The Dissertation Committee for Amit Partani
certifies that this is the approved version of the following dissertation:

Adaptive Jackknife Estimators for Stochastic Programming

Committee:

David P. Morton, Supervisor

Jonathan F. Bard

John J. Hasenbein

Elmira Popova

Thomas Sager

Adaptive Jackknife Estimators for Stochastic Programming

by

Amit Partani, B.Tech.; MSE

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

December 2007

Dedicated to my parents and beautiful wife Aditi

Acknowledgments

I first like to thank the Almighty for such a wonderful time I had during my PhD. The biggest source of motivation for me are my parents, my wife Aditi, my kins Arun and Anju. I thank their endless support and belief in me. I would not have done so much without the great effort that my parents put in me. I thank my wonderful supervisor Dr. Morton without whom my thesis would have lacked purpose. He has always been helpful and understanding. I am grateful to the wonderful faculty at UT for the knowledge that I have gained from them. A special thanks to Vishv Jeet for being a great colleague and all the motivational talks that he gave to me. Next comes a big list of my friends with whom I enjoyed every second of my time at UT specially Titash, Balaji, Burak, Ankur, Rohin, Mohit, Anubhav, Kranthi and many more. I thank everyone who has directly or indirectly been part of my life during my PhD.

Amit Partani

The University of Texas at Austin

December 2007

Adaptive Jackknife Estimators for Stochastic Programming

Publication No. ______________

Amit Partani, Ph.D.
The University of Texas at Austin, 2007

Supervisor: David P. Morton

Stochastic programming facilitates decision making under uncertainty. It is usually impractical or impossible to find the optimal solution to a stochastic program, and approximations are required. Sampling-based approximations are simple and attractive, but the standard point estimate of the optimal value of a stochastic program contains bias due to the interaction of optimization and the Monte Carlo approximation. We provide a method to reduce this bias, and hence provide a better, i.e., tighter, confidence interval on the optimal value and on a candidate solution's optimality gap. Our method requires less restrictive assumptions on the structure of the bias than previously-available estimators. Our estimators adapt to problem-specific properties, and we provide a family of estimators, which allows flexibility in choosing the level of aggressiveness for bias reduction. We establish desirable statistical

properties of our estimators and empirically compare them with known techniques on test problems from the literature.

Contents

List of Tables

List of Figures

Chapter 1

Introduction and Motivation

1.1 Introduction

We consider the problem of estimating the optimal value of a stochastic program. Estimates of the optimal value play a key role in assessing the quality of a candidate solution to a stochastic program, and sometimes the optimal value itself is of primary interest. The standard Monte Carlo estimate of the optimal value is biased, and this bias can hamper our ability to develop sufficiently tight interval estimates of a candidate solution's optimality gap and the stochastic program's optimal value. Standard bias correction procedures, such as the jackknife, can be difficult to apply because they require knowing the asymptotic form of the bias, and as we describe, this is not typically known in a stochastic program. So, we develop a class of adaptive jackknife estimators. They are adaptive in that they estimate the rate at which the bias shrinks to zero.

Mathematical programming models to capture optimization under uncertainty were pioneered by Dantzig [12] and Beale [4]. They proposed incorporating randomness in what have come to be known as stochastic programs with recourse. Since then the field of stochastic programming has added many dimensions. A well-known example of a stochastic program is the newsvendor problem with uncertain demand. The vendor does not know the demand before hand and must decide how many newspapers to buy. After this decision is made, the demand is realized and a loss in revenue or salvage cost is incurred in the case of excess demand or demand shortfall, respectively. The problem is to decide how many newspapers to buy to maximize the expected profit. Problems ranging from maximum-likelihood estimation to optimal portfolio selection can be viewed as stochastic programs. See Wallace & Ziemba [56] for a variety of applications where we can model the underlying problem using stochastic programming.

A wide class of stochastic programming problems involves making a decision at time 0 based on certain constraints. At time 1 we realize additional information and re-optimize the system.

Let $x \in X \subseteq \Re^{d_x}$ be a feasible first stage decision and its constraint set, $\tilde{\xi}$ be a d_{ξ}-dimensional random vector defined on probability space $\left(\Xi, \mathcal{F}, \mathcal{P}\right)$ and $f : X \times \Xi \rightarrow \Re$ be the cost function associated with decision x and a realization of the random vector $\tilde{\xi}$. Then a stochastic program can be formulated as

$$z^* = \min_{x \in X} Ef(x, \tilde{\xi}), \tag{1.1}$$

where E is the expectation operator. We will assume the following throughout this dissertation that:

(A1). *X is closed, nonempty, and compact.*

(A2). *$f(x, \cdot)$ is continuous on Ξ $\forall x \in X$, and $f(\cdot, \xi)$ is lower semicontinuous on X $\forall \xi \in \Xi$.*

(A3). $E \sup_{x \in X} \left[f(x, \xi)\right]^2 < \infty$.

Many variants of this problem can be formulated by

- replacing the expectation operator by some other real-valued operator on $f(x, \tilde{\xi})$ or
- imposing special structure on $f(x, \tilde{\xi})$ and X.

A special case of (1.1) is the two-stage stochastic linear program (SLP), where the constraint set X is a polyhedron and

$$\begin{aligned} f(x, \tilde{\xi}) = cx + \min_{y \geq 0} \ & \tilde{q}y \\ \text{s.t. } & \tilde{D}y = \tilde{B}x + \tilde{d}, \end{aligned} \tag{1.2}$$

where $\tilde{\xi} = (\tilde{q}, \tilde{D}, \tilde{B}, \tilde{d})$ is the random vector. Extensive research has been done on numerical solution procedures for such SLPs. As the number of realizations of $\tilde{\xi}$ grows large, or when $\tilde{\xi}$ does not have finite support, or the problem deviates from linearity then these algorithms may not apply. For a case where $d_\xi = 8$, and the components of $\tilde{\xi}$ are independent with 4 realizations each, we need to solve $4^8 = 65,536$ second stage problems (1.2) simply to evaluate $Ef(x, \tilde{\xi})$ for fixed $x \in X$. (Solving (1.1) with f defined in (1.2) is obviously more difficult.) Their sheer size demands special attention be paid to these kind of problems. To summarize, the potential difficulties associated with a stochastic program are:

- it can be computationally expensive to calculate $f(x, \tilde{\xi})$ for a given decision x and realization $\tilde{\xi}$ of the uncertainties;
- even if $f(x, \tilde{\xi})$ can be computed easily, many times it is impossible or impractical to calculate $Ef(x, \tilde{\xi})$ exactly; and,
- available solution techniques require evaluating the objective function at many feasible points.

A sampling-based approximation of the "true" problem (1.1) may be an appropriate way to attempt to overcome the above difficulties. Let $\tilde{\xi}^1, \ldots, \tilde{\xi}^n$ be a sample from the underlying probability distribution and $\hat{f}(x, \tilde{\xi}^1, \ldots, \tilde{\xi}^n)$ be the sampling-based estimator of $Ef(x, \tilde{\xi})$. Then, the sampling approximation problem can be stated as,

$$z_n^* = \min_{x \in X} \hat{f}(x, \tilde{\xi}^1, \ldots, \tilde{\xi}^n). \tag{1.3}$$

Let (x_n^*, z_n^*) and (x^*, z^*) denote a pair of optimal solution and optimal value to the sampling approximation problem (1.3) and the true problem (1.1), respectively. Assumptions (A1)-(A3) ensure that model (1.1) has a finite optimal solution which is achieved on X. We assume $\hat{f}(\cdot, , \tilde{\xi}^1, \ldots, \tilde{\xi}^n)$ is lower semicontinuous so that this holds for model (1.3) too. Then, the issues which need to be addressed include:

1. What is the limiting behavior of x_n^* and z_n^* as the sample size grows, relative to their counterparts in the true problem?

2. How should the approximation problem be solved?

3. How should solutions obtained from the approximation problem be validated?

As we review in the next section, the first two issues have been studied extensively. We plan to address the third issue, i.e., solution validation. The optimal value z_n^* of problem (1.3) provides a statistical estimate of the true optimal value z^*. However, this estimate is usually biased. We define the bias caused by the sampling approximation as,

$$b(z_n^*) = Ez_n^* - z^*. \tag{1.4}$$

Often in estimation, when bias arises, it shrinks to zero as $O(n^{-1})$. We show that for $b(\cdot)$ while $O(n^{-1})$ is possible, other rates are also possible. As in other areas of optimization, lower bounds in stochastic programming (for minimization problems) prove useful in validating the quality of a candidate solution. If $E\hat{f}(x, \tilde{\xi}^1, \ldots, \tilde{\xi}^n) = Ef(x, \tilde{\xi})$ $\forall x \in X$ then $Ez_n^* \leq z^*$, i.e., z_n^* is a lower bound estimator. However, we show that the bias of z_n^* can significantly degrade our ability to assess the quality of a candidate solution. Therefore, we develop techniques to reduce this bias. Our approach is rooted in jackknife estimators. Desirable asymptotic properties of our estimators are shown, and tested on some standard problems.

1.2 Background

We provide background on some existing results for sampling approximation methods in stochastic programming. There is a significant literature on these type of results, and it is not our purpose to give a comprehensive review. See, for example, the chapters of Shapiro [52] and Pflug [39].

1.2.1 Limiting properties of (x_n^*, z_n^*)

Even if $\hat{f}(\cdot, \tilde{\xi}^1, \ldots, \tilde{\xi}^n) \to Ef(\cdot, \tilde{\xi})$ pointwise w.p.1. (with probability 1), the convergence $z_n^* \to z^*$ is not guaranteed. For example, consider the problem where $f(x, \tilde{\xi}) = \tilde{\xi}x, X = \Re$, and $\tilde{\xi} \sim N(0,1)$. Let $\tilde{\xi}^1, \ldots, \tilde{\xi}^n$ be an i.i.d. (independent and identically distributed) sample of $\tilde{\xi}$ and define $\hat{f}(x, \tilde{\xi}^1, \ldots, \tilde{\xi}^n) = (\frac{1}{n}\sum_{i=1}^n \tilde{\xi}^i)x$. It is easy to see that $\hat{f}$ converges to Ef for any fixed x, however $z_n^* = -\infty$, w.p.1., for all n and thus does not converge to $z^* = 0$. Hence, further conditions must be imposed to guarantee the convergence of (x_n^*, z_n^*). Below, we give a consistency result of [52] when the sample $\tilde{\xi}^1, \ldots, \tilde{\xi}^n$ is i.i.d. as $\tilde{\xi}$.

Theorem 1 (Consistency, Theorem 10 [52])**.** *Consider problems (1.1) and (1.3). Let the sample $\tilde{\xi}^1, \ldots, \tilde{\xi}^n$ be i.i.d. as $\tilde{\xi}$, $\hat{f}(x, \tilde{\xi}^1, \ldots, \tilde{\xi}^n)$ be the sample mean estimator of $f(x, \tilde{xi})$ and z_n^* be the optimal value of (1.3). Suppose there exists a measurable function $K : \Xi \to \Re_+$ such that $E([K(\tilde{\xi})]^2)$ is finite and $|f(x_1, \xi) - f(x_2, \xi)| \leq K(\xi)||x_1 - x_2||$, for all $x_1, x_2 \in X$ and $\xi \in \Xi$. If (A1)-(A3) holds, then, $\lim_{n\to\infty} z_n^* = z^*$ w.p.1.*

Various conditions under which $(x_n^*,\ z_n^*)$ converge, in some sense, to true optimal points can be found in [18] and [52]. Shapiro [51] establishes central-limit-theorem (CLT) results for the optimal value z_n^*; these require stronger hypotheses than what we have assumed above. Under even stronger assumptions King and Rockafellar [31] provide CLT results for the optimal solution x_n^*. Large deviations theory leads to exponential rates of convergence of x_n^* to x^* for certain classes of problems; see [29, 32, 53].

1.2.2 Solution techniques for sampling approximation problems

Many natural estimates $\hat{f}(x, \tilde{\xi}^1, \ldots, \tilde{\xi}^n)$ take the form $E_{\mathcal{P}_n} f(x, \tilde{\xi})$ for a discrete approximating distribution $\mathcal{P}_n$. For example, if $\hat{f}(x, \tilde{\xi}^1, \ldots, \tilde{\xi}^n) = \frac{1}{n} \sum_{i=1}^{n} f(x, \tilde{\xi}^i)$ and the sampling is i.i.d., then $\mathcal{P}_n$ is simply the discrete empirical distribution, which puts weight $\frac{1}{n}$ on each point $\tilde{\xi}^i, i = 1, \ldots, n$. Once $\tilde{\xi}^1, \ldots, \tilde{\xi}^n$ have been generated we may view (1.3) as a stochastic program with a possibly modest number of realizations. For example, if f takes form (1.2) and X is polyhedral then (1.3) can be recast as a large-scale linear program, and we can attempt to solve it directly using commercially-available linear programming algorithms. Or, we can employ special purpose algorithms such as the L-shaped method [55] or enhancements of this decomposition algorithm [7, 46, 47]. If the first stage decisions x and/or the second stage decisions y are subject to integrality restrictions we can attempt to solve (1.3) as a large-scale mixed integer program or via special purpose algorithms. See, for example, the survey of Louveaux and Schultz [33]. In this type of approach we use the computational machinery which has been developed (independently of sampling based methods) over the last several decades to solve instances of (1.1) in which there are a modest number of realizations. Birge [6] and Kall and Mayer [28] survey the state-of-the art algorithms for solving (1.1) or its sampling approximation (1.3), given $\tilde{\xi}^1, \ldots, \tilde{\xi}^n$.

There is another approach that we will not explicitly discuss, which involves internal-sampling algorithms, i.e., algorithms in which Monte Carlo sampling is used *within* the algorithm to estimate function values and (sub)gradients. Examples of this approach include the classic adaptations of steepest descent by Robbins and Monro [45] and Kiefer and Wolfowitz [30] as well as modern adaptations of the L-shaped method by Higle and Sen [26] and by Dantzig, Glynn and Infanger [13, 14]. The latter set of methods can gain computational advantage by intelligent integration of optimization and sampling.

1.2.3 Validation of candidate solutions via optimality gap estimation

Given the consistency results and solution techniques for sampling-based approximations it becomes necessary to assess the quality of a candidate solution which is obtained with a finite sample size. The issue we face has analogs in other areas of optimization. In integer and nonlinear programming relaxation-based lower bounds (for minimization problems) are used to help bound the optimality gap of a candidate solution. Lower bounds obtained by relaxing integrality restrictions or complicating constraints in such settings are deterministically valid. Sometimes deterministically-valid lower bounds can be used in stochastic programs, e.g., via Jensen's inequality, but they require special structures and can be difficult to tighten.

Let $\hat{x} \in X$ be a given solution whose quality we wish to assess. We define quality in terms of the optimality gap, $\mu_{\hat{x}} = Ef(\hat{x}, \tilde{\xi}) - z^*$. The candidate solution could be obtained by solving a sampling-based problem of form (1.3). Or, $\hat{x} \in X$ could be obtained by running an internal-sampling algorithm. Or, it could be obtained by solving the expected-value problem, i.e., the single scenario problem with that scenario defined by $E\tilde{\xi}$ or some variant thereof [36]. The procedures we describe do not depend on the method by which $\hat{x}$ is found, although if it is found by a sampling-based algorithm, the sampling done in our estimation of the optimality gap will be independent of that done for obtaining $\hat{x}$. Our aim is to estimate the optimality gap, $Ef(\hat{x}, \tilde{\xi}) - z^*$, but we do not know z^*, and so we replace it with a lower bound provided by the following theorem [34, 35].

Theorem 2 (Lower Bound). *1. Let $\tilde{\xi}^1, \dots, \tilde{\xi}^n$ satisfy*

$$E\hat{f}(x, \tilde{\xi}^1, \dots, \tilde{\xi}^n) = Ef(x, \tilde{\xi}) \ \forall x \in X. \tag{1.5}$$

Then z_n^* as defined in (1.3) satisfies $Ez_n^* \leq z^*$.

2. *Let $\tilde{\xi}^1, \ldots, \tilde{\xi}^n$ be i.i.d. as $\tilde{\xi}$, and let $\hat{f}(x, \tilde{\xi}^1, \ldots, \tilde{\xi}^n) = \frac{1}{n}\sum_{i=1}^n f(x, \tilde{\xi}^i)$. Then,*

$$Ez_n^* \leq Ez_{n+1}^* \leq z^*.$$

Given that $\hat{f}$ is an unbiased estimator for a fixed value of x, the result in part 1 of Theorem 2 follows by sub-optimality of x^* in the sampling problem, and the bound may be viewed as a relaxation arising from exchanging the order of optimization and expectation. Part 2 of the theorem shows that under the standard sample mean estimator the bound tightens in expectation as the sample size n grows. These results hold quite generally, without requiring special structure of f, X or $\tilde{\xi}$ beyond existence of the associated expectations. We know from Theorem 1 that consistency of z_n^* follows under somewhat stronger assumptions. By Theorem 2 and $\hat{x} \in X$ we have that $Ef(\hat{x}, \tilde{\xi}) - Ez_n^* \geq \mu_{\hat{x}}$, and an estimate for this upper bound on the optimality gap is

$$G_n(\hat{x}) = \hat{f}(\hat{x}, \tilde{\xi}^1, \ldots, \tilde{\xi}^n) - \min_{x \in X} \hat{f}(x, \tilde{\xi}^1, \ldots, \tilde{\xi}^n), \tag{1.6}$$

where $\tilde{\xi}^1, \ldots, \tilde{\xi}^n$ satisfy (1.5). Note that both terms on the right-hand side of (1.6) use the same set of n observations, and as a result $G_n(\hat{x}) \geq 0$ due to suboptimality of $\hat{x}$ with respect to this sample. This is desirable property since we seek to bound $\mu_{\hat{x}} \geq 0$. Since the distribution of $G_n(\hat{x})$ is not known (in general, it can be non-normal), a confidence interval (CI) for $\mu_{\hat{x}}$ can be constructed by forming i.i.d. replications of $G_n(\hat{x})$ and using the standard central limit theorem. Confidence intervals based on single and multiple replication procedures are proposed by Bayraksan and Morton [3] and Mak et al. [34], respectively. We provide below the multiple replication procedure (MRP) given by Mak et al.

Multiple Replication Procedure (MRP^o)

Input: sample size n, replications m, confidence level $(1-\alpha)$, candidate solution $\hat{x} \in X$

Output: approximate $(1-\alpha)$-level CI on the optimality gap, $\mu_{\hat{x}}$

1. For $i = 1, \ldots, m$

   - Generate a random sample $\tilde{\xi}^1, \ldots, \tilde{\xi}^n$ satisfying (1.5)
   - Let $z_n^{i*} = \min_{x \in X} \hat{f}(x, \tilde{\xi}^1, \ldots, \tilde{\xi}^n)$
   - Let $G_n^i = \hat{f}(\hat{x}, \tilde{\xi}^1, \ldots, \tilde{\xi}^n) - z_n^{i*}$

2. Let
$$\bar{G}_m = \frac{1}{m}\sum_{i=1}^m G_n^i \quad \text{and} \quad s_m^2 = \frac{1}{m-1}\sum_{i=1}^m (G_n^i - \bar{G}_m)^2$$

3. Form CI for $\mu_{\hat{x}}$ as,
$$\left[0, \bar{G}_m + \frac{t_{m-1,\alpha} s_m}{\sqrt{m}}\right]$$

The samples generated in step 1 are independent in each of the m iterations, although the $\tilde{\xi}^1, \ldots, \tilde{\xi}^n$ within an iteration need not be independent. In step 3 of the MRP we use $t_{m-1,\alpha}$, which is the $(1-\alpha)$-quantile of a t random variable with $m-1$ degrees of freedom, i.e., $P(-t_{m-1,\alpha} \leq T_{m-1}) = 1-\alpha$. By the standard central limit theorem we infer that when m is sufficiently large

$$P(Ef(\hat{x}, \tilde{\xi}) \leq z^* + \tilde{\epsilon}_G) \geq 1 - \alpha, \tag{1.7}$$

where $\tilde{\epsilon}_G = \bar{G}_m + t_{m-1,\alpha} s_m / \sqrt{m}$. So, if the (random) CI width $\tilde{\epsilon}_G$ is sufficiently small we infer $\hat{x}$ is a high quality solution with (approximate) probability $1-\alpha$.

1.3 Motivation for reducing bias

Associated with the CI constructed by the MRP of the previous section is statement (1.7) regarding the quality of the given candidate solution $\hat{x}$. Tighter confidence interval widths $\tilde{\epsilon}_G$ allow us to make better assessments. This CI can be decomposed into three parts.

1. Suboptimality of the candidate solution, $Ef(\hat{x}, \tilde{\xi}) - z^*$
2. Bias induced by using the lower bound, $z^* - Ez_n^* = -b(z_n^*)$
3. Variance induced by sampling

The purpose of constructing the CI is to obtain an interval estimate of the first part, i.e., the optimality gap. Though it is desirable to find techniques which give better candidate solutions, i.e., solutions with smaller optimality gaps, the scope of this research is to focus on obtaining precise interval estimates for a given candidate solution. A number of authors have investigated techniques for reducing variance in Monte Carlo estimators for stochastic programming [2, 13, 15, 16, 24, 27, 38], i.e., to help reduce the contribution of issue 3. As we show later, the bias term sometimes dominates the width of the MRP confidence interval. So, our motivation lies in addressing issue 2, i.e., to reduce the bias. Of course, there is typically a trade-off between bias and sampling error, and we will investigate it via the mean square error. Taken together, Theorem 1 regarding consistency and Theorem 2 regarding bias suggest that the optimal value z_n^* of the sampling problem (1.3) converges to z^* from below. Often statistical estimation bias shrinks to zero as $O(n^{-1})$ as the sample size n grows. (In particular, this is true when the estimator may be viewed as a smooth nonlinear function of a sample mean.) However, we show in the following example that in our setting bias can shrink to zero at rate $O(n^{-p})$, where p can take on any value from 1/2 to ∞.

Example 1. *Consider the following instance of (1.1):*

$$z^* = \min_x \left(Ef(x, \tilde{\xi}) = E\left\{ \tilde{\xi}x + |x|^\delta \right\} \right), \tag{1.8}$$

where $\delta > 1$ and $\tilde{\xi} \sim N(0,1)$, i.e., $\tilde{\xi}$ is a standard normal random variable. Clearly, $z^* = 0$ and $x^* = 0$. However, if $\tilde{\xi}^1, \ldots, \tilde{\xi}^n$ are i.i.d. as $\tilde{\xi}$ and $\hat{f}$ is the sample mean we obtain (1.3) as:

$$z_n^* = \min_x \left(\hat{f}(x, \tilde{\xi}^1, \ldots, \tilde{\xi}^n) = \bar{\xi}_n x + |x|^\delta \right), \tag{1.9}$$

where $\bar{\xi}_n$ is a sample mean of n i.i.d. standard normals. Problem (1.9) has optimal solution and value

$$\begin{aligned} x_n^* &= (-1)^{I_{\{\bar{\xi}_n > 0\}}} \left(\frac{|\bar{\xi}_n|}{\delta} \right)^{\frac{1}{\delta-1}} \\ z_n^* &= -(\delta - 1) \left(\frac{|\bar{\xi}_n|}{\delta} \right)^{\frac{\delta}{\delta-1}} \sim -\delta^{-\frac{\delta}{\delta-1}} (\delta-1) \, |N(0,1)| \, n^{-p}, \end{aligned}$$

where $I_{\{\cdot\}}$ is the indicator function and $p = \frac{\delta}{2(\delta-1)}$. Taking expectations we obtain $b(z_n^*) = Ez_n^* = -an^{-p}$, where $a > 0$ is a constant independent of n. As $\delta \to \infty$, $p \downarrow 1/2$ and as $\delta \downarrow 1$, $p \to \infty$. $\square$

Example 1 shows that for a stochastic program of form (1.8), we can obtain a bias of $O(n^{-p})$ for any $p \in (1/2, \infty)$. $p = 1/2$ and ∞ can be obtained using i.i.d samples in examples $z^* = \min_{x \in [-1,1]} \left(Ef(x, \tilde{\xi}) = E[\tilde{\xi} x] \right)$, where $\tilde{\xi} \sim N(0,1)$, and $z^* = \min_{x \in [0,1]} \left(Ef(x, \tilde{\xi}) = E[\tilde{\xi} x] \right)$, where $\tilde{\xi} \sim N(1,1)$.

If the bias takes form $b(\cdot) = -an^{-p}$ for $p \in [1/2, \infty)$ then it would be natural to seek in a bias reduction technique an estimator that either: (i) effectively increases p, e.g., the new estimator has bias of form $O(n^{-(p+1)})$ or (ii) effectively shrinks the value of a. Typically one would prefer the former, i.e., an estimator that increases the rate at which the bias shrinks in n. However, in our case attempting to reduce the order of the bias can be too aggressive because we do not simply seek a better point estimate of z^*. Instead when using our estimator in a procedure for producing a confidence interval on the optimality gap we are averse to destroying the lower-bounding property of that estimator. We will return to this issue later.

1.4 Outline

Biased estimators arise frequently in statistics and simulation. Anderson et al. [1] describe a number of techniques for reducing bias. Jackknife and bootstrap estimates are widely used; see, e.g., [50]. There has been little work on bias reduction in stochastic programming, although exceptions include [11, 21].

In Chapter 2 we discuss the generalized jackknife estimator, which can be used to reduce $O(n^{-p})$ bias when p is known. The examples above show $O(n^{-p})$ bias can arise for a range of values of p when using z_n^* to estimate z^*. Unfortunately for a specific stochastic program we are unlikely to know the associated form of the bias. So, we propose a class of adaptive p-estimation jackknife estimators, which assume bias of the form $O(n^{-p})$ but does not require *a priori* specification of p. We compare the performance of these estimators on a simple asset allocation problem.

Unfortunately, these adaptive p-estimation estimators may fail to have basic properties like consistency. So, in Chapter 3 we lay the foundation for an adaptive estimator with properties that the p-estimation adaptive estimator can fail to exhibit. We provide consistency results and characterize bias properties of the new adaptive estimator. We also argue that under some mild conditions, the new adaptive estimator will preserve the conservative nature of naive estimators.

In Chapter 4 we develop three families of adaptive estimators, which allows one to choose more, or less, estimators as needed. We show that the most aggressive family completely eliminates bias under certain conditions. We then provide a number of properties that our estimators satisfy and discuss conditions when they outperform generalized jackknife estimators. We present numerical results to compare the family of estimators with the naive estimator and generalized jackknife estimator. We conclude and provide future research directions in Chapter 5.

1.5 Numerical performance measures

We use the following measures to compare the various estimators that we consider. Let θ be the parameter we wish to estimate and let the naive estimator, as in MRP^o, provide an estimator $\hat{\theta}$, which is a statistical upper bound on θ. To compare the statistical estimators that we consider in the following chapters we use the measures listed below.

1. Mean Square Error (MSE): $E(\hat{\theta} - \theta)^2$
2. Mean Square Error Positive (MSE$^+$): $E[(\hat{\theta} - \theta)^+]^2$
3. Mean Square Error Negative (MSE$^-$ = MSE - MSE$^+$): $E[(\hat{\theta} - \theta)^-]^2$
4. Probability (Estimator $< \theta$): $P(\hat{\theta} - \theta)$
5. Confidence Interval Widths
6. Schruben Coverage Plots

Usually a bias correction procedure leads to an increase in variance of the estimator and hence we consider MSE as a performance measure. However, as mentioned before, the naive estimators as used in MRP^o are statistical upper bounds on the true optimality gap. We would like to preserve this property and hence we also consider MSE$^-$ as a performance measure. MSE$^-$ measures the squared deviations of the estimators on the "wrong" side of true parameter. $P(\hat{\theta} - \theta)$ examines the frequency with which the estimators fall on the "wrong" side of the parameter in consideration. Schruben coverage plots compare the desired coverage from confidence intervals with the actual coverage that are obtained (see [48]). Our first four measures concern the point estimators and the last two measures involve the interval estimators.

Throughout this dissertation we will perform numerical experiments to assess the relative performance of the estimators we consider. Unless specified otherwise, these experiments were conducted using C code that employs callable libraries in CPLEX 9.0 and run on a computer with a 1.8GHz Xeon processor and 1GB of RAM.

Chapter 2

Generalized Jackknife Estimator

2.1 Generalized jackknife estimator

2.1.1 Introduction

The jackknife estimator was introduced by Quenouille [42, 43], with early work due to Durbin and Quenouille [19, 20, 44]. Tukey named the jackknife estimator and broadened its scope [54]. Since then it has been widely used in application areas including demographic and biological studies and used to improve tools in regression and simulation [5, 8, 17, 23, 40, 41]. Along with the bootstrap, it is one of the most commonly used resampling plans. The idea behind the jackknife estimator is to obtain two biased but highly correlated estimators of an unknown parameter and to try to remove or decrease the bias by subtracting one estimator from the other, using an appropriate proportionality constant to adjust the estimator towards the parameter of interest. The most widely-used form of the jackknife estimator assumes bias shrinks to zero as $O(n^{-1})$, but the generalized jackknife estimator allows for the bias to take different forms; see Gray and Schucany [22] and Shao and Tu [50]. Let θ be the parameter of interest. Let $\hat{\theta}^1$ and $\hat{\theta}^2$ be two *biased* estimators of θ. The generalized jackknife estimator is defined as,

$$J^G = \frac{\hat{\theta}^1 - R\hat{\theta}^2}{1 - R}, \tag{2.1}$$

where R does not depend on the sample observations. The degree (if any) of bias reduction depends on appropriate choices of the estimators $\hat{\theta}^1$ and $\hat{\theta}^2$ and on R. We now turn to a specific forms of (2.1) developed by Quenouille [44].

2.1.2 Delete-1 estimator

Let $\hat{\theta}_n$ be the naive estimator based on a "full" sample of size n. Let $\hat{\theta}^i_{n-1}$ be the same estimator based on the same sample but with the i^{th} observation deleted, and define $\hat{\phi}_{n-1} = \frac{1}{n}\sum_{i=1}^{n}\hat{\theta}^i_{n-1}$. Let $\bar{\phi}_{n-1}$ and $\bar{\theta}_n$ be the average of $\hat{\phi}_{n-1}$ and $\hat{\theta}_n$, respectively, over m i.i.d. replications. Choose $\theta^1 = \bar{\theta}_n$, $\theta^2 = \bar{\phi}_{n-1}$, and $R = \left(\frac{n-1}{n}\right)^q$, where q is representative of how fast the bias in $\hat{\theta}_n$ shrinks to zero. (We return to the issue of choosing q in detail below.) Substituting into (2.1), we obtain the jackknife estimator of Quenouille,

$$J_q^Q = \frac{n^q \bar{\theta}_n - (n-1)^q \bar{\phi}_{n-1}}{n^q - (n-1)^q}. \tag{2.2}$$

The estimators $\hat{\theta}^1$ and $\hat{\theta}^2$ used by this method are highly correlated, have similar bias structure and if q is chosen well then the bias of J_q^Q can be smaller than that of the original estimator $\hat{\theta}_n$. This estimator is called a delete-1 estimator.

We now demonstrate the effectiveness of the jackknife estimator (2.2), on some simple examples, with $m = 1$.

Example 2. *Let Y be a random variable with finite variance σ^2, and let $Y^1, \ldots, Y^n$ be i.i.d. as Y. We want to estimate the variance and $\sigma^2 = E(Y - EY)^2$ motivates use of $\hat{\theta}_n = \frac{1}{n}\sum_{i=1}^n \left(Y^i - \bar{Y}_n\right)^2$, where $\bar{Y}_n$ is the sample mean. We know $\hat{\theta}_n$ is a biased estimator of σ^2 with $E\hat{\theta}_n = \sigma^2\left(1 - \frac{1}{n}\right)$. We choose $q = 1$ and apply the jackknife estimator (2.2). After some algebraic simplifications, we obtain the jackknife estimator as $J_1^Q = \frac{1}{n-1}\sum_{i=1}^n \left(Y^i - \bar{Y}_n\right)^2$, which we know is an unbiased estimator of σ^2.* □

Example 3. *Let Y be a Bernoulli random variable with success probability α, and let $Y^1, \ldots, Y^n$ be i.i.d. as Y. We want to estimate α^2. A natural estimator is $\hat{\theta}_n = \left(\frac{\sum_{i=1}^n Y^i}{n}\right)^2$. This estimator is biased as $E\hat{\theta}_n = \alpha^2 + \frac{1}{n}(\alpha - \alpha^2)$. Choosing $q = 1$ and applying (2.2) we get $J_1^Q = \frac{\left(\sum_{i=1}^n Y^i\right)\left(\sum_{i=1}^n Y^i - 1\right)}{n(n-1)}$, which is an unbiased estimator of α^2.* □

Example 4. *Let Y be a uniform random variable with support $(0, \alpha)$, and let $Y^1, \ldots, Y^n$ be i.i.d. as Y. We want to estimate α, and a natural estimator is $\hat{\theta}_n = Y_{(n)} = \max\{Y^1, \ldots, Y^n\}$. However, it is biased as $E\hat{\theta}_n = \alpha - \left(\frac{1}{n+1}\right)\alpha$. We again choose $q = 1$ in (2.2) and obtain $J_1^Q = Y_{(n)} + \left(\frac{n-1}{n}\right)\left(Y_{(n)} - Y_{(n-1)}\right)$ with*

$EJ_1^Q = \alpha\left(1 - \frac{1}{n(n+1)}\right)$. *Though we are not able to remove the bias completely, we have increased the rate at which the bias shrinks to zero from* $O(n^{-1})$ *to* $O(n^{-2})$. $\square$

Example 5. *Let* $Y \sim N(0, \sigma^2)$, *and let* $Y^1, \ldots, Y^n$ *be i.i.d. as* Y. *We want to estimate* $(EY)^4$. *A natural estimator is* $\hat{\theta}_n = \left(\frac{1}{n}\sum_{i=1}^n Y^i\right)^4$. *If we choose* $q = 1$ *and apply (2.2), we obtain* $EJ_1^Q = -\frac{3\sigma^4}{n(n-1)}$. *Here by applying the usual jackknife estimator, i.e., with* $q = 1$, *we have actually increased the bias in magnitude and reversed the sign of the bias. If we choose* $q = 2$ *then* $EJ_2^Q = 0$. $\square$

For Examples 2 and 3, the standard jackknife estimators, J_1^Q, i.e., (2.2) with $q = 1$, completely eliminate bias. In Example 4 the bias is not eliminated, but the rate at which the jackknife estimate shrinks to zero is improved. However, in Example 5, J_1^Q reverses the sign of the bias and increases its magnitude. Hence, there is a danger of worsening the bias if we do not choose q appropriately. (In Example 5, choosing $q = 2$ in (2.2) completely eliminates the bias.) In all four of these examples we know the form of $b(\hat{\theta}_n)$ *a priori*, but of course, this may not be the case in general. We discuss results in the next subsection, which relate the amount of bias reduction to the value of q we choose.

2.1.3 Delete-half estimator

Usually when solving a sampling approximation problem we use a moderate-to-large sample size, n. As a result, the optimal value may change little when we delete just one observation, e.g., $z^*_{100} \approx z^*_{99}$. It may be preferable to delete more than one observation at a time. Computationally, it may also be preferable to delete multiple observations, particularly when the effort to solve (1.3) grows faster than linearly in n, as is often the case. So, our generalized jackknife estimator uses a batching scheme with two disjoint batches of size $n/2$, where n is even. Let $\hat{\theta}^1_{n/2}$ be the original estimator using the first half of the (randomly ordered) full sample n and $\hat{\theta}^2_{n/2}$ be the original estimator using the second half. Define $\hat{\phi}_{n/2} = \frac{1}{2}(\hat{\theta}^1_{n/2} + \hat{\theta}^2_{n/2})$ and let $\bar{\phi}_{n/2}$ be the average of $\hat{\phi}_{n/2}$ over m i.i.d. replications. The generalized jackknife estimator using this batching scheme can now be written as

$$J_q^B = \frac{n^q \bar{\theta}_n - (n/2)^q \bar{\phi}_{n/2}}{n^q - (n/2)^q}. \tag{2.3}$$

There are $\frac{n!}{2(n/2!)^2}$ different such partitions of a sample of size n. However, we will choose only one such partition as indicated above. We return to this issue later.

2.1.4 Limiting behavior of generalized jackknife estimator

We seek to understand how the asymptotic bias of the delete-1 and delete-half estimators compare with that of the original estimator $\hat{\theta}_n$. To compare the bias of the generalized jackknife estimator (2.2) to that of the original estimator $\hat{\theta}_n$, we define

$$\rho^Q = \lim_{n\to\infty} \frac{EJ_q^Q - \theta}{E\hat{\theta}_n - \theta}.$$

Here, ρ^Q is a measure of how fast the bias in J_q^Q shrinks to zero compared to that of $\hat{\theta}_n$. For Examples 2-4, when we consider J_1^Q, we obtain $\rho^Q = 0$, but for Example 5 we obtain $\rho^Q = -1$. If we instead consider J_2^Q for Example 5 then we obtain $\rho^Q = 0$. We now state results from Gray and Schucany [22], which relate the choice of q to ρ^Q.

Theorem 3 (Theorem 3.4 [22])**.** *Assume the bias of the original estimator $\hat{\theta}_n$ is of the form $an^{-p} + o(n^{-p})$ and $a \neq 0$. Then,*

1. *if $p \leq q$ then $0 \leq \rho^Q < 1$, with $\rho^Q = 0$ when $q = p$;*
2. *if $p/2 \leq q < p$, then $-1 \leq \rho^Q < 0$, with $\rho^Q = -1$ when $q = p/2$; and,*
3. *if $q < p/2$ then $\rho^Q < -1$.*

Theorem 3 provides valuable guidance for choosing q in (2.2). Part 1 of the theorem says we should select $q = p$ so that bias of J_q^Q shrinks to zero at a faster rate than that of $\hat{\theta}_n$. Of course, doing so requires knowing p. With a choice of $q > p$, J_q^Q guarantees a reduction in the magnitude of asymptotic bias without changing its order or sign. Parts 2-3 of the theorem indicate the effect of choosing q too small, presumably because p is unknown. A choice of $q < p$ reverses the sign of the bias in J_q^Q. If we select $q < p/2$ we increase the magnitude of the asymptotic bias. This suggests erring on the side of selecting q too large rather than too small, particularly because we wish to preserve the lower bounding property of z_n^*, i.e., we want to avoid reversing the sign of the bias. That said, as q grows large for fixed n, J_q^Q approaches $\hat{\theta}_n$, and the bias reduction benefits of jackknifing are lost. Below we extend this analysis to the delete-half estimator, defining the asymptotic bias ratio of the delete-half estimator with respect to the original estimator as,

$$\rho^B = \lim_{n\to\infty} \frac{EJ_q^B - \theta}{E\hat{\theta}_n - \theta}.$$

Theorem 4. *Assume the bias of the original estimator $\hat{\theta}_n$ is of the form $an^{-p} + o(n^{-p})$ and $a \neq 0$. Then,*

1. *if $p \leq q$ then $0 \leq \rho^B < 1$, with $\rho^B = 0$ when $q = p$;*

2. *if $\log_2 \frac{2^p+1}{2} \leq q < p$, then $-1 \leq \rho^B < 0$, with $\rho^B = -1$ when $q = \log_2 \frac{2^p+1}{2}$; and,*

3. *if $q < \log_2 \frac{2^p+1}{2}$ then $\rho^B < -1$.*

Proof. From (2.3), $EJ_q^B = \theta + \frac{2^q-2^p}{2^q-1}an^{-p} + o(n^{-p})$. Hence, $\rho^B = \frac{2^q-2^p}{2^q-1}$, and the results in each of the three parts follow. □

Theorem 4 gives similar guidance as that of Theorem 3, when choosing q in the delete-half estimator. Specifically, if q is above a particular threshold given by the first part of the theorem, then bias reduction is ensured (but the reduction becomes weaker as q grows). And, choosing q too small can reverse the sign of the bias, and, when it is sufficiently small, even increase its magnitude.

Note that $\log_2 \frac{2^p+1}{2} > \frac{p}{2}\ \forall p > 0$. Comparing the thresholds in Theorems 3 and 4 we see that this indicates the range of q for which the bias increases (with sign reversal) in the delete-half estimator is larger than that of the delete-1 estimator. This suggests that when choosing q in the delete-half estimator there is even more reason to exercise caution, in the sense of avoiding selection of a q that is too small.

2.1.5 Application to stochastic programming and potential issues

We now describe a multiple replication procedure to form a confidence interval on the optimality gap of a stochastic program using the jackknife estimator by the method of Quenouille. As indicated above, we should choose q close to p, but hedging to a larger value of q. Unfortunately, we are unlikely to know p *a priori*. So, a procedure by which we can estimate p would be valuable, and we pursue this in the following section. For now, however, we assume that an appropriate value for q has been selected. Below we state a multiple replication procedure for forming a confidence interval on the optimality gap using the delete-half estimator, (2.3), taking the value of q as input. A multiple replication procedure for delete-1 estimator can be produced along the same lines.

Multiple Replication Procedure with Delete-half Estimator **(MRP^q)**

Input: sample size n (even), replications m, confidence level $(1-\alpha)$, jackknife parameter q, candidate solution $\hat{x} \in X$

Output: approximate $(1-\alpha)$-level confidence interval on the optimality gap, $\mu_{\hat{x}}$

1. For $i = 1, \ldots, m$

   - Generate a random sample $\tilde{\xi}^1, \ldots, \tilde{\xi}^n$ which satisfies the unbiased condition (1.5), and also satisfies this condition when any of the observations are deleted

- Let $z_n^{i*} = \min_{x \in X} \hat{f}(x, \tilde{\xi}^1, \ldots, \tilde{\xi}^n)$ and $\hat{\theta}_n^i = \hat{f}(\hat{x}, \tilde{\xi}^1, \ldots, \tilde{\xi}^n) - z_n^{i*}$
- Let $z_{n/2}^{i1*} = \min_{x \in X} \hat{f}(x, \tilde{\xi}^1, \ldots, \tilde{\xi}^{n/2})$ and $\hat{\theta}_{n/2}^{i1} = \hat{f}(\hat{x}, \tilde{\xi}^1, \ldots, \tilde{\xi}^{n/2}) - z_n^{i1*}$
- Let $z_{n/2}^{i2*} = \min_{x \in X} \hat{f}(x, \tilde{\xi}^{n/2+1}, \ldots, \tilde{\xi}^n)$ and $\hat{\theta}_{n/2}^{i2} = \hat{f}(\hat{x}, \tilde{\xi}^{n/2+1}, \ldots, \tilde{\xi}^n) - z_{n/2}^{i2*}$
- Let $\hat{\phi}_{n/2}^i = \frac{\hat{\theta}_{n/2}^{i1} + \hat{\theta}_{n/2}^{i2}}{2}$

2. Let $\bar{\theta}_n = \frac{1}{m} \hat{\theta}_n^i$ and $\bar{\phi}_{n/2} = \frac{1}{m} \hat{\phi}_{n/2}^i$

3. Let

$$J_q^B = \frac{n^q \bar{\theta}_n - (n/2)^q \bar{\phi}_{n/2}}{n^q - (n/2)^q} \quad \text{and} \quad s_m^2 = \frac{1}{m-1} \sum_{i=1}^{m} \left(\frac{n^q \hat{\theta}_n^i - (n/2)^q \hat{\phi}_{n/2}^i}{n^q - (n/2)^q} - J_q^B \right)^2$$

4. Form CI for $\mu_{\hat{x}}$ as,

$$\left[0, J_q^B + \frac{t_{m-1,\alpha} s_m}{\sqrt{m}} \right]$$

The validity of the confidence interval produced in step 3 of the procedure follows from the standard central limit theorem because the $\tilde{\xi}^1, \ldots, \tilde{\xi}^n$ produced in each of the m replications in step 1 are independent. Hence, the observations $\hat{\theta}_n^i$ (and $\hat{\phi}_{n/2}^i$), $i = 1, \ldots, m$, are i.i.d.

2.2 A p-estimation adaptive procedure

2.2.1 Motivation

The generalized jackknife estimator (2.3) can work well in situations where we have some prior knowledge regarding the nature of the bias, e.g., that $b(\hat{\theta}_n) = O(n^{-p})$. As described in the previous section, if we know p, and choose q appropriately, we can remove the leading term in the bias. However, the J_q^B (and J_q^Q) estimators are less effective when we do not know the order of the bias *a priori*. We now describe an adaptive jackknife estimator which does not require knowing the order of the bias. Instead, both the order of the bias and the parameter of interest are found by the method. To motivate our procedure, assume $\hat{\theta}_n$ has bias of the form an^{-p}, and consider the following system of equations,

$$\begin{aligned} E\hat{\theta}_{n-1} &= \theta + a(n-1)^{-p} && \text{(2.4a)} \\ E\hat{\theta}_{n} &= \theta + an^{-p}. && \text{(2.4b)} \end{aligned}$$

Viewing θ and a as unknowns we can solve this system of linear equations to obtain

$$\theta = \frac{n^p E\hat{\theta}_n - (n-1)^p E\hat{\theta}_{n-1}}{n^p - (n-1)^p}. \tag{2.5}$$

We may view the derivation of θ in (2.5) from the equations in (2.4) as motivating the jackknife estimator (2.2) under the assumption that bias is $O(n^{-p})$ and p is known. An analogous derivation with $\hat{\theta}_n$ and $\hat{\theta}_{n/2}$ leads to (2.3). However, since we do not know p, we can write a similar set of three equations to be solved for three unknowns, i.e., θ, p and a. The result will motivate the definition of our adaptive jackknife estimator.

Again assume the bias in the original estimator $\hat{\theta}_n$ is of the form an^{-p}. Assume that n is a multiple of 4 and let $\hat{\theta}_{n/4}$, $\hat{\theta}_{n/2}$ and $\hat{\theta}_n$ be the original estimators based on the respective sample sizes of $n/4, n/2$ and n. Then we can write the following set of equations

$$\begin{aligned} E\hat{\theta}_{n/4} &= \theta + a(n/4)^{-p} && \text{(2.6a)} \\ E\hat{\theta}_{n/2} &= \theta + a(n/2)^{-p} && \text{(2.6b)} \\ E\hat{\theta}_{n} &= \theta + an^{-p}. && \text{(2.6c)} \end{aligned}$$

Solving the system of nonlinear equations in (2.6) yields

$$\theta = \frac{E\hat{\theta}_n E\hat{\theta}_{n/4} - \left(E\hat{\theta}_{n/2}\right)^2}{E\hat{\theta}_n + E\hat{\theta}_{n/4} - 2E\hat{\theta}_{n/2}} \tag{2.7a}$$

$$p = \log_2\left(\frac{E\hat{\theta}_{n/4} - E\hat{\theta}_{n/2}}{E\hat{\theta}_{n/2} - E\hat{\theta}_n}\right). \tag{2.7b}$$

The expressions in (2.7) are nonlinear functions of population means. To obtain an estimator motivated by the above development, we assume that m i.i.d. replications are performed.

2.2.2 p-estimation procedure

Our estimate of p replaces the expectations on the right-hand sides of (2.7) with sample means. We do this in the following manner: Given a sample $\tilde{\xi}^1, \ldots, \tilde{\xi}^n$ (n is a multiple of 4), we partition the sample into two subsamples of size $n/2$ and, in turn, partition those subsamples into four subsamples of size $n/4$. We then form a single observation of $\hat{\theta}_n$ based on the full sample, $\hat{\phi}_{n/2}$ as the average of two i.i.d. estimates of form $\hat{\theta}_{n/2}$, and $\hat{\phi}_{n/4}$ as the average of four i.i.d. estimates of form $\hat{\theta}_{n/4}$. Averaging over m i.i.d. replications we form estimators we denote $\bar{\theta}_n$, $\bar{\phi}_{n/2}$, and $\bar{\phi}_{n/4}$. There are some implementations issues for doing so, which we return to in Chapter 3. These sample-mean estimators replace their population counterparts in (2.7) to yield

$$\hat{\theta}^A = \frac{\bar{\theta}_n \bar{\phi}_{n/4} - \left(\bar{\phi}_{n/2}\right)^2}{\left(\bar{\phi}_{n/4} - \bar{\phi}_{n/2}\right) - \left(\bar{\phi}_{n/2} - \bar{\theta}_n\right)} \tag{2.8a}$$

$$p^A = \log_2\left(\frac{\bar{\phi}_{n/4} - \bar{\phi}_{n/2}}{\bar{\phi}_{n/2} - \bar{\theta}_n}\right), \tag{2.8b}$$

which are nonlinear functions, $p^A = f(\Theta_n)$, $\hat{\theta}^A = g(\Theta_n)$, of the three sample means $\Theta_n = (\bar{\phi}_{n/4}, \bar{\phi}_{n/2}, \bar{\theta}_n)$. Let $\hat{\Sigma}$ denote the standard sample covariance estimator of Θ_n.

Assume that $E(\hat{\theta}_n) \neq E\hat{\theta}_{n/2} \neq E\hat{\theta}_{n/4}$. We require this assumption to apply the delta theroem (see, e.g., Casella and Berger [10]) in order to estimate the variance s_p^2 of p^A. This assumption is satisfied, for example, when the bias of $\hat{\theta}_n$ is a strictly decreasing function of n. We return to this issue in Section 3.4. Under this assumption, using a first-order Taylor series expansion, we can estimate the variance of p^A as

$$s_p^2 = \nabla^T f(\Theta_n)\, \hat{\Sigma}\, \nabla f(\Theta_n).$$

We could form an adaptive jackknife procedure by simply using $q = p^A$ in MRP^q. We do not do so in an attempt to preserve the upper bound property of the estimators, i.e., we seek a conservative procedure. We instead let

$$q = \max\{p^A, \frac{1}{2}\} + t_{m-1,1-\beta} s_p, \tag{2.9}$$

where $0 \leq \beta \leq 1$. We know from Theorem 4 that our bias reduction is less aggressive for larger values of q and so as β goes to zero, and the corresponding t quantile grows, our procedure becomes more conservative. With β at our disposal we examine the performance of a family of adaptive jackknife estimators in the next section. Of course, the freedom to choose a parameter such as β can be disconcerting to some. In this case, we recommend choosing $\beta = \frac{1}{2} - \frac{\alpha}{2}$. This choice deflates the value of β relative to α, takes the correct value as α approaches one and only allows choosing $q \geq p^A$. Under relatively mild conditions (see, e.g., [52]), we know $b(z_n^*) = O(n^{-p})$ for $p \geq 1/2$, and hence we include the max operator in (2.9). We also note that s_p already includes the "$m^{-1/2}$" factor since $\hat{\Sigma}$ is the sample covariance of a vector of sample means.

2.2.3 Numerical comparison

We now compare the performance of the naive estimator from MRP^o, the standard jackknife estimatorJ_1^B and the p-estimation adaptive estimator J_q^B with q chosen according to (2.9). We do this comparison on an asset allocation problem on exchange traded funds from [37]. This single-period model has 14 assets representing Exchange-Traded Funds, i.e., funds designed to track indices such as the S&P 500, Russell 3000, and indices from industrial sectors like biotechnology and banking. We maximize expected utility using a so-called power utility function augmented by a penalty term that (mildly) discourages deviations from the investor's current portfolio. We assume the return distribution is multivariate normal and estimate the return's mean and covariance based on five years of monthly data from 1999 to 2004. Under the normal-distribution assumptions we can solve the problem exactly, and this allows us to assess the performance of our procedures, e.g., compute empirical coverage probabilities and values of the bias. We refer the reader to [37] for further details.

Figure 2.1 shows the confidence interval on $\mu_{\hat{x}}$ generated by MRP^o using $m = 40$, $\alpha = 0.95$, and varying n from 25 to 50 to 100. We obtained $\hat{x}$ by solving an instance of (1.3) with $n = 400$ i.i.d. samples, and we obtained the existing investor's portfolio, x^t, by solving a separate instance with $n = 400$ i.i.d. samples. The figure is based on averaging the output of MRP^o over $N = 2000$ runs. The CI width is partitioned into the three factors discussed in Section 1.3, namely the optimality gap, the sampling error and the bias. Here, the bias estimate is formed by subtracting the known optimality gap from the average of the $N = 2000$ point estimates $\bar{G}_m(\hat{x})$. We note that $z^* = 1.0015$. So, the $\hat{x}$ we are using is suboptimal by about 0.02%, and with $n = 100$ we are forming a 0.95-level CI on that optimality gap whose width is roughly 0.1% of z^*. We can clearly see from the figure that bias dominates the CI width. This motivates use of the bias reduction techniques we have proposed.

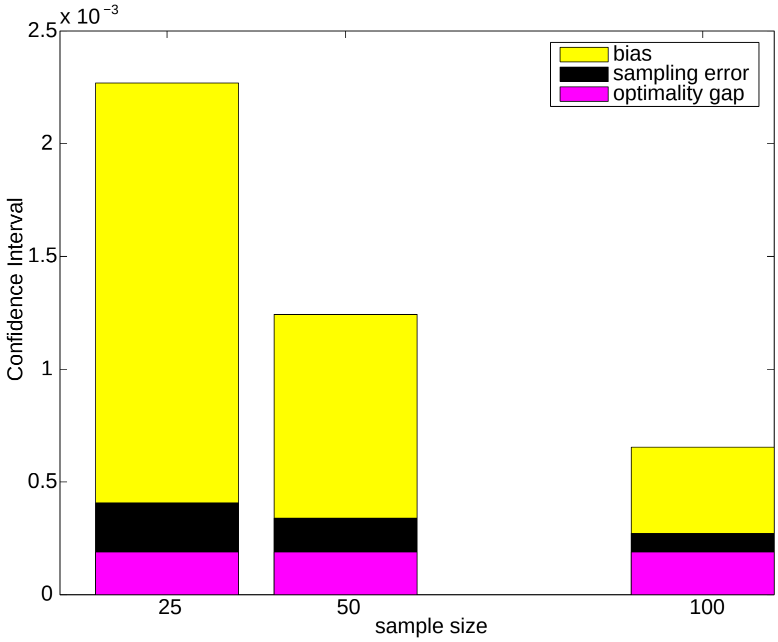

Figure 2.1: CI width versus sample size for MRP^o.

We assess the performance of three optimality-gap point and interval estimators, denoted as follows: (i) D_o, the point estimate and interval estimate of the MRP^o in Section 1.2.3 in which we do not attempt to reduce bias; (ii) J_1^B, the standard jackknife estimator with MRP^q of Section 2.1.5 with $q = 1$; and, (iii) D_β, the adaptive p-estimation estimator in which we choose q via (2.9) and use this in MRP^q. For D_β we consider β ranging from 0.3 down to 0.01. Throughout we use $n = 100$ and $m = 40$.

We begin by forming an empirical estimate of the mean-square error (MSE) of each estimator. We did so using $N = 2000$ i.i.d. runs of each procedure (i)-(iii) above. The estimated MSE of D_o and J_1^B were 4.5×10^{-7}, and 3.5×10^{-8}, respectively. Figure 2.2(a) shows the MSE of D_β for various values of β, and also includes those of D_o and J_1^B for reference. Figures 2.2(b)-(d) show the negative and positive part of MSE and the probability the gap point estimate is below $\mu_{\hat{x}}$. Because of the nature of our point and interval estimators, we prefer estimators

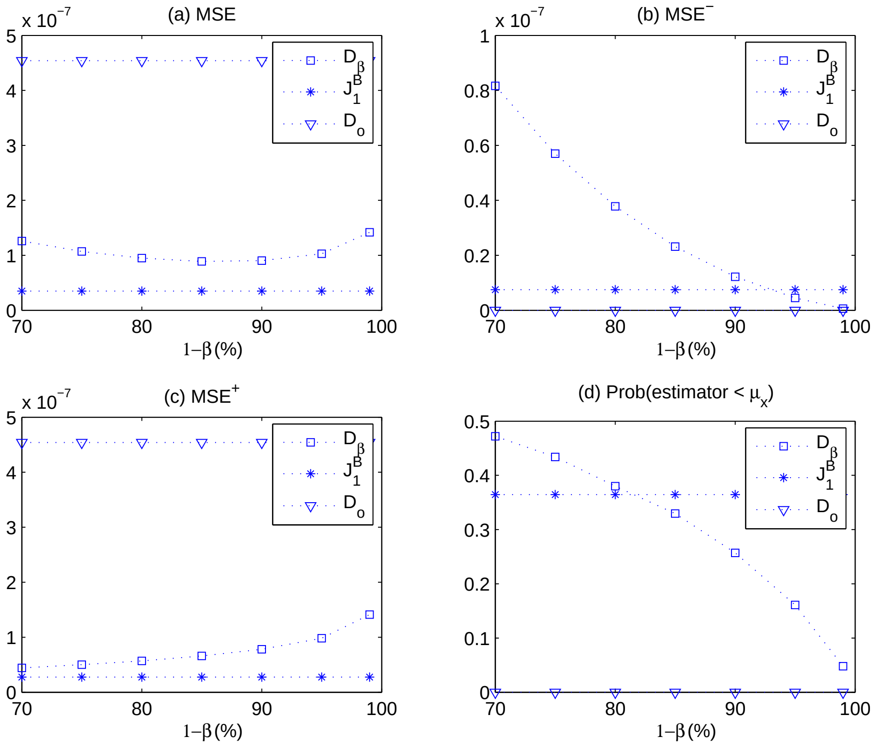

Figure 2.2: Empirical mean-square error (MSE), its negative and positive parts, and the probability the point estimate is below $\mu_{\hat{x}}$.

in which MSE^{-} and this probability are small. Restated, in choosing between two estimators, we may prefer an estimator with slightly larger MSE if these other two measures are smaller. The first observation is that all our jackknife estimators significantly reduce MSE. The standard jackknife estimator performs very well with respect to MSE. This is not surprising considering the estimates of p we obtained via p^A over the $N = 2000$ replications were 0.80 with a standard error of 0.30. So, $q = 1$ is arguably a reasonably conservative choice. That said, Figures 2.2(b)-(d) suggest that as β goes down the MSE^{-}, and probability of having an invalid upper bound point estimate, improve significantly while the relative increase in MSE^{+} is modest.

Figure 2.3 shows the empirical coverage function of the interval estimators

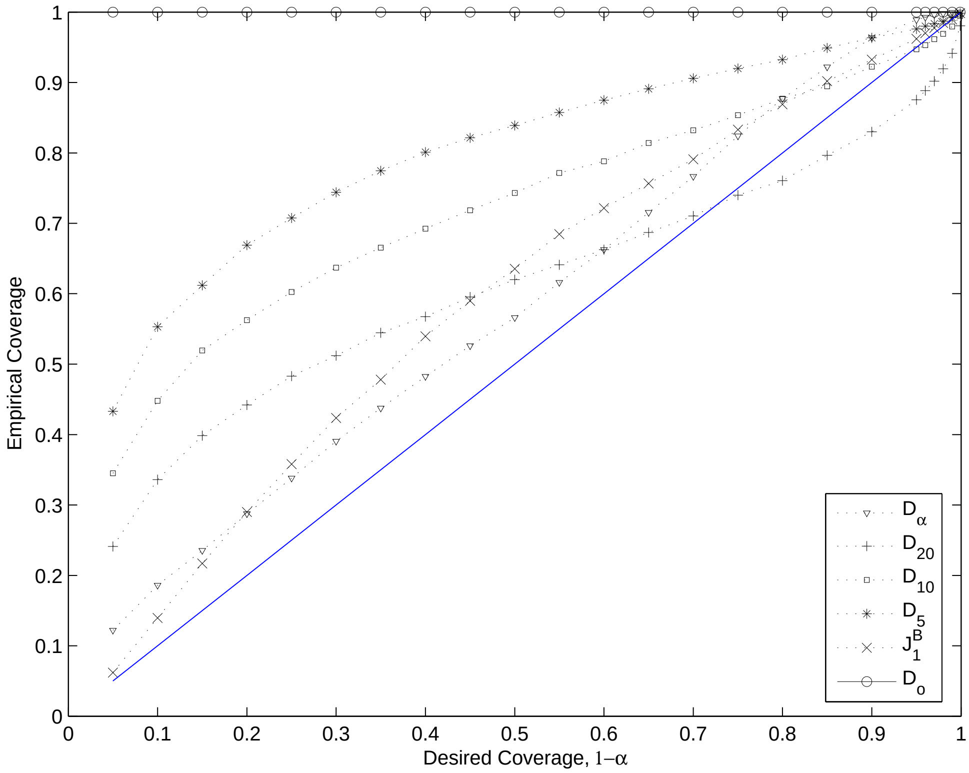

Figure 2.3: Coverage function plots are shown for the interval estimator without bias correction, the standard jackknife, the adaptive p-estimation estimators for $\beta = 0.20, 0.10$ and 0.05, and D_α, which adjusts the value of β according to $\beta = \frac{1}{2} - \frac{\alpha}{2}$ in the p-estimation estimator.

produced by our procedures, i.e., the Schruben coverage plots [48]. The original procedure, i.e., without bias reduction, produces an interval estimator with 100% coverage regardless of the value of α. (Of course, as α shrinks to zero this no longer holds but the smallest α in the plot is 0.05.) Using $\beta = 0.20$ yields an adaptive p-estimation estimator that has undercoverage for large values of α. The interval estimator of the standard jackknife and those associated with smaller values of $\beta = 0.10, 0.05$ and D_α (which is based on the $\beta = \frac{1}{2} - \frac{\alpha}{2}$ formula discussed earlier) all appear to perform well with respect to coverage.

2.3 Summary

The first half of this chapter is dedicated to application of generalized jackknife estimators in stochastic programming. We argued in Chapter 1 that the order of the bias is not known *a priori* and can be anything from 1/2 to ∞. During the second half of this chapter we present a new adaptive p-estimation jackknife estimator. The p-estimation scheme starts by estimating the order of bias and then uses the generalized jackknife procedure with the estimated order of bias as input. We compared the performance of the naive estimator of MRP^o, J_1^B in MRP^q and the p-estimation stage estimator on a static asset allocation model.

Our family of adaptive p-estimation jackknife estimators is parameterized by β. The simple asset allocation model has normally-distributed returns so that we could solve it exactly and compute the true optimality gap to better assess the performance of our estimators. In our simplest procedure, we do not attempt to reduce bias, and in this case the bias dominates the width of our confidence intervals. All of the jackknife estimators we consider significantly decrease mean-square error by reducing bias. When one seeks a conservative point estimate for use in a one-sided confidence interval our adaptive p-estimation jackknife with β = 0.05-0.10 provides significant improvement over neglecting bias entirely and may provide an attractive alternative to the standard jackknife.

Chapter 3

Adaptive Jackknife Estimator

The expressions for the p-estimation adaptive estimators in the previous chapter are nonlinear functions of the underlying estimators. The denominator in (2.8a) involves the difference of two small numbers. As a result these estimators can be poorly behaved, and even basic results like consistency can fail to hold. Still, the above derivation is instructive and motivates the scheme we pursue below.

3.1 An adaptive jackknife estimator

We seek an adaptive jackknife estimator that does not require *a priori* knowledge of p. We begin by revisiting (2.6a) and (2.6b) and solving to obtain

$$an^{-p} = \frac{(E\hat{\theta}_{n/2} - \theta)^2}{E\hat{\theta}_{n/4} - \theta}. \tag{3.1}$$

Of course, θ is unknown, but we can view our best estimate as being $\theta \approx E\hat{\theta}_n$. Substituting this into (3.1), we obtain

$$an^{-p} \approx \frac{(E\hat{\theta}_{n/2} - E\hat{\theta}_n)^2}{E\hat{\theta}_{n/4} - E\hat{\theta}_n}. \tag{3.2}$$

By (2.6c) the expression in (3.2) is an estimate of $\hat{\theta}_n$'s bias, i.e.,

$$\theta \approx E\hat{\theta}_n - r(E\hat{\theta}_{n/2} - E\hat{\theta}_n), \tag{3.3}$$

with

$$r = \frac{E\hat{\theta}_{n/2} - E\hat{\theta}_n}{E\hat{\theta}_{n/4} - E\hat{\theta}_n} = \frac{1}{1+2^p}, \tag{3.4}$$

where the final equality again assumes (2.6) holds. Note that since $p \geq 0$, $r \leq 1/2$. We now describe an adaptive jackknife estimator motivated by the above development. Above we assume that (2.6) holds but we will relax this in what follows. The following procedure is developed for a general underlying estimator $\hat{\theta}_n$

whose bias is of the form $O(n^{-p})$.

Adaptive jackknife procedure (AJP)

Input: sample size n which is a multiple of 4, replications m

Output: An adaptive estimator J^A of θ

1. For $i = 1, \ldots, m$
   - Generate a sample of size n indexed by N
   - Let N^j, $j = 1, \ldots, 4$, partition N, with $|N^j| = n/4$, $j = 1, \ldots, 4$
   - Let $\hat{\theta}_n^i$ be the underlying estimator based on the full sample N
   - Let $\hat{\theta}_{n/2}^{i1}, \hat{\theta}_{n/2}^{i2}$ be the underlying estimators using $N^1 \bigcup N^2$ and $N^3 \bigcup N^4$, respectively
   - Let $\hat{\theta}_{n/4}^{ij}$, $j = 1, \ldots, 4$, denote the estimators based on the respective N^j, $j = 1, \ldots, 4$
   - Let $\hat{\phi}_{n/4}^i = \dfrac{1}{4} \sum_{j=1}^{4} \hat{\theta}_{n/4}^{ij}$ and $\hat{\phi}_{n/2}^i = \dfrac{1}{2} \sum_{j=1}^{2} \hat{\theta}_{n/2}^{ij}$
2. Let $\bar{\phi}_{n/4} = \dfrac{1}{m} \sum_{i=1}^{m} \hat{\phi}_{n/4}^i$, $\bar{\phi}_{n/2} = \dfrac{1}{m} \sum_{i=1}^{m} \hat{\phi}_{n/2}^i$, and $\bar{\theta}_n = \dfrac{1}{m} \sum_{i=1}^{m} \hat{\theta}_n^i$
3. Define adaptive jackknife estimators,

$$J^A = \bar{\theta}_n - \hat{r}(\bar{\phi}_{n/2} - \bar{\theta}_n) \tag{3.5a}$$

$$\hat{r} = \frac{\bar{\phi}_{n/2} - \bar{\theta}_n}{\bar{\phi}_{n/4} - \bar{\theta}_n}. \tag{3.5b}$$

Note that the adaptive jackknife estimator in AJP was obtained by replacing $E\hat{\theta}_n$ with $\bar{\theta}_n$, $E\hat{\theta}_{n/2}$ with $\bar{\phi}_{n/2}$ and $E\hat{\theta}_{n/4}$ with $\bar{\phi}_{n/4}$ in equations (3.3) and (3.4). Figure 3.1 depicts the adaptive jackknife estimator and its common random number scheme. We use the convention that if $\bar{\theta}_n = \bar{\phi}_{n/2} = \bar{\phi}_{n/4}$, then $J^A = \bar{\theta}_n$. We justify this shortly, at least under certain conditions, by an inequality we will establish in

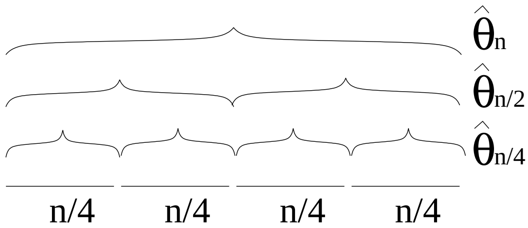

Figure 3.1: Common random number scheme.

Theorem 5 below. We call the jackknife estimator J^A adaptive because it does not require *a priori* knowledge of p. It only assumes bias is of the form $O(n^{-p})$ for some p.

3.2 Properties of the adaptive jackknife estimator

This section characterizes the adaptive jackknife estimator. We begin with an assumption that holds for the optimality gap estimator, i.e., when $\hat{\theta}_n = G_n(\hat{x})$, the focus of this research.

(A4). *Let $N = \{1, \ldots, n\}$ index a sample of size n. Let $N' \subset N$ and let $\bar{N}'$ and $N' \setminus \bar{N}'$ partition N' with $n' = |N'| \geq 2$, $\bar{n}' = |\bar{N}'|$, and $1 \leq \bar{n}' \leq n'$. Let $\hat{\theta}_{n'}$, $\hat{\theta}_{\bar{n}'}$ and $\hat{\theta}_{n'-\bar{n}'}$ be estimators defined on samples indexed by N', $\bar{N}'$ and $N' \setminus \bar{N}'$, respectively. Then*

$$\hat{\theta}_{n'} \leq \frac{1}{n'} \left(\bar{n}' \hat{\theta}_{\bar{n}'} + (n' - \bar{n}') \hat{\theta}_{n'-\bar{n}'} \right). \tag{3.6}$$

When $\hat{\theta}_n$ is a nonlinear function of a sample mean, (A4) holds provided the nonlinear function is convex. If that function is linear then (A4) holds with equality. In the context of stochastic programming, if we optimize sample means, i.e., $\hat{f}(x, \tilde{\xi}^1, \ldots, \tilde{\xi}^n) = \frac{1}{|N|} \sum_{i \in N} f(x, \tilde{\xi}^i) \equiv \hat{f}_N(x)$ and $z_n^* = \min_{x \in X} \hat{f}_N(x)$, then (A4) holds for the gap estimator $G_n(\hat{x}) = \hat{f}_N(\hat{x}) - z_n^*$. (A4) does not hold for z_n^* itself, rather it holds for $-z_n^*$. And, it holds for the optimal value of a maximization problem, i.e., for $\max_{x \in X} \hat{f}_N(x)$.

Theorem 5. *Let $\hat{\phi}_{n/4}$ and $\hat{\phi}_{n/2}$ be as defined in AJP. If (A4) holds, then $\bar{\phi}_{n/4} \geq \bar{\phi}_{n/2} \geq \bar{\theta}_n$, w.p.1, which further implies $0 \leq \hat{r} \leq 1$.*

Proof. $\bar{\phi}_{n/4}$ and $\bar{\phi}_{n/2}$ and $\bar{\theta}_n$ are defined as sample means of m i.i.d. replicates of $\hat{\phi}_{n/4}$ and $\hat{\phi}_{n/2}$ and $\hat{\theta}_n$, respectively, and it suffices to show $\hat{\phi}_{n/4} \geq \hat{\phi}_{n/2} \geq \hat{\theta}_n$ w.p.1. (A4) implies that $\hat{\theta}^1_{n/2} \leq \frac{2}{n}\left(\frac{n}{4}\hat{\theta}^1_{n/4} + \frac{n}{4}\hat{\theta}^2_{n/4}\right)$, $\hat{\theta}^2_{n/2} \leq \frac{2}{n}\left(\frac{n}{4}\hat{\theta}^3_{n/4} + \frac{n}{4}\hat{\theta}^4_{n/4}\right)$ and $\hat{\theta}_n \leq \frac{1}{n}\left(\frac{n}{2}\hat{\theta}^1_{n/2} + \frac{n}{2}\hat{\theta}^2_{n/2}\right)$. Combining these inequalities completes the proof. □

The next theorem provides consistency of the adaptive estimator, J^A.

Theorem 6. *Assume that the original estimator $\hat{\theta}_n$ is strongly consistent. Let J^A and $\hat{r}$ be as defined in AJP. If (A4) holds then $\lim\limits_{n\to\infty} J^A = \theta$, w.p.1.*

Proof. Consistency of $\hat{\theta}_n$ implies $\lim\limits_{n\to\infty} \bar{\phi}_{n/4} = \lim\limits_{n\to\infty} \bar{\phi}_{n/2} = \lim\limits_{n\to\infty} \bar{\theta}_n = \theta$, w.p1. From the proof of Theorem 5, we have $\bar{\phi}_{n/2} - \bar{\theta}_n \geq 0$, w.p.1, and $0 \leq \hat{r} \leq 1$, w.p.1. Combining the facts we have that $0 \leq \lim\limits_{n\to\infty} \hat{r}(\bar{\phi}_{n/2} - \bar{\theta}_n) \leq \lim\limits_{n\to\infty} (\bar{\phi}_{n/2} - \bar{\theta}_n) = 0$, w.p.1, implies that $\lim\limits_{n\to\infty} J^A = \theta$, w.p.1. □

We now compare the bias in the adaptive estimator, J^A, to that of the naive estimator, $\hat{\theta}_n$, and delete-half estimator, J^B_1.

Theorem 7. *Let J^A and $\hat{r}$ be as defined in AJP and J^B_q be as defined in (2.3). If (A4) holds then for all $q \leq 1$, $J^B_q \leq J^A \leq \bar{\theta}_n$ w.p.1 and $b(J^B_q) \leq b(J^A) \leq b(\hat{\theta}_n)$.*

Proof. From Theorem 5, we have $\bar{\phi}_{n/2} - \bar{\theta}_n \geq 0$, w.p.1, and $0 \leq \hat{r} \leq 1$, w.p.1. Combining these facts implies $\bar{\theta}_n - (\bar{\phi}_{n/2} - \bar{\theta}_n) \leq J^A \leq \bar{\theta}_n$ w.p.1. By definition $J^B_q = \frac{n^q\bar{\theta}_n - (n/2)^q\bar{\phi}_{n/2}}{n^q - (n/2)^q} = \bar{\theta}_n - \frac{(\bar{\phi}_{n/2} - \bar{\theta}_n)}{2^q - 1}$. This implies $J^B_q \leq J^A \leq \bar{\theta}_n$ w.p.1 if $q \leq 1$. Taking expectation and subtracting θ from the last inequality we have $b(J^B_q) \leq b(J^A) \leq b(\bar{\theta}_n) = b(\hat{\theta}_n)$. □

As indicated above, the choice of $q = 1$ is an often natural and hence popular choice for the generalized jackknife, but sometimes this choice is too aggressive and can reverse the sign of the bias and even increase its magnitude. Theorem 7 shows that the adaptive jackknife estimator is less aggressive than J_q^B for $q \leq 1$. This suggests that J^A may be appropriate in some settings when the standard jackknife is too aggressive.

The estimator J^A is a nonlinear function of sample means, and hence EJ^A is, in general, not equal to the right-hand side of (3.3), i.e., the expression of J^A with the estimators replaced by their population means. However, as the number of replications m grows large this equality does hold and it is instructive to analyze this approximation of EJ^A, i.e.,

$$EJ^A \approx E\hat{\theta}_n - \frac{E\hat{\theta}_{n/2} - E\hat{\theta}_n}{E\hat{\theta}_{n/4} - E\hat{\theta}_n}(E\hat{\theta}_{n/2} - E\hat{\theta}_n).$$

Now, assuming $E\hat{\theta}_n = \theta + an^{-p} + o(n^{-p})$ holds, we obtain

$$EJ^A \approx \theta + an^{-p}\frac{2}{2^p + 1} + o(n^{-p}). \tag{3.7}$$

This suggests that J^A will effectively reduce the coefficient a in the bias term and its ability to do so depends on p. Furthermore when m and n are large enough so that the above approximations are reasonable, and $b(\hat{\theta}_n) = O(n^{-p})$, this analysis suggests that J^A will not reverse the sign of the bias regardless of p's value.

3.3 Family of adaptive estimators

In this section we extend J^A by introducing a family of adaptive estimators which allow us to select the level of aggression when attempting to reduce bias. With γ

being a positive integer, we define the family of adaptive estimators as

$$J_\gamma^{A1} = \bar{\theta}_n - (\sum_{k=1}^{\gamma} \hat{r}^k)(\bar{\phi}_{n/2} - \bar{\theta}_n) \tag{3.8a}$$

$$J_\infty^{A1} = \begin{cases} \bar{\theta}_n - \frac{\hat{r}}{1-\hat{r}}(\bar{\phi}_{n/2} - \bar{\theta}_n) & \text{if } \hat{r} < 1 \\ \bar{\theta}_n & \text{if } \hat{r} = 1. \end{cases} \tag{3.8b}$$

Theorem 8. *Assume that the original estimator $\hat{\theta}_n$ is strongly consistent. Let γ be a positive integer and let J_γ^{A1} and $\hat{r}$ be as defined in AJP, except that (3.8a) replaces (3.5a). If (A4) holds then $\lim_{n\to\infty} J_\gamma^{A1} = \theta$, w.p.1.*

The proof of Theorem 8 is similar to that of Theorem 6 and hence is omitted. Also note that $J_{\gamma_1}^{A1} \le J_{\gamma_2}^{A1}$, w.p.1 for any positive integers $\gamma_1 \ge \gamma_2$, i.e., as γ grows we are more aggressive in reducing bias.

We can repeat the type of analysis we carried out at the end of Section 3.2. As the number of replications grow, we can approximate the expected value of J_γ^{A1} as

$$EJ_\gamma^{A1} \approx E\hat{\theta}_n - \frac{E\hat{\theta}_{n/2} - E\hat{\theta}_n}{E\hat{\theta}_{n/4} - E\hat{\theta}_{n/2}} \left(1 - \left(\frac{E\hat{\theta}_{n/2} - E\hat{\theta}_n}{E\hat{\theta}_{n/4} - E\hat{\theta}_n}\right)^\gamma\right) (E\hat{\theta}_{n/2} - E\hat{\theta}_n)$$

Now assuming $E\hat{\theta}_n = \theta + an^{-p} + o(n^{-p})$ holds, we obtain

$$EJ_\gamma^{A1} \approx \theta + an^{-p}\left(1 - \left(1 - \frac{1}{2^p}\right)\left(1 - \frac{1}{(2^p+1)^\gamma}\right)\right) + o(n^{-p}). \tag{3.9}$$

Of course, when $\gamma = 1$ in (3.9) we recover (3.7) from the end of Section 3.2. The last member of this family, i.e., J_∞^A, also follows (3.9) with $\gamma = \infty$ assuming $b(\hat{\theta}_n)$ is a strict decreasing function of n. Note that (3.9) again suggests that as γ grows bias reduction is more aggressive. However, it also suggests that the entire family inherits the same properties we described for the $\gamma = 1$ case at the end of Section 3.2. Namely, J_γ^A effectively reduces the coefficient a in the bias term but does not change its sign. Having a family of these estimators gives us flexibility in choosing the aggressiveness with which we seek to reduce the bias. Our selection of γ can also depend on other parameters. The sample size n needs to be large enough so that $b(\hat{\theta}_n) = O(n^{-p})$ holds for n, $n/2$ and $n/4$. Given this expression (3.9) holds provided the number of replications is sufficiently large, we have found that when m is larger it improves the performance of more aggressive members of the J_γ^A family.

3.4 Interval estimator

Our development so far has concentrated on an adaptive jackknife point estimate. We now discuss confidence interval construction. The estimator J_γ^{A1} in (3.8a) is a nonlinear function of a vector-valued sample mean and so we can repeat the type of analysis we performed in Section 2.2.2. Specifically, $\bar{\theta} = (\bar{\phi}_{n/4}, \bar{\phi}_{n/2}, \bar{\theta}_n)$ is a three-vector whose components are sample means formed from m i.i.d. observations of $\hat{\theta} = (\hat{\phi}_{n/4}, \hat{\phi}_{n/2}, \hat{\theta}_n)$, whose components are defined in AJP. Assuming finite second moments of $\hat{\theta}_n$, we have that $\bar{\theta}$ satisfies the following multivariate central limit theorem,

$$\sqrt{m}\left(\bar{\theta} - E\hat{\theta}\right) \Rightarrow N(0, C),$$

where C is $\hat{\theta}$'s covariance matrix. Now, $J_\gamma^{A1} = g_\gamma(\bar{\theta})$, where

$$g_\gamma(\theta_1, \theta_2, \theta_3) = \theta_3 - \sum_{k=1}^{\gamma} \left(\frac{\theta_2 - \theta_3}{\theta_1 - \theta_3}\right)^k (\theta_2 - \theta_3). \tag{3.10}$$

Here, $g_\gamma : H \to \Re$, where we can restrict the domain of g_γ to $H = \{(\theta_1, \theta_2, \theta_3) : \theta_1 \geq \theta_2 \geq \theta_3\}$ by (A4). Assume that $E(\hat{\theta}_n) \neq E\hat{\theta}_{n/2} \neq E\hat{\theta}_{n/4}$. Note that g_γ is twice continuously differentiable in a sufficiently small neighborhood of $E\hat{\theta}$ under this assumption. Hence, we can apply the delta theorem (see, e.g., Casella and

Berger [10]) to conclude

$$\sqrt{m}\left(J_\gamma^{A1} - g_\gamma(E\hat{\theta})\right) \Rightarrow N(0,\beta^2), \tag{3.11}$$

provided $\beta^2 = \nabla^T g_\gamma(E\hat{\theta}) C \nabla g_\gamma(E\hat{\theta}) > 0$. Thus the variance of the adaptive jackknife estimator can be estimated by $\hat{\beta}^2 = \nabla^T g_\gamma(\bar{\theta}) \hat{C} \nabla g_\gamma(\bar{\theta})$, where $\hat{C}$ is the sample covariance matrix from the replications of $\hat{\theta}$. Using the same type of Taylor series expansion that proves the delta theorem, we can correct for the bias introduced in J_γ^{A1} due to it being a nonlinear function of sample means. Specifically, we can replace J_γ^{A1} in (3.11) with

$$g_\gamma(\bar{\theta}) - \frac{1}{2m}\sum_{i=1}^{3}\sum_{j=1}^{3}\frac{\partial^2 g_\gamma}{\partial\theta_i\partial\theta_j}(\bar{\theta})\hat{C}_{ij}. \tag{3.12}$$

When doing so, (3.11) still holds and we reduce bias due to J_γ^{A1} being a nonlinear function of underlying estimators. This is justified under the assumption that $E(\hat{\theta}_n) \neq E\hat{\theta}_{n/2} \neq E\hat{\theta}_{n/4}$. In the numerical results we present in Section 3.6.5, the magnitude of the correction term, i.e., the second term in (3.12) is relatively small in all but one experiment. This empirical observation suggests that the approximation that J_γ^{A1} reduces the "a coefficient" in the bias term may be justified in these cases. Since the bias correction term from (3.12) is relatively small in all but one experiment, we simply use $J_\gamma^{A1} = g_\gamma(\bar{\theta})$. We return to this issue later.

3.5 Multiple replication procedure

We now extend the multiple replication procedure (MRP) for assessing solution quality in a stochastic program to incorporate the adaptive jackknife estimators we have developed above.

Multiple Replication Procedure with J_γ^{A1} (MRP^A)

Input: sample size n (multiple of 4), replications m, confidence level $(1-\alpha)$, candidate solution $\hat{x} \in X$, positive integer γ

Output: approximate $(1-\alpha)$-level CI on optimality gap, $\mu_{\hat{x}}$

1. For $i = 1, \ldots, m$.

   - Generate a random sample $\tilde{\xi}^1, \ldots, \tilde{\xi}^n$, indexed by N, which satisfies the unbiased condition (1.5), and also satisfies this condition when any of the observations are deleted
   - Let N^j, $j = \{1,2,3,4\}$ partion N, with $|N^j| = n/4$, $j = \{1,2,3,4\}$
   - Let $\hat{\theta}_n^i = \hat{f}(\hat{x}, N) - \min_{x \in X} \hat{f}(x, N)$
   - Let $\hat{\theta}_{n/2}^{i1} = \hat{f}(\hat{x}, N^1 \bigcup N^2) - \min_{x \in X} \hat{f}(x, N^1 \bigcup N^2)$
   - Let $\hat{\theta}_{n/2}^{i2} = \hat{f}(\hat{x}, N^3 \bigcup N^4) - \min_{x \in X} \hat{f}(x, N^3 \bigcup N^4)$
   - Let $\hat{\theta}_{n/4}^{ij} = \hat{f}(\hat{x}, N^j) - \min_{x \in X} \hat{f}(x, N^j)$, $j = \{1,2,3,4\}$
   - Let $\hat{\phi}_{n/2}^i = \frac{1}{2}\sum_{j=1}^2 \hat{\theta}_{n/2}^{ij}$ and $\hat{\phi}_{n/4}^i = \dfrac{1}{4}\sum_{j=1}^4 \hat{\theta}_{n/4}^{ij}$

2. Form $\bar{\theta} = \left(\dfrac{1}{m}\sum_{i=1}^m \hat{\phi}_{n/4}^i, \dfrac{1}{m}\sum_{i=1}^m \hat{\phi}_{n/2}^i, \dfrac{1}{m}\sum_{i=1}^m \hat{\theta}_n^i \right)$

3. Let $\hat{C}$ be the sample covariance matrix of $(\hat{\phi}_{n/4}^i, \hat{\phi}_{n/2}^i, \hat{\theta}_n^i)$, $i = 1, \ldots, m$

4. Form $J_\gamma^{A1} = g_\gamma(\bar{\theta})$, and $s^2 = \nabla^T g_\gamma(\bar{\theta}) \hat{C} \nabla g_\gamma(\bar{\theta})$, with g_γ as defined in (3.10)

5. $(1-\alpha)$ CI on $\mu_{\hat{x}}$ is,
$$\left[0, J_\gamma^{A1} + \frac{t_{m-1,\alpha} s}{\sqrt{m}}\right]$$

3.6 Numerical results

In this section, we compare the quality of point and interval estimators obtained by using the original, standard jackknife and adaptive jackknife estimators. We consider seven test problems in all. These include four two-stage SLPs from the literature, an asset allocation problem model from [37], an instance of Example 1 from Section 1.3 with $\delta = 4/3$, and an American-style option pricing problem from [9]. In the first six cases we form point and interval estimates on the optimality gap, but for the pricing problem we instead estimate the option's price, i.e., the optimal value z^*. We can solve six of these seven problems exactly, i.e., there is no need to employ sampling. Our purpose in examining these problems is to assess the estimators' relative performance when we can see, e.g., whether an interval estimator for $\mu_{\hat{x}}$ covers the true value of $\mu_{\hat{x}}$. We choose $\alpha = 0.05$, i.e., we form approximate 95% confidence intervals. We use $m = 30$ replications in all our experiments. Each experiment is repeated 1000 times and averaged values over these 1000 runs are reported for all of the performance measures.

3.6.1 Estimators and performance measures

In computing the relative performance of various estimators, we need to choose the parameter q for the generalized jackknife of Chapter 2 and the parameter γ for the adaptive jackknife estimator of this chapter. In the former case, we choose $q = 1$ as this is the most popular choice in the literature, and because (we assume) we do not know the values of p for our test problems. For this standard jackknife estimator we consider the delete-half estimator, i.e., J_1^B. For the adaptive jackknife estimator J_γ^{A1}, we will consider the first and last members of our family, i.e., J_1^{A1} and J_∞^{A1}, as well as J_γ^{A1} for intermediate values of γ. In addition of course, we will report results for the original estimator $\bar{\theta}_n$.

Typically, when reducing bias one increases sampling error and this occurs with our estimators, too. The standard way to measure this trade-off is via mean-square error (MSE). However, as we have already discussed, because of the one-sided nature of optimality gap estimator, we prefer to err on the side of not reversing the sign of the bias. So, we also examine the negative and positive parts of MSE, which we denote MSE^- and MSE^+, as well as the probability that the gap point estimate is below $\mu_{\hat{x}}$, Pr=P(estimator $< \mu_{\hat{x}}$) (or, P(estimator $> z^*$) when estimating the optimal value). While we prefer estimates with a smaller MSE, we are particularly concerned about controlling the values of MSE^- and Pr, i.e., we prefer to keep these values small.

To compare the interval estimators, we examine Schruben coverage plots [48] of the interval estimators and average 95% confidence interval widths. We again prefer to err on the conservative side, i.e., we wish to reduce the confidence interval widths without having under coverage.

3.6.2 Asset allocation model

Here we present numerical results on the simple asset allocation model that we considered in Section 2.2.3. Figure 3.2 compares our point estimators via MSE, MSE^-, MSE^+ and Pr. The figure plots these four measures with the horizontal axis denoting the aggressiveness, γ, of the adaptive family. We can see that both the standard jackknife and adaptive estimators reduce MSE significantly, but increase MSE^- and Pr relative to the original estimator of the MRP (Section 1.2.3). The standard jackknife estimator outperforms the adaptive jackknife estimators with respect to reducing MSE, but the associated values of MSE^- and Pr are substantially higher.

We next compare the Schruben coverage plots of the interval estimators in Figure 3.3(a). This figure shows that the original interval estimator is very conservative with nearly 100% coverage for the full range of α-values. The adaptive estimators J_1^{A1} and J_∞^{A1} are slightly less conservative but still quite conservative. The standard jackknife estimator performs very well on this problem. Figure 3.3(b) presents the average 95% CI width of the interval estimators, with $n = 120, m = 30$,

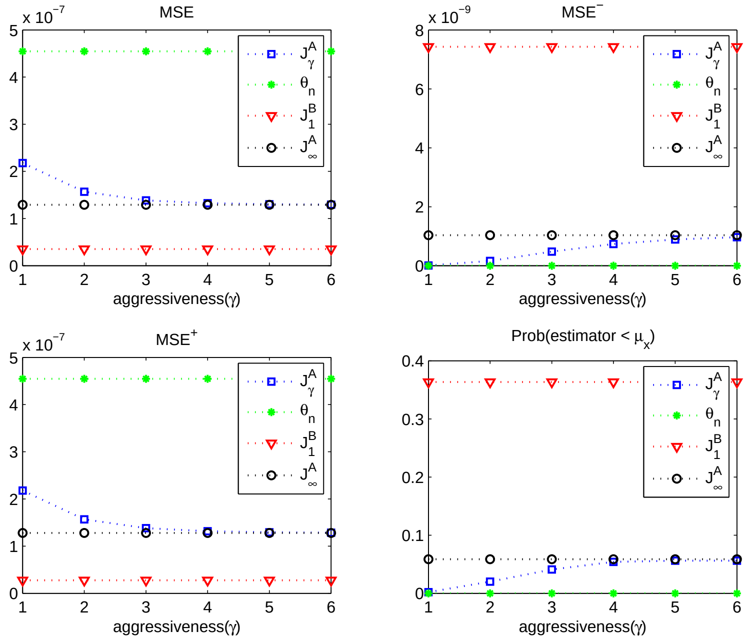

Figure 3.2: Point estimate comparison for asset allocation.

averaged over 2000 repetitions of the experiment. As earlier, we separate the contribution of the sampling error and the point estimate. For the original estimator, we can see that the point estimate's bias dominates the CI width. As the figure shows, J_1^B peforms very well, virtually eliminating the bias, albiet with an increase in sampling error. J_1^{A1} and J_∞^{A1} also reduce bias but not to the degree of J_1^B.

As we indicated above, the asset allocation model appears to have bias that shrinks with $O(n^{-1})$, i.e., $p = 1$ in our notation. In this case the choice of $q = 1$ is ideal for the generalized jackknife. Figure 3.3(b) shows the bias of J_1^{A1} is reduced to about two-thirds of that of $\bar{\theta}_n$ and the bias of J_∞^{A1} is half that of $\bar{\theta}_n$. This is consistent with the predictions of equation (3.9) with $p = 1$, and $\gamma = 1$ and $\gamma = \infty$ in these respective cases.

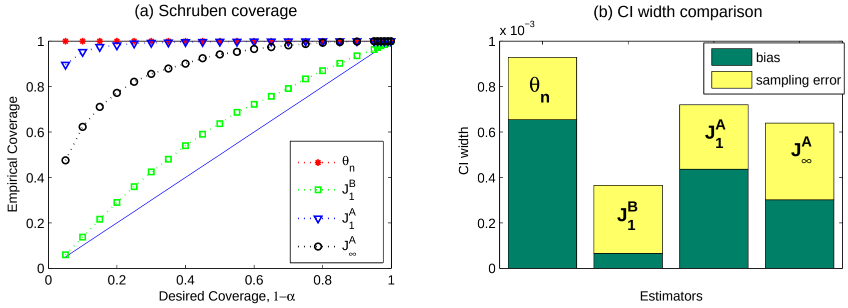

Figure 3.3: Schruben coverage plot in Figure 3.3(a) and CI width plot in Figure 3.3(b) for asset allocation, with $\mu_{\hat{x}} = 1.884 \times 10^{-4}$.

3.6.3 Example 1

The previous example shows the effectiveness when of the generalized jackknife estimator, J_q^B, when we choose q appropriately. Now, we consider an instance of Example 1 from Section 1.3 in which we choose $\delta = 4/3$, which implies the bias is of the form $O(n^{-2})$. Figure 3.4 plots our four performance measures for comparing point estimates. We can see that adaptive estimators reduce MSE significantly without significant increase in MSE^- and Pr. However, the standard jackknife estimator, does not decrease MSE significantly, but does increase MSE^- and Pr dramatically. Here we use $n = 80$ and, as above, $m = 30$ and we average results of 2000 experiments. This illustrates the poor behavior of the generalized jackknife estimator when q is not chosen appropriately.

We next compare the Schruben coverage plots of the interval estimators in Figure 3.5(a). As before, the original estimator is very conservative. The adaptive estimators are conservative, but less so than for the asset allocation model. Finally, the interval estimates for the standard jackknife, J_1^B, show substantial undercoverage. Figure 3.5(b) shows that the adaptive jackknife estimators tighten the average CI width again with a growth in sampling error. The standard jackknife's failure is clear as its point estimate has reversed the sign of the bias. The bias reduction of

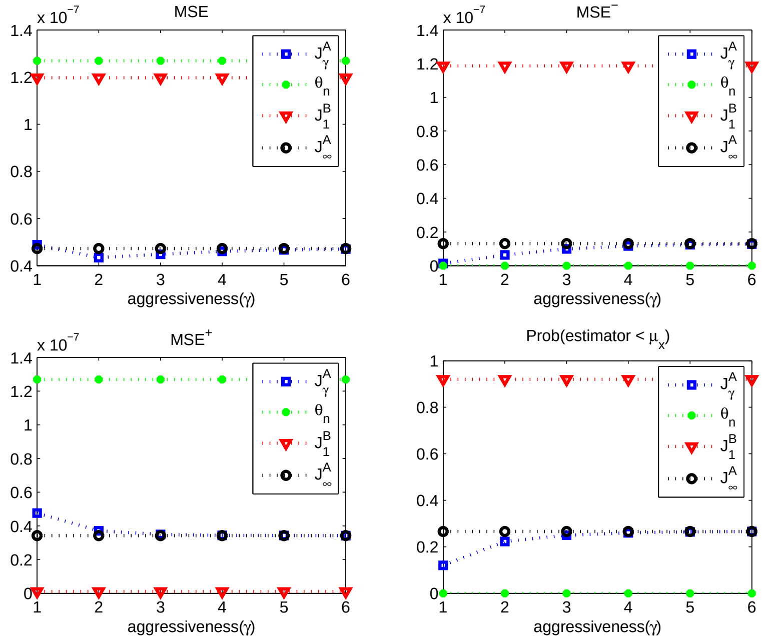

Figure 3.4: Point estimate comparison for Example 1.

J_1^{A1} and J_∞^{A1} in Figure 3.5(b) are again consistent with that predicated by equation (3.9), given that $p = 2$.

Example 1 illustrates the potential dangers associated with using a q in the generalized jackknife estimators that is too aggressive, in this case $q = 1 < p = 2$. The adaptive jackknife estimators are designed not to reverse the sign of the bias, at least when bias is of the form $O(n^{-p})$ for any p, i.e., without requiring any knowledge of p. Indeed, the adaptive estimators perform well on Example 1. The $|x|^\delta$ term that appears in Example 1's objective is similar to terms frequently used in optimization modeling to penalize deviations from a target. Depending on the effective value of δ, and the corresponding p, we may have $p > 1$. This example suggests adaptive estimators may outperform standard jackknifing in such cases. This example also shows that even when bias shrinks to zero as quickly as $O(n^{-2})$, and sampling error as $O(n^{-1/2})$, bias can still constitute the primary contribution to the CI width. This occurs, in part, because of the variance reduction provided in the various MRP

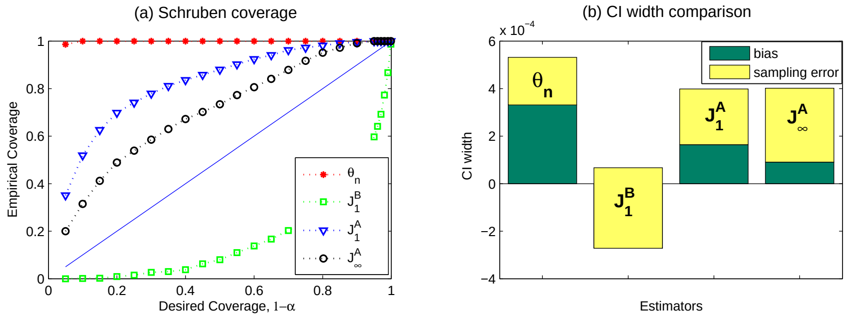

Figure 3.5: Schruben coverage plot in Figure 3.5(a) and CI width plot in Figure 3.5(b) for Example 1, with $\mu_{\hat{x}} = 0$.

schemes by using common random numbers (CRN).

3.6.4 American-style options pricing

An American option can be exercised at any time up to its expiration date. We consider a similar option, known as a Bermudan option, in which the opportunity to exercise is limited to certain pre-specified dates. (In some cases concerning periodic dividend payments, it is only optimal to exercise at certain dates and the American option becomes a Bermudan option.) We consider an instance of a Bermudan option from [9] in which the (single) underlying asset has an initial price of 110, a strike price of 110, a riskless interest rate of 5%, a dividend rate of 10% and the asset's price moves according to geometric Brownian motion (GBM) with a drift of -0.05 (riskless interest rate less dividend rate) and a volatility of 0.2. The option has three exercise opportunities at $T/3, 2T/3$, and T, where the option horizon is $T = 1$. Under the GBM assumption, the price of the option can be computed exactly as $z^* = 11.341$. In finance, it is the option's price as opposed to the optimal exercise strategy, that is of foremost importance and so we seek to estimate z^* rather than an optimality gap. This option pricing problem is a *multistage* stochastic optimization model, but we can still employ our adaptive jackknife estimators. We build an empirical

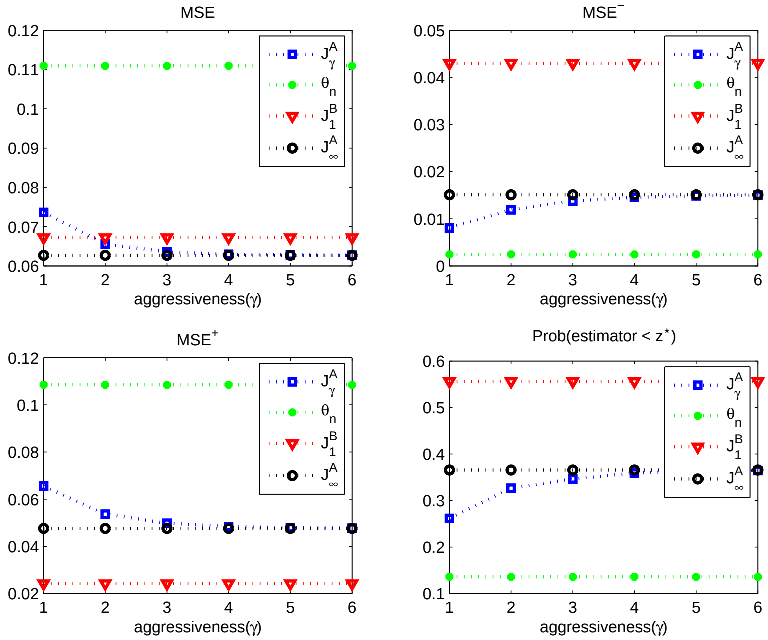

Figure 3.6: Point estimate comparison for pricing Bermudan call option.

(sampled) scenario tree with n branches at each node. This is used to form the original estimator $\hat{\theta}_n$. The estimators $\bar{\phi}_{n/2}$ and $\bar{\phi}_{n/4}$ are obtained from the same tree with (A4) maintained at every node on the tree.

Note that here we have a maximization problem and we estimate z^*. We use $n = 48$ so that the empirical trees (under a full sample) have 48^3 nodes. The CIs we form here are one-sided, e.g., we are confident at level 95% that z^* is no larger than $\bar{\theta}_n + t_{m-1,\alpha} s_m/\sqrt{m}$. Clearly, this is an incomplete analysis for pricing an option. In Broadie and Glasserman [9], they couple this type of one-sided CI with a lower-limit interval estimate, where the point estimate essentially comes from analyzing the performance of an exercise policy. We limit our discussion here to the one-sided upper-limit, and only indicate that a full analysis of the option's price would require the lower-limit, too.

Figure 3.6 plots our four performance measures. Here, we can see that again the adaptive jackknife estimators outperform the standard jackknife estimator. The

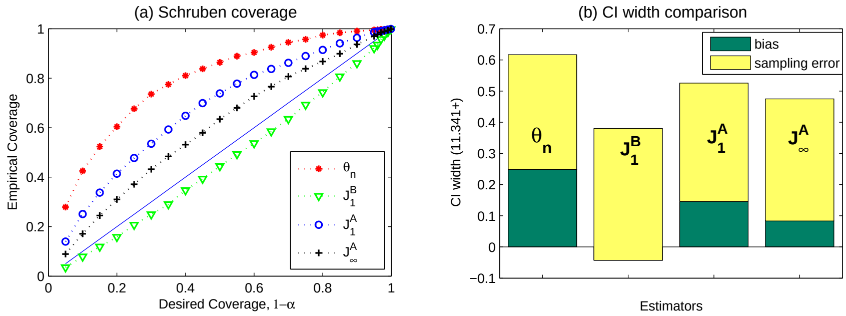

Figure 3.7: Schruben coverage plot in Figure 3.7(a) and CI width plot in Figure 3.7(b) for pricing Bermudan call option, with optimal price=11.341. In Figure 3.7(b) the CIs do not have a lower limit.

improvement in MSE of the adaptive jackknife estimators over J_1^B is modest but the differences in MSE$^-$ and Pr are significant.

We next compare the Schruben coverage plots of the interval estimators in Figure 3.7(a). The standard jackknife again exhibits undercoverage (although less dramatically than in Example 1), and the adaptive estimators perform well. Figure 3.7(b) illustrates CI widths. Here, our CI statement for the MRP procedure of Section 1.2.3 is $P(z^* \leq \bar{\theta}_n + t_{m-1,\alpha} s_m/\sqrt{m}) \approx 1-\alpha$, and Figure 3.7(b) plots the magnitude of $\bar{\theta}_n$, relative to z^*, as well as the sampling errors. Notably, the sampling errors are substantially larger relative to bias in Figure 3.7(b) compared to those we have examined for optimality gap estimation. This is largely due to the variance reduction achieved by using CRNs when forming the difference $\hat{f}(\hat{x}, \tilde{\xi}^1, \ldots, \tilde{\xi}^n) - \min_{x \in X} \hat{f}(x, \tilde{\xi}^1, \ldots, \tilde{\xi}^n)$ in the latter case. Restricting attention to bias we see that the adaptive estimators reduce bias significantly without tending to over-correct, but the standard jackknife estimator does reverse the sign of the bias.

3.6.5 Two-stage stochastic linear programs

Table 3.1 lists the results for three standard test problems from the stochastic programming literature, namely, APL1P, PGP2 and CEP1 (see [25].) These three SLPs are small with a modest number of scenarios, i.e., they can be solved exactly. For each problem we carry out optimality gap estimation using $\hat{x} = x^*$, where x^* is an optimal solution to (1.1). Thus, the averaged point estimates we report consist (in expectation) solely of bias. The results for APL1P and PGP2 are roughly similar to that of the asset allocation problem. However, CEP1 presents a case where all the jackknife estimators fail to reduce MSE. This happens because the bias in the optimality gap (equivalently in z_n^*) does not appear to shrink to zero as $O(n^{-p})$. For some stochastic programs this bias can shrink to zero at a rate faster than $O(n^{-p})$ for any finite p. In our experiments with CEP1, $\bar{\theta}_{30}$ and $\bar{\theta}_{60}$ had significant bias, but $\bar{\theta}_{120}$ took value zero in 90% of the repetitions. Of course, when the bias shrinks to zero this quickly, one might argue that bias-correcting estimators are not necessary. Still, this example suggests that even though the adaptive estimators appear conservative in the other cases we have examined (i.e., not over-correcting bias), this is not universally true.

Note that since the adaptive estimators are formed as a nonlinear function of sample means, we could use a second-order Taylor series bias correction as described in Section 3.4. The last two columns in Table 3.1 show the bias corrected values of MSE and point estimates of adaptive estimators. The bias correction for APL1P and PGP2 is relatively small, however, for CEP1 it is a large percentage of the point estimate. We conjecture that the Taylor series bias correction may serve as an indicator as to whether the adaptive estimators are in danger of failing, but we do not pursue this issue further here.

Table 3.1: Stochastic linear programs with sample size, $n = 120$

	MSE	MSE^{+}	MSE^{-}	Pr	Pt. Est.	Samp. Err.	Coverage	Bias corrected	
								MSE	Pt. Est.
APL1P, $\mu_{\hat{x}} = 0$, $z^* = 24,642.32$									
$\hat{\theta}_n$	1529.62	1529.62	0.00	0.000	37.73	16.46	1.000		
J_1^{A1}	787.06	787.00	0.06	0.004	25.54	19.05	1.000	808.14	25.97
J_∞^{A1}	547.99	533.21	14.78	0.097	18.04	25.00	1.000	595.16	19.80
J_1^{B}	196.07	146.76	49.31	0.402	3.91	22.08	0.977		
PGP2, $\mu_{\hat{x}} = 0$, $z^* = 447.32$									
$\hat{\theta}_n$	8.89	8.89	0.00	0.000	2.93	0.95	1.000		
J_1^{A1}	3.70	3.70	0.00	0.010	1.79	1.18	1.000	3.75	1.80
J_∞^{A1}	1.69	1.50	0.19	0.183	0.81	1.70	0.992	1.80	0.88
J_1^{B}	0.84	0.70	0.14	0.287	0.43	1.37	0.976		
CEP1, $\mu_{\hat{x}} = 0$, $z^* = 355,164.46$									
$\hat{\theta}_n$	34.18	34.18	0.00	0.000	1.22	1.92	1.000		
J_1^{A1}	110.66	22.60	88.06	0.697	-4.04	11.65	1.000	83.52	-2.51
J_∞^{A1}	457.88	20.82	437.06	0.705	-7.39	21.64	1.000	125.58	-1.99
J_1^{B}	697.26	9.58	687.68	0.775	-17.47	24.03	0.937		

3.6.6 Computational time comparison

We now analyze the computational effort required to achieve similar confidence interval widths, using adaptive jackknife estimators J_γ^{A1} and using $\hat{\theta}_n$. We choose a larger two-stage SLP from the stochastic programming literature, SSN [49] as well as the Bermudan option pricing problem for this analysis. For SSN we compare the performance of estimators, J_1^{A1} and J_∞^{A1} based on a sample size of $n = 8000$ against to that of the naive estimator with double the sample size, $n = 16000$. For the Bermudan option pricing problem we compare the estimators J_1^{A1} and J_∞^{A1}, based on a sample size of $n = 240$, against the naive estimator with a sample size of $n = 300$. We use the regularized decomposition (RD) as implemented by Ruszczyński and Świetanowski [47] to solve instances of SSN. The RD algorithm couples a multi-cut Benders' decomposition with a quadratic proximal term for two-stage SLPs. All problem instances are solved from scratch, i.e., we do not exploit any warm starts. The results are listed in Table 3.2. For SSN, J_∞^{A1} provides an improvement in CI width that is 70% of what we can obtain by doubling the sample size in the naive estimator, with similar computational effort. However, there is significant potential for accelerating the computation time of the adaptive estimators by reusing cut information when solving the seven related instances of SSN (one with sample size n, two with sample size $n/2$ and four with sample size $n/4$) required for obtaining one estimate of J_γ^{A1}. Moreover, if general purpose commercial solvers are used to solve problem instances, then increasing (doubling) the sample size may not be a viable option. The results we have presented are preliminary and further work is required to develop efficient procedures for re-using cut information when estimating J_γ^{A1}. For Bermudan options pricing, it is clear that the adaptive estimators require significantly less time to provide similar CI width estimates obtained by increasing the sample size in the naive estimator.

Table 3.2: Computational time comparison.

	SSN				Bermudan option pricing			
	$\hat{\theta}_{8000}$	J_1^{A1}	J_∞^{A1}	$\hat{\theta}_{16000}$	$\hat{\theta}_{240}$	J_1^{A1}	J_∞^{A1}	$\hat{\theta}_{300}$
Point Estimate	0.502	0.473	0.454	0.453	11.3671	11.3524	11.3462	11.3699
Sampling Error	0.029	0.031	0.036	0.020	0.1760	0.1777	0.1787	0.1599
CI Width	0.531	0.503	0.489	0.473	11.5431	11.5301	11.5249	11.5298
% Δ(CI width)	–	5.160	7.840	10.910	–	0.113	0.160	0.115
MSE					0.0126	0.0123	0.0123	0.0107
MSE$^+$					0.0089	0.0074	0.0069	0.0080
MSE$^-$					0.0037	0.0048	0.0054	0.0028
Time	49 hrs	121 hrs	121 hrs	114 hrs	5 min	5 min	5 min	10 min

3.7 Summary

We have developed a new adaptive jackknife estimator in this chapter, and we provided conditions when consistency in the underlying estimator is inherited by the adaptive jackknife estimator. We further developed results comparing the bias of the standard jackknife estimator with that of our adaptive estimator. We then presented a family of adaptive estimators in which more aggressive bias reduction can be obtained by choosing larger values of the aggressiveness parameter γ. We argued that using generalized jackknife estimators (e.g., with the most popular choice of $q = 1$) may backfire and that our adpative estimators provide an attractive alternative for bias reduction, at least when bias is of form $O(n^{-p})$. The numerical results we obtained on a range of problems were encouraging. The adaptive estimators performed well in cases where the bias is of the form $O(n^{-p})$, and outperformed generalized jackknife estimators when the associated choice of q was too aggressive. The adaptive estimator J_γ^{A1} uses a particular functional form $h_\gamma(\hat{r}) = \sum_{k=1}^{\gamma} \hat{r}^k$ as a multiplier in the expression $J_\gamma^{A1} = \bar{\theta}_n - h_\gamma(\hat{r})(\bar{\phi}_{n/2} - \bar{\theta}_n)$. As we will see in the next chapter, other functional forms for $h_\gamma(\hat{r})$ are possible.

Chapter 4

Families of Adaptive Jackknife Estimators, with Enhancements

In Chapter 3 we argued that adaptive jackknife estimators can provide an attractive alternative to the generalized jackknife estimators for bias reduction. The adaptive family that we described in Chapter 3 can be used to reduce bias with little danger of reversing the sign of the bias. We now describe two families of adaptive jackknife estimators that can be used for more aggressive bias reduction. We also describe how the common random number scheme described in the AJP can be enhanced to improve bias reduction. At the end of the chapter, we provide some recommendation on selecting from our families of adaptive estimators, including selection of the associated parameters that we describe shortly.

4.1 Two more families of adaptive jackknife estimator

In Chapter 3 we developed a family of adaptive jackknife estimators, J_γ^{A1}, in (3.8) that have form $J_\gamma^{A1} = \bar{\theta}_n - h_\gamma(\hat{r})(\bar{\phi}_{n/2} - \bar{\theta}_n)$, where $h_\gamma(\hat{r}) = \sum_{k=1}^{\gamma} \hat{r}^k$, and where $\hat{r}$ is defined in (3.5). In this chapter we introduce two more families, J_γ^{A2} and J_γ^{A3}, with different choices of $h_\gamma(\hat{r})$. Our motivation is the following: At the close of Section 3.2 we argued that as the number of replications m grows large the bias of J_1^{A1} is of the form $\frac{2}{2^p+1} a n^{-p} + o(n^{-p})$, and that more generally (see Section 3.3) the bias of J_γ^{A1} has form

$$a\left(1-\left(1-\frac{1}{2^p}\right)\left(1-\frac{1}{(2^p+1)^\gamma}\right)\right)n^{-p}+o(n^{-p}).$$

To make this concrete, if $p = 1$ and we choose $\gamma = 1$ we will have effectively replaced the "a" in $\bar{\theta}_n$'s bias formula by $\frac{2}{3}a$, or if we choose $\gamma = \infty$, by $\frac{1}{2}a$. If instead the underlying estimator has $p = 2$ these respective values are $\frac{2}{5}a$ and $\frac{1}{4}a$. Finally, if $p = 1/2$ the respective values are about $0.83a$ and $0.71a$. So, we are motivated to choose functional forms $h_\gamma(\hat{r})$ that can improve upon those reductions in the effective value of a. At the same time, we still wish to be conservative in the sense that we do not want to over-correct, i.e., change the sign of the bias. Using two different forms of $h_\gamma(\hat{r})$ we now propose the following two new adaptive jackknife families:

$$J_\gamma^{A2} = \bar{\theta}_n - \left(\sum_{k=1}^{\gamma} k\hat{r}^k\right)(\bar{\phi}_{n/2} - \bar{\theta}_n) \tag{4.1a}$$

$$J_\infty^{A2} = \begin{cases} \bar{\theta}_n - \frac{\hat{r}}{(1-\hat{r})^2}(\bar{\phi}_{n/2} - \bar{\theta}_n) & \text{if } \hat{r} < 1 \\ \bar{\theta}_n & \text{if } \hat{r} = 1 \end{cases} \tag{4.1b}$$

$$J_\gamma^{A3} = \bar{\theta}_n - \left(\sum_{k=1}^{\gamma} \frac{1}{2}(2\hat{r})^k\right)(\bar{\phi}_{n/2} - \bar{\theta}_n) \tag{4.2a}$$

$$J_\infty^{A3} = \begin{cases} \frac{\bar{\phi}_{n/4}\bar{\theta}_n - \bar{\phi}_{n/2}^2}{(\bar{\phi}_{n/4} - \bar{\phi}_{n/2}) - (\bar{\phi}_{n/2} - \bar{\theta}_n)} & \text{if } \hat{r} < 1 \\ \bar{\theta}_n & \text{if } \hat{r} = 1. \end{cases} \tag{4.2b}$$

In Theorem 6 we established consistency of J_1^{A1} and we extended that to J_γ^{A1} in Theorem 8. The key to these proofs is that $h_\gamma(\hat{r}) = \sum_{k=1}^{\gamma} \hat{r}^k$ is bounded w.p.1. That property is ensured by assumption (A4). Theorem 9 below extends these consistency results to J_γ^{A2} and J_γ^{A3}. Its proof hinges on boundedness of $h_\gamma(\hat{r})$, as illustrated in the proof of Theorem 6, and is hence omitted.

Theorem 9. *Assume that the original estimator $\hat{\theta}_n$ is consistent. Let γ be a positive integer. Let J_γ^{A2}, J_γ^{A3} and $\hat{r}$ be as defined in AJP, except that (4.1a) and (4.2a), respectively, replace (3.5a). If (A4) holds, then, $\lim_{n\to\infty} J_\gamma^{A2} = \theta$, w.p.1, and $J_\gamma^{A3} = \theta$, w.p.1.*

We now mimic the analysis at the end of Section 3.3 to understand the asymptotic form of the bias. In particular, assuming $E\hat{\theta}_n = \theta + an^{-p} + o(n^{-p})$

and letting the number of the replications, m, grow large we obtain

$$EJ_\gamma^{A2} \approx \theta + an^{-p}\left(1 - \left(1 - \frac{1}{4^p}\right)\left(1 - \frac{1+(\gamma+1)2^p}{(2^p+1)^{\gamma+1}}\right)\right) + o(n^{-p}) \quad (4.3a)$$

$$EJ_\gamma^{A3} \approx \theta + an^{-p}\left(\frac{2}{2^p+1}\right)^\gamma + o(n^{-p}). \quad (4.3b)$$

We now consider how the asymptotic bias in (4.3) behaves. First, note from (3.8), (4.1) and (4.2) that $J_1^{A1} = J_1^{A2} = J_1^{A3}$, i.e., these three estimators are idenctical and hence so is the nature of their bias reduction. However, for J_γ^{A2} we see from (4.3a) that as γ grows large the "a" in the bias formula of original estimator is replaced by $\frac{1}{2}a$, $\frac{1}{4}a$ and $\frac{1}{16}a$ in the respective cases of $p = 1/2$, 1, and 2, and in the sense is more aggressive than the most aggressive member of the J_γ^{A1} family. From (4.3b) we see that as γ grows large that the leading order term of the bias is eliminated. Note that for any $p > 0$ and positive γ the coefficient multiplying an^{-p} lies in (0,1) and hence may provide bias reduction without reversing its sign.

As already indicated, $J_1^{A1} = J_1^{A2} = J_1^{A3}$. For and $\hat{r} > 0$, when $\gamma = 2$ we have $J_2^{A1} \geq J_2^{A2} = J_2^{A3}$, and when $\gamma > 2$ we have $J_\gamma^{A1} > J_\gamma^{A2} > J_\gamma^{A3}$. This indicates as we move from family 1 to family 2 to family 3 we are more aggressive in reducing bias for a fixed value of γ. And, within a family we are more aggressive in reducing bias as γ grows. In fact, as $\lim_{\gamma\to\infty} J_\gamma^{A3} = J_\infty^{A3}$ as defined in (4.2b), and we see that this has the same form as $\hat{\theta}^A$ defined in (2.8a), which arose by solving the nonlinear system of three equations in three unknowns given by (2.6). We know from Chapter 2 that even basic properties like consistency can therefore fail to hold for J_∞^{A3}. This further suggests that even though J_γ^{A3} is consistent when γ is finite (Theorem 9) that J_γ^{A3} may be poorly behaved when γ is large.

We now further pursue what happens as γ grows large in the adaptive estimators. It is reasonable anticipate that $b(\hat{\theta}_n)$ shrinks to zero at least as quickly as $O(n^{-p})$ for $p = 1/2$. Said another way, if $b(\hat{\theta}_n) = O(n^{-p})$ for $p < 1/2$ then $\sqrt{n}(\hat{\theta}_n - \theta)$ will explode, and so if our naive estimator satisfies a CLT with scaling factor $\sqrt{n}$

then $b(\hat{\theta}_n)$ must satisfy the above condition. From equation (3.4) we know that for $p \geq 1/2$ we have $r \leq \frac{1}{1+\sqrt{2}} \approx 0.41$ as long as bias shrinks to zero (i.e., $p > 0$) we have $r \leq 1/2$. Of course, when we estimate r via $\hat{r}$, it is possible $\hat{r} > 1/2$ and all we are assured under (A4) is that $0 \leq \hat{r} \leq 1$ from Theorem 5. Still, if $\hat{r}$ is to estimate r well, we should anticipate $\hat{r} \leq 1/2$. Now consider $h_\gamma(\hat{r})$ for three families, which we denote $h_\gamma^{Ai}(\hat{r})$, $i = 1, 2, 3$. In all cases as $\hat{r}$ grows $h_\gamma^{Ai}(\hat{r})$ grows and the term we subtract from $\bar{\theta}_n$ to form the adaptive estimator grows, i.e., bias reduction is more aggressive. For all three families as $\hat{r} \to 1$ we have, $h_\gamma^{Ai}(\hat{r}) \to \infty$ with $h_\gamma^{Ai}(\hat{r})$'s growth being faster as we move from family 1 to family 2 to family 3. In such cases the associated estimates will fail. Now compare $h_2^{A2}(\hat{r}) = h_2^{A3}(\hat{r}) = \hat{r} + 2\hat{r}^2$ and consider the most aggressive member of family 1, $h_\infty^{A1}(\hat{r}) = \frac{\hat{r}}{1-\hat{r}}$. The range of values for which $J_2^{A2} = J_2^{A3}$ is more aggressive in reducing bias than J_∞^{A1} is $\hat{r} + 2\hat{r}^2 > \frac{\hat{r}}{1-\hat{r}}$, i.e., $\hat{r} < 1/2$. This suggests (given our above discussion) that when $\hat{r}$ is well behaved, i.e., less than 1/2, $J_2^{A2} = J_2^{A3}$ should outperform J_∞^{A1}. And, when $\hat{r}$ is poorly behaved and near 1 that J_∞^{A1} could fail. This suggests there is merit in investing the performance of families 2 and 3, particularly when γ is small. However, because $h_\gamma^{Ai}(\hat{r})$ grows large more quickly when $\hat{r}$ is large and this effect is more pronounced with the more aggressive families, it suggests we may need to exercise caution, i.e., not choose γ too large. We will investigate this in our computation results, but first we discuss another "parameter" that has been implicit in the development of our estimators so far.

4.2 Generalized partition factor, k

In forming the system of nonlinear equations (2.6) that motivated our development of adaptive jackknife estimators, we partitioned a sample of size n into two subsamples of size $n/2$ and further partitioned those into further subsamples of size $n/4$. (See Figure 3.1.) Instead we could have partitioned the original sample into three subsamples of size $n/3$ and repeated that to obtain 9 subsamples of size $n/9$. Pursuing this generalization with integer $k \geq 2$ leads to the following variant of (2.6)

$$\begin{aligned} E\hat{\theta}_{n/k^2} &= \theta + a\left(\frac{n}{k^2}\right)^{-p} && (4.4a) \\ E\hat{\theta}_{n/k} &= \theta + a\left(\frac{n}{k}\right)^{-p} && (4.4b) \\ E\hat{\theta}_{n} &= \theta + an^{-p}, && (4.4c) \end{aligned}$$

We can now mimic the development of J_1^{A1} in Section 3.1 except that equations (2.6) are replaced by (4.4) and we can corresspondingly modify AJP to include the generalized partition scheme. Below we present AJP^k, which generalized AJP to use integer partition factor $k \geq 2$.

<u>Adaptive jackknife procedure with generalized partitions</u> (AJP^k) Input: partition factor k, sample size n which is a multiple of k^2, replications m

Output: An adaptive estimator J^A of θ

1. For $i = 1, \ldots, m$
   - Generate a sample of size n indexed by N
   - Let N^j, $j = 1, \ldots, k^2$, partition N, with $|N^j| = n/k^2$, $j = 1, \ldots, k^2$
   - Let $\hat{\theta}_n^i$ be the underlying estimator based on the full sample N
   - Let $\hat{\theta}_{n/k}^{ij}$, $j = 1, \ldots, k$, be the underlying estimators using $\displaystyle\bigcup_{t=(j-1)k+1}^{jk} N^t$, $j = 1, \ldots, k$, respectively
   - Let $\hat{\theta}_{n/k^2}^{ij}$, $j = 1, \ldots, k^2$, be the underlying estimators using N^j, $j = 1, \ldots, k^2$, respectively
   - Let $\hat{\phi}_{n/k^2}^i = \dfrac{1}{k^2}\displaystyle\sum_{j=1}^{k^2} \hat{\theta}_{n/k^2}^{ij}$ and $\hat{\phi}_{n/k}^i = \dfrac{1}{k}\displaystyle\sum_{j=1}^{k} \hat{\theta}_{n/k}^{ij}$

2. Let $\bar{\phi}_{n/k^2} = \frac{1}{m}\sum_{i=1}^{m} \hat{\phi}^i_{n/k^2}$, $\bar{\phi}_{n/k} = \frac{1}{m}\sum_{i=1}^{m} \hat{\phi}^i_{n/k}$, and $\bar{\theta}_n = \frac{1}{m}\sum_{i=1}^{m} \hat{\theta}^i_n$

3. Define adaptive jackknife estimators,

$$J_1^{A1}(k) = \bar{\theta}_n - \hat{r}(\bar{\phi}_{n/k} - \bar{\theta}_n) \tag{4.5a}$$

$$\hat{r}(k) = \frac{\bar{\phi}_{n/k} - \bar{\theta}_n}{\bar{\phi}_{n/k^2} - \bar{\theta}_n} \tag{4.5b}$$

AJP^k defines $J_1^{A1}(k)$, i.e., extends $J_1^{A1} \equiv J_1^{A1}(2)$ to use a general partition factor $k = 2, 3, \ldots$. The extensions to handle $J_\gamma^{A1}(k)$ with $\gamma \geq 1$ as well as $J_\gamma^{A2}(k)$ and $J_\gamma^{A3}(k)$ are straight forward. In Sections 3.2 and 3.3 we established a consistency result for (what we now label) $J_\gamma^{A1}(2)$. This result relies on two hypotheses: consistency of the original estimator $\hat{\theta}_n$ and (A4). Assuming (A4) holds, we can apply it recursively to obtain $k = 3$, or $k = 4$, etc., terms on the right-hand side of (3.6). This observation is key to establishing consistency of $J_\gamma^{A1}(k)$, $J_\gamma^{A2}(k)$ and $J_\gamma^{A3}(k)$. In Chapter 3 we also showed there is a relationship between the bias of: The original estimator $\hat{\theta}_n$, the generalized jackknife estimator with parameter $q \leq 1$ and $J_1^{A1}(2)$. The following theorem summarizes the above discussion on consistency to our new families.

Theorem 10. *Assume that the original estimator $\hat{\theta}_n$ is strongly consistent. Let $\gamma \geq 1$ and $k \geq 2$ be integers and let $J_\gamma^{A1}(k)$, $J_\gamma^{A2}(k)$, and $J_\gamma^{A3}(k)$ be defined through AJP^k's extension of (3.8a), (4.1a) and (4.2a), respectively. If (A4) holds then $\lim_{n\to\infty} J_\gamma^{A1}(k) = \lim_{n\to\infty} J_\gamma^{A2}(k) = \lim_{n\to\infty} J_\gamma^{A3}(k) = \theta$, w.p.1.*

Assuming $E\hat{\theta}_n = \theta + an^{-p} + o(n^{-p})$ holds, under the generalized partition factor k, as the number of the replications, m, increases, we have

$$EJ_1^{A1}(k) \approx \theta + an^{-p}\frac{2}{k^p + 1} + o(n^{-p}). \tag{4.6}$$

We can similarly extend the expression for $EJ_\gamma^{A1}(k)$ from that of $EJ_\gamma^{A1}(2)$ in (3.9) as well as $EJ_\gamma^{A2}(k)$ and $EJ_\gamma^{A3}(k)$ from (4.3a) and (4.3b). (4.6) suggests that increasing the partition factor k may reduce the bias. However, in what follows we restrict ourselves to $k \in \{2, 3, 4\}$ for the following two reasons: (i) The total number of stochastic programs we must solve in every replication grows quadratically in k, as $1 + k + k^2$; and (ii) As we increase k, the smallest stochastic programs have sample size $\frac{n}{k^2}$ and if k is too large the sample size may not be large enough so that the bias is effectively of form $O(n^{-p})$. We discuss these issues in more detail in the Section 4.4.

4.3 Rotation policies

In Sections 3.1 and 3.2 we motivated the use of common random number (CRN) streams in forming $\bar{\theta}_n, \bar{\phi}_{n/2}$, and $\bar{\phi}_{n/4}$ in AJP by the desirable property it yielded in Theorem 5, which in turn was key to establishing consistency of $J_\gamma^{A1}(2)$ in Theorems 6 and 8. CRN streams also have the benefit of reducing variance. We now describe what we call the rotation policy (RP) that can help further reduce variance.

Rotation Policy. *We return to the development of $\hat{\phi}_{n/2}$ in the AJP of Section 3.1. Let $\hat{\theta}^{ij}_{n/2}$ be the underlying estimator based on samples in partition $N^i \cup N^j$, $i < j$, $i, j \in \{1, 2, 3, 4\}$. Redefine $\hat{\phi}_{n/2}$ as*

$$\hat{\phi}_{n/2} = \frac{1}{6} \sum_{i=1}^{3} \sum_{j=i+1}^{4} \hat{\theta}^{ij}_{n/2}. \tag{4.7}$$

In understanding the rotation policy, it helps to refer to Figure 3.1. The estimator $\hat{\phi}_{n/2}$ in the original AJP of Section 3.1 is based on averaging what we now label $\hat{\theta}^{12}_{n/2}$ and $\hat{\theta}^{34}_{n/2}$, i.e., the former is based on $N^1 \cup N^2$, the first $n/2$ observations and the latter is based on $N^3 \cup N^4$, the last $n/2$ observations. In the RP we form

$\begin{pmatrix} 4 \\ 2 \end{pmatrix} = 6$ observations of $\hat{\theta}_{n/2}$ by using all combinations of different subsamples of size $n/4$. So, relative to AJP under the original non-rotation policy, we now must solve 4 additional stochastic programs with sample size $n/2$. The expression for $\hat{\phi}_{n/2}$ in (4.7) has the same expectation as that of the original AJP and so the bias of $J_{\gamma}^{A1}(2)$ in unchanged under RP, i.e., Theorem 7 still holds. Furthermore, we can extend Theorem 5 to establish that $\bar{\phi}_{n/4} \geq \bar{\phi}_{n/2} \geq \bar{\theta}_n$, w.p.1, when $\bar{\phi}_{n/2}$ is formed using RP, and as a result we also obtain consistency of $J_{\gamma}^{A1}(2)$ under RP. We can also extend RP so that it applies to $J_{\gamma}^{A1}(k)$ with $k \geq 2$, and instead of forming and averaging k observations of $\hat{\theta}_{n/k}$ to form $\hat{\phi}_{n/k}$ we form and average $\begin{pmatrix} k^2 \\ k \end{pmatrix}$ observations of $\hat{\theta}_{n/k}$. This again suggests we should keep k moderate in size. The RP further extends, to $J_{\gamma}^{A2}(k)$ and $J_{\gamma}^{A3}(k)$ in a straightforward way.

4.4 Numerical results

In this section we present numerical results to compare bias reduction capabilities of $\bar{\theta}_n$, J_1^B, $J_{\gamma}^{A1}(k)$, $J_{\gamma}^{A2}(k)$, and J_{γ}^{A3} for $k = 2, 3, 4$. We also present numerical results after incorporating Latin hypercube sampling for variance reduction in our underlying estimators. In this section we point out shortcomings of the adaptive estimators, and using the enhancements we also suggest changes to overcome these shortcomings. We begin with the sample size $n = 144$ and number of replications, $m = 30$. We present results which are based on an average of 1000 repitition of the estimators. We choose $\alpha = 0.05$, i.e., we form approximate 95% confidence intervals. We use a subset of the test problems described in Chapter 3.

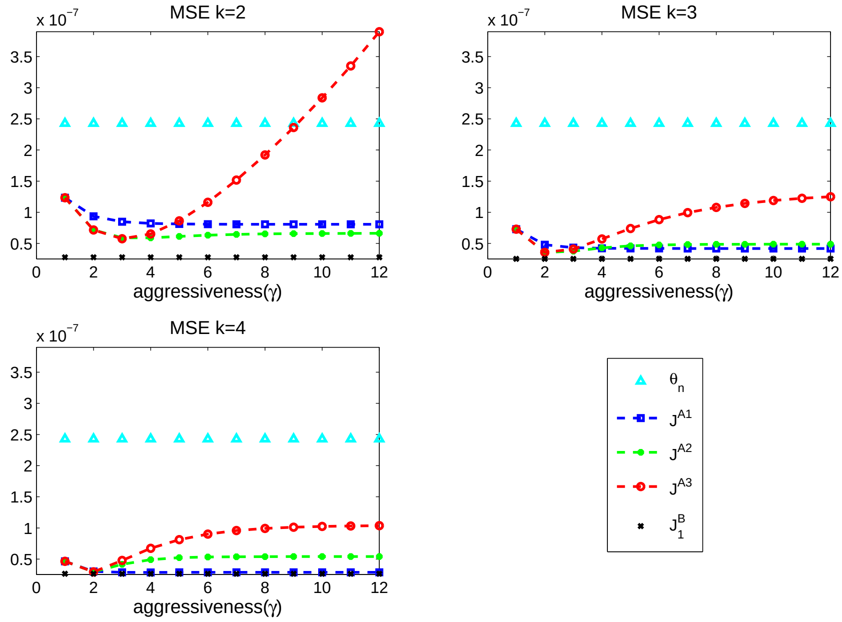

Figure 4.1: MSE plots without bias correction for asset allocation model.

4.4.1 Asset allocation model

Figure 4.1 contains the MSE plots without any bias correction (see equation (3.12)) for the asset allocation model. (Throughout when we refer to MSE it represents MSE of the estimator without bias correction unless otherwise specified.) Figure 4.2 contains MSE plots with bias correction. In Section 3.6.5, we suggested that the magnitude of the bias correction term in point estimate may indicate whether use of the adaptive estimator is appropriate. Figures 4.3 and 4.4 contain the point estimate plots for asset allocation model without and with bias correction respectively. From Figures 4.1-4.4 we see that for $k = 2$ with larger values of γ, there are significant differences in point estimates of $J_\gamma^{A3}(2)$ and its MSE with and without bias correction. The MSE for J_γ^{A3} and $k = 2$ is exploding as we increase γ, suggesting failure of this adaptive jackknife estimator.

A potential cause for this behavior may be larger variance in the underlying

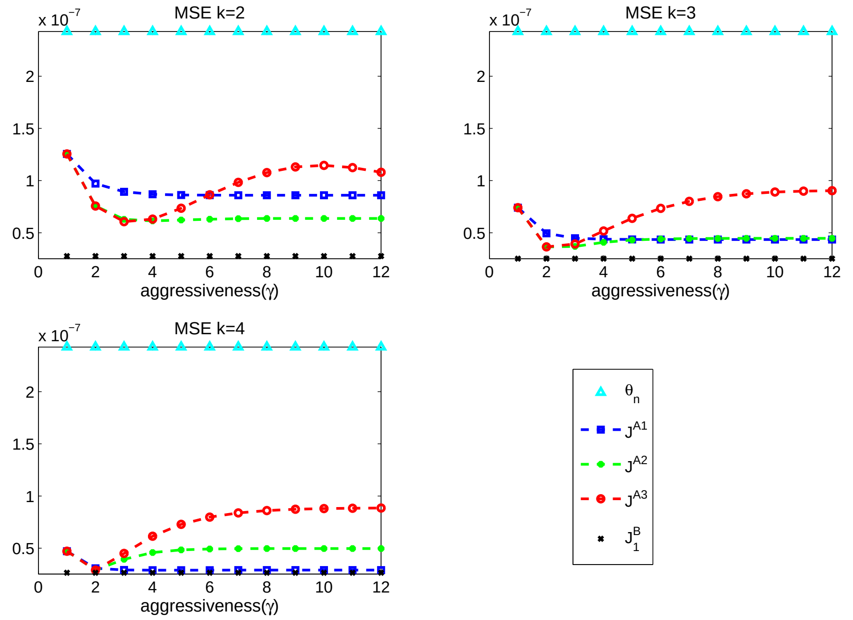

Figure 4.2: MSE plots with bias correction for asset allocation model.

estimators. To help reduce the variance of the estimator with $k = 2$, we investigate the application of the rotation policy, RP. Note that for values of $k = 3$ (or 4), the estimators $\bar{\phi}_{n/k}$ and $\bar{\phi}_{n/k^2}$ are formed by averaging 3 (or 4) and 9 (or 16) underlying estimators, hence they have reduced variance compared to the estimator with $k = 2$. This explains the relative stability of the estimators shown in Figures 4.1-4.4 for larger values of k. Figure 4.5 presents MSE plots using RP for $k = 2$. We see that RP has helped stabilize $J^{A3}_{\gamma}(k)$, i.e., MSE plots, with and without bias correction, after using RP are very close. With use of RP, not only do the MSE plots become stable, but families 2 and 3 almost attain the same MSE as J^B_1. For the asset allocation model, $p \approx 1$, and hence, J^B_1 almost completely removes the bias. In what follows, we do not use RP unless otherwise specified.

We now refer to the MSE plots in Figure 4.1 for further analysis. Putting aside $J^{A3}_{\gamma}(2)$ for large values of γ, the other combinations of k, γ and family type

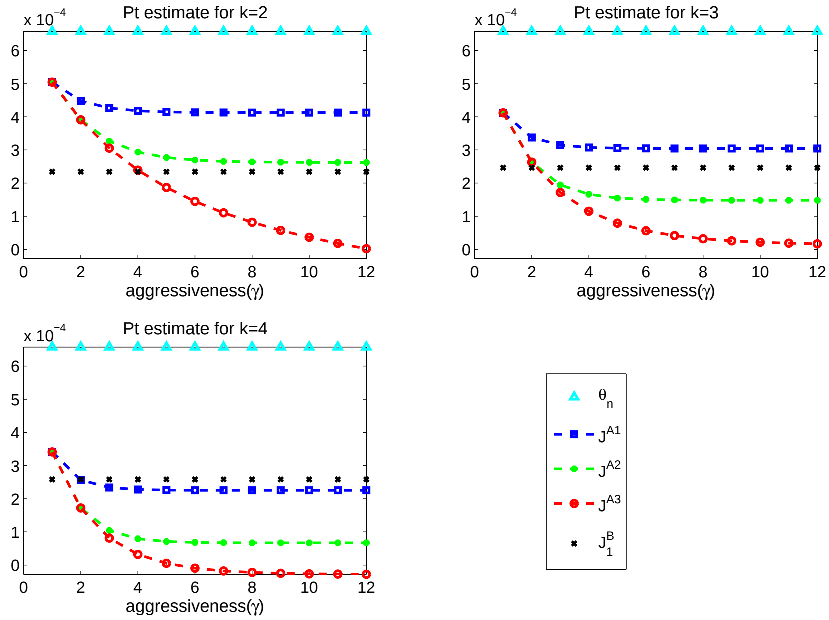

Figure 4.3: Point estimate plots for asset allocation problem without bias correction; $\mu_{\hat{x}} = 1.884 \times 10^{-4}$.

appear well behaved. We can see that the three adaptive jackknife families and the standard jackknife jackknife estimator reduce MSE significantly as compared to the naive estimator $\bar{\theta}_n$.

For any fixed γ, the MSE decreases within all the families as we increase k. This is consistent with our discussion in Section 4.2 suggesting that increasing the partition factor k may reduce bias.

For $k = 2$ and 3, the MSE of J_γ^{A1} and J_γ^{A2} decreases as we increase γ, however for J_γ^{A3}, the MSE first decreases and then increases after $\gamma \approx 2 - 3$. As we increase k to 4, J_γ^{A1} still shows a decrease in MSE as we increase γ, but J_γ^{A2} and J_γ^{A3}'s MSE grows after $\gamma = 2$. We conjecture the cause of this behavior is due to the following: For a sample size of $n = 144$, the smallest stochastic programs we solve (i.e., for forming $\hat{\theta}_{n/k^2}$) have sample sizes of 36, 16 and 9 for $k = 2$, 3 and 4, respectively. Estimators obtained with these smaller sample sizes may deviate from the $O(n^{-p})$

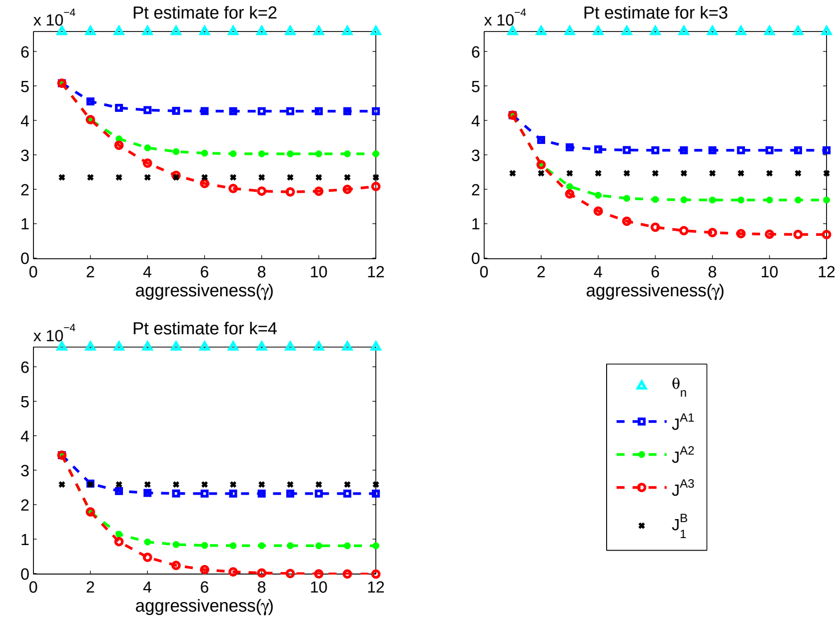

Figure 4.4: Point estimate plots for asset allocation problem with bias correction; $\mu_{\hat{x}} = 1.884 \times 10^{-4}$.

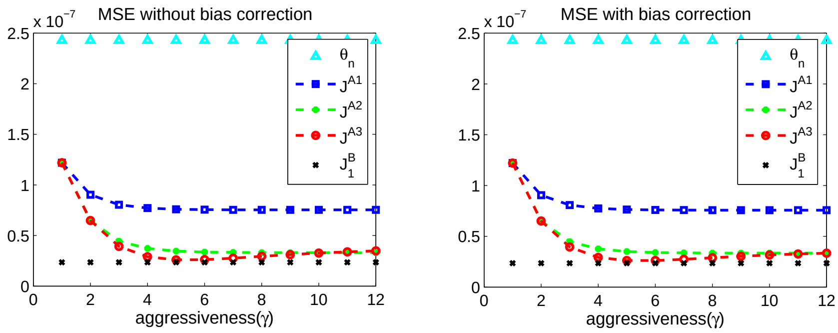

Figure 4.5: MSE plots for asset allocation model with RP and $k = 2$.

nature of bias, and this may explain the increase of MSE for higher values of γ and $k = 4$ for J_γ^{A2} and J_γ^{A3}. To obtain the asymptotic results we presented in (3.9) and (4.3) one should have a large enough sample size, n such that the $O(n^{-p})$ form of the bias approximately holds for problems with sample size n/k^2. The effect of smaller sample sizes is amplified when we increase the aggressiveness, either by increasing γ in a family or moving from family 1 to family 2 to family 3. We now increase n to 576 so that when forming $\hat{\theta}_{n/k^2}$ we have sample sizes of 144, 64 and 36 for $k = 2$, 3 and 4, respectively. Figure 4.6 has the corressponding MSE plots. We see that for $k = 3$ and 4, as we increase γ, J_γ^{A2} and J_γ^{A3} have relatively less increase in MSE, compared to that of Figure 4.1.

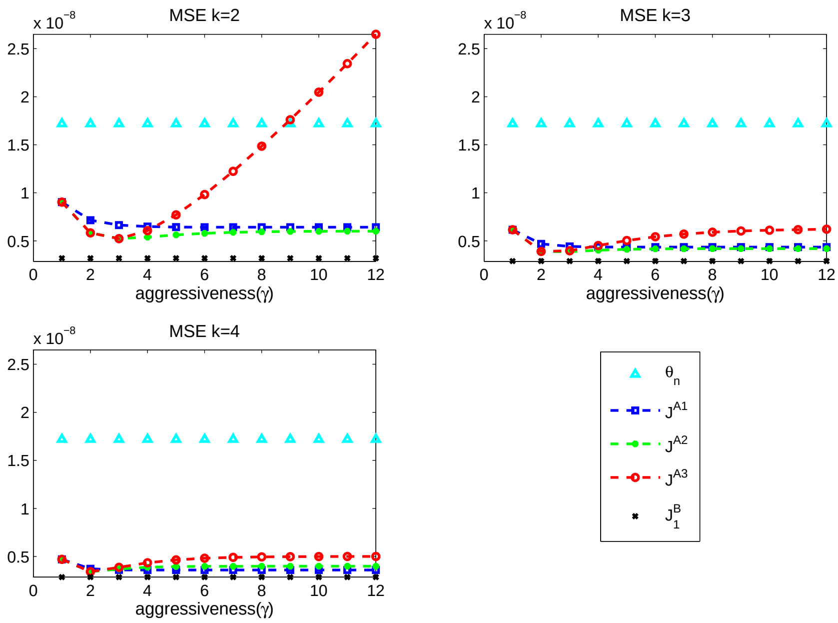

Figure 4.6: MSE plots without bias correction for asset allocation model with $n = 576$.

Though J_γ^{A1} is expected to be the least aggressive of three adaptive families, it seems to work well, at least when k is large. $J_\gamma^{A1}(4)$ has MSE similar to that of J_1^B. As mentioned before, p for the asset allocation model is close to 1, and hence the standard jackknife estimator with $q = 1$ eliminates most of the bias. We achieve similar results with $J_\gamma^{A1}(4)$ without any *a priori* assumption on p.

Figure 4.7 provides MSE$^-$ plots for asset allocation model. We see that more aggressiveness in the adaptive estimators leads to an increase in MSE$^-$, as expected.

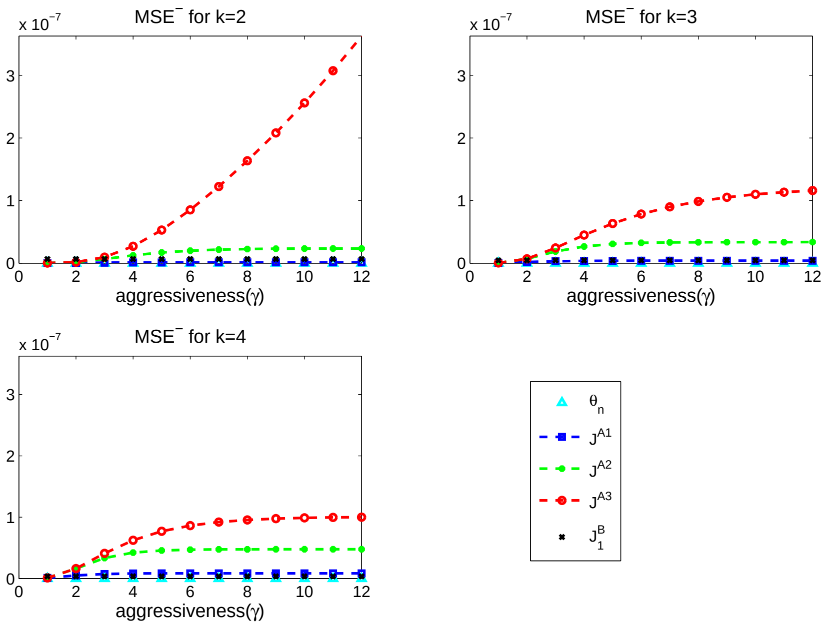

Figure 4.7: MSE^{-} plots for asset allocation model.

$J_{\gamma}^{A1}(k)$ has MSE^{-} comparable to that of J_1^B for all values of k. Figure 4.8 contains the Pr plots for asset allocation model. As we increase the aggressiveness of bias reduction, either by increasing k or increasing γ, Pr increases for all the families. Figure 4.9 contains CI width plots. If we compare the point estimate plots in Figure 4.3 with that of the CI width plots, we find that the biggest contributor to the CI width is sampling error for families 2 and 3.

Figures 4.10 and 4.11 contain the Schruben coverage plots with sample sizes, $n = 144$ and 576, respectively. We compare Schruben coverage provided by estimators: $\bar{\theta}_n$, $J_1^{A1}(k)$ $(= J_1^{A2}(k) = J_1^{A3}(k))$, $J_2^{A2}(k)$ $(= J_2^{A3}(k))$, $J_{12}^{A1}(k)$, $J_{12}^{A2}(k)$, $J_{12}^{A3}(k)$ and J_1^B. For the plots with $n = 144$, the coverage becomes tighter as we move from family 1 to family 2 to family 3. However, for $k = 3$ and $k = 4$, family 2 and 3 with high values of γ ($\gamma = 12$) tend to give undercoverage. A reason for this might be that the estimators $\hat{\theta}_{n/k^2}$ for $k = 3$ or 4 may not have bias which is of the form

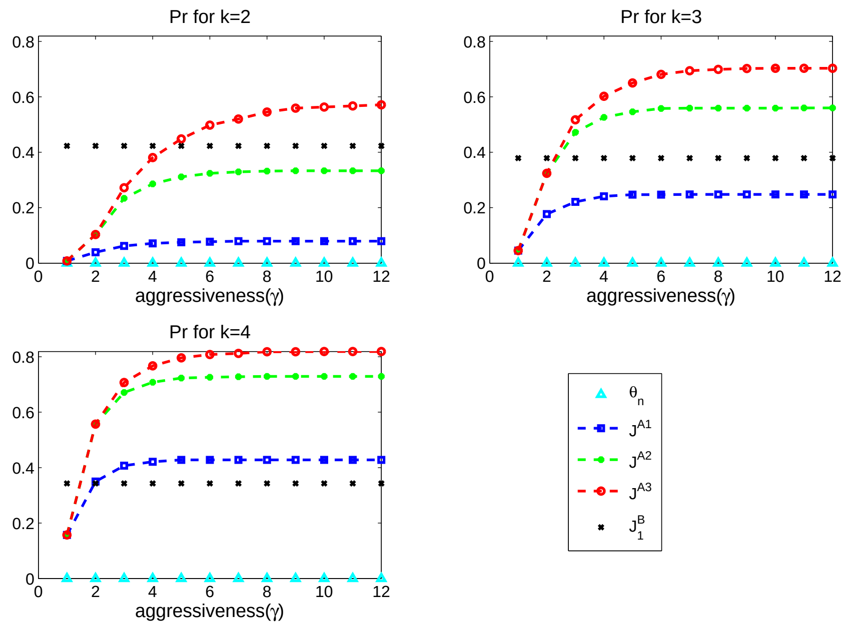

Figure 4.8: Pr plots for asset allocation model.

$O(n^{-p})$ because of very small sample sizes involved. This undercoverage effect is considerably less pronounced when we increase the sample size to 576. Note that the coverage provided by $J_2^{A2}(k)$ is similar to but a bit tighter than that of $J_{12}^{A1}(k)$, which agrees with our discussion in Section 4.1.

4.4.2 Example 1

In this section we again consider Example 1 from Section 1.3 with $p = 2$. Example 1 exhibits bias of the form $O(n^{-p})$ for any value of n. Figures 4.12-4.16 contain the MSE, MSE$^-$, Pr, CI width and Schruben coverage plots, respectively, for Example 1.

We see that for $k = 2$ and large values of γ, the MSE of $J_\gamma^{A3}(2)$ is exploding. The plots of $J_\gamma^{A2}(k)$ and $J_\gamma^{A3}(k)$ for $k = 3$ and 4 are well-behaved. This is consistent with our observation in the previous section. We can see that other than the com-

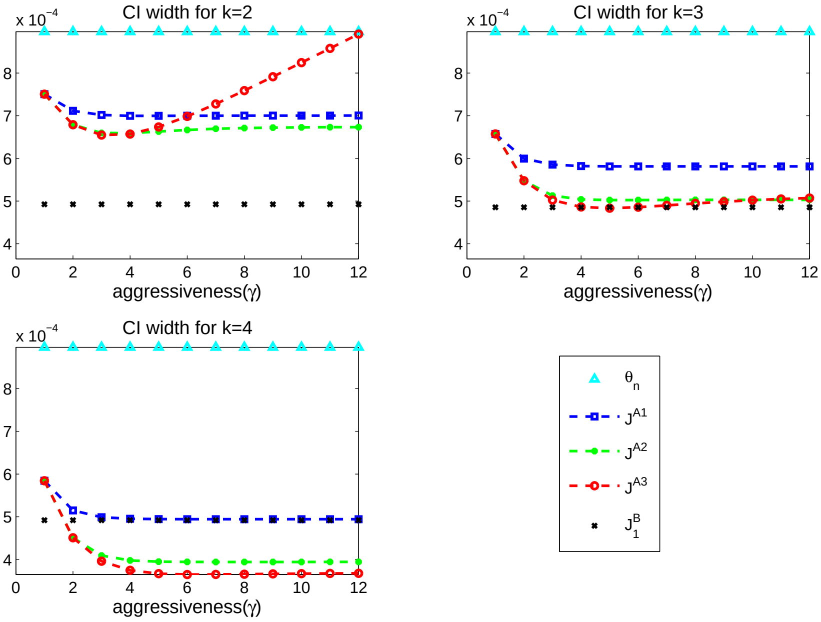

Figure 4.9: CI width plots for asset allocation model with $\mu_{\hat{x}} = 1.884 \times 10^{-4}$.

bination of $k = 2$, large γ and family 3, all other combinations of k, γ and adaptive family type reduce the MSE without much increase in MSE$^-$ (see Figure 4.13) or Pr (see Figure 4.14). The Schruben coverage plots for the adaptive estimators are much tighter than that of the naive estimator. The adaptive estimators do not provide undercoverage in any case, but J_1^B gives significant undercoverage in all the cases. Using RP again circumvents the undesirable behavior of family 3 adaptive estimators with large γ and $k = 2$, which is consistent with our conjecture in Section 4.4.1.

4.4.3 Incorporating variance reduction techniques

So far we have formed the underlying estimator $\hat{\theta}_n$ using i.i.d. sampling. When $\hat{\theta}_n = G_n(\hat{x}) = \frac{1}{n}\sum_{i=1}^n f(\hat{x}, \tilde{\xi}^i) - \min_{x \in X} \frac{1}{n}\sum_{i=1}^n f(x, \tilde{\xi}^i)$, we have used common random number streams, $\tilde{\xi}^1, \ldots, \tilde{\xi}^n$, to reduce variance, but that stream has been i.i.d. In

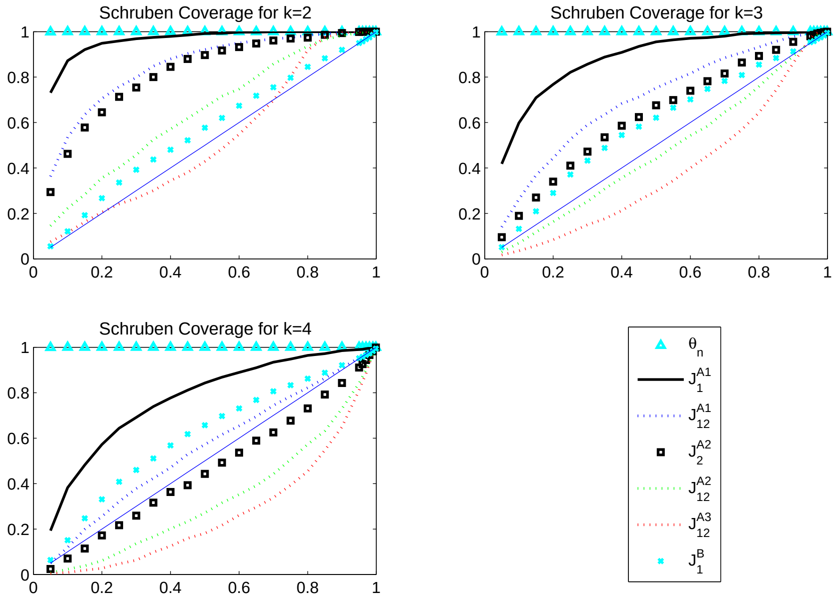

Figure 4.10: Schruben coverage plots for asset allocation model.

this section we discuss how we can apply other sampling strategies in selecting $\tilde{\xi}^1, \dots, \tilde{\xi}^n$ within the context of our adaptive jackknife estimation procedures. We begin with a more general discussion and then apply Latin hypercube sampling in particular.

Suppose we select $\tilde{\xi}^1, \dots, \tilde{\xi}^n$ in a non-i.i.d. manner in an attempt to reduce the variance of $\hat{\theta}_n$. We choose the partition factor $k = 2$ to simplify the discussion. In order to form adaptive jackknife estimators we partition the original sample, with index set N, into $k^2 = 4$ subsamples indexed by N^j with $|N^j| = n/4$, $j = 1, \dots, 4$. With these subsamples we can form $\hat{\theta}_{n/4}$, $j = 1, \dots, 4$ and under RP the estimators $\hat{\theta}^{ij}_{n/2}$, $i < j$, $i, j = 1, \dots, 4$. In this setting we require the following:

(i) Each estimator $\hat{\theta}^{j}_{n/4}$, $\hat{\theta}^{ij}_{n/2}$ and $\hat{\theta}_n$ is strongly consistent as $n \to \infty$;

(ii) Each estimator $\hat{\theta}^{j}_{n/4}$, $\hat{\theta}^{ij}_{n/2}$ and $\hat{\theta}_n$ should have bias satisfying (2.4), at least within the addition of $o(n^{-p})$ terms; and,

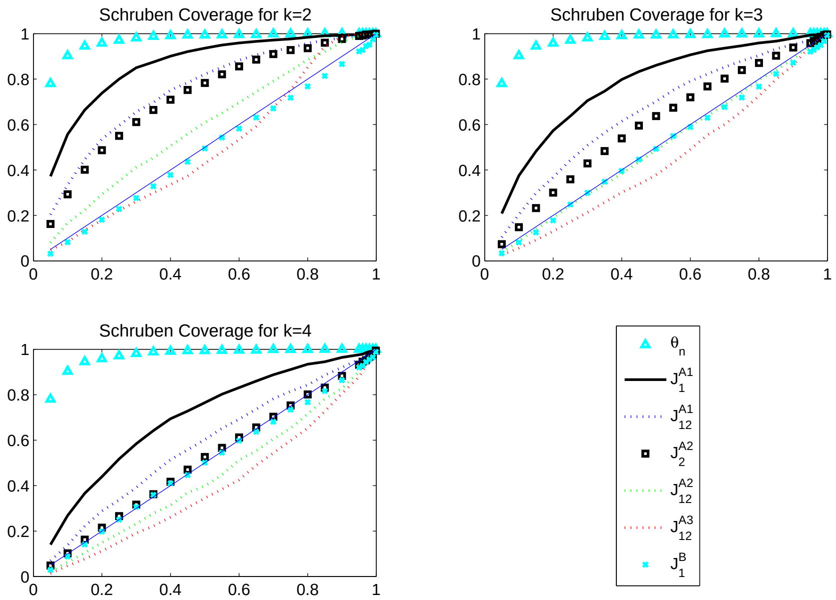

Figure 4.11: Schruben coverage plots for asset allocation model with increase sample size, $n = 576$.

(iii) (A4) holds, in turn, with $N' = N$ and $\bar{N}' = N^i \cup N^j$, $i < j$, $i, j = 1, \ldots, 4$, and with $N' = N^i \cup N^j$ and $\bar{N}' = N^i$, $i < j$, $i, j = 1, \ldots, 4$.

The non-i.i.d. sampling scheme should satisfy (i)-(iii) and, of course, should reduce variance.

We now apply Latin hypercude sampling (LHS) to a two-stage stochastic program. LHS (essentially) requires that the components of $\tilde{\xi}$ be independent. LHS carries out stratified sampling on each of the components of $\tilde{\xi}$ separately and then randomly shuffles their order when combining to form observations of the vector $\tilde{\xi}$. If we were to employ LHS to form $\tilde{\xi}^1, \ldots, \tilde{\xi}^n$ and then construct subsamples of size $n/4$ then, in general, we could not expect $\hat{\theta}_{n/4}$ and $\hat{\theta}_{n/2}$ to even satisfy the consistency condition (i) above. Instead we apply LHS four times to form four independent subsamples indexed by N^j with $|N^j| = n/4$, $j = 1, \ldots, 4$. Then, under mild conditions on the underlying stochastic program (see Chapter 1) consistency

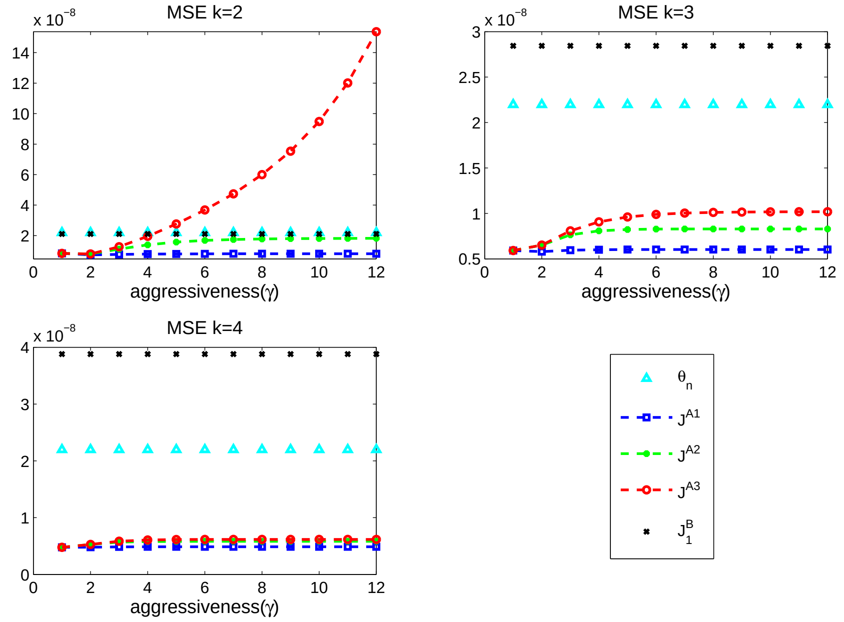

Figure 4.12: MSE plots for Example 1.

holds as required by (i). Condition (1.5) holds for each of the subsamples N^j, $j = 1, \ldots, 4$, $N^i \cup N^j$, $i < j$, $i, j = 1, \ldots, 4$, and $N = N^1 \cup \ldots \cup N^4$ and hence the variant of (A4) described above in (iii) holds.

We now apply LHS to PGP2. This test problem has a three-dimensional random vector $\tilde{\xi}$ with independent components, and from Table 3.1 we see that, percentage-wise, PGP2 has the largest contribution of sampling error to its CI width for the naive estimator $\bar{\theta}_n$. For the same reasons discussed in Section 4.4.1 for asset allocation model, we choose $n = 576$. Figures 4.17-4.21 contain the MSE, MSE$^-$, Pr, CI width and Schruben coverage plots, respectively, for PGP2 with LHS. We performed Latin hypercube sampling on each successive sample of size $576/k^2$ for $k = 2, 3, 4$.

Figures 5.6-5.10 in Appendix contain the MSE, MSE$^-$, Pr, CI width and Schruben coverage plots, respectively, for PGP2 without LHS and sample size $n =$

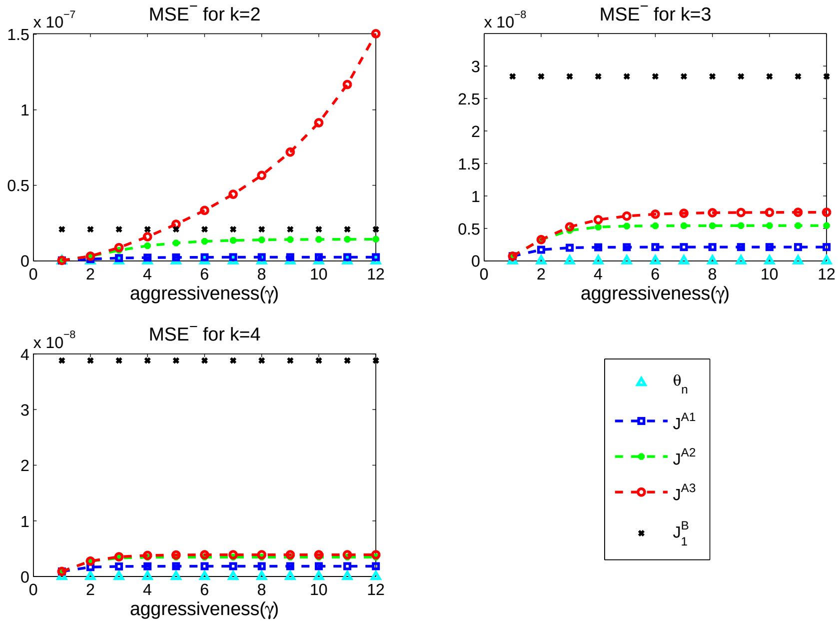

Figure 4.13: MSE$^-$ plots for Example 1.

576. From figures 4.17 and 5.6 we see that MSE of the underlying estimator $\bar{\theta}_n$ decreased from 0.8 to 0.7 after using LHS for variance reduction. We see that not only the underlying estimator, but the adaptive estimators too get a benefit in MSE by applying LHS. We see almost no change in Schruben coverage plots of all the estimators after applying LHS. The behavior of adaptive estimators with or without use of LHS is similar to the asset allocation model and Example 1. We can see that it may be possible to successfully combine variance reduction techniques and adaptive jackknife estimators.

4.5 Recommendations and summary

In this chapter we introduced two new families of adaptive jackknife estimators. If the sample size n and γ are chosen wisely, then adaptive families 2 and 3 may provide increased bias reduction over family 1. That said, we observed that family

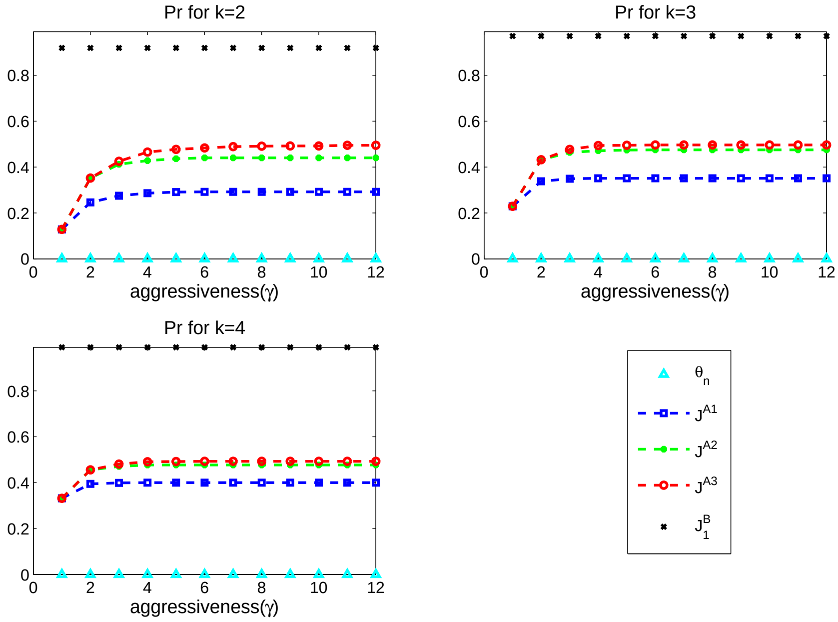

Figure 4.14: Pr plots for Example 1.

1 is a conservative choice that works well in almost all cases. We also introduced the concept of generalized partition factor, k. Larger values of the partition factor k may lead to more aggressive bias reduction, but we must ensure the sample size n is sufficiently large. Use of the rotation policy, RP, may help stabilize the adaptive estimators. We also investigated how variance reduction techniques can be incorporated in our adaptive procedure. Our empirical results suggest that our adaptive estimators can significantly reduce bias withour any *a priori* knowledge of the order of the bias, p. Moreover, the aggressiveness parameter γ, the partition factor k and three adaptive families provide flexibility in formulating a bias reduction strategy.

In this chapter we analyzed three test problems: The asset allocation model, Example 1, and PGP2. Chapter 3 include four additional test problems: Bermudan option pricing, APL1P, CEP1. The three problems we considered here are representative of the other results. Appendix A includes results for all of the test problems

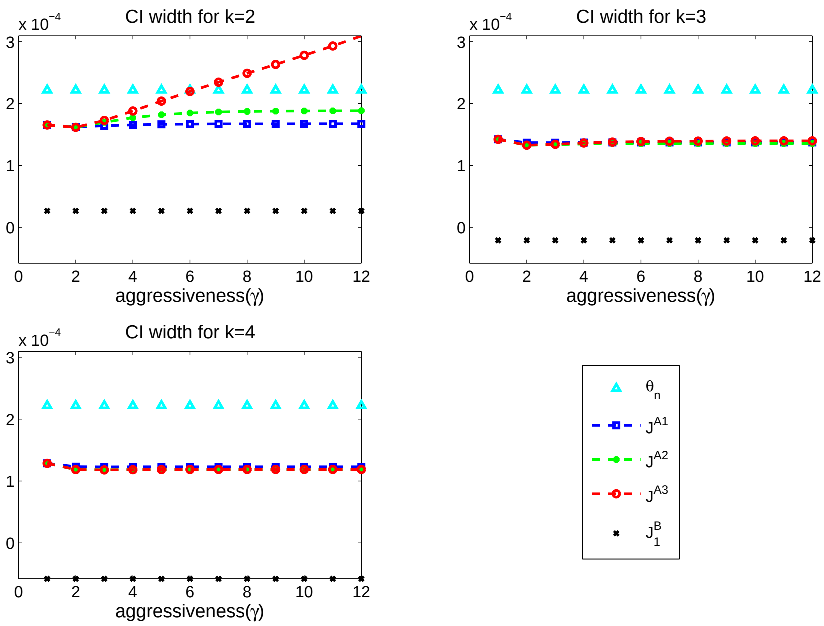

Figure 4.15: CI width plots for Example 1 with $\hat{\mu}_x = 0$.

except for CEP1. (For CEP1 even the least aggressive family could not be applied and so we do not apply the more aggressive estimators of family 2 and 3.) We next provide some recommendations for using adaptive estimators.

- Family 1 of the adaptive estimators is the least aggressive but appears to perform reliably for all values of k and even with smaller sample sizes, n. When p is near 1, the standard jackknife estimator J_1^B works well. Family 1 with $k = 4$ seems to work as well as standard jackknife, J_1^B when p is near 1.

- Family 2 seems to work well for all values of γ when $k = 2$ or 3. However, if $k = 4$, we recommend using values of γ close to 1.

- Family 3 with $k = 3$ and 4 significantly reduces MSE, but one has to be careful that n is large enough so that the estimators with sample size n/k^2 satisfy the required $O(n^{-p})$ form of the bias. If using family 3: a) Value of γ should be

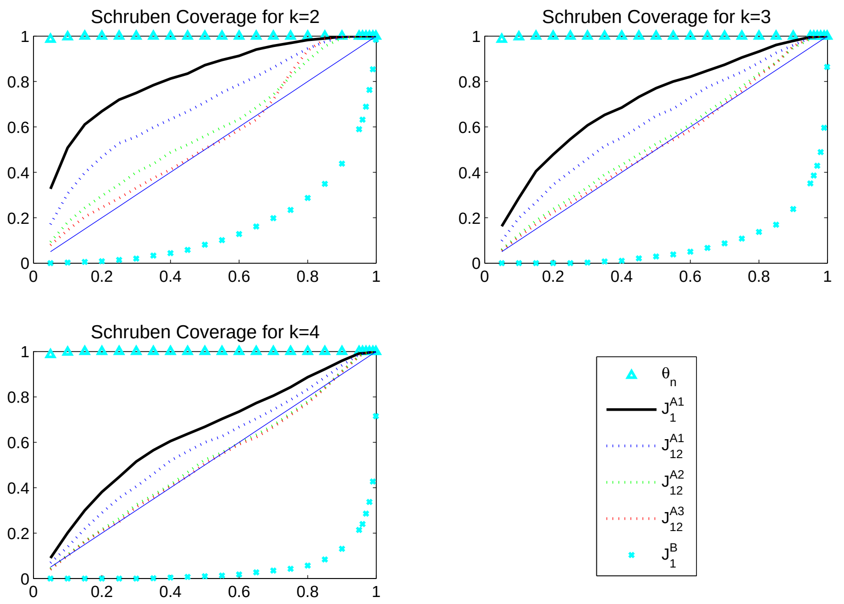

Figure 4.16: Schruben coverage plots for Example 1.

chosen close to 1, and b) k should be chosen as 3 or 4.

- Use of the rotation policy, RP stabilizes the adaptive estimators, especially for the combination of $k = 2$ and family 3.

- For almost complete bias reduction, we recommend using the combination of family 3, values of γ close to 4 or 5, RP and $k = 4$.

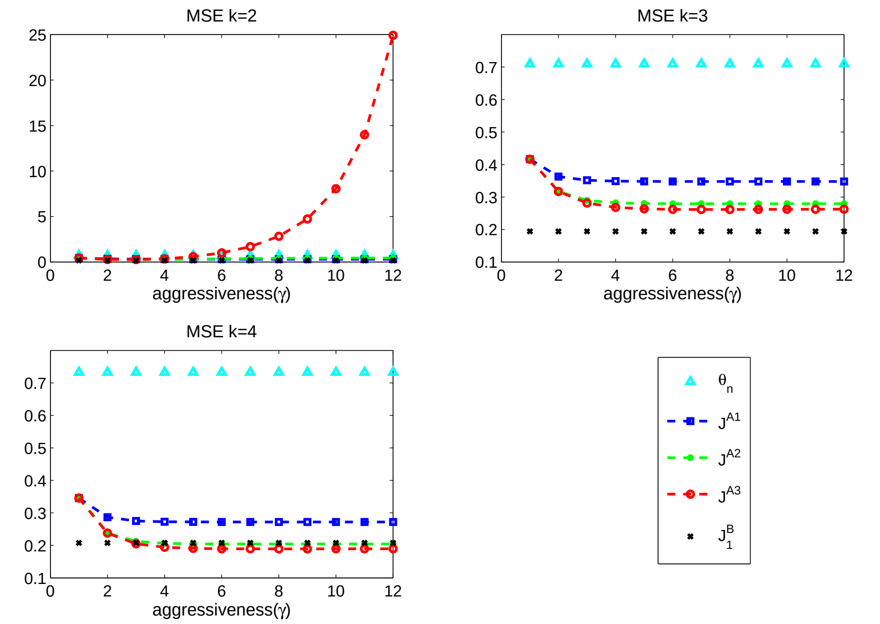

Figure 4.17: MSE plots for PGP2 with Latin hypercube sampling.

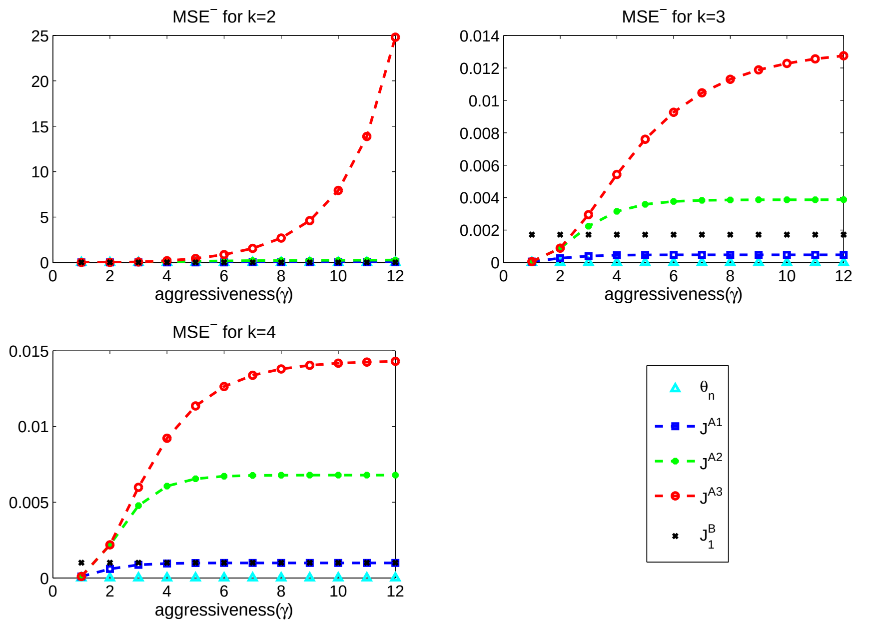

Figure 4.18: MSE^{-} plots for PGP2 with Latin hypercube sampling.

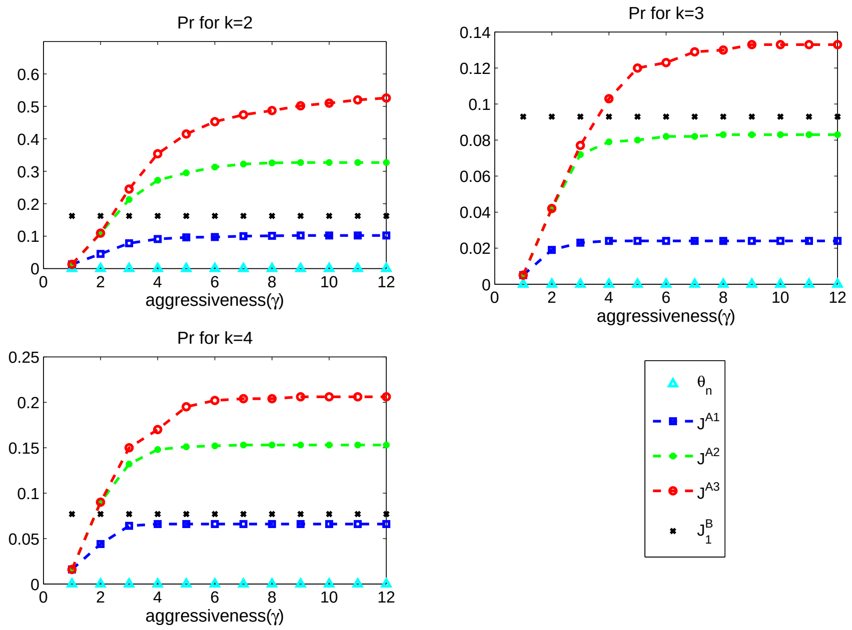

Figure 4.19: Pr plots for PGP2 with Latin hypercube sampling.

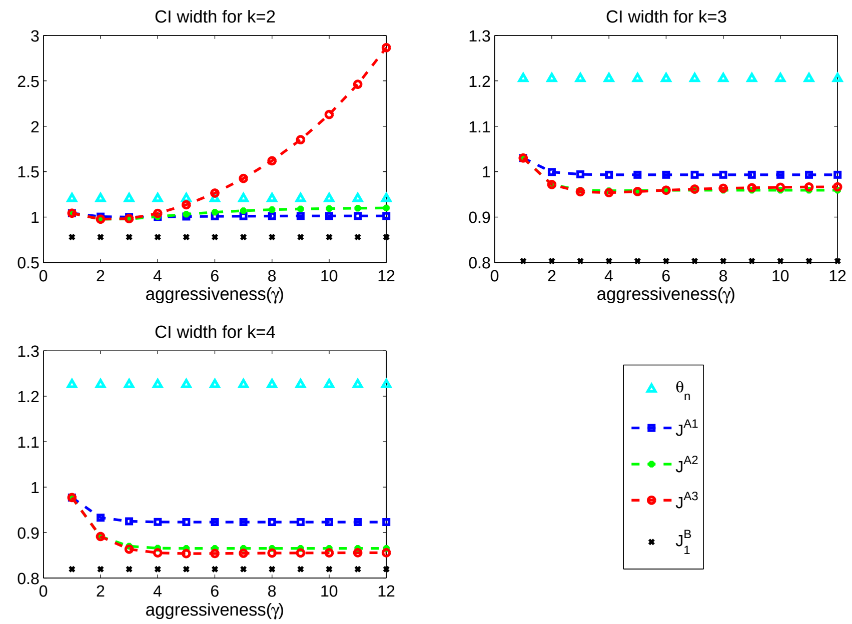

Figure 4.20: CI width plots for PGP2 with Latin hypercube sampling with $\hat{\mu}_x = 0$.

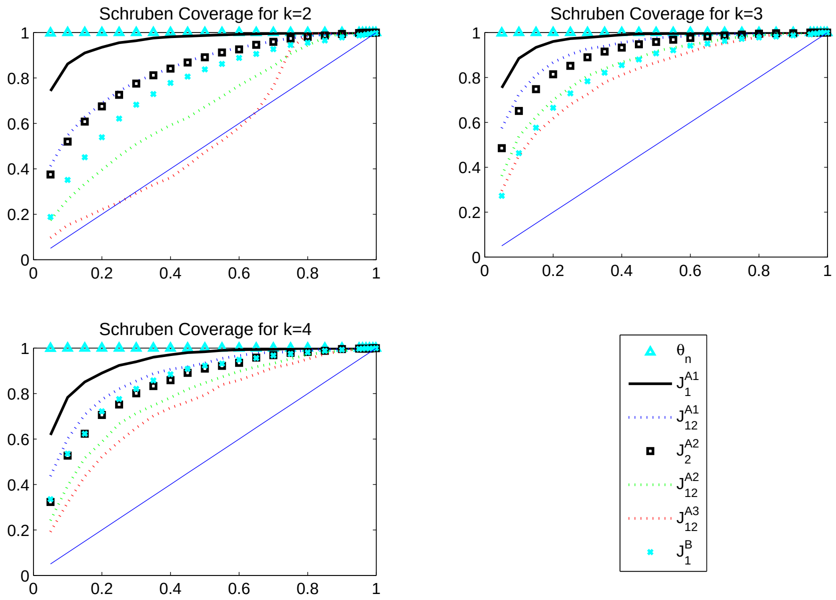

Figure 4.21: Schruben coverage plots for PGP2 with Latin hypercube sampling.

Chapter 5

Conclusions and Future Research

5.1 Conclusions

Many important problems involve decision making under uncertainty. Stochastic programming is a powerful tool for modeling these problems. Usually, it is impossible to solve a stochastic program exactly. Sampling-based approximations provide an attractive approach to estimate the true optimal solution and value of a stochastic program. However, the estimates of the latter are biased under quite general conditions. A point estimate or a confidence interval estimator of the optimality gap or the optimal value can be obtained by using a multiple replication procedure. To assess the quality of an available candidate solution or estimate the true optimal value, it is desirable to have the bias as low as possible. This dissertation aimed to address bias reduction procedures in stochastic programming.

In Chapter 2 we discussed an available technique for bias reduction from the literature known as generalized jackknife estimators. Generalized jackknife estimators work well when the order of the bias is known *a priori*. We discussed how generalized jackknife estimators can be used in stochastic programming to reduce bias in estimators obtained by sampling-based approximations. Unfortunately, prior information about the order of the bias is unavailable for most problems. Example 1 from Chapter 2 showed that bias can shrink to zero as $O(n^{-p})$ where p can be anything from $1/2$ to ∞. Incorrect estimates for the order of the bias may lead to over-correction when using generalized jackknife estimators. We then presented a p-estimation procedure in which the order of the bias, p, is estimated adaptively prior to applying a generalized jackknife estimator.

When estimating an optimality gap or optimal value of a stochastic program we prefer to err on the conservative side, i.e., prevent over-correction. This motivated our development of adaptive jackknife estimators, which do not require *a priori* knowledge of the order of the bias. In Chapter 3 we started with development of our most basic adaptive jackknife estimator. We then presented results regarding consistency of this estimator and also argued that for a "well-defined" problem, the adaptive estimators eliminate the possibility of over-correction. We extended the development from a single adaptive jackknife estimator to a family of adaptive estimators, parameterized by a positive integer γ. The members of the adaptive family provide more aggressive bias reduction as γ grows.

In Chapter 4, we presented two more families of adaptive estimators. The aggressiveness of bias reduction grows as we move from family 1 to family 2 to family 3. We also argued that the limiting member of family 3 may completely remove the leading term in the bias in the underlying estimator. We introduced the concept of a generalized partition factor, k, for adaptive jackknife estimators. Larger values of k provide more aggressive bias reduction. We also introduced the concept of a rotation policy, which may help reduce the variance in adaptive estimators, which in turn provides smoother bias reduction using adaptive estimators. Finally, we showed that the adaptive estimators can be combined with variance reduction techniques, i.e., the bias reduction can be achieved on top of variance reduction techniques.

At the end of each of the Chapters 2-4, we compared the performance of the adaptive estimators with the traditional generalized jackknife estimator and the underlying estimator. We used several performance measures including MSE, MSE^-, CI widths and Schruben coverage plots on a variety of problems. The results were encouraging and supported our development of adaptive estimators. In Chapter 4, we identified the cases when adaptive estimators might fail and provided remedies.

Finally, at the end of Chapter 4, we provided some recommendations on the parameter settings for the adaptive estimators. The numerical results we obtained were consistent with our recommendations. To summarize, in this dissertation we developed the concept of adaptive jackknife estimators for bias reduction. The adaptive estimators do not require *a priori* knowledge about the order of the bias in the underlying estimators. The adaptive estimators that we developed are parameterized by family type, partition factor k, and aggressiveness parameter γ. These parameters give flexibility in designing a bias reduction procedure.

5.2 Future research

The concept of adaptive bias-reducing estimators opens a new line of research. The key to use of adaptive estimators is satisfying assumption (A4), or its variants, which ensures a form of convexity of the estimators in the sample size. The use of adaptive jackknife estimators may be applied to areas other than stochastic programming. One future research direction is to identify problems where we can guarantee variants of (A4) and apply the adaptive jackknife estimators. An example of such problem is when the estimator is a convex function of sample means. We can show that in such a case (A4) is satisfied.

We also plan to compare the performance of adaptive jackknife estimators with other bias reduction strategies, e.g., bootstrap and the Taylor series approach. Using the adaptive jackknife estimator requires estimation of a group of underlying estimators on a sample and its subsets. The computational efficiency of the adaptive procedure can be significantly improved with intelligent re-estimation schemes. In the context of two-stage stochastic programs, these re-estimations can be done efficiently using a variant of Benders' decomposition scheme, and this is one promising future research direction.

Another challenge concerns indentifying when bias shrinks to zero more quickly than $O(n^{-p})$ because in such cases even our adaptive estimators are too aggressive. One might argue that in such cases there is limited need for reducing bias. Still, a systematic way to identify such situations would be valuable. In Section 3.6.5 we pointed to one possible approach to this problem.

Appendix

In this Appendix we provide the numerical results for the test problems APL1P, PGP2, and pricing Bermudan call option. We present numerical results to compare bias reduction capabilities of $\bar{\theta}_n$, J_1^B, $J_\gamma^{A1}(k)$, $J_\gamma^{A2}(k)$, and J_γ^{A3} for $k = 2, 3, 4$. We present results which are based on an average of 1000 repititions of the estimators. We choose number of replications, $m = 30$, and $\alpha = 0.05$, i.e., we form approximate 95% confidence intervals. For use sample size, $n = 144, 576, 144$ for APL1P, PGP2, and pricing Bermudan call option, respectively. The numerical results reported without bias correction (see equation (3.12)), without RP and without LHS.

Figures 5.1-5.5, 5.6-5.10, and 5.11-5.15 contain MSE, MSE$^-$, Pr, CI width and Schruben coverage plots for APL1P, PGP2 and pricing Bermudan call option, respectively. We can see that numerical results presented in Figures 5.1-5.15 are consistent with what we presented in Section 4.4, and hence are not discussed further.

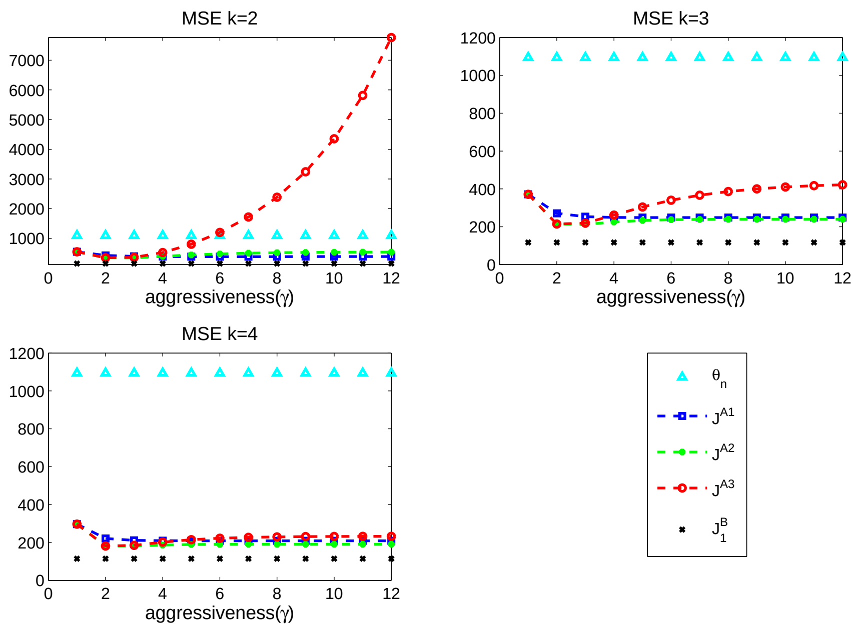

Figure 5.1: MSE plots for APL1P.

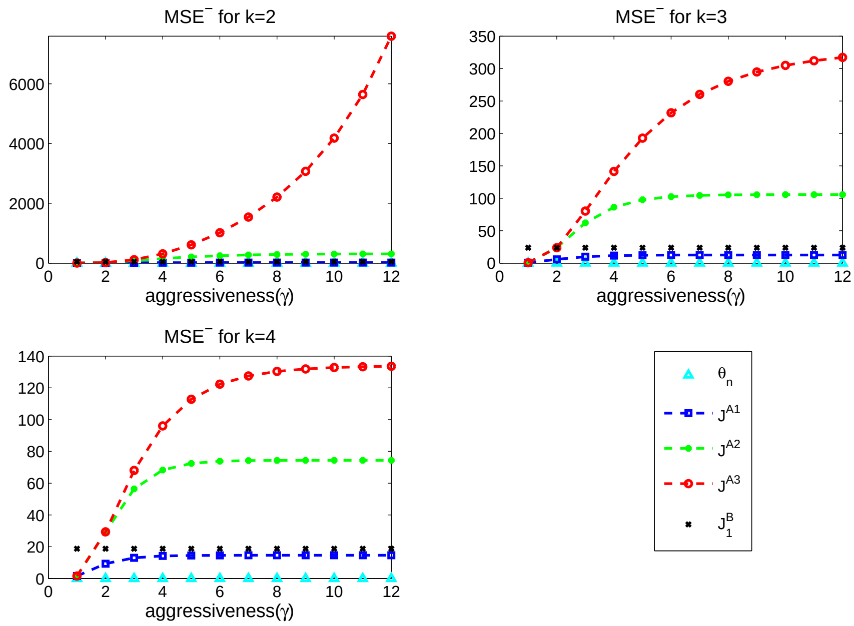

Figure 5.2: MSE^{-} plots for APL1P.

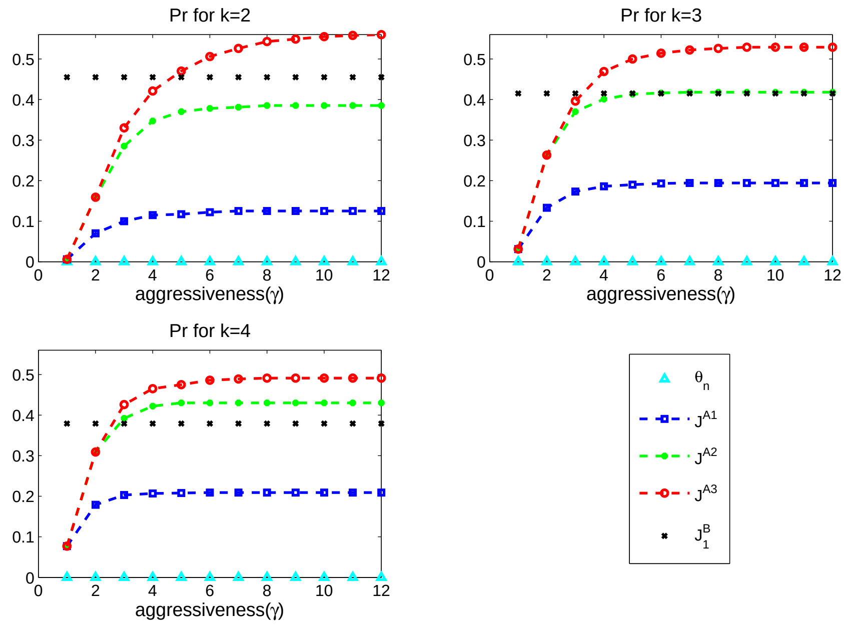

Figure 5.3: Pr plots for APL1P.

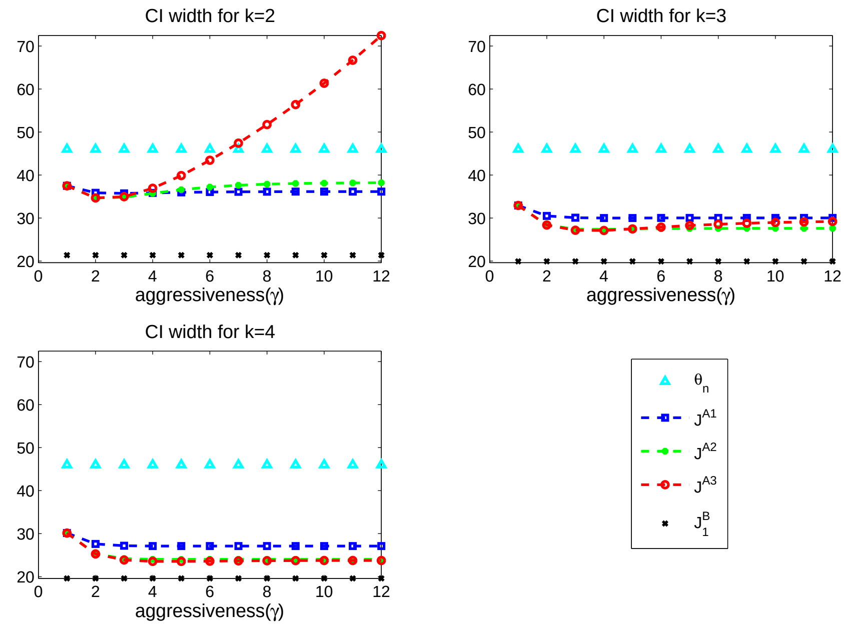

Figure 5.4: CI width plots for APL1P with $\hat{\mu}_x = 0$.

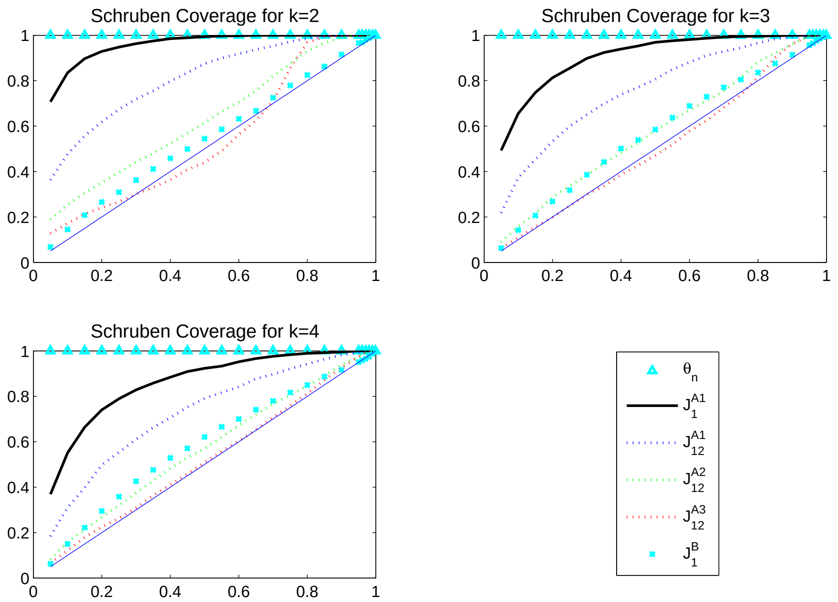

Figure 5.5: Schruben coverage plots for APL1P.

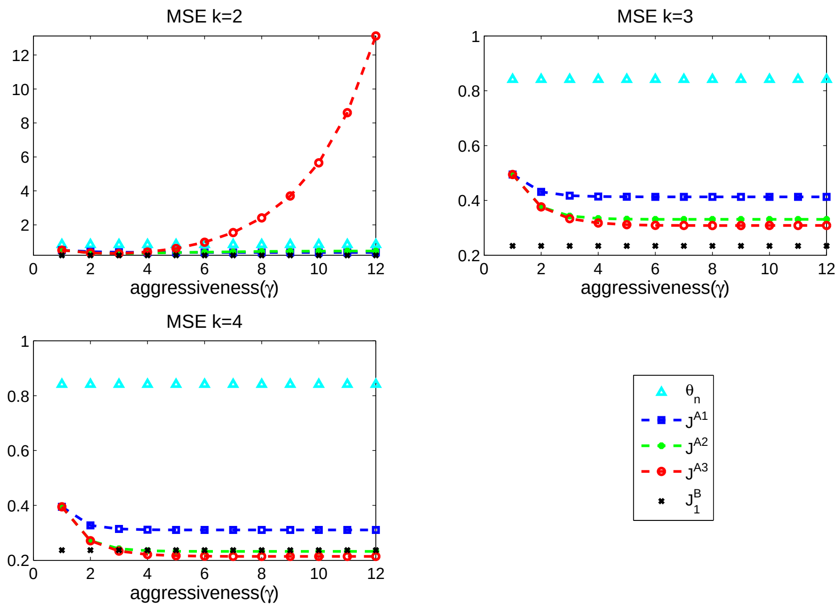

Figure 5.6: MSE plots for PGP2.

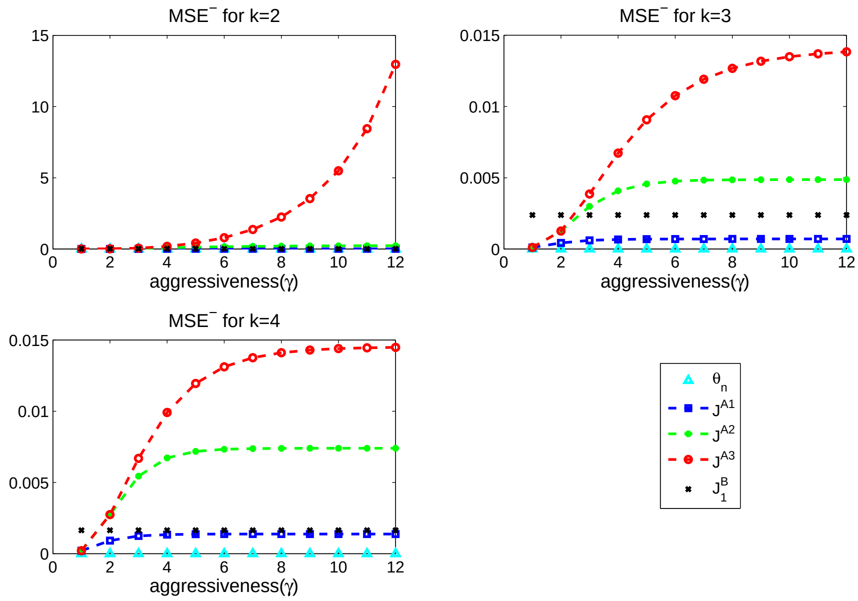

Figure 5.7: MSE^- plots for PGP2.

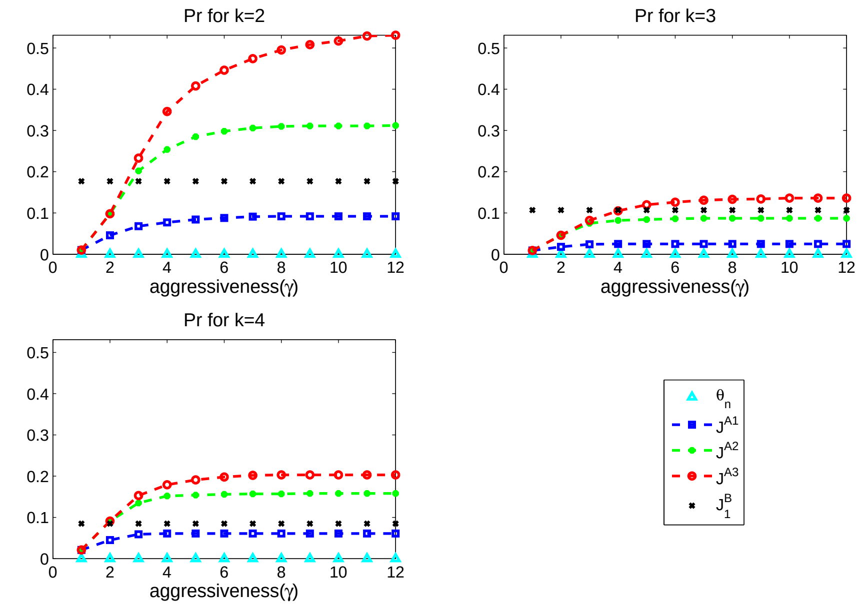

Figure 5.8: Pr plots for PGP2.

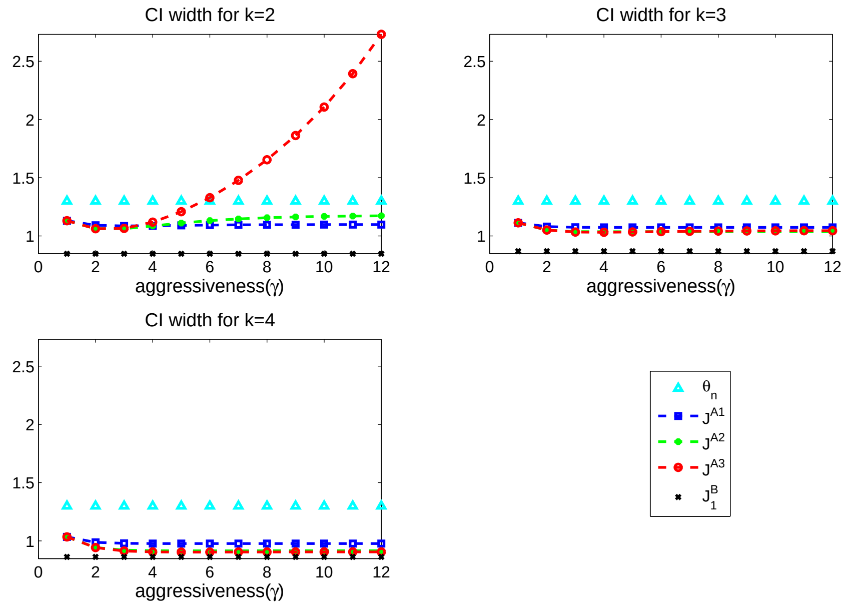

Figure 5.9: CI width plots for PGP2 with $\hat{\mu}_x = 0$.

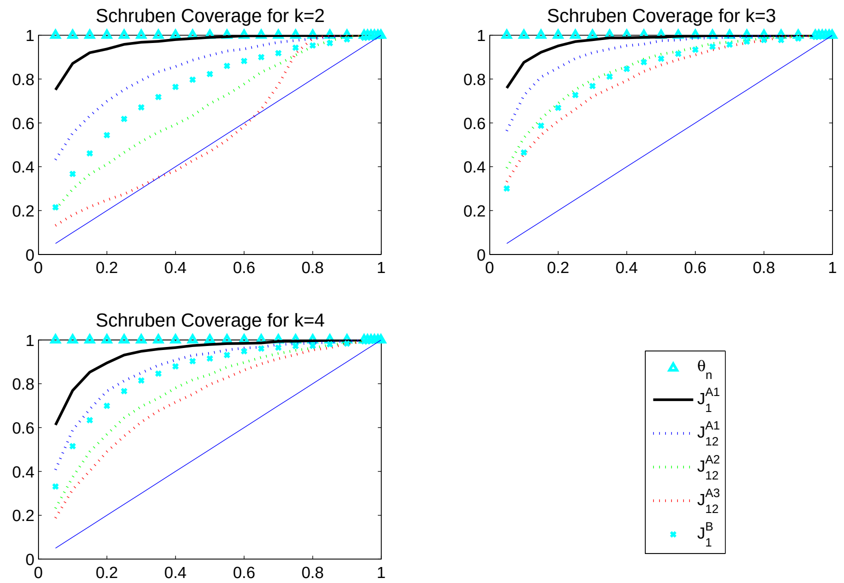

Figure 5.10: Schruben coverage plots for PGP2.

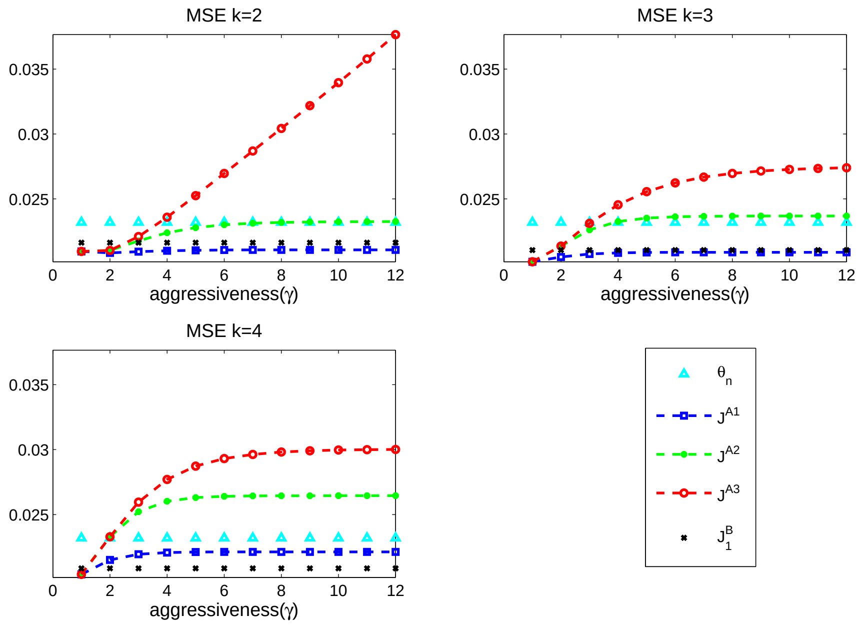

Figure 5.11: MSE plots for pricing Bermudan call option.

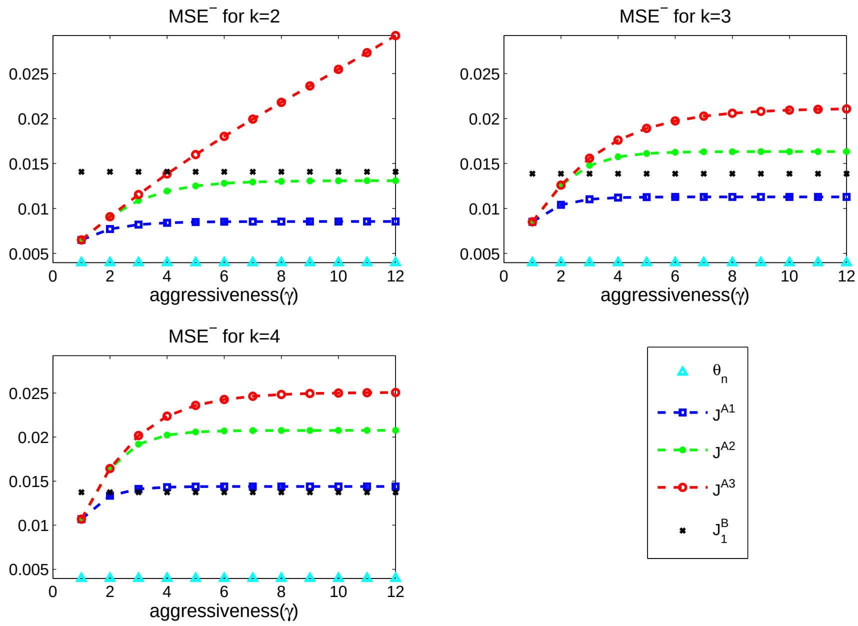

Figure 5.12: MSE^{-} plots for pricing Bermudan call option.

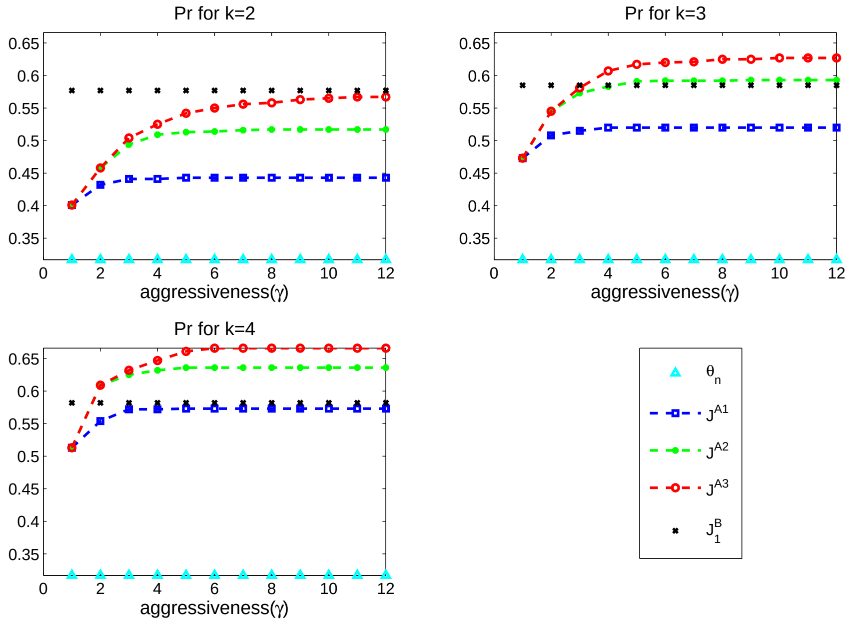

Figure 5.13: Pr plots for pricing Bermudan call option.

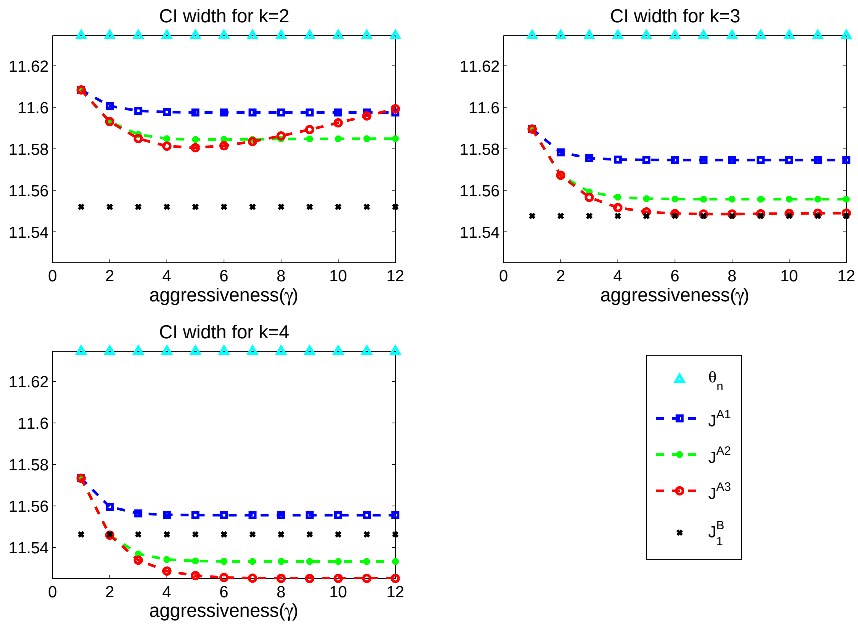

Figure 5.14: CI width plots for pricing Bermudan call option with optimal price = 11.341.

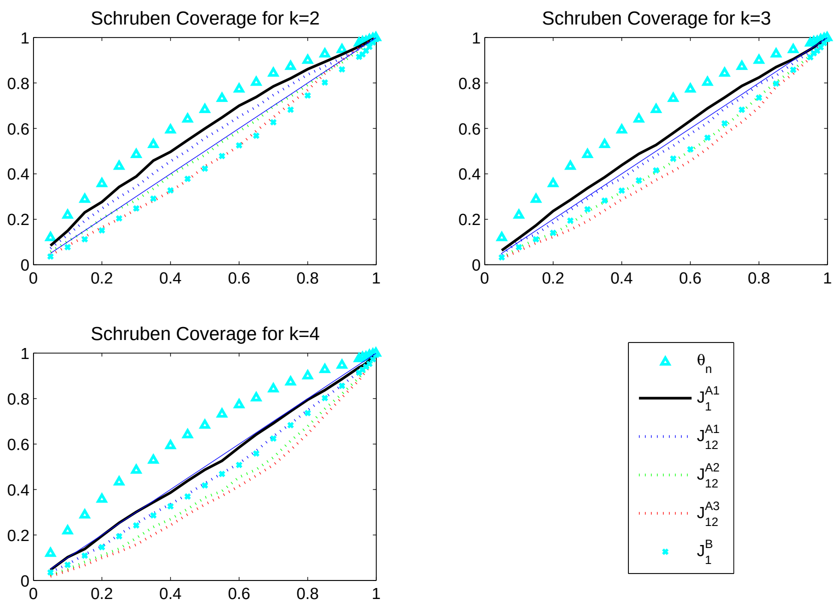

Figure 5.15: Schruben coverage plots for pricing Bermudan call option.

Bibliography

[1] S. Anderson, A. Auquier, W.W. Hauck, D. Oakes, W. Vandaele, and H.I. Weisberg. *Statistical Methods for Comparative Studies: Techniques for Bias Reduction*. John Wiley & Sons, New York, 1980.

[2] T.G. Bailey, P.A. Jensen, and D.P. Morton. Response surface analysis of two-stage stochastic linear programming with recourse. *Naval Research Logistics*, 46:753–778, 1999.

[3] G. Bayraksan and D.P. Morton. Assessing solution quality in stochastic programs. *Mathematical Programming*, 108:495–514, 2006.

[4] E.M.L. Beale. On minimizing a convex function subject to linear inequalities. *Journal of the Royal Statistical Society*, 17B:173–184, 1955.

[5] J.L. Bienias, P.S. Kott, and D.A. Evans. Application of the delete-a-group jackknife variance estimator to analyses of data from a complex longitudinal survey. *Proceedings of Statistics Canada Symposium*, 2002.

[6] J.R. Birge. Stochastic programming computation and applications. *INFORMS Journal on Computing*, 9:111–133, 1997.

[7] J.R. Birge and F.V. Louveaux. A multicut algorithm for two-stage stochastic linear programs. *European Journal of Operational Research*, 34:384–392, 1988.

[8] L.C. Briand, K.E. Emam, B. Freimut, and O. Laitenberger. Quantitative evaluation of capture-recapture models to control software inspections. *Proceedings of the Eighth International Symposium on Software Reliability Engineering*, pages 234–244, 1997.

[9] M. Broadie and P. Glasserman. Pricing American-style securities using simulation. *Journal of Economic Dynamics and Control*, 21:1323–1352, 1997.

[10] G. Casella and R.L. Berger. *Statistical Inference.* Duxbury Press, Belmont, California, 1990.

[11] A. Chiralaksanakul. *Monte Carlo Methods for Multi-stage Stochastic Programs.* PhD thesis, The University of Texas at Austin, May 2003.

[12] G.B. Dantzig. Linear programming under uncertainty. *Management Science*, 1:197–206, 1955.

[13] G.B. Dantzig and P.W. Glynn. Parallel processors for planning under uncertainty. *Annals of Operations Research*, 22:1–21, 1990.

[14] G.B. Dantzig and G. Infanger. Large-scale stochastic linear programs: importance sampling and Benders' decomposition. In C. Brezinski and U. Kulisch, editors, *Computational and Applied Mathematics I—Algorithms and Theory, Proceedings of the 13th IMCAS World Congress*, pages 111–120, Dublin, Ireland, 1991.

[15] S. Drew and T. Homem-de-Mello. Some large deviations results for Latin hypercube sampling. In *Proceedings of the 2005 Winter Simulation Conference*, pages 673–681, 2005.

[16] S. Drew and T. Homem-de-Mello. Quasi-Monte Carlo strategies for stochastic optimization. In *Proceedings of the 2006 Winter Simulation Conference*, pages 774–782, 2006.

[17] G.T. Duncan. An empirical study of jackknife-constructed confidence regions in nonlinear regression. *Technometrics*, 20:123–129, 1978.

[18] J. Dupačová and R.J.-B. Wets. Asymptotic behavior of statistical estimators and of optimal solutions of stochastic optimization problems. *The Annals of Statistics*, 16:1517–1549, 1988.

[19] J. Durbin. Sampling theory for estimates based on fewer individuals than the number of selected. *Bulletin of the International Statistical Institute*, 36:113–119, 1958.

[20] J. Durbin. A note on the application of Quenouille's method of bias reduction to the estimation of ratios. *Biometrika*, 46:477–480, 1959.

[21] M.B. Freimer, D.J. Thomas, and J.T. Linderoth. Reducing bias in stochastic linear programs with sampling methods. Technical report, Industrial and Systems Engineering, Lehigh University, 2005.

[22] H.L. Gray and W.R. Schucany. *The Generalized Jackknife Statistic*. Marcel Dekker, Inc., New York, 1972.

[23] J. Hahn and W. Newey. Jackknife and analytical bias reduction for nonlinear panel models. *Econometrica*, 72:1295–1319, 2004.

[24] J.L. Higle. Variance reduction and objective function evaluation in stochastic linear programs. *INFORMS Journal on Computing*, 10:236–247, 1998.

[25] J.L. Higle and S. Sen. Stochastic decomposition: An algorithm for two-stage linear programs with recourse. *Mathematics of Operations Research*, 16:650–669, 1991.

[26] J.L. Higle and S. Sen. *Stochastic Decomposition: A Statistical Method for Large*

Scale Stochastic Linear Programming. Kluwer Academic Publishers, Dordrecht, 1996.

[27] G. Infanger. Monte Carlo (importance) sampling within a Benders' decomposition algorithm for stochastic linear programs. *Annals of Operations Research*, 39:69–95, 1992.

[28] P. Kall and J. Mayer. *Stochastic Linear Programming: Models, Theory and Computation.* Springer, New York, 2005.

[29] Y.M. Kaniovski, A.J. King, and R.J.-B.Wets. Probabilistic bounds (via large deviations) for the solutions of stochastic programming problems. *Annals of Operations Research*, 56:189–208, 1995.

[30] J. Kiefer and J. Wolfowitz. Stochastic estimation of the maximum of a regression function. *Annals of Mathematical Statistics*, 23:462–466, 1952.

[31] A.J. King and R.T. Rockafellar. Asymptotic theory for solutions in statistical estimation and stochastic programming. *Mathematics of Operations Research*, 18:148–162, 1993.

[32] A.J. Kleywegt, A. Shapiro, and T. Homem-de-Mello. The sample average approximation method for stochastic discrete optimization. *SIAM Journal on Optimization*, 12:479–502, 2002.

[33] F.V. Louveaux and R. Schultz. Stochastic integer programming. In A. Ruszczyński and A. Shapiro, editors, *Stochastic Programming: Handbooks in Operations Research and Management Science.* Elsevier, Amsterdam, 2003.

[34] W.K. Mak, D.P. Morton, and R.K. Wood. Monte Carlo bounding techniques for determining solution quality in stochastic programs. *Operations Research Letters*, 24:47–56, 1999.

[35] V. I. Norkin, G. Ch. Pflug, and A. Ruszczyinski. A branch and bound method for stochastic global optimization. *Mathematical Programming*, 83:425–450, 1998.

[36] A. Partani. Upper and lower bounds on the value of the stochastic solution in stochastic programming. The University of Texas at Austin, December 2005. Master's Report.

[37] A. Partani, D.P. Morton, and I. Popova. Jackknife estimators for reducing bias in asset allocation. In *Proceedings of the 2006 Winter Simulation Conference*, 2006.

[38] T. Pennanen and M. Koivu. An adaptive importance sampling technique. In *Proceedings of Monte Carlo and Quasi-Monte Carlo Methods 2004*, pages 443–455, 2004.

[39] G.Ch. Pflug. Stochastic optimization and statistical inference. In A. Ruszczyński and A. Shapiro, editors, *Stochastic Programming: Handbooks in Operations Research and Management Science*. Elsevier, Amsterdam, 2003.

[40] I. Pigeot. A jackknife estimator of a combined odds ratio. *Biometrics*, 47:373–381, 1991.

[41] I. Pigeot. The jackknife and bootstrap in biomedical research- common principles and possible pitfalls. *Drug Information Journal*, 35:1431–1443, 2001.

[42] M.H. Quenouille. Approximate tests of correlation in time-series. *Journal of the Royal Statistical Society, Series B*, 11:68–84, 1949.

[43] M.H. Quenouille. Problems in plane sampling. *Annals of Mathematical Statistics*, 20:355–375, 1949.

[44] M.H. Quenouille. Notes on bias in estimation. *Biometrika*, 43:353–360, 1956.

[45] H. Robbins and S. Monro. A stochastic approximation method. *Annals of Mathematical Statistics*, 22:400–407, 1951.

[46] A. Ruszczyński. A regularized decomposition method for minimizing a sum of polyhedral functions. *Mathematical Programming*, 35:309–333, 1986.

[47] A. Ruszczyński and A. Świetanowski. Accelerating the regularized decomposition method for two stage stochastic linear problems. *European Journal of Operational Research*, 101:328–342, 1997.

[48] L. Schruben. Coverage function of interval estimators of simulation response. *Management Science*, 26:18–27, 1980.

[49] S. Sen, R.D. Doverspike, and S. Cosares. Network planning with random demand. *Telecommunication Systems*, 3:11–30, 1994.

[50] J. Shao and D. Tu. *The Jackknife and Bootstrap*. Springer-Verlag, New York, 1995.

[51] A. Shapiro. Asymptotic analysis of stochastic programs. *Annals of Operations Research*, 30:169–186, 1991.

[52] A. Shapiro. Monte Carlo sampling methods. In A. Ruszczyński and A. Shapiro, editors, *Stochastic Programming: Handbooks in Operations Research and Management Science*. Elsevier, Amsterdam, 2003.

[53] A. Shapiro, T. Homem-de-Mello, and J. Kim. Conditioning of convex piecewise linear stochastic programs. *Mathematical Programming*, 94:1–19, 2002.

[54] J. Tukey. Bias and confidence in not quite large samples. *Annals of Mathematical Statistics*, 29:614, 1958.

[55] R.M. van Slyke and R.J.-B. Wets. L-shaped linear programs with applications to optimal control and stochastic programming. *SIAM Journal on Applied Mathematics*, 17:638–663, 1969.

[56] S. W. Wallace and W. T. Ziemba, editors. *Applications of Stochastic Programming*. Number 5 in MPS-SIAM Series on Optimization. SIAM, Philadelphia, 2005.

Vita

Amit Partani was born in the family of Mrs. Bhagyawati and Mr. Sitaram Partani at Malpura, Rajasthan, India on December 11, 1978. He did his schooling at Adarsh Vidhya Mandir, Malpura (till 2^{nd} standard) and Kendriya Vidhyalaya, Malpura (3^{rd}-10^{th} standard). He moved to Birla Public School, Pilani, India for high school studies. He admitted to Indian Institute of Technology (IITB), Mumbai, India in July 1997 and received his B.Tech. in Mechanical Engineering from IITB in May 2001. As an undergraduate he won three prestigious awards at IITB for his extra-curricular and sports endeavors. Immediately after his graduation he was admitted to the University of Texas at Austin in 2001 and received his MSE in Operations Research & Industrial Engineering in December 2005. He married beautiful and lovely Aditi Jakhetiya on January 22, 2006. In 2007, he won prestigious ICS Best Student Paper Award and was awarded Honorable Mention for the G.E. Nicholson Best Student Paper award for his PhD research. He graduated with his PhD in Operations Research & Industrial Engineering from the University of Texas at Austin in December 2007. He joined Wellington Management Company, Boston, MA as Assistant Vice President in January 2008.

Permanent Address: s/o Mr. Sitaram Partani,
Ward no. 10, Nai Nagri,
Malpura, Dist. Tonk
Rajasthan - 304502, India

This dissertation was typeset with LaTeX 2_ε[1] by the author.

[1] LaTeX 2_ε is an extension of LaTeX. LaTeX is a collection of macros for TeX. TeX is a trademark of the American Mathematical Society. The macros used in formatting this dissertation were written by Dinesh Das, Department of Computer Sciences, The University of Texas at Austin, and extended by Bert Kay, James A. Bednar, and Ayman El-Khashab.