NORTHWESTERN UNIVERSITY

Advances in Risk Management Simulation

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Industrial Engineering and Management Sciences

By

R. Evren Baysal

EVANSTON, ILLINOIS

December 2008

ABSTRACT

Advances in Risk Management Simulation

R. Evren Baysal

Risk measurement involves estimating some functional of a loss distribution. This calls for nested simulation, in which risk factors are sampled at an outer level of simulation, while the inner level of simulation provides estimates of loss given each realization of the risk factors. Assessing the statistical uncertainty of estimates of risk measures at the outer simulation level is crucial in designing computationally efficient two-level simulation procedures for risk management applications. Confidence intervals for the risk measure of interest provide information about this statistical uncertainty and we provide asymptotically valid confidence intervals and confidence regions involving value at risk, conditional tail expectation, and expected shortfall (conditional value at risk), based on three different methodologies. One is an extension of previous work based on robust statistics, the second is a straightforward application of bootstrapping, and we derive the third using empirical likelihood. We then evaluate the small-sample coverage of the confidence intervals and regions in simulation experiments using financial examples. We find that the coverage probabilities are approximately nominal for large sample sizes, but are

noticeably low when sample sizes are too small (roughly, less than 500 here). The new empirical likelihood method provides the highest coverage at moderate sample sizes in these experiments.

Nested simulations can also be used in evaluating trading and hedging strategies. Suppose that one wishes to evaluate the distribution of profit and loss (P&L) resulting from a dynamic trading strategy. A straightforward method, then, is to simulate thousands of paths (i.e., time series) of relevant financial variables and to track the resulting P&L at every time at which the trading strategy rebalances its portfolio. In many cases, this requires numerical computation of portfolio weights at every rebalancing time on every path by a nested simulation performed conditional on market conditions at that time on that path. Such a two-level simulation could involve many millions of simulations to compute portfolio weights, and thus be too computationally expensive to attain high precision. We show that response surface methodology enables a more efficient simulation procedure: in particular, it is possible to do far fewer simulations by using kriging to model portfolio weights as a function of underlying financial variables.

Table of Contents

List of Tables

List of Figures

CHAPTER 1

Empirical Likelihood for Value at Risk and Expected Shortfall

1.1. Introduction

We want to measure the risk of a given portfolio which has random profits at the end of a predetermined investment period. We can sample from the distribution of the portfolio's profits using Monte Carlo simulation based on a stochastic model of financial markets. Our focus will be on estimating risk measures for our portfolio based on simulated profits and providing information in the form of confidence intervals and regions about the statistical uncertainty of these estimates. We address only this Monte Carlo sampling error in estimating risk, not the model risk that includes errors introduced by using an incorrect model of financial markets and statistical error in estimating the model's parameters from data. We will emphasize moderate Monte Carlo sample sizes, which are appropriate when it is computationally expensive to simulate financial scenarios and determine the value of the portfolio in each scenario.

Define V to be the random profit of the given portfolio at a specific investment horizon. The 95% Value at Risk ($\text{VaR}_{95\%}$) of the portfolio is the 95% quantile of the loss $-V$. A related risk measure is the 95% Conditional Tail Expectation ($\text{CTE}_{95\%}$), which is

$$\text{CTE}_{95\%} = \mathbf{E}[-V | -V \geq \text{VaR}_{95\%}].$$

Another closely related risk measure is Expected Shortfall ($\mathrm{ES}_{95\%}$), which is

$$\mathrm{ES}_{95\%} = -\frac{1}{0.05}\left(\mathbf{E}[V\mathbf{1}_{\{V\le v_{0.05}\}}] + v_{0.05}(0.05 - \mathbf{P}[V \le v_{0.05}])\right),$$

where $\nu_{0.05}$ is the lower 5% quantile (Def. 1.5.2) of the distribution of V. Under continuity conditions on the loss distribution, CTE equals ES ([**2**]). ES always equals Conditional Value-at-Risk (CVaR), which is coherent ([**2**], [**37**]). A risk measure is coherent if it satisfies certain axioms of translation invariance, subadditivity, positive homogeneity and monotonicity ([**6**]). We use the term "expected shortfall" here because ES includes an expectation, which is closely related to simulation, on which we focus, while CVaR is closely associated with a minimization formula due to [**36, 37**].

Our goal is to construct confidence intervals and regions for the above risk measures based on a simulated sample $V_1, \ldots, V_k$ of independent profits with common distribution F_0. Let $V_{[1]}, \ldots, V_{[k]}$ be ascending order statistics. The obvious point estimators of VaR and CTE at the $(1-p)$ level, assuming kp is an integer, are

$$\widehat{\mathrm{VaR}}_{1-p,k} = -V_{[kp]} \quad \text{and} \tag{1.1}$$

$$\widehat{\mathrm{CTE}}_{1-p,k} = -\frac{1}{kp}\sum_{i=1}^{kp} V_{[i]} \tag{1.2}$$

respectively. Other point estimators are discussed in Section 1.7.

Here we focus on constructing a confidence interval for ES and a confidence region for VaR and CTE simultaneously. We consider three methods of constructing them. To facilitate comparisons between their error rates, we also compare the three methods' confidence intervals for VaR to a standard confidence interval for VaR. This standard is the

binomial confidence interval for a quantile ([**14**]), which we summarize in Section 1.2. In Section 1.3, we construct confidence intervals and regions by extending results of [**45**] and [**30**] based on the influence function used in robust statistics. Section 1.4 briefly discusses how to construct them by bootstrapping. The major new results are in Section 1.5, where we show how to construct them using empirical likelihood (Owen 2001). In Section 1.6, we present computer simulation experiments which show that these confidence intervals and regions achieve close to nominal coverage for large sample sizes, but not for moderate sample sizes that are too small. Empirical likelihood provides the highest coverage at moderate sample sizes in these experiments, for the most part.

One contribution of this paper is simply in providing the first test (known to us) of the coverage of confidence regions and intervals involving CTE on financial examples. This provides some guidance about how large the sample size must be before the coverage is adequate, or how low the coverage might be at low sample sizes. Through a nontrivial application of empirical likelihood, we provide a method for generating confidence regions and intervals with higher coverage. The empirical likelihood approach is also useful in enabling risk measurement procedures that can cope with the need to use simulation at two levels: in sampling from a distribution of risky scenarios, and in estimating the portfolio loss in each of those scenarios ([**29**]).

1.2. Binomial Confidence Intervals for VaR

There is a well-known confidence interval for quantiles ([**14**]), and thus VaR, based on the binomial distribution of the number of losses $N(q) := \sum_{i=1}^{k} \mathbf{1}\{-V_i \geq q\}$ that exceed a threshold q. The lower and upper limits of a two-sided confidence interval for VaR_{1-p}

with $(1-\alpha)$ nominal coverage probability are respectively

$$\inf\left\{q \mid \sum_{n=N(q)+1}^{k}\binom{k}{n}p^n(1-p)^{k-n} \geq \frac{\alpha}{2}\right\} \quad \text{and} \quad \sup\left\{q \mid \sum_{n=0}^{N(q)}\binom{k}{n}p^n(1-p)^{k-n} \geq \frac{\alpha}{2}\right\}.$$

The limit of a one-sided upper confidence interval for VaR_{1-p} with $(1-\alpha)$ nominal coverage probability is $\sup\{q \mid \sum_{n=0}^{N(q)}\binom{k}{n}p^n(1-p)^{k-n} \geq \alpha\}$.

These endpoints of the confidence interval equal order statistics of the data sample, i.e. quantiles of the empirical distribution function. It is not generally possible to get exactly nominal coverage for the confidence interval because of the discreteness of the empirical distribution function, or, viewed differently, because of the discreteness of the binomial distribution ([**3**]). Nonetheless, these confidence intervals are often called "exact" because they are related to an exact hypothesis test for the value of the quantile. The justification of these confidence intervals does not involve the convergence of a statistic's distribution to a limiting distribution as sample size k grows, as do the methods described in later sections.

1.3. Influence Function

The approach based on the influence function in the theory of robust statistics allows us to compute the variances of the asymptotic normal distributions of the estimators in Equations (1.1)–(1.2). As [**30**, note 6] state, under regularity conditions discussed in [**43**],

$$k\mathbf{Var}(\widehat{\mathrm{CTE}}_{1-p,k}) \to \frac{\mathbf{Var}(-V \mid -V > \mathrm{VaR}_{1-p}) + p\,(\mathrm{CTE}_{1-p} - \mathrm{VaR}_{1-p})^2}{(1-p)}, \tag{1.3}$$

$$k\mathbf{Var}(\widehat{\mathrm{VaR}}_{1-p,k}) \to \frac{p\,(1-p)}{f^2(\mathrm{VaR}_{1-p})}, \quad \text{and} \tag{1.4}$$

$$k\mathbf{Cov}(\widehat{\mathrm{CTE}}_{1-p,k}, \widehat{\mathrm{VaR}}_{1-p,k}) \to \frac{p\,(\mathrm{CTE}_{1-p} - \mathrm{VaR}_{1-p})}{f(\mathrm{VaR}_{1-p})}, \tag{1.5}$$

where $f(\mathrm{VaR}_{1-p})$ is the value of the probability density of the underlying distribution at the quantile. Similar results appear in [**45**], but complicated by a truncation argument. [**45**] report confidence intervals for VaR and ES, but not a confidence region for both simultaneously. It also remains to show how to estimate the unknown quantities in Equations (1.3)–(1.5) to construct a confidence interval or region.

[**30**] propose the following estimates of asymptotic variances and covariances:

$$\widehat{\mathbf{Var}}_k(\widehat{\mathrm{CTE}}_{1-p,k}) = \frac{\frac{1}{kp-1}\sum_{i=1}^{kp}(\widehat{\mathrm{CTE}}_{1-p,k} + V_{[i]})^2 + p\,(\widehat{\mathrm{CTE}}_{1-p,k} + V_{[kp]})^2}{k\,(1-p)}, \tag{1.6}$$

$$\widehat{\mathbf{Var}}_k(\widehat{\mathrm{VaR}}_{1-p,k}) = \frac{p\,(1-p)}{k\hat{f}^2(-V_{[kp]})}, \quad \text{and} \tag{1.7}$$

$$\widehat{\mathbf{Cov}}_k(\widehat{\mathrm{CTE}}_{1-p,k}, \widehat{\mathrm{VaR}}_{1-p,k}) = \frac{p\,(\widehat{\mathrm{CTE}}_{1-p,k} + V_{[kp]})}{k\,\hat{f}(-V_{[kp]})}, \tag{1.8}$$

where$\hat{f}(-V_{[kp]})$ is an estimate of the probability density. [**30**] propose to use

$$\hat{f}(-V_{[kp]}) = \frac{\xi}{F_k^{-1}(p) - F_k^{-1}(p-\xi)}$$

where $F_k(x) = (1/k)\sum_{i=1}^{k} 1_{\{V_i \le x\}}$ is the empirical distribution derived from the sample of size k and ξ is chosen to be a small number. Note that the choice of ξ affects the empirical

density function estimate $\hat{f}$ considerably, especially for small samples. Hence, we propose to use the kernel method to estimate f via a Gaussian kernel estimator function:

$$\hat{f}_k(-V_{[kp]}) = \frac{1}{kh}\sum_{i=1}^{k}\Phi'\left(\frac{-V_{[kp]}+V_{[i]}}{h}\right) \tag{1.9}$$

where $h = (\frac{4}{3k})^{1/5}\sigma$, $\Phi'(u) = (2\pi)^{-1/2}\exp(-u^2/2)$ and the sample standard deviation can be used for σ.

We extend the results above to create confidence intervals and regions. We define $Y := \begin{pmatrix} \widehat{\mathrm{VaR}}_{1-p,k} \\ \widehat{\mathrm{CTE}}_{1-p,k} \end{pmatrix}$ based on a sample of size k. This is asymptotically normal with mean $y_0 := \begin{pmatrix} \mathrm{VaR}_{1-p} \\ \mathrm{CTE}_{1-p} \end{pmatrix}$ and covariance matrix $\mathbf{\Sigma}$ described by Equations (1.3)–(1.5). There exists a unique symmetric positive definite matrix $\mathbf{A}$ such that $\mathbf{A}^\top\mathbf{A} = \mathbf{\Sigma}^{-1}$. We define $Z := \mathbf{A}(Y - y_0)$ whose components are independent and asymptotically standard normal. Then, the quadratic form $(Y-y_0)^\top\mathbf{\Sigma}^{-1}(Y-y_0) = \mathrm{Z}^\top\mathrm{Z}$ is distributed asymptotically as χ^2 with two degrees of freedom. Note that the asymptotic formulae (1.6)–(1.8) can be used to construct $\widehat{\mathbf{\Sigma}}$ as an estimate of the covariance matrix $\mathbf{\Sigma}$. With probability one, $\widehat{\mathbf{Var}}_k(\widehat{\mathrm{CTE}}_{1-p,k})$ converges to $\mathbf{Var}(\widehat{\mathrm{CTE}}_{1-p,k})$ ([**27**]). Weak convergence results for the kernel density estimate $\hat{f}_k$ (1.9) are given by [**42**] and [**11**]. Hence, $\widehat{\mathbf{\Sigma}}$ is a consistent estimator of $\mathbf{\Sigma}$ and by the converging-together lemma of [**18**] an asymptotically valid $(1-\alpha)$ confidence region for VaR_{1-p} and CTE_{1-p}, is an elliptical region centered at $(\widehat{\mathrm{VaR}}_{1-p,k}, \widehat{\mathrm{CTE}}_{1-p,k})$ and is given by

$$\left\{ y_0 \,\middle|\, (Y-y_0)^\top\widehat{\mathbf{\Sigma}}^{-1}(Y-y_0) \le \chi^2_{(2),1-\alpha} \right\} \tag{1.10}$$

where $\chi^2_{(2),1-\alpha}$ is the $1-\alpha$ quantile of the chi-squared distribution with two degrees of freedom. By applying the converging-together lemma to $\widehat{\mathrm{CTE}}_{1-p,k}$ and $\widehat{\mathbf{Var}}_k(\widehat{\mathrm{CTE}}_{1-p,k})$, one can show that where $Z_{1-\alpha/2}$ is the $1-\alpha/2$ quantile of the standard normal distribution,

$$\left\{\mu_0 \left| |\widehat{\mathrm{CTE}}_{1-p,k} - \mu_0| \leq Z_{1-\alpha/2}\sqrt{\widehat{\mathbf{Var}}_k(\widehat{\mathrm{CTE}}_{1-p,k})} \right.\right\} \tag{1.11}$$

is a two-sided confidence interval for CTE_{1-p} ([**27**]). Correspondingly,

$$\left\{\mu_0 \left| \mu_0 \leq \widehat{\mathrm{CTE}}_{1-p,k} + Z_{1-\alpha}\sqrt{\widehat{\mathbf{Var}}_k(\widehat{\mathrm{CTE}}_{1-p,k})} \right.\right\} \tag{1.12}$$

is a one-sided upper confidence interval for CTE_{1-p}. Likewise,

$$\left\{q_0 \left| |\widehat{\mathrm{VaR}}_{1-p,k} - q_0| \leq Z_{1-\alpha/2}\sqrt{\widehat{\mathbf{Var}}_k(\widehat{\mathrm{VaR}}_{1-p,k})} \right.\right\} \tag{1.13}$$

is a two-sided confidence interval for VaR_{1-p} and

$$\left\{q_0 \left| q_0 \leq \widehat{\mathrm{VaR}}_{1-p,k} + Z_{1-\alpha}\sqrt{\widehat{\mathbf{Var}}_k(\widehat{\mathrm{VaR}}_{1-p,k})} \right.\right\}. \tag{1.14}$$

is a one-sided upper confidence interval for VaR_{1-p}.

1.4. Bootstrapping

The idea of bootstrapping to create confidence intervals for CTE was suggested by [**17**] and [**24**]. Bootstrap methods are in general motivated by the need to evaluate the accuracy of an estimate in the absence of distributional assumptions ([**13**]). [**39**] discuss in detail the application of bootstrap methods to hypothesis testing and confidence interval

estimation for various statistics including quantiles. The logic behind bootstrapping for quantile estimation is readily applicable to estimating VaR, CTE, and ES.

As before, let $V_{[1]} \leq \ldots \leq V_{[k]}$ be order statistics, sorted after sampling profits independently from the common distribution F_0. We assume kp is an integer. To estimate VaR_{1-p} and CTE_{1-p} based on this sample, we compute the obvious estimators previously mentioned, $\widehat{\mathrm{VaR}}_{1-p,k} = -V_{[kp]}$ and $\widehat{\mathrm{CTE}}_{1-p,k} = -\frac{1}{kp}\sum_{i=1}^{kp} V_{[i]}$. We will denote them by $\hat{q}_k(p)$ and $\hat{\mu}_k(p)$, respectively, to emphasize their dependence on the initial sample of size k. Because kp is an integer, the estimate $\hat{\mu}_k(p)$ of CTE_{1-p} is also an estimate of ES_{1-p}.

In order to assess the uncertainty associated with these estimates, we generate B i.i.d. bootstrap samples by resampling from the empirical distribution function F_k of the initial Monte Carlo sample. For risk management applications, resampling may be considerably faster than generating samples from the original distribution F_0. We denote the bootstrap samples by $\hat{V}^b_{[1]}, \ldots, \hat{V}^b_{[k]}$ for $b = 1, \ldots, B$. From the bth bootstrap sample, we compute the estimates

$$\hat{\mu}_b(p) = -\frac{1}{kp}\sum_{i=1}^{kp} \hat{V}^b_{[i]} \quad \text{and} \quad \hat{q}_b(p) = -\hat{V}^b_{[kp]}.$$

Note that we only need $\hat{V}^b_{[1]}, \ldots, \hat{V}^b_{[kp]}$ to compute $\hat{q}_b(p)$ and $\hat{\mu}_b(p)$, and the bootstrap sample for the first kp order statistics can be generated efficiently by $\hat{V}^b_{[j]} = F_k^{-1}(U_{[j]})$ where $U_{[1]}, \ldots, U_{[k]}$ are the order statistics of an i.i.d. sample of size k from the standard uniform distribution. The following algorithm of order $O(kp)$ from [**15**] can be used to

generate $U_{[1]}, \ldots, U_{[kp]}$:

$$
\begin{aligned}
&U_{[0]} = 0, \\
&\text{for } i = 1 \text{ to } kp \\
&\quad \text{generate } V_i \sim \text{Uniform}[0,1], \\
&\quad U_{[i]} = 1 - (1 - U_{[i-1]}) V_i^{\frac{1}{kp-i+1}}, \\
&\text{end for.}
\end{aligned}
$$

1.4.1. Bootstrap Confidence Intervals for VaR and ES

There are various methods for constructing asymptotically valid confidence intervals for VaR_{1-p} and ES_{1-p} from $\hat{q}_1(p), \ldots, \hat{q}_B(p)$ and $\hat{\mu}_1(p), \ldots, \hat{\mu}_B(p)$, such as the bootstrap t, the bootstrap percentile, the bootstrap bias-corrected percentile, and the bootstrap bias-corrected/accelerated (BCa) percentile methods ([**39**]). We use the `bootci` function of the MATLAB Statistical Toolbox to construct BCa intervals in our experiments. We set the upper confidence limits of one-sided $100(1-\alpha)\%$ upper confidence intervals to the upper limits of the corresponding two-sided confidence intervals with $(1-2\alpha)$ nominal coverage probability.

1.4.2. Bootstrap Confidence Regions for VaR and CTE

[**16**] suggests basing a joint bootstrap confidence region for a vector parameter y_0 on the quadratic form

$$\mathcal{Q} = (Y - y_0)^\top \widehat{\boldsymbol{\Sigma}}^{-1} (Y - y_0)$$

where Y is an estimate of y_0 and $\widehat{\mathbf{\Sigma}}$ is the estimated covariance matrix of Y. When Y is approximately normal, $\mathcal{Q}$ will be approximately $\chi^2_{(2)}$. Its distribution can be assessed by bootstrapping instead.

As in Section 1.3, we let $y_0 = \begin{pmatrix} \mathrm{VaR}_{1-p} \\ \mathrm{CTE}_{1-p} \end{pmatrix}$, $Y = \begin{pmatrix} -V_{[kp]} \\ -\frac{1}{kp}\sum_{i=1}^{kp} V_{[i]} \end{pmatrix}$, and $\widehat{\mathbf{\Sigma}}^{-1}$ be the influence function estimate of the covariance matrix of Y, as in Equations (1.6)–(1.8). We calculate

$$\mathcal{Q}^b = (Y_b - Y)^\top \widehat{\mathbf{\Sigma}}_b^{-1}(Y_b - Y)$$

for each bootstrap sample $b = 1, \ldots, B$, yielding an estimate Y_b of CTE_{1-p} and an estimated covariance matrix $\widehat{\mathbf{\Sigma}}_b^{-1}$. We denote the ordered bootstrap values as $\mathcal{Q}^b_{[1]} \leq \cdots \leq \mathcal{Q}^b_{[B]}$. Then a bootstrap confidence region for the vector parameter y_0 is the set

$$\left\{ y_0 \mid (Y - y_0)^\top \widehat{\mathbf{\Sigma}}^{-1}(Y - y_0) \leq \mathcal{Q}^b_{[B(1-\alpha)]} \right\}, \tag{1.15}$$

which is similar to Equation (1.10) but with $\mathcal{Q}^b_{[B(1-\alpha)]}$ replacing $\chi^2_{(2),1-\alpha}$.

1.5. Empirical Likelihood

Empirical likelihood (EL) is a nonparametric method for hypothesis testing (and therefore for confidence region construction) that is similar to the usual parametric likelihood ratio approach, which rejects a hypothesis when its likelihood ratio is too low. The empirical likelihood ratio, instead of being constructed from a parametric family of distributions, considers the family $\mathcal{F}_k := \{F \mid F \ll F_k\}$ of discrete distributions absolutely continuous with respect to the empirical cumulative distribution F_k whose support equals the observed data points. Such a distribution $F \ll F_k$ puts weights (i.e., probability mass)

$w_1, \ldots, w_k$ on order statistics $V_{[1]}, \ldots, V_{[k]}$, where the weights must be nonnegative and sum to 1. The empirical likelihood of F is $\prod_{i=1}^{k} w_i$ and the empirical likelihood ratio of F is defined as $R(F) := \prod_{i=1}^{k}(kw_i)$, since the maximum likelihood member of $\mathcal{F}_k$ is the empirical distribution, F_k, which has all weights equal to $1/k$ and thus has empirical likelihood k^{-k}.

Let $T(\cdot)$ be some statistical functional of the distribution F_0, where F_0 is the true distribution of portfolio profit V. The nonparametric maximum likelihood estimate of $T(F_0)$ is $T(F_k)$ and sets of the form

$$\{T(F) \mid R(F) \geq r,\ F \in \mathcal{F}_k\} \tag{1.16}$$

can be used as confidence regions for $T(F_0)$ where r is chosen appropriately to get the right asymptotic coverage, as $k \to \infty$ ([**32**]).

In particular, the empirical likelihood confidence interval for VaR coincides with the binomial confidence interval of Section 1.2 (Owen 2001, Sec. 3.6).

1.5.1. A Nonparametric Confidence Region for VaR and CTE

Definition 1.5.1. For any $0 < p < 1$, any value Q^p such that $\mathbf{P}(V \leq Q^p) \geq p$ and $\mathbf{P}(V \geq Q^p) \geq 1 - p$ is a p-quantile of F_0 ([**34**]).

We defined the 95% Value at Risk of our portfolio as the 95% quantile of the loss given by $Q_{-V}^{95\%}$. Using the definition above, we see that this is equivalent to the negative of the 5% quantile of the profit, which is given by $-Q_V^{5\%}$. Then, the 95% Conditional Tail Expectation of our portfolio is $\mathbf{E}[-V \mid -V \geq Q_{-V}^{95\%}] = -\mathbf{E}[V \mid V \leq Q_V^{5\%}]$.

Our goal is to construct an empirical likelihood confidence region for $\mathrm{VaR}_{1-p} = -Q_V^p$ and $\mathrm{CTE}_{1-p} = -\mathbf{E}[V \mid V \leq Q_V^p]$ and to provide asymptotic coverage probability results for such confidence regions.

Definition 1.5.2. The lower and upper p-quantiles of any distribution F are defined as $\nu_p := \inf\{v|F(v) \geq p\}$ and $\nu^p := \inf\{v|F(v) > p\}$, respectively ([**2**]).

Definition 1.5.3. The expected shortfall at level $1-p$ of V is defined as

$$\mathrm{ES}_{1-p} := -p^{-1}\left(\mathbf{E}[V\mathbf{1}_{\{V \leq v_p\}}] + v_p\left(p - \Pr[V \leq v_p]\right)\right)$$

where ν_p is the lower p-quantile of the distribution of V ([**2**]).

Because it is not, in general, uniquely defined, it is not possible to write Q_V^p of Definition 1.5.1 as a statistical functional $T(F_0)$. This poses a problem for constructing confidence regions of the form (1.16). However, if F_0 is continuous and strictly increasing at Q_V^p, then Q_V^p is unique and is equal to ν_p. Furthermore, $\mathbf{P}(V \leq Q_V^p) = \mathbf{P}(V \leq \nu_p) = p$, which by Definition 1.5.3 implies $\mathrm{ES}_{1-p} = \mathrm{CTE}_{1-p}$. Under this simple restriction on F_0, the empirical likelihood results for M-estimates ([**33**]) can be used to construct empirical likelihood confidence regions for VaR_{1-p} and ES_{1-p}.

Definition 1.5.4. An M-estimate is a statistical functional defined as a root $t = T_\psi(F)$ of

$$\int \psi(V,t)F(dV) = 0 \tag{1.17}$$

where $V \sim F$ ([**33**]).

Proposition 1.5.1. *If F_0 is continuous and strictly increasing at its p-quantile, the functional T_ψ defined by Equation (1.17) is an M-estimate for the vector (VaR_{1-p}, ES_{1-p}) where the function $\psi : \mathcal{R} \times \mathcal{R}^2 \to \mathcal{R}^2$ is given by $\psi(V,(q,\mu)) := (p - \mathbf{1}_{\{V \le -q\}}, \mu + \frac{1}{p} V \mathbf{1}_{\{V \le -q\}})$.*

Proof: The unique root of Equation (1.17) with $F = F_0$ is $(\mathrm{VaR}_{1-p}, \mathrm{ES}_{1-p})$, as follows. First, $\int (p - \mathbf{1}_{\{V \le -q\}}) F_0(dV) = p - F_0(-q) = 0$ which implies $F_0(-q) = p$. Since we assumed F_0 has a unique p-quantile with $F_0(Q_V^p) = p$, we find $Q_V^p = -q$ and $\mathrm{VaR}_{1-p} = -Q_V^p = q$. Second, $\int (\mu + \frac{1}{p} V \mathbf{1}_{\{V \le -q\}}) F_0(dV) = \mu + \frac{1}{p} \int_{-\infty}^{-q} V F_0(dV) = 0$ which implies $\mu = -\mathbf{E}[V \mid V \le -q]$. Again by uniqueness of $Q_V^p = \nu_p = -q$ and therefore of $\mathrm{ES}_{1-p} = -\mathbf{E}[V \mid V \le v_p]$, we get $\mu = \mathrm{ES}_{1-p}$. $\square$

Note that for $\psi(V,(q,\mu))$ defined as in Proposition 1.5.1 and $T_\psi(F)$ defined as in Definition 1.5.4, the set $\{T_\psi(F) \mid F \ll F_k, R(F) \ge r\}$ equals the confidence region $\{(q,\mu) \mid \int \psi(V,(q,\mu)) F(dV) = 0, F \ll F_k, R(F) \ge r)\}$ for VaR_{1-p} and ES_{1-p} depicted in Figure 1.1.

Proposition 1.5.2. *For ψ defined as in Proposition 1.5.1, if F_0 is continuous and strictly increasing at its p-quantile, and if $V \mathbf{1}_{\{V \le Q_V^p\}}$ is not a constant and $\mathbf{E}[V^2 \mathbf{1}_{\{V \le Q_V^p\}}] < \infty$, then $\{T_\psi(F) \mid F \ll F_k, R(F) \ge \exp(-\frac{1}{2}\chi^2_{(2),1-\alpha})\}$ is a confidence region for VaR_{1-p} and CTE_{1-p} with $(1-\alpha)$ asymptotic coverage probability.*

Proof: By Proposition 1.5.1, $T_\psi(F_0)$ exists and is unique if F_0 is continuous and increasing at ν_p, which we already assumed. The assumption that $V \mathbf{1}_{\{V \le Q_V^p\}}$ is not a constant and $\mathbf{E}[V^2 \mathbf{1}_{\{V \le Q_V^p\}}] < \infty$ implies that the rank of $\mathbf{Var}[\psi(V,t)]$ is two. Then, we

can use Theorem 3 of [**33**] to show that $\Pr\left[T_\psi(F_0) \notin \{T_\psi(F) \mid F \ll F_k, R(F) \geq r\}\right] \to \alpha$ as $k \to \infty$ if we pick $r = \exp(-\frac{1}{2}\chi^2_{(2),1-\alpha})$. $\square$

While computing $\{T_\psi(F) \mid F \ll F_k, R(F) \geq r\}$, we must restrict our attention to F within the family $\mathcal{F}_k$ such that for some l, W_l defined by

$$W_l := \sum_{i=1}^{l} w_i \tag{1.18}$$

is equal to p. Otherwise, $T_\psi(F)$ does not exist. It is worth noting that $T_\psi(F)$ is not unique for such $F \in \mathcal{F}_k$ since for any q such that $-q \in \left[V_{[l]}, V_{[l+1]}\right)$, $(q, -\frac{1}{p}\sum_{i=1}^{l} w_i V_{[i]})$ is a root of $\int \psi(V,t)dF(V) = 0$; however, we require only $T_\psi(F_0)$ to be unique.

A confidence region with $(1-\alpha)$ asymptotic coverage probability can be written as

$$\begin{aligned}
CR_{1-\alpha} &= \left\{ t \,\middle|\, \int \psi(V,t)dF(V) = 0,\ F \ll F_k,\ R(F) \geq r \right\} \\
&= \bigcup_{l=1}^{k-1} \left\{ \mu \,\middle|\, \prod_{i=1}^{k}(kw_i) \geq r, \sum_{i=1}^{l} w_i = p, \mu = -\frac{1}{p}\sum_{i=1}^{l} V_{[i]}w_i, w_i \geq 0, \sum_{i=1}^{k} w_i = 1 \right\} \times \\
&\qquad \left(-V_{[l+1]}, -V_{[l]}\right] \\
&= \bigcup_{l=1}^{k-1} \left(-V_{[l+1]}, -V_{[l]}\right] \times \left\{ \mu \mid R_l^\psi(\mu) \geq r \right\}
\end{aligned}$$

where $R_l^\psi(\mu) := \max_w \left\{ \prod_{i=1}^{k}(kw_i) \mid \sum_{i=1}^{l} w_i = p, \mu = -\frac{1}{p}\sum_{i=1}^{l} V_{[i]}w_i, w_i \geq 0, \sum_{i=1}^{k} w_i = 1 \right\}$. Appendix 1 up through Lemma 1.1 shows that this maximum is attained at $\{w_i^*\}_{i=1,\ldots,k}$ given by

$$w_i^* = \begin{cases} \frac{p}{l}[1 - (V_{[i]} + \mu)\lambda^*]^{-1} & \text{for } i = 1, \ldots, l \\ \frac{1-p}{k-l} & \text{for } i = l+1, \ldots, k \end{cases} \tag{1.19}$$

where λ^* is the unique solution to

$$\sum_{i=1}^{l} \frac{V_{[i]} + \mu}{1 - (V_{[i]} + \mu)\lambda^*} = 0$$

which can be computed by numerical root finding within the interval $\left[\frac{1-1/l}{\mu+V_{[1]}}, \frac{1-1/l}{\mu+V_{[l]}}\right]$. By

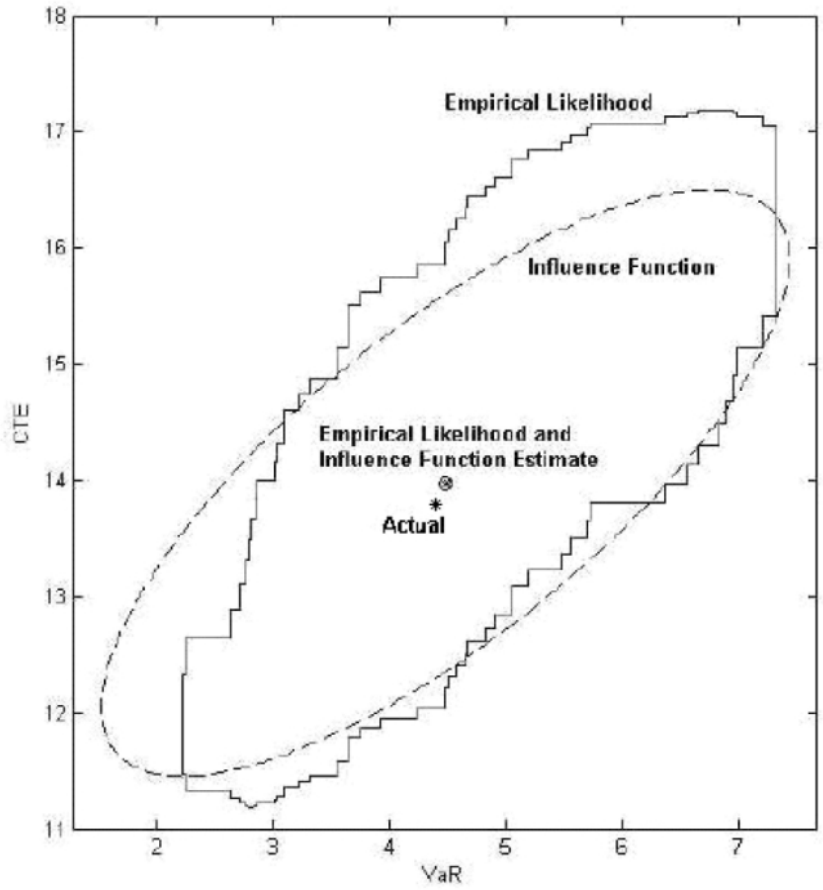

Figure 1.1. Influence Function and Empirical Likelihood Confidence Regions

Lemma 1.2 in Appendix 1, for each l, $R_l^{\psi}(\mu)$ is single peaked at $-\frac{1}{l}\sum_{i=1}^{l} V_{[i]}$ and continuous and monotone on either side of this peak. This implies $I_l^{\psi} := \{\mu \mid R_l^{\psi}(\mu) \geq r\}$ is an

interval if it is not empty. Because $\max_\mu R_l^\psi(\mu) = k^k \left(\frac{p}{l}\right)^l \left(\frac{1-p}{k-l}\right)^{k-l}$, I_l^ψ is non-empty if $k \log k + l \log \frac{p}{l} + (k-l) \log \frac{1-p}{k-l} \geq -\frac{1}{2}\chi^2_{(2),1-\alpha}$. Therefore $I_l^\psi \subseteq \left[-V_{[l]}, V_{[1]}\right]$ can be computed as $I_l^\psi = \left[\mu_l^{lo}, \mu_l^{hi}\right]$ where

$$\mu_l^{lo} \text{ is the unique root of } R_l^\psi(\mu) = r \text{ in } \left[-V_{[l]}, -\frac{1}{l}\sum_{i=1}^{l} V_{[i]}\right]$$

and

$$\mu_l^{hi} \text{ is the unique root of } R_l^\psi(\mu) = r \text{ in } \left[-\frac{1}{l}\sum_{i=1}^{l} V_{[i]}, -V_{[1]}\right].$$

Finally, we compute $CR_{1-\alpha} = \bigcup_{l=1}^{k-1} \left(-V_{[l+1]}, -V_{[l]}\right] \times \left[\mu_l^{lo}, \mu_l^{hi}\right]$. Figure 1.1 compares the shape of such a confidence region to the shape of a confidence region constructed by the influence function approach.

1.5.2. Nonparametric Confidence Intervals for Expected Shortfall

Complications arise when we try to compute a confidence interval for CTE even if we restrict our attention to continuous distributions for which CTE is coherent. This is because $\psi(V, (q, \mu))$ is a non-smooth function of (q, μ) and hence theoretical justification is lacking to profile out either component of (q, μ) to get a confidence interval for the other. We, therefore, turn our attention to expected shortfall (ES), for which we can use empirical likelihood theory to compute an asymptotically valid confidence interval. Note that ES is still coherent even if the profit distribution F_0 is not continuous or strictly increasing at Q^p.

Empirical likelihood most naturally produces two-sided confidence intervals, and we will focus on these in this section. We produce one-sided confidence intervals according

to the following suggestion of Owen (2001, Sec. 2.7). Where (L, U) is a two-sided $100(1-2\alpha)\%$ empirical likelihood confidence interval, $(-\infty, U)$ can be used as a one-sided $100(1-\alpha)\%$ confidence interval.

[**2**] show that ES_{1-p} of Definition 1.5.3 can be represented as a functional T by $T(F_0) := -\frac{1}{p}\int_0^p \mathrm{F}_0^{-1}(u)du$ where $F_0^{-1}(u) := \inf\{v \mid \mathrm{F}_0(v) \geq u\}$. The empirical likelihood ratio of the hypothesis $\mu = T(F_0)$ is defined as

$$R(\mu) := \max\left\{\prod_{i=1}^{k}(kw_i) \mid T(F) = \mu, F \ll F_k\right\}$$

where F has weights $\{w_i\}_{i=1,\ldots,k}$ and, with W defined as in Equation (1.18),

$$T(F) = -\frac{1}{p}\left\{\sum_{i=1}^{l-1}\int_{W_{i-1}}^{W_i} V_{[i]}du + \int_{W_{l-1}}^{p} V_{[l]}du\right\} = -\frac{1}{p}\left\{\sum_{i=1}^{l-1} w_i V_{[i]} + (p - W_{l-1})V_{[l]}\right\}$$

with l determined by $W_l \geq p$ and $W_{l-1} < p$.

Proposition 1.5.3. *If $|F_0^{-1}(u)|$ is $O(u^{-\frac{1}{3}+\epsilon})$ as $u \to 0$, for some $\epsilon > 0$, then a confidence interval for ES_{1-p} with $100(1-\alpha)\%$ asymptotic coverage probability is*

$$\left\{\mu \mid F \ll F_k, R(\mu) \geq \exp(-\frac{1}{2}\chi^2_{(1),1-\alpha})\right\}.$$

Proof: We start by writing $\mathrm{ES}_{1-p} = T(F_0) = \int_0^1 F_0^{-1}(u)g(u)du$ where $g(u) = -\frac{1}{p}\mathbf{1}_{\{u\leq p\}}$. Note that $T(F)$ produces an L-estimator when we plug in the c.d.f. F_k for F. According to Theorem 10.2 of [**34**], $\Pr\left[T(F_0) \notin \left\{\mu \mid F \ll F_k, R(\mu) \geq \exp(-\frac{1}{2}\chi^2_{(1),1-\alpha})\right\}\right] \to$

α as $k \to \infty$ if for some $c \in (0, \infty)$, some $M \in (0, \infty)$, and some $d \in (\frac{1}{6}, \frac{1}{2})$, both

$$|g(u)| \leq M[u(1-u)]^{\frac{1}{c}-\frac{1}{2}+d} \quad \text{and}$$

$$|F_0^{-1}(u)| \leq M[u(1-u)]^{-\frac{1}{c}}$$

hold for all $0 < u < 1$. In our case, $g(u) = 0$ for $u > p$, so only the left tail behavior is relevant. That is, we are only concerned with the behavior of F_0^{-1} as $u \to 0$, because our L-estimator uses only values less than the median of the data sample. We will show that

$$|g(u)| \leq M u^{\frac{1}{c}-\frac{1}{2}+d} \quad \text{and} \tag{1.20}$$

$$|F_0^{-1}(u)| \leq M u^{-\frac{1}{c}} \tag{1.21}$$

hold for suitable values of c, d, and M, given the assumption that $F_0^{-1}(u)$ is $O(u^{-\frac{1}{3}+\epsilon})$ as $u \to 0$.

The interesting case is when losses are unbounded, in which case $\epsilon < 1/3$. Take $c = 1/(1/3-\epsilon)$ and $d = 1/6+\epsilon$. Then (1.21) holds for sufficiently large M by assumption and (1.20) holds for $M \geq 1/p$ because $1/c - 1/2 + d = 0$ and $|g(u)| = 1/p$ for $u < p$.

If losses are bounded, take M to be the maximum of the bound and $1/p$. Take $c = 3$ and $d = 1/3$. Then (1.21) holds because $|F_0^{-1}(u)| \leq M \leq M u^{-\frac{1}{3}}$ and (1.20) holds because $|g(u)| \leq M \leq M u^{-\frac{1}{6}}$. $\square$

We compute $R(\mu)$ by $R(\mu) = \max_{l=1,\ldots,k} R_l(\mu)$ where

$$\begin{aligned} R_l(\mu) &= \sup\left\{\prod_{i=1}^{k}(kw_i) \mid \mu = T(F), W_l \geq p, W_{l-1} < p, W_k = 1, w_i \geq 0\right\} \\ &= \max\left\{R_l^{\psi}(\mu), R_l^{int}(\mu)\right\} \end{aligned}$$

and $R_l^{int}(\mu)$ is defined as

$$R_l^{int}(\mu) := \sup\left\{\prod_{i=1}^{k}(kw_i) \mid \mu = T(F), W_l > p, W_{l-1} < p, W_k = 1, w_i \geq 0\right\}.$$

We observe that $R_l^{\psi}(\mu) = \max\{\prod_{i=1}^{k}(kw_i) \mid \mu = T(F), W_l = p, W_k = 1, w_i \geq 0\}$ is as defined in the previous section because for $W_l = p$, we get $T(F) = -\frac{1}{p}\sum_{i=1}^{l} w_i V_{[i]}$ with $W_{l-1} < p$, optimally. As $W_l \to p$ and as $W_{l-1} \to p$, limits of feasible points in this maximization converge to feasible points in the maximizations of the previous section whose optimal values are, respectively, $R_l^{\psi}(\mu)$ and $R_{l-1}^{\psi}(\mu)$. This reasoning shows that

$$R(\mu) = \max_{l=1,\ldots,k}\left\{\max\left\{R_l^{\psi}(\mu), R_l^{int}(\mu)\right\}\right\} = \max\left\{\max_{l=1,\ldots,k} R_l^{\psi}(\mu), \max_{l\in L^{int}(\mu)} R_l^{int}(\mu)\right\} \tag{1.22}$$

where $l \in L^{int}(\mu)$ if and only if $R_l^{int}(\mu)$ is attained at an interior solution characterized by $W_l > p$ and $W_{l-1} < p$, since otherwise $R_l^{int}(\mu) = R_l^{\psi}(\mu)$ or $R_{l-1}^{\psi}(\mu)$.

Since we have already found a way to compute $R_l^{\psi}(\mu)$, we need only concern ourselves with interior solutions $R_l^{int}(\mu)$ with $l \in L^{int}(\mu)$ to the following problem:

$$\begin{aligned} &\text{maximize} && \textstyle\prod_{i=1}^{k}(kw_i) \\ &\text{subject to} && \mu = -V_{[l]} - \tfrac{1}{p}\textstyle\sum_{i=1}^{l-1} w_i(V_{[i]} - V_{[l]}) \\ &&& W_l > p \text{ and } W_{l-1} < p \\ &&& W_k = 1 \text{ and } w_i \geq 0 \end{aligned} \tag{1.23}$$

which we will refer to as Maximization Problem II. It is maximization of a concave objective with linear constraints and nonzero Hessian, so there is an interior solution if and only if there is a solution to the two first-order conditions in two unknowns, which are

$$W_{l-1}^* - \sum_{i=1}^{l-1} g_i(W_{l-1}^*, \lambda^*) = 0 \tag{1.24}$$

and

$$\sum_{i=1}^{l-1} g_i(W_{l-1}^*, \lambda^*)(V_{[l]} - V_{[i]}) - p(\mu + V_{[l]}) = 0 \tag{1.25}$$

where g_i is a function specifying the optimal weight w_i for $i = 1, \ldots, l-1$. In Appendix 2, we show that $g_i(W_{l-1}^*, \lambda^*) := \left[\frac{k-l+1}{1-W_{l-1}^*} + \lambda^*(V_{[l]} - V_{[i]})\right]^{-1}$, so the optimal weights are

$$w_i^* = \begin{cases} \left[\frac{k-l+1}{1-W_{l-1}^*} + \lambda^*(V_{[l]} - V_{[i]})\right]^{-1} & \text{for } i = 1, \ldots, l-1 \\ \frac{1-W_{l-1}^*}{k-l+1} & \text{for } i = l, \ldots, k \end{cases} \tag{1.26}$$

where λ^* and $W_{l-1}^* \in \left(p - \frac{1-p}{k-l}, p\right)$ solve the first order conditions.

We construct a confidence interval with $100(1-\alpha)\%$ asymptotic coverage probability for expected shortfall as $CI_{1-\alpha} := \{\mu \mid F \ll F_k, R(\mu) \geq r\}$. By Equation (1.22), μ is in

$CI_{1-\alpha}$ if and only if $R_l^{\psi} \geq r$ for some l or $R_l^{int}(\mu) \geq r$ for some $l \in L^{int}(\mu)$. Then, $CI_{1-\alpha}$ can be computed as

$$CI_{1-\alpha} = \left(\bigcup_{l=1}^{k} I_l^{\psi}\right) \bigcup \left(\bigcup_{l=1}^{k} I_l^{int}\right)$$

where we define $I_l^{int} := \{\mu \mid l \in L^{int}(\mu), R_l^{int}(\mu) \geq r\} = \{\mu \mid \mu \in M_l^{int}, R_l^{int}(\mu) \geq r\}$ and $M_l^{int} = \{\mu \mid l \in L^{int}(\mu)\}$ is the set of μ such that Equations (1.24) and (1.25) have an interior solution. We show by Lemma 2.2 of Appendix 2 that M_l^{int} is an open interval whose lower endpoint m_l^{lo} satisfies Equations (1.24) and (1.25) with $W_{l-1}^* = p - \frac{1-p}{k-l}$ and whose upper endpoint m_l^{hi} satisfies Equations (1.24) and (1.25) with $W_{l-1}^* = p$.

We have already shown how to calculate I_l^{ψ} in Section 1.5.1 and it remains to compute I_l^{int}. By definition, I_l^{int} is a subset of $M_l^{int} = \left(m_l^{lo}, m_l^{hi}\right)$ where m_l^{lo} and m_l^{hi} can be found by solving Equation (1.24) for λ^* with $W_{l-1}^* = p - \frac{1-p}{k-l}$ and $W_{l-1}^* = p$, respectively and then by solving Equation (1.25) for μ with these W_{l-1}^* and λ^*. Continuity of R_l^{int} and Lemma 2.3 of Appendix 2 justify the following procedure:

(1) If $l \leq kp$: If $R_l^{int}(m_l^{lo}) < r$, then I_l^{int} is empty. Otherwise, the lower endpoint of I_i^l is m_l^{lo} and the upper endpoint of I_l^{int} is the root of $R_l^{int}(\mu) - r = 0$ on $\left(m_l^{lo}, m_l^{hi}\right)$.

(2) If $kp < l < kp+1$: The roots of $R_l^{int}(\mu) - r = 0$ on $\left(m_l^{lo}, T(F_k)\right)$ and $\left(T(F_k), m_l^{hi}\right)$ are the lower and upper endpoints of I_l^{int}.

(3) If $l \geq kp+1$: If $R_l^{int}(m_l^{hi}) < r$, then I_l^{int} is empty. Otherwise, the upper endpoint of I_i^l is m_l^{hi} and the lower endpoint of I_l^{int} is the root of $R_l^{int}(\mu) - r = 0$ on $\left(m_l^{lo}, m_l^{hi}\right)$.

Finally, since both I_l^{int} and I_l^{ψ} are intervals, we compute

$$CI_{1-\alpha} = \left(\bigcup_{l=1}^{k} I_l^{\psi}\right) \bigcup \left(\bigcup_{l=1}^{k} I_l^{int}\right) = \left[\mu^{lo}, \mu^{hi}\right]$$

by setting μ^{hi} equal to the maximum of the upper endpoints of I_l^{ψ} and of I_l^{int} and likewise by setting μ^{lo} equal to the minimum of the lower endpoints of I_l^{ψ} and I_l^{int}.

1.6. Experimental Results

We use the following two examples from [**30**] to test the performances of our confidence intervals and regions.

(1) **Put Option:** The owner of the portfolio has issued an in-the-money European put option and we use Monte Carlo simulation to estimate risk measures of this simple portfolio. The put option matures in 10 years with a strike price of \$110. The current stock price is \$100 and assumed to follow a lognormal return process with drift 8% and volatility 15%. The continuous discount rate is 6%.

(2) **Pareto Distribution:** The loss is assumed to have a Pareto distribution, whose tail behavior is similar to that observed in some applications. The Pareto distribution is tractable enough for obtaining closed form expressions for the variance of the CTE estimator. We use Monte Carlo simulation to estimate risk measures for losses generated by a heavy-tailed Pareto distribution with shape and scale parameters set to 2.5 and 25, respectively.

These are simple examples, but the results should be indicative of the coverage we would expect these procedures to provide for similar, larger examples. The simulations reported here do not use variance reduction. It is not straightforward to combine variance

reduction techniques, such as those applied to this problem by [**30**], with the methods for constructing confidence intervals and regions.

To evaluate the procedures for generating confidence intervals and regions, we run each of them 10,000 or 50,000 times. Each of these N macroreplications contains k simulated losses, where the sample size k is 500, 1,000, 2,000, or more in the experiments whose results are depicted in the following figures. From each macroreplication, we calculate one-sided and two-sided confidence intervals for $\mathrm{ES}_{0.95}$ and confidence regions for $\mathrm{VaR}_{0.95}$ and $\mathrm{CTE}_{0.95}$ at a nominal confidence level of 95% by the influence function, bootstrap, and empirical likelihood methods. We also calculate one-sided confidence intervals for $\mathrm{VaR}_{0.95}$ at a nominal confidence level of 95% by the binomial, influence function, and bootstrap methods. The number of bootstrap samples B we use is either 2,000 or 10,000. We compute, for each sample size k, the observed coverage probabilities of confidence intervals or regions

$$(1-\hat{\alpha}) := \#\{\text{confidence intervals or regions that include the true value}\}/N$$

where the true values are computed according to the formulae given by [**30**]. The coverage results for confidence intervals and regions are summarized in the following figures. The error bars in these figures represent 95% binomial confidence intervals for coverage probabilities based on observing N macroreplications, each of which is a success if the true value is included, a failure otherwise.

We first consider the example of selling a put option in the Black-Scholes model. We examine one-sided confidence intervals for VaR in Figure 1.2 to see how the methods under consideration differ in the well-studied setting of quantile estimation. As has been

documented by [**3**], the one-sided binomial confidence interval shows modest overcoverage for sample sizes between 500 and 2,000. The bootstrap and influence function methods show modest undercoverage, but attain coverage above 94% by sample size 4,000. Bootstrapping is slightly better than the influence function method at small sample sizes.

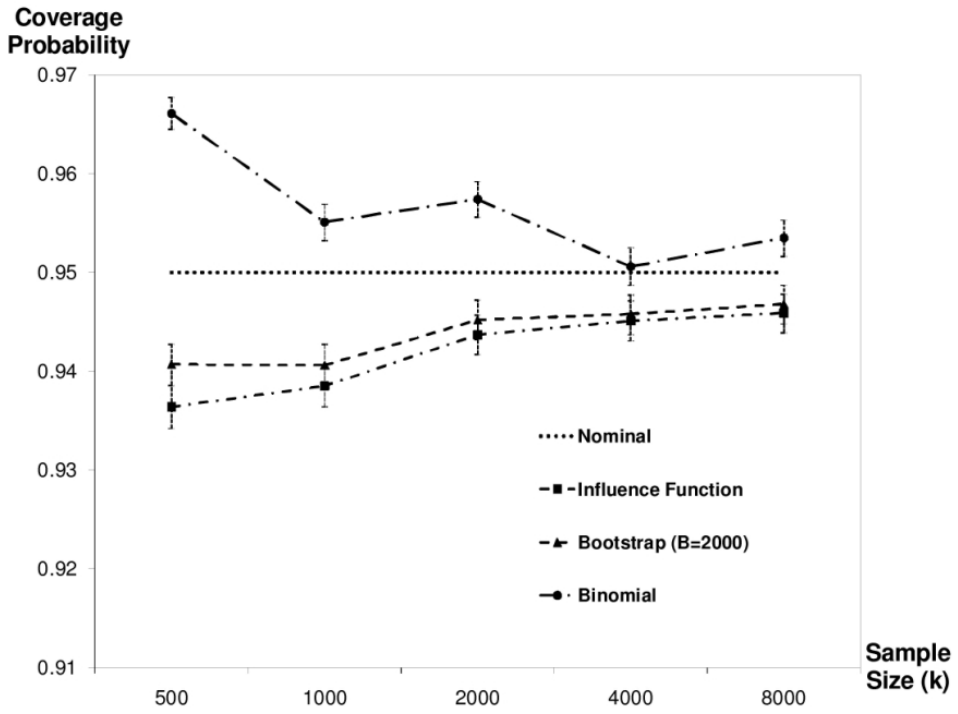

Figure 1.2. **(Put Option)** Coverage Probability of One-Sided Confidence Intervals for Value at Risk ($N = 50,000$ Macroreplications)

Next we turn to one-sided confidence intervals for ES in Figure 1.3. Again bootstrapping shows modest undercoverage, but for ES it attains nominal coverage by sample size 4,000. Empirical likelihood provides somewhat worse undercoverage until sample size

4,000. The influence function method has the worst undercoverage and has not attained nominal coverage even by sample size 8,000.

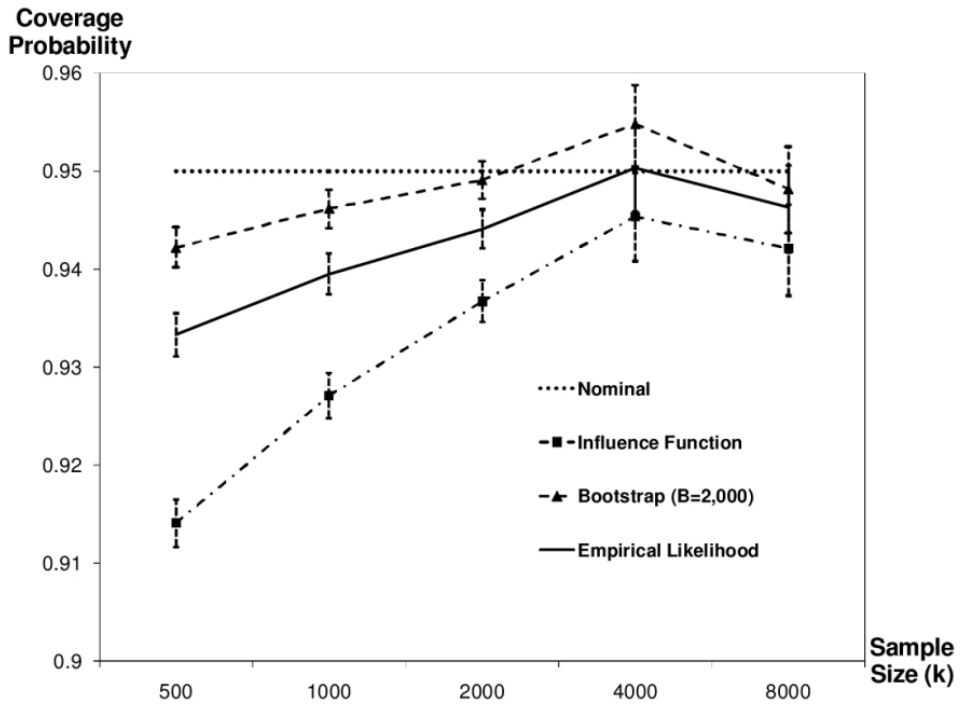

Figure 1.3. **(Put Option)** Coverage Probability of One-Sided Confidence Intervals for Expected Shortfall ($N = 50,000$ Macroreplications)

Figure 1.4 shows the coverage of two-sided confidence intervals for ES. The results are qualitatively similar to those for one-sided confidence intervals, but as usual, the two-sided confidence intervals have less undercoverage. This figure and the next also show that the bootstrap sample size $B = 2,000$ we use elsewhere is adequate: the improvement in coverage created by using a bootstrap sample size of $B = 10,000$ is negligible.

In Figure 1.5 we investigate the coverage of the confidence regions for VaR and CTE. The empirical likelihood method attains nominal coverage by sample size 2,000, while

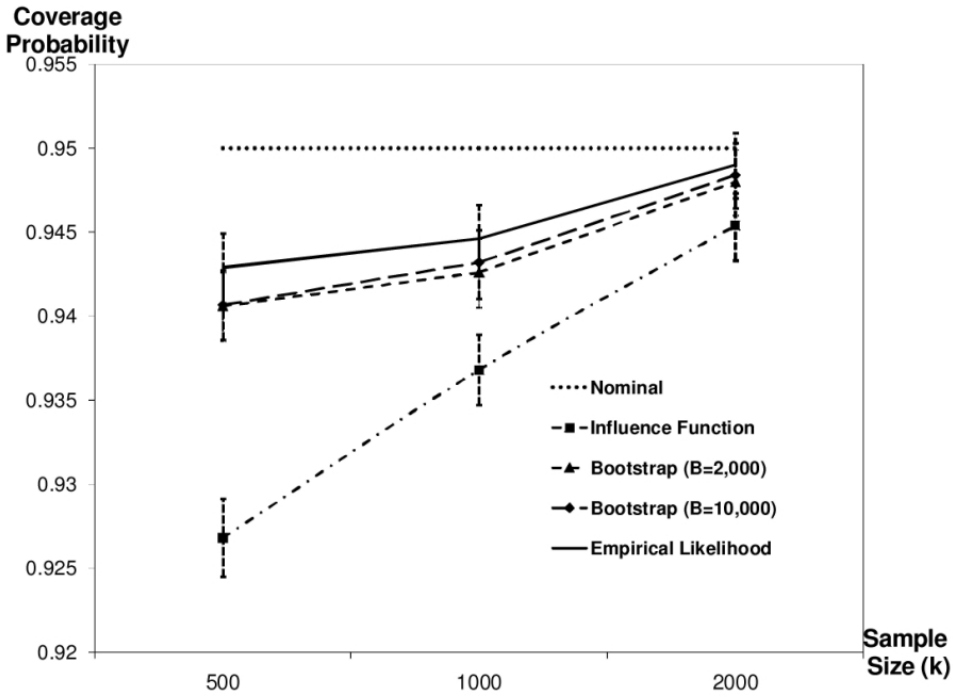

Figure 1.4. **(Put Option)** Coverage Probability of Two-Sided Confidence Intervals for Expected Shortfall ($N = 50,000$ Macroreplications)

the bootstrap and influence function methods produce disastrous undercoverage at these small sample sizes. We suspect that this deficiency is due to the difficulty of density estimation, resulting in poor covariance matrix estimates.

The subsequent figures portray the results of experiments on the Pareto distribution example, which serve to illustrate how well the methods perform when the loss distribution's tail is heavy instead of light. We focus on one-sided confidence intervals for ES in this example. Figure 1.6 shows that this example is much more challenging. All the methods produce severe undercoverage at small sample sizes, where bootstrapping is slightly better than empirical likelihood, which is in turn much better than the influence

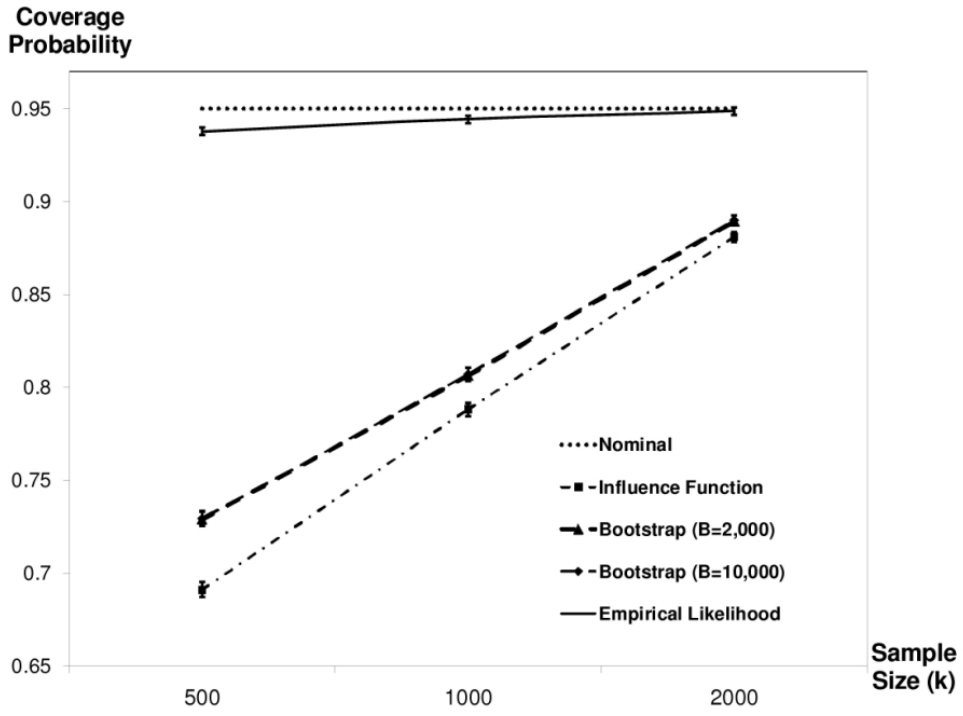

Figure 1.5. **(Put Option)** Coverage Probability of Confidence Regions for VaR and CTE ($N = 50,000$ Macroreplications)

function method. At large sample sizes, bootstrapping and empirical likelihood perform similarly. They still undercover somewhat even at a sample size of $k = 128,000$, but they are greatly superior to the influence function method.

Considering that confidence intervals fail to produce nearly nominal coverage even for very large sample sizes when the distribution is heavy-tailed, we investigate empirically how quickly the coverage converges to the nominal level. Figure 1.7 is a log-log plot of coverage error, defined as the absolute value of the difference between observed coverage and nominal coverage $|\hat{\alpha} - \alpha|$, against sample size k. For each method, the slope of the curve indicates how quickly the coverage converges to the nominal level. For example, the

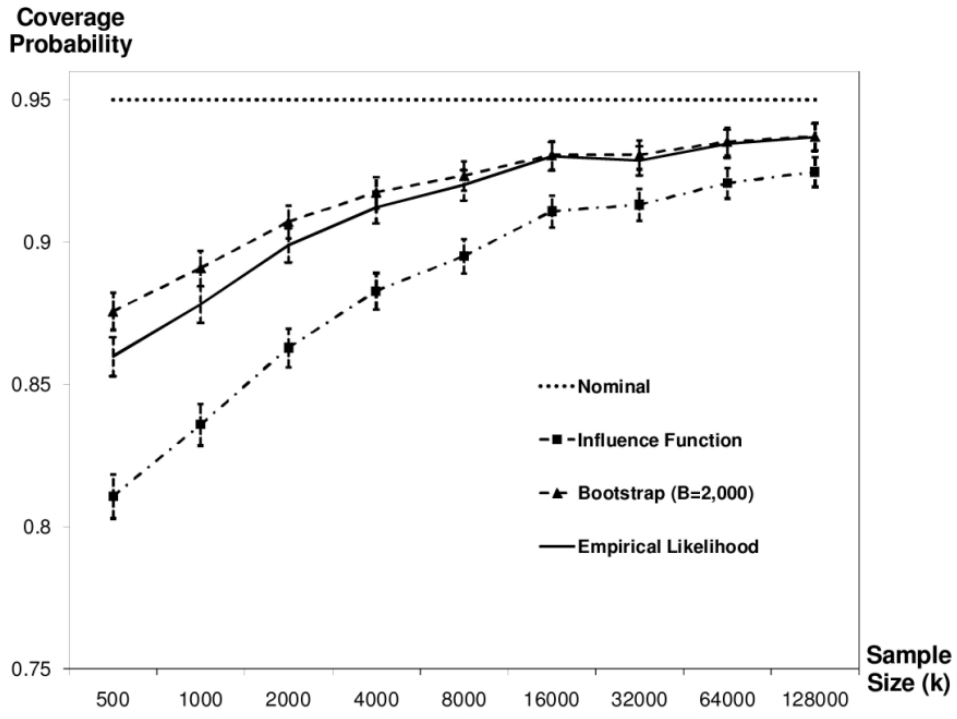

Figure 1.6. **(Pareto Distribution)** Coverage Probability of One-Sided Confidence Intervals for Expected Shortfall ($N = 10,000$ Macroreplications)

coverage error for one-sided confidence intervals is typically $\mathcal{O}(k^{-1/2})$ when produced by empirical likelihood (Owen 2001, Sec. 2.7), and $\mathcal{O}(k^{-1})$ when produced by the BCa bootstrapping method (Owen 2001, Sec. A.6). This implies that on a log-log plot of coverage error versus sample size, these methods should yield curves whose slopes approach -0.5 and -1, respectively, for large sample size. It is possible to correct empirical likelihood one-sided confidence intervals so that their coverage error is also $\mathcal{O}(k^{-1})$ (Owen 2001, Ch. 13).

However, far from finding that BCa bootstrapping dominates empirical likelihood asymptotically, we found that as sample size increases, the empirical likelihood method catches up with bootstrapping. Also, the influence function method becomes increasingly uncompetitive. We can see this in Figure 1.7, where we estimated slopes on the log-log plot of coverage error versus sample size of -0.34 for the influence function method, -0.42 for the empirical likelihood method, and -0.38 for bootstrapping, over a range of sample sizes from 500 to 128,000. The slope of -0.42 for empirical likelihood is not too far from the theoretical asymptotic slope of -0.5, but the slope of -0.38 for BCa bootstrapping is far from the typical theoretical asymptotic slope of -1. Of course, for finite sample sizes, the slope may differ from the asymptotic slope as sample size goes to infinity. We conjecture that there is another reason that the slope is far from -1 in Figure 1.7 for the coverage error of the BCa bootstrap one-sided confidence interval. In this example, the loss distribution is extremely heavy-tailed: the Pareto distribution with shape parameter 2.5 has first and second moments, but no third moment. Because the BCa method is based on a skewness correction, one would not expect it to work if skewness does not exist.

1.7. Conclusions and Future Research

Based on empirical likelihood, we have developed an asymptotically valid confidence interval for ES and confidence region for VaR and CTE. In Monte Carlo experiments, we found that they have coverage close to nominal for moderate sample sizes: about 1,000 samples in a financial example in which losses are light-tailed, and somewhat more in an example in which the loss distribution is Pareto. The confidence interval based on

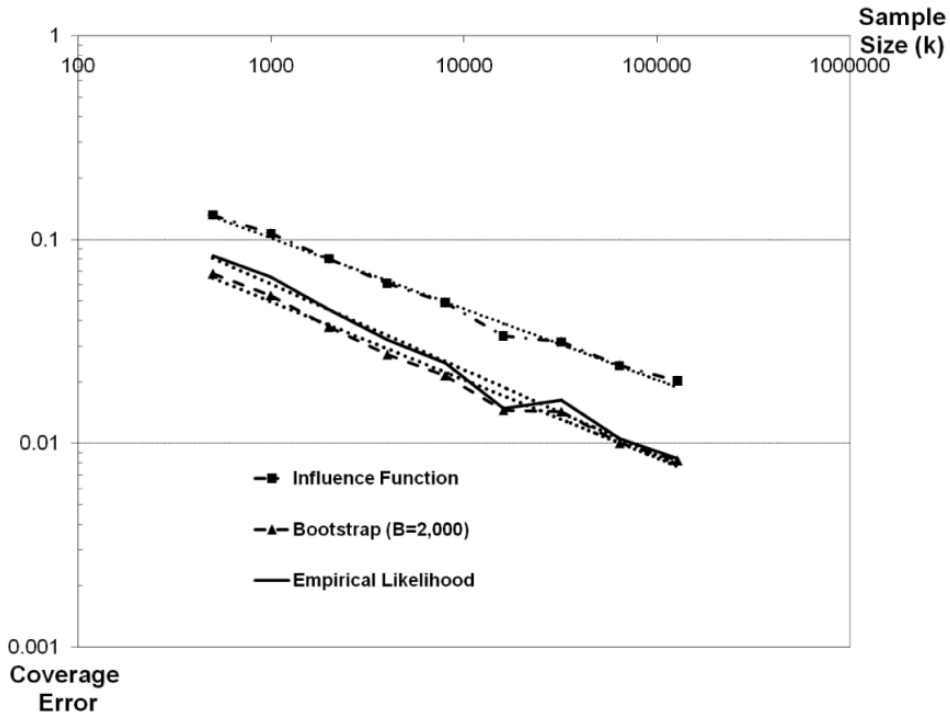

Figure 1.7. **(Pareto Distribution)** Log-Log Plot of Coverage Error of One-Sided 95% Confidence Intervals for Expected Shortfall vs. Sample Size ($N = 10,000$ Macroreplications)

empirical likelihood performed about as well as one based on bootstrapping and better than one based on the influence function. The confidence region based on empirical likelihood performed better than both its competitors.

The confidence intervals and regions discussed here are based on the most straightforward point estimators of VaR and CTE or ES. The most straightforward point estimator of VaR, which is a quantile, is a sample quantile. There is a large literature on quantile estimation which shows that more complicated estimators, such as kernel estimators and the Harrell-Davis estimator, can outperform the sample quantile [**11, 41**].

In this study, we have applied the basic version of empirical likelihood, but more advanced versions could be applied to the same problem. Methods such as Bartlett correction can improve the coverage of empirical likelihood confidence intervals (Owen 2001, Ch. 13). It has been found that smoothed or adjusted empirical likelihood methods can produce confidence intervals for quantiles with improved coverage [**12, 46**]. It is also possible to apply data tilting methods, which are generalizations of empirical likelihood, to construct confidence intervals for quantiles. [**35**] do this for extreme quantiles by explicitly estimating the tail index of the loss distribution. This method may also be applied to CTE or ES.

As suggested by [**17**], the techniques described here could be applied to any spectral measure of risk ([**2**]) as well as to ES. Another direction for future research is to show how to construct confidence intervals and regions when variance reduction techniques are used in the Monte Carlo sampling. This would yield smaller confidence intervals and regions given the same amount of computational effort.

CHAPTER 2

Response Surface Methodology for Simulating Hedging and Trading Strategies: A Simple Kriging Approach

2.1. Introduction

Suppose that one wishes to evaluate the distribution of profit and loss (P&L) resulting from a dynamic trading strategy. For example, one may be interested in a strategy that aims to make a profit by trading stocks and options, or in a hedging strategy that is intended to reduce the risk associated with selling over-the-counter derivative securities. One would like to know the distribution of P&L that results from a strategy so as to decide whether or not to adopt the strategy, or which of several rival strategies to implement.

Monte Carlo simulation is an appealing tool to use for this purpose: given a stochastic model of the market's dynamic behavior, one can simulate multiple scenarios for the market's behavior, determine the strategy's P&L in each scenario, and consider the resulting empirical distribution of P&L over all scenarios. Because of our interest in dynamic trading strategies, each scenario takes the form of a time series, a sequence of snapshots of the market's state at successive moments in time, which we will call a *path*. Given a path, we need to be able to compute the strategy's P&L along the path. This may require computation of the portfolio weights that the strategy chooses at each time and the values of financial securities at those times.

For some problems, those computations are easy, in which case the simulation approach is relatively straightforward (§2.4.1). The motivation for this article is the case in which substantial computational effort is required to approximate the portfolio weights or security values, and the accuracy of the approximations is proportional to the computational effort expended. This happens frequently in practice. For example, it happens when the portfolio weights are hedge ratios for which no formula exists. We will focus on the use of simulation to compute portfolio weights and security values, but the framework we propose in this chapter is applicable even if they are computed by other methods such as trees or numerical solution of partial differential equations.

The absence of formulae for portfolio weights and security values leads to a two-level simulation procedure (§2.4.2) which is used in practice: at the *outer* level, simulate paths; at the *inner* level, use simulation to compute portfolio weights and security values at every time step on every path. The inner-level simulation typically involves simulating many terminal payoffs of relevant securities. We refer to the Monte Carlo samples used to generate these payoffs as *pricing replications*. It may be necessary to use tens of thousands of pricing replications at every time step on every path, and thousands of paths, to attain high accuracy. As there are often dozens to thousands of times at which the portfolio is rebalanced, the total number of pricing replications required can be many millions or even billions. Thus, the disadvantage of this method, which we call "full simulation," is that it can be very slow. Our goal is to create a more efficient simulation procedure.

The central insight is that the full simulation method described above does too much work in pricing replications. For example, suppose that the problem involves hedging an option on a single stock using that stock and a riskless money market account (as in §2.3).

When the full simulation method needs to compute portfolio weights at time step 3 given a stock price of \$100, it does so using a full set of entirely new pricing replications, even if the procedure had previously computed portfolio weights on a different path in which the stock price was \$100.10 at time step 3. For many securities—although not for all securities, such as barrier options, whose value functions can be discontinuous—it seems that we should be able to use information from nearby paths to reduce the computational burden. The same insight underlies the use of regression Monte Carlo methods in pricing American options (see, e.g., [**7**], §8.6). Those methods use regression to approximate the option's value as a function of underlying state variables, such as the stock price in our example, for each potential exercise time separately. In our context, where there may be many rebalancing times, we expand on this idea by modeling security values and portfolio weights as functions of time and other state variables jointly.

We apply response surface methodology to each security value and portfolio weight that we need to know. Each of them is modeled as a function of time and state variables. We call a combination of time and state variables (such as stock price) a *point*. Response surface methodology in simulation works by performing simulations that estimate the response only at certain *design points*, and then attempting to infer the value of the response at other points. Our simulation procedure based on this idea (§2.4.3) first performs pricing replications at perhaps a few hundred design points, then constructs response surfaces, and finally uses the response surfaces to approximate portfolio weights and security values at thousands of points.

In this chapter, we show that response surface methodology can yield a more efficient simulation procedure for evaluating hedging and trading strategies. The response surface

methodology we apply to this problem is kriging and the experimental design we use for choosing design points is based on a Latin hypercube, with some modifications due to the structure of this financial problem. These issues are discussed in more detail in §2.4. First we give a formal statement of the problem and present our notation in §2.2 and describe a simple example in §2.3. Finally, we present results of computer experiments in §2.5.

2.2. Problem Formulation

Let Π be a random variable representing the strategy's terminal P&L. We focus on learning the univariate distribution F_Π of terminal P&L, although our method is also applicable to studying the joint distribution of P&L at various times. We do not know F_Π or how to sample from it directly; we only know how to sample from the distribution of paths: the outer level of simulation. P&L is some function of path, but we do not know this function either. P&L is given by Equation (2.1) below, but to use it, we need to compute security values and portfolio weights, which are also not known as functions of the path. We do know how to estimate these quantities by an inner level of simulation. One way to think of the situation is that we can approximately sample from F_Π, by sampling a path and then doing high-precision inner-level simulation to approximate the P&L on this path. [**29**] have discussed a rigorous framework for understanding two-level simulation and shown how to get confidence intervals from it for functionals of F_Π. Another recent contribution to two-level simulation in financial risk analysis is [**23**]. Here we will focus on point estimation of the mean and variance of F_Π, corresponding to evaluation of the strategy's average profitability and its risk.

We will use the following notation:

- T: the time horizon over which the strategy is to be evaluated.
- s: the number of times at which the portfolio can change.
- $t_0, t_1, \ldots, t_s$: times at which the portfolio is analyzed. A portfolio is set up at $t_0 = 0$, final P&L is measured at $t_s = T$, and $t_1, t_2, \ldots, t_{s-1}$ are the times at which the portfolio is rebalanced.
- $\mathbf{S}(t)$: the state vector of market risk factors at time t.
- $\mathbf{S}_i = \mathbf{S}(t_i)$: condensed notation for the state vector of market risk factors at time t_i.
- $\mathbf{V}_i$: the vector of values of the securities at time t_i.
- $\mathbf{f}$: the vector function of security values, so that the value of security l at time i is $V_{il} = f_l(t_i, \mathbf{S}_i)$.
- θ_i: the vector of portfolio weights at time t_i.
- $\mathbf{g}$: the vector function of portfolio weights, so that the number of shares of security l in the portfolio at time i is $\theta_{il} = g_l(t_i, \mathbf{S}_i)$.
- Π_i: the cumulative P&L up to time t_i.
- k: the number of paths simulated.
- m: the number of replications used in inner-level simulation at any point.
- A superscript indicates the realization of a random variable on a particular path: for example, $\mathbf{S}_i^j$ is the vector of market risk factors at time t_i on path j, and $\mathbf{S}^j$ is the jth sample path of the discrete-time vector Markov process $\mathbf{S}$.

The P&L that occurs over step i, that is, between times t_{i-1} and t_i, is the sum of the changes in value of the portfolio's holdings in each security. The number of shares of each security remains constant over this time step, so step i's contribution to P&L is

$\theta_{i-1}^\top(\mathbf{V}_i - \mathbf{V}_{i-1})$. The cumulative P&L up to time t_i is thus

$$\Pi_i = \sum_{i'=1}^{i} \theta_{i'-1}^\top(\mathbf{V}_{i'} - \mathbf{V}_{i'-1}). \tag{2.1}$$

In mathematical finance, it is standard to consider *self-financing* strategies, which have the property that the portfolio's value does not change when the portfolio is rebalanced:

$$\theta_{i-1}^\top \mathbf{V}_i = \theta_i^\top \mathbf{V}_i \tag{2.2}$$

for all $i = 1, 2, \ldots, s$. For self-financing strategies, P&L also equals the change in the portfolio's value across all steps to date, which is

$$\Pi_i = \theta_i^\top \mathbf{V}_i - \theta_0^\top \mathbf{V}_0. \tag{2.3}$$

To estimate the P&L of a strategy at times $t_0 = 0, t_1, \ldots, t_s = T$, we simulate a path of market risk factors $\mathbf{S}_0, \mathbf{S}_1, \ldots, \mathbf{S}_s$ at each of these times. We need to estimate security values $\mathbf{V}_i$ and portfolio weights θ_i at each time $i = 0, 1, \ldots, s$ to use Equation (2.1) to find P&L at each time. Section 2.3 describes some simplifications that are possible in a typical application of hedging a derivative security. We simulate k paths to get a sample $\Pi^1, \Pi^2, \ldots, \Pi^k$. Its sample average is an estimate of the mean terminal P&L, while its sample standard deviation is an estimate of the standard deviation of terminal P&L.

Estimation of security values and portfolio weights is founded on inner-level simulation. We assume that we are in the typical complete-markets, Markov-process setting in which each security's value at any time t is the conditional expectation of its payoff under

the risk-neutral probability measure given the value of $\mathbf{S}(t)$ (see, e.g.,[**31**], §5.8). For simplicity, we also assume that all securities have maturity $T = t_s$ and are path-independent. In this case, the inner-level simulation need only sample values of $\mathbf{S}(T) = \mathbf{S}_s$. Otherwise, the inner-level simulation must sample state vectors at the maturities of all securities. If securities are path-dependent, then the inner-level simulation must sample paths of the state vector observed at all relevant times.

Consider an inner-level simulation at some point $x = (t^*, \mathbf{S}^*)$ which is a combination of time t^* and state $\mathbf{S}^*$. The inner-level simulation samples *inner-level replications* $\mathbf{S}_s^1(x), \mathbf{S}_s^2(x), \ldots, \mathbf{S}_s^m(x)$ from the risk-neutral conditional distribution of $\mathbf{S}_s = \mathbf{S}(T)$ given $\mathbf{S}(t^*) = \mathbf{S}^*$. The terminal payoff function $g(T, \cdot)$ of all securities at maturity is known. This allows us to estimate the security value $g_l(t^*, \mathbf{S}^*)$ at time t^* by $\sum_{j=1}^m g_l(T, \mathbf{S}_s^j(x))/m$ for each security l ([**22**], §1.2).

Portfolio weights in hedging strategies often arise as sensitivities of security values to risk factors. See [**22**] Ch. 7 on Monte Carlo methods for estimating sensitivities such as delta. In general, the method we describe can be applied if the portfolio weights can be computed by some method founded on simulation: for example, if the portfolio weights arise from a portfolio optimization problem at each point, then they can be estimated by optimization-via-simulation, and the method applies.

We give a simple, concrete example of a hedging strategy in this framework in §2.3. In §2.4, we describe three different methods for estimating security values $\mathbf{V}_i$ and portfolio weights θ_i at each time $i = 0, 1, \ldots, s$.

2.3. Delta-Hedging A European Put Option

Our computational experiments feature the example of delta-hedging a European put option on a stock under the Black-Scholes model (see, e.g., [**31**], Ch. 6 for further background). Under the Black-Scholes model, the stock price S is geometric Brownian motion. The securities in the hedging portfolio include a riskless money-market account and the underlying stock. Their values at time t_i are respectively $V_{i1} = e^{rt_i}$ where r is the interest rate and $V_{i2} = S_i$. The put option (denoted as security 0) is hedged from the time it is sold until its maturity, when it pays off $V_{s0} = f_0(T, S_s) = \max\{K - S_s, 0\}$, where K is the strike price.

For $i = 0, 1, \ldots, s-1$, the number θ_{i2} of shares of stock to hold at time t_i is set equal to the negative of the first-order sensitivity $\partial V_{i0}/\partial S_i$ of the put option value with respect to the stock price at that point in time, which is called the *delta* of the option. At time $t_s = T$, the option matures and the hedge is unwound, so $\theta_{s2} = 0$. The number of shares in the money-market account is set so that the hedging strategy is self-financing:

$$\theta_{i1} = \theta_{i-1,1} + (\theta_{i-1,2} - \theta_{i2})V_{i2}/V_{i1} \tag{2.4}$$

for $i = 1, 2, \ldots, s$, based on Equation (2.2). The initial number of shares in the money-market account is set so that the value of the initial portfolio is zero:

$$\theta_{01} = -(V_{00} + \theta_{02}V_{02})/V_{01}. \tag{2.5}$$

At any time t_i, $i = 0, 1, \ldots, s$, there is $\theta_{i0} = 1$ share of the put option. Because the total initial portfolio value is zero, the final P&L according to Equation (2.3) is

$$\Pi_s = V_{s0} + \theta_{s1} V_{s1}. \tag{2.6}$$

Some simplifications of the general framework for P&L presented in §2.2 apply in many examples, including this one. Suppose that we have a self-financing strategy for hedging a derivative security until maturity, the derivative security's initial value is known, and we are only interested in terminal P&L. Then we can use Equation (2.3) and we need only compute portfolio weights, not security values. The reason is that underlying security values (e.g., of the stock and money market account) are known functions of the path, and the value of the derivative security (e.g., put option) is known at the times that we need it. We do not need to know the derivative security value at intermediate times, its initial value is given, and its terminal value is always a known function of the path, such as $\max\{K - S_s, 0\}$ in the example of the European put option. Even though only the initial and terminal portfolio weight vectors θ_0 and θ_s appear in Equation (2.3), we do need to know the portfolio weight vectors at intermediate times too. They are required to enforce the self-financing condition. For our put option example, we can use Equation (2.4) recursively with Equation (2.5) as its initial condition, use $V_{i1} = e^{rt_i}$ and $V_{i2} = S_i$, plug into Equation (2.6), and get

$$\Pi_s = V_{s0} + \sum_{i=1}^{s} (\theta_{i-1,2} - \theta_{i2}) S_i e^{r(T-t_i)} - (V_{00} + \theta_{02} S_0) e^{rT}.$$

This shows that we need to know delta at all steps $i = 0, 1, \ldots, s-1$, to get θ_{i2}, but the only option values we need are V_{00}, the given initial value, and V_{s0}, which is the payoff

$\max\{K - S_s, 0\}$. We use pathwise derivative estimation (Glasserman 2003, §7.2) for delta:

$$\hat{\theta}_{i2} = e^{-r(T-t_i)} \frac{1}{m} \sum_{j=1}^{m} \mathbf{1}\{S_s^j < K\} \frac{S_s^j}{S_i^j}.$$

In the particular example we use in our computational experiments, the stock price follows a geometric Brownian motion with initial value $S_0 = \$100$, drift $\mu = 8\%$, and volatility $\sigma = 15\%$. The put option has maturity of $T = 1$ years and strike price $K = \$110$. The interest rate on the money-market account is $r = 5\%$. There are $s = 60$ rebalancing times, and $t_i = iT/s$ for $i = 1, 2, \ldots, s$.

2.4. Methods

In this section, we describe three simulation procedures used in our experimental results. The procedures differ only in how they estimate security values and portfolio weights at times after t_0: by using a formula, by using a nested simulation at every time step on every path, or by using response surface modeling. We assume that security values and portfolio weights are known at t_0 even in procedures that do not rely on formulae for them. In practice, even if security values and portfolio weights at t_0 need to be estimated by simulation, they have usually been estimated to very high precision; it is only security values and portfolio weights in hypothetical scenarios at future times that we can not afford to estimate to high precision.

For similar reasons, we regard the P&L distribution produced by the formula-based procedure (§ 2.4.2) as the true P&L distribution. Even if formulae are not available, the decision-maker will estimate $\mathbf{f}(t_i, \mathbf{S}_i)$ and $\mathbf{g}(t_i, \mathbf{S}_i)$ to high precision when in state $\mathbf{S}_i$ at time t_i. The actual P&L arising on a path $\mathbf{S}_1, \mathbf{S}_2, \ldots, \mathbf{S}_s$ will be nearly the P&L assigned

to it by the formula-based procedure, which is the limit of the P&L assigned to it by the full nested simulation procedure (§ 2.4.1) as the number m of inner-level replications goes to infinity.

2.4.1. Formula-Based

The example described in §2.3 is so simple that we actually do have formulae for the put option's value and delta as functions of time and stock price (see, e.g., [**31**], §6.6). This means that the security values and portfolio weights are known as functions of the path. By using these functions, we can avoid any inner-level simulation and eliminate all associated statistical error. In general, these functions are unknown, but using them in this simple example allows us to study the statistical error associated with inner-level simulation in one of the following two methods. The formula-based procedure for sampling P&L Π_i^j for every step $i = 1, 2, \ldots, s$ and on paths $j = 1, 2, \ldots, k$ is as follows.

- for $j = 1, 2, \ldots, k$,
  - $\Pi_0^j = 0$
  - for $i = 1, 2, \ldots, s$,
    - sample $\mathbf{S}_i^j$ from the conditional distribution of $\mathbf{S}_i$ given $\mathbf{S}_{i-1} = \mathbf{S}_{i-1}^j$
    - $\theta_i^j = \mathbf{g}(t_i, \mathbf{S}_i^j)$ and $\mathbf{V}_i^j = \mathbf{f}(t_i, \mathbf{S}_i^j)$
    - $\Pi_i^j = \Pi_{i-1}^j + (\theta_{i-1}^j)^\top (\mathbf{V}_i^j - \mathbf{V}_{i-1}^j)$
- next j

2.4.2. Full Nested Simulation

In the absence of formulae for these functions, we require estimates of security values and portfolio weights at each of s time steps on each of k paths. One way to get them is from fully nested simulation: at each of ks points, do an inner-level simulation to estimate security values and portfolio weights there. For simplicity, we suppose there are the same number m of inner-level replications at all points. The full nested simulation procedure is as follows.

- for $j = 1, 2, \ldots, k$,
  - $\Pi_0^j = 0$
  - for $i = 1, 2, \ldots, s$,
    - sample $\mathbf{S}_i^j$ from the conditional distribution of $\mathbf{S}_i$ given $\mathbf{S}_{i-1} = \mathbf{S}_{i-1}^j$
    - for $h = 1, 2, \ldots, m$,
      - sample $\mathbf{S}_s^h(t_i, \mathbf{S}_i^j)$ from the conditional risk-neutral distribution of $\mathbf{S}_s$ given $\mathbf{S}_i = \mathbf{S}_i^j$
    - using this inner-level simulation, estimate θ_i^j by $\hat{\theta}_i^j$ and $\mathbf{V}_i^j$ by $\hat{\mathbf{V}}_i^j$
    - $\Pi_i^j = \Pi_{i-1}^j + (\hat{\theta}_{i-1}^j)^\top (\hat{\mathbf{V}}_i^j - \hat{\mathbf{V}}_{i-1}^j)$
- next j

2.4.3. Response Surface Modeling

Instead of performing inner-level simulation at each of the ks points at which we want to know security values and portfolio weights, this method reads estimates at each of those points from functions called *response surfaces*. These response surfaces are in turn computed as the result of inner-level simulation, which is performed only at certain *design*

points. [**20**] and [**40**] have also applied response surface modeling in financial risk analysis. In our experiments, we use a response surface methodology called *kriging*, which chooses the value of a response surface at any point that is not among the design points by interpolating among the values observed at design points. Figure 2.1 shows a response surface for the value of the put option constructed by kriging based on inner-level simulation, and the error due to kriging in estimating the response surface given by the Black-Scholes formula in the example of §2.3. On the kriging method and its use in the design and analysis of computer experiments see, e.g., [**19**], [**38**], and [**44**].

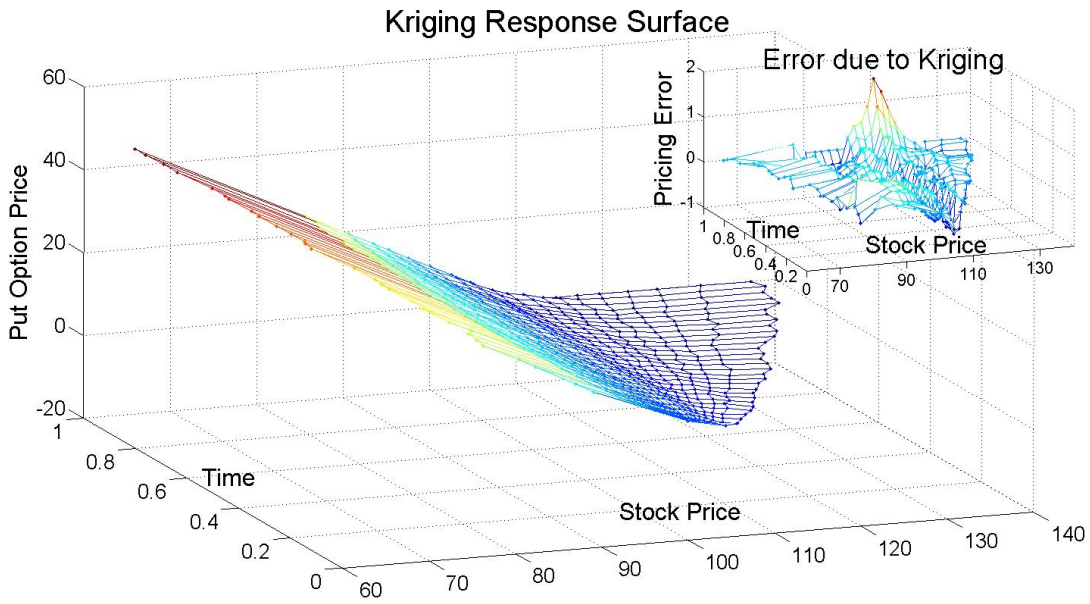

Figure 2.1. A response surface estimated by kriging, and its difference from the Black-Scholes formula, for the price of the put option in §2.3.

By choosing a set of $n \ll ks$ design points, the response surface modeling method can be faster than full nested simulation with the same numbers k of paths, s of time steps, and m of inner-level replications per point. This speed can come at the price of

accuracy in assessing the distribution of P&L, because the response surface models of security values and portfolio weights at a point $(t_i, \mathbf{S}^*)$ tend not to be as accurate as estimates provided by inner-level simulation performed conditional on $\mathbf{S}_i = \mathbf{S}^*$; that is, interpolating estimates from other points is not as accurate as simulating at this point. The goal is to trade off a little bit of accuracy for a lot of speed to get a more efficient procedure.

Our response surface modeling procedure has three parts. First, we choose a set $\{x_1, x_2, \ldots, x_n\}$ of design points and perform inner-level simulation at each of them to estimate the security values and portfolio weights there. Second, use kriging to build response surfaces for each component, separately, of the security value function $\mathbf{f}$ and the portfolio weight function $\mathbf{g}$. Third, compute P&L on each path by using the response surfaces to provide estimates of security values and portfolio weights.

The choice of experimental design (the set of design points) can have a very large impact on the fidelity of the response surface generated by kriging. In most applications of kriging, one knows in advance a finite region within which one wishes to estimate a response. However, in the financial examples we consider here, this is not so: for example, asset prices are typically unbounded. We solve this problem by first simulating all the paths we need, and then modeling responses over a finite region that contains all the simulated data. In the example of §3, this region is contained in the Cartesian product $[t_0, t_{s-1}] \times [S_{\min}, S_{\max}]$ where $S_{\min} = \min\{S_i^j | j = 1, 2, \ldots, k,\ i = 1, 2, \ldots, s\}$ and $S_{\max} = \max\{S_i^j | j = 1, 2, \ldots, k,\ i = 1, 2, \ldots, s\}$ are the smallest and largest observed stock prices. The reason that $t_s = T$ is excluded from response surface modeling is twofold. We do not need a response surface model at maturity because security values

then are payoffs, which are known functions, and likewise portfolio weights then do not need to be estimated. There is also a discontinuity at maturity in the put option's delta, and hence in g_2, the number of shares of stock to hold: $\lim_{S\uparrow K} g_2(T,S) = 1$ while $\lim_{S\downarrow K} g_2(T,S) = 0$. In general, response surface modeling can encounter severe problems when the true function is discontinuous.

Latin hypercube designs (see, e.g., [**22**], §4.4) have been reported to be more effective in kriging than uniform grid designs; our experiments confirmed this. Our experimental design for the example of §2.3, based on a Latin hypercube, is constructed as follows.

- Partition $[t_0, t_{s-1}]$ and $[S_{\min}, S_{\max}]$ into d intervals of equal width.
- Sample one point uniformly within each interval.
- Randomly pair the time and stock price values to get d design points.
- Add the four points $(t_0, S_{\min})$, $(t_0, S_{\max})$, $(t_{s-1}, S_{\min})$ and $(t_{s-1}, S_{\max})$ to the design.

The last step is important: without the addition of the four corner points there can be a point (t_i, S_i^j) on some path that falls outside the convex hull of the Latin hypercube design. This causes problems because kriging, as an interpolation method, performs quite badly when it is used to extrapolate outside the convex hull of design points. An illustration of this experimental design appears in Figure 2.2.

We feed estimated security values and portfolio weights at the design points to kriging, which constructs response surfaces by interpolating among them. The interpolated value at a point x is a weighted average of the values observed at all design points. The weight a design point x_j receives depends on its distance to x; the basic idea behind kriging is that nearer design points are more highly correlated (in a loose sense) with the point of interest,

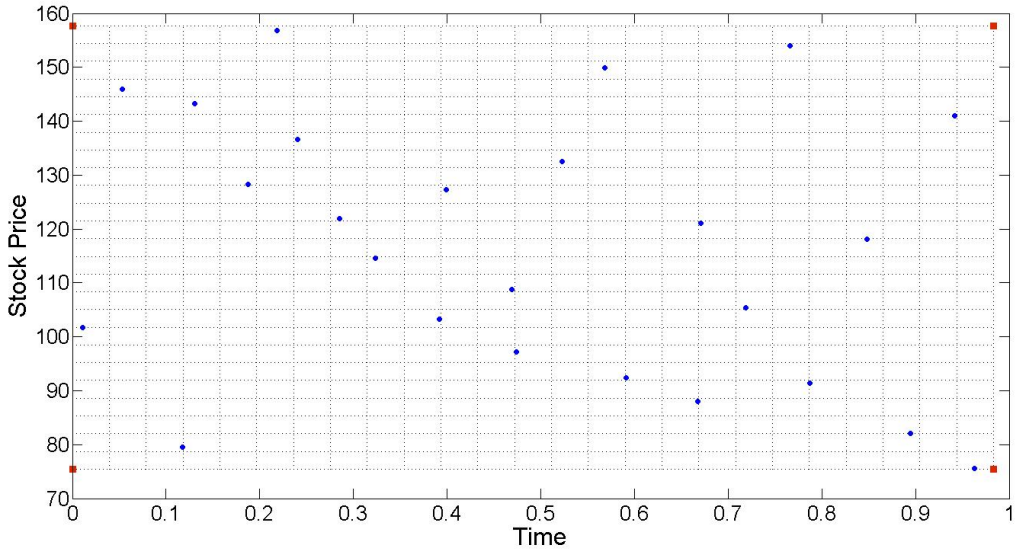

Figure 2.2. An experimental design for the example of §2.3 with 29 design points. The 4 red squares are corner points and the 25 blue dots form a Latin hypercube design.

and so get greater weight. The kriging model requires a choice of a correlation function that governs how weights diminish with distance. We used the exponential correlation function.

The response surface modeling procedure is as follows.

- for $j = 1, 2, \ldots, k$,
  - for $i = 1, 2, \ldots, s$,
    - sample $\mathbf{S}_i^j$ from the conditional distribution of $\mathbf{S}_i$ given $\mathbf{S}_{i-1} = \mathbf{S}_{i-1}^j$
  - next i
- lay down a design of points $x_1 = (t_1^*, \mathbf{S}_1^*)$, $x_2 = (t_2^*, \mathbf{S}_2^*)$, $\ldots$, $x_n = (t_n^*, \mathbf{S}_n^*)$
- for $j = 1, 2, \ldots, n$,
  - for $h = 1, 2, \ldots, m$,

        * sample $\mathbf{S}_s^h(x_j)$ from the conditonal risk-neutral distribution of $\mathbf{S}_s$ given $\mathbf{S}(t_j^*) = \mathbf{S}_j^*$
    - using this inner-level simulation, estimate $\theta(x_j)$ by $\hat{\theta}(x_j)$ and $\mathbf{V}(x_j)$ by $\hat{\mathbf{V}}(x_j)$
- next j
- for each security l,
    - build response surfaces $\hat{f}_l$ for f_l using $\hat{V}_l(x_1), \hat{V}_l(x_2), \ldots, \hat{V}_l(x_n)$ and $\hat{g}_l$ for g_l using $\hat{\theta}_l(x_1), \hat{\theta}_l(x_2), \ldots, \hat{\theta}_l(x_n)$
- next l
- for $j = 1, 2, \ldots, k$,
    - $\Pi_0^j = 0$
    - for $i = 1, 2, \ldots, s$,
        * $\Pi_i^j = \Pi_{i-1}^j + (\hat{\mathbf{g}}(t_i, \mathbf{S}_{i-1}^j)^\top (\hat{\mathbf{f}}(t_i, \mathbf{S}_i^j) - \hat{\mathbf{f}}(t_i, \mathbf{S}_{i-1}^j))$
- next j

2.5. Experiments

We performed experiments in MATLAB, implementing kriging with the `DACE` toolbox and generating Latin hypercube designs with the `lhsdesign` function. We used $k = 1,000$ paths and $m = 1,000$ inner-level replications in the example described in §2.3. Our experimental designs had $n = 104$ or $n = 404$ design points. This means that the full nested simulation procedure required 60 million inner-level replications while the response surface modeling procedure required only 104,000 or 404,000. Of course, performing kriging takes time too. However, the relative effort required for inner-level replications, sampling paths, and performing kriging varies with the problem and the

computing platform. Because we expect that the time spent on inner-level replications will dominate unless m is small and the stochastic model of the market is simple, we ignore the other computational costs.

To assess the accuracy of these simulation procedures, we ran 100 macro-replications, each an independent run of the entire procedure, involving independently generated paths and design points. However, within a single macro-replication, all procedures used the same paths. This produced 100 estimates of the mean P&L and the standard deviation of P&L at each time step, for each procedure: that is, each of 100 runs produces an estimate of $\mathbf{E}[\Pi_i]$ and an estimate of $\sqrt{\mathbf{Var}[\Pi_i]}$ for each time step $i = 0, 1, \ldots, s$. We depict the average and the sample root mean squared error (RMSE) of these 100 estimates in Figures 2.3–2.6. The error used in computing RMSE is the difference between one macro-replication's estimate of $\mathbf{E}[\Pi_i]$ or $\sqrt{\mathbf{Var}[\Pi_i]}$ and our best estimate of $\mathbf{E}[\Pi_i]$ or $\sqrt{\mathbf{Var}[\Pi_i]}$, which we get by combining all 100,000 replications generated by the formula-based procedure.

It is a feature of the financial example, not a problem with the simulation procedure, that the mean P&L $\mathbf{E}[\Pi_i]$ is decreasing in the time step i for the formula-based procedure in Figure 2.3. Because of the option's convexity, the discrete-time hedging strategy that is delta-neutral at the beginning of each time step holds too little stock, on average, between portfolio re-balancing times. Because the stock's expected return exceeds the interest rate, holding too little stock gives the strategy a negative expected return.

In Figure 2.4, even the formula-based procedure yields an estimate with positive RMSE because of outer-level sampling error: each macro-replication has a different set of $k = 1,000$ paths.

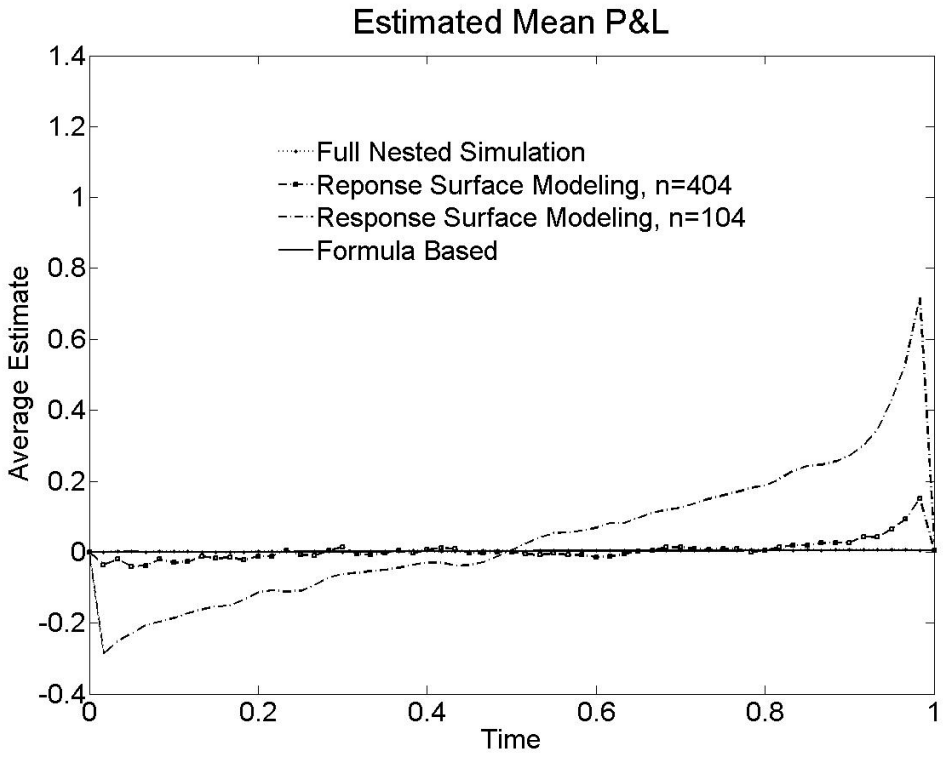

Figure 2.3. The average, over 100 macro-replications, of four simulation methods' estimates of mean P&L at each time step.

Figures 2.3 and 2.4 show that response surface modeling can introduce substantial bias and variance into estimating mean P&L, but the problem is ameliorated by using a moderate number of design points. With $n = 104$ design points, kriging tends to generate a response surface with poor fidelity. Particularly noticeable is the bias just before maturity, where the put option's price and delta are badly behaved functions, as explained in §2.3. Increasing the number of design points to $n = 404$ greatly reduces this bias. With 404 design points, response surface modeling estimates mean P&L with a fair degree of accuracy, compared to the standard deviation of P&L portrayed in Figure 2.5. The bias and variance introduced by response surface modeling are much greater at time

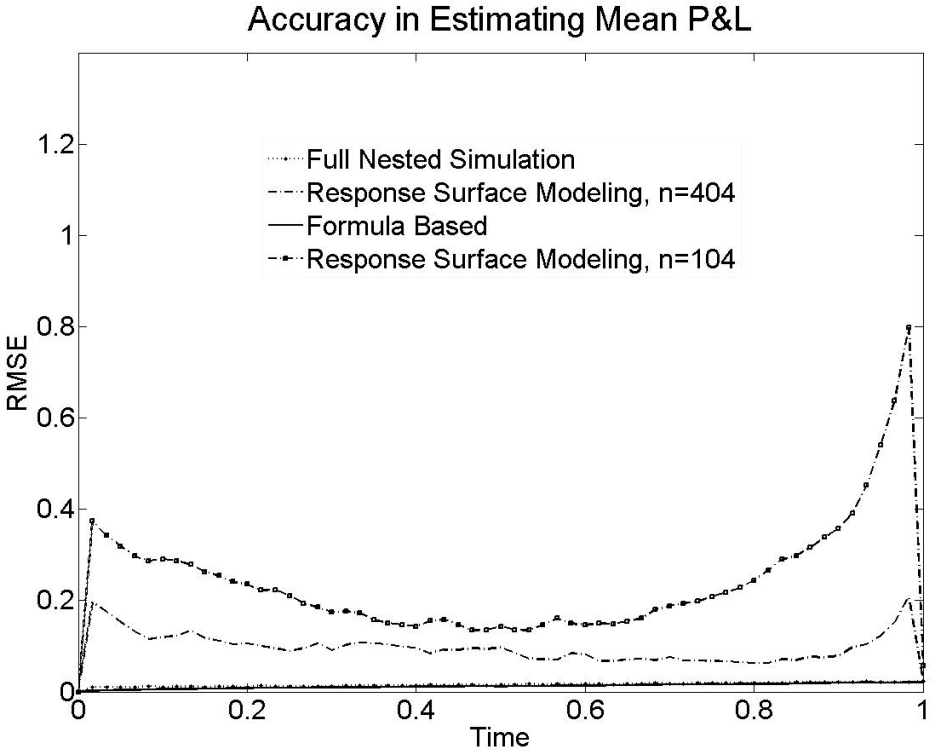

Figure 2.4. The sample root mean squared error, over 100 macro-replications, of four simulation methods' estimates of mean P&L at each time step.

t_{s-1} than at time $t_s = T$ and at time t_1 than at time $t_0 = 0$ (where P&L is zero by construction). This happens because the simulation procedure relies on the response surface for the put option price to produce P&L at times t_{s-1} and t_1 but not at times t_s and t_0.

Figures 2.5 and 2.6 show that response surface modeling with $n = 404$ design points performs similarly to the other procedures in estimating the standard deviation of P&L, except at early times, when it is not as accurate as the formula-based procedure, but more accurate than full nested simulation. Again, the jump from time t_0 to time t_1 in RMSE of

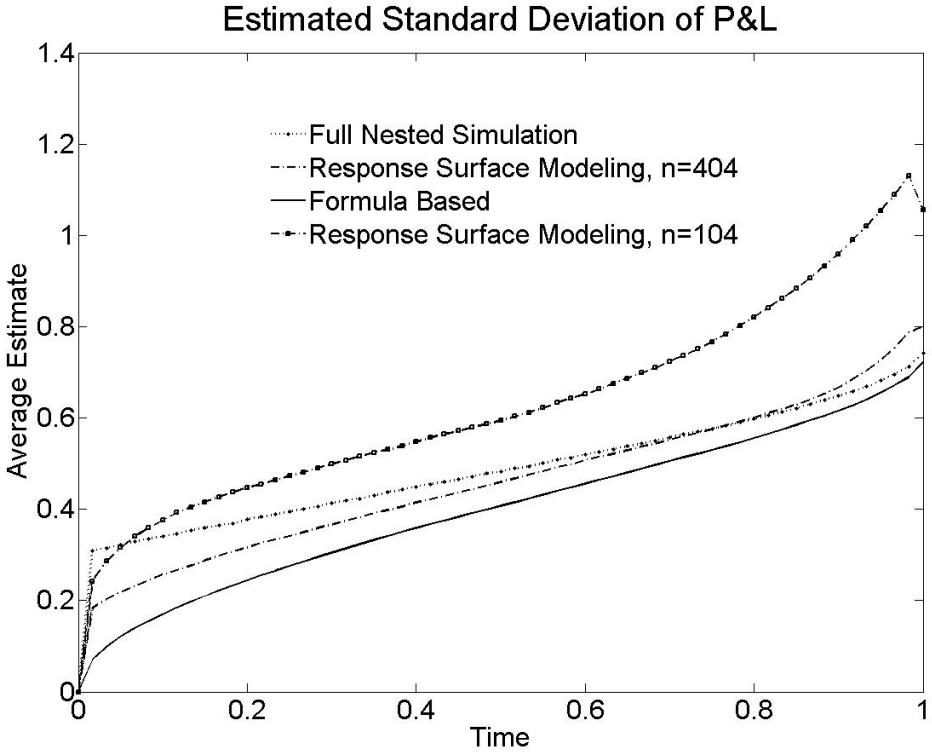

Figure 2.5. The average, over 100 macro-replications, of four simulation methods' estimates of the standard deviation of P&L at each time step.

estimates from methods other than the formula-based method is explained by inner-level sampling error in estimating the put option value function $f_0(t_1, \cdot)$.

Next we focus on the distribution of terminal P&L generated by each of the simulation methods. We pool the terminal P&L values on $k = 1,000$ paths in each of 100 macro-replications to produce a picture of the probability density of terminal P&L in Figure 2.7. Table 2.1 summarizes the results in Figures 2.3–2.6 for P&L at time $T = 1$ only. It also shows the standard errors in estimating these quantities with 100 macro-replications. The standard error in estimating RMSE is calculated by the delta method (see, e.g., [**7**], §III.3). From the table we see that the response surface modeling procedure with $n = 404$

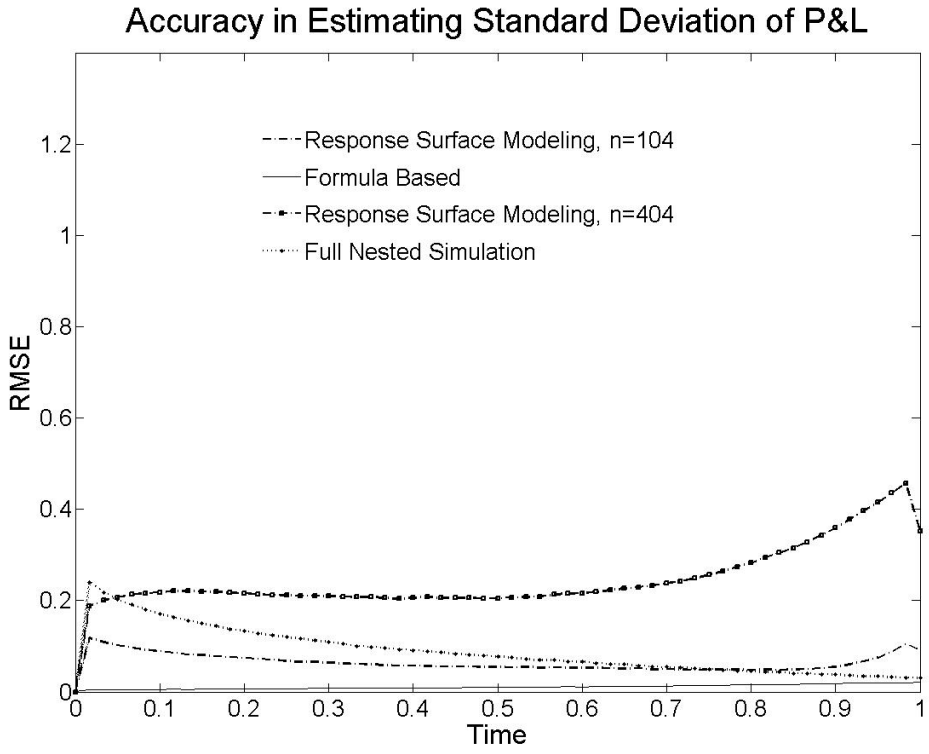

Figure 2.6. The sample root mean squared error, over 100 macro-replications, of four simulation methods' estimates of the standard deviation of P&L at each time step.

design points is quite accurate in estimating the mean and standard deviation of terminal P&L, compared to full nested simulation, but it is on the order of 100 times faster.

2.6. Conclusions and Future Research

We have found, in exploring one example that using kriging to create response surfaces for the price and delta of a put option enables fast, accurate estimation of the mean and standard deviation of the terminal P&L of a hedging strategy for the put option. This suggests that kriging can be an effective tool for reducing the computational cost of nested simulations of hedging and trading strategies. However, the method's performance needs

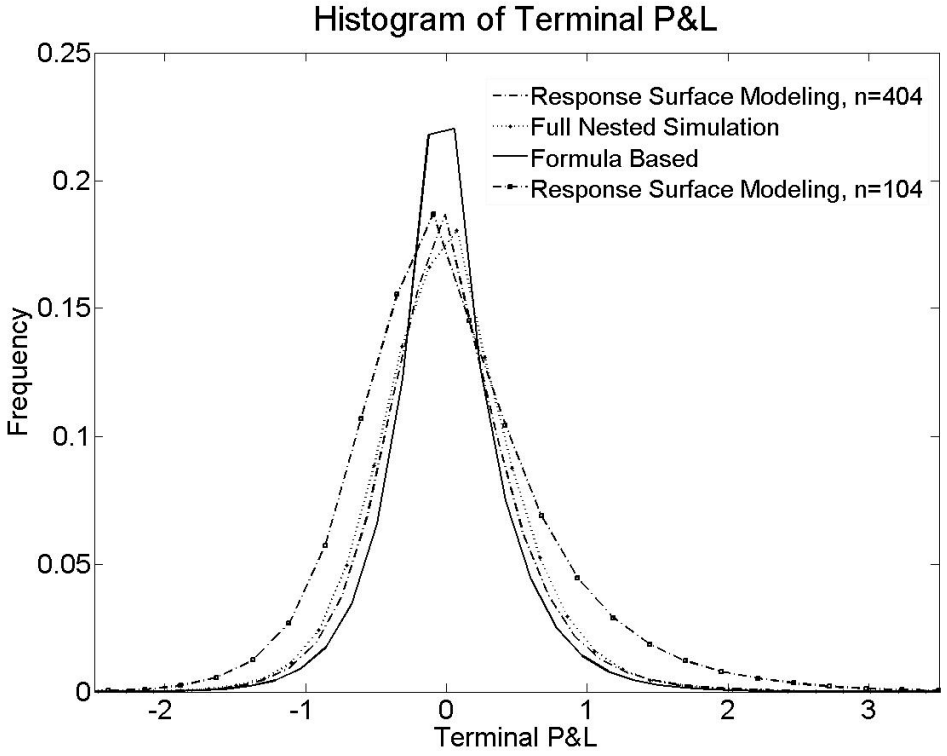

Figure 2.7. Approximate probability density functions of terminal P&L generated by four simulation methods, based on $100,000$ samples of terminal P&L.

Table 2.1. The average, standard error (SE) and the root mean square error (RMSE) statistics of the mean estimates of terminal P&L.

	Mean P&L		Std. Dev. of P&L	
Method	Average	RMSE	Average	RMSE
Formula	-0.030	0.0573	0.792	0.6136
Based	*(0.006)*	*(0.007)*	*(0.051)*	*(0.035)*
Full Nested	0.017	0.0651	0.854	0.6152
Simulation	*(0.004)*	*(0.005)*	*(0.047)*	*(0.037)*
RSM, n=104	0.037	0.1748	0.970	0.6895
	(0.016)	*(0.030)*	*(0.046)*	*(0.036)*
RSM, n=404	0.013	0.1231	1.014	0.9429
	(0.012)	*(0.012)*	*(0.076)*	*(0.080)*

to be investigated in other examples, especially examples in which response surfaces are built over a higher-dimensional space. Kriging in higher dimensions is more challenging, and the example we used here has only two dimensions, time and stock price.

There are also possibilities for improving the performance of kriging in this setting. It is well-known in kriging that experimental design has a major impact on performance, and we found the same in our example, having tried other designs in computational experiments not reported here. One could look for designs better than the one illustrated in Figure 2.2.

CHAPTER 3

Response Surface Methodology for Simulating Trading and Hedging Strategies: Stochastic Kriging and Trend Modeling Applications

3.1. Introduction

Chapter 2 suggests that simple kriging can be an effective tool for reducing the computational cost of nested simulations of hedging and trading strategies. In addition to problem specific modifications to the way kriging is applied on a problem, such as design specifications depending on a qualitative analysis of the actual response surface, we can also use other versions of the kriging methodology which might work better in practice for simulating and trading simulations.

One improvement to simple kriging, in which the response surface is formed simply by interpolating among observed values, is the theory of kriging that incorporates the modeling of trends in the response surface, as in regression. For example, one might model the response surface as a linear function of the spatial variables plus local deviations from this trend: then the spatial correlation model applies only to the local deviations. We introduce the concept of trend modeling in §3.2 and explore its effectiveness on a financial example in §3.3.

Kriging was developed for analyzing the results of deterministic experiments, for example, physical experiments in geology or deterministic computer experiments such as

finite-element codes: if these experiments were to be repeated, they would yield the same results, at least roughly. Kriging interpolates between the values observed at design points because it assumes that these values are the truth. This is not so in our financial framework, in which the values observed at design points include inner-level sampling error. [**5**] describes a stochastic kriging procedure which takes account of this uncertainty when constructing the response surface. We briefly describe this procedure in §3.4 and apply it on a financial example in §3.5. We compare the performances of the resulting response surfaces with surfaces generated by a simple kriging procedure.

3.2. Response Surface Modeling with Trend

Kriging is based on an assumption that the response surface has a location-invariant correlation structure: roughly speaking, the values of the response surface at locations separated by the same distance have the same degree of similarity regardless of where you look. Figure 2.1 suggests that this stationarity assumption is untrue in our example of a put option since the variability of the price of the option changes considerably in different regions of the (time $\times$ stock price) design space. To be more specific, at a particular time the put option's value is almost constant at nearly zero as long as the option is deep out of the money, i.e the stock price is larger than the strike price, while the option's value varies a lot in the subregion of the design space where the option is deep in the money where the stock price effects the value of the option considerably. For the particular example in §2.3, the value of the option at time stock price pairs (0.8, 120) and (0.8, 130) are very similar (nearly zero), but the option values at (0.8, 80) and (0.8, 90) are quite different (by about \$10). Several approaches have been used in the literature to enhance random function

modeling while retaining some of the theoretical simplifications that stationarity provides. The most frequently used of these techniques is to let the mean of the stochastic process generating the response depend on the risk factors in a standard regression manner while assuming that the residual variation still follows a stationary Gaussian random process. We will model the response $Y(x)$ at a design point x by

$$Y(x) = \sum_{j=1}^{b} B_j(x)\beta_j + Z(x) = \mathbf{B}^\top \beta + Z(x) \tag{3.1}$$

where $B_1, \ldots B_b$ are *known* basis functions to model the trend in the mean of the stochastic process, $\beta = (\beta_1, \ldots, \beta_b)^\top$ is a vector of *unknown* regression coefficients and $Z(\cdot)$ is a zero mean Gaussian random field with a location-invariant correlation structure over the domain of the risk factors x.

Given the responses at a number of design points, an estimate of the regression coefficient vector β can be calculated as a function of the covariance structure of the Gaussian random process using the method of generalized least squares. The details on how to estimate regression coefficients together with the parameters of the Gaussian random process can be found in [**38**] and [**19**].

In the literature on design and analysis of computer experiments, trend modeling is frequently reported not to work well in practice. However, because we have many more design points in our financial examples than are typical in that literature, it is possible that trend modeling may improve the performance of kriging when used as the response surface methodology for hedging and trading simulations. We investigate, in the next section, the

effect of using basis functions on response surface modeling within the context of another hedging simulation example.

3.3. Delta-Hedging a European Call Option under Stochastic Volatility

Our experiments in Chapter 3 featured the example of delta-hedging a European put option on a stock under the Black-Scholes model. We looked at three different simulation procedures which differed in how they estimated the security values and portfolio weights at each re-hedging time on each path of the outer hedging simulation. Both security values and portfolio weights were functions of time and the stock price, which resulted in a two-dimensional design space. We can increase the dimensionality of our design space by introducing stochastic volatility to the evolution of the stock price so that security values and portfolio weights depend not only on stock price and time but also on the level of stochastic volatility given the stock price at that time.

One of the most well-known and popular of all stochastic volatility models is the Heston model ([**26**]). Under the Heston model, the stock price follows a geometric Brownian motion while the squared volatility of the stock price has a mean reverting square root diffusion process.

Our experiment will be on delta-hedging a European call option where outer simulation paths are sampled under the Heston model. Our hedging portfolio will include a riskless money-market account and the underlying stock. The number of shares of stock to hold at any re-hedging time is equal to the delta of the option. Both the delta and value of the option are calculated under the same stochastic volatility process which is used to generate outer simulation paths.

In the particular example we use in our computational experiments the stock price price follows a Heston stochastic volatility process with initial value $S_0 = 100$, drift $\mu = 0.08$, initial squared volatility $\nu = 0.0174$, long-term mean squared volatility $\bar{\nu} = 0.0354$, volatility of volatility $\eta = 0.3877$, speed of reversion of squared volatility to its long-term mean $\lambda = 1.3253$, correlation between stock returns and changes in squared volatility $\rho = -0.7165$. The interest rate on the money-market account is $r = 0.0422$. The call option has maturity of $T = 1$ years and strike price K =100.

The paper of ([**28**]) proposes a nice numerical computation procedure, which we use in our formula-based approach described in §2.4.1 to calculate the value and delta of a European call option by Heston's formula ([**26**]). For the full nested simulation approach of §2.4.2, we use the Quadratic-Exponential scheme of [**4**] as the discretization scheme for the Monte Carlo simulation of the Heston stochastic volatility model. This scheme is straightforward, easier to implement and execute compared to the bias-free algorithm of [**10**], which is too complex and computationally expensive for the purposes of our analysis. Calculation of the delta of the option with Monte Carlo simulation will be done by using the path-wise estimation method as described in [**9**].

The response surface modeling approach of §2.4.3 should be modified to take into account the additional squared volatility component of the design space. The region which contains all the simulated data is now the Cartesian product $[t_0, t_{s-1}] \times [S_{\min}, S_{\max}] \times [V_{\min}, V_{\max}]$, where $V_{\min} = \min\{V_i^j | j = 1, 2, \ldots, k,\ i = 1, 2, \ldots, s\}$ and $V_{\max} = \max\{V_i^j | j = 1, 2, \ldots, k,\ i = 1, 2, \ldots, s\}$ are the smallest and largest observed squared volatility values. Then, our experimental design is also updated as follows:

- Partition $[t_0, t_{s-1}]$, $[S_{\min}, S_{\max}]$ and $[V_{\min}, V_{\max}]$ into d intervals of equal width.

- Sample one point uniformly within each interval.
- Randomly combine the time, stock price and squared volatility values to get d design points.
- Add the eight points $(t_0, S_{\min}, V_{\min})$, $(t_0, S_{\max}, V_{\min})$,$(t_{s-1}, S_{\min}, V_{\min})$, $(t_{s-1}, S_{\max}, V_{\min})$, $(t_0, S_{\min}, V_{\max})$, $(t_0, S_{\max}, V_{\max})$, $(t_{s-1}, S_{\min}, V_{\max})$ and $(t_{s-1}, S_{\max}, V_{\max})$ to the design.

We performed experiments in MATLAB, implementing kriging with the `PERK` toolbox and generating Latin hypercube designs with the `lhsdesign` function. We used $k = 500$ paths, $m = 5,000$ inner-level replications and 150 inner-level discretization steps for our simulation study. Our experimental designs had $n = 208$ design points. Using the same experimental design (both the design points and the responses at these points), we fit two kriging surfaces each, for both the value and the delta of the option. For fitting one of the value (delta) surfaces, the value (delta) of the call option under the Black-Scholes model is calculated at each of the design points and is used as a basis function for trend modeling in addition to a vector of 1's, which is used to estimate the mean of the residuals of the fitted Gaussian random process ($B_1(S,t,V) = 1$ and $B_2(S,t,V) = BSPrice(S,t,V)$ for price and $B_2(S,t,V) = BSDelta(S,t,V)$) for delta response surface. For a call option, the Black-Scholes model call prices and deltas are

$$\begin{aligned} BSCall(S,t,V) &= S\Phi(d_1) - K\exp(-r(T-t))\Phi(d_2), \\ BSDelta(S,t,V) &= \Phi(d_1) \end{aligned}$$

where

$$d_1 = \frac{\log(S/K) + (\sigma - \frac{r^2}{2})(T-t)}{\sigma\sqrt{T-t}},$$

$$d_2 = d_1 - \sigma\sqrt{T-t} \text{ and } \phi \text{ is the standard normal cumulative distribution function.}$$

The second value (delta) surface is generated by using just a vector of 1's as inputs at design points to estimate the mean of the fitted Gaussian random process for the value (delta) of the call option.

Figure 3.1 (Figure 3.3) shows the absolute error of the predicted values (deltas) of the call option at each re-hedging time along each hedging path by the response surface fitted with Black-Scholes value (delta) function as the basis function. The absolute error at a prediction point is calculated by taking the absolute difference of the value (delta) of the option predicted by the response surface and value (delta) of the option given by the formula based approach. Figure 3.2 (Figure 3.4) shows the absolute error at prediction points for the response surface created with no basis function other than a constant function, which is equal to 1. The colors of the points in these graphs are used to represent the level of the squared volatility at the prediction points.

Figure 3.1 (Figure 3.3) and Figure 3.2 (Figure 3.4) show that introduction of a reasonable basis function such as the value (delta) of the option under the Black-Scholes model helps us to reduce the substantial errors of the response surfaces particularly shortly before maturity, where both the value and delta of the option are badly behaved functions. We hope to reduce the error of the response surfaces even further by improving our experimental design. To assess the accuracy of the simulation procedures described in this section and also to produce graphs for the evolution of the mean and standard deviation of

the $P\&L$ during the hedging horizon, we propose to run a number of macro-replications, involving independently generated paths and design points, but with all procedures using the same paths in a specific macro-replication as part of future research.

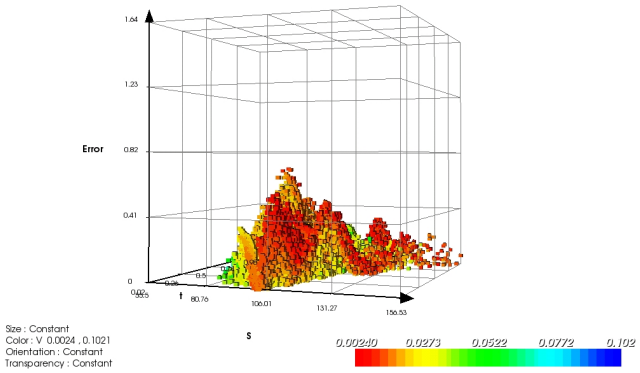

Figure 3.1. The absolute error of the price response surface estimated by kriging with Black-Scholes function for price used as a basis function

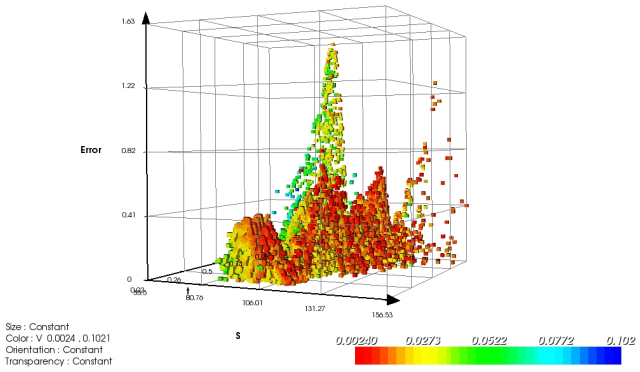

Figure 3.2. The absolute error of the price response surface estimated by simple kriging without any trend

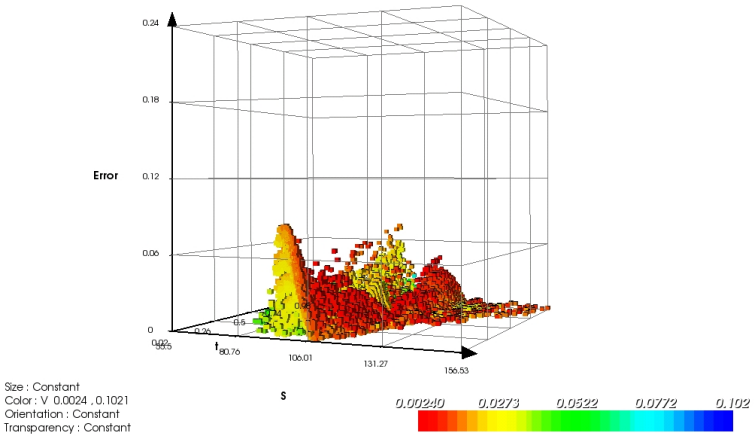

Figure 3.3. The absolute error of the delta response surface estimated by kriging with Black-Scholes function for delta used as a basis function

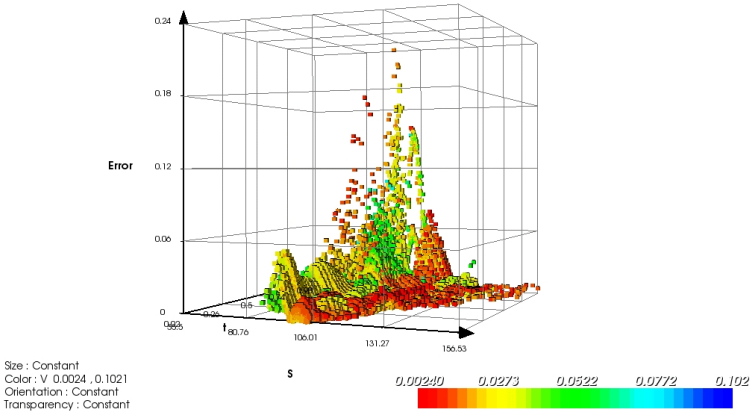

Figure 3.4. The absolute error of the delta response surface estimated by simple kriging without any trend

3.4. Response Surface Modeling with Stochastic Kriging

The application of response surface methodology to hedging and trading simulations can be improved by including the effect of inner-sampling error inherent in the calculation of the response at the design points. When portfolio weights or security values at a specific design point are estimated using Monte Carlo simulation, the intrinsic variance of the response due to sampling error will in general depend on where the design point is. As an example, the standard error of the Monte Carlo estimate of the value of a European call option will be high if the option is deep in or out of the money, while that of the Monte Carlo estimate of the delta of a European call option will be very small.

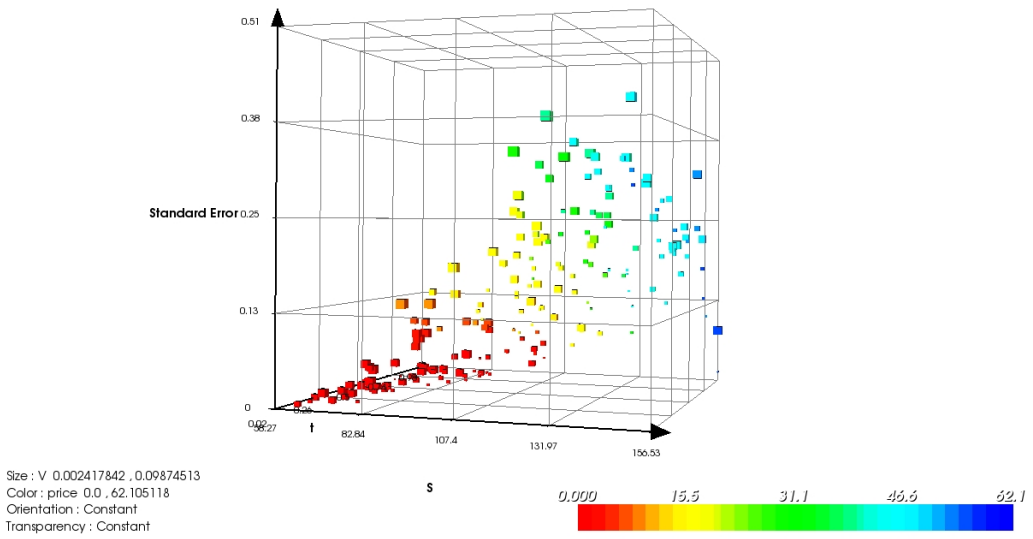

Figure 3.5. The standard errors of the Monte Carlo estimates of the price of the European call option at the design points in §3.3

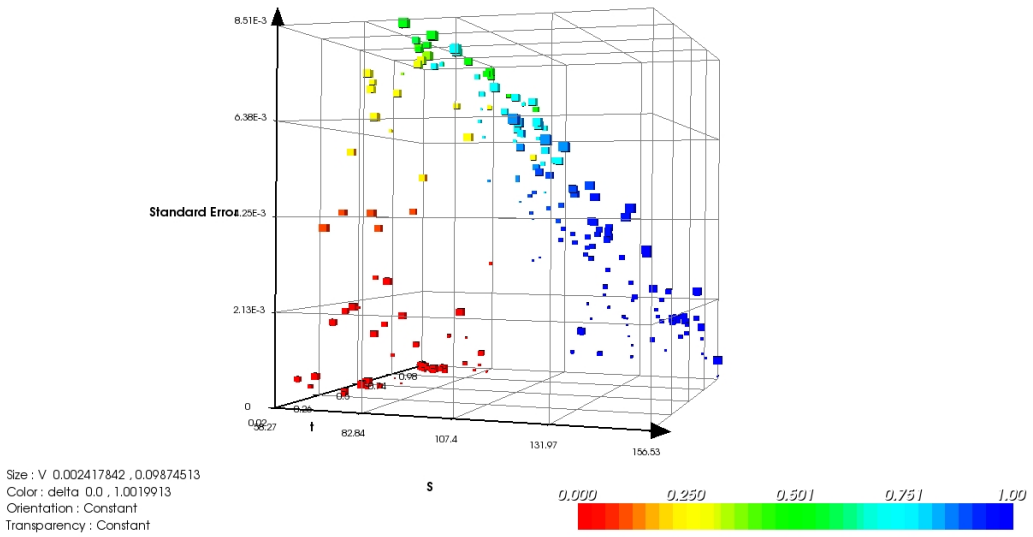

Figure 3.6. The standard errors of the Monte Carlo estimates of the delta of the European call option at the design points in §3.3

Figure 3.5 and Figure 3.6 demonstrate perfectly how the intrinsic sampling errors of estimates of the value and delta of the European call option under Heston model in §3.3, change according to where in the design space these responses are simulated.

[**5**] describes a stochastic kriging procedure, which accounts for both sampling and response surface uncertainty, and shows through a queueing model example how including sampling variability has an impact on experiment design, response surface estimation and inference. They model the response $Y_j(x)$ from the jth replication of a simulation study at a design point x by

$$Y_j(x) = \mathbf{B}^\top \beta + Z(x) + \varepsilon_j(x) \tag{3.2}$$

where the only difference from the model of the previous section is the intrinsic variance term $\varepsilon_j(x)$. The intrinsic noise $\varepsilon_1(x), \varepsilon_2(x), \ldots$ at a design point x is naturally independent and identically distributed across simulation replications but is not constant over the design space, i.e. the intrinsic variance depends on the design point at which the simulation is performed. Given the responses and the intrinsic variances associated with these responses at a number of design points, estimates of the regression coefficients β and the parameters of the Gaussian random process which determine the spatial correlation structure can be computed by the method of maximum likelihood estimation([**5**]). [**8**] describes an efficient procedure for parameter estimation of stochastic kriging.

In the next section, we apply the same stochastic kriging procedure of [**5**], implemented as described in [**8**], on a financial example, the hedging of a simple portfolio of options, and compare the results with those of a regular kriging procedure applied on the same problem.

3.5. Delta-Hedging a Portfolio of Options

Our computational experiments feature the example of delta-hedging a portfolio consisting of a European put option and a European call option on the same stock with different strike prices under the Black-Scholes model. Under the Black-Scholes model, the stock price S is geometric Brownian motion. The securities in the hedging portfolio include a riskless money-market account and the underlying stock. Their values at time t_i are respectively $V_{i1} = e^{rt_i}$ where r is the interest rate and $V_{i2} = S_i$. The value V_{i0} of the portfolio at any time is equal to the sum of the values of the put option and the call option at that time.

The portfolio (denoted as security 0) is hedged from the time it is sold until its maturity, when it pays off $V_{s0} = f_0(T, S_s) = \max\{K_1 - S_s, 0\} + \max\{S_s - K_2, 0\}$, where K_1 are the strike prices of the put and the call options, respectively.

For $i = 0, 1, \ldots, s-1$, the number θ_{i2} of shares of stock to hold at time t_i is set equal to the negative of the first-order sensitivity $\partial V_{i0}/\partial S_i$ of the portfolio value with respect to the stock price at that point in time, which is equal to the negative of the sum of the *deltas* of the two options.

At time $t_s = T$, the options mature and the hedge is unwound, so $\theta_{s2} = 0$. The number of shares in the money-market account is set so that the hedging strategy is self-financing as in §2.3.

In the particular example we use in our computational experiments, the stock price follows a geometric Brownian motion with initial value $S_0 = \$100$, drift $\mu = 8\%$, and volatility $\sigma = 15\%$. Both options have maturities of $T = 0.5$ years and strike price of the put option is $K_1 = \$110$ and the strike price of the call option is $K_2 = \$100$. The interest rate on the money-market account is $r = 5\%$. There are $s = 20$ re-balancing times, and $t_i = iT/s$ for $i = 1, 2, \ldots, s$.

We can evaluate the P&L distribution resulting from delta-hedging our portfolio, which contains two securities, the values of which, together with their sensitivities to the underlying stock on any path, can be calculated through both the formula based and full nested simulation approaches of §2.4. For response surface modeling approach, we try two different experimental setups for each of the two different response surface methodologies: stochastic kriging and regular kriging. The first setup is the original experimental setup of the second chapter, where the response (the value or the sensitivity) is a function of

time and the underlying stock price. So, for a single outer hedging simulation, a single value and a single sensitivity response surface is generated over the (time)×(stock price) domain. The second setup is a new experimental setup, we propose to cope with the non-stationarity characteristics of our financial example, where a different value and a different sensitivity response surface over the (stock price) domain is generated at each re-hedging time for an outer hedging simulation.

The response surface modeling approaches with stochastic kriging require us to estimate through inner-simulations not only the value of the portfolio and its sensitivity with respect to the underlying stock, but also the sampling variabilities of the estimates of these quantities. Note that for each response surface methodology, only two response surfaces are fit: one for the value of the portfolio and one for the delta of the portfolio. Another approach would be to create response surfaces for the values and deltas of the options separately and use these response surfaces to estimate the value of the portfolio and its delta.

We performed our experiments in MATLAB, implementing regular kriging with the `PERK` toolbox. The stochastic kriging code in MATLAB was written based on the procedure described in [**8**]. Generating Latin hypercube designs was performed by the `lhsdesign` function.

We used $k = 1,000$ paths for hedging and $m = 500$ inner-level replications for pricing Our experimental design had $n = 114$ design points for the original setup and $n = 6$ design points for the new setup. Note that the new setup requires generating response surfaces at each of the $20 - 1 = 19$ re-hedging times, therefore an equal number $19 \times 6 = 114$ of inner level simulations are performed for the new experimental setup compared to the

original setup. We subtract one from the number of re-hedging times since we do not need response surfaces at the last re-hedging time corresponding to the maturity of the options in our portfolio.

To assess the accuracy of the four simulation procedures, we ran 100 macro-replications, each an independent run of the entire procedure, involving independently generated paths and design points. However, within a single macro-replication, all procedures used the same paths and the response surfaces were fit using stochastic and regular kriging based on the same design. This produced 100 estimates of the mean P&L and the standard deviation of P&L at each time step, for each procedure: that is, each of 100 runs produces an estimate of $\mathbf{E}[\Pi_i]$ and an estimate of $\sqrt{\mathbf{Var}[\Pi_i]}$ for each time step $i = 0, 1, \ldots, s$. We depict the average and the sample root mean squared error (RMSE) of these 100 estimates in Figures 3.9–3.12.

Figures 3.9 and 3.10 show that response surface modeling with ordinary kriging can introduce substantial bias and variance into estimating mean P&L. With $n = 114$ design points and not taking into account intrinsic variability, ordinary kriging tends to generate a response surface with poor fidelity.

Particularly noticeable is the bias just before maturity, where the portfolio's price and delta are badly behaved functions, for reasons explained in §2.3 for the value and the delta of the put option. We see that this bias is considerably smaller for the response surface modeling approach using the stochastic kriging procedure. This is due to the fact that the response surfaces for both the price and delta of the portfolio generated by stochastic kriging take into account the variability of the standard errors of the estimates of the price and delta of the portfolio over the design space, which can be observed in Figure 3.7 and

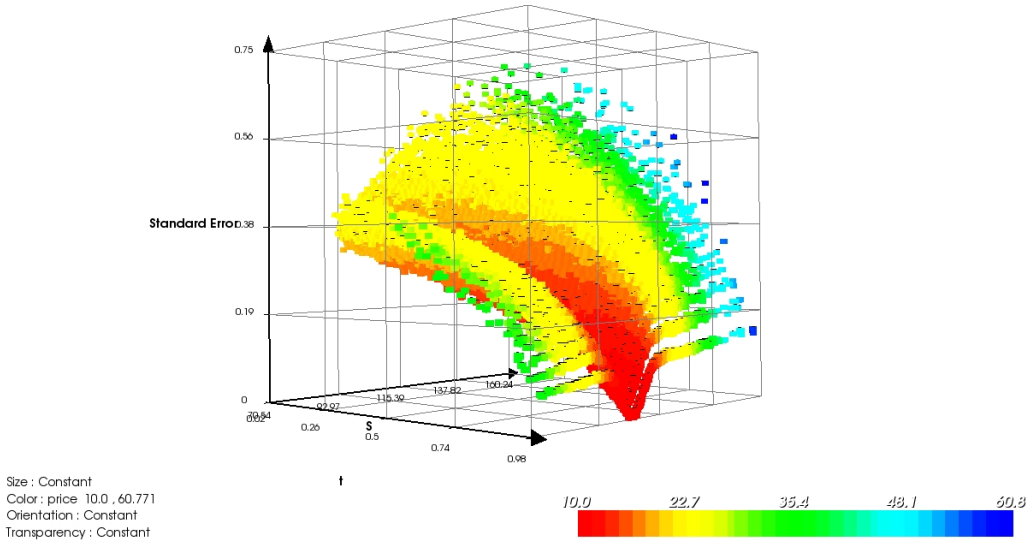

Figure 3.7. The variability in the standard error of the value of the portfolio estimated through inner-simulations for a single macro-replication

Figure 3.8 for a single macro-replication. Furthermore, estimating a different response surface at each re-hedging time does a poor job in terms of estimating the mean P&L compared to the original setup especially when ordinary kriging methodology is used. The stochastic kriging methodology gives significantly better results with both experimental setups, which indicates that including an intrinsic variance term varying over the design space for a *global* response surface is more useful for this example than estimating different *local* response surface over different sub-domains, i.e. each re-hedging time.

We conjecture that increasing the number of design points as in §2.5 can bring the bias of response surface modeling estimates of mean P&L for both methods at intermediary times to a fair degree of accuracy, compared to the standard deviation of P&L portrayed in Figure 3.11.

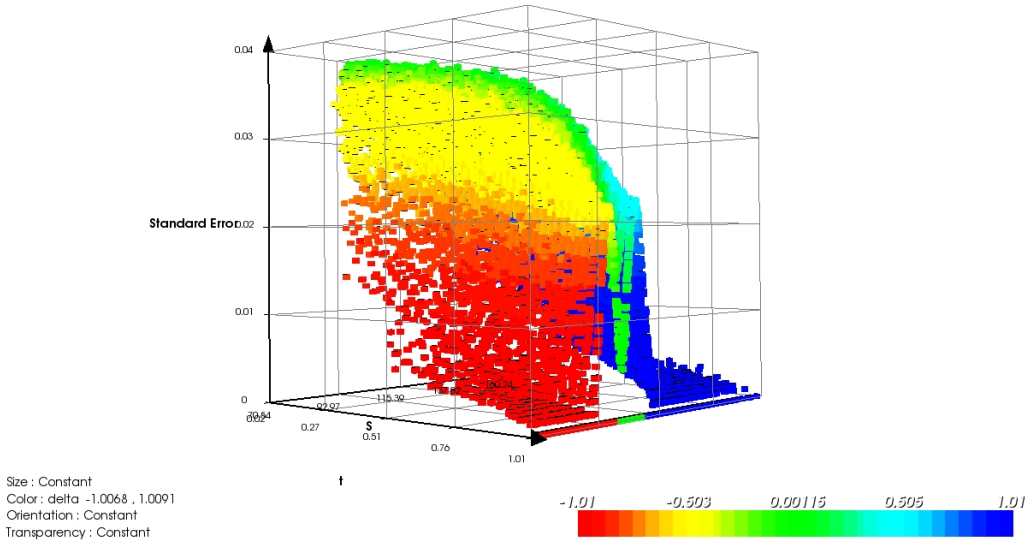

Figure 3.8. The variability in the standard error of the delta of the portfolio estimated through inner-simulations for a single macro-replication

Figures 3.11 and 3.12 show that response surface modeling with stochastic kriging procedure performs similarly under both experimental setups, the original setup slightly better than the new setup towards maturity, in estimating the standard deviation of P&L, while the ordinary kriging method, especially with the new experimental setup, fails to give good estimates of the standard deviation P&L.

The bias and variance introduced by response surface modeling are much greater at time t_{s-1} than at time $t_s = T$ and at time t_1 than at time $t_0 = 0$ (where P&L is zero by construction). This happens because the simulation procedure relies on the response surface for the portfolio values to produce P&L at times t_{s-1} and t_1 but not at times t_s and t_0. The jump from time t_0 to time t_1 in RMSE of estimates from methods other than

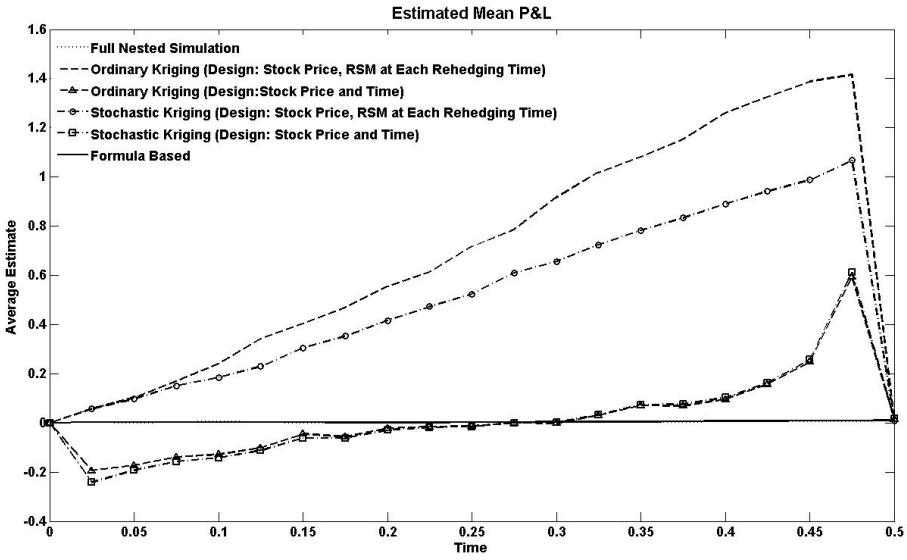

Figure 3.9. The average, over 100 macro-replications, of four simulation methods' estimates of mean P&L at each time step

the formula-based method is explained by inner-level sampling error in estimating the put option value function $f_0(t_1, \cdot)$.

Next we focus on the distribution of terminal P&L generated by each of the simulation methods. We pool the terminal P&L values on $k = 1,000$ paths in each of 100 macro-replications to produce a picture of the probability density of terminal P&L in Figure 3.13. Table 3.1 summarizes the results in Figures 3.9–3.12 for P&L at time $T = 0.5$ only. It also shows the standard errors in estimating these quantities with 100 macro-replications. The standard error in estimating RMSE is calculated by the delta method (see, e.g., [**7**], §III.3). From the table we see that the response surface modeling procedure with stochastic kriging is quite accurate in estimating the mean and standard deviation of terminal P&L.

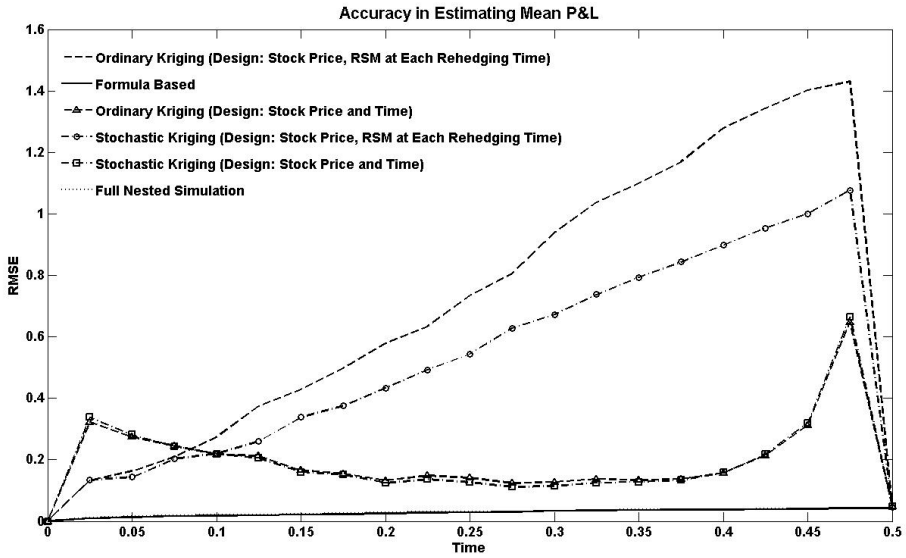

Figure 3.10. The sample root mean squared error, over 100 macro-replications, of four simulation methods' estimates of mean P&L at each time step

Table 3.1. The average, standard error (SE) and the root mean square error (RMSE) statistics of the mean estimates of terminal P&L.

	Mean P&L		Std. Dev. of P&L	
Method	Average	RMSE	Average	RMSE
Formula	0.010	0.0428	1.376	0.4613
Based	*(0.004)*	*(0.003)*	*(0.003)*	*(0.003)*
Full Nested	0.009	0.0430	1.406	0.4920
Simulation	*(0.004)*	*(0.004)*	*(0.004)*	*(0.004)*
Ordinary Kriging	0.011	0.0435	1.492	0.5782
(RSM at Each Time)	*(0.004)*	*(0.004)*	*(0.004)*	*(0.004)*
Ordinary Kriging	0.015	0.0484	1.438	0.5237
	(0.005)	*(0.004)*	*(0.004)*	*(0.004)*
Stochastic Kriging	0.009	0.0457	1.450	0.5357
(RSM at Each Time)	*(0.005)*	*(0.004)*	*(0.004)*	*(0.004)*
Stochastic Kriging	0.016	0.0482	1.437	0.5231
	(0.005)	*(0.004)*	*(0.004)*	*(0.004)*

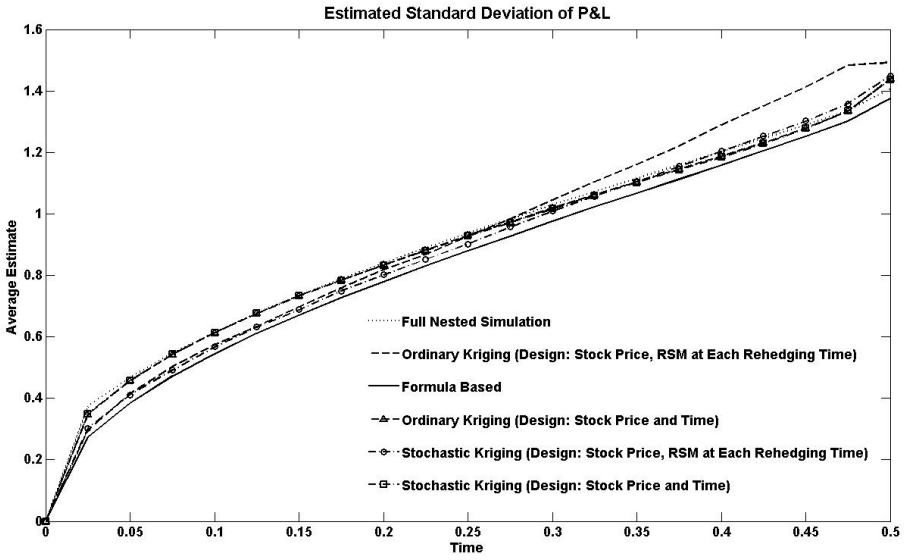

Figure 3.11. The average, over 100 macro-replications, of four simulation methods' estimates of the standard deviation of P&L at each time step

3.6. Conclusion and Future Research

We have seen that versions of kriging method with trend modeling can perform better compared to simple kriging for financial applications due to large number of design points. Although the responses surfaces created for the price and delta of a European call option under the Heston model were relatively satisfactory, the performance of using the Black-Scholes value and delta functions as basis functions for this specific example should be further tested by performing experiments with macro-replications similar to those of §2.5 and §3.5. It should be noted that such experiments are computationally expensive since a single macro-replication as in §3.3 requires the numerical computation and Monte Carlo estimation of the values and deltas of the European option under the Heston model at ks risk factor combinations.

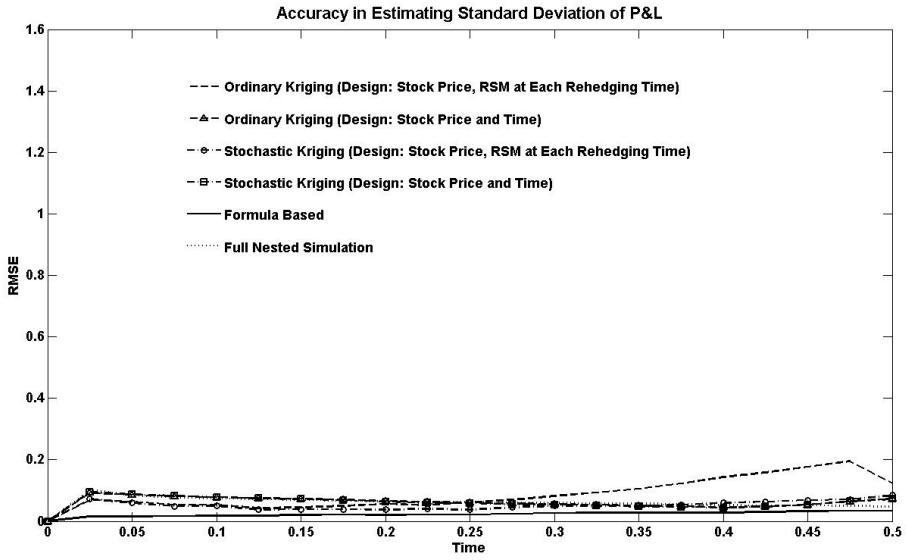

Figure 3.12. The sample root mean squared error, over 100 macro-replications, of four simulation methods' estimates of the standard deviation of P&L at each time step

We have also seen that stochastic kriging by taking into account intrinsic variability of the responses at design points performs considerably better than regular kriging for hedging simulations where the responses such as the prices and price sensitivities of securities with respect to risk factors are estimated using Monte Carlo simulation. In addition to better fitting the hedging simulation framework through allowing design dependent intrinsic variance analysis, stochastic kriging has computational advantages over regular kriging, especially for parameter estimation. A detailed description and analysis of computational aspects of the efficient stochastic kriging code used for our analysis here can be found in [**8**]. Using the algorithm of this code to compare the efficiency gains due to making fewer simulations to the computational effort required to fit a response surface

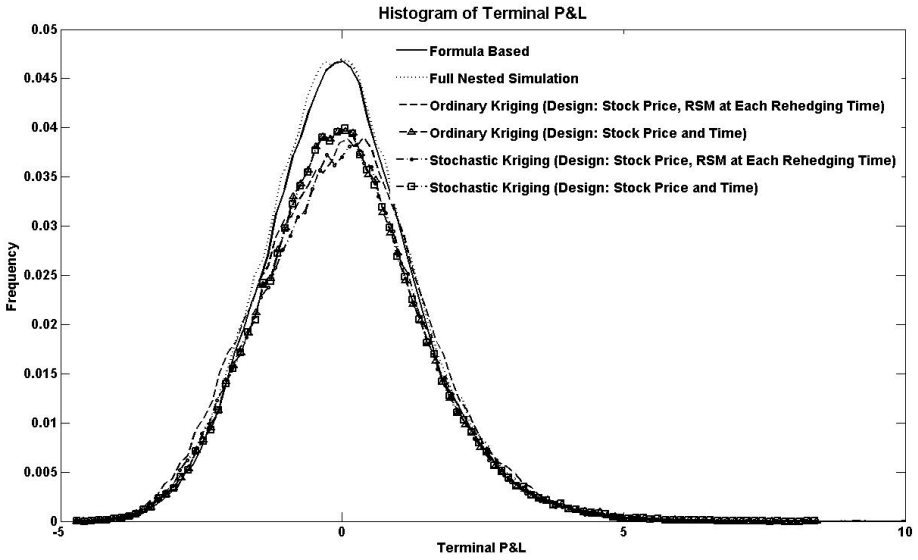

Figure 3.13. Approximate probability density functions of terminal P&L generated by four simulation methods, based on $100,000$ samples of terminal P&L.

and to predict the responses at unexplored cites using the fitted response surface is part of future research.

Another issue that we would like to analyze for future research is the effect of the choice of correlation function for the kriging procedure on the performance of response surface modeling for financial applications. We have used in all of our experiments correlation functions from the family of product exponential functions, i.e. correlation dies out exponentially in distance along each dimension and the correlation between two points is calculated by taking the product of the correlation components along each dimension. This is definitely a restriction especially for financial applications, which involve in general an interaction between dimensions and where the correlation between two design points depends on not only how apart they are spaced but also where in the design space they

are located. We have observed especially for stochastic kriging that the scale parameter of the power exponential correlation function, which controls how fast correlation dies out in distance, along the time dimension is quite small. This implies that the variability in the response surface is mostly explained by the spot price, which on average is true, and is *homogeneous* along the time dimension. On the contrary, we know that the price of an at-the-money option will be different according to how much time there is until maturity while time to maturity has negligible effect if the option is deep in or out of the money. Therefore, if we were able to focus on a subregion of the design space, where the option is slightly in or out of the money, we could actually see an increase in the estimated scale parameter for the correlation function along the time dimension. Note that this is a better problem specific approach than dividing the problem domain into sub-domains at each re-hedging time, which did not work well for the portfolio example in this chapter. Since the limits of our design space is controlled by the outer level simulations, it might be appropriate to look for improvements in how we design the two-level simulations as well as improvements in the kriging procedure itself such as choice of correlation function (e.g. non-stationary correlation functions with interaction terms) and design.

References

[1] Acerbi, C. 2002 . Spectral measures of risk: a coherent representation of subjective risk aversion. *Journal of Banking and Finance*, **26**, 1505–1518.

[2] Acerbi, C., and Tasche, D. 2002 . On the coherence of expected shortfall. *Journal of Banking and Finance*, **26**, 1487–1503.

[3] Agresti, A., and Coull, B. A. 1998 . Approximate is better than exact for interval estimation of binomial proportions. *The American Statistician*, **52**(2), 119–126.

[4] Andersen, L. 2008 . Efficient simulation of the Heston stochastic volatility model. *Journal of Computational Finance*, **11**(3).

[5] Ankenman, B., Nelson, B. L., and Staum, J. 2008 . Stochastic kriging for simulation metamodeling. *Working Paper 08-01, Dept. of IEMS, Northwestern University.*

[6] Artzner, P., Delbaen, F., Eber, J., and Heath, D. 1999 . Coherent Measures of Risk. *Mathematical Finance*, **9**, 203–228.

[7] Asmussen, S., and P. W. Glynn. 2007 . *Stochastic Simulation.* Springer-Verlag.

[8] Baysal, E., and Staum, J. 2008. Mechanics of Krigign for Simulation Metamodeling: Implementation and Analysis *Working Paper 08-04, Dept. of IEMS, Northwestern University*

[9] Broadie M. and Kaya, O. 2004 . Exact simulation of option greeks under stochastic volatility and jump diffusion models. *Proceedings of the 2004 Winter Simulation Conference ed. Ingalls, R. D., Rossetti, M. D., Smith, J. S. and Peters, B. A.*

[10] Broadie M. and Kaya, O. 2006 . Exact simulation of stochastic volatility and other affine jump diffusion processes. *Operations Research* **54**(2).

[11] Chang, Y. P., Hung, M. C., and Wu, Y. F. 2003 . Nonparametric estimation for risk in value-at-risk estimator. *Communications in Statistics: Simulation and Computation*, **3**, 1041–1064.

[12] Chen, S. X., and Hall, P. 1993 . Smoothed empirical likelihood confidence intervals for quantiles. *Annals of Statistics*, **21**(3), 1166–1181.

[13] Chernick, M. R. 1999 . *Bootstrap Methods: A Practitioner's Guide*. Wiley Series in Probability and Statistics.

[14] Clopper, C. J., and Pearson, E. S. 1934 . The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika*, **26**(4), 404–413.

[15] Dagpunar, J. 1988 . *Principles of Random Variate Generation*. Clarendon Press - Oxford.

[16] Davison, A. C., and Hinkley, D. V. 1997 . *Bootstrap Methods and their Application*. Cambridge University Press.

[17] Dowd, K. 2005 . Estimating risk measures. *Financial Engineering News*, **43**, 13.

[18] Durrett, R. 1996 . *Probability: Theory and Examples*. Duxbury Press.

[19] Fang, K. T., Li,R. and Sudjianto, A. 2006 . *Design and Modeling for Computer Experiments*. Chapman & Hall/CRC.

[20] Frye, J. 1998 . Monte Carlo by day. *Risk Magazine*, Nov. 1998, 66–71.

[21] Gatheral, J. 2006 . *The Volatility Surface A Practitioner's Guide*. John Wiley & Sons.

[22] Glasserman, P. 2003 . *Monte Carlo Methods in Financial Engineering*. Springer-Verlag.

[23] Gordy, M. B., and Juneja, S. 2008 . Nested simulation in portfolio risk measurement. *Working Paper 2008-21, Finance and Economics Discussion Series, Federal Reserve Board*.

[24] Hardy, M. 2006 . Simulating VaR and CTE. *Financial Engineering News*, **47**, 17.

[25] Hesterberg, T. C., and Nelson, B. L. 1998 . Control Variates for Probability and Quantile Estimation. *Management Science*, **44**(9), 1295-1312.

[26] Heston, S. L. 1993 . A closed-form solution of options with stochastic volatility with applications to bond and currency options. *The Review of Financial Studies*. **6** (2), 327–343.

[27] Hong, L. J. 2006 . Estimating value and sensitivities of conditional value-at-risk. Unpublished manuscript, Department of Industrial Engineering and Logistics Management, Hong Kong University of Science and Technology.

[28] Kahl, C. and Jackel, P. 2005 . Not-so-complex logarithms in the Heston model. *Wilmott Magazine* September, 2005.

[29] Lan, H., Nelson, B. L., and Staum, J. 2007 . Two-level simulations for risk management. *Proceedings of the 2007 INFORMS Simulation Society Research Workshop, ed. Chick, S., Chen, C. H., Henderson, S. and Yücesan, E.*, 102–107. INSEAD.

[30] Manistre, B. J., and Hancock, G. H. 2005 . Variance of the CTE estimator. *North American Actuarial Journal*, **9**, 129–154.

[31] Nielsen, L. T. 1999 . *Pricing and Hedging of Derivative Securities*. Oxford University Press.

[32] Owen, A. B. 1988 . Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*, **75**, 237–249.

[33] Owen, A. B. 1990 . Empirical likelihood ratio confidence regions. *The Annals of Statistics*, **18**, 90–120.

[34] Owen, A B. 2001 . *Empirical Likelihood*. Chapman & Hall/CRC.

[35] Peng, L., and Qi, Y. 2006 . Confidence regions for high quantiles of a heavy tailed distribution. *Annals of Statistics*, **34**, 1964–1986.

[36] Rockafellar, R. T., and Uryasev, S. 2000 . Optimization of conditional value-at-risk. *Journal of Risk*, **2**, 21–41.

[37] Rockafellar, R. T., and Uryasev, S. 2002 . Conditional value-at-risk for general loss distributions. *Journal of Banking and Finance*, **26**, 1443-1471.

[38] Santner, T. J., Williams, B. J. and Notz, W. I. 2003 . *The Design and Analysis of Computer Experiments*. Springer-Verlag.

[39] Shao, J., and Tu, D. 1995 . *The Jacknife and Bootstrap*. Springer Series in Statistics.

[40] Shaw, J. 1999 . Beyond VAR and stress testing. *Monte Carlo: Methodologies and Applications for Pricing and Risk Management, ed. Dupire, B.*, 231–244. Risk Books.

[41] Sheather, S. J., and Marron, J. S. 1990 . Kernel quantile estimators. *Journal of the American Statistical Association*, **85**, 410–416.

[42] Silverman, B. W. 1978 . Weak and uniform consistency of the kernel estimate of a density function and its derivatives. *The Annals of Statistics*, **6**, 177–184.

[43] Staudte, R. G., and Sheather, S. J. 1990 . *Robust Estimation and Testing*. John Wiley & Sons.

[44] Stein, M. L. 1999 . *Interpolation of Spatial Data: Some theory for Kriging*. Springer-Verlag.

[45] Yamai, Y., and Yoshiba, T. 2002 . Comparative analyses of expected shortfall and value-at-risk: Their estimation error, decomposition, and optimization. *Monetary and Economic Studies*, **20**(1), 87–121.

[46] Zhou, W., and Jing, B. Y. 2003 . Adjusted empirical likelihood method for quantiles. *Annals of the Institute of Statistical Mathematics*, **55**(4), 689–703.

APPENDIX A

Empirical Likelihood for VaR and ES

1. Maximization Problem I

The problem of computing $R_l^{\psi}(\mu)$ given by

$$(.1) \qquad R_l^{\psi}(\mu) = \max\left\{\prod_{i=1}^{k}(kw_i) \mid \sum_{i=1}^{l} w_i = p, \mu = -\frac{1}{p}\sum_{i=1}^{l} V_{[i]}w_i, w_i \geq 0, \sum_{i=1}^{k} w_i = 1\right\}$$

reduces to solving the following problem referred to as Maximization Problem I:

$$(.2) \qquad \begin{aligned} &\text{maximize} && \textstyle\sum_{i=1}^{l}\log(kw_i) + (k-l)\log(k\frac{1-p}{k-l}) \\ &\text{subject to} && -\tfrac{1}{p}\textstyle\sum_{i=1}^{l} w_i V_{[i]} = \mu \\ & && \textstyle\sum_{i=1}^{l} w_i = p \end{aligned}$$

since $W_l = \sum_{i=1}^{l} w_i$ is restricted to be exactly equal to p by (.1) and in this case $R_l^{\psi}(\mu)$ is achieved by assigning equal weights to the remaining $k-l$ portfolio values $V_{[l+1]}, \ldots, V_{[k]}$.

Note that the first equation in (.2) can be written as $\sum_{i=1}^{l} w_i(\mu + V_{[i]}) = 0$ by $p\mu + \sum_{i=1}^{l} w_i V_{[i]} = (\sum_{i=1}^{l} w_i)\mu + \sum_{i=1}^{l} w_i V_{[i]}$.

Since a strictly concave function is maximized on a linear set of equality constraints, the solution to this maximization problem will be found by using the Lagrangian function

$$\mathcal{L} = \sum_{i=1}^{l}\log(kw_i) + (k-l)\log(k\frac{1-p}{k-l}) + \frac{l}{p}\lambda\sum_{i=1}^{l} w_i(V_{[i]} + \mu) + \gamma(\sum_{i=1}^{l} w_i - p)$$

and the first order conditions

$$\frac{\partial \mathcal{L}}{\partial w_i^*} = \frac{1}{w_i^*} + \frac{l}{p}\lambda^*(V_{[i]} + \mu) + \gamma^* = 0 \qquad \forall i = 1, \ldots, l \tag{.3}$$

are sufficient.

Using equations (.3), we get $\sum_{i=1}^{l} w_i^* \frac{\partial \mathcal{L}}{\partial w_i^*} = 0$, which leads together with constraints in (.2) to $l + 0 + \gamma^* p = 0$ and hence $\gamma^* = -l/p$. Plugging the value of γ^* back into equations (.3), we get

$$w_i^* = \frac{p}{l}\left[1 - (V_{[i]} + \mu)\lambda^*\right]^{-1} \qquad \forall i = 1, \ldots, l. \tag{.4}$$

Plugging the values of w_i^* calculated above into the first constraint in (.2), we get

$$\sum_{i=1}^{l} \frac{V_{[i]} + \mu}{1 - (V_{[i]} + \mu)\lambda^*} = 0. \tag{.5}$$

Lemma 1.1. *If $\mu \in (-V_{[l]}, -V_{[1]})$, then Equation (.5) is satisfied for some*

$$\lambda^* \in \left(\frac{l-1}{l(\mu + V_{[1]})}, \frac{l-1}{l(\mu + V_{[l]})}\right).$$

Proof: Define $f_i(\mu, \lambda) := \frac{V_{[i]} + \mu}{1 - (V_{[i]} + \mu)\lambda}$ and $f_i^\mu(\lambda) := f_i(\mu, \lambda)$. Each f_i^μ has one discontinuity at $(V_{[i]} + \mu)^{-1}$, where f_i^μ is not defined. Therefore $\sum_{i=1}^{l} f_i^\mu(\lambda)$ is continuous on $\left((V_{[1]} + \mu)^{-1}, (V_{[l]} + \mu)^{-1}\right)$. Because the partial derivative

$$\frac{\partial f_i}{\partial \lambda} = \frac{(V_{[i]} + \mu)^2}{[1 - (V_{[i]} + \mu)\lambda]^2}$$

is positive unless $\mu = -V_{[i]}$, in which case it is zero, we can see that $\sum_{i=1}^{l} f_i^{\mu}(\lambda)$ is increasing in λ on $\left((V_{[1]}+\mu)^{-1}, (V_{[l]}+\mu)^{-1}\right)$. Because

$$\lim_{\lambda \uparrow (V_{[l]}+\mu)^{-1}} f_l^{\mu}(\lambda) = \infty \quad \text{and} \quad \lim_{\lambda \downarrow (V_{[1]}+\mu)^{-1}} f_1^{\mu}(\lambda) = -\infty,$$

there exists $\lambda^* \in \left((V_{[1]}+\mu)^{-1}, (V_{[l]}+\mu)^{-1}\right)$ such that $\sum_{i=1}^{l} f_i^{\mu}(\lambda^*) = 0$.

In fact, we can find tighter bounds for λ^* than $(V_{[1]}+\mu)^{-1}$ and $(V_{[l]}+\mu)^{-1}$. Because $w_1^* = \frac{p}{l}[1-\lambda^*(V_{[1]}+\mu)]^{-1} \leq p$ and $V_{[1]} < -\mu \Rightarrow V_{[1]}+\mu < 0$, $\lambda^* > \frac{l-1}{l(\mu+V_{[1]})}$. Likewise, using $w_l^* = \frac{p}{l}[1-\lambda(V_{[l]}+\mu)]^{-1} \leq p$ and $-\mu < V_{[l]} \Rightarrow V_{[l]}+\mu > 0$, we find $\lambda^* < \frac{l-1}{l(\mu+V_{[l]})}$. $\square$

Lemma 1.2. *R_l^{ψ} is increasing on $\left(-V_{[l]}, \mu_l^*\right)$ and decreasing on $\left(\mu_l^*, -V_{[1]}\right)$, where μ_l^* is defined as $\mu_l^* := -\frac{1}{l}\sum_{i=1}^{l} V_{[i]}$.*

Proof: The empirical likelihood ratio R_l^{ψ} is maximized at μ_l^*, where the solution to Maximization Problem I with $\mu = \mu_l^*$ involves $\lambda^* = 0$, so the optimal weights are p/l for $i = 1, \ldots, l$ and are $(1-p)/(k-l)$ for $i = l+1, \ldots, k$. Consider some $\mu \in (-V_{[l]}, \mu_l^*)$. Let F_{μ}, with weights $\{w_i^{\mu}\}_{i=1,\ldots,k}$, be the distribution at which $R_l^{\psi}(\mu)$ is attained. Because $\mu < \mu_l^*$, Equations (.2) and (.4) imply that $\lambda_{\mu}^* > 0$ at the solution to Maximization Problem I. This makes the optimal weights $\{w_i^{\mu}\}_{i=1,\ldots,l}$ increasing in i. In the trivial case where $V_{[i]}$ is the same for all $i = 1, \ldots, l$, the conclusion of the lemma holds; we henceforth assume that there exist $m < n \leq l$ such that $V_{[m]} < V_{[n]}$. Because the weights are increasing, for some $\epsilon > 0$, $w_m^{\mu} = w_n^{\mu} - \epsilon$. For any $\mu' \in (\mu, \mu + (\epsilon/2)(V_{[n]} - V_{[m]})$, let $\delta = (\mu' - \mu)/(V_{[n]} - V_{[m]})$. Construct F' with weights $\{w_i'\}_{i=1,\ldots,k}$ such that $F' = F_{\mu}$ except $w_m' = w_m^{\mu} + \delta$ and $w_n' = w_n^{\mu} - \delta$. Because $\delta \in \left(0, \frac{\epsilon}{2}\right)$, $w_m' w_n' > w_m^{\mu} w_n^{\mu}$, so $R(F') > R(F_{\mu})$. This leads to the conclusion $R_l^{\psi}(\mu') \geq R(F') > R(F_{\mu}) = R_l^{\psi}(\mu)$. We therefore showed that

for all $\mu \in (-V_{[l]}, \mu_l^*)$, R_l^ψ is increasing on a nonempty open interval whose left endpoint is μ, which in turn proves that R_l^ψ is increasing on $\left(-V_{[l]}, \mu_l^*\right)$. A similar analysis for $\mu > \mu_l^*$, involving $\lambda_\mu^* < 0$, proves that R_l^ψ is decreasing on $\left(\mu_l^*, -V_{[1]}\right)$. $\square$

2. Maximization Problem II

Maximization Problem II is more complicated due to the inequality constraints for W_l and W_{l-1}. We write $w_l = p - W_{l-1} + \delta$ where $\delta > 0$ so that the constraint $W_l > p$ is satisfied. The expected shortfall constraint can be written as $p(\mu + V_{[l]}) = \sum_{i=1}^{l-1} w_i (V_{[l]} - V_{[i]})$. Since w_l does not appear in the modified expected shortfall constraint, $R_l^{int}(\mu)$ will be attained when w_l is as close as possible to $w_{l+1} = \ldots = w_k = \frac{1-W_{l-1}}{k-l+1}$ and this implies $W_{l-1} > p - \frac{1-p}{k-l}$ since δ is defined to be positive. Then, the problem of finding $R_l^{int}(\mu)$ reduces to solving the following maximization problem:

$$
\begin{array}{ll}
\text{maximize} & \sum_{i=1}^{l-1} \log(k w_i) + (k-l+1) \log(k \frac{1-W_{l-1}}{k-l+1}) \\
\text{subject to} & \sum_{i=1}^{l-1} w_i (V_{[l]} - V_{[i]}) = p(\mu + V_{[l]}) \\
& W_{l-1} = \sum_{i=1}^{l-1} w_i \\
& W_{lb} < W_{l-1} < W_{ub}
\end{array}
\tag{.6}
$$

where $W_{lb} := p - \frac{1-p}{k-l}$ and $W_{ub} := p$. The Hessian of the above objective function is an l-dimensional diagonal matrix with $\{-\frac{1}{w_i^2}\}_{i,\ldots,l-1}$ and $-\frac{1}{(1-W_{l-1})^2}$ as the diagonal entries and therefore is negative definite. Since a concave function is maximized subject to linear constraints, there exists a unique global optimum for the above maximization problem.

This maximum can be computed by using the Lagrangian function

$$\begin{aligned}\mathcal{L} &= \sum_{i=1}^{l-1}\log(kw_i) + (k-l+1)\log(k\frac{1-W_{l-1}}{k-l+1}) - \lambda\left(\sum_{i=1}^{l-1} w_i(V_{[l]}-V_{[i]}) - p\,(\mu+V_{[l]})\right)\\ &\quad -\gamma\left(W_{l-1}-\sum_{i=1}^{l-1} w_i\right)\end{aligned}$$

and the first order conditions

$$(.7)\qquad 0 = \frac{\partial\mathcal{L}}{\partial w_i^*} = \frac{1}{w_i^*} - \lambda^*(V_{[l]}-V_{[i]}) + \gamma^*\,,\ \forall i = 1,\ldots,l-1 \text{ and}$$

$$(.8)\qquad 0 = \frac{\partial\mathcal{L}}{\partial W_{l-1}^*} = -\frac{k-l+1}{1-W_{l-1}^*} - \gamma^* = 0$$

are sufficient. Using equations (.7) and (.8), we write

$$(.9)\qquad w_i^* := g_i(W_{l-1}^*,\lambda^*) = \left[\frac{k-l+1}{1-W_{l-1}^*} + \lambda^*(V_{[l]}-V_{[i]})\right]^{-1}, \forall i = 1,\ldots,l-1$$

and get the following system of nonlinear equations in two unknowns:

$$W_{l-1}^* - \sum_{i=1}^{l-1} g_i(W_{l-1}^*,\lambda^*) = 0 \text{ and}$$

$$\sum_{i=1}^{l-1} g_i(W_{l-1}^*,\lambda^*)(V_{[l]}-V_{[i]}) - p(\mu+V_{[l]}) = 0.$$

Lemma 2.1. *The function g_i defined in Equation (1.19) is strictly decreasing in each of its arguments.*

Proof: The partial derivatives of g_i with respect to W_{l-1} and λ are

$$\frac{\partial g_i(W_{l-1},\lambda)}{\partial W_{l-1}} = -\frac{(k-l+1)w_i^2}{(1-W_{l-1})^2} \quad \text{and} \tag{.10}$$

$$\frac{\partial g_i(W_{l-1},\lambda)}{\partial \lambda} = -(V_{[l]}-V_{[i]})w_i^2 \tag{.11}$$

which are negative everywhere. □

The following lemma provides bounds on μ for which $R_l^{int}(\mu)$ can be found by solving the first order conditions for Maximization Problem II. Define M_l^{int} as a set which contains μ if and only if Equations (1.24) and (1.25) have a solution (W_{l-1}^*, λ^*) with $W_{l-1}^* \in (W_{lb}, W_{ub})$.

Lemma 2.2. *The set M_l^{int} is an interval $\left(m_l^{lo}, m_l^{hi}\right)$ such that Equations (1.24) and (1.25) can be solved for $\mu = m_l^{lo}$ and $W_{l-1} = W_{lb}$ and for $\mu = m_l^{hi}$ and $W_{l-1} = W_{ub}$. If (W_{l-1}, λ) and $(\tilde{W}_{l-1}, \tilde{\lambda})$ satisfy the first order conditions (1.24) and (1.25) for μ and $\tilde{\mu}$, respectively, while $W_{lb} \leq W_{l-1} < \tilde{W}_{l-1} \leq W_{ub}$, then $\tilde{\lambda} < \lambda$ and $\tilde{\mu} > \mu$.*

Proof: The first statement follows from the second, whose proof follows. Suppose that a pair (W_{l-1}, λ) with $W_{l-1} < p$ solves Equations (1.24) and (1.25) for some μ. This implies $w_i = g_i(W_{l-1}, \lambda)$ as in Equation (1.19). If W_{l-1} is increased by δ to $\tilde{W}_{l-1} = W_{l-1} + \delta < p$, then $g_i(\tilde{W}_{l-1}, \lambda) < g_i(W_{l-1}, \lambda), \forall i = 1, \ldots, l-1$ and $\sum_{i=1}^{l-1} g_i(\tilde{W}_{l-1}, \lambda) < \sum_{i=1}^{l-1} g_i(W_{l-1}, \lambda) = W_{l-1}$. By Lemma 2.1, $\sum_{i=1}^{l-1} g_i(\tilde{W}_{l-1}, \lambda') = W_{l-1}$ can be satisfied only for a unique $\lambda' < \lambda$. Since $V_{[1]}, \ldots, V_{[k]}$ are sorted in ascending order, $(V_{[l]} - V_{[i]})$ is decreasing in i and the derivative of $g_i(W_{l-1}, \lambda)$ with respect to λ given in Equation (.11) is increasing in i. Therefore, for $\lambda' < \lambda$, $i < j$ implies $g_i(\tilde{W}_{l-1}, \lambda') - g_i(\tilde{W}_{l-1}, \lambda) >$

$g_j(\tilde{W}_{l-1}, \lambda') - g_j(\tilde{W}_{l-1}, \lambda)$. If $\sum_{i=1}^{l-1} g_i(\tilde{W}_{l-1}, \lambda') = \sum_{i=1}^{l-1} g_i(W_{l-1}, \lambda)$ is satisfied, then there exists $i_\delta < l-1$ such that $g_i(\tilde{W}_{l-1}, \lambda') \geq g_i(W_{l-1}, \lambda)$ for $i \leq i_\delta$ and $g_i(\tilde{W}_{l-1}, \lambda') < g_i(W_{l-1}, \lambda)$ for $i > i_\delta$. We define $\Delta := \sum_{i=1}^{i_\delta} g_i(\tilde{W}_{l-1}, \lambda') - \sum_{i=1}^{i_\delta} g_i(W_{l-1}, \lambda)$ to be the total increase of $w_1, \ldots, w_{i_\delta}$ and $\sum_{i=i_\delta+1}^{l-1} g_i(W_{l-1}, \lambda) - \sum_{i=i_\delta+1}^{l-1} g_i(\tilde{W}_{l-1}, \lambda') = -\Delta$ follows from the fact that both the new and original weights add up to W_{l-1}. We, then, plug the new weights in equation (1.25) to find the expected shortfall μ' corresponding to $(\tilde{W}_{l-1}, \lambda')$ by $p\left(\mu' + V_{[l]}\right) = \sum_{i=1}^{l-1} g_i(\tilde{W}_{l-1}, \lambda')(V_{[l]} - V_{[i]})$. Note that

$$
\begin{aligned}
p\left(\mu' + V_{[l]}\right) - p\left(\mu + V_{[l]}\right) &= \sum_{i=1}^{k} [g_i(\tilde{W}_{l-1}, \lambda') - g_i(W_{l-1}, \lambda)](V_{[l]} - V_{[i]}) \\
&= \sum_{i=1}^{i_\delta} [g_i(\tilde{W}_{l-1}, \lambda') - g_i(W_{l-1}, \lambda)](V_{[l]} - V_{[i]}) \\
&\quad + \sum_{i=i_\delta+1}^{k} [g_i(\tilde{W}_{l-1}, \lambda') - g_i(W_{l-1}, \lambda)](V_{[l]} - V_{[i]}) \\
&\geq \sum_{i=1}^{i_\delta} [g_i(\tilde{W}_{l-1}, \lambda') - g_i(W_{l-1}, \lambda)](V_{[l]} - V_{[i_\delta]}) \qquad (.12) \\
&\quad + \sum_{i=i_\delta+1}^{k} [g_i(\tilde{W}_{l-1}, \lambda') - g_i(W_{l-1}, \lambda)](V_{[l]} - V_{[i_\delta+1]}) \\
&= \Delta(V_{[l]} - V_{[i_\delta]}) - \Delta(V_{[l]} - V_{[i_\delta+1]}) \\
&= \Delta(V_{[i_\delta+1]} - V_{[i_\delta]}) \geq 0
\end{aligned}
$$

which implies $\mu' \geq \mu$. Inequality (.12) follows since

$$
(V_{[l]} - V_{[i_\delta]}) \leq (V_{[l]} - V_{[i]}) \text{ and } g_i(\tilde{W}_{l-1}, \lambda') - g_i(W_{l-1}, \lambda) \geq 0\,,\ \forall i \leq i_\delta \text{ and}
$$

$$
(V_{[l]} - V_{[i_\delta+1]}) \geq (V_{[l]} - V_{[i]}) \text{ and } g_i(\tilde{W}_{l-1}, \lambda') - g_i(W_{l-1}, \lambda) \leq 0\,,\ \forall i \geq i_\delta + 1.
$$

Equation (1.24) for $\tilde{W}_{l-1}$ becomes $\sum_{i=1}^{l-1} g_i(\tilde{W}_{l-1}, \tilde{\lambda}) = \tilde{W}_{l-1} > W_{l-1}$, which can be satisfied only for a unique $\tilde{\lambda} < \lambda'$. Due to monotonicity, $g_i(\tilde{W}_{l-1}, \tilde{\lambda}) > g_i(\tilde{W}_{l-1}, \lambda')\,, \forall i = 1, \ldots, l-1$ and this implies for the expected shortfall $\tilde{\mu}$ corresponding to $(\tilde{W}_{l-1}, \tilde{\lambda})$

$$\begin{aligned} p\left(\tilde{\mu} + V_{[l]}\right) &= \sum_{i=1}^{l-1} g_i(\tilde{W}_{l-1}, \tilde{\lambda})(V_{[l]} - V_{[i]}) \\ &> \sum_{i=1}^{l-1} g_i(\tilde{W}_{l-1}, \lambda')(V_{[l]} - V_{[i]}) \\ &= p\left(\mu' + V_{[l]}\right) \\ &\geq p\left(\mu + V_{[l]}\right). \end{aligned}$$

Hence, we showed that if (W_{l-1}, λ) and $(\tilde{W}_{l-1}, \tilde{\lambda})$ satisfy the first order conditions (1.24) and (1.25) for μ and $\tilde{\mu}$, respectively, while $\tilde{W}_{l-1} > W_{l-1}$, then $\tilde{\mu} > \mu$.

We complete the proof of the lemma by showing that there exist $W_0 \in (W_{lb}, W_{ub})$, λ_0, and μ_0 such that Equations (1.24) and (1.25) are solved with $(W_{l-1}, \lambda) = (W_0, \lambda_0)$ and $\mu = \mu_0$; this proves that M_l^{int} is nonempty. Define $f(W_{l-1}, \lambda) := W_{l-1} - \sum_{i=1}^{l-1} g_i(W_{l-1}, \lambda)$, which is increasing in λ by Lemma 2.1. For any $W_0 \in (W_{lb}, W_{ub})$, $\lim_{\lambda \to \infty} f(W_0, \lambda) = W_0$. Let λ_{lb} be the solution of $g_1(W_0, \lambda_{lb}) = W_0$. Then $0 < g_i(W_0, \lambda_{lb}) \leq W_0\,,\ \forall i > 1$ because the absolute value of the derivative of $g_i(W_0, \lambda)$ with respect to λ given in equation (.11) is decreasing in i. Consequently, $f(W_0, \lambda_{lb}) = W_0 - W_0 - \sum_{i=2}^{l-1} g_i(W_0, \lambda) < 0$. Because f is continuous in its second argument over the range $[\lambda_{lb}, \infty)$, there exists λ_0 such that $f(W_0, \lambda_0) = 0$, i.e. Equation (1.24) holds for $W_{l-1} = W_0$ and $\lambda = \lambda_0$. Then μ_0 is chosen to satisfy Equation (1.25). $\square$

The following lemma justifies the way in which root-finding is used to determine the endpoints of confidence intervals and the rectangles that make up confidence regions.

Lemma 2.3. *If*

- $l \leq kp$*:* R_l^{int} *is decreasing on* (m_l^{lo}, m_l^{hi})*, and the supremum of* R_l^{int} *on* (m_l^{lo}, m_l^{hi}) *is* $R_l^{int}(m_l^{lo})$*, where the first-order conditions with* $\mu = m_l^{lo}$ *are solved at* $W_{l-1}^* = W_{lb} = p - (1-p)/(k-l)$ *and* $\lambda^* < 0$*.*
- $kp < l < kp+1$*:* R_l^{int} *is increasing on* $(m_l^{lo}, T(F_k))$ *and decreasing on* $(T(F_k), m_l^{hi})$*, and the supremum of* R_l^{int} *on* (m_l^{lo}, m_l^{hi}) *is* $R_l^{int}(T(F_k)) = 1$*, where the first-order conditions with* $\mu = T(F_k)$ *are solved at* $W_{l-1}^* = (l-1)/k$ *and* $\lambda^* = 0$*.*
- $l \geq kp + 1$*:* R_l^{int} *is increasing on* (m_l^{lo}, m_l^{hi}) *and the supremum of* R_l^{int} *on* (m_l^{lo}, m_l^{hi}) *is* $R_l^{int}(m_l^{hi})$*, where the first-order conditions with* $\mu = m_l^{hi}$ *are solved at* $W_{l-1}^* = W_{ub} = p$ *and* $\lambda^* > 0$*.*

Proof: Consider the set $\mathcal{F}_l^*$ of all points $(W_{l-1}^*, \lambda^*, \mu)$ such that the first-order conditions of Maximization Problem II are satisfied, for $W_{l-1}^* \in (0,1)$. This includes the point $(\frac{l-1}{k}, 0, T(F_k))$, corresponding to equal weights $w_1^*, \ldots, w_k^* = 1/k$.

This point is feasible iff $\frac{l-1}{k} \in (W_{lb}, W_{ub})$, that is, $kp < l < kp + 1$, and in this case $R_l^{int}(T(F_k)) = 1$, the largest possible empirical likelihood ratio. It follows from Lemma 2.2 that if (W_{l-1}, λ, μ) and $(\tilde{W}_{l-1}, \tilde{\lambda}, \tilde{\mu})$ are in $\mathcal{F}_l^*$ while $\mu < \tilde{\mu} < T(F_k)$, then $W_{l-1} < \tilde{W}_{l-1} < \frac{l-1}{k}$ and $\lambda > \tilde{\lambda} > 0$. Because $W_{l-1} < \tilde{W}_{l-1} < \frac{l-1}{k}$, the average weight in the tail (i.e. w_i for $i = 1, \ldots, l-1$) is "too small", that is, less than $1/k$, in the solution to Maximization Problem II with $\tilde{\mu}$ and it is even less in the solution with μ. Because $\lambda > \tilde{\lambda} > 0$, the weights in the tail are unequal to each other in the solutions to Maximization Problem II

with μ or with $\tilde{\mu}$, and they are more distorted in the solution with μ. Both of these effects cause $R_l^{int}(\mu) < R_l^{int}(\tilde{\mu})$. Thus R_l^{int} is increasing on $(-V_{[k]}, T(F_k)]$. By similar reasoning, it is decreasing on $[T(F_k), -V_{[1]})$.

Next consider the case $l \geq kp + 1$. In this case, $(\frac{l-1}{k}, 0, T(F_k))$ is infeasible because $\frac{l-1}{k} \geq p = W_{ub}$. By Lemma 2.2, $(m_l^{lo}, m_l^{hi}) \subset (-V_{[k]}, T(F_k)]$, and the conclusion follows.

Finally, consider the case $l \leq kp$, where $(\frac{l-1}{k}, 0, T(F_k))$ is infeasible because $\frac{l-1}{k} \leq W_{lb}$. The conclusion follows in a similar manner from $(m_l^{lo}, m_l^{hi}) \subset (T(F_k), -V_{[1]})$. $\square$