

## CONTROL VARIATES FOR SENSITIVITY ESTIMATION

Tarik Borogovac

Dept. of Electrical & Computer Engineering  
Boston University  
Boston, MA 02225

Na Sun

Division of Systems Engineering  
Boston University  
Boston, MA 02225

Pirooz Vakili

Division of Systems Engineering &  
Mechanical Engineering Department  
Boston University  
Boston, MA 02225

### ABSTRACT

We adapt a newly proposed generic approach to control variate selection to the problem of efficient estimation of sensitivity of financial security prices to model parameters, the so-called Greeks. We show that estimators based on pathwise and likelihood ratio methods can be cast in a general setting where generic control variates can be systematically defined for their estimation. In general, the means of such controls cannot be exactly calculated. One can use the Biased or Estimated Control Variates approach and estimate the means via simulation, or use the approach of DataBase Monte Carlo (DBMC) which also requires estimation of control means via simulation. We consider a parametric setting where price sensitivities need to be estimated repeatedly at multiple parameters. The fact that the same controls can be used for multiple estimation problems can justify the setup cost. The approach is illustrated via simple examples and preliminary computational results are provided.

### 1 INTRODUCTION

Sensitivity estimation plays an important part in the analysis, optimization, and control of general stochastic systems. While the approach of this paper is broadly applicable to efficient sensitivity estimation problems, to fix ideas, we limit ourselves to the case of estimating sensitivity of financial security prices with respect to model parameters, the so called “Greeks.” These sensitivities are used in hedging and risk management in finance. In contrast to security prices, price sensitivities are not market observable and they need to be estimated using stochastic models of relevant state variables. Given a stochastic model, Monte Carlo simulation is a general and flexible tool for sensitivity estimation. On the other hand, it is observed that the computational effort needed for accurate estimation of price sensitivities via Monte Carlo can be substantial, “easily 10 to 100 times greater than that for estimating security prices with the same level of accuracy” (Chen and Glasserman 2007). This motivates our work which aims to improve the efficiency of Monte Carlo for sensitivity estimation.

There are two broad methods of direct sensitivity estimation via Monte Carlo, namely, the pathwise method and the likelihood ratio method (see, e.g., (Asmussen and Glynn 2007) and (Glasserman 2004)). More recently, price sensitivity estimators based on Malliavin calculus have been proposed (see, e.g., (Fournie, Lasry, and Lebuchous 1999), (Fournie et al. 2001)). Malliavin estimators are defined in continuous time. To evaluate them via Monte Carlo, they need to be discretized in time. It is shown that the discretized versions are closely related to combinations of pathwise and likelihood ratio derivatives (Chen and Glasserman 2007). We, therefore, focus on the two methods of pathwise and likelihood ratio derivatives. Our approach extends to Malliavin based estimators estimated via simulation.

In this paper, we consider the variance reduction technique of Control Variates (CV) in order to increase the efficiency of Monte Carlo for sensitivity estimation. To use the CV technique, the first step is to identify a set of control variates. In common applications of the CV technique this step requires a tailor-made approach for each problem: effective controls can be found if one can identify and exploit special features of the problem. We follow the approach proposed in (Borogovac and Vakili 2009) that is more systematic in control variate selection. This approach is applicable when the estimation problem depends on some model or decision

parameter and can be represented as estimating

$$J(\theta) = E[Y(\theta, \omega)]$$

where  $\{Y(\theta); \theta \in \Theta\}$  is a family of random variables defined on the same probability space  $(\Omega, \mathcal{B}, P)$ . It is assumed that the probability measure  $P$  does not depend on parameter  $\theta$ . In this case, given  $\omega \in \Omega$ ,  $Y(\theta, \omega)$  is a deterministic function of  $\theta \in \Theta$ . (Borogovac and Vakili 2009) shows that common deterministic function approximation methods that are linear in their input data imply control variates that can be very effective. These include, for example, Taylor expansion and linear interpolation. It is worth noting that the approach does not require the user to commit to or use any of the deterministic function approximation methods; the approximation methods simply provide a theoretical basis for the use of the implied control variates.

Once the control variates are identified, the process of controlled estimation is fairly standard and straightforward (see, e.g., (Nelson 1990)). In common applications of the CV technique, the control variates are limited to those whose means can be precisely computed. In almost all cases we consider, the means of the control variates cannot be exactly computed and need to be estimated. The user can follow the Estimated/Biased Control Variate approach (see (Schmeiser, Taaffe, and Wang 2001), (Pasupathy et al. 2008), (Emsermann and Simon 2002)) where the means of the control variates are approximated or estimated via Monte Carlo. Alternatively, the DataBase Monte Carlo (DBMC) implementation proposed in (Borogovac and Vakili 2008, Borogovac and Vakili 2009) can be used, where, again, the means of the control variates need to be estimated via Monte Carlo. In either case, the implementation requires some computational setup cost. The justification or payback for this expenditure is the sum of efficiency gains when using the same set of controls to estimate sensitivities at many parameter values. Price sensitivity estimation of financial securities is a particularly well-suited application for such an implementation for at least two reasons: Repeated sensitivity estimation at different parameters is of practical interest (Staum 2009) and the relevant range of parameter values is sufficiently small to make controls based on quantities at a select set of parameter values effective for estimation at most other parameters.

In this paper we cast the sensitivity estimation via pathwise and likelihood ratio estimators in a setting where the generic control variate selection approach of (Borogovac and Vakili 2009) can be applied. We also consider generic control variates specifically used for derivative estimation (Borogovac 2009). We use the DBMC approach for controlled estimation and use simple examples to illustrate the approach. We use some examples from (Brodie and Glasserman 1996) in order to provide a basis for comparison with our preliminary computational experiments.

The rest of the paper is organized as follows. In Section 2, following (L'Ecuyer 1990), we give a representation of sensitivity estimation that reflects both pathwise and likelihood ratio estimation problems. A brief description of the control variate technique is included in Section 3 before we present a systematic approach to control variate selection in Section 4. We discuss two possible implementations of the CV technique that require estimating the mean of the control variates in Section 5. Experimental results are presented in Section 6 and we conclude in Section 7.

## 2 A UNIFIED REPRESENTATION OF SENSITIVITY ESTIMATORS

We begin with a brief review of sensitivity estimators that are used in Monte Carlo simulation. For more details, see, e.g., (Asmussen and Glynn 2007) and (Glasserman 2004).

Let  $\{\Psi(\theta), \theta \in \Theta\}$  denote a family of random variables, where  $\Psi(\theta)$  represents the sample performance of a stochastic system or the sample discounted payoff of a financial security;  $\theta$  is model or decision variable. To simplify the discussion, assume  $\theta$  is a scalar. Let

$$\alpha(\theta) = E[\Psi(\theta)]$$

denote the expected system performance or the price of a financial security. Sensitivity estimation is the problem of estimating

$$\alpha'(\theta) = \frac{d\alpha}{d\theta} = \frac{d}{d\theta} E[\Psi(\theta)].$$

The most straightforward and longstanding sensitivity estimators are those based on the finite-difference approach. For example,

$$\hat{\alpha}'(\theta) = \frac{\Psi(\theta+h) - \Psi(\theta)}{h}$$

represents the forward-difference estimator ( $h > 0$ ). These are sometimes called indirect estimators since to evaluate them we need to generate samples from a system with perturbed parameters. If sensitivities with respect to  $k$  parameters are required, we need to simulate samples at  $k$  different perturbed systems in addition to sampling at the nominal parameter.

Direct sensitivity estimators that require sampling at a single parameter only fall into two broad categories of path or pathwise derivatives—some times called Infinitesimal Perturbation Analysis (IPA) estimators—and likelihood ratio (LR) derivatives. Following (L'Ecuyer 1990), let  $\{P_\theta; \theta \in \Theta\}$  be a family of probability measures on the same measurable space  $(\Omega, \mathcal{B})$ . Let  $G$  be a probability measure that dominates all  $P_\theta$ , i.e.,  $P_\theta$  is absolutely continuous with respect to  $G$  for all  $\theta \in \Theta$ . Then  $\alpha(\theta)$  can be written as

$$\alpha(\theta) = E_\theta[\Psi(\theta)] = \int_\Omega \Psi(\theta, \omega) dP_\theta(\omega) = \int_\Omega \Psi(\theta, \omega) L(G, \theta, \omega) dG(\omega) = E_G[\Psi(\theta)L(G, \theta)]$$

where  $L(G, \theta, \omega) = (dP_\theta/dG)(\omega)$  is the likelihood ratio. Then, subject to validity of the interchange of differentiation and integration, we have

$$\alpha'(\theta) = E_G\left[\frac{d\Psi(\theta)}{d\theta}L(G, \theta) + \Psi(\theta)\frac{dL(G, \theta)}{d\theta}\right].$$

The above identity implies the following unbiased estimator for  $\alpha'(\theta)$

$$\widehat{\alpha}'(\theta) = \frac{d\Psi(\theta)}{d\theta}L(G, \theta) + \Psi(\theta)\frac{dL(G, \theta)}{d\theta}. \tag{1}$$

In most stochastic models of practical interest either the sample performance  $\Psi$  or the probability measure  $P$  depends on  $\theta$  and not both. If this is not the case, in almost all cases, the user can select an equivalent representation for which this is true.

If  $\theta$  is only a parameter of the sample performance  $\Psi$  and  $P = G$ , then the above estimator is the so-called *pathwise estimator*, i.e.,

$$\widehat{\alpha}'_{PW}(\theta) = \frac{d\Psi}{d\theta}(\theta). \tag{2}$$

On the other hand, if the sample performance is independent of  $\theta$  and the sampling measure  $G$  is equal to  $P_{\theta_0}$  for a nominal parameter  $\theta_0$ , then the estimator is the so-called *likelihood ratio estimator*

$$\widehat{\alpha}'_{LR}(\theta_0) = \Psi \cdot \frac{d \ln(dP_\theta)}{d\theta}(\theta_0). \tag{3}$$

More generally, if we fix  $G = P_{\theta_0}$ , then

$$\widehat{\alpha}'_{LR}(\theta) = \Psi \cdot \frac{dL(\theta_0, \theta)}{d\theta}. \tag{4}$$

Therefore, for both cases of pathwise and likelihood ratio derivatives we can write

$$\alpha'(\theta) = J(\theta) = E[Y(\theta, \omega)] \tag{5}$$

where  $\{Y(\theta); \theta \in \Theta\}$  is an appropriately defined family of random variables defined on the same probability space  $(\Omega, \mathcal{B}, P)$  for an appropriately defined probability measure  $P$ . Note that this representation is not unique.

Given such a representation, we can use the approach proposed by (Borogovac and Vakili 2009) for control variate selection. Before describing the approach we briefly review the CV technique.

### 3 CONTROL VARIATE TECHNIQUE

Assume our goal is to estimate the unknown mean of an estimation variable  $Y$ ,  $J = E[Y]$ . Let  $Y$  be defined on the probability space  $(\Omega, \mathcal{B}, P)$  and let  $\{Y_1, \dots, Y_n\}$  be an i.i.d. sample of  $Y$ . Then the standard/crude estimator

for  $J = E[Y]$  is the sample average

$$\hat{J}(n) = \bar{Y}(n) = \frac{1}{n}(Y_1 + \dots, Y_n).$$

Under mild conditions we have the central limit theorem

$$\sqrt{n} \frac{\hat{J}(n) - J}{S(n)} \Rightarrow N(0, 1)$$

that provides the basis for constructing asymptotically valid confidence intervals for  $J$  ( $S(n)$  is the sample standard deviation and  $\Rightarrow$  denotes weak convergence).

Assume  $k$  controls  $X(1), \dots, X(k)$ , defined on the same probability space  $(\Omega, \mathcal{B}, P)$ , have somehow been identified. Let  $\mathbf{X} = (X(1), \dots, X(k))$ . (All vectors are assumed to be column vectors. They are written as row vectors for ease of presentation.) Assume the mean of  $X(i)$  is known for all  $i = 1, \dots, k$ . Without loss of generality we can assume  $E[X(i)] = 0, i = 1, \dots, k$ . For any set of scalars  $\beta_1, \dots, \beta_k$  define the following *controlled* variable

$$Z(\beta) = Y - (\beta_1 X(1) + \dots + \beta_k X(k)) = Y - \beta^\top \mathbf{X}$$

where  $\beta = (\beta_1, \dots, \beta_k)$  and  $\top$  denotes transpose.  $Z(\beta)$  is an unbiased estimator of  $J$  for any  $\beta \in \mathbb{R}^k$ , i.e.,

$$E[Z(\beta)] = J.$$

Assume  $Y$  and  $X(i) \in L^2(\Omega, \mathcal{B}, P)$  for all  $i$  where  $L^2(\Omega, \mathcal{B}, P)$  (or  $L^2$  for simplicity) is the Hilbert space of random variables on  $(\Omega, \mathcal{B}, P)$  with finite second moment. Furthermore assume the covariance matrix of  $\mathbf{X}$ ,  $\Sigma_{\mathbf{X}}$ , is nonsingular. Then, there exists a variance minimizing coefficient vector  $\beta^* = \Sigma_{\mathbf{X}}^{-1} \Sigma_{\mathbf{X}Y}$ . Let  $Z^* = Z(\beta^*)$ . The minimized variance is

$$\sigma_{Z^*}^2 = (1 - R_{\mathbf{X}Y}^2) \sigma_Y^2.$$

where  $R_{\mathbf{X}Y}^2$  is the squared correlation coefficient given by  $R_{\mathbf{X}Y}^2 = \Sigma_{\mathbf{X}Y}^\top \Sigma_{\mathbf{X}}^{-1} \Sigma_{\mathbf{X}Y} / \sigma_Y^2$ .

The variance reduction ratio due to using samples of the optimally controlled  $Z^*$  as opposed to the uncontrolled  $Y$ , denoted by VRR, is given by

$$VRR = \frac{\sigma_Y^2}{\sigma_{Z^*}^2} = (1 - R_{\mathbf{X}Y}^2)^{-1}.$$

In almost all applications of the CV method  $\beta^*$  needs to be estimated due to the fact that all or some components of  $\Sigma_{\mathbf{X}Y}$  and  $\Sigma_{\mathbf{X}}^{-1}$  are not known in advance and need to be estimated from i.i.d. samples  $(Y_1, \mathbf{X}_1), \dots, (Y_n, \mathbf{X}_n)$ .

Let  $\mathcal{S}$  denote the closed linear subspace of  $L^2$  generated by  $X(1), \dots, X(k)$  i.e.,

$$\mathcal{S} = \{\beta^\top \mathbf{X}; \beta \in \mathbb{R}^k, \mathbf{X} = (X(1), \dots, X(k))\}.$$

Then  $\mathbf{X}^* = \beta^{*\top} \mathbf{X}$  is the perpendicular projection of  $Y - E[Y]$  onto  $\mathcal{S}$  and we have the following variance decomposition.

$$\text{Var}(Y) = \text{Var}(\mathbf{X}^*) + \text{Var}(Y - \mathbf{X}^*)$$

A portion of  $\text{Var}(Y)$ , i.e.,  $\text{Var}(\mathbf{X}^*)$ , can be explained and removed by the control variates. Furthermore,  $Z^*$  is perpendicular to  $\mathcal{S}$  hence uncorrelated to every element of  $\mathcal{S}$ . As a result its variance cannot be further reduced using the controls. For a Hilbert space-based exposition of the CV technique, derivations, and further results, see, e.g., (Szechtman 2006).

Let  $\mathbf{A}$  be a non-singular  $k \times k$  matrix and  $\mathbf{A}\mathbf{X}$  a nonsingular transformation of  $\mathbf{X}$ . Components of  $\mathbf{A}\mathbf{X}$  produce a new set of controls that span the same linear subspace as that spanned by components of  $\mathbf{X}$  and yield the same optimal controlled estimator  $Z^*$ . Therefore, they form a vector of controls as effective as the original. In other words, the effectiveness of a vector of controls is a property of the *linear subspace generated by its components* rather than the controls themselves. Thus, an effective vector of controls is one where the

linear subspace generated by its components is *close* to the estimation variable  $Y$ . This notion of *proximity* is the basis for the controls proposed in the next section.

#### 4 SYSTEMATIC CONTROL VARIATE SELECTION

To present the systematic control variate selection approach, we return to the representation (5) of Section 2. Our objective is to estimate

$$J(\theta) = E[Y(\theta, \omega)] \tag{6}$$

where  $\{Y(\theta); \theta \in \Theta\}$  is a family of random variables defined on the same probability space  $(\Omega, \mathcal{B}, P)$ . The probability measure  $P$  does not depend on  $\theta$ .

We begin with an example to describe the basic idea underlying the approach. Assume  $\theta$  is a scalar parameter and  $\omega \in \Omega$ . Then,  $Y(\theta, \omega)$  is a deterministic function of  $\theta$ . Let  $\theta_1 < \theta < \theta_2$  and assume  $Y(\theta_1, \omega)$  and  $Y(\theta_2, \omega)$  are computed. One way to approximate  $Y(\theta, \omega)$  is to use interpolation based on values  $Y(\theta_1, \omega)$  and  $Y(\theta_2, \omega)$ . In this case we have

$$\widehat{Y}(\theta, \omega) = \frac{\theta - \theta_1}{\theta_2 - \theta_1} Y(\theta_1, \omega) + \frac{\theta_2 - \theta}{\theta_2 - \theta_1} Y(\theta_2, \omega) = \beta_1 Y(\theta_1, \omega) + \beta_2 Y(\theta_2, \omega).$$

It is important to note that  $\beta_1, \beta_2$  are the same for all  $\omega \in \Omega$ .

Define

$$\widehat{Y}(\theta) = \beta_1 Y(\theta_1) + \beta_2 Y(\theta_2)$$

pointwise, i.e., for each sample  $\omega' \in \Omega$ . Then, every realization of  $\widehat{Y}(\theta)$  also is a deterministic approximation of the realization of  $Y(\theta)$  based on interpolating realizations of  $Y(\theta_1)$  and  $Y(\theta_2)$ . In other words,

$$\widehat{Y}(\theta, \omega') = \beta_1 Y(\theta_1, \omega') + \beta_2 Y(\theta_2, \omega') \quad \text{for all } \omega' \in \Omega.$$

Now define  $X(i) = Y(\theta_i) - E[Y(\theta_i)]$  for  $i = 1, 2$ . The following argument suggests that  $X(1)$  and  $X(2)$  can be effective control variates for estimating  $J(\theta) = E[Y(\theta)]$ .

Let

$$\varepsilon(\omega) = |\widehat{Y}(\theta, \omega) - Y(\theta, \omega)|$$

denote the deterministic approximation error for sample  $\omega$ . Furthermore, let

$$Z(\theta) = Y(\theta) - (\beta_1 X(1) + \beta_2 X(2))$$

be a controlled estimator of  $J(\theta)$  using control variates  $X(1)$  and  $X(2)$  and coefficients implied by the linear interpolation. Then we have,

$$\text{Var}(Z(\theta)) = \text{Var}(Y(\theta) - (\beta_1 X(1) + \beta_2 X(2))) = \text{Var}(Y(\theta) - \widehat{Y}(\theta)) = E[\varepsilon^2].$$

Therefore, if “on ( $L^2$ ) average” the samplewise deterministic approximation error is small, then the controlled estimator of  $J(\theta)$  will have a small variance.

We would like to highlight the following.

- The coefficients of the optimally controlled estimator of  $J(\theta)$  are not necessarily  $\beta_1$  and  $\beta_2$  and in general they are not.
- $E[\varepsilon^2]$  is an upper bound of the variance of the optimally controlled estimator.
- Linear interpolation has no practical role in the implementation of CV technique using control variates  $X(1)$  and  $X(2)$ . We do not need to know  $\beta_1$  and  $\beta_2$  that are sample-wise quantities. The linear interpolation simply provides a theoretical justification for using control variates  $X(1)$  and  $X(2)$ . The appropriate coefficients in CV implementation are determined on a “population-wise” basis. In this sense, we say that  $X(1)$  and  $X(2)$  are control variates implied or suggested by linear interpolation.
- Linear interpolation or other deterministic function approximation methods that imply control variates can play a role in the analysis of the properties of the controlled estimators. See, for example, the proposition given below.

(Borogovac and Vakili 2009) gives the following more general result based on polynomial rather than linear interpolation.

**Proposition.** Let  $\theta \in [a, b] \subset \mathbb{R}$ . Assume the following are in force.

1.  $P(Y(\theta) \in C^k([a, b])) = 1$ .
2. There exists a random variable  $K$  on  $(\Omega, \mathcal{B}, P)$  such that  $Y^{(k)}(\theta) \leq K$  uniformly over  $\theta \in [a, b]$ .
3.  $K \in L^2(\Omega, \mathcal{B}, P)$ .

Then let  $X(i) = Y(\theta_i) - E[Y(\theta_i)]$  ( $i = 1, \dots, k$ ,  $\theta_i \in (a, b)$ ) be a set of controls for estimating  $E[Y(\theta)]$  and  $Z^*$  the optimal controlled estimator. Then we have, for some  $C > 0$ ,

$$VRR^*(\theta) \geq C(\min_i |\theta - \theta_i|)^{-2}.$$

( $C^k([a, b])$  is the set of  $k$ -times continuously differentiable functions on  $[a, b]$ .  $VRR^*(\theta)$  is variance reduction ratio of optimally controlled estimator compared to crude Monte Carlo.

A generalization of the argument we presented above is that all deterministic function approximation methods that are linear in their input data and where the linear coefficients of the approximations are independent of the realization of the random function being estimated, imply sets of controls defined by their input data (see, (Borogovac and Vakili 2009)). This is true for multidimensional as well as scalar  $\theta$ . This observation suggests a large class of control variates that can be considered for parametric estimation in general and sensitivity estimation at multiple parameters in particular. In our experiments we consider the following two types

- (PL)  $X(i) = Y(\theta_i)$ ,  $i = 1, \dots, k$ , implied by polynomial interpolation.  
 (TY)  $X(i) = Y^i(\theta_0)$ ,  $i = 0, \dots, k$ , implied by Taylor expansion at  $\theta_0$ . ( $Y^i(\theta_0)$  is the  $i$ th derivative of  $Y$  at  $\theta_0$ .)

So far we have not taken into account the fact that  $Y$  itself is a derivative of another function, say  $Y = d\Phi/d\theta$ . Our discussion so far applies equally to the estimation of the expected performance or the price of a security as to the estimation of their sensitivities to model parameters. We now specifically look at what can be said when  $Y = d\Phi/d\theta$  for some function  $\Phi$ . Again, to simplify, assume  $\theta$  is a scalar.

Note that  $Y$  can be approximated by a finite difference:

$$Y(\theta, \omega) \approx \frac{\Phi(\theta + h, \omega) - \Phi(\theta, \omega)}{h} = \frac{1}{h}\Phi(\theta + h, \omega) - \frac{1}{h}\Phi(\theta, \omega) = \beta_1\Phi(\theta + h, \omega) + \beta_2\Phi(\theta, \omega).$$

This finite difference approximation satisfies the criterion we specified above. Namely, it is linear in its input data, i.e.,  $\Phi(\theta + h, \omega)$  and  $\Phi(\theta, \omega)$ , and the linear coefficients of the approximation, i.e.,  $1/h$  and  $-1/h$  are independent of the realization of the random function being estimated. Therefore, this approximation implies that  $\Phi(\theta + h)$ , and  $\Phi(\theta)$  can be effective control variates. This argument generalizes to estimating higher order derivatives of  $\Phi$  (see (Borogovac 2009)).

Based on the above discussion, we consider the following class of control variates in our experiments

- (FD)  $X(i) = \Phi(\theta_i)$ ,  $i = 1, \dots, k$ , implied by finite difference method.

Here, to illustrate, we give an example that will also be used in some of the experiments presented in Section 6. This example and others we consider in Section 6 are from (Brodie and Glasserman 1996).

#### 4.1 Illustrating example

Assume our objective is to estimate the *delta* of a European call option on a dividend paying stock that satisfies the Black-Scholes (BS) model. Delta is the sensitivity of the price of the option to the initial value of the stock.

Assume the stock price process,  $\{S_t; t \in [0, T]\}$ , follows the Black-Scholes model under the risk-neutral measure. Namely,

$$\frac{dS_t}{S_t} = (r - \delta)dt + \sigma dW_t$$

where  $r$  is the risk-free interest rate,  $\delta$  is the rate of dividend payment,  $\sigma$  is the volatility parameter, and  $\{W_t\}$  is a standard Brownian motion. Let  $K$  be the strike price. In general, the parameters of interest are  $S_0, r, \sigma, K$ , and  $T$  and we can take  $\theta = (S_0, r, \sigma, K, T)$ . To simplify the discussion, we limit ourselves to the initial stock price and assume  $\theta = S_0$ .

Let  $\Phi$  denote the discounted payoff of the option, i.e.,

$$\Phi(\theta) = e^{-rT} (S_T - K)^+$$

where  $S_T = S_0 \cdot \exp((r - \delta - \sigma^2/2)T + \sigma W_T)$ .

The price of the option is given by

$$p(\theta) = E[\Phi(\theta)]$$

and the delta is

$$\frac{dP}{d\theta}(\theta).$$

We give the pathwise and likelihood ratio estimators of delta and the proposed control variates for each estimator below. For derivation of the pathwise and likelihood ratio estimators, see (Brodie and Glasserman 1996).

The *pathwise estimator* of delta is

$$D_{PW}(\theta) \triangleq \frac{\partial \Phi}{\partial \theta}(\theta) = Y(\theta) = e^{-rT} I_{\{S_T \geq K\}} \frac{S_T}{S_0}.$$

- **PL** control variates. These controls are simply of the form  $X(1) = D_{PW}(S_0(1)), \dots, X(k) = D_{PW}(S_0(k))$ , where  $S_0(i)$  is the  $i$ th initial stock price. These initial values are selected by the user.
- **TY** control variates. To define a Taylor expansion we note that the derivative of  $D_{PW}(\theta)$  with respect to  $\theta$  is zero almost everywhere. As a result, we do not consider these controls for this case.
- **FD** control variates. In this case we have  $X(i) = \Phi(\theta_i), i = 1, \dots, k$ , where  $\theta_i$ 's are again selected by the user.

The *likelihood ratio estimator* of delta is

$$D_{LR}(\theta) = Y(\theta) = \Phi(\theta) \cdot \frac{\ln(S_T/S_0) - (r - \delta - \sigma^2/2)T}{S_0 \sigma^2 T}.$$

- **PL** control variates. The controls are of the form  $X(1) = Y(S_0(1)), \dots, X(k) = Y(S_0(k))$  where  $S_0(i)$  is the  $i$ th initial stock price. These initial values are selected by the user.
- **TY** control variates. The controls are  $X(1) = D_{LR}(\theta_0), X(2) = \frac{d}{d\theta} D_{LR}(\theta_0)$ , etc.
- **FD** control variates. In this case we need to find  $\Upsilon(\theta)$  such that  $\frac{d\Upsilon}{d\theta}(\theta) = Y(\theta)$ . It can be verified that

$$\Upsilon(\theta) = \Phi(\theta_0) \frac{g(x, \theta)}{g(x, \theta_0)}$$

where  $g(x, \theta)$  is the density of terminal stock price  $S_T$  when initial stock price is  $\theta$ . Then, we have  $X(i) = \Upsilon(\theta_i), i = 1, \dots, k$ , where  $\theta_i$ 's are again selected by the user.

## 5 IMPLEMENTATION

As stated earlier, the means of the control variates proposed in the previous section can rarely be computed exactly. They need to be estimated. In our experiments in Section 6 we use the DataBase Monte Carlo (DBMC) approach for this purpose. This is not the only possible implementation available. In this section, we give a brief description of the DBMC approach and comment on the alternative implementation. See, (Borogovac and Vakili 2008) and (Borogovac and Vakili 2009) for more details.

Assume our objective is to estimate

$$J(\theta) = E[Y(\theta)]$$

for some  $\theta \in \Theta$ .

We draw  $N$  i.i.d. samples from  $\Omega$  according to probability measure  $P$ . The size of  $N$  is user defined. Assume that  $N$  is “very large,” taken to be significantly larger than the number of samples generally used for estimating  $J(\theta)$ .

Let  $\Omega_{DB} = \{\omega_1, \dots, \omega_N\}$  denote the samples generated. In what follows, we assume  $\Omega_{DB}$  is fixed. We refer to  $\Omega_{DB}$  as the *database*. Let  $\tilde{P}$  denote the empirical measure associated with samples  $\omega_1, \dots, \omega_N$ . We consider solving an approximate problem defined as estimating

$$\tilde{J}(\theta) = E_{\tilde{P}}[Y(\theta)] = \tilde{E}[Y(\theta)].$$

In other words,  $\tilde{J}(\theta)$  is the expectation of  $Y(\theta)$  with respect to the empirical measure  $\tilde{P}$ .

Note that the empirical measure  $\tilde{P}$  is simply the uniform measure on  $\Omega_{DB} = \{\omega_1, \dots, \omega_N\}$  assuming that if identical samples are generated, they are kept as separate elements. Therefore,

$$\tilde{J}(\theta) = \frac{1}{N} \sum_{j=1}^N Y(\theta, \omega_j).$$

In other words,  $\tilde{J}(\theta)$  is simply the average of  $N$  i.i.d. samples of  $Y(\theta, \omega_j)$ . Given our assumption of very large  $N$ , and under some regularity assumptions on  $Y(\theta, \omega)$ , we can expect  $\tilde{J}(\theta)$  to be a very good approximation of  $J(\theta)$  with high probability.

Let  $X(1), \dots, X(k)$  be a set of control variates as proposed in the previous section. We can directly calculate the means of these controls with respect to the empirical measure  $\tilde{P}$  as

$$\tilde{\mu}(i) = \frac{1}{N} \sum_{j=1}^N X(i)(\omega_j).$$

We will not evaluate  $\tilde{J}(\theta)$  exactly for other values of  $\theta$ . Rather, we use the CV technique to find an estimate of it. We use the classical control variate technique to estimate  $\tilde{J}(\theta)$ . Figure (1) gives the outline of this implementation.

1. For  $j = 1, \dots, n$ 
  - (a) Generate a sample  $\omega_j$  uniformly from the database.
  - (b) Evaluate  $Y(\theta, \omega_j)$  and  $X(i)(\theta, \omega_j)$ ,  $i = 1, \dots, k$ .
2. The controlled estimator for  $\tilde{J}(\theta)$  is  $\bar{Z}(\theta)$  defined as

$$\bar{Z}(\theta) = \bar{Y}(\theta) + \sum_{i=1}^k \hat{\beta}_i^* (\bar{X}(i) - \tilde{\mu}(i))$$

Figure 1: Implementation of DBMC

$\bar{Y}(\theta)$  and  $\bar{X}(i)$  are the sample averages of the above  $n$  samples and  $\hat{\beta}_i^*$  is the estimate of the optimal coefficient based on these samples.

An alternative implementation is the following. Once the means of the controls are computed, we discard the samples  $\omega_1, \dots, \omega_N$ , and in step (a) of Figure 1, we generate new samples from  $\Omega$  according to the original



probability measure  $P$ . The next steps follow as in Figure 1. This is the method of Estimated Control Variate (Pasupathy et al. 2008).

Both implementations produce biased estimators of  $J(\theta)$ . The bias can be reduced by increasing  $N$ . We expect the statistical properties of the estimators resulting from the two implementations to be quite similar. For a discussion of the statistical properties of the estimators resulting from the two implementations, see (Borogovac and Vakili 2008), (Borogovac and Vakili 2009), and (Pasupathy et al. 2008).

Both implementation incur a setup cost, as should be clear from the above description. The cost can be justified if the control variates are used for estimation of  $J(\theta)$  for a large number of  $\theta \in \Theta$ , for example.

## 6 EXPERIMENTAL RESULTS

In this section, we provide some preliminary experimental results for the estimation of price sensitivities when using the control variates proposed in Section 4. To provide a basis for comparison, we have selected the set of problems included in (Brodie and Glasserman 1996). A few general comments about this comparison is in order.

- We do not run all the experiments that are included in (Brodie and Glasserman 1996); rather, for each of the three models they consider, we select the estimation of one of the price sensitivities. The results give an overall picture of the effectiveness of the controls we propose.
- In (Brodie and Glasserman 1996) for all experiments  $S_T$ , the terminal stock price (in a B-S model) is used as a control. The price sensitivity estimates and the standard errors based on crude and control estimation calculated in (Brodie and Glasserman 1996) are included in the tables to follow and identifies by a B&G designation.
- As we have argued in this paper, the proposed controls can be used for estimation of price sensitivities at many model parameter values, varying, for example, initial stock price, interest rate, volatility, strike price, or any combinations of these. (Brodie and Glasserman 1996) report experimental results for the same model for three different initial stock prices only. As a result, we only report the use of the control variates for these parametric estimations. Note that the same control variates are used in all three cases. These control variates can be used to estimate price sensitivities at other parameter values as well.
- The overall message of the following experiments is that the proposed control variates are effective. This can be taken as a proof of concept. We have not attempted to select the best possible control variates (nor have Brodie and Glasserman apparently). To assess the degree of effectiveness of the proposed control variates requires further experimentation and analysis.

### 6.1 Delta of European call option

As in Section 4 let  $\{S_t; t \in [0, T]\}$  denote a dividend paying stock price that under the risk-neutral measure satisfies the following stochastic differential equation

$$\frac{dS_t}{S_t} = (r - \delta)dt + \sigma dW_t$$

where  $r$  is the risk-free interest rate,  $\delta$  is the rate of dividend payment,  $\sigma$  is the volatility parameter, and  $W_t$  is a standard Brownian motion. Assume  $K$  be the strike price and the payoff is given by

$$\Phi(\theta) = e^{-rT} (S_T - K)^+$$

The pathwise estimator of the delta is

$$D_{PW}(\theta) \triangleq \frac{\partial \Phi}{\partial \theta}(\theta) = Y(\theta) = e^{-rT} I_{\{S_T \geq K\}} \frac{S_T}{S_0}.$$

and the likelihood ratio estimator is given by

$$D_{LR}(\theta) = Y(\theta) = \Phi(\theta) \cdot \frac{\ln(S_T/S_0) - (r - \delta - \sigma^2/2)T}{S_0 \sigma^2 T}.$$

We run  $N = 1000,000$  samples to construct the database. The estimators are based on  $n = 10,000$  samples as in (Brodie and Glasserman 1996). The parameters of the experiment and the results of the experiment are given in Table 1. All three types of control variates are used in this case for estimating  $D_{LR}$  and PL and FD control variates are used for estimating  $D_{PW}$ .

Table 1: European Call Options on Dividend Paying Assets - Delta

	Initial Asset Price( $S_0$ )					
	90	(Std. Err)	100	(Std.Err)	110	(Std. Err)
<i>Delta(dp/dS<sub>0</sub>)</i>						
Exact	0.2219		0.5684		0.8443	
Pathwise estimate(B&G)	0.217	0.005	0.561	0.005	0.844	0.004
Pathwise with control(B&G)	0.221	0.003	0.566	0.003	0.848	0.002
Pathwise estimate	0.2277	0.0046	0.5708	0.0054	0.8430	0.0040
Pathwise with PL control	0.2191	0.0032	0.5709	0.0034	0.8472	0.0023
Pathwise with FD control	0.2207	0.0029	0.5667	0.0026	0.8428	0.0025
LR estimate(B&G)	0.215	0.008	0.551	0.013	0.817	0.017
LR with Control(B&G)	0.220	0.006	0.562	0.008	0.834	0.010
Likelihood Ratio estimate	0.2153	0.0078	0.5674	0.0127	0.8491	0.0172
LR estimate with PL control	0.2230	0.0006	0.5696	0.0001	0.8467	0.0005
LR estimate with TY control	0.2223	0.0005	0.5686	0.0003	0.8448	0.0013
LR estimate with FD control	0.2228	0.0017	0.5698	0.0007	0.8454	0.0018

Parameter settings:  $\sigma_0 = 0.25, r_0 = 0.1, \delta = 0.03, T = 0.2, K = 100$ .  
 $S_1 = 95, S_2 = 105$  for PL and FD controls and  $S_0 = 99$  for TY control.

### 6.2 Vega of an Asian option

Consider the same stock price dynamics as in the previous experiment. Let  $T$  be the maturity of the option written on the average of the last  $m$  daily closing prices. Let  $S_i$  be the price at time  $t_i = T - (m - i)/365.25$ , i.e., assume the maturity is greater than the averaging period and  $t_1$  is larger than the increment between averaged prices (which is one day in this example). Let  $\bar{S} = \frac{1}{m} \sum_{i=1}^m S_i$  denote the average price.

Then

$$p(S_0, r, \sigma) = E[e^{-rT} (\bar{S} - K)^+].$$

Vega of Asian option is defined as the derivative of  $p$  with respect to  $\sigma$ , that is  $dp/d\sigma$ . The pathwise estimator of vega can be represented as

$$V_{pw}(S_0, r, \sigma) \triangleq \frac{\partial Y}{\partial \sigma}(S_0, r, \sigma) = e^{-rT} I_{\{\bar{S} \geq K\}} \frac{1}{m\sigma} \sum_{i=1}^m S_i (\ln(S_i/S_0) - (r - \delta + \sigma^2/2)t_i).$$

The likelihood ratio estimator of Vega for Asian option is

$$V_{lr}(S_0, r, \sigma) \triangleq e^{-rT} (\bar{S} - K)^+ \sum_{i=1}^m \left( -d_i \frac{\partial d_i}{\partial \sigma} - \frac{1}{\sigma} \right),$$

where

$$d_i = (\ln(S_i/S_{i-1}) - (r - \delta - \sigma^2/2)\Delta t_i) / (\sigma \sqrt{t_i}),$$

$$\partial d_i / \partial \sigma = (\ln(S_i/S_{i-1}) - (r - \delta + \sigma^2/2)\Delta t_i) / (\sigma^2 \sqrt{t_i}).$$

In this case we use only the Polynomial interpolation based and Taylor expansion based control variates. The Taylor expansion involves the sample value and the derivative of the sample value with respect to  $S_0$ . Results are provided in Table 2.

Table 2: Asian Call Options on Dividend Paying Assets - Vega

	Initial Asset Price( $S_0$ )					
	90	(Std. Err)	100	(Std.Err)	110	(Std. Err)
<i>Vega(dp/dσ)</i>						
Pathwise estimate(B&G)	8.871	0.225	15.190	0.251	8.843	0.345
Pathwise with control(B&G)	8.747	0.167	15.024	0.149	8.618	0.212
Pathwise estimate	8.9783	0.2268	15.1133	0.2522	8.8553	0.3448
Pathwise with PL control	8.8399	0.0877	14.9193	0.0255	8.5651	0.1153
Pathwise with TY control	8.8136	0.0721	14.9370	0.0086	8.6457	0.1095
Likelihood Ratio estimate(B&G)	9.161	0.962	13.844	2.516	5.338	4.812
LR estimate with control(B&G)	8.984	0.937	13.511	2.483	4.584	4.777
Likelihood Ratio estimate	8.0329	0.9318	12.4801	2.3837	5.5009	4.6012
LR estimate with PL control	8.8738	0.2780	14.6782	0.2361	9.1057	0.5895
LR estimate with TY control	8.9551	0.3695	14.9540	0.0345	9.5064	1.2236

Parameter settings:  $\sigma_0 = 0.25, r_0 = 0.1, \delta = 0.03, T = 0.2, m = 30, K = 100$ .  
 $S_1 = 95, S_2 = 105$  for PL controls;  $S_0 = 99$  for TY control.

### 6.3 Vega of a model with stochastic volatility

In this section, we consider the European call option on a dividend paying stock whose price follows the following stochastic volatility model :

$$dS_t = S_t[(r - \delta)dt + \sigma_t dZ_t],$$

$$d\sigma_t = \sigma_t[\mu dt + \xi dW_t],$$

where  $Z$  and  $W$  are correlated Brownian motion process with correlation  $\rho$ , initial price  $S_0$  and initial volatility  $\sigma_0$  are given. The price of European call option with strike price  $K$  can be similarly written as

$$p(S_0, r, \sigma_0, \mu, \xi) = E[e^{-rT}(S_T - K)^+] = E[Y(S_0, r, \sigma_0, \mu, \xi)].$$

There are three *vegas* in this case denoted by  $Vega = dp/d\sigma_0$ ,  $Vega1 = dp/d\xi$ , and  $Vega2 = dp/d\mu$ . We denote the pathwise estimator of these sensitivities as functions of the parameters by  $V_{pw}, V_{pw}^1, V_{pw}^2$ , respectively. They are given by

$$V_{pw}(S_0, r, \sigma_0, \mu, \xi) = e^{-rT} I_{\{S_T \geq K\}} \frac{S_T}{\sigma_0} \sum_{i=1}^m \left( 1 - \frac{(1 + (r - \delta)dt)S_{i-1}}{S_i} \right),$$

$$V_{pw}^1(S_0, r, \sigma_0, \mu, \xi) = e^{-rT} I_{\{S_T \geq K\}} \frac{S_T}{\xi} \sum_{i=1}^m \left( 1 - \frac{(1 + (r - \delta)dt)S_{i-1}}{S_i} \right) \times \left( \sum_{k=1}^{i-1} \left[ 1 - \frac{(1 + \mu dt)\sigma_{k-1}}{\sigma_k} \right] \right)$$

$$V_{pw}^2(S_0, r, \sigma_0, \mu, \xi) = e^{-rT} I_{\{S_T \geq K\}} S_T dt \sum_{i=1}^m \left( 1 - \frac{(1 + (r - \delta)dt)S_{i-1}}{S_i} \right) \times \left( \sum_{k=1}^{i-1} \frac{\sigma_{k-1}}{\sigma_k} \right).$$

where  $dt = T/m$  is the discretization step size. In this case we only use polynomial interpolation based estimators. Results are provided in Table 3.

Table 3: Call Options on Divident Paying Assets with Stochastic Volatility - Vega's

	Initial Asset Price( $S_0$ )					
	90	(Std. Err)	100	(Std.Err)	110	(Std. Err)
<i>Vega</i> ( $dp/d\sigma$ )						
Pathwise estimate(G&B)	11.640	0.268	16.932	0.294	10.720	0.390
Pathwise with control(G&B)	11.887	0.175	17.236	0.155	11.111	0.220
Pathwise estimate	12.0641	0.2858	17.3462	0.3141	10.8687	0.4111
Pathwise with PL control	12.0408	0.0772	17.3440	0.0313	10.9154	0.11533
<i>Vega1</i> ( $dp/d\varepsilon$ )						
Pathwise estimate(G&B)	0.339	0.025	0.004	0.031	-0.201	0.037
Pathwise with control(G&B)	0.332	0.023	-0.003	0.030	-0.209	0.035
Pathwise estimate	0.3446	0.0243	0.0007	0.0304	-0.2435	0.0363
Pathwise with PL control	0.3459	0.0076	0.0013	0.0070	-0.2442	0.0096
<i>Vega2</i> ( $dp/d\mu$ )						
Pathwise estimate(G&B)	0.302	0.008	0.429	0.009	0.273	0.012
Pathwise with control(G&B)	0.298	0.005	0.423	0.006	0.266	0.009
Pathwise estimate	0.3003	0.0078	0.4201	0.0094	0.2589	0.0123
Pathwise with PL control	0.3013	0.0027	0.4241	0.0023	0.2663	0.0035

Parameter settings:  $\sigma_0 = 0.25, \mu = -0.1, \xi = 0.3, \rho = 0.5, r_0 = 0.1, \delta = 0.03, T = 0.2, m = 130, K = 100$ .  
 $S_1 = 94, S_2 = 106$  for PL controls.

## 7 CONCLUSIONS

In this paper we have proposed a systematic way of selecting control variates for controlled sensitivity estimation. While the approach is broadly applicable, we have focused on estimating price sensitivities of financial securities to model parameters, known as the "Greeks." The proposed control variates are implied by a set of deterministic function approximation methods. There is no need to use the deterministic approximations themselves; they simply provide justifications for using the proposed controls and can help in analyzing their effectiveness.

There are many deterministic approximation methods that can be considered and there are corresponding classes of control variates that are implied. The issue of which control variates work better is most likely problem dependent and need to be investigated. We provide some preliminary experimental results for a number of price sensitivity problems and compare our results with those reported in the literature. The overall message is that the proposed control variates are effective. The have a better assessment of the degree of their effectiveness and how to best select them from the set of possible choices require more experimentation and analysis and are left for future research.

## ACKNOWLEDGMENTS

Research supported in part by the National Science Foundation grants CMMI-0620965 and DGE-0221680.

## REFERENCES

- Asmussen, S., and P. Glynn. 2007. *Stochastic Simulation: Algorithms and Analysis*. Springer.
- Borogovac, T. 2009, May. *Constructive and Generic Control Variates for Monte Carlo Estimation*. Ph. D. thesis, Boston University.
- Borogovac, T., and P. Vakili. 2008. Control Variate Technique: A constructive approach. In *Proceedings of the 2008 Winter Simulation Conference*, ed. S. J. Mason, R. R. Hill, L. Moench, and O. Rose, 320–327.
- Borogovac, T., and P. Vakili. 2009. DataBase Monte Carlo (DBMC) & Generic Control Variates for Parametric Estimation. Technical report, Boston University College of Engineering.

- Brodie, M., and P. Glasserman. 1996. Estimating security price derivatives using simulation. *Management Science* 42:269–285.
- Chen, N., and P. Glasserman. 2007. Malliavin greeks without malliavin calculus. *Management Science* 117:1689–1723.
- Emsermann, M., and B. Simon. 2002. Improving Simulation Efficiency with Quasi Control Variates. *Stochastic Models* 18:425–448.
- Fournie, E., J. Lasry, and J. Lebuchous. 1999. Applications of Malliavin Calculus to Monte Carlo methods in finance. *Finance and Stochastics* 3:391–412.
- Fournie, E., J. Lasry, J. Lebuchous, and P. Lions. 2001. Applications of Malliavin Calculus to Monte Carlo methods in finance. *Finance and Stochastics* 5:201–236.
- Glasserman, P. 2004. *Monte Carlo Methods in Financial Engineering*. Springer Verlag.
- L'Ecuyer, P. 1990. A Unified View of the IPA, SF, and LR Gradient Estimation Techniques. *Management Science* 36:1364.
- Nelson, B. L. 1990. Control Variate Remedies. *Operations Research* 38:974–992.
- Pasupathy, R., B. W. Schmeiser, M. R. Taaffe, and J. Wang. 2008. Control-variate estimation using estimated control means. *IIE Transactions*, to appear.
- Schmeiser, B., M. R. Taaffe, and J. Wang. 2001. Biased control-variate estimation. *IIE Transactions* 33:219–228.
- Staum, J. 2009. Monte Carlo Computation in Finance. In *Monte Carlo and Quasi-Monte Carlo Methods 2008*, ed. P. L'Ecuyer and A. B. Owen, 19–42. Springer.
- Szechtman, R. 2006. A Hilbert Space Approach to Variance Reduction. In *Handbook in OR and MS*, ed. S. G. Henderson and B. L. Nelson, Volume 13, Chapter 10, 259–289. Elsevier B.V.

#### AUTHOR BIOGRAPHIES

**Tarik Borogovac** is a Postdoctoral Research Associate in Electrical Engineering at Boston University, and in the Smart Lighting Engineering Research Center. His current research interests include optical communications systems, optical modeling, and Monte Carlo simulation. His e-mail address is <tarikb@bu.edu>.

**Na Sun** is a Ph.D. student of Systems Engineering at Boston University. Her current research interests include efficient Monte Carlo simulation in the areas of computational finance. Her e-mail address is <sunna@bu.edu>.

**Pirooz Vakili** is an Associate Professor in the Division of Systems Engineering and the Department of Mechanical Engineering at Boston University. His research interests include Monte Carlo simulation, optimization, computational finance, and bioinformatics. His email address is <vakili@bu.edu>.