Operations Research
Vol. 53, No. 2, March–April 2005, pp. 298–312
ISSN 0030-364X | EISSN 1526-5463 | 05 | 5302 | 0298

DOI 10.1287/opre.1040.0148
© 2005 INFORMS

Large Sample Properties of Weighted Monte Carlo Estimators

Paul Glasserman, Bin Yu
Graduate School of Business, Columbia University, New York, New York 10027
{pg20@columbia.edu, by52@gsb.columbia.edu}

A general approach to improving simulation accuracy uses information about auxiliary control variables with known expected values to improve the estimation of unknown quantities. We analyze weighted Monte Carlo estimators that implement this idea by applying weights to independent replications. The weights are chosen to constrain the weighted averages of the control variables. We distinguish two cases (unbiased and biased), depending on whether the weighted averages of the controls are constrained to equal their expected values or some other values. In both cases, the number of constraints is usually smaller than the number of replications, so there may be many feasible weights. We select maximally uniform weights by minimizing a separable convex function of the weights subject to the control variable constraints. Estimators of this form arise (sometimes implicitly) in several settings, including at least two in finance: calibrating a model to market data (as in the work of Avellaneda et al. 2001) and calculating conditional expectations to price American options. We analyze properties of these estimators as the number of replications increases. We show that in the unbiased case, weighted Monte Carlo reduces asymptotic variance, and that all convex objective functions within a large class produce estimators that are very close to each other in a strong sense. In contrast, in the biased case the choice of objective function does matter. We show explicitly how the choice of objective determines the limit to which the estimator converges.

Subject classifications: simulation: statistical analysis; finance: asset pricing.
Area of review: Simulation.
History: Received October 2002; revisions received June 2003, December 2003; accepted December 2003.

1. Introduction

One of the most fundamental ideas in stochastic simulation is that the accuracy of simulation estimates can often be improved by taking advantage of known properties of a simulated model. This principle underlies the method of control variables, which is among the most widely used variance reduction techniques. *Weighted Monte Carlo* (WMC) provides another mechanism for using information about a model to improve accuracy. This paper analyzes the convergence of WMC estimators and investigates their relation to more traditional control variable estimators.

In its simplest form, the method of linear control variables (LCV) relies on knowing the expectation of an auxiliary simulated random variable, called a control. The known expectation is compared with the estimated expectation obtained by simulation. The observed discrepancy between the two is then used to adjust estimates of other (unknown) quantities that are the primary focus of the simulation. The adjustment made is usually linear in the difference between the estimated and exact value of the expectation of the control variable.

The method of WMC also requires knowing the correct expectation of one or more control variables, but WMC forces the simulated and true values to agree, even over a finite number of replications. It accomplishes this by applying weights to the replications so that the *weighted* averages of simulated control variables match their true values (or possibly other given values, as we will see). The same weights are then used in estimating unknown quantities. In practice, the number of controls is much smaller than the number of replications, so the constraints on the weighted averages of the controls do not determine a unique set of weights. If the replications are independent and identically distributed (i.i.d.), it is natural to try to select weights that are as uniform as possible. This can be made precise by choosing weights that minimize a separable convex function, subject to the control variable constraints. This is the approach we analyze.

We distinguish two cases based on the constraints imposed on the control variables. In the *unbiased* case, the weighted average of each control is constrained to equal its expected value, whereas in the *biased* case it may be constrained to equal some other value. The purpose of WMC in the unbiased case is variance reduction: Constraining weighted averages of the controls to their expected values should reduce the variability of other weighted averages calculated from the same replications.

In the biased case, the objective of WMC is not so much to improve precision given a simulated model, but rather to correct the model itself. In this respect, it differs fundamentally from techniques usually studied in the simulation literature. Estimators of this form have been advocated by Avellaneda and colleagues (Avellaneda 1998, Avellaneda et al. 2001), and also arise as a key step in the method of Broadie et al. (2000). (These applications are the primary motivation for this work; we discuss them in §2.) Whereas in the unbiased case the constraints become nonbinding as the number of replications grows (by the strong law of large numbers), they remain binding even in the limit for the biased case. Indeed, the key issue in the biased case is understanding to what value a WMC estimator converges, rather than whether or not it reduces variance.

Our main results for the two settings are as follows. In the unbiased case, we show that WMC does indeed reduce variance asymptotically and that the variance reduction achieved is exactly the same as that achieved using an ordinary LCV estimator. Indeed, we show that for a large class of objective functions (used in selecting weights), the WMC estimator and LCV estimator are very close in a strong sense. As a step in establishing this we show that the LCV estimator is the WMC estimator for a quadratic penalty on the weights. These results for the unbiased case may be viewed as negative in the sense that they show that any advantage WMC estimators may have over LCV estimators must be limited to small samples.

For the biased case, we identify the limit to which WMC estimators converge. In contrast to the unbiased case, here the choice of objective function used to select weights does matter. Different choices of objective function correspond to different ways of adjusting or correcting a model based on side information. Our results do not provide grounds for preferring one objective function to another—no such comparison seems possible in general. Instead, our results show how the choice of objective determines the adjustment imposed on a model.

The rest of this paper is organized as follows. Section 2 gives a general formulation of WMC estimators and discusses the applications that motivate our investigation. Section 3 considers the unbiased case and formulates our main result for this setting, stating that a large class of objective functions results in nearly identical estimators. Section 4 considers the biased case; our main result in this setting shows how the choice of objective function determines the estimator's limit. The results of §§3 and 4 are illustrated with option-pricing examples. Section 5 summarizes our conclusions. Most proofs are collected in Appendix A. Our results use an assumption on the Lagrange multipliers that determine optimal weights; Appendix B provides conditions supporting this assumption.

2. Weighted Monte Carlo Estimators

2.1. Formulation

Throughout this article, $(X, Y), (X_1, Y_1), (X_2, Y_2), \ldots$ denote i.i.d., square-integrable simulation outputs with Y scalar and X taking values in $\Re^d$. The objective of the simulation is (at least initially) to estimate $\mathsf{E}[Y]$. The expectation $\mathsf{E}[X]$ is assumed known and, without loss of generality, equal to zero. Denote by

$$\begin{pmatrix} \Sigma_X & \Sigma_{XY} \\ \Sigma_{YX} & \sigma_Y^2 \end{pmatrix}$$

the covariance matrix of (X, Y), which we assume is nonsingular.

The sample mean $\bar{Y} = (Y_1 + \cdots + Y_n)/n$ converges to $\mathsf{E}[Y]$ almost surely and is asymptotically normal,

$$\sqrt{n}(\bar{Y} - \mathsf{E}[Y]) \Rightarrow N(0, \sigma_Y^2),$$

"$\Rightarrow$" denoting convergence in distribution and $N(a, b)$ denoting a normal random variable with mean a and variance b. The LCV estimator is

$$\hat{Y}_{\text{cv}} = \bar{Y} - \hat{\beta}^\top \bar{X}, \tag{1}$$

where $\bar{X}$ is the sample mean of $X_1, \ldots, X_n$ and $\hat{\beta}$ denotes the vector of coefficients in an ordinary least squares regression of $Y_1, \ldots, Y_n$ against $X_1, \ldots, X_n$. (We take vectors to be column vectors by default and use "$\top$" for transpose.) More explicitly,

$$\hat{\beta} = \frac{1}{n} M^{-1} \left[\sum_{i=1}^{n} X_i Y_i - n\bar{X}\bar{Y} \right], \tag{2}$$

with $X_i = (X_{i1}, \ldots, X_{id})^\top$ and M the $d \times d$ matrix with (k, l)th entry

$$M_{kl} = \frac{1}{n} \sum_{i=1}^{n} (X_{ik} - \bar{X}_k)(X_{il} - \bar{X}_l), \tag{3}$$

$\bar{X} = (\bar{X}_1, \ldots, \bar{X}_d)^\top$. Because M converges to Σ_X, it is invertible for all sufficiently large n.

The LCV estimator converges to $\mathsf{E}[Y]$ almost surely and it satisfies (see, e.g., Nelson 1990)

$$\sqrt{n}(\hat{Y}_{\text{cv}} - \mathsf{E}[Y]) \Rightarrow N(0, \sigma_Y^2(1 - R^2)), \tag{4}$$

with

$$R^2 = \Sigma_{YX} \Sigma_X^{-1} \Sigma_{XY} / \sigma_Y^2.$$

Asymptotically (as $n \to \infty$), linear controls never increase variance because $0 \leqslant R^2 \leqslant 1$, and they produce strict variance reduction unless $R^2 = 0$.

Now let $h: \Re \to \Re \cup \{+\infty\}$ be a strictly convex function and consider the optimization problem

$$\min_{\omega_{1,n}, \ldots, \omega_{n,n}} \sum_{i=1}^{n} h(\omega_{i,n}) \tag{5}$$

subject to

$$\frac{1}{n} \sum_{i=1}^{n} \omega_{i,n} = 1, \tag{6}$$

$$\frac{1}{n} \sum_{i=1}^{n} \omega_{i,n} X_i = c_X \tag{7}$$

for some fixed $c_X \in \Re^d$. The objective is strictly convex and the constraints are linear, so if there is a feasible solution with a finite objective function value, then there is a unique optimal solution $\omega_{1,n}, \ldots, \omega_{n,n}$. This optimum defines the WMC estimator

$$\hat{Y}_{\text{WMC}} = \frac{1}{n}\sum_{i=1}^{n} \omega_{i,n} Y_i.$$

We investigate the convergence and asymptotic normality of this type of estimator.

It is convenient to allow h to take the value $+\infty$. This allows us to include, for example, $h(w) = -\log(w)$ for all $w \in \Re$ through the convention that $-\log(w) = +\infty$ if $w \leqslant 0$. This also allows us to require positive or nonnegative weights without adding constraints by making h infinite on $(-\infty, 0]$ or $(-\infty, 0)$. While allowing $h(w)$ to be infinite for negative w is useful, there is little practical loss of generality in requiring that h be finite on $(0,\infty)$, so we assume this holds. The set on which a convex function h is finite (its effective domain) is denoted by $\text{dom}(h)$, so this condition states that $(0,\infty) \subseteq \text{dom}(h)$.

If h does take infinite values, then we need to consider whether there is a finite feasible solution to (5)–(7). Under the assumption that $(0,\infty) \subseteq \text{dom}(h)$, a finite feasible solution exists so long as c_X is in the relative interior of the convex hull of $X_1, \ldots, X_n$, because c_X is then a positive convex combination of these points (cf., Rockafellar 1970, p. 50). This almost surely holds for all sufficiently large n if c_X is in the interior of the support of X. If $\text{dom}(h)$ also includes 0, then a finite feasible solution exists if c_X is anywhere in the convex hull of $X_1, \ldots, X_n$, because it is then a (not necessarily positive) convex combination of these points.

The weights derived from (5)–(7) can be made more explicit by introducing the Lagrangian

$$L_n = \sum_{i=1}^{n} h(\omega_{i,n}) + \lambda_n\left(n - \sum_{i=1}^{n}\omega_{i,n}\right) + \mu_n^\top\left(nc_X - \sum_{i=1}^{n}\omega_{i,n}X_i\right),$$

with $\lambda_n \in \Re$ and $\mu_n \in \Re^d$. If h is differentiable, the first-order conditions characterizing the optimum are

$$0 = h'(\omega_{i,n}) - \lambda_n - \mu_n^\top X_i.$$

Because h is strictly convex, h' is strictly increasing on $\text{dom}(h)$ and has an inverse H. The optimal weights are then given by

$$\omega_{i,n} = H(\lambda_n + \mu_n^\top X_i), \tag{8}$$

with λ_n, μ_n chosen to ensure that constraints (6) and (7) are satisfied. This is a general description of the solution of a separable convex minimization problem with linear equality constraints. The computational effort required to compute a solution depends on h. Three cases are of particular interest:

Quadratic objective. Not surprisingly, the case $h(w) = w^2/2$ is particularly tractable. The optimal weights are $\omega_{i,n} = \alpha_{i,n}$, with

$$\alpha_{i,n} = 1 + \bar{X}^\top M^{-1}(\bar{X} - X_i), \tag{9}$$

where M is the $d \times d$ matrix defined in (3); see Proposition 1 below. The solution (9) requires that M be invertible, which holds for all sufficiently large n.

Empirical likelihood objective. This case takes $h(w) = -\log w$ and produces weights of the form

$$\omega_{i,n} = -\frac{1}{\lambda_n + \mu_n^\top X_i}, \tag{10}$$

with the Lagrange multipliers chosen to satisfy the constraints. The resulting WMC estimator arises as a nonparametric estimator of $\mathsf{E}[Y]$ through the empirical likelihood method in Owen (1990, 2001). This case is analyzed in the simulation setting by Szechtman and Glynn (2001), who also show that it is asymptotically equivalent to an LCV estimator. The application of empirical likelihood estimation to simulation is also considered in an unpublished work by Jing Qin (private communication). The log objective is a special case of the Cressie-Read family studied in Baggerly (1998).

Entropy objective. Setting $h(w) = w \log w$ gives the $\omega_{i,n}$ an interpretation as maximum entropy weights. This objective is also a member of the Cressie-Read family. We follow the usual convention that $0 \log 0 = 0$. The optimal weights are

$$\omega_{i,n} = \frac{n\exp(-\mu_n^\top X_i)}{\sum_{j=1}^{n}\exp(-\mu_n^\top X_j)}, \tag{11}$$

which has the form of an exponential change of distribution. This case is put forward in Avellaneda and colleagues (Avellaneda 1998, Avellaneda et al. 2001), where it is also given Bayesian and information-theoretic interpretations. Jourdain and Nguyen (2001) prove a convergence result for the measures defined by the weights using this objective.

An alternative formulation of the WMC optimization problem omits the factor $1/n$ from (6) and (7). This version can be handled through an analysis that parallels the one we develop here, and the two problems are equivalent (their optimal weights differing only by a factor of n) if h satisfies

$$h'(aw) = p(a)h'(w) + q(a) \tag{12}$$

for some functions p, q, and all aw and w in $\text{dom}(h)$. To avoid redundancy we consider only the formulation in (6)–(7). The quadratic objective, the empirical likelihood objective, and the entropy objective all satisfy (12).

Weights chosen by minimizing a separable convex objective, as in (5), are, in a sense, maximally uniform. This idea can be made precise through the *majorization* ordering, a partial order on vectors. (See, e.g., Marshall and Olkin 1979 for a definition.) To say that one vector is majorized by another is to say that its elements are more uniform.

As discussed in Marshall and Olkin (1979), for any convex h, the mapping from $(\omega_{1,n}, \ldots, \omega_{n,n})$ to $h(\omega_{1,n}) + \cdots + h(\omega_{n,n})$ is *Schur convex* and thus increasing in the majorization order. A solution to (5)–(7) is therefore a minimal element of the feasible set in the sense that it does not majorize any other feasible element. This makes precise the idea that the optimization problem (5)–(7) selects the most uniform weights consistent with the control variable constraints (7). This is intuitively appealing because in the absence of side information there is no reason to give some replications more weight than others. Different choices of h correspond to different ways of penalizing deviations from uniformity.

2.2. Applications

As noted in the introduction, we distinguish two types of WMC estimators based on the control variable constraints (7). In the unbiased case, $c_X = 0$. Recall that we assume the controls have been centered so that $\mathsf{E}[X] = 0$; taking $c_X = 0$ therefore corresponds to constraining the sample weighted averages of the controls to equal their population means, and the purpose of the weights is variance reduction. In the biased case, $c_X \neq 0$ and the weights are best viewed as attempting to correct a simulated model rather than reduce variance. We illustrate this idea with three examples arising in financial applications.

Example 1 (Model Calibration). Avellaneda and colleagues (Avellaneda 1998, Avellaneda et al. 2001) suggest the use of weighted estimators in the pricing of derivative securities. In the setting they consider, the market prices of some actively traded securities are readily observable—these are the controls—and simulation is to be used to price less-liquid securities. The simulation is based on a model of the dynamics of underlying assets, but because no model is perfect, the simulation-based price of a control security may differ from its market price, even in the limit of infinitely many replications. This in turn casts doubt on prices computed by simulation for securities for which no market price is available.

The problem, then, is to modify or adjust the simulated model to make it consistent with observed prices before trying to price a new security; this is the calibration step. Assuming the simulated model has been chosen carefully, it is natural to look for a minimal adjustment to the model in trying to match market prices. Avellaneda et al. (2001) propose calibrating the model by assigning weights to replications. In the framework of the optimization problem (5)–(7), the controls correspond to the securities for which market prices are observed. By taking c_X to be the difference between the market price and the model price, the WMC estimator attempts to calibrate the model to the market. The premise of this approach is that the price estimate $\hat{Y}_{\text{WMC}}$ is more consistent with the market prices of the control securities than is the ordinary sample mean $\bar{Y}$. Avellaneda et al. (2001) advocate using the entropy objective in selecting weights (as do Stutzer 1996 and Buchen and Kelly 1996 in closely related applications), based in part on Bayesian and information-theoretic interpretations. In §4.1, we illustrate this approach with a numerical example.

Example 2 (Reducing Discretization Bias). Many of the models commonly used to describe the dynamics of asset prices are based on diffusion processes. In practice, these are usually simulated through a discrete-time approximation, and this introduces discretization bias in prices computed by simulation. The same issue arises in virtually all applications that require simulating diffusion processes.

In some cases, the expectation of a related function of the diffusion is available in closed form and offers a potential control variable. Sometimes, the control is more easily computed in the diffusion than in the simulated approximation. For example, the expected maximum of a standard Brownian motion over $[0, T]$ is $\sqrt{2T/\pi}$, whereas the expected maximum of an approximating normal random walk is not nearly so tractable. In cases where the control is tractable for both the discrete-time and continuous-time processes, one must choose between the two values. Using the diffusion value as the "true" value for a control changes both the mean and variance of other quantities estimated from a discrete-time simulation, whereas using the discrete-time value changes the variance but has a negligible effect on the mean (at least in large samples) because it is an unbiased control.

In the WMC formulation of (5)–(7), using the diffusion value as the "true" value for a control corresponds to taking c_X equal to the difference between the diffusion and discrete-time values. In this case, the constraints (7) may be viewed as correcting (or attempting to correct) for known discretization error. An idea of this type is tested in Fu et al. (1999) and found to be advantageous. They consider only the case of a LCV, but the idea can also be applied in the WMC formulation of (5)–(7).

Related estimators using biased control variables are analyzed by Schmeiser et al. (2001), though not specifically focused on discretization error. In their setting, the simulated model is assumed to be correct, and bias results from using a numerical approximation to $\mathsf{E}[X]$, whereas in this example and the previous one $\mathsf{E}[X]$ is assumed known, but the simulated model has some error, and the value of c_X is intended to reduce the error.

Example 3 (Approximating Conditional Expectations). A European option can be exercised only at a fixed date, but an American option can be exercised at any time up to its expiration date. Valuing an American option entails finding the optimal exercise policy, and this is usually formulated as a dynamic programming problem. This makes pricing American options by simulation a challenge.

In a discrete-time formulation of the problem, the holder of the option must choose at each step whether to exercise or to continue. The payoff upon exercise is usually specified explicitly, but the continuation value is given only as a conditional expectation—the expected value of continuing given the current state—and must be calculated. Computing conditional expectations at each exercise opportunity is the main obstacle in valuing American options by simulation.

Various methods have been proposed to address this problem; the approach in Broadie et al. (2000) (and, less directly, that in Tsitsiklis and Van Roy 2001) uses weights to estimate conditional expectations from paths generated unconditionally. We describe a simplified version of the problem.

Let $(U, V), (U_1, V_1), \ldots, (U_n, V_n)$ be i.i.d. pairs; think of U and V as consecutive states of a Markov chain. Suppose we want to estimate a conditional expectation $\mathsf{E}[f(V) \mid U = u]$ for some fixed u and function f. (Think of this as the expected value of continuing rather than exercising in state u.) With $Y_i = f(V_i)$, the sample mean $\bar{Y}$ converges instead to the unconditional expectation $\mathsf{E}[f(V)]$. The goal then is to weight the Y_i to bring their weighted average closer to the conditional expectation.

Suppose we know $\mathsf{E}[g(V) \mid U = u]$ for some other function g (vector-valued) and set $X_i = g(V_i) - \mathsf{E}[g(V)]$, $X = g(V) - \mathsf{E}[g(V)]$. (We subtract $\mathsf{E}[g(V)]$ to be consistent with our convention that the controls have mean 0.) By taking $c_X = \mathsf{E}[g(V) \mid U = u] - \mathsf{E}[g(V)]$ we force the weighted average of the X_i to equal the conditional expectation of X given $U = u$. (Note that the value of c_X depends on u.) More explicitly, (7) becomes equivalent to

$$\frac{1}{n}\sum_{i=1}^{n} \omega_{i,n} g(V_i) = \mathsf{E}[g(V) \mid U = u].$$

The premise of this method is that the weighted average of the Y_i should then be closer to the conditional expectation of Y given $U = u$. This is exact if f is a linear combination of the components of g, and it should be a good approximation if f is nearly such a linear combination.

3. Variance Reduction in the Unbiased Case

3.1. Quadratic Case

In this section, we investigate WMC estimators defined with $c_X = 0$ (i.e., $c_X = \mathsf{E}[X]$) in the constraint (7). We show that this type of estimator does indeed reduce asymptotic variance, but that the estimator is nearly the same as the LCV estimator. As a first step, we show that barring degeneracies in the controls, the LCV estimator is the WMC estimator produced by a quadratic objective. "Sufficiently large n" in the following result means large enough for M in (3) to be invertible.

Proposition 1. *If $h(w) = w^2/2$, then $\hat{Y}_{\mathrm{WMC}} = \hat{Y}_{\mathrm{cv}}$ for all sufficiently large n.*

Proof. With the qualification that n be sufficiently large, we may assume that the matrix M in (3) is invertible. The function H (the inverse of h') reduces to the identity so the first-order conditions yield $\omega_{i,n} = \lambda_n + \mu_n^\top X_i$. Substituting these weights into constraints (6)–(7) yields

$$1 = \lambda_n + \mu_n^\top \bar{X},$$
$$0 = \lambda_n \bar{X} + \left(\frac{1}{n}\sum_{i=1}^{n} X_i X_i^\top\right)\mu_n.$$

These equations are solved by $\mu_n = -M^{-1}\bar{X}$ and $\lambda_n = 1 + \bar{X}^\top M^{-1}\bar{X}$. With these substitutions we find that $\omega_{i,n}$ indeed equals $\alpha_{i,n}$ defined in (9). It is also known (see, for example, Hesterberg and Nelson 1998 or Equation (7.10) of Rao and Toutenburg 1995 in the regression setting) that with some algebraic rearrangement using (2), the LCV estimator (1) can be expressed as a weighted average,

$$\hat{Y}_{\mathrm{cv}} = \frac{1}{n}\sum_{i=1}^{n} \alpha_{i,n} Y_i,$$

using the same $\alpha_{i,n}$. Thus, the two estimators coincide in this case. □

It is worth noting that although they coincide, the two estimators in the proposition result from two distinct quadratic optimization problems. For the LCV estimator, one selects $\hat{\beta}$ to minimize the sum of squared regression residuals; for the WMC estimator, one minimizes the sum of squared weights subject to constraints. The first problem involves the Y_i, whereas the second does not.

3.2. General Convex Objective

Our analysis of general WMC estimators relies on properties of the Lagrange multipliers (λ_n, μ_n). For this we digress briefly and consider constants (λ_0, μ_0) defined by the equations

$$\mathsf{E}[H(\lambda_0 + \mu_0^\top X)] = 1, \tag{13}$$

$$\mathsf{E}[H(\lambda_0 + \mu_0^\top X)X] = 0. \tag{14}$$

At this point, H could be arbitrary so long as these equations have a unique solution. Consider the problem of estimating (λ_0, μ_0) from observations $X_1, \ldots, X_n$. A natural approach would be to define estimators (λ_n, μ_n) as roots to the sample counterparts of (13)–(14), given by

$$\frac{1}{n}\sum_{i=1}^{n} H(\lambda_n + \mu_n^\top X_i) = 1, \tag{15}$$

$$\frac{1}{n}\sum_{i=1}^{n} H(\lambda_n + \mu_n^\top X_i) X_i = 0. \tag{16}$$

Estimators defined as roots of equations in this way are often called M-estimators (see, e.g., Huber 1981); they include maximum likelihood estimators as a special case.

There is a large literature showing that under various conditions such estimators are consistent and asymptotically normal in the sense that

$$\sqrt{n}[(\lambda_n, \mu_n^\top) - (\lambda_0, \mu_0^\top)] \Rightarrow N(0, \Sigma_L) \tag{17}$$

for some $(d+1) \times (d+1)$ covariance matrix Σ_L; see, e.g., Hansen (1982), Heyde (1997), Huber (1981), or Serfling (1980).

In light of (8), Equations (15)–(16) are precisely the equations defining the Lagrange multipliers (λ_n, μ_n) through constraints (6)–(7) in the unbiased case $c_X = 0$. When H is the inverse of h' and $\mathsf{E}[X] = 0$, (13) and (14) are solved explicitly by $(\lambda_0, \mu_0) = (h'(1), 0)$. In Appendix B, we detail specific conditions ensuring (17). As these are somewhat technical and as (17) is the *typical* case, we proceed under the assumption that it holds. (See Owen 1990 for a specific analysis of the multipliers under the empirical likelihood objective.) Actually, we require a somewhat weaker property. To state it we need some notation: A sequence of random variables ξ_n is $O_p(a_n)$ if for all $\epsilon > 0$ there is a constant K such that $P(|\xi_n| \geqslant K a_n) < \epsilon$ for all sufficiently large n. The sequence is $o_p(a_n)$ if $P(|\xi_n| \geqslant a_n \epsilon) \to 0$ for all $\epsilon > 0$. We apply these symbols to random vectors if the properties hold componentwise.

Condition (A). $(\lambda_n, \mu_n^\top) = (\lambda_0, \mu_0^\top) + O_p(1/\sqrt{n})$, with (λ_0, μ_0) the unique solution to (13)–(14) and, for all sufficiently large n, (λ_n, μ_n) the unique solution to (15)–(16).

The next result verifies that the uniqueness imposed by Condition (A) is meaningful. We define a class of admissible objective functions h through the following properties:

(i) $h\colon \Re \mapsto \Re \cup \{+\infty\}$ is convex, dom(h) includes $(0, \infty)$, and h is strictly convex on dom(h);

(ii) h is continuously differentiable on dom(h) and $h'(1) < \infty$.

Lemma 1. *If Σ_X is nonsingular and h satisfies properties* (i)–(ii), *then* $(\lambda_0, \mu_0) \equiv (h'(1), 0)$ *solves* (13)–(14) *uniquely*; *and for all sufficiently large n,* (15)–(16) *are solved uniquely by some* (λ_n, μ_n).

Proof. That $(h'(1), 0)$ is a solution follows from $H(h'(1)) = 1$ and $\mathsf{E}[X] = 0$. Now suppose (λ_1, μ_1) and (λ_2, μ_2) both solve (13)–(14). Because h is convex and continuously differentiable, H is strictly increasing, and thus

$$[H(\lambda_1 + \mu_1^\top X) - H(\lambda_2 + \mu_2^\top X)] \cdot [(\lambda_1 + \mu_1^\top X) - (\lambda_2 + \mu_2^\top X)] \geqslant 0.$$

The expected value of this product is zero, in view of (13)–(14), so the product must in fact be zero almost surely. However, then $H(\lambda_1 + \mu_1^\top X) = H(\lambda_2 + \mu_2^\top X)$ almost surely, which further implies $\lambda_1 + \mu_1^\top X = \lambda_2 + \mu_2^\top X$ almost surely, because H is strictly increasing. Rewriting this as

$$(\mu_2 - \mu_1)^\top X = (\lambda_1 - \lambda_2)$$

and taking the variance of both sides, we get $(\mu_2 - \mu_1)^\top \cdot \Sigma_X(\mu_2 - \mu_1) = 0$. This implies that $(\mu_2 - \mu_1) = 0$ because Σ_X is positive definite, and this in turn implies $(\lambda_1 - \lambda_2) = 0$.

For all sufficiently large n, the relative interior of the convex hull of $X_1, \ldots, X_n$ contains the origin $(= \mathsf{E}[X])$; this follows from the central limit theorem. Thus, for all sufficiently large n, (15)–(16) has at least one solution (λ_n, μ_n). Also, for all sufficiently large n, the matrix M in (3) is nonsingular. Uniqueness now follows through the argument used for (13)–(14), replacing expected values with averages over $i = 1, \ldots, n$. □

For our main result in the unbiased setting, we strengthen (ii) above to the following condition:

(ii′) h is three times continuously differentiable on dom(h) and $0 < h''(1) < \infty$.

We prove the following result in Appendix A. We use $\|\cdot\|$ to indicate the Euclidean norm.

Theorem 1. *Suppose that h satisfies conditions* (i) *and* (ii′) *and $c_X = 0$ in* (7). *Suppose that $\mathsf{E}[\|X\|^3] < \infty$, $\mathsf{E}[|Y|^3] < \infty$, Σ_X is nonsingular, and Condition* (A) *holds. Then, $\hat{Y}_{\mathrm{WMC}} = \hat{Y}_{\mathrm{cv}} + O_p(1/n)$.*

This result makes precise the idea that the WMC and LCV estimators are very close. It also indicates that for a broad class of objective functions h, all WMC estimators are nearly the same when $c_X = 0$. In particular, as a consequence of Theorem 1 we get a central limit theorem showing that WMC estimators achieve exactly the same asymptotic variance reduction as the LCV estimator:

Corollary 1. *Under the conditions of Theorem* 1,

$$\sqrt{n}(\hat{Y}_{\mathrm{WMC}} - \mathsf{E}[Y]) \Rightarrow N(0, \sigma_Y^2(1 - R^2)).$$

Proof. It follows from Theorem 1 that $\sqrt{n}(\hat{Y}_{\mathrm{WMC}} - \hat{Y}_{\mathrm{cv}}) \Rightarrow 0$. The result now follows from (4) and Theorem 4.1 of Billingsley (1968, p. 25). □

In a different setting, Glynn and Whitt (1989) show that a large class of nonlinear control variable estimators are asymptotically equivalent to LCV estimators. The estimators they consider are nonlinear functions of sample means and thus distinct from those considered here.

3.3. Numerical Example

To illustrate Theorem 1, we test WMC estimators in a standard option-pricing problem. We apply the estimators to the pricing of a call option on the average level of an underlying asset—an Asian option. The price of the option is $\mathsf{E}[Y]$, with

$$Y = e^{-rT} \max\{0, \bar{S} - K\},$$

where the constants T, r, and K are the maturity, interest rate, and strike price. The random variable $\bar{S}$ is the average level

$$\bar{S} = \frac{1}{m} \sum_{j=1}^{m} S(t_j)$$

of the underlying asset S over a fixed set of dates $0 < t_1 < \cdots < t_m = T$. See, e.g., Hull (2000) for background.

We model the dynamics of the underlying asset using geometric Brownian motion. More explicitly, as in, e.g., Hull (2000), p. 238, we have

$$S(t) = S(0)\exp\big(\big[r - \tfrac{1}{2}\sigma^2\big]t + \sigma W(t)\big), \tag{18}$$

where $S(0)$ is a fixed initial price, σ is the asset's volatility, and W is a standard Brownian motion. To simulate $\bar{S}$, it suffices to simulate $S(t_1), \ldots, S(t_m)$ and thus to simulate W at times $t_1, \ldots, t_m$. This is accomplished by the recursion

$$W(t_i) = W(t_{i-1}) + \sqrt{t_i - t_{i-1}}Z_i,$$

with $Z_1, \ldots, Z_m$ independent standard normal random variables and $W(0) = 0$.

A consequence of (18) is that $\mathsf{E}[S(t_i)] = S(0)e^{rt_i}$; this follows from the more basic identity $\mathsf{E}[\exp(aZ)] = \exp(a^2/2)$ for the moment-generating function of a standard normal random variable Z. We may therefore take the control variable X to be a vector of $d = m$ components, with ith component equal to

$$S(t_i) - S(0)e^{rt_i}.$$

These are not necessarily the most effective controls for this problem, but our objective is to provide a simple illustration of the WMC estimators rather than to find the best way to price the option.

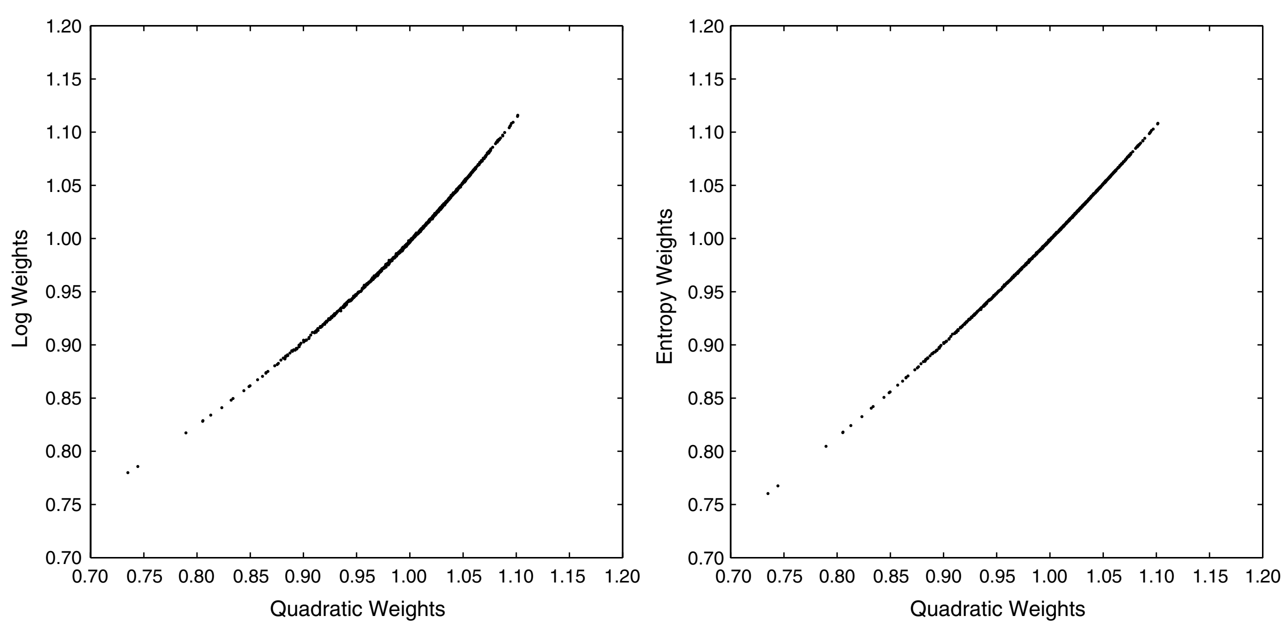
For the numerical tests we use $S(0) = 50$, $r = 5\%$, $\sigma = 0.20$, $K = 54$, and $T = 5$, with time measured in years. We consider $m = 1, 2$, and 4, the case $m = 1$ corresponding to an ordinary European rather than Asian option. We compare WMC estimators based on the quadratic, empirical likelihood (log), and entropy objectives.

As a first illustration, we compare the weights calculated using the three objectives. We generate $n = 1{,}000$ independent replications of (X_i, Y_i) with $m = 4$ and solve the optimization problem (5)–(7), with $c_X = 0$, for each of the three objective functions. In each case we get a sequence of n weights, $\omega_{1,n}, \ldots, \omega_{n,n}$.

Figure 1 shows scatter plots comparing the weights found under the three objectives using the same simulated values $X_1, \ldots, X_n$. The left panel plots the empirical likelihood weights against the quadratic weights, and the right panel plots the entropy weights against the quadratic weights. Under each objective, we get 1,000 weights—one for each replication. The coordinates of each point in the scatter plots are the weights assigned to the same replication using two different objectives. In both panels of the figure, the weights under the two objectives are strikingly close to each other, especially near the limiting value of 1. This is an even stronger relation than the one in Theorem 1, which refers only to the averages $\hat{Y}_{\mathrm{WMC}}$, and not to the individual weights themselves. From (9) we see that the quadratic weights are $1 + O_p(1/\sqrt{n})$; the analysis in Appendix A suggests (heuristically) that the differences between the weights are $O_p(1/n)$, so that they are closer to each other (under different objectives) than they are to 1. The figures are consistent with this suggestion because the points fall near the 45° line.

Next, we compare point estimates and variances of the three WMC estimators. We know from Theorem 1 and

Figure 1. Comparison of optimal weights calculated using empirical likelihood (log) and quadratic objectives (left) and entropy and quadratic objectives (right).



Corollary 1 that these should coincide at large n, so the main purpose of numerical tests is to see what happens in small samples. To estimate $\mathsf{Var}[\hat{Y}_{\mathrm{WMC}}]$ at a sample size of n, we generate 100 batches, each of n independent replications. We solve the optimization problem for each batch to get a value of $\hat{Y}_{\mathrm{WMC}}$ for each batch. We then use the sample variance over the 100 batches as our estimate of $\mathsf{Var}[\hat{Y}_{\mathrm{WMC}}]$. Through batching, we avoid having to deal with the dependence between observations in a single sample of size n introduced by the weights. (Recall that $\hat{Y}_{\mathrm{WMC}}$ is not just the average of independent replications.) The same procedure can be used to compute confidence intervals.

Results for $m = 1$, 2, and 4 controls are displayed in Table 1 for the quadratic, log, and entropy objectives, along with results for ordinary (equally weighted) Monte Carlo (MC). The variances of the WMC estimators are indeed very close, in most cases for sample sizes as small as 100. The point estimates are also close. The WMC variances are appreciably smaller than those for ordinary Monte Carlo.

To solve the optimization problem under the log and entropy objectives, we used nonlinear least squares to find Lagrange multipliers satisfying the first-order conditions (8) and (6)–(7) in the specific forms (10) and (11). At small values of n, this sometimes required restarting at multiple initial values for the Lagrange multipliers to converge to the optimum. We found that the entropy objective took about twice as long as the quadratic objective, and the log objective took a bit more than twice as long as the entropy objective. However, we did not attempt to optimize our implementation of these calculations. Owen (2001) discusses algorithmic issues for these types of problems.

4. Convergence in the Biased Case

We turn now to the biased case in which the weighted averages of the controls are constrained in (7) to equal some value $c_X \neq 0$ different from $\mathsf{E}[X]$. (We continue to assume that X has been centered so that $\mathsf{E}[X] = 0$.) Recall from §2.2 that in this setting the weighting scheme is best viewed as a mechanism for adjusting or correcting a simulation model rather than reducing variance in an estimator based on a model. The most relevant issue in this case is therefore identifying the value to which $\hat{Y}_{\mathrm{WMC}}$ converges. We will see that, in contrast to the unbiased case, here the choice of convex function h does affect the limit.

We replace (13)–(14) with

$$\mathsf{E}[H(\lambda+\mu^\top X)] = 1, \qquad \mathsf{E}[H(\lambda+\mu^\top X)X] = c_X, \qquad (19)$$

and similarly define (λ_n, μ_n) using (15)–(16), with 0 replaced by c_X in (16). We replace Condition (A) with

Condition (B). $(\lambda_n, \mu_n^\top) = (\lambda, \mu^\top) + O_p(1/\sqrt{n})$, with (λ, μ) uniquely solving (19) and, for all sufficiently large n, (λ_n, μ_n) uniquely solving (6), (7), and (8).

In the unbiased case ($c_X = 0$) the existence of solutions is essentially automatic, but here it depends on the choice of c_X and h. If dom(h) is the entire real line (as in the quadratic case), then any c_X is feasible; if dom(h) includes $[0, \infty)$, then any c_X in the interior of the support of X is feasible. Given existence of a solution, uniqueness follows through exactly the argument used in Lemma 1.

Table 1. Point estimates and variances for ordinary Monte Carlo (MC) and weighted Monte Carlo estimators using three objective functions for an Asian option with $m = 1$, 2, and 4 controls.

	Mean				Variance			
n	MC	Quadratic	Log	Entropy	MC	Quadratic	Log	Entropy
$m = 1$								
50	12.810	12.484	12.620	12.553	7.987	0.489	0.436	0.453
100	13.156	12.574	12.663	12.619	4.719	0.186	0.179	0.178
200	13.109	12.754	12.794	12.774	2.096	0.167	0.164	0.165
400	12.679	12.717	12.737	12.727	1.003	0.070	0.068	0.069
800	12.695	12.703	12.713	12.708	0.478	0.034	0.034	0.034
1,600	12.646	12.704	12.710	12.707	0.294	0.016	0.017	0.017
$m = 2$								
50	9.654	8.991	9.106	9.050	5.102	0.344	0.339	0.335
100	9.390	9.085	9.155	9.120	2.343	0.175	0.175	0.172
200	9.108	9.120	9.151	9.136	0.917	0.074	0.072	0.073
400	9.128	9.110	9.125	9.117	0.451	0.052	0.051	0.052
800	9.050	9.109	9.119	9.114	0.285	0.025	0.025	0.025
1,600	9.120	9.136	9.140	9.138	0.114	0.010	0.010	0.010
$m = 4$								
50	7.809	7.244	7.337	7.296	3.229	0.433	0.384	0.401
100	7.469	7.406	7.461	7.434	1.367	0.150	0.148	0.148
200	7.437	7.376	7.408	7.392	0.601	0.075	0.077	0.075
400	7.271	7.368	7.387	7.377	0.396	0.045	0.046	0.046
800	7.373	7.395	7.404	7.400	0.170	0.019	0.019	0.019
1,600	7.413	7.392	7.397	7.394	0.091	0.010	0.010	0.010

Our next result shows that, under appropriate conditions, the WMC estimator $\hat{Y}_{\mathrm{WMC}}$ converges in probability to

$$c_Y = \mathsf{E}[H(\lambda + \mu^\top X)Y], \tag{20}$$

and is asymptotically normal. For this we need some further notation. Define

$$d_\lambda = \mathsf{E}[H'(\lambda + \mu^\top X)Y], \qquad d_\mu = \mathsf{E}[H'(\lambda + \mu^\top X)XY].$$

We will require that these be finite. Also, set

$$C_H = \mathsf{E}[H'(\lambda + \mu^\top X)XX^\top] - \frac{\mathsf{E}[H'(\lambda + \mu^\top X)X]\,\mathsf{E}[H'(\lambda + \mu^\top X)X]^\top}{\mathsf{E}[H'(\lambda + \mu^\top X)]}; \tag{21}$$

we will require that this $d \times d$ matrix be finite and nonsingular. The following result is proved in Appendix A.

Theorem 2. *Suppose that h is strictly convex and three times continuously differentiable on* dom(h), $\mathsf{E}[\|X\|^3] < \infty$ *and* $\mathsf{E}[|Y|^3] < \infty$, *and Σ_X is nonsingular. Suppose that d_λ and d_μ are finite, C_H is finite and nonsingular, and $H(\lambda + \mu^\top X)$ and $H(\lambda + \mu^\top X)X$ are square integrable. Suppose Condition* (B) *holds. Then, $\hat{Y}_{\mathrm{WMC}}$ converges to c_Y in probability and*

$$\sqrt{n}[\hat{Y}_{\mathrm{WMC}} - c_Y] \Rightarrow N(0, \sigma_H^2),$$

with $\sigma_H^2 < \infty$.

An expression for the limiting variance parameter σ_H^2 is derived in the proof of Theorem 2. Though it involves quantities that would typically be unknown, it can be estimated from the simulated data. It differs from

$$\sigma_W^2 = \mathsf{Var}[H(\lambda + \mu^\top X)Y],$$

the variance parameter that would apply to estimation of c_Y using i.i.d. replications if λ and μ were known. In fact, we will show that

$$\begin{aligned}\sqrt{n}(\hat{Y}_{\mathrm{WMC}} - c_Y) &= \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} H(\lambda + \mu^\top X_i)Y_i - c_Y\right) \\ &\quad + \sqrt{n}(\lambda_n - \lambda)d_\lambda + \sqrt{n}(\mu_n - \mu)^\top d_\mu + o_p(1).\end{aligned} \tag{22}$$

The first term on the right is asymptotically normal with variance σ_W^2, and this representation shows that $\hat{Y}_{\mathrm{WMC}}$ is asymptotically equivalent to applying the Lagrange multipliers as LCV to the first term.

Theorem 2 reveals a resemblance between WMC and importance sampling. The value c_Y in (20) to which $\hat{Y}_{\mathrm{WMC}}$ converges looks like the expectation of an importance-sampling estimator if we interpret $H(\lambda + \mu^\top X)$ as a likelihood ratio. This interpretation is supported by the first equation in (19)—the expected value of a likelihood ratio equals 1—but must be qualified by the fact that H may take negative values.

Expressions for the *sensitivity* of c_Y to c_X are derived in Avellaneda and Gamba (2000) for general convex objectives h. These involve covariance terms weighted by H' similar to (21). Like Theorem 2, these expressions show the effect of different choices of objective.

We now revisit the specific choices of h introduced in §2.1.

Quadratic objective. In this case, (19) can be solved explicitly to get $\lambda = 1$ and $\mu = \Sigma_X^{-1} c_X$, with Σ_X the covariance matrix of X, the WMC estimator converges to

$$\begin{aligned}c_Y &= \mathsf{E}[(1 + c_X^\top \Sigma_X^{-1} X)Y] \\ &= \mathsf{E}[Y] + \beta^\top c_X,\end{aligned}$$

with $\beta = \Sigma_X^{-1}\Sigma_{XY}$. Thus, c_Y is the value fitted at c_X by the least squares regression line. If $\mathsf{E}[Y \mid X = x]$ is linear in x (for example, if (X, Y) is multivariate normal) then $c_Y = \mathsf{E}[Y \mid X = c_X]$.

Entropy objective. Define the cumulant generating function of X by $\psi_X(\theta) = \log \mathsf{E}[\exp(\theta^\top X)]$, $\theta \in \Re^d$, and let $\nabla\psi_X$ denote its gradient. Then, (19) reduces to solving $\nabla\psi_X(-\mu) = c_X$ and then setting $\lambda = \psi_X(-\mu) - 1$. Define a new probability measure $P_{-\mu}$ and expectation operator $\mathsf{E}_{-\mu}$ through the likelihood ratio $\exp(-\mu^\top X + \psi_X(-\mu))$. In particular, this means that under the new measure, Y has distribution

$$P_{-\mu}(Y \leqslant y) = \mathsf{E}\big[\mathbf{1}\{Y \leqslant y\} e^{-\mu^\top X + \psi_X(-\mu)}\big],$$

with $\mathbf{1}\{\cdot\}$ the indicator of the event in braces. The WMC estimator converges to

$$c_Y = \mathsf{E}\big[Y e^{-\mu^\top X + \psi_X(-\mu)}\big] = \mathsf{E}_{-\mu}[Y].$$

Here the connection with importance sampling is explicit: c_Y is the expected value of Y under an exponential change of measure defined by X, a standard transformation in importance sampling. (See, e.g., Asmussen 1987.) This value of c_Y has an interpretation as the limit of $\mathsf{E}[\bar{Y} \mid \bar{X} = c_X]$ as the number of replications n increases; see Example 8 of Zabell (1980).

Empirical likelihood objective. In this case we get

$$c_Y = \mathsf{E}\left[\left(\frac{1}{\lambda + \mu^\top X}\right)Y\right],$$

with

$$\mathsf{E}[1/(\lambda + \mu^\top X)] = 1, \qquad \mathsf{E}[X/(\lambda + \mu^\top X)] = c_X,$$

assuming these expectations are well defined and finite. The weight $1/(\lambda + \mu^\top X)$ could take negative values. This case emerges as an approximation to the previous one through the approximation $\exp(-x) \approx 1/(1 + x)$.

4.1. Numerical Example

We illustrate the use of biased WMC in an application to model calibration as suggested by Avellaneda and colleagues (Avellaneda 1998, Avellaneda et al. 2001). (For numerical examples applying the method to estimate conditional expectations in pricing American options, see Broadie et al. 2000.) We consider the problem of adjusting an imperfect option-pricing model to match a set of market prices and applying the adjusted model to value a different option for which no market price is available. The calibration aspect of this problem lies in ensuring consistency with the observed market prices.

We consider a simple option-pricing example that contains the essential features of more complex and more realistic problems that motivate the approach of Avellaneda and colleagues (Avellaneda 1998, Avellaneda et al. 2001). Our starting point is the Black-Scholes model (as in, e.g., Hull 2000), but we suppose that the observed market prices of options are incompatible with this model. The goal is to price another option (for which no market price is available) in a way that is consistent with observed market prices. This requires imposing some adjustment or correction to a simulation of the Black-Scholes model. The observed market prices of options provide the WMC constraints.

In more detail, we consider options on an underlying asset initially at $S(0) = 50$. We assume an interest rate of $r = 5\%$ and a maturity of $T = 5$ years. We suppose that we observe the market prices of four call options at this maturity, differing in their strike prices; the strikes we use are

$$(K_1, K_2, K_3, K_4) = (0.85, 0.95, 1.05, 1.15)e^{rT}S(0)$$
$$= (54.57, 60.99, 67.41, 73.83).$$

Suppose that the observed market prices of these four options are

$$(q_1, q_2, q_3, q_4) = (13.97, 9.93, 7.07, 5.15).$$

Suppose that simulation is to be used to price a "deep out-of-the-money" option with a strike price of $K = 1.4e^{rT}S(0) = 89.88$, relative to the other options for which market prices are assumed known. The simulated model for the underlying asset price S is geometric Brownian motion, as in (18), with the same parameters as in §3.3. We set $Y = e^{-rT}\max\{0, S(T) - K\}$.

The simulated model is incompatible with the observed market prices because the simulated model is consistent with the Black-Scholes formula but the market prices are not. Option prices computed by simulation will converge to Black-Scholes prices for $S(0) = 50$, $r = 5\%$, $T = 5$, $\sigma = 0.20$, and $K_1, \ldots, K_4$ as above. These Black-Scholes prices are

$$(p_1, p_2, p_3, p_4) = (12.46, 9.93, 7.87, 6.21),$$

and do not match the market prices listed above. The inconsistency results from the fact that the simulated model (geometric Brownian motion) is not an exact representation of the market. Indeed, there is no value of σ (the only adjustable parameter) that will make all four Black-Scholes prices match the four market prices.

In the approach of Avellaneda et al. (2001) (and related methods of Stutzer 1996 and Buchen and Kelly 1996), we use weights to adjust the simulated model to match the market prices and then use the weighted value rather than $\mathsf{E}[Y]$ as the price for the new option. Thus, we take the vector of controls X to have four components corresponding to the four given option prices; the ith component of X is the difference

$$e^{-rT}\max\{0, S(T) - K_i\} - p_i,$$

with p_i the Black-Scholes price given above. (Subtracting p_i gives the control variable an expected value of 0.) For c_X in constraint (7), we take a vector of four components in which the ith component is the difference $q_i - p_i$ between the market and model prices of the ith option. Constraining the weighted average of replications of X to equal c_X may thus be interpreted as calibrating the simulation to the market prices.

In this discussion, we have subtracted the model prices p_i in X and c_X to be consistent with the formulation in (5)–(7), where we assume $\mathsf{E}[X] = 0$. The expectations cancel, so in practice we omit the expectations from both X and c_X. Indeed, it is not even necessary that the model prices p_i be known.

Table 2 shows results under the quadratic, log, and entropy objectives at various sample sizes n. For each n, we generate 500 batches and calculate the mean and standard error (SE) over these batches, as in §3.3. The standard errors in the table are very similar for the three objectives, indicating that their variances are very close, though this is not guaranteed by Theorem 2. More importantly, the limiting means are appreciably different for the three objectives; a difference of more than 3% (as in the quadratic and log cases) could be financially as well as statistically significant.

The analysis in this paper does not provide grounds for preferring any one of the three price estimates in Table 2 over the others. It does, however, clarify the implications of the choice of objective function. The entropy objective has received particular attention, but a compelling reason for

Table 2. Comparison of WMC estimates of an option price based on calibration to four other prices.

	Quadratic		Log		Entropy	
n	Mean	SE	Mean	SE	Mean	SE
100	2.25	0.012	2.36	0.010	2.30	0.010
200	2.28	0.008	2.38	0.007	2.33	0.007
400	2.28	0.006	2.38	0.005	2.33	0.005
800	2.28	0.004	2.37	0.004	2.32	0.004
1,600	2.28	0.003	2.37	0.003	2.33	0.002
10,000	2.27	0.001	2.37	0.001	2.32	0.001

focusing on this case in practice remains to be established. The fact that it produces positive weights is appealing, as are some of its information-theoretic interpretations, but these features must be balanced against the computational advantage of the quadratic objective.

5. Concluding Remarks

This article establishes two main results on the large sample properties of WMC estimators. The two results address the biased and unbiased formulations of WMC estimators. We have shown that in the unbiased case, a large class of WMC estimators are nearly identical to LCV estimators and thus achieve exactly the same asymptotic variance reduction. In the biased case, we have identified the limit to which WMC estimators converge and we have shown that—in contrast to the unbiased case—here the choice of objective function affects the large sample properties. We have illustrated these ideas with applications and numerical examples.

Our main result for the unbiased case indicates that, at least in large samples, there is no advantage to considering WMC estimators other than the usual LCV estimator: The asymptotic variance reduction achieved is the same, and the LCV estimator is computationally less demanding and its statistical properties are better understood. There may still be advantages to the WMC formulation of LCV estimators (as in the proof of Proposition 1). If, for example, the same controls are to be applied to estimating multiple quantities, the weights need to be computed just once. See also the application to quantile estimation in Hesterberg and Nelson (1998).

For the biased case, we have shown how the choice of objective function affects the model adjustment imposed through WMC: The inverse of the derivative of the objective appears in the limit of the WMC estimator as a type of weighting function. Our results do not provide a basis for preferring one objective to another without further criteria. Different choices of objective correspond to different ways of extrapolating from constraints on control variables (e.g., observed market prices) to quantities estimated by simulation (e.g., prices of securities for which no market value is available).

Our main result for the biased case reveals a link between WMC and importance sampling, and this suggests a direction for further investigation. In some applications where one might want to use importance sampling, finding an appropriate change of distribution or the associated likelihood ratio may be difficult. A potential alternative, then, is to use weights determined through a constrained optimization problem in place of a genuine likelihood ratio. This, in fact, is the motivation for using weights in Broadie et al. (2000), but the technique may have broader applicability.

Appendix A. Convergence Proofs

Proof of Theorem 1. Recalling the first-order conditions (8) and the convergence of λ_n to λ_0 and μ_n to 0, we use a Taylor approximation to H to write the optimal weights as

$$\omega_{i,n} = H(\lambda_0) + H'(\delta_{i,n})((\lambda_n - \lambda_0) + \mu_n^\top X_i),$$

where

$$\delta_{i,n} = \lambda_0 + \kappa_{i,n}((\lambda_n - \lambda_0) + \mu_n^\top X_i)$$

for some $0 \leqslant \kappa_{i,n} \leqslant 1$. We use this and $H(\lambda_0) = 1$ to write the scaled difference between the two estimators as follows:

$$\begin{aligned} n(\hat{Y}_{\text{WMC}} - \hat{Y}_{\text{cv}}) &= \sum_{i=1}^n (\omega_{i,n} - \alpha_{i,n}) Y_i \\ &= \sum_{i=1}^n H(\lambda_0) Y_i + \sum_{i=1}^n H'(\delta_{i,n})((\lambda_n - \lambda_0) + \mu_n^\top X_i) Y_i \\ &\quad - \sum_{i=1}^n Y_i - \sum_{i=1}^n \bar{X}^\top M^{-1} (\bar{X} - X_i) Y_i \\ &= (\lambda_n - \lambda_0) \sum_{i=1}^n H'(\delta_{i,n}) Y_i + \mu_n^\top \sum_{i=1}^n H'(\delta_{i,n}) X_i Y_i \\ &\quad - \sum_{i=1}^n \bar{X}^\top M^{-1} (\bar{X} - X_i) Y_i. \end{aligned} \tag{23}$$

We analyze each of these terms through a series of lemmas, starting with the following useful result proved in Owen (1990, p. 98):

Lemma 2. *If $V_1, V_2, \ldots$ are i.i.d. with finite second moment, then* $\max_{1\leqslant i\leqslant n} |V_i| = o(\sqrt{n})$, *with probability* 1.

We will remove the $\delta_{i,n}$ from (23) using the following lemma.

Lemma 3. *For $Z_i = Y_i$, $X_i Y_i$, $X_i X_i^\top$, or* 1,

$$\left\| \sum_{i=1}^n \left(H'(\delta_{i,n}) - \frac{1}{h''(1)} \right) Z_i \right\| = O_p(n^{1/2}),$$

with $\|\cdot\|$ denoting the usual Euclidean norm in the vector and matrix cases and absolute value in the scalar case.

Proof. First note that $H'(\lambda_0) = 1/h''(1)$. The condition that h is C^3 implies that H is C^2 in a neighborhood of λ_0. Applying a Taylor approximation to H' we get, for some $\nu_{i,n} = \lambda_0 + \tau_{i,n}(\delta_{i,n} - \lambda_0)$, $0 \leqslant \tau_{i,n} \leqslant 1$,

$$\begin{aligned} &\left\| \sum_{i=1}^n \left(H'(\delta_{i,n}) - \frac{1}{h''(1)} \right) Z_i \right\| \\ &= \left\| \sum_{i=1}^n H''(\nu_{i,n}) \left(\delta_{i,n} - \lambda_0 \right) Z_i \right\| \\ &\leqslant \max_{1\leqslant i\leqslant n} |H''(\nu_{i,n})| \sum_{i=1}^n \left(|\lambda_n - \lambda_0| + |\mu_n^\top X_i| \right) \|Z_i\| \\ &\leqslant \max_{1\leqslant i\leqslant n} |H''(\nu_{i,n})| \left(|\lambda_n - \lambda_0| \sum_{i=1}^n \|Z_i\| + \|\mu_n\| \sum_{i=1}^n \|X_i\| \|Z_i\| \right) \\ &= \max_{1\leqslant i\leqslant n} |H''(\nu_{i,n})| \left(O_p(n^{-1/2}) O(n) + O_p(n^{-1/2}) O(n) \right). \end{aligned}$$

The inequality follows from

$$|\nu_{i,n}-\lambda_0|\leqslant|\delta_{i,n}-\lambda_0|\leqslant|\lambda_n+\mu_n^\top X_i-\lambda_0|, \tag{24}$$

and the order symbols follow from Condition (A) and the strong law of large numbers. (Hölder's inequality and our moment conditions on X and Y ensure that $\mathsf{E}[\|X_i\|\|Z_i\|]<\infty$ for each case of Z_i.) The lemma will be proved once we show that

$$\max_{1\leqslant i\leqslant n}|H''(\nu_{i,n})|=O_p(1). \tag{25}$$

Again, using (24) and applying the triangle inequality, the Cauchy-Schwarz inequality, and Lemma 2, we find that

$$\begin{aligned}\max_{1\leqslant i\leqslant n}|\nu_{i,n}-\lambda_0| &\leqslant \max_{1\leqslant i\leqslant n}|\lambda_n+\mu_n^\top X_i-\lambda_0|\\ &\leqslant|\lambda_n-\lambda_0|+\max_{1\leqslant i\leqslant n}|\mu_n^\top X_i|\\ &\leqslant|\lambda_n-\lambda_0|+\|\mu_n^\top\|\max_{1\leqslant i\leqslant n}\|X_i\|\\ &=O_p(n^{-1/2})+O_p(n^{-1/2})o(\sqrt{n})\\ &=o_p(1).\end{aligned} \tag{26}$$

The continuity of H'' in a neighborhood of λ_0 implies that for any sufficiently small $\varepsilon>0$, we may define a finite K by setting

$$K=\max_{|v-\lambda_0|\leqslant\varepsilon}|H''(v)|.$$

Then,

$$P\Big(\max_{1\leqslant i\leqslant n}|H''(\nu_{i,n})|>K\Big)\leqslant P\Big(\max_{1\leqslant i\leqslant n}|\nu_{i,n}-\lambda_0|>\varepsilon\Big)$$

which, by (26), is less then ε for all sufficiently large n. This verifies (25). □

Combining (23) and Lemma 3, we arrive at the representation

$$\begin{aligned}n(\hat{Y}_{\text{WMC}}-\hat{Y}_{\text{cv}})&=\frac{(\lambda_n-\lambda_0)}{h''(1)}\sum_{i=1}^n Y_i+\frac{\mu_n^\top}{h''(1)}\sum_{i=1}^n X_iY_i\\ &\quad-\sum_{i=1}^n\bar{X}^\top M^{-1}(\bar{X}-X_i)Y_i+O_p(1).\end{aligned} \tag{27}$$

To deal with the first term in (27), we strengthen part of Condition (A):

Lemma 4. $\lambda_n-\lambda_0=O_p(1/n)$.

Proof. From the constraint (6), we get

$$\begin{aligned}1&=\frac{1}{n}\sum_{i=1}^n H(\lambda_n+\mu_n^\top X_i)\\ &=\frac{1}{n}\sum_{i=1}^n H(\lambda_0)+\frac{1}{n}\sum_{i=1}^n H'(\delta_{i,n})((\lambda_n-\lambda_0)+\mu_n^\top X_i).\end{aligned}$$

However, $H(\lambda_0)=1$, so using Lemma 3 we get

$$\begin{aligned}0&=\frac{\lambda_n-\lambda_0}{n}\sum_{i=1}^n H'(\delta_{i,n})+\frac{\mu_n^\top}{n}\sum_{i=1}^n H'(\delta_{i,n})X_i\\ &=\frac{\lambda_n-\lambda_0}{n}\left(\frac{n}{h''(1)}+O_p(n^{1/2})\right)\\ &\quad+\frac{\mu_n^\top}{n}\left(\frac{1}{h''(1)}\sum_{i=1}^n X_i+O_p(n^{1/2})\right),\end{aligned} \tag{28}$$

and then by Condition (A) and the central limit theorem for $\bar{X}$,

$$\begin{aligned}\lambda_n-\lambda_0&=-\mu_n^\top\bar{X}+O_p(1/n)\\ &=O_p(n^{-1/2})O_p(n^{-1/2})+O_p(1/n)=O_p(1/n).\end{aligned} \quad □$$

For the other Lagrange multiplier, we have

Lemma 5. $\mu_n=-h''(1)M^{-1}\bar{X}+O_p(1/n)$.

Proof. From the constraint (7), we get

$$\begin{aligned}0&=\frac{1}{n}\sum_{i=1}^n H(\lambda_n+\mu_n^\top X_i)X_i\\ &=\frac{1}{n}\left(\sum_{i=1}^n H(\lambda_0)X_i+\sum_{i=1}^n X_iH'(\delta_{i,n})((\lambda_n-\lambda_0)+\mu_n^\top X_i)\right)\\ &=\bar{X}+\frac{\lambda_n-\lambda_0}{n}\sum_{i=1}^n X_iH'(\delta_{i,n})+\sum_{i=1}^n H'(\delta_{i,n})X_iX_i^\top\frac{\mu_n}{n}\\ &=\bar{X}+\frac{\lambda_n-\lambda_0}{nh''(1)}\sum_{i=1}^n X_i+\frac{1}{nh''(1)}\sum_{i=1}^n X_iX_i^\top\mu_n+O_p(1/n),\end{aligned} \tag{29}$$

the last equality following from Lemma 3 and Condition (A). Using Lemma 4, this becomes

$$0=\bar{X}+\frac{1}{nh''(1)}\sum_{i=1}^n X_iX_i^\top\mu_n+O_p(1/n),$$

and then because $\bar{X}=O_p(n^{-1/2})$, we also have

$$0=\bar{X}+\frac{1}{h''(1)}\left(\frac{1}{n}\sum_{i=1}^n X_iX_i^\top-\bar{X}\bar{X}^\top\right)\mu_n+O_p(1/n).$$

The matrix multiplying μ_n in this expression is M. □

We can now conclude the proof of the theorem by showing that (27) is $O_p(1)$. Using Lemma 4 and the strong law of large numbers, we find that the first term in (27) is

$$\frac{(\lambda_n-\lambda_0)}{h''(1)}\sum_{i=1}^n Y_i=O_p(1/n)O(n)=O_p(1).$$

Using Lemma 5 and the symmetry of M, the second term in (27) becomes

$$\begin{aligned}\frac{\mu_n^\top}{h''(1)}\sum_{i=1}^n X_iY_i&=-\bar{X}^\top M^{-1}\sum_{i=1}^n X_iY_i+O_p(1/n)\sum_{i=1}^n X_iY_i\\ &=-\bar{X}^\top M^{-1}\sum_{i=1}^n X_iY_i+O_p(1).\end{aligned}$$

By adding all terms in (27), we therefore get

$$n(\hat{Y}_{\text{WMC}} - \hat{Y}_{\text{cv}}) = n\bar{X}^\top M^{-1}\bar{X}\bar{Y} + O_p(1) = O_p(1),$$

the last step following from the central limit theorem for $\bar{X}$ and the convergence of M^{-1} to Σ_X^{-1}.

Proof of Theorem 2. As in the proof of Theorem 1, we use (8) and a Taylor approximation to write

$$\omega_{i,n} = H(\lambda+\mu^\top X_i) + H'(\delta_{i,n})((\lambda_n+\mu_n^\top X_i) - (\lambda+\mu^\top X_i)),$$

with

$$\delta_{i,n} = (\lambda+\mu^\top X_i) + \theta_{i,n}((\lambda_n+\mu_n^\top X_i) - (\lambda+\mu^\top X_i))$$

for some $0 \leqslant \theta_{i,n} \leqslant 1$. We therefore have

$$\begin{aligned}
&\sqrt{n}(\hat{Y}_{\text{WMC}} - c_Y) \\
&= \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n H(\lambda+\mu^\top X_i)Y_i - c_Y\right) \\
&\quad + \sqrt{n}(\lambda_n-\lambda)\frac{1}{n}\sum_{i=1}^n \big(H'(\delta_{i,n}) - H'(\lambda+\mu^\top X_i)\big)Y_i \\
&\quad + \sqrt{n}(\lambda_n-\lambda)\frac{1}{n}\sum_{i=1}^n H'(\lambda+\mu^\top X_i)Y_i \\
&\quad + \sqrt{n}(\mu_n-\mu)^\top\frac{1}{n}\sum_{i=1}^n \big(H'(\delta_{i,n}) - H'(\lambda+\mu^\top X_i)\big)X_iY_i \\
&\quad + \sqrt{n}(\mu_n-\mu)^\top\frac{1}{n}\sum_{i=1}^n H'(\lambda+\mu^\top X_i)X_iY_i \\
&= \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n H(\lambda+\mu^\top X_i)Y_i - c_Y\right) \\
&\quad + \sqrt{n}(\lambda_n-\lambda)\frac{1}{n}\sum_{i=1}^n H'(\lambda+\mu^\top X_i)Y_i \\
&\quad + \sqrt{n}(\mu_n-\mu)^\top\frac{1}{n}\sum_{i=1}^n H'(\lambda+\mu^\top X_i)X_iY_i + o_p(1). \qquad (30)
\end{aligned}$$

The last equality follows from Condition (B) and the following counterpart to Lemma 3:

Lemma 6. *For $Z_i = Y_i,\ X_iY_i,\ X_iX_i^\top$, or 1,*

$$\left\|\sum_{i=1}^n \big(H'(\delta_{i,n}) - H'(\lambda+\mu^\top X_i)\big)Z_i\right\| = O_p(n^{1/2}).$$

The proof of this result is exactly the same as that of Lemma 3, with λ_0 replaced by λ and μ_n replaced by $\mu_n - \mu$.

Applying the law of large numbers to the last two sums in (30) yields (22). Under Condition (B), this proves that $\hat{Y}_{\text{WMC}} \Rightarrow c_Y$. To get the asymptotic normality of the estimator, (22) shows that we need to establish a joint central limit theorem for

$$W_n = \frac{1}{n}\sum_{i=1}^n H(\lambda+\mu^\top X_i)Y_i - c_Y, \qquad (31)$$

$\lambda_n - \lambda$, and $\mu_n - \mu$. We therefore focus on the Lagrange multipliers next.

Lemma 7.

$$\frac{1}{n}\sum_{i=1}^n (H(\lambda+\mu^\top X_i) - 1) + \frac{(\lambda_n-\lambda)}{n}\sum_{i=1}^n H'(\lambda+\mu^\top X_i) + \frac{1}{n}\sum_{i=1}^n H'(\lambda+\mu^\top X_i)X_i^\top(\mu_n-\mu) = O_p(1/n),$$

$$\frac{1}{n}\sum_{i=1}^n (H(\lambda+\mu^\top X_i)X_i - c_X) + \frac{(\lambda_n-\lambda)}{n}\sum_{i=1}^n H'(\lambda+\mu^\top X_i)X_i + \frac{1}{n}\sum_{i=1}^n H'(\lambda+\mu^\top X_i)X_iX_i^\top(\mu_n-\mu) = O_p(1/n).$$

Proof. The proof is very similar to those given for Lemmas 4 and 5 in the unbiased case. The first equation results from constraint (6), followed by Taylor expansion of $H(\lambda_n + \mu_n^\top X_i)$ around $H(\lambda + \mu^\top X_i)$, just as in (28). The second equation similarly results from constraint (7), just as in (29). In both cases, the Taylor approximation involves evaluating $H'(\delta_{i,n})$ for some point $\delta_{i,n}$ on the line segment joining $\lambda + \mu^\top X_i$ and $\lambda_n + \mu_n^\top X_i$. These terms get replaced by terms of the form $H'(\lambda + \mu^\top X_i)$ through Lemma 6. □

Define random variables L_n, S_n, random d-vectors U_n, V_n, and random $d \times d$ matrices A_n to rewrite the two equations in Lemma 7 as

$$L_n + (\lambda_n - \lambda)S_n + V_n^\top(\mu_n - \mu) = O_p(1/n),$$
$$U_n + (\lambda_n - \lambda)V_n + A_n(\mu_n - \mu) = O_p(1/n),$$

and observe that each is an average of i.i.d. random elements. Some algebra yields

$$\left(A_n - \frac{1}{S_n}V_nV_n^\top\right)(\mu_n - \mu_0) = \frac{L_n}{S_n}V_n - U_n + O_p(1/n).$$

By the strong law of large numbers,

$$S_n \to s \equiv \mathsf{E}[H'(\lambda+\mu^\top X)],$$
$$V_n \to v \equiv \mathsf{E}[H'(\lambda+\mu^\top X)X],$$

and

$$\left(A_n - \frac{1}{S_n}V_nV_n^\top\right) \to C_H,$$

with probability 1. By the central limit theorem,

$$\left(A_n - \frac{1}{S_n}V_nV_n^\top\right) = C_H + O_p(n^{-1/2}).$$

Because C_H is nonsingular, the approximating matrices are nonsingular for all sufficiently large n. We may therefore solve the equations above to get

$$\sqrt{n}(\mu_n - \mu) = \sqrt{n}C_H^{-1}\left(\frac{L_n}{S_n}V_n - U_n\right) + o_p(1)$$

and

$$\sqrt{n}(\lambda_n-\lambda) = \sqrt{n}\left(-\frac{L_n}{S_n}-\frac{1}{S_n}V_n^\top C_H^{-1}\left(\frac{L_n}{S_n}V_n-U_n\right)\right)+o_p(1).$$

The components of the vector $(W_n, L_n, U_n^\top)$ are i.i.d. averages of functions of the $(X_i^\top, Y_i)$, $i=1,\ldots,n$, with mean zero and finite second moments, so we have

$$\sqrt{n}(W_n, L_n, U_n^\top)\Rightarrow(W, L, U^\top),$$

with $(W, L, U^\top)$ multivariate normal. This limiting random vector has mean zero and its covariance matrix is the covariance matrix of

$$(H(\lambda+\mu^\top X)Y, H'(\lambda+\mu^\top X), H'(\lambda+\mu^\top X)X^\top).$$

Write this $(d+2)\times(d+2)$ covariance matrix in block form as

$$\begin{pmatrix}\sigma_W^2 & \sigma_{WL} & \Sigma_{WU}\\ \sigma_{LW} & \sigma_L^2 & \Sigma_{LU}\\ \Sigma_{UW} & \Sigma_{UL} & \Sigma_{UU}\end{pmatrix},$$

where, for example, σ_W^2 and σ_L^2 are scalars and Σ_{UU} is $d\times d$. Because $S_n\to s$ and $V_n\to v$, $\lambda_n-\lambda$ and $\mu_n-\mu$ are asymptotically linear transformations of $(W, L, U^\top)$, and

$$\sqrt{n}(W, \lambda_n-\lambda, \mu_n^\top-\mu^\top)\Rightarrow(W, La+U^\top b, Lb^\top-U^\top C_H^{-1}),$$

with

$$a=-\frac{1}{s}-\frac{1}{s^2}v^\top C_H^{-1}v, \qquad b=\frac{1}{s}C_H^{-1}v.$$

Finally, from (22), we see that $\sqrt{n}(\hat{Y}_{\mathrm{WMC}}-c_Y)$ is asymptotically a linear combination of W, $\lambda_n-\lambda$, and $\mu_n-\mu$ and converges in distribution to

$$W+(ad_\lambda+b^\top d_\mu)L+(d_\lambda b^\top-d_\mu^\top C_H^{-1})U.$$

This limit is normal with mean zero and variance

$$\begin{aligned}\sigma_H^2 &= \sigma_W^2+(ad_\lambda+b^\top d_\mu)^2\Sigma_{LL}\\ &\quad+(d_\lambda b^\top-d_\mu^\top C_H^{-1})\Sigma_{UU}(d_\lambda b-C_H^{-1}d_\mu)\\ &\quad+2(ad_\lambda+b^\top d_\mu)\Sigma_{WL}+2\Sigma_{WU}(d_\lambda b-C_H^{-1}d_\mu)\\ &\quad+2(ad_\lambda+b^\top d_\mu)\Sigma_{LU}(d_\lambda b-C_H^{-1}d_\mu).\end{aligned}$$

Appendix B. Convergence of the Multipliers

This appendix provides sufficient conditions for the Lagrange multipliers (λ_n, μ_n) to converge at rate $O_p(n^{-1/2})$; in fact, these conditions also imply that they are asymptotically normal. The conditions are based on Hansen (1982). This appendix is self-contained in the sense that all required conditions are stated explicitly, even those introduced earlier in the paper.

(a-i) $(X, Y), (X_1, Y_1), (X_2, Y_2), \ldots$ are i.i.d., with $X\in\Re^d$, $Y\in\Re$, $\mathsf{E}[X]=0$, $\mathsf{E}[\|X\|^2]<\infty$, and $\mathsf{E}[Y^2]<\infty$.

(a-ii) h is convex and twice continuously differentiable on $\mathrm{dom}(h)$, with $0<h''(1)<\infty$; there exists a continuously differentiable function H on the range of h' such that $H(h'(x))=x$ for all x in $\mathrm{dom}(h)$.

(a-iii) There is a bounded open set $S\subset\Re^{d+1}$ such that $(\lambda_n, \mu_n^\top)\in S$ for all sufficiently large n, with probability 1.

Define $f: \Re\times\Re^d\times\Re^d\to\Re^{d+1}$ via

$$f(\lambda, \mu, x)=\begin{pmatrix}H(\lambda+\mu^\top x)\\ H(\lambda+\mu^\top x)x\end{pmatrix}.$$

(a-iv) $\mathsf{E}[f(\lambda, \mu, X)]$ exists and is finite for all pairs $(\lambda, \mu^\top)\in S$.

(a-v) $(\lambda_*, \mu_*^\top)\in S$ uniquely solves $\mathsf{E}[f(\lambda_*, \mu_*, X)]^\top=(1, c_X^\top)$ (where c_X may or may not be the zero vector), and $\mathsf{E}[\|f(\lambda_*, \mu_*, X)\|^2]<\infty$.

For $(\lambda, \mu)\in S$ and $\delta>0$, define

$$\epsilon_1(\lambda, \mu, \delta)=\sup\{\|f(\lambda', \mu', X)-f(\lambda, \mu, X)\|: \|(\lambda', \mu')-(\lambda, \mu)\|<\delta\}.$$

We require

(a-vi) For every $(\lambda, \mu)\in S$, there is some $\delta>0$ for which $\mathsf{E}[\epsilon_1(\lambda, \mu, \delta)]<\infty$.

Define the derivative map

$$Df(\lambda, \mu, x)=\begin{pmatrix}H'(\lambda+\mu^\top x) & H'(\lambda+\mu^\top x)x^\top\\ H'(\lambda+\mu^\top x)x & H'(\lambda+\mu^\top x)xx^\top\end{pmatrix}$$

and define $\epsilon_2(\lambda, \mu, \delta)$ by replacing f with Df in the definition of ϵ_1.

(a-vii) $\mathsf{E}[Df(\lambda_*, \mu_*, X)]$ exists, is finite, and has full rank.

(a-viii) For every $(\lambda, \mu)\in S$, there is some $\delta>0$ for which $\mathsf{E}[\epsilon_2(\lambda, \mu, \delta)]<\infty$.

In the unbiased case $c_X=0$, condition (a-vii) simplifies to a full-rank requirement for Σ_X, the covariance matrix of X, because in this case $(\lambda_*, \mu_*)=(h'(1), 0)$ and

$$Df(\lambda_*, \mu_*, X)=\frac{1}{h''(1)}\begin{pmatrix}1 & x^\top\\ x & xx^\top\end{pmatrix},$$

so

$$\mathsf{E}[Df(\lambda_*, \mu_*, X)]=\frac{1}{h''(1)}\begin{pmatrix}1 & 0\\ 0 & \Sigma_X\end{pmatrix}.$$

More generally, the condition reduces to the requirement that the matrix C_H in (21) be finite and nonsingular. We now have

Proposition 2. *If Conditions* (a-i)–(a-viii) *hold, then* $(\lambda_n, \mu_n^\top)\to(\lambda_*, \mu_*^\top)$ *with probability* 1, *and* $\sqrt{n}[(\lambda_n, \mu_n^\top)-(\lambda_*, \mu_*^\top)]$ *converges in distribution to a multivariate normal. In particular,* $(\lambda_n, \mu_n^\top)=(\lambda_*, \mu_*^\top)+O_p(n^{-1/2})$.

Proof. Conditions (a-i)–(a-viii) imply the assumptions of Theorem 2.1 of Hansen (1982), establishing almost sure convergence, and they imply the assumptions of Theorem 3.1 of Hansen (1982) ensuring asymptotic normality. □

Acknowledgments

The authors thank Alexander Shapiro for pointing out the relevance of Huber's book (1981) in analyzing the Lagrange multipliers. Marco Avellaneda, Art Owen, and two anonymous reviewers provided additional references and useful comments on an earlier version of this paper. This work was supported in part by NSF grant DMS0074637.

References

Asmussen, S. 1987. *Applied Probability and Queues*. Wiley, Chichester, UK.

Avellaneda, M. 1998. Minimum-relative-entropy calibration of asset-pricing models. *Internat. J. Theoret. Appl. Finance* **1** 447–472.

Avellaneda, M., R. Gamba. 2000. Conquering the Greeks in Monte Carlo: Efficient calculation of the market sensitivities and hedge-ratios of financial assets by direct numerical simulation. M. Avellaneda, ed. *Quantitative Analysis in Financial Markets*, Vol. III. World Scientific, Singapore, 336–356.

Avellaneda, M., R. Buff, C. Friedman, N. Grandchamp, L. Kruk, J. Newman. 2001. Weighted Monte Carlo: A new technique for calibrating asset-pricing models. *Internat. J. Theoret. Appl. Finance* **4** 1–29.

Baggerly, K. A. 1998. Empirical likelihood as a goodness-of-fit measure. *Biometrika* **85** 535–547.

Billingsley, P. 1968. *Convergence of Probability Measures*. Wiley, New York.

Broadie, M., P. Glasserman, Z. Ha. 2000. Pricing American options by simulation using a stochastic mesh with optimized weights. S. Uryasev, ed. *Probabilistic Constrained Optimization: Methodology and Applications*. Kluwer, Norwell, MA, 32–50.

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

Fu, M. C., D. Madan, T. Wang. 1999. Pricing continuous Asian options: A comparison of Monte Carlo and Laplace transform inversion methods. *J. Comput. Finance* **2**(Winter) 49–74.

Glynn, P. W., W. Whitt. 1989. Indirect estimation via $L = \lambda W$. *Oper. Res.* **37** 82–103.

Hansen, L. P. 1982. Large sample properties of generalized method of moments estimators. *Econometrica* **50** 1029–1054.

Hesterberg, T. C., B. L. Nelson. 1998. Control variates for probability and quantile estimation. *Management Sci.* **44** 1295–1312.

Heyde, C. C. 1997. *Quasi-Likelihood and Its Applications*. Springer, New York.

Huber, P. J. 1981. *Robust Statistics*. Wiley, New York.

Hull, J. 2000. *Options, Futures, and Other Derivative Securities*. Prentice-Hall, Upper Saddle River, NJ.

Jourdain, B., L. Nguyen. 2001. Minimisation de l'entropie relative par méthode de Monte-Carlo. *C.R. Acad. Sci. Paris* **332**(Série 1) 345–350.

Marshall, A. W., I. Olkin. 1979. *Inequalities: Theory of Majorization and Its Applications*. Academic Press, New York.

Nelson, B. L. 1990. Control variate remedies. *Oper. Res.* **38** 974–992.

Owen, A. B. 1990. Empirical likelihood ratio confidence regions. *Ann. Statist.* **18** 90–120.

Owen, A. B. 2001. *Empirical Likelihood*. Chapman & Hall, London, UK.

Rao, C. R., H. Toutenburg. 1995. *Linear Models: Least Squares and Alternatives*. Springer-Verlag, New York.

Rockafellar, R. T. 1970. *Convex Analysis*. Princeton University Press, Princeton, NJ.

Schmeiser, B. W., M. R. Taaffe, J. Wang. 2001. Biased control-variate estimation. *IIE Trans.* **33** 219–228.

Serfling, R. J. 1980. *Approximation Theorems of Mathematical Statistics*. Wiley, New York.

Stutzer, M. 1996. A simple nonparametric approach to valuing derivative securities. *J. Finance* **51** 1633–1652.

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

Tsitsiklis, J., B. Van Roy. 2001. Regression methods for pricing complex American-style options. *IEEE Trans. Neural Networks* **12** 694–703.

Zabell, S. L. 1980. Rates of convergence for conditional expectations. *Ann. Probab.* **8** 928–941.