*Quantitative Finance*, 2013
http://dx.doi.org/10.1080/14697688.2013.822989

Routledge
Taylor & Francis Group

# Robust risk measurement and model risk

PAUL GLASSERMAN† and XINGBO XU*‡

†Columbia Business School, Columbia University, New York 10027, NY, USA
‡IEOR Department, School of Engineering and Applied Sciences, Columbia University, 500 West 120th Street, New York 10027, NY, USA

(*Received 18 September 2012; accepted 2 July 2013*)

Financial risk measurement relies on models of prices and other market variables, but models inevitably rely on imperfect assumptions and estimates, creating model risk. Moreover, optimization decisions, such as portfolio selection, amplify the effect of model error. In this work, 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.

*Keywords*: Risk measures; Validation of pricing models; Derivatives risk management; Risk management

*JEL Classification*: C63, D81



## 1. Introduction

Risk measurement relies on modelling assumptions. Errors in these assumptions introduce errors in risk measurement. This makes risk measurement vulnerable to *model risk*.

This paper develops tools for quantifying model risk and making risk measurement robust to modeling errors. Simplifying assumptions are inherent to all modelling, so the first goal of model risk management is to assess vulnerabilities to model errors and their potential impact. We develop the following objectives:

- to bound the effect of model error on specific measures of risk, given a baseline nominal model for measuring risk and
- 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 on this risk measure.

For the first objective, we calculate an upper or lower bound (or both) on the range of risk values that can result over a range of model errors within a certain 'distance' of a nominal model. These bounds are somewhat analogous to a confidence interval; but whereas a confidence interval quantifies the effect of sampling variability, the *robustness* bounds we develop quantify the effect of model error.

For the second objective, we identify the changes to a nominal underlying model that attain the bounds on the measure of risk—in other words, we identify the worst-case error in the nominal model. This step is crucial. Indeed, simply quantifying the potential magnitude of model risk would be of limited value if we could not point to the sources of model vulnerability that lead to the largest errors in measuring risk.

A simple example should help illustrate these ideas. Standard deviation is a conventional measure of risk for portfolio returns. Measuring standard deviation prospectively requires assumptions about the joint distribution of the returns of the assets in the portfolio. For the first objective listed above, we would want to bound the values of standard deviation that can result from a reasonable (in a sense to be quantified) degree of model error. For the second objective, we would want to identify which changes in the assumed joint distribution of returns have the largest impact on the portfolio standard deviation.

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

*Corresponding author. Email: xx2126@columbia.edu

different markets. More often, if it is considered at all, model risk is investigated by varying model parameters. Importantly, the tools developed here go beyond parameter sensitivity to consider the effect of changes in the probability law that defines an underlying model. This allows us to identify vulnerabilities to model error 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.

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 following Hansen and Sargent (2007). 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. It is also a convenient choice because the worst-case alternative within a relative entropy constraint is typically given by an exponential change of measure. Indeed, relative entropy has been applied for model calibration and estimation in numerous sources, including Avellaneda (1998), Avellaneda *et al.* (2000, 1997), Buchen and Kelly (1996), Cont and Deguest (2013), Cont and Tankov (2004, 2006), Gulko (1999, 2002), and Segoviano and Goodhart (2009). In working with heavy-tailed distributions, for which relative entropy may be undefined, we use a related notion of $\alpha$-divergence, as do Dey and Juneja (2010) in a portfolio selection problem.

The tools we develop for risk measurement 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. El Ghaoui *et al.* (2003) 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 (forthcoming). Related techniques are used in Boyarchenko *et al.* (2012) and Meucci (2008). 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.

Our approach combines conveniently with Monte Carlo simulation for risk measurement. 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 adversary's change of probability measure. To understand how the adversary's choice changes the model, we need to simulate under the worst-case model. This is again straightforward because simulating under the original model and then multiplying any output by the adversary's 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 focus, as already noted, is on bounding worst-case model error. 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). For risk measurement, the added conservatism of considering the worst case is often appropriate and can be controlled through the parameter that controls the degree of robustness by penalzing or constraining relative entropy.

The rest of the paper is organized as follows. Section 2 provides an overview of our approach and develops the main supporting theoretical tools. In Section 3, we discuss the implementation of the approach through a set of techniques we call robust Monte Carlo. The remainder of the paper is devoted to illustrative applications: Section 4 considers portfolio variance; Section 5 considers conditional value-at-risk; Section 6 examines the Gaussian copula model of portfolio credit risk; Section 7 investigates delta hedging, comparing the worst-case hedging error with various specific sources of model error; and Section 8 studies model risk in the dependence between exposures and default times in credit valuation adjustment (CVA).

## 2. Overview of the approach

We begin by introducing the main ideas of the paper 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 $f$ and $\tilde{f}$ with a well-defined likelihood ratio $m = \tilde{f}/f$, we define the relative entropy of $\tilde{f}$ with respect to $f$ to be

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

†In some references, relative entropy is defined under the alternative model $\tilde{f}$ by a change of measure, i.e. $\mathcal{R}(f, \tilde{f}) = E[m \log m] = \int \log \frac{\tilde{f}(x)}{f(x)} \tilde{f}(x)\, dx$, which is equilavent to our definition.

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 $\tilde{f}$ preferable to a nominal model $f$.

It is easy to see that $\mathcal{R} \geq 0$, and $\mathcal{R}(f, \tilde{f}) = 0$ only if $\tilde{f}$ and $f$ coincide almost everywhere (with respect to $f$). Relative entropy is not symmetric in $f$ and $\tilde{f}$ and does not define a distance in the usual sense, but $\mathcal{R}(f, \tilde{f})$ is nevertheless interpreted as a measure of how much the alternative $\tilde{f}$ deviates from $f$. (Our views of $f$ and $\tilde{f}$ are generally not symmetric either: we favour the nominal model $f$ but wish to consider the possibility that $\tilde{f}$ 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 $\tilde{f}(x)$ to be $m(x) f(x)$. With the nominal model $f$ fixed, we write $\mathcal{R}(m)$ instead of $\mathcal{R}(f, \tilde{f})$.

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' $\eta$ of the original model. We then seek to evaluate, in addition to the nominal risk measure $E[V(X)]$, the bounds

$$\inf_{m \in \mathcal{P}_\eta} E[m(X)V(X)] \text{ and } \sup_{m \in \mathcal{P}_\eta} E[m(X)V(X)]. \tag{1}$$

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

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

The bounds in (1) thus bound the range of possible values for the risk measure consistent with a degree of model error bounded by $\eta$.

The standard approach to the maximization problem in (1) is to form the dual problem

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

(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

$$m^*_\theta = \frac{\exp(\theta V(X))}{E[\exp(\theta V(X))]}, \tag{2}$$

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 (1) is solved the same way but with $\theta < 0$. The explicit solution we get in (2) is the greatest advantage of working with relative entropy to quantify model error. In Section 3, we will apply (2) at multiple values of $\theta$ to trace out bounds at multiple levels of relative entropy.

### 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 modelled 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

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

We are interested in the worst-case variance

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

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

From (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 $N(\mu, \tilde{\Sigma})$,

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

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

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

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 $2\theta$ times the square of the original variance.

This simple example illustrates ideas that recur throughout this paper. 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 $\Sigma$. 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 (1) measure the greatest error the adversary can introduce, subject to this constraint.

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

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

The relative entropy associated with this change of measure evaluates to

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

By inverting the mapping $\theta \mapsto \eta(\theta)$, we can find the worst-case $\theta$ associated with any relative entropy budget $\eta$. In most of our examples, it is easier to evaluate model error at various values of $\theta$ 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 3.

### 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

$$\inf_{a\in A} E[V_a(X)], \tag{3}$$

for some parameter $a$ ranging over a parameter set $A$. 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 $X$.

We formulate a robust version of the optimization problem (3) as

$$\inf_a \sup_{m\in\mathcal{P}_\eta} E[mV_a(X)]. \tag{4}$$

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

$$\inf_a \inf_{\theta>0} \sup_m E\left[mV_a(X) - \frac{1}{\theta}(m\log m - \eta)\right]. \tag{5}$$

Proposition 2.1 *Under Assumptions* A.1–A.2 *introduced in* Appendix A*, problem* (5) *is equivalent to*

$$\inf_{\theta>0} \inf_a \sup_m E\left[mV_a(X) - \frac{1}{\theta}(m\log m - \eta)\right]. \tag{6}$$

*For fixed* $\theta \in (0, \theta^*_{\max})$*, the corresponding optimal objective function of inner* $\inf_a \sup_m$ *in* (6) *becomes*

$$H(\theta) + \frac{\eta}{\theta} := \inf_a \sup_m E\left[mV_a(X) - \frac{1}{\theta}m\log m + \frac{\eta}{\theta}\right] = \frac{1}{\theta}\log E\left[\exp(\theta V_{a^*(\theta)}(X))\right] + \frac{\eta}{\theta}, \tag{7}$$

*where the optimal decision is*

$$a^*(\theta) = \arg\inf_a \frac{1}{\theta}\log E[\exp(\theta V_a(X))], \tag{8}$$

*and the worst-case change of measure is*

$$m^*_\theta = \exp\left(\theta V_{a^*(\theta)}(X)\right) / E\left[\exp(\theta V_{a^*(\theta)}(X))\right]. \tag{9}$$

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 $\frac{1}{\theta}\log E[\exp(\theta V_a(X))]$ is continuous both in $a$ and $\theta$, we can define the optimal decision and objective function when $\theta$ approaches 0 as follows:

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

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 (8) is convex as well. The constrained problem (4) is equivalent to

$$\inf_{\theta>0} H(\theta) + \frac{\eta}{\theta}. \tag{10}$$

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, (10) has a solution $\theta > 0$ and the optimal value and solution solve the original constraint problem (4).

Proposition 2.2 *With Assumption* A.2 *introduced in* Appendix A*, the objective function in* (8) *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 (8) is convex as well. □

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

$$\eta = E\left[m^*\left(\theta^*_\eta, a^*(\theta^*_\eta)\right)\log m^*\left(\theta^*_\eta, a^*(\theta^*_\eta)\right)\right], \tag{11}$$

i.e. the uncertainty upper bound is reached at the optimal perturbation. So with $\theta^*_\eta > 0$, and the adversary's optimal choice as (9), the original constraint problem (4) has the optimal objective

$$E\left[m^*\left(\theta^*_\eta, a^*\left(\theta^*_\eta\right)\right)V_{a^*(\theta^*_\eta)}(X)\right] = \frac{E\left[V_{a^*(\theta^*_\eta)}(X)\exp\left(\theta^*_\eta V_{a^*(\theta^*_\eta)}(X)\right)\right]}{E\left[\exp\left(\theta^*_\eta V_{a^*(\theta^*_\eta)}(X)\right)\right]}, \tag{12}$$

which differs from the objective function of the penalty version (7) 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 $\eta$. In this case, rather than repeat the procedure above multiple times to solve (10), we can work directly with multiple values of $\theta > 0$ and evaluate $\eta(\theta)$ with each, as in (11). Working with a range of values of $\theta$, this allows us to explore the relationship between $\eta$ and the worst-case error (and this is the approach we use in our numerical examples). This method requires that $\eta$ be an

increasing function of $\theta$, a property we have observed numerically in all of our examples.

### 2.3. Robustness with heavy tails: extension to $\alpha$-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 A.1 introduced in Appendix A. To deal with heavy-tailed distribution, we can use an extension of relative entropy called $\alpha$-divergence and defined as (see also Rényi (1961) and Tsallis (1988))

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

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

With relative entropy replaced by $\alpha$-divergence, the constraint problem (4) becomes

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

The corresponding penalty problem is

$$\begin{aligned} &\inf_a \inf_{\theta>0} \sup_m \quad E\left[ m V_a(X) - \frac{1}{\theta}(\mathcal{D}_\alpha(m) - \eta) \right] \\ &= \inf_{\theta>0} \inf_a \sup_m \; E\left[ m V_a(X) - \frac{1}{\theta}(\mathcal{D}_\alpha(m) - \eta) \right]. \end{aligned} \tag{13}$$

The supremum is taken over valid likelihood ratios—non-negative 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.

PROPOSITION 2.3 *Suppose Assumption* A.3 *introduced in* Appendix A *holds. For any $a \in A$, $\theta > 0$ 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* (13)*:*

$$m^*(\theta, \alpha, a) = (\theta(\alpha - 1)V_a(X) + c(\theta, \alpha, a))^{\frac{1}{\alpha-1}}, \tag{14}$$

*for some constant $c(\theta, \alpha, a)$, such that*

$$\theta(\alpha - 1)V_a(X) + c(\theta, \alpha, a) \geq 0, \tag{15}$$

*and*

$$E\left[ (\theta(\alpha - 1)V_a(X) + c(\theta, \alpha, a))^{\frac{1}{\alpha-1}} \right] = 1. \tag{16}$$

*Proof* The objective of (13) 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

$$\begin{aligned} K(t) &:= E\Bigg[ ((1-t)m^* + tm)V_a \\ &\qquad + \frac{1}{\theta\alpha(1-\alpha)}((1-t)m^* + tm)^\alpha \Bigg] + \frac{\eta}{\theta}, \\ K'(0) &= E\left[ \left\{ V_a + \frac{1}{\theta(1-\alpha)}(m^*)^{\alpha-1} \right\} (m - m^*) \right]. \end{aligned} \tag{17}$$

In order to have $K'(0) = 0$ for any $m$, we need the term inside braces in (17) 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 $\alpha \geq 0$, (15) 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 (15).

A feasible likelihood ratio exists in a neighbourhood 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 (15) and (16) exists for any $\theta \in [0, \theta_0)$. Once $c(\theta, \alpha, a)$ is found, (14) gives an optimal change of measure (not necessarily unique). The optimal decision becomes

$$\begin{aligned} a^*(\theta) = \arg\min_a \frac{\alpha - 1}{\alpha} E\Bigg[ \big(\theta(\alpha - 1)V_a(X) \\ + c(\theta, \alpha, a)\big)^{\frac{1}{\alpha-1}} V_a(X) \Bigg] + \frac{c(\theta, \alpha, a)}{\theta\alpha(1-\alpha)}. \end{aligned} \tag{18}$$

In contrast to the relative entropy case, it is not clear whether the objective in (18) is convex in $a$.

Measuring potential model error through $\alpha$-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

$$\tilde{f}_X(x) \approx c x^{k/(\alpha-1)} f_X(x), \tag{19}$$

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

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 $m$. Relative entropy and $\alpha$-divergence are special cases.

## 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.

### 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 $\theta$ 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 $m_\theta$ by generating the $X_i$ from the original (nominal) measure and forming the estimator

$$\frac{\sum_{i=1}^N V(X_i)\exp(\theta V(X_i))}{\sum_{i=1}^N \exp(\theta V(X_i))}, \tag{20}$$

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 $\theta$ as

$$\hat{\eta}(\theta) = \frac{1}{N}\sum_{i=1}^N \hat{m}_{\theta,i}\log \hat{m}_{\theta,i}. \tag{21}$$

Thus, we can easily estimate $(\eta(\theta), E[m_\theta V(X)])$ across multiple values of $\theta$. Given a relative entropy budget $\eta$, 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 $\eta$. We will illustrate this procedure through several examples.

Just as importantly, we can use the same simulation to analyse 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

$$\frac{1}{N}\sum_{i=1}^N \hat{m}_{\theta,i} h(X_i). \tag{22}$$

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 $\theta$, 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, 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

$$\frac{\sum_{i=1}^N V_{\hat{a}^*}(X_i)\exp(\theta V_{\hat{a}^*}(X_i))}{\sum_{i=1}^N \exp(\theta V_{\hat{a}^*}(X_i))}, \tag{23}$$

where the estimated optimal decision parameter is

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

and the estimated optimal likelihood ratio is

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

By continuous mapping theorem, for given $\hat{a}^*_N$ and any $\theta \in [0, \theta_{\max})$, the averages of both numerator and denominator of (23) are consistent estimators. That is,

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

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

Hence, (23) is a consistent estimator for (12) 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 $\alpha$-divergence framework. For given $\theta > 0, \alpha > 1$ and $a$, we estimate the worst-case likelihood ratio as

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

for some constant $\hat{c}(\theta, a)$, s.t.

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

with

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

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

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

The robust estimator for the objective becomes

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

### 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 $\eta_i$.

Such constraints can be imposed as part of an iterative evaluation of model risk. In (22), we showed how a change of measure selected by an adversary can be analysed 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

$$\sup_{m\in\mathcal{P}_{\mathcal{M}}} E[mV(X)], \tag{24}$$

where

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

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 $\lambda_i$ and transform (24) 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 $V$:

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

When $\theta$ and the $\lambda_i$ are fixed, the problem can be treated as before in (6).

PROPOSITION 3.1 *For fixed $\theta > 0$ and $\lambda_i > 0$, $i = 1,\dots,n_M$, such that*

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

*The worst change of measure is*

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

*The optimization over $(\theta,\lambda_i)$ becomes*

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

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

This is a standard result on constraints in exponential families of probability measures. It is used in Avellaneda *et al.* (2000) 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) analyse the convergence of Monte Carlo estimators in which constraints are imposed by applying weights to the replications.

For an optimization problem as in (3), adding constraints entails solving another layer of optimization. For example, if the original problem is a minimization problem as in (3), then for given $(\theta,\lambda_i)$, the optimal decision becomes

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

### 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 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

$$\inf_a\sup_m E\Bigg[m(X,Y)V_a(X,Y) - \frac{1}{\theta}\big(m(X,Y)\log m(X,Y) - \eta\big)\Bigg] \tag{25}$$

$$\text{s.t.}\quad E[m(X,Y)|Y=y] = 1, \quad \forall y \tag{26}$$

$$m(x,y) \ge 0 \quad \forall x,y.$$

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 3.2 *For problem* (25) *with* $\theta > 0$ *and* $E[\exp(\theta V_a(X,Y))|Y = y] < \infty$ *for all* $y$*:*

(1) *Any likelihood ratio that satisfies* (26) *preserves the law of* $Y$.
(2) *For any* $a$*, the likelihood ratio*

$$m^*(x,y) = \frac{\exp(\theta V_a(x,y))}{E[\exp(\theta V_a(X,Y))|Y=y]}, \tag{27}$$

*is an optimal solution to the maximization part of problem* (25).
(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

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

where $m$ is an arbitrary likelihood ratio satisfying (26). Obviously, $m^*$ satisfies (26). Taking the derivative of $\bar{K}$ at zero and substituting for $m^*$, we get

$$\begin{aligned}\bar{K}'(0) &= E\left[\left(V_a(X,Y) - \frac{1}{\theta}\log m^* - \frac{1}{\theta}\right)(m^* - m)\right] \\ &= E\left[E\left[\left(V_a(X,Y) - \frac{1}{\theta}\log m^* - \frac{1}{\theta}\right)(m^* - m)|Y\right]\right] \\ &= E\Bigg[\frac{1}{\theta}\left(\log E[\exp(\theta V_a(X,Y))|Y] - 1\right) \\ &\qquad \times E\left[(m^* - m)|Y\right]\Bigg]. \end{aligned} \tag{28}$$

By constraint (26), for any $Y = y$, the conditional expectation $E[(m^* - m)|Y]$ in (28) equals zero, so $\bar{K}'(0) = 0$. Hence $m^*$ is an optimal solution satisfying constraint (26).

Next, we show that any likelihood ratio satisfying (26) 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 (27), we need to generate multiple copies $X_1, \ldots, X_N$ for each outcome of $y$ and then form the Monte Carlo counterpart of (27),

$$\hat{m}^*(x, y) = \frac{\exp(\theta V_a(x, y))}{\sum_{i=1}^N \exp(\theta V_a(X_i, y))/N} \tag{29}$$

**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 (20), we can estimate the worst-case error at each level of $\theta$. We can do this across multiple values of $\theta$ at minimal computational cost. By also estimating $\eta(\theta)$ as in (21), we can plot the worst-case error as a function of relative entropy.
- To examine the effect of the change of measure defined by $\theta$, we can estimate moments and the expectations of other auxiliary functions using (22). We can also sample directly from the measure defined by $\theta$ 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 3.1. This technique can be used iteratively to constrain the potential model error if the values estimated through (22) appear implausible.
- Using Proposition 3.2 and (29), 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.

## 4. Portfolio variance

The rest of the paper deals with applications of the ideas developed in the previous sections. In Section 2.1, we illustrated the key ideas of robust risk measurement through an application to portfolio variance. Here, we expand on this example.

### 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

$$\inf_a -E\left[a^\top X - \frac{\gamma}{2} a^\top (X - E[X])(X - E[X])^\top a\right]. \tag{30}$$

As before, $a$ denotes a vector of portfolio weights. To illustrate the method of Section 3.2, we constrain the mean vector and limit uncertainty to the covariance matrix, which leads to the robust problem

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

Following the argument in Section 3.2, for some $a$, $\theta > 0$ and $\lambda$, the worst-case likelihood ratio is

$$m^* \propto \exp\left(\theta\left(V_a(X) - \lambda^\top (X - \mu)\right)\right) \tag{31}$$

where $\lambda$ solves

$$\inf_\lambda \frac{1}{\theta}\log E\left[\exp\left(\theta\left[V(X) - \lambda^\top X\right]\right)\right] + \lambda^\top \mu.$$

Proceeding as in Section 2.1, we find that the worst-case change of measure preserves the normality of $X$. The term with $\lambda$ is linear in $X$ and, therefore, affects only the mean of $X$. Because we have constrained the mean, $m^*$ satisfies

$$m^* \propto \exp\left(\frac{\theta\gamma}{2} a^\top (X-\mu)(X-\mu)^\top a\right), \tag{32}$$

Matching (31) and (32), we find that $\lambda = a$.

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

$$\begin{aligned}a^*(\theta) &= \arg\inf_{a\in\mathcal{A}(\theta)} \frac{1}{\theta}\log E\left[\exp\left(\theta\left[V(X) - \lambda^\top X\right]\right)\right] + \lambda^\top \mu \\ &= \arg\inf_{a\in\mathcal{A}(\theta)} \frac{1}{\sqrt{\det(I - \theta\gamma aa^\top \Sigma)}} + a^\top \mu. \end{aligned} \tag{33}$$

The corresponding relative entropy is

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

To illustrate, we consider an example with 10 assets, where $\mu_i = 0.1$, $\sigma_{ii} = 0.3$ and $\rho_{ij} = 0.25$ for $i \neq j$, $i, j = 1, \ldots, 10$ and $\gamma = 1$. We refer to the optimal portfolio at these parameter values as the nominal portfolio (NP). At each

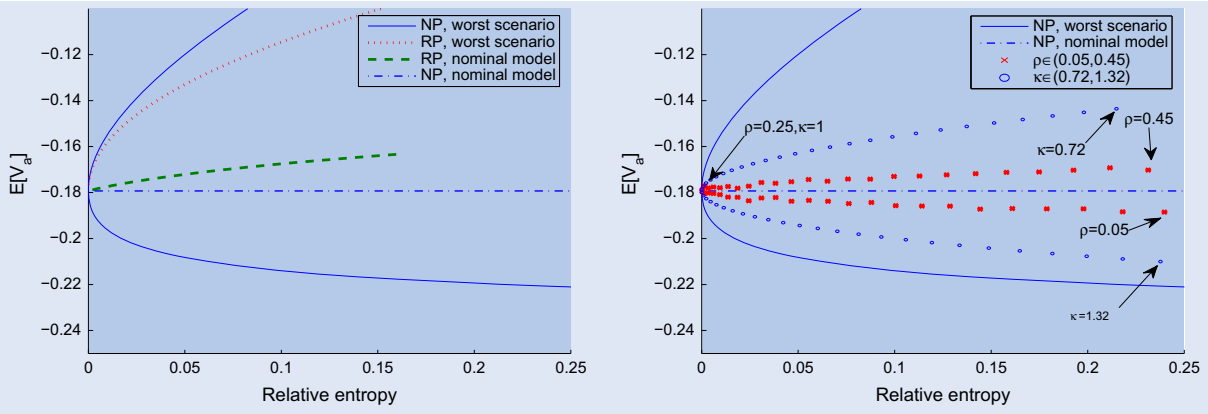



Figure 1. Expected performance vs. relative entropy. The left panels shows the performance of the nominal portfolio (NP) and the robust portfolio (RP) under the nominal and worst-case models. The right panel shows the performance of the nominal portfolio under perturbations in model parameters. Higher values on the vertical scale indicate worse performance.

Table 1. Realized and forecast variance with model uncertainty.

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

$\theta$ value, we compute the robust portfolio (RP), meaning the one that is optimal under the change of measure defined by $\theta$. In the left panel of figure 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 $\theta$). 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 paper 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 $\rho$ moves through this range. We also examine the effect of multiplying the covariance matrix of the assets by $\kappa \in (0.72, 1.32)$. 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 (forthcoming) 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.

### *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 1 January, 1990, to 31 December, 2011.

We first estimate the mean $\mu$ and covariance $\Sigma$ 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

$$a = (\gamma\Sigma)^{-1}(\mu - \lambda I) \qquad (34)$$
$$\text{where } \lambda = (I^\top(\gamma\Sigma)^{-1}\mu - 1)/(I^\top(\gamma\Sigma)^{-1}I)$$

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 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

Table 2. Worst-case portfolio variance at different levels of $\theta$ and $\alpha$. 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.

| $\theta$ | $\alpha = 2$ | $\alpha = 2.5$ | $\alpha = 2.5$, 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 |

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.

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 $\theta$. 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 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 $0.21 \times 10^{-3}$. 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 $0.10 \times 10^{-3}$ and $0.25 \times 10^{-3}$ for $\theta = 500$ and 900, respectively, and they change to $0.10 \times 10^{-3}$ and $0.36 \times 10^{-3}$ 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.

### 4.3. *The heavy-tailed case*

To illustrate the use of $\alpha$-divergence in the heavy-tailed setting, we now suppose that the vector of asset returns is given by $X \sim \mu + Z$, where $Z \sim t_\nu(\Sigma, \nu)$ 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 $\alpha$-divergence as an uncertainty measure. With a fixed portfolio weight vector $a$, Proposition 2.3 yields the worst-case likelihood ratio

$$m^*(\theta,\alpha) = (\theta(\alpha-1)V_a(X) + c(\theta,\alpha))^{\frac{1}{\alpha-1}} \quad (35)$$
$$\text{with } c(\theta,\alpha) \text{ s.t.} \quad E[m^*(\theta,\alpha)] = 1$$
$$\text{where } V_a(X) = a^\top(X-\mu)(X-\mu)^\top a.$$

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 $i, j = 1, \ldots, n$ and $i \neq j$. We use a randomly generated portfolio weight vector

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

and simulate $N = 10^7$ samples to examine the worst-case scenario. Table 2 shows the portfolio variance across various values of $\theta$ and $\alpha$, with $\theta = 0$ corresponding to the baseline nominal model. For fixed $\alpha$, increasing $\theta$ 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.

We saw in (19) that the choice of $\alpha$ influences the tail of $V(X)$ under the worst-case change of measure. A smaller $\alpha$ in table 2 yields a heavier tail, but this does not necessarily imply a larger portfolio variance. To contrast the role of $\alpha$ with $\theta$, we can think of choosing $\alpha$ based on an assessment of how heavy the tail might be and then varying $\theta$ 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 $\alpha$ parameter, we examine the tail of the portfolio excess return, $r = a^\top X - \mu$. Figure 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 $f_{|r|}(x)$, is asymptotically linear

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

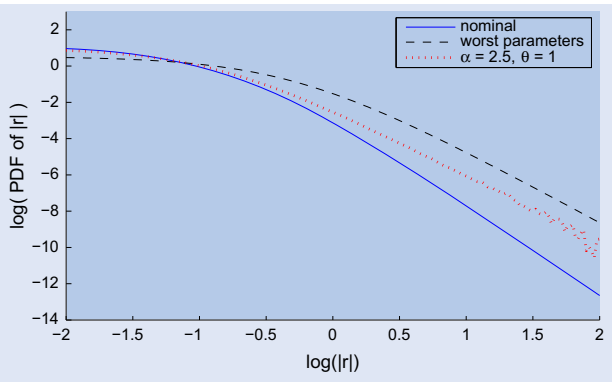



Figure 2. Tail density of absolute returns $|r|$.

Table 3. Difference of slopes $k_{\theta,\alpha} - k_{no}$ of the worst-case and nominal densities, as in figure 2.

| | $\alpha = 2$ | $\alpha = 2.5$ | $\alpha = 3$ | $\alpha = 3.5$ |
|---|---|---|---|---|
| $2/(\alpha - 1)$ | 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 |

Using the fact that

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

we find (as in (19)) that

$$\log\left(\tilde{f}_{|r|}(x)\right) - \left(\log f_{|r|}(x)\right) \approx \frac{2}{\alpha-1}\log x, \text{ for } x \gg 0 \quad (36)$$

where $\tilde{f}_{|r|}$ is the density of $|r|$ under the change of measure. This suggests that the difference of the slopes in figure 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 $\nu - 2/(\alpha-1)$.

We fit linear functions to the curves in figure 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 3 lists the differences $k_{\theta,\alpha} - k_{no}$; as we increase $\theta$, the difference of slopes gets closer to the limit $2/(\alpha-1)$ in (36).

By reweighting 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

$$\nu_{\alpha,\theta} = \nu + k_{no} - k_{\theta,\alpha} \quad (37)$$

and estimate the covariance matrix as

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

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, we show the estimated $\nu_{\alpha,\theta}$ using (37) 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 (37) in sampling under the worst-case parameter values and in figure 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 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 4, we see that the $\alpha$-divergence under the approximating $t$ is much larger. Thus, the adversary has economized the use of $\alpha$-divergence to magnify the portfolio variance without making the tail heavier than necessary.

Table 4. Comparison of $\alpha$-divergence using the worst-case change of measure and the approximating $t$ distribution from the worst case.

| $\theta$ | $\alpha = 2.5$ | 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 |

## 5. Conditional value at risk

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

$$CVaR_\beta = E[X|X > VaR_\beta],$$
$$\text{where } VaR_\beta \text{ satisfies } 1-\beta = P(X > VaR_\beta).$$

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

$$\min_a \frac{1}{1-\beta}E[(X-a)_+] + a, \quad (38)$$

for which the optimal $a$ is $VaR_\beta$.

To put this problem in our general framework, we set $V_a(X) = (1-\beta)^{-1}(X-a)_+ + a$. 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 $\alpha$-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 $\alpha$-divergence.

### 5.1. *Relative entropy uncertainty*

Suppose $X$ follows a double exponential distribution $DE(\mu, b)$ with location parameter $\mu$ and scale parameter $b$, meaning that

its density function is

$$f(x) \propto \exp\left(-\frac{|x-\mu|}{b}\right).$$

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

$$\tilde{f}(x) = m^*_{\theta,a}(x) f(x) \propto \exp\left(-\frac{|x-\mu|}{b} + \frac{\theta}{1-\beta}(x-a)_+\right).$$

The values of $a$ and $\beta$ are connected by $P(X > a) = 1-\beta$ under the nominal distribution. Because $\theta/(1-\beta) > 0$, we need $1/b > \theta/(1-\beta)$ to ensure this density function is well defined. The exponent is a piecewise linear function of the argument $x$, so $\tilde{f}$ 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 (8):

$$E[\exp(\theta V_a(X))] = \begin{cases} \exp(\theta a)\left[1 + \frac{1}{2}\left(\frac{1}{1-\frac{\theta b}{1-\beta}} - 1\right)\exp\left(\frac{\mu-a}{b}\right)\right], & \text{if } a > \mu \\ \frac{1}{2}\exp(\theta a)\left[\left(\frac{1}{1-\frac{\theta b}{1-\beta}} + \frac{1}{1+\frac{\theta b}{1-\beta}}\right)\exp\left(\frac{\theta}{1-\beta}(\mu-a)\right)\right. & (39) \\ \quad \left. + \left(1 - \frac{1}{1+\frac{\theta b}{1-\beta}}\right)\exp\left(\frac{a-\mu}{b}\right)\right], & \text{else.} \end{cases}$$

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

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

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

$$\frac{d}{da}E\big[\exp(\theta V_a(X))\big]\Big|_{a=\mu} = \frac{\theta}{2}\left(2 + \frac{\theta b - 1}{1-\beta-\theta b}\right)\exp(\theta\mu).$$

This is positive provided $\beta > 1/2$, so we can solve the first order condition for $a > \mu$ to get

$$a^*(\theta) = \mu - b\log\left(\frac{2(1-\beta-\theta b)}{1-\theta b}\right),$$

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

The VaR for the nominal model is

$$VaR_\beta = \mu - b\log(2(1-\beta)).$$

and the nominal CVaR is

$$CVaR_\beta = VaR_\beta + b = \mu - b\log(2(1-\beta)) + b.$$

Under the worst-case change of measure at parameter $\theta$, the CVaR becomes

$$CVaR_{\beta,\theta} = a^*(\theta) + \frac{1}{\frac{1}{b} + \frac{\theta}{1-\beta}}.$$

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

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

Figure 4 shows the nominal and worst-case densities starting from a nominal density that is $DE(0,1)$, using $\beta = 95\%$ 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 $\theta$.

### 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 (39) and following do not apply. In this case, we need to use $\alpha$-divergence as the uncertainty measure. With $\alpha > 1$, $\theta > 0$ and $a$ fixed, the worst case likelihood ratio now becomes

$$m^*_{\theta,a}(X) = (\theta(\alpha-1)V_a(X) + c(\theta,\alpha,a))^{\frac{1}{\alpha-1}}, \qquad (40)$$

for some constant $c(\theta,\alpha,a)$ satisfying (15) and (16).

If the density function of $X$ under the nominal distribution is regularly varying with index $\rho$, i.e. $\lim_{x\to\infty} f(tx)/f(x) = t^\rho$ for any $t > 0$ and some index $\rho < 0$, then under the worst-case change of measure it is regularly varying with index $\rho + 1/(\alpha-1)$, as suggested by (19). We require $\rho + 1/(\alpha-1) < 0$ 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

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

or a generalized extreme value distribution with density

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

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

Figure 3 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 4, 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.

A recurring and inevitable question in incorporating robustness into risk measurement is how much uncertainty to allow—in other words, where to set $\theta$ or $\alpha$. If the distribution of $X$ is

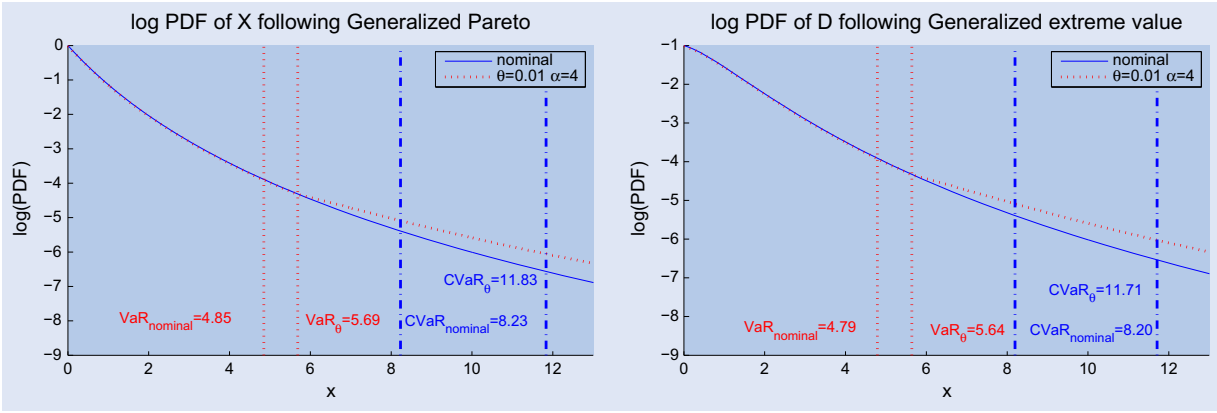



Figure 3. Density of $X$. The nominal distribution is generalized Pareto (left) or generalized extreme value (right), with parameters $b_{gp} = 1$ (scale), $\xi_{gp} = 0.3$ (shape), and $\xi_{gev} = 0.3$ (shape). Other parameters are $\theta = 0.01$, $\alpha = 4$, and $\beta = 95\%$.

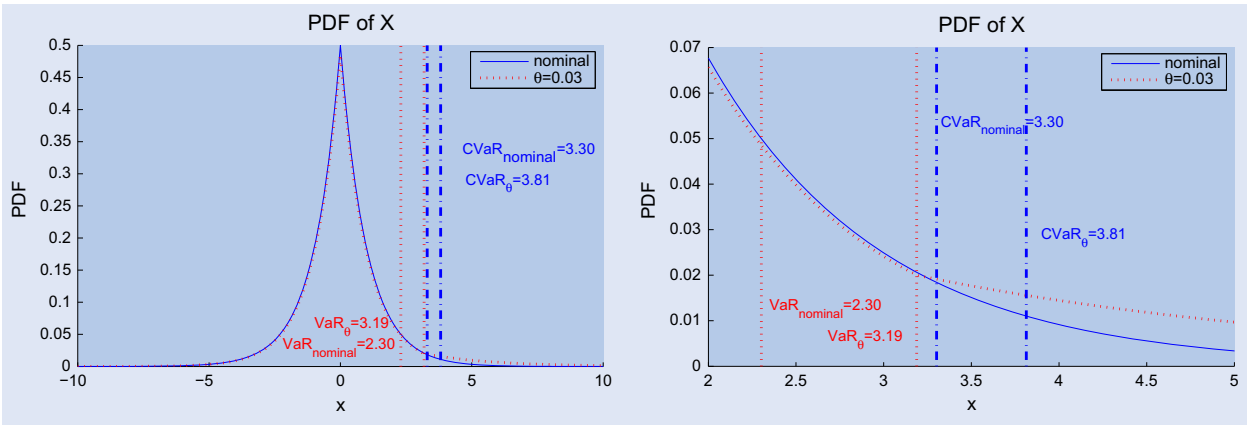



Figure 4. The dotted red line shows the worst-case density, with $\beta = 95\%$ and $\theta = 0.03$, relative to a $DE(0, 1)$ nominal density (the solid blue line). The right panel gives a magnified view of the right tail.



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 4.2. Also, the Monte Carlo approach presented in Section 3 illustrates how auxiliary quantities (for example, moments of $X$) can be calculated under the worst-case change of measure to gauge its plausibility.

## 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.

### 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 $Y_i$ denote the default indicator for $i$th obligor, meaning that

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

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

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

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

In the Gaussian copula model, each default indicator $Y_i$ is represented through the indicator of an event $\{X_i > x_i\}$, where $X_i$ has a standard normal distribution, and the threshold $x_i$ is chosen so that $P(Y_i = 1) = P(X_i > x_i) = p_i$, for a given default probability $p_i$. Dependence between default indicators is introduced through correlations between the $X_i$.

For simplicity, we focus on a single-factor homogeneous model in which the $X_i$ are given by

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

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 $\epsilon_i$ is an idiosyncratic risk associated with the $i$th obligor only. We have $n = 100$ obligors, each with a 1% default probability $p_i$, so $x_i = 2.33$. The loss given default is $c_i \equiv 1$ for all $i = 1, \ldots, n$.

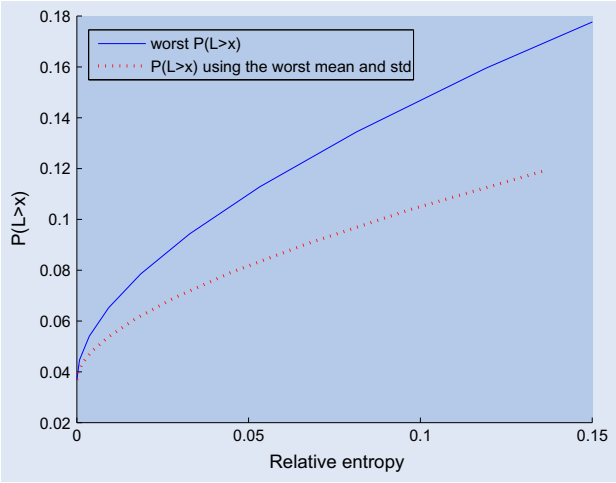



Figure 5. Loss probability as a function relative entropy. The solid blue line shows results under the worst-case change of measure. The dotted red line shows results using parameter values estimated from the worst-case change of measure. The comparison shows that the vulnerability to model error goes well beyond errors in parameters.

### 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 $P(L > x)$. 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 $P(L > x)$ means taking $V(Z, \epsilon_1, \ldots, \epsilon_n) = I_{L>x}$, so the worst-case change of measure at parameter $\theta$ is

$$m_\theta^* \propto \exp(\theta I_{L>x}).$$
$$\Rightarrow \tilde{P}(L \in dl) = \begin{cases} \frac{\exp(\theta)}{C} P(L \in dl) & \text{if } l > x; \\ \frac{1}{C} P(L \in dl) & \text{otherwise.} \end{cases} \quad (41)$$

Here, $C > 1$ is a normalization constant. This change of measure lifts the probabilities of losses greater than $x$ and lowers the probability of all other scenarios. Equivalently, we can say that the probability of any outcome of the default indicators $(Y_1, \ldots, Y_n)$ is increased by $\exp(\theta)/C$ if it yields a loss greater than $x$ and is lowered by a factor of $C$ otherwise.

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

Figure 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 (41). 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 (41); 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.

A comparison of the lines in figure 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 5 reports parameter estimates obtained under the worst-case model at two values of $\theta$. They indicate, in particular, that the parameters of the $\epsilon_i$ are affected very little by the change in distribution. Indeed, with 95% confidence, Jarque-Bera and Anderson-Darling test reject normality of $Z$ at $\theta \geq 1$ but fail to reject normality of the $\epsilon_i$ 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.

To gain further insight into the worst-case change of distribution, we examine contour plots in figure 6 of the joint density function of $\epsilon_{100}$ and $Z$. The joint density function is derived by using the original joint density function and the likelihood ratio $m_\theta^*$. The leftmost figure shows $\theta = 0.5$, and the next two correspond to $\theta = 2$. The increase in $\theta$ shifts probability mass of $Z$ to the right but leaves the joint distribution of the $\epsilon_i$ 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 (41). In the middle panel of figure 6, we see a slight asymmetry in the upper right corner, reflecting the fact that defaults are more likely when both the $\epsilon_i$ and $Z$ are increased. The left panel of figure 7 shows contours of the joint density of $(X_{99}, X_{100})$ 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 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 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

Table 5. Statistics of $\epsilon_j$ and $Z$ under the worst-case change of measure.

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

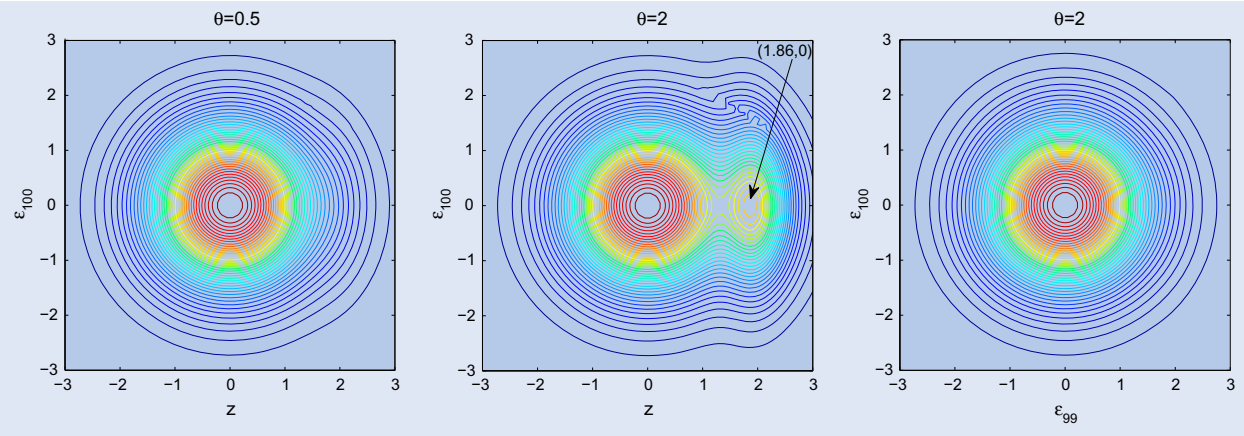



Figure 6. Contours of joint densities of $(Z, \epsilon_{100})$ with $\theta = 0.5$ (left) and $\theta = 2$ (middle), and joint density of $(\epsilon_{99}, \epsilon_{100})$ at $\theta = 2$ (right).

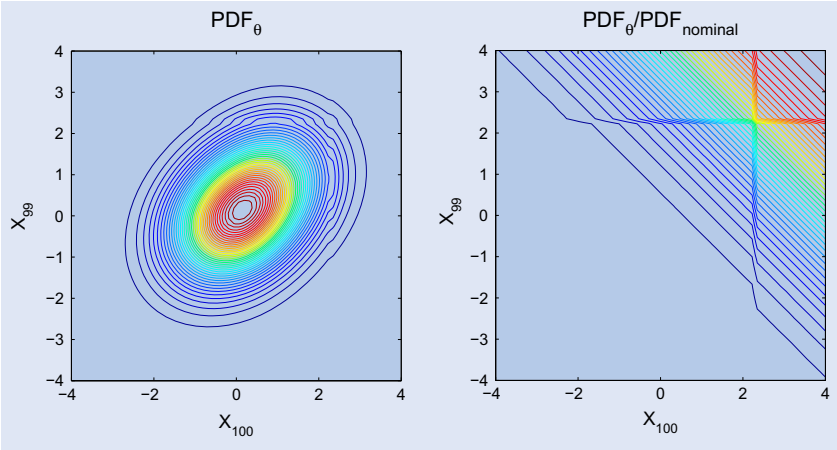



Figure 7. Contours of the joint density of $(X_{99}, X_{100})$ under the worst scenario $\theta = 2$ (left), and the ratio of the worst-case joint density to the nominal density (right).

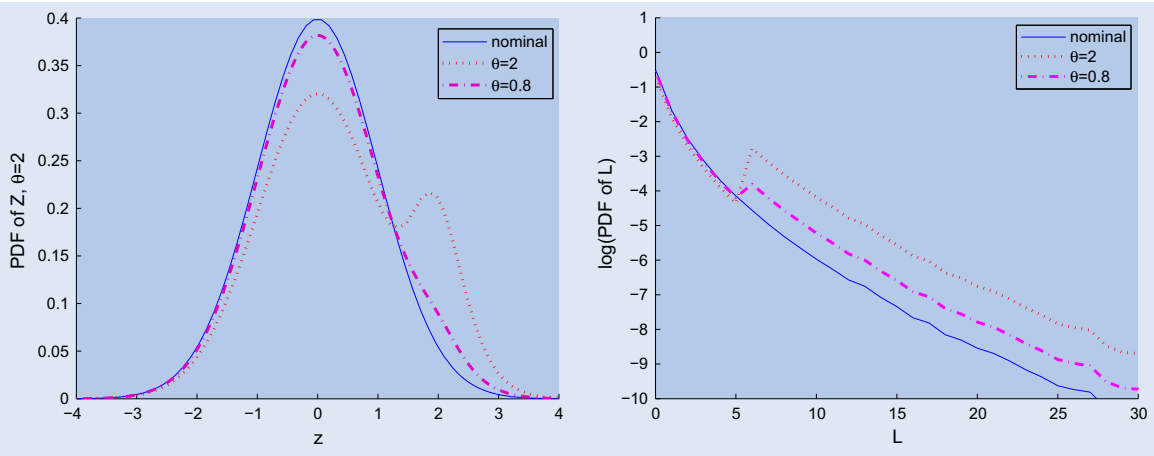



Figure 8. Marginal density of $Z$ and $L$ under worst scenario with $\theta = 0.8$ and $\theta = 2$ vs. nominal model.

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 6. Figure 9 compares the marginal distribution of $Z$ 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 $P(L > x)$ decreases as we add constraints.

Table 6. Default probability for unconstrained and constrained cases. The values of $\theta$ for the constrained cases are chosen to keep the relative entropy fixed across all three cases.

| | $P(L > x)$ |
|---|---|
| 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 |

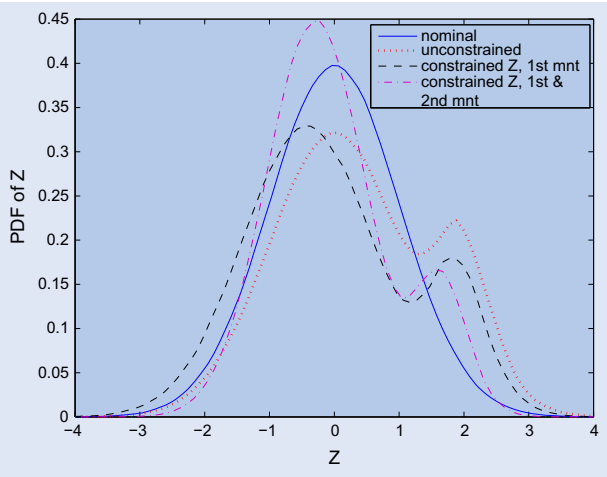



Figure 9. Density of $Z$ under the nominal, unconstrained worst-case and constrained worst-case measures.

We can further restrict the marginal distribution of $Z$ through the method of Section 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 $\epsilon$ for each realization of $Z$, we report the resulting probability in the last row of table 6, taking $\theta = 4$ to make the 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 first and second moments.

Figure 10 shows the distribution of $\epsilon$ 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 $\epsilon_{99}$ or $\epsilon_{100}$. The worst-case marginal density of $\epsilon_{100}$ changes in a way similar to the marginal density of $Z$ in figures 8 and 9.

## 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 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.

### 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 $\mu_n$. 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 $t$. At time 0, the portfolio's cash and stock values 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

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

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

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

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

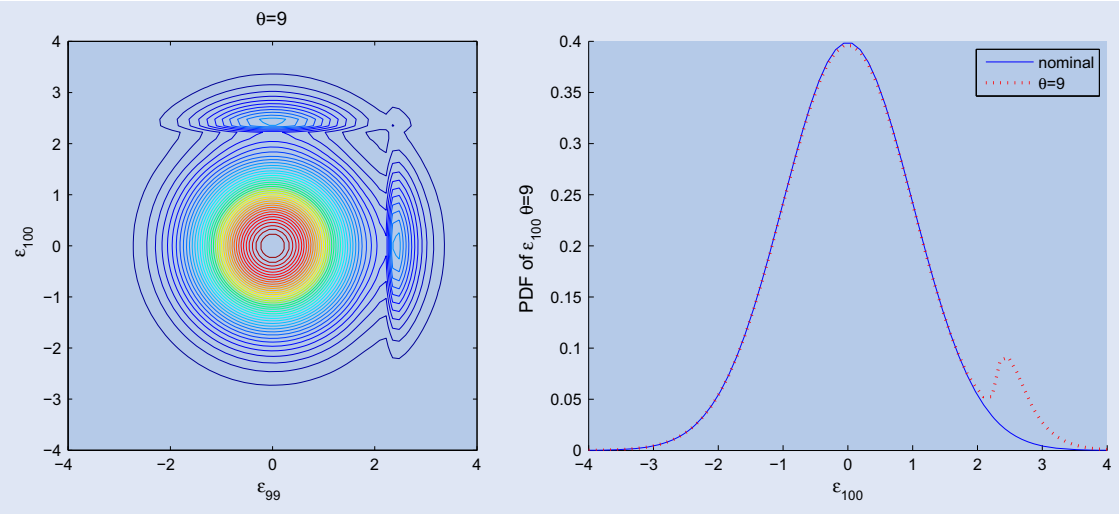



Figure 10. The marginal distribution of $Z$ is fixed. The left figure is the joint density of $(\epsilon_{99}, \epsilon_{100})$ under the worst scenario, and the right figure is the marginal density of $\epsilon_{100}$ under the worst scenario. Both figures have $\theta = 9$.



hedging error under a jump-diffusion model for the underlying asset. Bertsimas *et al.* (2000) analyse 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 $\sigma_n$. Under our formulation of robustness with discrete delta hedging, we can calculate a robust value of this input $\sigma_n$ in the sense of minimizing the maximum value of the hedging error $E[|H_e|]$ at given value of $\theta$. The result is illustrated in figure 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 $\sigma_n$ 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 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.* (2011), among others, also find that Black–Scholes delta hedging performs surprsingly well, even when its underlying assumptions are not satisfied.

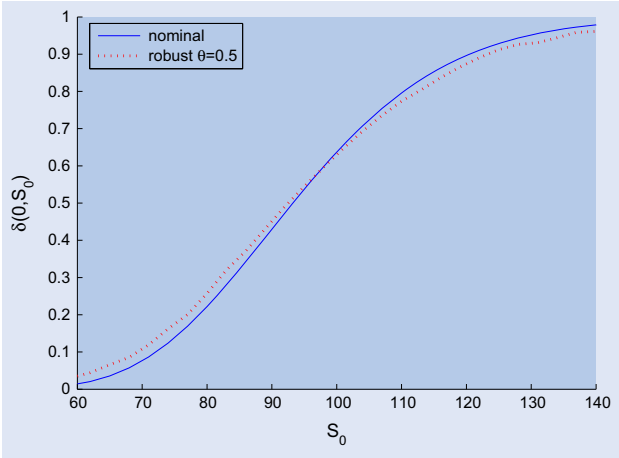



Figure 11. Optimal delta vs. $S_0$ with $\theta = 0.5$.

### *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 (9), we can estimate the optimal likelihood ratio $m^*_\theta$ 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 12, where we can see that the change of measure makes the right tail heavier. In figure 12, the tail

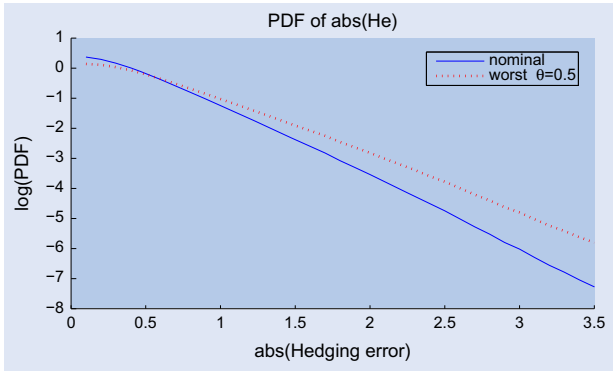



Figure 12. Density of absolute hedging error under nominal and worst scenario, with $\theta = 0.5$.

is fit through a non-parametric method, using the 'ksdensity' command in MATLAB with a normal kernel and bandwidth 0.1.

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

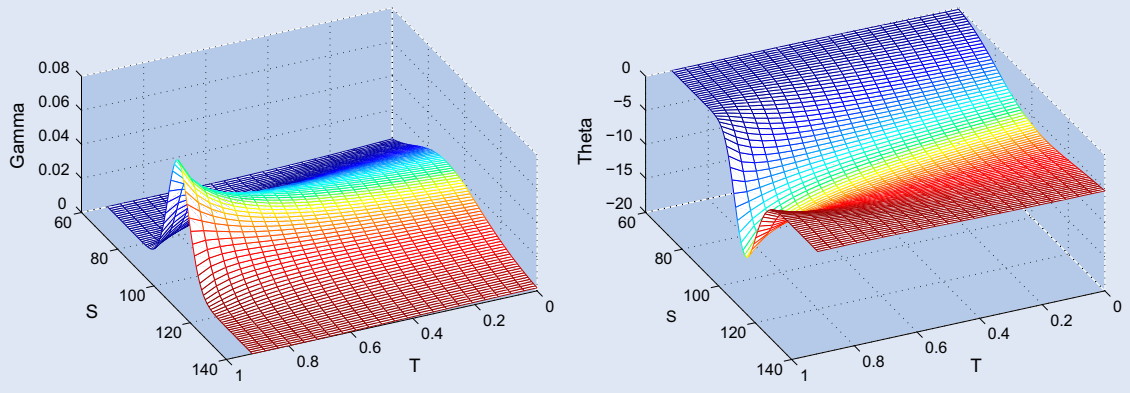



Figure 13. Gamma and Theta for European call option.

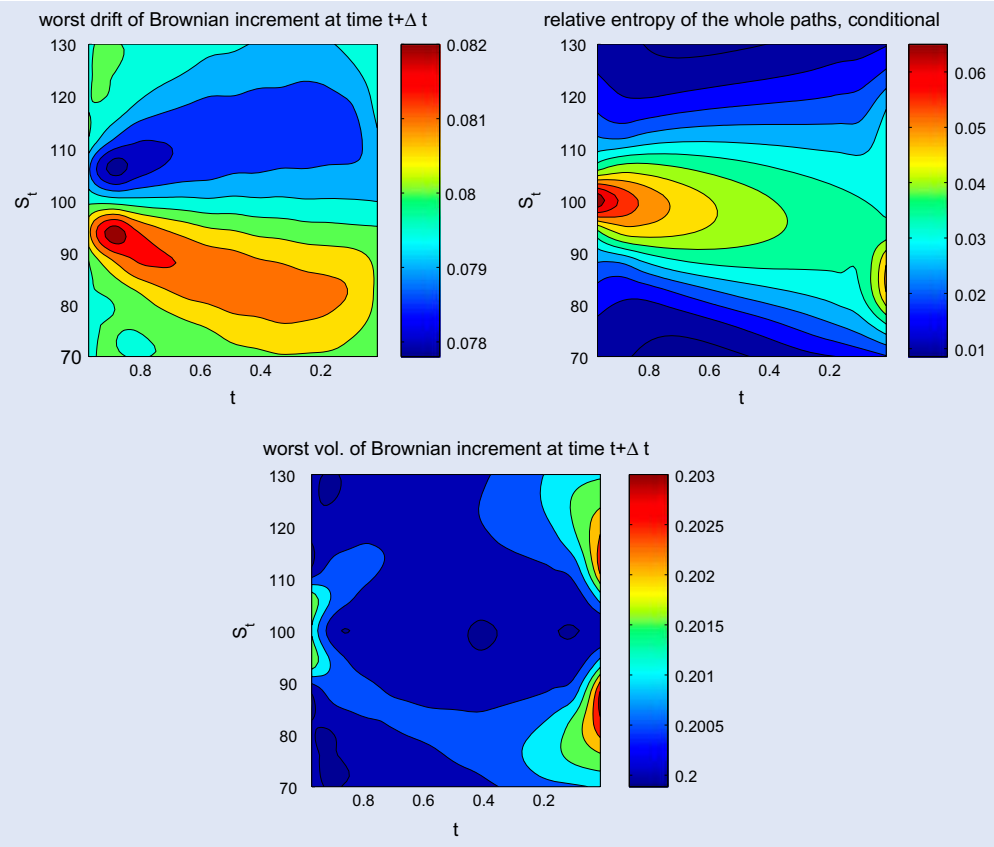



Figure 14. Conditional on $S_t$, the worst-case drift (upper left), relative entropy (upper right), and worst-case volatility (lower left), all using $\theta = 0.5$.

reaching points $(t, S_t)$. For every $t = T(2 + 8k/N_T)$, for $k = 1, \ldots, 12$, and every $S_t = 70 + 6l$ for $l = 1, \ldots, 10$, 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 $i$th simulated path, then the conditional likelihood given $S_t = x$ is

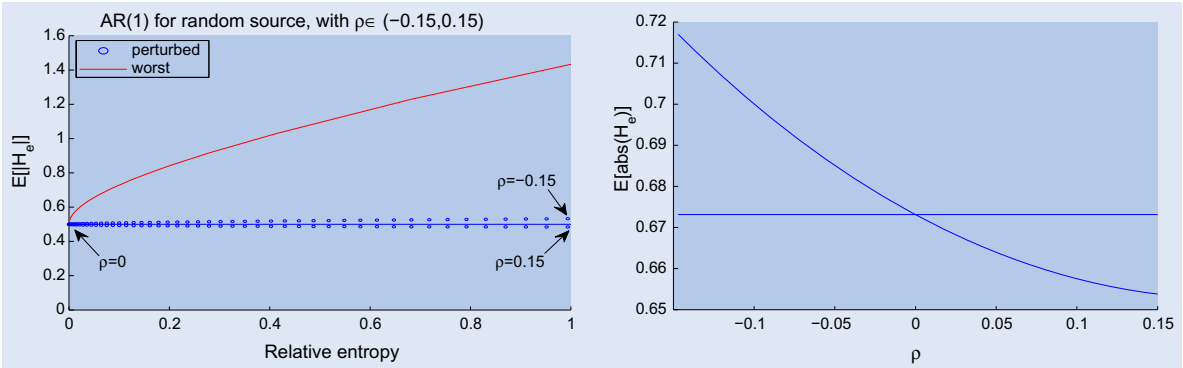
$$\begin{aligned} m^*(\texttt{path}_i | S_t = x) &= \frac{\tilde{f}(\texttt{path}_i | S_t = x)}{f(\texttt{path}_i | S_t = x)} \\ &= \frac{\tilde{f}(\texttt{path}_i)}{f(\texttt{path}_i)} \frac{f(S_t \in (x, x + dx))}{\tilde{f}(S_t \in (x, x + dx))} \\ &\propto \frac{\tilde{f}(\texttt{path}_i)}{f(\texttt{path}_i)}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

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

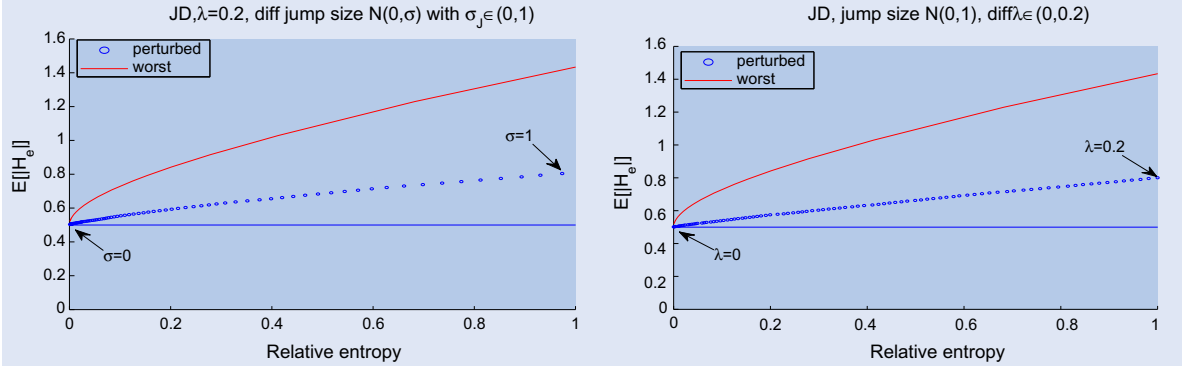
$$m^*(\texttt{path}_i | S_t = x) = \frac{m^*(\texttt{path}_i)}{\sum_{j=1}^{N} m^*(\texttt{path}_j)/N}$$

across the $N$ simulated paths.

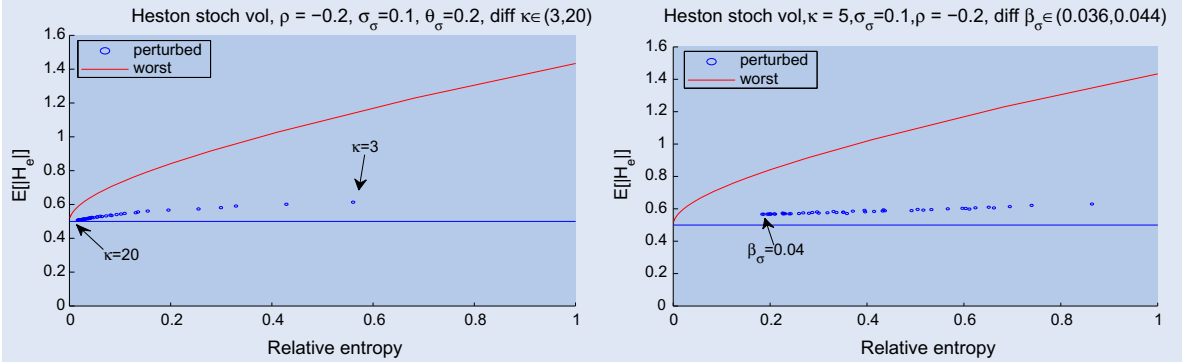
As a point of comparison for the simulation results, it is useful to consider potential sources of hedging error. With discrete rebalancing, we would expect a large move in the underlying asset to produce a large hedging error. Figure 13 plots the option gamma and the time-decay theta, and these suggest that the hedging error is particularly vulnerable close to maturity when the underlying is near the strike. (Time in



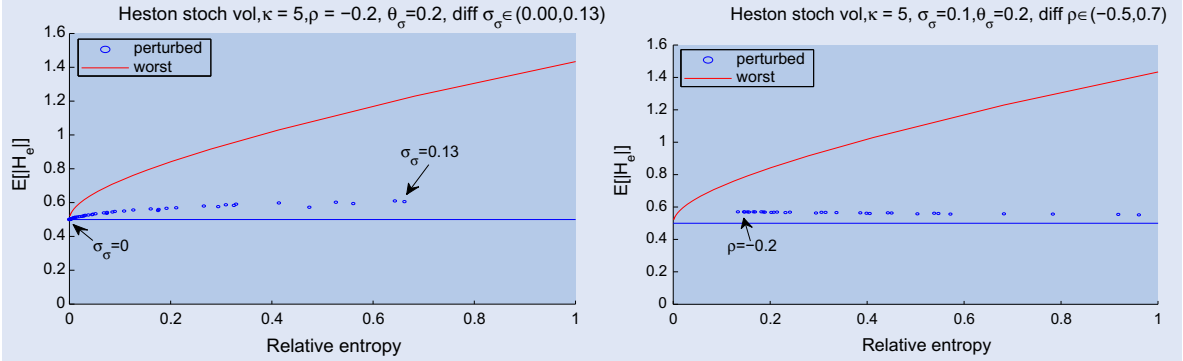

(a) AR(1) perturbation (left) and the effect of varying $\rho$ (right).




(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 $\kappa$ (left) and different long term limit $\beta_\sigma$ (right).




(d) Heston stochastic volatility perturbation with different volatility of volatility $\sigma_\sigma$ (left) and correlation of diffusion $\rho$ (right).

Figure 15. Hedging error under various changes in the underlying dynamics.

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 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 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 centred 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 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 14, we used kernel smoothing. The smoothed value at $(s, t)$ is a weighted average of results at $(s_i, t_i)$, $i = 1, \ldots, n$, using a kernel $K(.,.) > 0$,

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

In particular, we used $K((s',t'),(s,t)) = \phi(\|(s',t') - (s,t)\|/a)$, with $\phi$ 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 neighbouring nodes $(s,t)$ and $(s',t')$ on the grid, $\|(s,t) - (s',t')\|/a = 1$.

### 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 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 $\theta$ in $[0, 0.23]$.

In panel (a) of figure 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 $\epsilon_t$ are independent and normally distributed with mean 0 and variance $\Delta t$. 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 $\rho$ 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 15, we use Merton's jump-diffusion model,

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

where $J$ is a compound Poisson, $J_t = \sum_{i=1}^{N_t} \exp(Y_i)$, with $N_t$ a Poisson process with intensity $\lambda$, and $Y_i$ i.i.d. $N(0, \sigma_J)$, with $\sigma_J = 1$. When increasing $\sigma_J$ from 0 to 1, or the jump intensity $\lambda$ 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 15 test the Heston stochastic volatility model, in which the square of volatility $v_t = \sigma^2$ follows the dynamics

$$dv_t = \kappa(\beta_\sigma - v_t)dt + \sigma_\sigma\sqrt{v_t}dW_t^v, \tag{42}$$

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

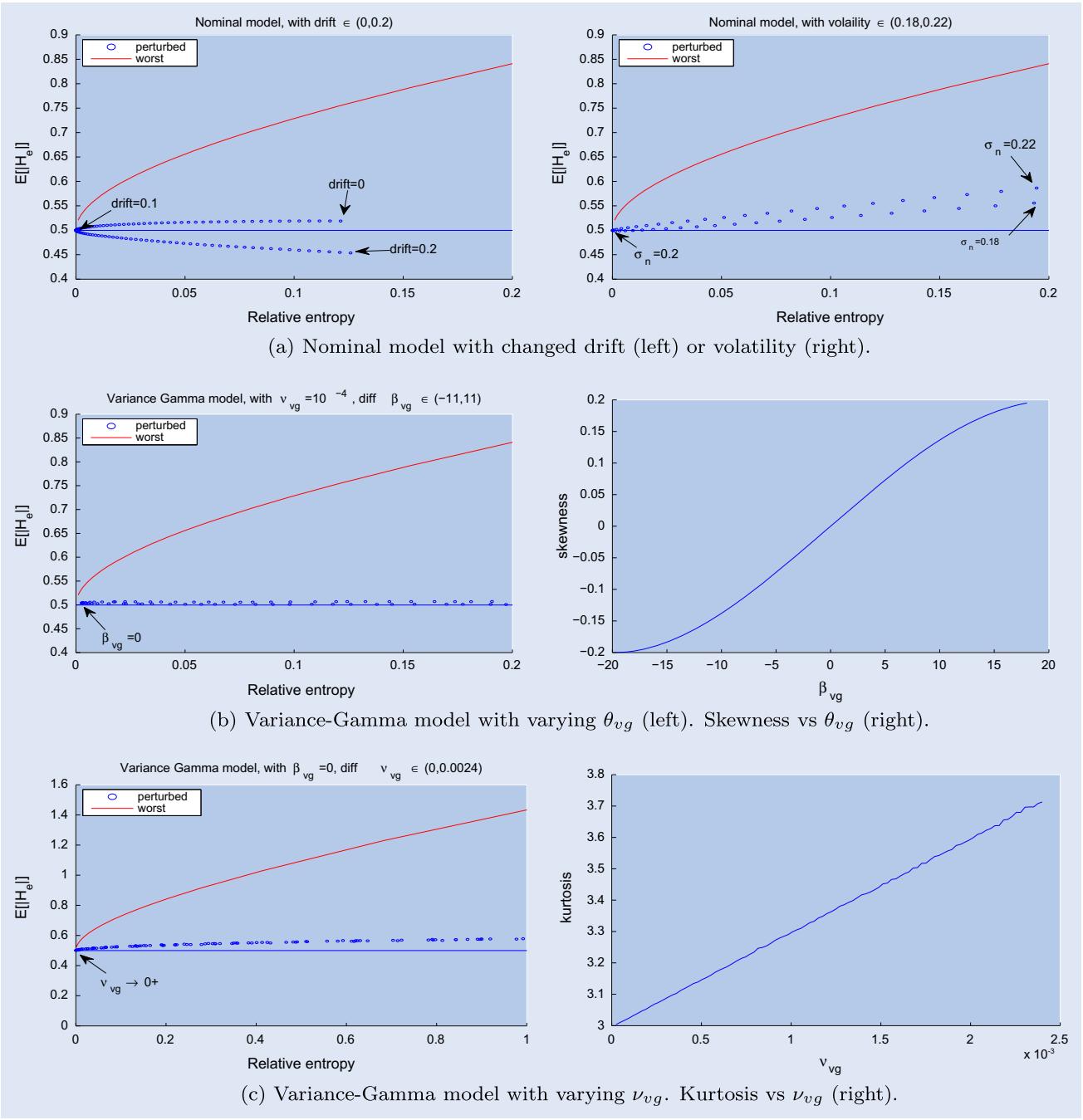
When discretized to dates $t_i = i\Delta t$, $i = 1, \ldots, N_T$, the likelihood ratio for the price process becomes

$$m(s_{t_1}, \ldots, s_{t_{N_T}}) = \frac{\tilde{f}(s_{t_1}, \ldots, s_{t_{N_T}})}{f(s_{t_1}, \ldots, s_{t_{N_T}})} = \frac{E_v[\tilde{f}(s_{t_1}, \ldots, s_{t_{N_T}}|v_{t_1}, \ldots, v_{t_N})]}{f(s_{t_1}, \ldots, s_{t_{N_T}})}$$

where $f$ and $\tilde{f}$ are the joint density functions of prices under the nominal and Heston models, respectively. In the second equality, $\tilde{f}(.|.)$ 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 $\kappa$ changes from 3 to 20, the relative entropy and the expected hedging error decrease. As $\kappa$ becomes larger, the expected hedging error gets closer to the nominal value, while relative entropy appears to converge to some positive value. With a large $\kappa$, any deviation from the nominal variance decays quickly, leaving only a short-term deviation introduced by the diffusion term of (42).

As the long-run limit $\beta_\sigma$ 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 $\beta_\sigma < 0.04$ and $\beta_\sigma > 0.04$. As the volatility of volatility $\sigma_\sigma$ varies from 0 to 0.13, both relative entropy and expected hedging error increase. As $\sigma_\sigma$ gets closer to zero, the volatility behaves more like a constant, which is the nominal model. And, as the correlation $\rho$ 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 $\rho$ is close to nominal value $-0.2$. The relative entropy reaches the minimum value when $\rho$ equals the nominal value $-0.2$.




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

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

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

Figure 16. Hedging errors under various changes in the underlying dynamics.

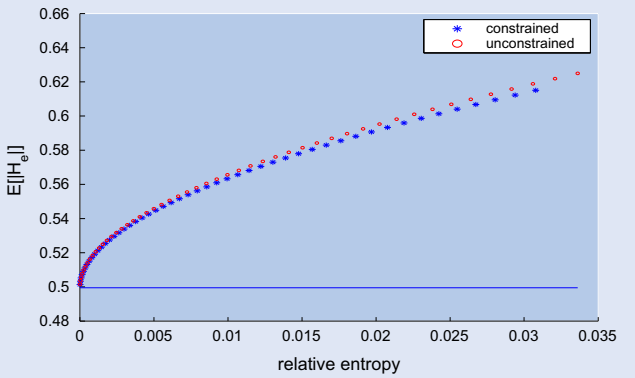
For our last comparison, we use the variance-gamma model of Madan *et al.* (1998),

$$\begin{aligned} S_t &= S_0 \exp((\mu+\omega)t + X_t) \\ \text{where } X_t &= \beta_{vg}\gamma(t;1,\nu_{vg}) + \sigma W_{\gamma(t;1,\nu_{vg})} \\ \omega &= \frac{1}{\nu_{vg}} \log\left(1 - \beta_{vg}\nu_{vg} - \frac{1}{2}\sigma^2\nu_{vg}\right) \end{aligned}$$

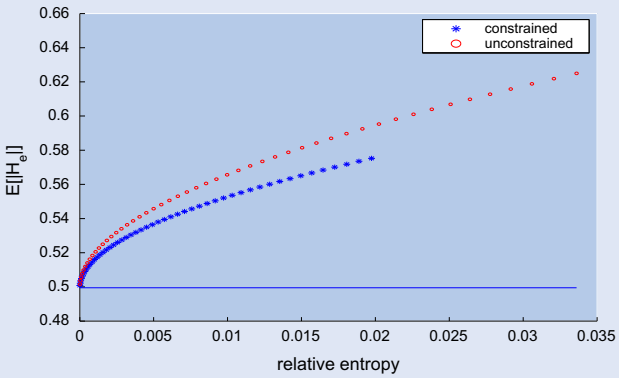
where $\gamma(t;1,\nu_{vg})$ is the gamma process with unit mean rate. Parameter $\beta_{vg}$ controls the skewness of return and $\nu_{vg}$ controls the kurtosis; see panels (b) and (c) of figure 16. The figure suggests that skewness and kurtosis have limited impact on hedging error.

It is noteworthy that in most of the examples in figures 15 and 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 $\Delta W_i$ 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 17(a) shows that this has only a minor effect on the worst-case hedging error. In figure 17(b), we constrain




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




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

Figure 17. The blue dots are for constraint cases, and the red dots are for the unconstrained case.

Table 7. Worst-case results and parameters for CVA example.

| | | $\theta$ | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Nominal $\rho$ | | $-12$ | $-9$ | $-6$ | $-3$ | 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 |
| $-0.3$ | $\rho$ | $-0.299$ | $-0.299$ | $-0.299$ | $-0.299$ | $-0.299$ | $-0.299$ | $-0.299$ | $-0.299$ | $-0.299$ |
| 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 |
| $-0.3$ | $\sigma_x$ | 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 |
| $-0.3$ | $\sigma_\lambda$ | 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 |

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 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 15 and 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 14 is that the greatest vulnerabilities occur close to maturity and, to a lesser extent, at inception.

## 8. Credit valuation adjustment

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

### 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 $X_t$,

$$dX_t = \kappa_x(\mu_x - X_t)dt + \sigma_x dW_t^x. \tag{43}$$

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 modelled through a stochastic default intensity $\lambda_t$, 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 $W^\lambda$ are Brownian motions with correlation $\rho$, and $J_t$ is a compound Poisson process with jump intensity $\nu_j$ 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 $\tau$ is

$$\tau = \Lambda^{-1}(\xi), \quad \text{where} \Lambda(t) = \int_0^t \lambda_s\, ds \quad \text{and} \quad \xi \sim Exp(1), \tag{44}$$

meaning that $\xi$ 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 (43) 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 $\xi$ in (44) remain unchanged: the adversary can change the dynamics of the default intensity (as well as the exposure), but having $\xi$ be a unit-mean exponential in (44) is part of what it means for $\lambda$ to be the default intensity, so this element is not subject to model error.

We enforce this condition through the method in Section 3.3. We simulate $N = 10^4$ sample paths for $X$ and $\lambda$, and use $N_\xi = 10^4$ samples of $\xi$. For each realization of $\xi$, all $N$ paths of $\lambda$ are generated using (44), yielding a total of $N \times N_\xi$ paths. (Paths of $X$ and $\lambda$ are generated using an Euler approximation.) For path $(X^i, \lambda^i)$ and given $\xi$,

$$\hat{m}^*(X^i, \lambda^i, \xi) = \frac{\exp(\theta V(X^i, \lambda^i, \xi))}{\sum_{j=1}^N \exp(\theta V(X^j, \lambda^j, \xi))/N} \tag{45}$$

$$\text{where } V(X^i, \Lambda^i, \xi) = (1 - R)e^{-r\tau} I_{\tau<T} \max(X_\tau^i, 0)$$
$$\text{and } \tau = \Lambda_i^{-1}(\xi),\ \Lambda_i(t) = \int_0^t \lambda_i(s)ds.$$

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

### 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 $\rho$ 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 $k$th moment of the increments of $W_x$, we use

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

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 7. The columns show estimates for different values of $\theta$, with $\theta = 0$ corresponding to the nominal value. Positive larger $\theta$ corresponds to possible larger CVA in the worst-case scenario, while negative smaller $\theta$ 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 $\theta$. The impact is asymmetric, with positive $\theta$ values having a greater effect than smaller $\theta$ values, particularly at larger values of $\rho$. This is at least partly explained by controlling for differences in relative entropy $\mathcal{R}(m_\theta)$.

The estimates of $\rho$, $\sigma_x$ and $\sigma_\lambda$ 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 $\theta$, though the magnitude of change is small. In short, the wide range of CVA values are all consistent with nearly the

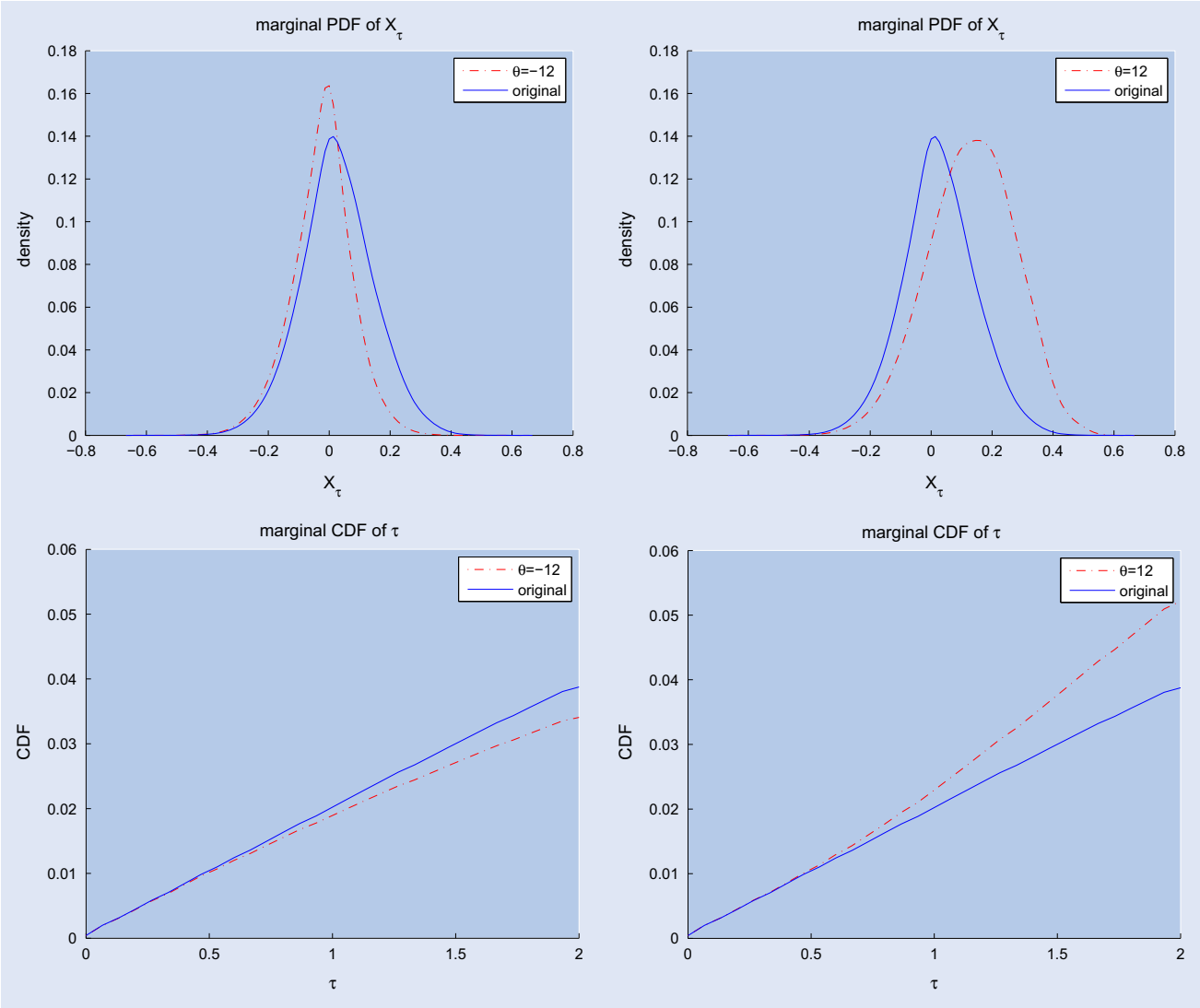



Figure 18. Marginal distribution of $X_\tau$ and $\tau$ for $\tau < T$.

Table 8. Correlations between $X_\tau$ and $\tau$, conditional on $\tau < T$.

| $\theta$ | −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 |

same parameters values; changes in parameter values are not the primary source of model risk.

Next, we consider changes in the marginal distributions of $X_\tau$ and $\tau$, considering only outcomes in which $\tau < T$. The upper panels of figure 18 show the marginal density of $X_\tau$. 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 $\tau$. Here, a larger $\theta$ value makes default more likely within the horizon (thus increasing the CVA), whereas a smaller $\theta$ makes default less likely.

The most interesting aspect of the worst-case change of measure is the effect on the dependence between $\tau$ and $X_\tau$. We can get a first indication of this dependence from the correlations estimated at different $\theta$ values reported in table 8. The correlations consistently increase with $\theta$.

To further examine the dependence, in figure 19 we plot contours of the joint density of $\tau$ and $X_\tau$ for different values of $\theta$, taking $\rho = 0.3$. Despite this correlation in the driving Brownian motions, we do not observe much dependence between $X_\tau$ and $\tau$ 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 has a chance to accumulate. In other words, the least costly way for the adversary to generate a large CVA is to push $\tau$ toward $T$ and push $X$ upward. In the lower right corner, we show the joint distribution obtained using the parameter values estimated at

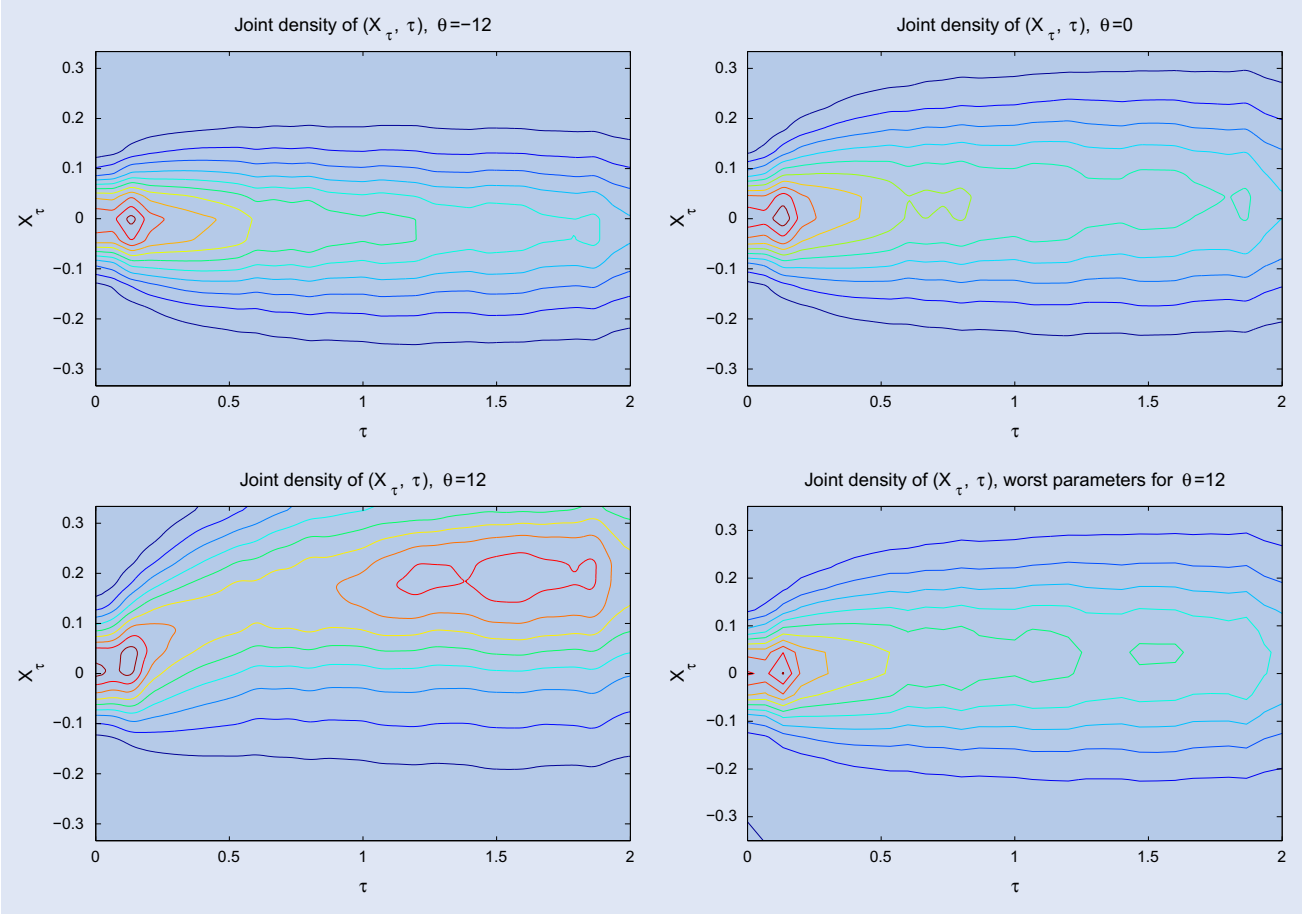



Figure 19. Joint density of $X_\tau$ and $\tau$ for $\tau < T$, $T = 2$.

Table 9. CVA and $\rho_{X_\tau,\tau}$ using parameters estimated from the worst-case scenario, the worst-case scenario, and scenarios with perturbed 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 |

$\theta = 12$. 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 $\tau$ and $X_\tau$; here, the adversary tries to generate a small CVA with a quick default near $X = 0$ or no default at all.

In figure 20, we plot contours of the joint density of $(\xi, X_\tau)$ for the same cases that appear in figure 19. These are consistent with the pattern in figure 19, but recall that the marginal distribution of $\xi$ does not change so the pattern here is more purely determined by the change in dependence. This is further illustrated in figure 21, which shows contours of the copula for $X_\tau$ and $\xi$.

In figure 22, we revisit the comparisons of figures 18–20, 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 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.

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.

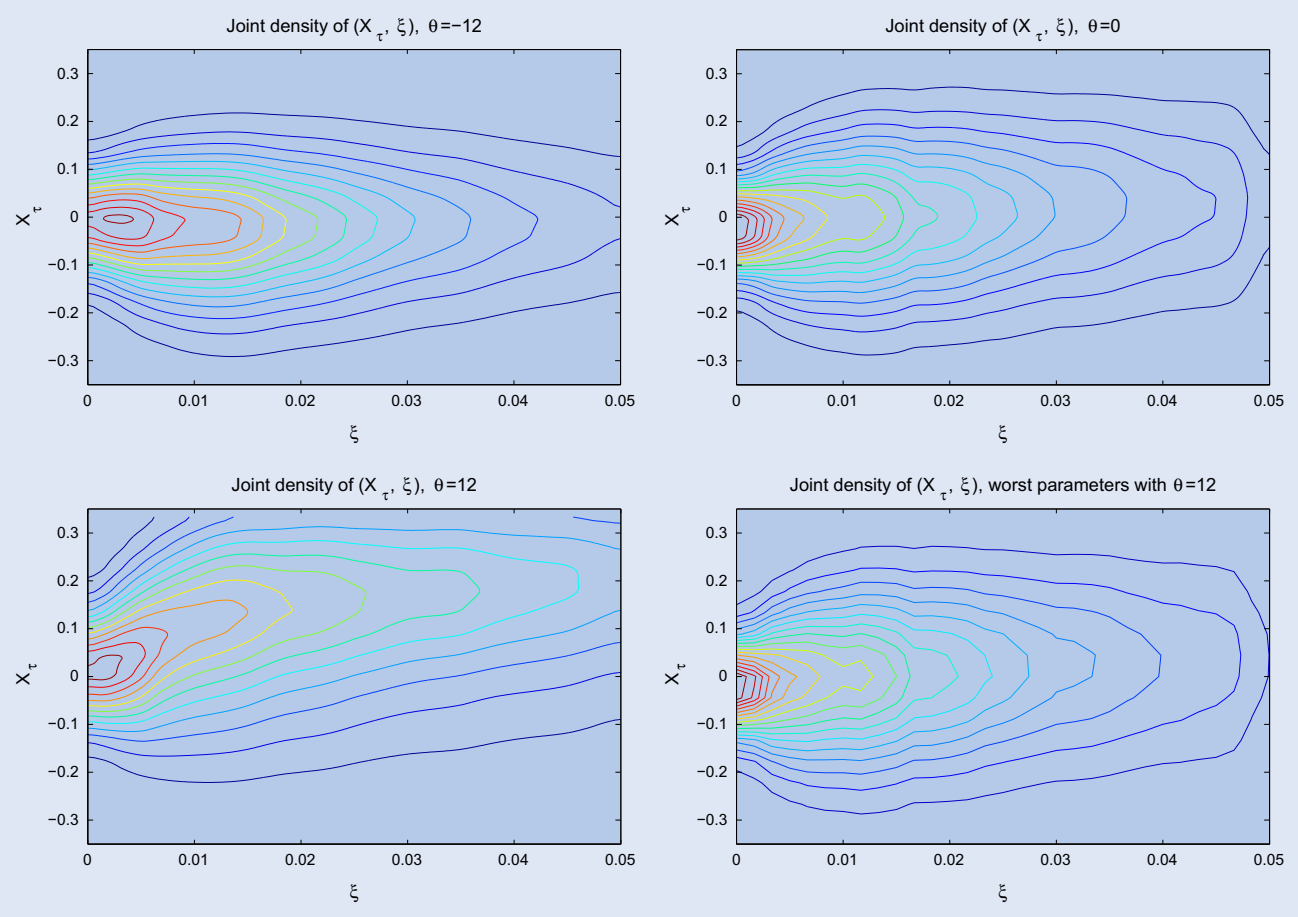



Figure 20. Joint density of $X_\tau$ and $\xi$ for $\tau < T$, $T = 2$.

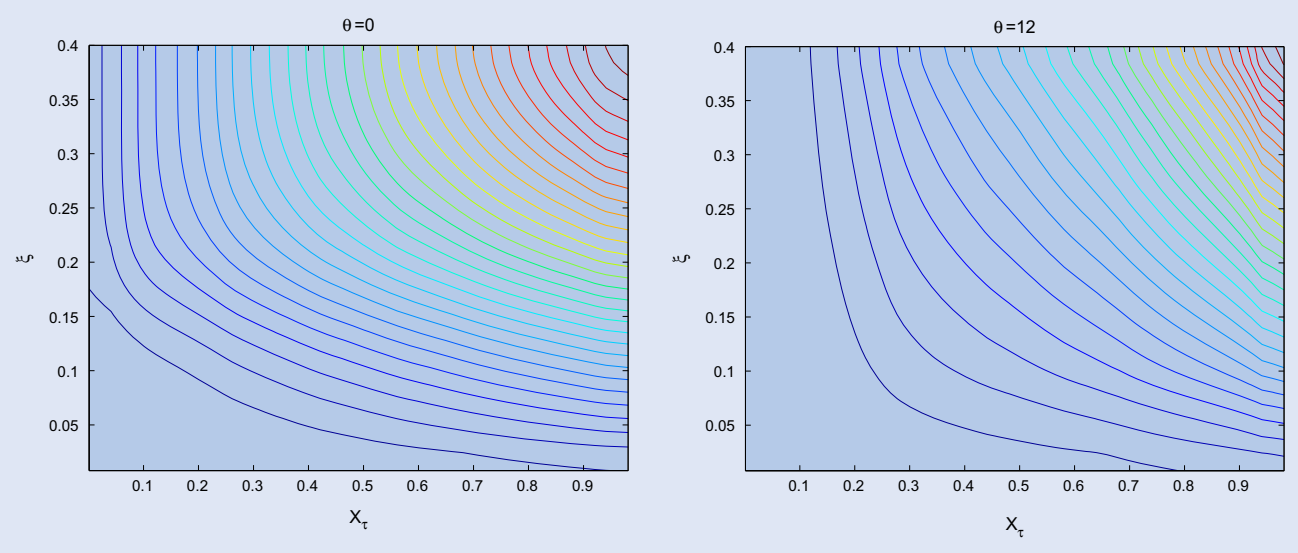



Figure 21. Copula of $(X_\tau, \xi)$.

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 $\lambda$ and $X$.

In table 9, we estimate CVA at different values of $\theta$ 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 parameters—increasing $\rho$ to 0.95 or increasing $\rho$ together with

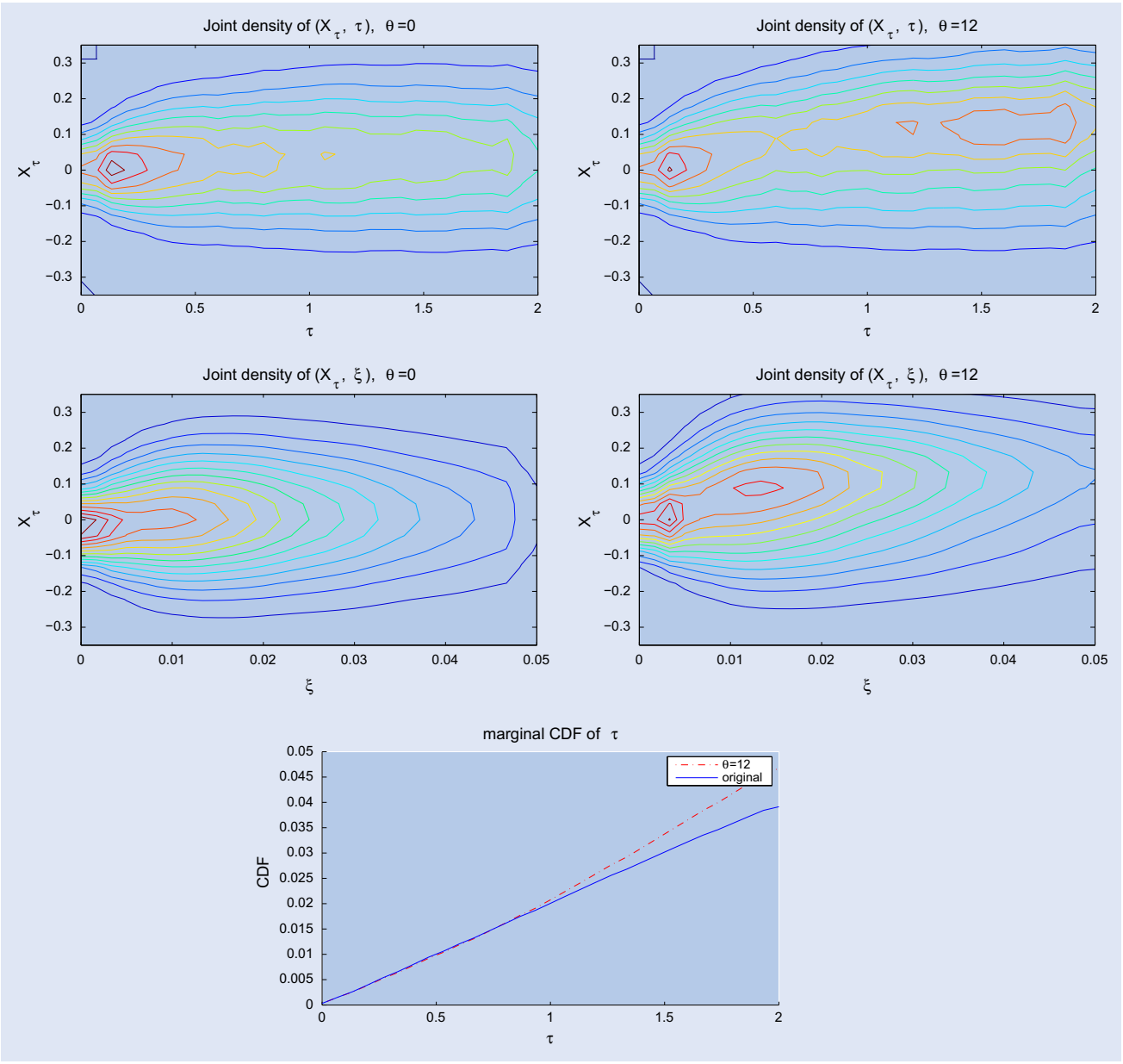



Figure 22. Marginal densities of $\tau$ and $X_\tau$(first row), joint density of $(X_\tau, \tau)$ (second row) and joint density of $(X_\tau, \xi)$ (third row). The dynamics of $X$ are fixed.

$\mu_\lambda$. Once again, we find that the worst-case model error is not simply described by a change in parameters.

By simply increasing $\rho$, 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 $\mu_\lambda$ is increased, the CVA increases, but when $\mu_\lambda \geq 0.06$, $\rho_{X_\tau,\tau}$ turns negative. A possible explanation is that as $\mu_\lambda$ becomes very large, those paths with small realizations of $W^\lambda$ also default before $T$, contributing small values of $X_\tau$, 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 $\rho = 0.8$.

Compared to figure 19, figure 22 shows much less distortion in the joint density of $(X_\tau, \tau)$ and $(X_\tau, \xi)$, 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 $\lambda$, 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 $\tau$ is similar to what it was before early in the horizon.

We have also tested the case with jumps in the dynamics of $\lambda$, 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 23, the dynamics of $W_x$ and $W_\lambda$ have very minor changes even before defaults.

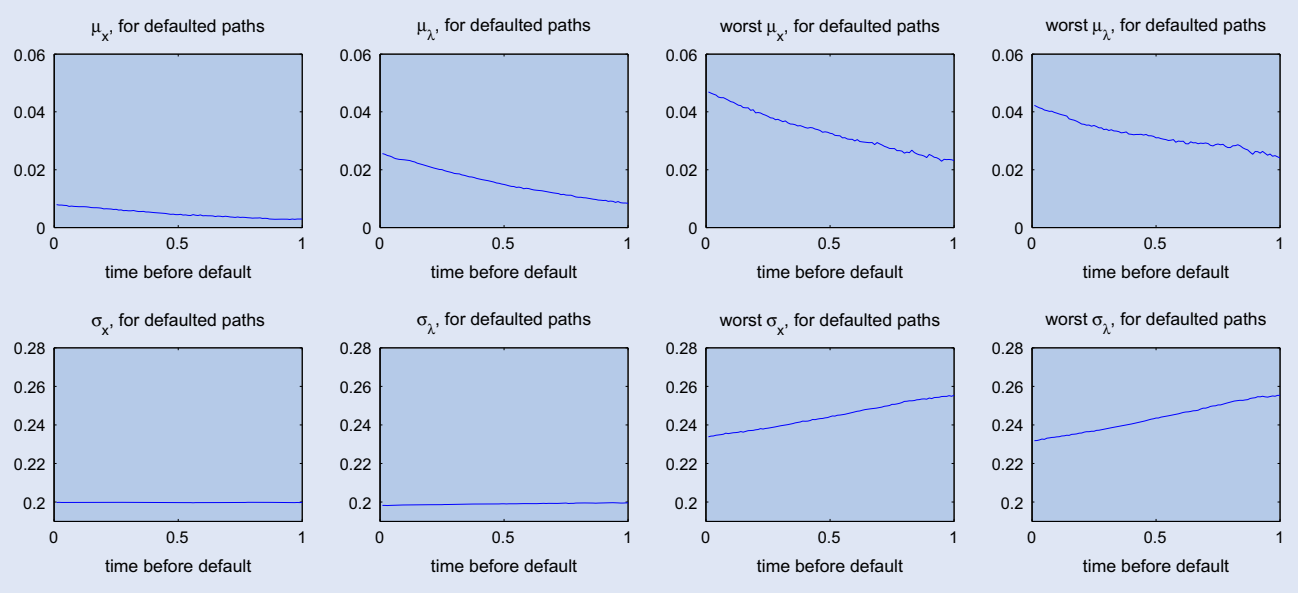



Figure 23. Statistics for increments before defaults using $\theta = 0$ (left) and $\theta = 12$ (right).

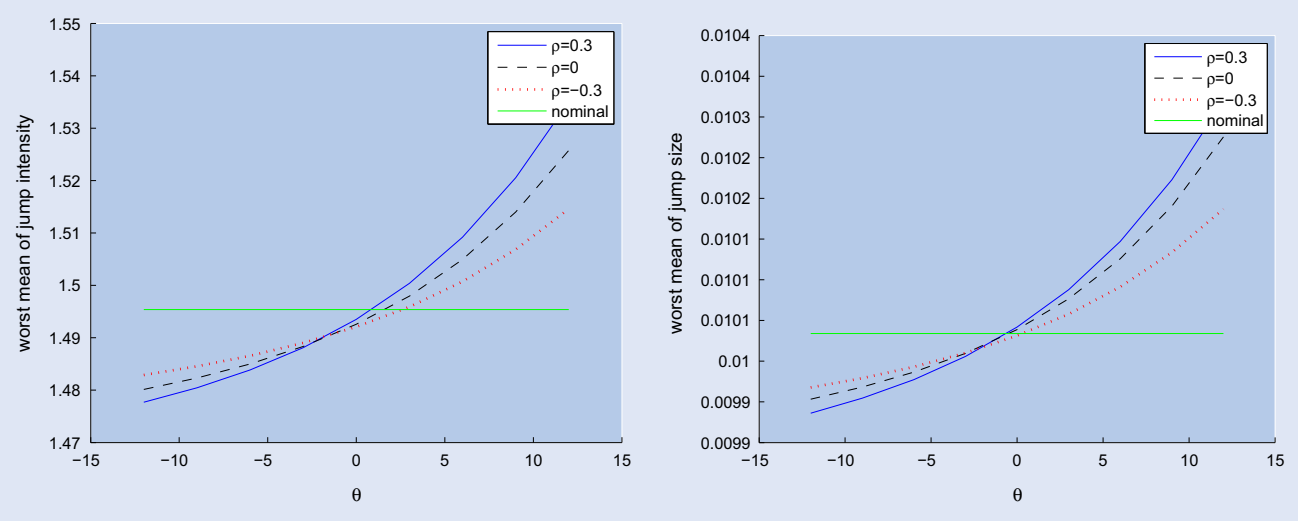



Figure 24. Worst-case jump intensity and worst-case mean jump size.

In figure 24, we plot the worst-case jump intensity and the worst-case mean jump sizes. Both increase with $\theta$, but the magnitudes of the changes are small.

## 9. Concluding remarks

This paper 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 $\alpha$-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 CVA. 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.

## Acknowledgements

We thank Xuedong He for helpful discussion. This research is supported by NSF grant DMS09-14539 and a fellowship from the Centre for Risk Studies at the University of Cambridge.

## References


Avellaneda, M., Minimum-relative-entropy calibration of asset pricing models. *Int. J. Theor. Appl. Finance*, 1998, **1**(4), 447–472.

Avellaneda, M., Buff, R., Friedman, C., Grandchamp, N., Kruk, L. and Newman, J., Weighted Monte Carlo: A new technique for calibrating asset-pricing models. *Int. J. Theor. Appl. Finance*, 2000, **4**(1), 1–29.

Avellaneda, M., Friedman, C., Holmes, R. and Samperi, D., Calibrating volatility surfaces via relative-entropy minimization. *Appl. Math. Finance*, 1997, **4**(1), 37–64.

Avellaneda, M., Levy, A. and Paras, A., Pricing and hedging derivative securities in markets with uncertain volatilities. *Appl. Math. Finance*, 1995, **2**(2), 73–88.

Avramov, D., Stock return predictability and model uncertainty. *J. Financ. Econ.*, 2002, **64**(3), 423–458.

Ben-Tal, A., Margalit, T. and Nemirovski, A., Robust modeling of multi-stage portfolio problems. In *High Performance Optimization*, edited by H. Frenk, K. Roos, T. Terlaky and S. Zhang, pp. 303–328, 2000 (Kluwer Academic: Dordrecht).

Bertsimas, D., Kogan, L. and Lo, A.W., When is time continuous? *J. Financ. Econ.*, 2000, **55**(2), 173–204.

Bertsimas, D. and Pachamanova, D., Robust multiperiod portfolio management in the presence of transaction costs. *Comput. Oper. Res.*, 2008, **35**(1), 3–17.

Boyarchenko, N., Cerrato, M., Crosby, J. and Hodges, S., No good deals-no bad models. FRB of New York Staff, Report 589, 2012.

Branger, N., Krautheim, E., Schlag, C. and Seeger, N., Hedging under model misspecification: All risk factors are equal, but some are more equal than others. *J. Futures Markets*, 2011, **32**(5), 397–430.

Brigo, D., Mauri, G. and Mercurio, F., Lognormal-mixture dynamics and calibration to market volatility smiles. *Int. J. Theor. Appl. Finance*, 2002, **5**(4), 427–446.

Buchen, P.W. and Kelly, M., The maximum entropy distribution of an asset inferred from option prices. *J. Financ. Quant. Anal.*, 1996, **31**(01), 143–159.

Cont, R. and Deguest, R., Equity correlations implied by index options: Estimation and model uncertainty analysis. *Math. Finance*, 2013, **23**(3), 496–530.

Cont, R. and Minca, A., Recovering portfolio default intensities implied by CDO quotes. *Math. Finance*, 2013, **23**(1), 94–121.

Cont, R. and Tankov, P., Non-parametric calibration of jump-diffusion option pricing models. *Comput. Finance*, 2004, **7**, 1–49.

Cont, R. and Tankov, P., Retrieving Lévy processes from option prices: Regularization of an ill-posed inverse problem. *SIAM J. Control Optim.*, 2006, **45**(1), 1–25.

Dey, S. and Juneja, S., Entropy approach to incorporate fat tailed constraints in financial models. Working Paper, 2010. Available online at: ssrn.com/abstract=1647048

Draper, D., Assessment and propagation of model uncertainty. *J. R. Stat. Soc. Ser. B (Methodological)*, 1995, **57**(1), 45–97.

El-Bachir, N. and Brigo, D., An analytically tractable time-changed jump-diffusion default intensity model. ICMA Centre Discussion Papers in Finance, 2008.

El Ghaoui, L., Oks, M. and Oustry, F., Worst-case value-at-risk and robust portfolio optimization: A conic programming approach. *Oper. Res.*, 2003, **51**(4), 543–556.

Geweke, J. and Amisano, G., Optimal prediction pools. *J. Econom.*, 2011, **164**(1), 130–141.

Glasserman, P. and Xu, X., Robust portfolio control with stochastic factor dynamics. *Oper. Res.*, forthcoming.

Glasserman, P. and Yu, B., Large sample properties of weighted Monte Carlo estimators. *Oper. Res.*, 2005, **53**(2), 298–312.

Goldfarb, D. and Iyengar, G.N., Robust portfolio selection problems. *Math. Oper. Res.*, 2003, **28**, 1–38.

Gulko, L., The entropy theory of stock option pricing. *Int. J. Theor. Appl. Finance*, 1999, **2**(3), 331–355.

Gulko, L., The entropy theory of bond option pricing. *Int. J. Theor. Appl. Finance*, 2002, **5**(4), 355–384.

Hansen, L.P., Sargent, T.J., Turmuhambetova, G. and Williams, N., Robust control and model misspecification. *J. Econ. Theory*, 2006, **128**(1), 45–90.

Hansen, L.P. and Sargent, T.J., *Robustness*, 2007 (Princeton University Press: Princeton, NJ).

Iyengar, G.N., Robust dynamic programming. *Math. Oper. Res.*, 2005, **30**, 257–280.

Jabbour, C., Pena, J.F., Vera, J.C. and Zuluage, L.F., An estimation-free, robust conditional value-at-risk portfolio allocation model. *J. Risk*, 2008, **11**(1), 57–78.

Ledoit, O. and Wolf, M., Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *J. Empir. Finance*, 2003, **10**(5), 603–621.

Madan, D.B., Carr, P.P. and Chang, E.C., The variance gamma process and option pricing. *Eur. Finance Rev.*, 1998, **2**(1), 79–105.

Meucci, A., Fully flexible views: Theory and practice. *Fully Flexible Views: Theory Pract. Risk*, 2008, **21**(10), 97–102.

Morini, M., *Understanding and Managing Model Risk*, 2011 (Wiley: Chichester).

Mykland, P.A., Conservative delta hedging. *Ann. Appl. Probab.*, 2000, **10**(2), 664–683.

Natarajan, K., Pachamanova, D. and Sim, M., Incorporating asymmetric distributional information in robust value-at-risk optimization. *Manage. Sci.*, 2008, **54**(3), 573–585.

Pesaran, M.H., Schleicher, C. and Zaffaroni, P., Model averaging in risk management with an application to futures markets. *J. Empir. Finance*, 2009, **16**(2), 280–305.

Petersen, I.R., James, M.R. and Dupuis, P., Minimax optimal control of stochastic uncertain systems with relative entropy constraints. *IEEE Trans. Autom. Control*, 2000, **45**(3), 398–412.

Raftery, A.E., Madigan, D. and Hoeting, J.A., Bayesian model averaging for linear regression models. *J. Am. Stat. Assoc.*, 1997, **92**, 179–191.

Rényi, A., On measures of entropy and information. In *Fourth Berkeley Symposium on Mathematical Statistics and Probability*, pp. 547–561, 1961 (University of California Press).

Rockafellar, R.T. and Uryasev, S., Conditional value-at-risk for general loss distributions. *J. Bank. Finance*, 2002, **26**(7), 1443–1471.

Segoviano, M.A. and Goodhart, C., Bank stability measures. Working Paper WP-09-4, International Monetary Fund, Washington, DC, 2009.

Shapiro, A., Dentcheva, D. and Ruszczyński, A., *Lectures on Stochastic Programming: Modeling and Theory*, vol. 9, 2009 (Society for Industrial and Applied Mathematics: Philadelphia, PA).

Szechtman, R. and Glynn, P.W., Constrained Monte Carlo and the method of control variates. In *Proceedings of the 2001 Winter Simulation Conference*, pp. 394–400, 2001 (IEEE Press: Piscataway, NJ).

Tankov, P. and Voltchkova, E., Jump-diffusion models: a practitioner's guide. Banque et Marchés, March–April, 2009.

Tsallis, C., Possible generalization of Boltzmann–Gibbs statistics. *J. Stat. Phys.*, 1988, **52**(1), 479–487.

Zhu, S. and Fukushima, M., Worst-case conditional value-at-risk with application to robust portfolio management. *Oper. Res.*, 2009, **57**(5), 1155–1168.

Zhu, S. and Pykhtin, M., A guide to modeling counterparty credit risk. *GARP Risk Rev.*, 2007, **37**, 16–22.


## Appendix A: Technical Assumptions

**Assumption A.1** *For the minimization problem* (3)

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

(2) *For all $a \in A$, the moment generating function $\hat{F}_a(\theta) = E[\exp(\theta V_a(X))]$ exists for $\theta$ 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 $\hat{F}_a(\theta)$ 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 $\hat{F}_a(\theta)$ is finite and thus an exponential change of measure defined by $\exp(\theta V_a(X))$ is well defined.

For any $\theta > 0$ and decision parameter $a$, if part (2) of Assumption A.1 is satisfied, the optimal change of measure for the adversary is described by the likelihood ratio

$$m^*_{\theta,a} = \exp(\theta V_a(X))/E[\exp(\theta V_a(X))], \tag{46}$$

where we need $\theta \in (0, \theta_{\max}(a))$ . By substituting (46) into (5), we get

$$\inf_a \inf_{\theta>0} \frac{1}{\theta} \log E[\exp(\theta V_a(X))] + \frac{\eta}{\theta}. \tag{47}$$

If $\theta_{\max}(a) < \infty$, then as $\theta \uparrow \theta_{\max}(a)$, the objective function in (47) goes to infinity, so the infimum over $\theta$ will automatically make the optimal $\theta$ smaller than $\theta_{\max}$. That is, we can safely consider $\theta < \infty$ instead of $\theta \in (0, \theta_{\max})$. This allows us to change the order of $\inf_a$ and $\inf_\theta$ in (5), whereas $\theta_{\max}(a)$ depends on the decision $a$. Now, we can relax the constraints for $\theta$ in both (5) and (47) to $\theta > 0$. Assumption A.1 is relevant to the $\inf_a$ and $\inf_\theta$ ordered as (47). To swap the order, we need the following assumption.

**Assumption A.2**

(1) *If* $(\theta^*, a^*, m^*)$ *solves* (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 any* $\theta \in [0, \theta^*_{\max})$*,* $E[\exp(\theta V_a(X))]$ *is lower semi-continuous in* $a$.

Because $E[\exp(\theta V_a(X))]$ is not necessarily continuous in $a$, the lower semi-contintuity condition in Assumption A.2 is needed to guarantees that the infimum in (8) can be attained.

**Assumption A.3** *For any a,* $V_a(X) \geq 0$ *almost surely under the nominal measure, and* $E[V_a(X)^{\frac{\alpha}{\alpha-1}}] < \infty$.