2 + \frac{1}{2}\sigma_w^2 + \frac{1}{2}\|\sigma_i - \bar{\sigma}\|^2)dt + (\sigma_i - \bar{\sigma})^\top dW_t + dJ_t^i \\ &= a_i dt + b_i^\top dW_t + d\tilde{J}_t^i, \end{aligned}$$

where $a_i = \mu_i - \mu_w - \frac{1}{2}\|\sigma_i\|^2 + \frac{1}{2}\sigma_w^2 + \frac{1}{2}\|\sigma_i - \bar{\sigma}\|^2 + \lambda\mu_i^y$ and $b_i = \sigma_i - \bar{\sigma}$.

For our analysis, we need some standard properties of predictable quadratic variations: $< t, t >= 0$, $< t, W_t^i >= 0$ and $< t, \tilde{J}_t^j >= 0$ for all i and j; $< W^i, W^j >_t = \delta_{ij}t$, and $< \tilde{J}^j, \tilde{J}^i >_t = m_{ij}t$, for constants m_{ij}. To derive the appropriate constants, we observe that

$$\begin{aligned}
E[\tilde{J}_t^i \tilde{J}_t^j] &= E[[\tilde{J}^i, \tilde{J}^j]_t] \\
&= \frac{1}{4}E[[\tilde{J}^i + \tilde{J}^j, \tilde{J}^i + \tilde{J}^j]_t - [\tilde{J}^i - \tilde{J}^j, \tilde{J}^i - \tilde{J}^j]_t] \\
&= \frac{1}{4}E[\sum_{0<s<t}(\tilde{J}_s^i - \tilde{J}_{s-}^i + \tilde{J}_s^j - \tilde{J}_{s-}^j)^2 - \sum_{0<s<t}(\tilde{J}_s^i - \tilde{J}_{s-}^i - \tilde{J}_s^j + \tilde{J}_{s-}^j)^2] \\
&= E[\sum_{0<s<t}((\tilde{J}_s^i - \tilde{J}_{s-}^i)(\tilde{J}_s^j - \tilde{J}_{s-}^j))] \\
&= \lambda t E[\bar{Y}^i \bar{Y}^j].
\end{aligned}$$

The third equality is due to the fact that a compound Poisson process $\sum_{i=1}^{N(t)} Z_i$ has quadratic variation $\sum_{i=1}^{N(t)} Z_i^2$ ((Cont and Tankov 2003, Example 8.4)). Thus, we need $m_{ij} = \lambda E[\bar{Y}^i \bar{Y}^j]$.

B.3 Strong Approximation for the Jump-Diffusion model

We now use the strong approximation scheme of order 3/2 to prove Theorem 4.2.3 and 4.2.4. First we write

$$X_{i,N}(\Delta t) = 1 + \zeta_{1/2,N}^i + \zeta_{1,N}^i + \zeta_{3/2,N}^i + r_N^i,$$

where $\zeta^i_{\cdot,N}$ are defined as follows. First,

$$\zeta^i_{1/2,N} = \int_0^{\Delta t} b_i^\top dW + \int_0^{\Delta t} d\tilde{J}^i = b_i^\top \Delta W + \Delta \tilde{J}^i.$$

(From now on we drop the limits of integration for iterated integrals taken over $[0, \Delta t]$. An integral of the form $\int g\, d\tilde{J}^i$ should be understood as $\int g(t-) d\tilde{J}^i(t)$.) Continuing, we have

$$\begin{aligned}\zeta^i_{1,N} = a^i \int dt + \int\int b_i^\top dW b_i^\top dW + \int\int b_i^\top dW d\tilde{J}^i \\ + \int\int d\tilde{J}^i b_i^\top dW + \int\int d\tilde{J}^i d\tilde{J}^i\end{aligned} \tag{B.1}$$

and

$$\begin{aligned}\zeta^i_{3/2,N} = a^i \int\int b_i^\top dW dt + a^i \int\int dt b_i^\top dW + a^i \int\int d\tilde{J}^i dt + a^i \int\int dt d\tilde{J}^i \\ + \int\int\int b_i^\top dW b_i^\top dW b_i^\top dW + \int\int\int b_i^\top dW b_i^\top dW d\tilde{J}^i \\ + \int\int\int b_i^\top dW d\tilde{J}^i b_i^\top dW + \int\int\int d\tilde{J}^i b_i^\top dW b_i^\top dW + \int\int\int b_i^\top dW d\tilde{J}^i d\tilde{J}^i \\ + \int\int\int d\tilde{J}^i d\tilde{J}^i b_i^\top dW + \int\int\int d\tilde{J}^i b_i^\top dW d\tilde{J}^i + \int\int\int d\tilde{J}^i d\tilde{J}^i d\tilde{J}^i.\end{aligned} \tag{B.2}$$

By observing that $\sum w_i b_i = 0$ and $\sum w_i \tilde{J}^i = 0$, we find that $\sum w_i \zeta^i_{1/2,N} = 0$. For the next

term, we have

$$\sum_i w_i \zeta^i_{1,N} = \sum_i w_i [\epsilon_{n,N} + b_i^\top \Delta W \Delta \tilde{J}^i + \int\int d\tilde{J}^i d\tilde{J}^i].$$

Here, $\epsilon_{n,N}$ is the corresponding error term in the absence of jumps; the last two terms are the difference between the continuous and jump-diffusion cases.

It is now easy to see that $\| \sum w_i \zeta^i_{1,N} \| = O(\Delta t)$, and similarly $\| \sum w_i \zeta^i_{3/2,N} \| = O(\Delta t^{3/2})$. Now we need to show that the remainder r^i_N satisfies $\|r^i_N\| = O(\Delta t^2)$.

Lemma B.3.1. *(Modified from (Studer 2001, Lemma 3.42).) Given an adapted caglad (left continuous with right limits) process g(t), with $\int_0^t E[g(s)^2]ds = K < \infty$, then*

$$E[(\int_0^t g(s) dM_s)^2] \leq \begin{cases} tK, & \text{if } M_t = t; \\ K, & \text{if } M_t = W^i_t; \\ m_{ii}K, & \text{if } M_t = \tilde{J}^i_t. \end{cases}$$

(The integrand should be understood as $g^\top$ when $M = W$.)

Proof. The result and proof are the same as in Studer (2001). □

To bound the error when we truncate a strong approximation, we can apply a result of (Studer 2001, Proposition 3.43), or s similar result of (Bruti-Liberati and Platen 2005, Theorem 6.1). Out setting is simpler than theirs because of the special form of the dynamics in the JD model.

Lemma B.3.2. *(Modified from (Studer 2001, Proposition 3.43).) Under our assumptions (4.2.3)*

and (4.2.4) for the JD model, there exist some constants C_1 and C_2 such that for any $i = 1, ..., d$

$$E[(X_{i,N}(t) - \sum_{\kappa \in \mathcal{A}_k} I_\kappa[f_\kappa(0, X_{i,N}(0))])^2] \leq C_1 (C_2 t)^{2k+1}.$$

Proof. Since $f(t, x) = x$, the conditions in (Studer 2001, Proposition 3.43) (and those in (Bruti-Liberati and Platen 2005, Theorem 6.1)) are satisfied. Thus, for any $\omega \in \mathcal{B}(\mathcal{A}_k)$, we can find some constant C_3

$$\sup_{0 \leq t \leq T} E[(f_\omega(t, X_{i,N}(t)))^2] \leq C_3.$$

Denote $n_i(\omega)$ be the number of components for $\tilde{J}^i$ in ω. By induction and the previous lemma, we have for any $\omega \in \mathcal{B}(\mathcal{A}_k)$, we can find some constant C_4

$$\begin{aligned} E[I_\omega[f_\omega(., X_{i,N}(.))]_t^2] &\leq t^{n(\omega)} (m_{ii})^{n_i(\omega)} C_3 t^{l(\omega)} \\ &\leq C_3 C_4^{2k+1} t^{l(\omega)+n(\omega)}; \end{aligned}$$

and $|\mathcal{B}(\mathcal{A}_k)| \leq (3d+3)^{k+1}$, therefore,

$$\begin{aligned} E[(X_t - \sum_{\kappa \in \mathcal{A}_k} I_\kappa[f_\kappa(0, X_0)])^2] &\leq (\sum_{\omega \in \mathcal{B}(\mathcal{A}_k)} (E[I_\omega[f_\omega(., X)]])^{1/2})^2 \\ &\leq (\sum_{\omega \in \mathcal{B}(\mathcal{A}_k)} (C_3 C_4^{2k+1} t^{l(\omega)+n(\omega)})^{1/2})^2 \\ &\leq C_1 (C_2 t)^{2k+1}. \end{aligned}$$

□

As a consequence, for our setting we get

Lemma B.3.3. $\|r_N^i\| = \|X_{i,N} - 1 - \zeta_{1/2,N}^i - \zeta_{1,N}^i - \zeta_{3/2,N}^i\| = O(\Delta t^2)$.

B.4 Correlation Between ζ_1^i and $\zeta_{3/2}^i$

In this section, we show that the terms $\sum w_i \zeta_{1,N}^i$ and $\sum w_i \zeta_{3/2,N}^i$ are uncorrelated. Before specializing to our setting, we derive some general properties used extensively in this subsection.

To calculate the covariance between iterated integrals, from (Cont and Tankov 2003, Proposition 8.11) we have (using the notation of Lemma B.3.1)

$$\begin{aligned} E[I_{\omega_1} I_{\omega_2}] &= E[\int I_{\omega_1 -} dM_1 \int I_{\omega_2 -} dM_2] \\ &= E[\int I_{\omega_1 -} I_{\omega_2} dM_1 + \int I_{\omega_2 -} I_{\omega_1} dM_2 + \int I_{\omega_1 -} I_{\omega_2 -} d[M_1, M_2]] \\ &= E[\int I_{\omega_1 -} I_{\omega_2} dM_1 + \int I_{\omega_2 -} I_{\omega_1} dM_2 + \int I_{\omega_1 -} I_{\omega_2 -} d < M_1, M_2 >], \end{aligned} \tag{B.1}$$

where

$$M_i(t) = \begin{cases} t & \text{if } r(\omega_i) = 0; \\ W_t & \text{if } r(\omega_i) = 1; \\ \tilde{J}_t^k & \text{if } r(\omega_i) = k < 0, \end{cases}$$

with $r(\omega_i)$ the rightmost element of ω_i. As before, when $M_i = \tilde{J}^k$ for some i and k, we use the left-continuous version of the integrand. When $M_i = W$, we take its transpose in the integrand. Here we use the square bracket and sharp bracket to denote quadratic variation and predictable quadratic variation as introduced towards the end of Section B.2.

When $r(\omega_i) \neq 0$ for both $i = 1$ and 2, $M_{r(\omega_i)}$ is a martingale, so after taking expectations, the first two terms in (B.1) vanish. Assumption (4.2.4) implies square integrability of these iterated integrals, which contain jump terms. Otherwise, when they consist of only dt or dW, their integrability is immediate. Thus, we have the following possible combinations:

When $r(\omega_1) > 0$ and $r(\omega_2) = -j < 0$, M_1 and M_2 are uncorrelated martingales, so the expectation of their product is 0. Thus, we have:

$$E[\int I_{\omega_1-}^{\top} dW \int I_{\omega_2-} d\tilde{J}^j] = 0. \tag{B.2}$$

When $r(\omega_1) = r(\omega_2) = 1$,

$$E[\int I_{\omega_1-}^{\top} dW \int I_{\omega_2-}^{\top} dW] = \int E[I_{\omega_1-}^{\top} I_{\omega_2-}] dt$$

and when $r(\omega_1) = -i$, $r(\omega_2) = -j$,

$$E[\int I_{\omega_1-} d\tilde{J}^i \int I_{\omega_2-} d\tilde{J}^j] = \int E[I_{\omega_1-} I_{\omega_2-}] m_{ij} dt. \tag{B.3}$$

When $r(\omega_1) = 0$ and $r(\omega_2) \neq 0$,the second and the third term in (B.1) vanish, leaving

$$E[\int I_{\omega_1-} dt \int I_{\omega_2-} dM_2] = \int E[I_{\omega_1-} I_{\omega_2}] dt.$$

Now we apply these results to analyze the correlation between ζ_1 and $\zeta_{3/2}$. Let $\mathcal{B}_{l,n} = \{\gamma | l(\gamma) = l, n(\gamma) = n\}$. All strings in $\mathcal{B}_{l,n}$ are of the same length l and have the same number of zeros n. We

observe from (B.1) and (B.2) that $\zeta^i_{1,N}$ is a linear combination of elements in $\mathcal{B}(\mathcal{A}_1)$ and $\zeta^i_{3/2,N}$ is a linear combination of elements of $\mathcal{B}(\mathcal{A}_{3/2})$. From here until the end of this subsection, we let ω and κ be strings with $l(\omega) = 1$ and $l(\kappa) = 3/2$, and we treat all possible combinations of values of $n(\omega)$ and $n(\kappa)$:

(a) If $n(\omega) = 0$ and $n(\kappa) = 0$ — that is, neither contains dt integrals — then (B.2)–(B.3) show that $E[I_\omega I_\kappa]$ equals to an integral against dt with its integrand either zero or $E[I_{\omega-} I_{\kappa-}]$. Applying the same argument again, so we can say that $E[I_{\omega-} I_{\kappa-}]$ is again an integral against dt with its integrand either zero or $E[I_{\omega--} I_{\kappa--}]$, which is zero, since $l(\omega) = 1$. So $E[I_\omega I_\kappa] = 0$ for any $\omega \in \mathcal{B}_{1,0}$ and $\kappa \in \mathcal{B}_{3/2,0}$. Hence any linear combination of elements of $\{I_\omega : \omega \in \mathcal{B}_{1,0}\}$ and any linear combination of elements of $\{I_\kappa : \omega \in \mathcal{B}_{3/2,0}\}$ are uncorrelated.

(b) If $l(\omega) = n(\omega) = 1$, but $n(\kappa) = 0$, then I_ω is actually deterministic. So $E[I_\omega I_\kappa] = I_\omega E[I_\kappa] = 0$, since I_κ is a martingale. Hence any linear combination of elements of $\{I_\omega : \omega \in \mathcal{B}_{1,1}\}$ and any linear combination of elements of $\{I_\kappa : \omega \in \mathcal{B}_{3/2,0}\}$ are uncorrelated.

(c) For the case $n(\omega) = 0$ and $n(\kappa) = 1$, we observe that in our particular setting, for any $i \neq 0$, $I_{(i,0)}$ and $I_{(0,i)}$ always appear in pairs in ζ^i and have the same coefficients. Using integration by parts we can consider them in pairs, for $i \neq 0$, to get

$$I_{(i,0)} + I_{(0,i)} = \int d(tM_i) = \Delta t \Delta M_i,$$

so

$$E[I_\omega (I_{(i,0)} + I_{(0,i)})] = \Delta t E[I_\omega \Delta M_i] = 0,$$

the last equality following from the same argument as (a). Hence, any linear combination of elements of $\{I_\omega : \omega \in \mathcal{B}_{1,0}\}$ and any linear combination of elements of $\{I_\kappa : \omega \in \mathcal{B}_{3/2,1}\}$ are uncorrelated.

(d) If $n(\omega) = 1$, and $n(\kappa) = 1$, then $I_\omega = \Delta t$, which is deterministic, and $I_{(i,0)} + I_{(0,i)} = \Delta t \Delta M_i$ has zero mean. Hence any linear combination of elements of $\{I_\omega : \omega \in \mathcal{B}_{1,1}\}$ and any linear combination of elements of $\{I_\kappa : \omega \in \mathcal{B}_{3/2,1}\}$ are uncorrelated.

To summarize, we have proved

Lemma B.4.1. *$\sum w_i \zeta^i_{1,N}$ and $\sum w_i \zeta^i_{3/2,N}$ are uncorrelated.*

B.5 Convergence Proofs

Using our analysis of the strong approximation for the jump-diffusion case, we can now prove Theorems 4.2.3 and 4.2.4.

B.5.1 Proof of Theorem 4.2.3

Proof. We have

$$\frac{\hat{R}_{n,N}}{R_{n,N}} = 1 + \sum_i w_i \zeta^i_{1,N} + \sum_i w_i \zeta^i_{3/2,N} + \sum_i w_i r^i_N.$$

We have shown that $\| \sum w_i \zeta^i_{1,N} \| = O(\Delta t)$, $\| \sum w_i \zeta^i_{3/2,N} \| = O(\Delta t^{3/2})$, $\| \sum w_i r^i_N \| = O(\Delta t^2)$, that $E[\sum w_i \zeta^i_{1,N}] = 0$ and $E[\sum w_i \zeta^i_{3/2,N}] = 0$, and that $\sum w_i \zeta^i_{1,N}$ and $\sum w_i \zeta^i_{3/2,N}$ are uncorrelated. We can now follow the argument used in (Glasserman 2012, Proposition 1) to prove (4.2.8).

Next we calculate the variance of the relative error. To condense (4.2.9), let $E_1 = \sum w_i [b_i^\top \Delta W \Delta \tilde{J}^i]$,

and $E_2 = \sum w_i[\int \int d\tilde{J}^i d\tilde{J}^i]$. By following steps similar to those used to prove Lemma B.4.1, we can show that the pairwise correlations between $\epsilon_{n,N}$, E_1, and E_2 are all zero. Thus,

$$Var[\tilde{\epsilon}_{n,N}] = Var[\epsilon_{n,N}] + Var[E_1] + Var[E_2].$$

We need to calculate the last two terms on the right. For E_1, we have

$$\begin{aligned} Var[E_1] = E[E_1^2] &= E[(\sum w_i b_i^\top \Delta W \Delta \tilde{J}^i)^2] \\ &= E[(\Delta W^\top b \Omega \Delta \tilde{J})^2] \\ &= E[(\sum_i \Delta W_i^2) \Delta \tilde{J}^\top (\Omega b^\top b \Omega) \Delta \tilde{J}] \\ &= \Delta t^2 (w^\top (b^\top b \circ M) w). \end{aligned}$$

For E_2, we have

$$\begin{aligned} Var[E_2] = E[E_2^2] &= E[(\sum w_i \int \int d\tilde{J}^i d\tilde{J}^i)^2] \\ &= \sum_{i,j} w_i w_j \int E[< \tilde{J}^i, \tilde{J}^j >_s] m_{ij} ds \\ &= \frac{\Delta t^2}{2} w^\top M \circ M w. \end{aligned}$$

□

B.5.2 Proof of Theorem 4.2.4

Proof. First, from the expression of the asymptotics of the relative error in (4.2.9), the contribution of the compensation terms in the jump terms are of lower order, so we can replace $\tilde{J}_n^i$ and $\tilde{J}^i$ with J_n^i and J^i respectively throughout (4.2.9) and (4.2.8) still holds. That is,

$$E[(\frac{\hat{V}(T) - V(T)}{V(T)} - \sum_{n=1}^{N} \bar{\epsilon}_{n,N})^2] = O(\Delta t^2)$$

$$\text{where } \bar{\epsilon}_{n,N} = \epsilon_{n,N} + \sum_{i=1}^{d} w_i [b_i^\top \Delta W_n \Delta J_n^i + \int_{(n-1)\Delta t}^{n\Delta t} \int_{(n-1)\Delta t}^{s-} dJ^i(r) dJ^i(s)]. \tag{B.1}$$

The last term in (B.1) is nonzero only when there are at least two jumps in the period $[(n-1)\Delta t, n\Delta t]$, which has probability $O(\Delta t^2)$. Since the number of jumps in different periods are i.i.d., the probability that none of the time intervals has more than one jump is of order $1 - O(\Delta t)$, so

$$\sqrt{N} \sum_{n=1}^{N} \sum_{i=1}^{d} w_i \int_{(n-1)\Delta t}^{n\Delta t} \int_{(n-1)\Delta t}^{s-} dJ^i(r) dJ^i(s) \Rightarrow 0.$$

For the same reason, we can ignore multiple jumps in each Δt interval in (B.1). More precisely,

$$\sqrt{N} \sum_{j=1}^{N(T)} \left(\sum_{i=1}^{d} w_i b_i^\top \Delta W_{n(j)} \bar{Y}_{n(j)}^i - \sum_{i=1}^{d} w_i b_i^\top \Delta W_{n(j)} \Delta J_{n(j)}^i \right) \Rightarrow 0, \tag{B.2}$$

where $n(j)$ is the index of the interval when j^{th} jump takes place.

To analyze the limit of (B.2), we rewrite it as

$$\sqrt{N} \sum_{\substack{n \neq n(j) \\ j=1,\ldots,N(T)}} \epsilon_{n,N} + \sqrt{N} \sum_{\substack{n=n(j) \\ j=1,\ldots,N(T)}} \epsilon_{n,N} + \sqrt{N} \sum_{j=1}^{N(T)} \sum_{i=1}^{d} w_i b_i^\top \Delta W_{n(j)} \bar{Y}^i_{n(j)}. \tag{B.3}$$

Let $N \to \infty$, noting that $N(T)$ remains fixed. In (B.3), the first term is independent of the other two terms, and it converges to $\underline{X} \sim N(0, \sigma_L^2 T)$, as shown in Theorem 4.2.1. The second term in (B.3) converges to zero in L_2 and thus in probability. Thus (B.3) converges in distribution to

$$\underline{X} + \sum_{j=1}^{N(T)} \sum_{i=1}^{d} w_i b_i^\top \xi_j \bar{Y}^i_j,$$

where ξ_j are i.i.d. standard normal random variables independent of everything else. The limit does not hold in L_2, since the L_2-norm of the third term in (B.1) has order $O(\Delta t^2)$, as shown in the proof of Theorem 4.2.3.

□

B.6 Strong Approximation for the Mean-Reverting Case

In this section, we prove Theorem 4.2.1. We build on the strong approximation technique introduced in Section B.3, but the argument will be somewhat simpler because we no longer have jump terms.

B.6.1 Proof of Theorem 4.2.1

Proof. The value of the discretely rebalanced portfolio at Δt is given by

$$\underline{\hat{V}}(\Delta t) = \sum_i w_i \exp\{(\mu_i - \frac{1}{2}\|\sigma_i\|^2)\Delta t + \sigma_i \int_0^{\Delta t} e^{-\kappa(\Delta t - s)} dW_s + (1 - e^{-\kappa \Delta t})\mu_i^l\},$$

and the ratio of the discrete portfolio value to the continuous portfolio value is given by

$$\begin{aligned}
\frac{\underline{\hat{R}}_N}{\underline{R}_N} &= \frac{\sum_i w_i \exp\{(\mu_i - \frac{1}{2}\sigma_i^2)\Delta t + \sigma_i^\top \int_0^{\Delta t} e^{\kappa(s-\Delta t)} dW_s + (1 - e^{-\kappa \Delta t})\mu_i^l\}}{\exp\{(\mu_w - \frac{1}{2}\sigma_w^2)\Delta t + \bar{\sigma} \int_0^{\Delta t} e^{\kappa(s-\Delta t)} dW_s + (1 - e^{-\kappa \Delta t})\bar{\mu}^l\}} \\
&= \sum_i w_i \exp\{(\mu_i - \mu_w - \frac{1}{2}(\|\sigma_i\|^2 - \sigma_w^2))\Delta t + (\sigma_i - \bar{\sigma})^\top \int_0^{\Delta t} e^{\kappa(s-\Delta t)} dW_s \\
&+ (1 - e^{-\kappa \Delta t})(\mu_i^l - \bar{\mu}^l)\} \\
&=: \sum_i w_i D_i(\Delta t),
\end{aligned}$$

where each D_i satisfies

$$\begin{aligned}
dD_i &= D_i[(\mu_i - \mu_w - \frac{1}{2}(\|\sigma_i\|^2 - \sigma_w^2 - \|\sigma_i - \bar{\sigma}\|^2))dt + d\bar{L}_i] \\
d\bar{L}_i &= \kappa(\mu_i^l - \bar{\mu}^l - \bar{L}_i)dt + (\sigma_i - \bar{\sigma})^\top dW.
\end{aligned}$$

Using strong approximation as introduced in Section B.3, we get (with all iterated integrals

taken from 0 to Δt):

$$D_i(\Delta t) = 1 + (\mu_i - \mu_w - \frac{1}{2}(\|\sigma_i\|^2 - \sigma_w^2 - \|\sigma_i - \bar{\sigma}\|^2))(\Delta t + \int\int d\bar{L}_i dt + \int\int dt d\bar{L}_i) + \Delta\bar{L}_i + \int\int d\bar{L}_i d\bar{L}_i + \int\int\int d\bar{L}_i d\bar{L}_i d\bar{L}_i + O(\Delta t^2),$$

where

$$\begin{aligned}\Delta\bar{L}_i &= (\sigma_i - \bar{\sigma})^\top e^{-\kappa\Delta t}\int_0^{\Delta t} e^{\kappa s} dW_s + (1 - e^{-\kappa\Delta t})(\mu_i^l - \bar{\mu}^l) \\ &= (\sigma_i - \bar{\sigma})^\top(\Delta W - \kappa\int_0^{\Delta t} W_s ds) + \kappa(\mu_i^l - \bar{\mu}^l)\Delta t + O(\Delta t^2).\end{aligned}$$

Expanding the iterated integrals of $\bar{L}_i$ and substituting, we get

$$\begin{aligned}D_i(\Delta t) = 1 &+ (\sigma_i - \bar{\sigma})^\top\Delta W + [(\mu_i - \mu_w - \frac{1}{2}(\|\sigma_i\|^2 - \sigma_w^2))\Delta t \\ &+ \frac{1}{2}\Delta W^\top\bar{\Sigma}_i\Delta W] + [\frac{1}{6}\Delta W^\top\bar{\Sigma}_i\Delta W(\sigma_i - \bar{\sigma})^\top\Delta W \\ &+ (\mu_i - \mu_w - \frac{1}{2}(\|\sigma_i\|^2 - \sigma_w^2))(\sigma_i - \bar{\sigma})^\top\Delta W\Delta t - (\sigma_i - \bar{\sigma})^\top(\kappa\int_0^{\Delta t} W_s ds)] \\ &+ \kappa(\mu_i^l - \bar{\mu}^l)\Delta t + \kappa(\mu_i^l - \bar{\mu}^l)(\sigma_i - \bar{\sigma})^\top\Delta W\Delta t + O(\Delta t^2),\end{aligned}$$

where $\bar{\Sigma}_i = (\sigma_i - \bar{\sigma})(\sigma_i - \bar{\sigma})^\top$, and we drop the term $\Delta W^\top\bar{\Sigma}_i\int_0^{\Delta t} s dW_s$ because its L_2-norm is $O(\Delta t^2)$.

Now taking the weighted sum of the D_i, we get

$$\begin{aligned}\sum_i w_i D_i(\Delta t) &= 1 + 0 + [-\frac{1}{2}(\sum_i w_i \|\sigma_i\|^2 - \sigma_w^2)\Delta t + \frac{1}{2}\Delta W^\top \bar{\Sigma} \Delta W] \\ &\quad + \sum_i w_i [\frac{1}{6}\Delta W^\top \bar{\Sigma}_i \Delta W (\sigma_i - \bar{\sigma})^\top \Delta W \\ &\quad + (\mu_i - \mu_w - \kappa(\mu_i^l - \bar{\mu}^l) - \frac{1}{2}(\|\sigma_i\|^2 - \sigma_w^2))(\sigma_i - \bar{\sigma})^\top \Delta W \Delta t] + O(\Delta t^2) \\ &\quad =: 1 + \zeta_1^N + \zeta_{3/2}^N + \underline{r}\end{aligned}$$

where $\bar{\Sigma} = \sum_i w_i \bar{\Sigma}_i$ and $\|\underline{r}\| = O(\Delta t^2)$.

Following essentially the same arguments used in the jump-diffusion case, it is now easy to show that $\|\zeta_1^N\| = O(\Delta t)$ and $\|\zeta_{3/2}^N\| = O(\Delta t^{3/2})$, and also that ζ_1^N and $\zeta_{3/2}^N$ are uncorrelated, leading to

$$\|\frac{\hat{\underline{V}}(T)}{\underline{V}(T)} - 1 - \sum_{n=1}^N \zeta_{1,n}^N\| = O(\Delta t).$$

At the same time,

$$\zeta_{1,n}^N = \frac{1}{2}(\Delta W^\top \bar{\Sigma} \Delta W - Tr(\bar{\Sigma})\Delta t) = \epsilon_{n,N},$$

coincides with the $\epsilon_{n,N}$ in the case of multivariate geometric Brownian motion considered in Glasserman (2012). The same limit therefore applies here.

□

Given the representation in Theorem 4.2.1, the proof of Theorem 4.2.2 is the same as that of Theorem 1 in Glasserman (2012).

B.7 Analysis of the Volatility Adjustments

B.7.1 Proof of Proposition 4.3.2: the Jump-Diffusion Case

Proof. With

$$X_N = \sqrt{N}\sum_{n=0}^{N-1}(\frac{\hat{V}((n+1)\Delta t)}{V((n+1)\Delta t)} - \frac{\hat{V}(n\Delta t)}{V(n\Delta t)})$$

we can write $Cov[\log V(T), X_N]$ as

$$\begin{aligned} & Cov[\log V(T), X_N] \\ = \; & \sqrt{N}\sum_{k=1}^{N}\sum_{n=0}^{N-1} E[(\bar{\sigma}^\top \Delta W_k \\ & + \sum_{j=N(k)+1}^{N(k+1)} (\log\sum_i w_i Y_j^i - \mu_J))(\frac{\hat{V}((n+1)\Delta t)}{V((n+1)\Delta t)} - \frac{\hat{V}(n\Delta t)}{V(n\Delta t)})] \end{aligned}$$

where, as before, $\mu_J = E[\log\sum_i w_i Y_j^i]$. If we interchange the order of summation and fix a value of n, we need to evaluate

$$E[(\bar{\sigma}^\top \Delta W_k + \sum_{j=N(k)+1}^{N(k+1)} (\log\sum_i w_i Y_j^i - \mu_J))(\frac{\hat{V}((n+1)\Delta t)}{V((n+1)\Delta t)} - \frac{\hat{V}(n\Delta t)}{V(n\Delta t)})], \tag{B.1}$$

for which we have three cases:

(1) $k \geq n+2$. In this case, we have

$$E[(\bar{\sigma}^\top \Delta W_k + \sum_{j=N(k)+1}^{N(k+1)} (\log \sum_i w_i Y_j^i - \mu_J))(\frac{\hat{V}((n+1)\Delta t)}{V((n+1)\Delta t)} - \frac{\hat{V}(n\Delta t)}{V(n\Delta t)})] = 0,$$

because $W(k)$ and $\sum_{j=N(k)+1}^{N(k+1)} (\log \sum_i w_i Y_j^i)$ are both independent of

$(\hat{V}(n\Delta t), V(n\Delta t), \hat{V}((n+1)\Delta t), V((n+1)\Delta t))$.

(2) $k = n+1$. (B.1) becomes

$$E\left[\frac{\hat{V}(n\Delta t)}{V(n\Delta t)}\right] E\left[(\bar{\sigma}^\top \Delta W_{n+1} + \sum_{j=N(k)+1}^{N(k+1)} (\log \sum_i w_i Y_j^i - \mu_J))\frac{\hat{R}_{n+1}}{R_{n+1}}\right]. \tag{B.2}$$

Multiplying the factors inside the last expectation produces two terms. For the first, we have

$$\begin{aligned}
&E[\bar{\sigma}^\top \Delta W_{n+1} \frac{\hat{R}_{n+1}}{R_{n+1}}] \\
&= \sum_i w_i E[\bar{\sigma}^\top \Delta W_{n+1} \exp\{(\mu_i - \mu_w - \frac{1}{2}\|\sigma_i\|^2 + \frac{1}{2}\sigma_w^2)\Delta t + (\sigma_i - \bar{\sigma})^\top \Delta W_{n+1}\} \\
&\prod_{j=N(n+1)+1}^{N(n+2)} \frac{Y_j^i}{\sum w_l Y_j^l}] \\
&= \sum_i w_i (\bar{\sigma}^\top \sigma_i - \sigma_w^2)\Delta t \exp\{(\mu_i - \mu_w + \sigma_w^2 - \sigma_i^\top \bar{\sigma})\Delta t + \lambda \Delta t(\mu_i^y)\} \\
&= \gamma_L \Delta t^2 + \sum_i w_i \bar{\sigma}^\top \sigma_i \lambda \mu_i^y \Delta t^2 + O(\Delta t^3).
\end{aligned} \tag{B.3}$$

For the other term, from (B.2) we have

$$
\begin{aligned}
&E[\sum_{j=N(k)+1}^{N(k+1)}(\log\sum_i w_i Y_j^i-\mu_J)\frac{\hat{R}_{n+1}}{R_{n+1}}]\\
&=\sum_i w_i\exp\{(\mu_i-\mu_w+\sigma_w^2-\sigma_i^\top\bar{\sigma})\Delta t\}\\
&\times E[(\prod_{r=N(k)+1}^{N(k+1)}(\bar{Y}^i+1))(\sum_{j=N(k)+1}^{N(k+1)}(\log\sum_i w_i Y_j^i-\mu_J))],
\end{aligned}
\tag{B.4}
$$

where

$$
\begin{aligned}
&E[(\prod_{r=N(k)+1}^{N(k+1)}(\bar{Y}_r^i+1))(\sum_{j=N(k)+1}^{N(k+1)}(\log\sum_i w_i Y_j^i-\mu_J))]\\
&=\sum_{n=1}^{\infty}e^{-\lambda\Delta t}\frac{(\lambda\Delta t)^n}{n!}\sum_{j=1}^{n}E[\prod_{k=1}^{n}\left(\bar{Y}^i+1\right)(\log\sum_l w_l Y_j^l-\mu_J)]\\
&=\exp\{\lambda\Delta t\mu_i^y\}\Delta t\lambda E[(\bar{Y}^i+1)(\log\sum_l w_l Y^l-\mu_J)].
\end{aligned}
\tag{B.5}
$$

Substituting (B.5) into (B.4), we get

$$
\begin{aligned}
&E[\sum_{j=N(k)+1}^{N(k+1)}(\log\sum_i w_i Y_j^i-\mu_J)(\frac{\hat{R}_{n+1}}{R_{n+1}})]\\
&=\sum_i w_i\exp\{(\mu_i-\mu_w+\sigma_w^2-\sigma_i^\top\bar{\sigma})\Delta t\}\exp\{\lambda\Delta t\mu_i^y\}\\
&\times\Delta t\lambda E[(\bar{Y}^i+1)(\log\sum_l w_l Y^l-\mu_J)].
\end{aligned}
\tag{B.6}
$$

Applying a Taylor expansion to the exponential part under assumptions (4.2.4) and (4.2.5), (B.6)

becomes

$$\sum_i w_i \lambda(\mu_i - \sigma_i^\top \bar{\sigma} + \lambda \mu_i^y) E[\bar{Y}^i (\log \sum_l w_l Y^l - \mu_J)] \Delta t^2 + O(\Delta t^3). \tag{B.7}$$

Using (B.3) and (B.7) we have for (B.2)

$$E[(\bar{\sigma}^\top \Delta W_{n+1} + \sum_{j=N(k)+1}^{N(k+1)} (\log \sum_i w_i Y_j^i - E))(\frac{\hat{R}_{n+1}}{R_{n+1}})] = \tilde{\gamma}_L \Delta t^2 + O(\Delta t^3).$$

(3) $k < n+1$. The same argument applies in this case, and we have

$$E[(\bar{\sigma}^\top \Delta W_{n+1} + \sum_{j=N(k)+1}^{N(k+1)} (\log \sum_i w_i Y_j^i - \mu_J))(\frac{\hat{V}((n+1)\Delta t)}{V((n+1)\Delta t)} - \frac{\hat{V}(n\Delta t)}{V(n\Delta t)})] = O(\Delta t^4).$$

Hence we have

$$N^{-1/2} Cov[\log T(T), X_N] = \frac{\tilde{\gamma}_L T^2}{N} + O(N^{-2}).$$

(ii) For the second part of the proposition, we need to show that

$$E[(\bar{V}(T) - \hat{V}(T))^2] = O(N^{-2}).$$

By following the steps of a similar proof in Glasserman (2012), it suffices to show $E[V(T)^2 X_N^2] < \infty$.

We can write

$$V^2(T) = \exp\{2\mu_w T + \sigma_w^2 T\} \exp\{2\bar{\sigma}^\top W(T) - 2\sigma_w^2 T\} \exp\{-(\lambda - \tilde{\lambda})T\}$$
$$\times \exp\{(\lambda - \tilde{\lambda})T\} \prod_{j=1}^{N(t)} (\sum_i w_i Y_j^i)^2,$$

and now we would like to use the following as a Radon-Nikodym derivative:

$$\exp\{2\bar{\sigma}^\top W(T) - 2\sigma_w^2 T\} \exp\{(\lambda - \tilde{\lambda})T\} \prod_{j=1}^{N(t)} (\sum_i w_i Y_j^i)^2. \tag{B.8}$$

The first exponential term is itself a Radon-Nikodym derivative for the diffusion process. From assumption (4.2.4), we have $E[(Y^i)^2] < \infty$, so we can choose an appropriate $\tilde{\lambda}$ such that $\tilde{f}(y) = \lambda y^2 f(y)/\tilde{\lambda}$ is a well-defined density function, where $f(.)$ and $\tilde{f}(.)$ are the density functions for $\sum_i w_i Y^i$ under the original probability and the new probability measure, respectively. Therefore, (B.8) is indeed a Radon-Nikodym derivative, and, under the probability measure it defines, each asset's drift is changed from μ_i to $\mu_i + 2\sigma_i^\top \bar{\sigma}$, and the $\sum_i w_i Y^i$ now have density $\tilde{f}$.

From Theorem 4.2.3, the convergence of the second moment of X_N holds as long as the drifts and Poisson rate are constant, and assumption (4.2.3) and the first inequality of (4.2.4) hold under the new measure. Because of absolute continuity, (4.2.3) will still hold. For (4.2.4)

$$\tilde{E}[|\bar{Y}^k + 1|^3] = \exp\{(\lambda - \tilde{\lambda}T)\} E[|Y^k|^2 |\bar{Y}^k + 1|]$$
$$\leq \exp\{(\lambda - \tilde{\lambda}T)\} \|\bar{Y}^k + 1\|_3 \|Y^k\|_3^2 < \infty.$$

Hence we have proved the second part of the proposition.

□

B.7.2 Proof of Proposition 4.3.1: the Mean-Reverting Case

Proof. Under assumption (4.2.3), first we focus on the case of only one jump

$$
\begin{aligned}
&P(I_d^n = 1) \leq P(I_d^n = 1, N(\Delta t) = 1) + P(N(\Delta t) > 1) \\
&= P(\sum_i w_i \exp\{(\mu_i - \frac{1}{2}\sum_{j=1}^{d} \sigma_{ij}^2)s + \sigma_i^\top W(s)\} \prod_{j=1}^{N(s)} Y_j^i < 0, \\
&\text{for some } s \in [0, \Delta t] | N(\Delta t) = 1) P(N(\Delta t) = 1) + O(\Delta t^2). \qquad \text{(B.9)}
\end{aligned}
$$

The last term $O(\Delta t^2)$ is from the probability of more than one jump within the time interval. Now we simplify the first term by using the fact of having only one jump, and also apply a first-order Taylor expansion to the exponential:

$$
\begin{aligned}
&P(\sum_i w_i \exp\{(\mu_i - \frac{1}{2}\sum_{j=1}^{d} \sigma_{ij}^2)s + \sigma_i^\top W(s)\} \prod_{j=1}^{N(s)} Y_j^i < 0, \\
&\text{for some } s \in [0, \Delta t] | N(\Delta t) = 1) \\
&= P(\sum_i w_i (1 + \sigma_i^\top W(s) + \bar{r}_i(s)) Y^i < 0, \text{ for some } s \in [0, \Delta t]),
\end{aligned}
$$

where $\bar{r}_i$ is the remainder in the Taylor approximation, with L_2-norm $O(\Delta t)$. Then

$$
\begin{aligned}
&P(\sum_i w_i(1+\sigma_i^\top W(s)+\bar{r}_i(s))Y^i<0, \text{ for some } s\in[0,\Delta t])\\
&\leq P(\max(|\sigma_i^\top W(s)+\bar{r}_i(s)|)\sum_i |w_iY^i|>\sum_i w_iY^i, \text{ for some } s\in[0,\Delta t])\\
&\leq P(\sum_i |\sigma_i^\top W(s)+\bar{r}_i(s)|>\frac{\sum_i w_iY^i}{\sum_i |w_iY^i|}, \text{ for some } s\in[0,\Delta t]).
\end{aligned}
$$

Conditioning on the Y^i and applying Chebyshev's inequality yields

$$
\begin{aligned}
&P(\sum_i |\sigma_i^\top W(s)+\bar{r}_i(s)|>\frac{\sum_i w_iY^i}{\sum_i |w_iY^i|}, \text{ for some } s\in[0,\Delta t])\\
&\leq E[Var(\sum_i |\sigma_i^\top W(s)+\bar{r}_i(s)|)\left(\frac{\sum_i w_iY^i}{\sum_i |w_iY^i|}\right)^2, \text{ for some } s\in[0,\Delta t]]\\
&\leq Var(\sum_i |\sigma_i^\top W(s)+\bar{r}_i(s)|)E\left[\left(\frac{\sum_i w_iY^i}{\sum_i |w_iY^i|}\right)^2\right]\\
&=O(\Delta t).
\end{aligned}
$$

Substituting these results in (B.9) concludes the proof.

□

B.8 Proof of Lemma B.8.1

Lemma B.8.1. *Given assumption (4.2.3) ,* $P(I_d^n=0)=O(\Delta t^2)$.

Proof. (i) With

$$X_N = \sqrt{N}\sum_{n=0}^{N-1}(\frac{\hat{\underline{V}}((n+1)\Delta t)}{\underline{V}((n+1)\Delta t)} - \frac{\hat{\underline{V}}(n\Delta t)}{\underline{V}(n\Delta t)}),$$

we have

$$\begin{aligned} &Cov[\log \underline{V}(T), X_N] \\ &= \sqrt{N}\sum_{k=1}^{N}\sum_{n=0}^{N-1} E[\bar{\sigma}^\top e^{-\kappa}\int_{(k-1)\Delta t}^{k\Delta t} e^{\kappa s} dW_s(\frac{\hat{\underline{V}}((n+1)\Delta t)}{\underline{V}((n+1)\Delta t)} - \frac{\hat{\underline{V}}(n\Delta t)}{\underline{V}(n\Delta t)})]. \end{aligned} \tag{B.1}$$

For $k \geq n+2$,

$$E[\bar{\sigma}^\top e^{-\kappa}\int_{(k-1)\Delta t}^{k\Delta t} e^{\kappa s} dW_s(\frac{\hat{\underline{V}}((n+1)\Delta t)}{\underline{V}((n+1)\Delta t)} - \frac{\hat{\underline{V}}(n\Delta t)}{\underline{V}(n\Delta t)})] = 0,$$

For $k = n+1$, we have

$$\begin{aligned} &E[\bar{\sigma}^\top e^{-\kappa}\int_{(n+1)\Delta t}^{(n+2)\Delta t} e^{\kappa s} dW_s(\frac{\hat{\underline{V}}((n+1)\Delta t)}{\underline{V}((n+1)\Delta t)} - \frac{\hat{\underline{V}}(n\Delta t)}{\underline{V}(n\Delta t)})] \\ &= E[\frac{\hat{\underline{V}}(n\Delta t)}{\underline{V}(n\Delta t)}]E[\bar{\sigma}^\top e^{-\kappa}\int_{(n+1)\Delta t}^{(n+2)\Delta t} e^{\kappa s} dW_s \frac{\hat{\underline{R}}_{n+1}}{\underline{R}_{n+1}}]. \end{aligned}$$

$$
\begin{aligned}
&E[\bar{\sigma}^\top e^{-\kappa}\int_{(n+1)\Delta t}^{(n+2)\Delta t} e^{\kappa s} dW_s \frac{\hat{\underline{R}}_{n+1}}{\underline{R}_{n+1}}] \\
&= \sum_i w_i \bar{\sigma}^\top(\sigma_i - \bar{\sigma}) e^{-\kappa(1+\Delta t)} \int_0^{\Delta t} e^{\kappa s} ds \ \exp\{(\mu_i - \mu_w - \frac{1}{2}(\|\sigma_i\|^2 - \sigma_w^2))\Delta t \\
&\quad + \frac{1}{2}\|\sigma_i - \bar{\sigma}\|^2 e^{-2\kappa\Delta t}\int_0^{\Delta t} e^{2\kappa s} ds + (1 - e^{-\kappa\Delta t})(\mu_i^l - \bar{\mu}^l))\}.
\end{aligned}
\tag{B.2}
$$

We only need its coefficient on Δt^2, which is

$$
\begin{aligned}
&\sum_i w_i(\bar{\sigma}^\top\sigma_i)e^{-\kappa}(\mu_i - \mu_w - \frac{1}{2}(\|\sigma_i\|^2 - \sigma_w^2) + \frac{1}{2}\|\sigma_i - \bar{\sigma}\|^2 + \kappa(\mu_i^l - \bar{\mu}^l)) \\
&= \sum_i w_i(\bar{\sigma}^\top\sigma_i)e^{-\kappa}(\mu_i - \mu_w + \sigma_w^2 - \sigma_i^\top\bar{\sigma} + \kappa(\mu_i^l - \bar{\mu}^l)) \\
&= e^{-\kappa}(\gamma_L + \sum_i w_i(\bar{\sigma}^\top\sigma_i)\kappa(\mu_i^l - \bar{\mu}^l)).
\end{aligned}
$$

For the first factor in (B.2), we have

$$
E[\frac{\hat{\underline{V}}(n\Delta t)}{\underline{V}(n\Delta t)}] = \prod_{k=1}^n E[\frac{\hat{\underline{R}}_{n+1}}{\underline{R}_{n+1}}] = \prod_{k=1}^n(1 + O(\Delta t^2)) = 1 + O(\Delta t).
$$

So, we have

$$
E[\bar{\sigma}^\top e^{-\kappa}\int_{(n+1)\Delta t}^{(n+2)\Delta t} e^{\kappa s} dW_s(\frac{\hat{\underline{V}}((n+1)\Delta t)}{\underline{V}((n+1)\Delta t)} - \frac{\hat{\underline{V}}(n\Delta t)}{\underline{V}(n\Delta t)})] = \underline{\gamma}_L\Delta t^2 + O(\Delta t^3).
$$

For the case $k \leq n$, following the same argument as in the proof of (Glasserman 2012, Prop.

4), we get

$$E[\bar{\sigma}^\top e^{-\kappa}\int_{k\Delta t}^{(k+1)\Delta t} e^{\kappa s} dW_s(\frac{\hat{\underline{V}}((n+1)\Delta t)}{\underline{V}((n+1)\Delta t)} - \frac{\hat{\underline{V}}(n\Delta t)}{\underline{V}(n\Delta t)})] = O(\Delta t^4),$$

and then (B.1) becomes

$$N^{-1/2}Cov(\log \underline{V}(T), X_N) = \frac{\underline{\gamma}_L T^2}{N} + O(N^{-2}).$$

The proof for part (ii) follows the same line as the one in Glasserman (2012). The only modification needed is that now the Girsanov transformation is a little more general, the change of measure now changing the standard Brownian motion $W(T)$ to a Gaussian process $\int_0^T e^{\kappa s} W(s)$.

□

B.9 Proof of Proposition 4.5.1

Proof. With N fixed, since $\{\hat{R}_{n,N} : n = 1, ..., N\}$ are i.i.d., from (4.5.1) and the surrounding discussion, the number of intervals n until $I_d^n = 1$ has a geometric distribution, and

$$P(I_d^n = 1 \text{ for some } n = 1, \ldots, N) = O(\Delta t).$$

If the discrete portfolio defaults, $\hat{\underset{\sim}{V}}(T) = 0$ and $\frac{\hat{\mathcal{V}}(T) - \mathcal{V}(T)}{\mathcal{V}(T)} = -1$, so

$$\|\frac{\hat{\underset{\sim}{V}}(T) - \underset{\sim}{V}(T)}{\underset{\sim}{V}(T)} I_{\{I_d^n = 1 \text{ for some } n=1,\ldots,N\}}\| = O(\Delta t).$$

□

Bibliography

Asness, C., T. Moskowitz, L. Pedersen. 2009. Value and momentum everywhere. *AFA 2010 Atlanta Meetings Paper*.

Avellaneda, M. 1998. Minimum-relative-entropy calibration of asset-pricing models. *International Journal of Theoretical and Applied Finance* **1**(04) 447–472.

Avellaneda, M., R. Buff, C. Friedman, N. Grandechamp, L. Kruk, J. Newman. 2001. Weighted monte carlo: a new technique for calibrating asset-pricing models. *International Journal of Theoretical and Applied Finance* **4**(1) 91–119.

Avellaneda, M., C. Friedman, R. Holmes, D. Samperi. 1997. Calibrating volatility surfaces via relative-entropy minimization. *Applied Mathematical Finance* **4**(1) 37–64.

Avellaneda, M., A. Levy, A. Parás. 1995. Pricing and hedging derivative securities in markets with uncertain volatilities. *Applied Mathematical Finance* **2**(2) 73–88.

Avellaneda, M., S. Zhang. 2010. Path-Dependence of Leveraged ETF Returns. *SIAM Journal on Financial Mathematics* **1** 586–603.

Basak, S., G. Chabakauri. 2010. Dynamic mean-variance asset allocation. *Review of Financial Studies* **23**(8) 2970–3016.

Basel Committee on Banking Supervision. 2007. Guidelines for computing capital for incremental default risk in the trading book (bcbs 134). *Bank for International Settlements, Basel, Switzerland. Available at www.bis.org* .

Basel Committee on Banking Supervision. 2009. Guidelines for computing capital for incremental risk in the trading book (bcbs 149). *Bank for International Settlements, Basel, Switzerland. Available at www.bis.org* .

Basurto, M.A.S., C.A.E. Goodhart. 2009. *Banking stability measures*. International Monetary Fund.

Ben-Tal, A., T. Margalit, A. Nemirovski. 2000. Robust modeling of multi-stage portfolio problems. *High Performance Optimization* 303–328.

Bertsekas, D.P. 2007. *Dynamic programming and optimal control*. Athena Scientific Belmont, MA.

Bertsimas, D., L. Kogan, A.W. Lo. 2000. When is time continuous? *Journal of Financial Economics* **55** 173–204.

Bertsimas, D., D. Pachamanova. 2008. Robust multiperiod portfolio management in the presence of transaction costs. *Computers & Operations Research* **35**(1) 3–17.

Bielecki, T.R., S.R. Pliska. 1999. Risk-sensitive dynamic asset management. *Applied Mathematics & Optimization* **39**(3) 337–360.

Bielecki, T.R., S.R. Pliska, S.J. Sheu. 2005. Risk sensitive portfolio management with cox–ingersoll–ross interest rates: The HJB equation. *SIAM Journal on Control and Optimization* **44**(5) 1811–1843.

Boyd, S., L. Vandenberghe. 2004. *Convex optimization*. Cambridge University Press.

Boyle, P.P., D. Emanuel. 1980. Discretely adjusted option hedges. *Journal of Financial Economics* **8** 259–282.

Branger, N., E. Krautheim, C. Schlag, N. Seeger. 2012. Hedging under model misspecification: All risk factors are equal, but some are more equal than others. *Journal of Futures Markets* **32**(5) 397–430.

Bruti-Liberati, N., E. Platen. 2005. On the strong approximation of jump-diffusion processes. *Technical Report, Quantitative Finance Research Papers 157, University of Technology, Sydney, Australia* .

Bruti-Liberati, N., E. Platen. 2007. Approximation of jump diffusions in finance and economics. *Computational Economics* **29** 283–312.

Buchen, P.W., M. Kelly. 1996. The maximum entropy distribution of an asset inferred from option prices. *Journal of Financial and Quantitative Analysis* **31**(1).

Campbell, J.Y., L.M. Viceira. 2002. *Strategic asset allocation: portfolio choice for long-term investors*. Oxford University Press, USA.

Cont, R., R. Deguest. 2010. Equity correlations implied by index options: estimation and model uncertainty analysis. *Mathematical Finance* .

Cont, R., P. Tankov. 2003. *Financial Modelling with Jump Processes*. Chapman & Hall/CRC.

Cont, R., P. Tankov. 2004. Non-parametric calibration of jump-diffusion option pricing models.

Cont, R., P. Tankov. 2006. Retrieving lévy processes from option prices: Regularization of an ill-posed inverse problem. *SIAM Journal on Control and Optimization* **45**(1) 1–25.

Das, S.R., R. Uppal. 2004. Systemic risk and international portfolio choice. *The Journal of Finance* **59**(6) 2809–2834.

Delage, E., Y. Ye. 2010. Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations Research* 1–18.

Dempster, Mitra G. Pflug G., M.A.H. 2009. *Quantitative Fund Management*. Chapman & Hall/CRC Press, Boca Raton, Florida.

Dey, S., S. Juneja. 2010. Entropy approach to incorporate fat tailed constraints in financial models. *Available at SSRN 1647048* .

Draper, D. 1995. Assessment and propagation of model uncertainty. *Journal of the Royal Statistical Society. Series B (Methodological)* 45–97.

Duffie, D., T.S. Sun. 1990. Transactions costs and portfolio choice in a discrete-continuous-time setting. *Journal of Economic Dynamics and Control* **14** 35–51.

El Bachir, N., D. Brigo. 2008. An analytically tractable time-changed jump-diffusion default intensity model. *ICMA Centre Discussion Papers in Finance DP2008-6* .

Erb, C.B., C.R. Harvey. 2006. The strategic and tactical value of commodity futures. *Financial Analysts Journal* **62**(2) 69–97.

Fleming, W.H., S.J. Sheu. 2001. Risk-sensitive control and an optimal investment model. *Mathematical Finance* **10**(2) 197–213.

Garleanu, N.B., L.H. Pedersen. 2012. Dynamic trading with predictable returns and transaction costs. *The Journal of Finance* Forthcoming.

Glasserman, P. 2012. Risk horizon and rebalancing horizon in portfolio risk measurement. *Mathematical Finance* **22**(2) 215–249.

Glasserman, P., X. Xu. 2010. Importance sampling for tail risk in discretely rebalanced portfolio. *Proceedings of the Winter Simulation Conference* 2655–2665.

Glasserman, P., X. Xu. 2013. Robust portfolio control with stochastic factor dynamics. *Operations Research* Forthcoming.

Glasserman, P., B. Yu. 2005. Large sample properties of weighted monte carlo estimators. *Operations Research* **53**(2) 298–312.

Goh, J., M. Sim. 2010. Distributionally robust optimization and its tractable approximations. *Operations Research* **58**(4-Part-1) 902–917.

Goldfarb, D., G. Iyengar. 2003. Robust portfolio selection problems. *Mathematics of Operations Research* **28**(1) 1–38.

Guasoni, P., G. Huberman, Z. Wang. 2011. Performance maximization of actively managed funds. *Journal of Financial Economics* **101**(3) 574–595.

Gulko, L. 1999. The entropy theory of stock option pricing. *International Journal of Theoretical and Applied Finance* **2**(03) 331–355.

Gulko, L. 2002. The entropy theory of bond option pricing. *International Journal of Theoretical and Applied Finance* **5**(04) 355–383.

Hansen, L.P., T.J. Sargent. 1995. Discounted linear exponential quadratic gaussian control. *Automatic Control, IEEE Transactions on* **40**(5) 968–971.

Hansen, L.P., T.J. Sargent. 2007. *Robustness*. Princeton University Press.

Hansen, L.P., T.J. Sargent, G. Turmuhambetova, N. Williams. 2006. Robust control and model misspecification. *Journal of Economic Theory* **128**(1) 45–90.

Haugh, M. 2011. A note on constant proportion trading strategies. *Operations Research Letters* **39**(3) 172–179.

Iyengar, G.N. 2005. Robust dynamic programming. *Mathematics of Operations Research* **30**(2) 257–280.

Jabbour, C., J.F. Peña, J.C. Vera, L.F. Zuluaga. 2008. An estimation-free, robust conditional value-at-risk portfolio allocation model. *The Journal of Risk* **11**(1) 57–78.

Jacod, J., P. Protter. 1998. Asymptotic error distributions for the Euler method for stochastic differential equations. *Annals of Probability* **26** 267–307.

Jessen, C. 2010. Constant proportion portfolio insurance: Discrete-time trading and gap risk coverage. *23rd Australasian Finance and Banking Conference* .

Kallsen, J. 2000. Optimal portfolios for exponential lévy processes. *Mathematical methods of operations research* **51**(3) 357–374.

Kim, T.S., E. Omberg. 1996. Dynamic nonmyopic portfolio behavior. *Review of Financial Studies* **9**(1) 141.

Kloeden, P.E., E. Platen. 1992. *Numerical Solution of Stochastic Differential Equations*. Springer-Verlag, Berlin.

Kurtz, T., P. Protter. 1991. Wong-Zakai corrections, random evolutions and numerical schemes for SDEs. *Stochastic Analysis, Academic Press, Boston,MA* 331–346.

Ledoit, O., M. Wolf. 2003. Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *Journal of Empirical Finance* **10**(5) 603–621.

Leland, H.E. 1985. Option pricing and replication with transactions costs. *Journal of Finance* **40** 1283–1301.

Lim, A.E.B., J.G. Shanthikumar, T. Watewai. 2011. Robust asset allocation with benchmarked objectives. *Mathematical Finance* **21**(4) 643–679.

Madan, D.B., P.P. Carr, E.C. Chang. 1998. The variance gamma process and option pricing. *European Finance Review* **2**(1) 79–105.

Markowitz, H. 1952. Portfolio selection. *The journal of finance* **7**(1) 77–91.

Meyer, C.D. 2001. *Matrix analysis and applied linear algebra*, vol. 2. Society for Industrial Mathematics.

Moallemi, C., M. Saglam. 2012. Dynamic portfolio choice with linear rebalancing rules. *working paper* .

Morini, M. 2011. *Understanding and Managing Model Risk: A Practical Guide for Quants, Traders and Validators*. The Wiley Finance Series, Wiley.

Morton, A.J., S.R. Pliska. 1995. Optimal portfolio management with fixed transaction costs. *Mathematical Finance* **5** 337–356.

Moskowitz, T., Y.H. Ooi, L.H. Pedersen. 2012. Time series momentum. *Journal of Financial Economics* **104** 228C250.

Muñoz, D.F., P.W. Glynn. 1997. A batch means methodology for estimation of a nonlinear function of a steady-state mean. *Management Science* **43**(8) 1121–1135.

Mykland, P.A. 2000. Conservative delta hedging. *The Annals of Applied Probability* **10**(2) 664–683.

Natarajan, K., D. Pachamanova, M. Sim. 2008. Incorporating asymmetric distributional information in robust value-at-risk optimization. *Management Science* **54**(3) 573–585.

Natarajan, K., M. Sim, J. Uichanco. 2010. Tractable robust expected utility and risk models for portfolio optimization. *Mathematical Finance* **20**(4) 695–731.

Nilim, A., L. El Ghaoui. 2005. Robust control of markov decision processes with uncertain transition matrices. *Operations Research* **53**(5) 780–798.

Pesaran, M.H., C. Schleicher, P. Zaffaroni. 2009. Model averaging in risk management with an application to futures markets. *Journal of Empirical Finance* **16**(2) 280–305.

Pesaran, M.H., A. Timmermann. 2012. Predictability of stock returns: Robustness and economic significance. *The Journal of Finance* **50**(4) 1201–1228.

Petersen, I.R., M.R. James, P. Dupuis. 2000. Minimax optimal control of stochastic uncertain systems with relative entropy constraints. *Automatic Control, IEEE Transactions on* **45**(3) 398–412.

Platen, E. 1982. A generalized Taylor formula for solutions of stochastic equations. *Sankhya: The Indian Journal of Statistics* **44, No. 2** 163–172.

Raftery, A.E., D. Madigan, J.A. Hoeting. 1997. Bayesian model averaging for linear regression models. *Journal of the American Statistical Association* **92**(437) 179–191.

Rényi, A. 1961. On measures of entropy and information. *Fourth Berkeley Symposium on Mathematical Statistics and Probability*. 547–561.

Rockafellar, R.T., S. Uryasev. 2002. Conditional value-at-risk for general loss distributions. *Journal of Banking & Finance* **26**(7) 1443–1471.

Ruszczyński, A. 2010. Risk-averse dynamic programming for markov decision processes. *Mathematical Programming* **125**(2) 235–261.

Sepp, A. 2012. An approximate distribution of delta-hedging errors in a jump-diffusion model with discrete trading and transaction costs. *Quantitative Finance* **12**(7) 1119–1141.

Shapiro, A., D. Dentcheva, A. Ruszczyński. 2009. *Lectures on stochastic programming: modeling and theory*, vol. 9. Society for Industrial and Applied Mathematics.

Skiadas, C. 2003. Robust control and recursive utility. *Finance and Stochastics* **7**(4) 475–489.

Studer, M. 2001. Stochastic Taylor Expansions and Saddlepoint Approximations for Risk Management. *Dissertation No. 14242, Department of Mathematics, ETH Zurich, Available at e-collection.ethbib.ethz.ch* .

Szechtman, R., P.W. Glynn. 2001. Constrained monte carlo and the method of control variates. *Proceedings of the 33nd conference on Winter simulation*. IEEE Computer Society, 394–400.

Tankov, P., E. Voltchkova. 2009. Asymptotic analysis of hedging errors in models with jumps. *Stochastic Processes and their Applications* **119** 2004–2027.

Tsallis, C. 1988. Possible generalization of boltzmann-gibbs statistics. *Journal of Statistical Physics* **52**(1) 479–487.

Weber, E.J. 2003. The misuse of central bank gold holdings. *Gold and the Modern World Economy* **38** 64–81.

Whittle, P. 1981. Risk-sensitive linear/quadratic/gaussian control. *Advances in Applied Probability* **13**(4) 764–777.

Whittle, P., P.R. Whittle. 1990. *Risk-sensitive optimal control*. Wiley Chichester.

Zhu, S., M. Pykhtin. 2007. A guide to modeling counterparty credit risk. *GARP Risk Review, July/August* .