Proceedings of the 2015 Winter Simulation Conference
L. Yilmaz, W. K. V. Chan, I. Moon, T. M. K. Roeder, C. Macal, and M. D. Rossetti, eds.

CONTROL VARIATE METHODS FOR PERFORMANCE EVALUATION OF HEURISTIC INVENTORY CONTROL POLICIES

Sechan Oh

IBM Research
1101 Route 134 Kitchawan Rd
Yorktown Heights, NY 10598, USA

ABSTRACT

Development of heuristic policies is a common solution approach for stochastic inventory control problems that are computationally intractable. In the inventory control literature, Monte Carlo simulation has been widely used to measure the performance of heuristic policies. In this paper, we propose control variate methods for the estimation of heuristic inventory control policies. Our methods construct control variates using the optimal control policies of relaxed optimization problems. We apply the methods to two inventory control problems. Numerical experiments demonstrate the effectiveness of our methods.

1 INTRODUCTION

Stochastic inventory control theory, which is concerned with how to procure, assemble, distribute, and price inventories under demand uncertainty, is a central research topic in operations management. Early research on stochastic inventory control theory has shown that the optimal inventory control policy in single-product, single-location systems has a simple structure under increasingly general settings (Arrow, Harris, and Marchak 1951, Veinott 1965). In those problems, the optimal control parameters and the resulting performance metrics such as expected inventory costs are often given in simple formulas.

Stochastic inventory control problems, however, are computationally intractable when there are multiple products in the system or when the supply chain consists of multiple locations. In such cases, little is known about the structure of the optimal policy. Due to the large dimension of the state space, one cannot numerically compute the optimal solutions and the resulting performance metrics. Inventory control problems in distribution systems (Federgruen and Zipkin 1984) and assemble-to-order systems (Doğru, Reiman, and Wang 2010), and joint pricing and inventory control problems with a positive procurement leadtime (Pang, Chen, and Feng 2012) are some of the examples that are known to be intractable.

Developing heuristic policies is a common solution approach for intractable stochastic inventory problems. There are two common approaches to demonstrate the performance of heuristic policies. The first approach is to provide analytic bounds on performance metrics by focusing on asymptotic behaviors. The second approach is to measure the performance of a heuristic policy via Monte Carlo simulation. By comparing the performance metrics under the heuristic policy and other reference policies, one can show how well the heuristic policy performs.

This paper proposes control variate methods for the estimation of the performance of heuristic inventory control policies. Our methods construct control variates using the reference control policies to which heuristic policies are compared. If a heuristic policy were to perform well, the decisions made under this policy should not be too different from the decisions made under reference policies. Our methods construct control variates using the similarity between the decisions made under heuristic policies and those under reference policies.

978-1-4673-9743-8/15/$31.00 ©2015 IEEE

In order for a statistic to be used as a control variate, its expected value should be known without simulation. In stochastic inventory control problems, a widely used method to define a reference control policy for performance measure is to obtain a tractable problem by relaxing certain constraints from the original optimization problem (Gallego, Özer, and Zipkin 2007). In many cases, the optimal policy of a relaxed problem and the resulting performance metrics are given in functional forms, and thus their expected values can be computed without simulation.

In this paper, we revisit two stochastic inventory control problems studied in the literature. The first problem is the classical inventory control problem in distribution systems. The second problem is a joint pricing and inventory control problem in an assemble-to-order system. For each problem, we introduce a heuristic control policy and a relaxed optimization problem proposed by Federgruen and Zipkin (1984) and Oh, Sourirajan, and Ettl (2014), and apply our control variate methods. Numerical results are provided to show that the control variate methods significantly reduce the variances of estimators. Under the same model parameters used for the numerical results in Federgruen and Zipkin (1984) and Oh, Sourirajan, and Ettl (2014), our methods reduce the variances of the estimators by at least 94%.

2 INVENTORY CONTROL IN DISTRIBUTION SYSTEMS

In this section, we consider the classical inventory control problem in distribution systems. Among many papers on this topic, we use the model and heuristic policy of Federgruen and Zipkin (1984) for our study.

2.1 Problem Formulation

Consider a stationary, periodic-review distribution system consisting of one central warehouse and J identical retailers at which random demands for a single product need to be fulfilled. At each period, a central decision maker places a procurement order for the product. This order arrives at the warehouse after L periods, and is distributed to the retailers immediately. Allocated units to each retailer reach them after l periods. Unmet demands are backlogged, and incur penalty costs. Inventories incur holding costs. The central decision maker's objective is to minimize the average inventory holding and shortage costs.

The sequence of events and notations are as follows. At the beginning of each period t, the central decision maker orders y_t units of the product, and receives the items ordered at period $t-L$, i.e., y_{t-L}. The central decision maker then determines how to distribute the y_{t-L} units of products to the J retailers. The amounts allocated to each retailer at period t are denoted by $z_t = (z_{t,1}, , z_{t,2}, \dots, z_{t,J})$. Next, the retailers receive the units allocated to them at period $t-l$, i.e., z_{t-l}. The net inventory levels at the retailers before receiving the last shipment z_{t-l} are denoted by $x_t = (x_{t,1}, x_{t,2}, \dots, x_{t,J})$. Finally, demands for the product $D_t = (D_{t,1}, D_{t,2}, \dots, D_{t,J})$ are realized, and unmet demands are backlogged. An inventory holding cost h is incurred for each unit of inventory, and a penalty cost p is incurred for each unit of backlogged demand. The demand at each retailer is normally distributed with the mean value of μ and the standard deviation of σ. Demands at different retailers and different time periods are independent.

This problem can be formulated as a dynamic program. To determine how much inventory to allocate to each retailer, the decision maker needs to know how much inventory is being held at or being shipped to each retailer, i.e., the inventory position at each retailer $\hat{x}_{t,j} = x_{t,j} + \sum_{k=1}^{l} z_{t-k,j}$. To make a procurement decision, the decision maker needs to know the inventory positions at each retailer and the items that can be distributed to them in each of the upcoming periods, i.e., $\tilde{y}_t = (y_{t-1}, y_{t-2}, \dots, y_{t-L})$. In other words, there are $J+L$ state variables $\hat{x}_t$ and $\tilde{y}_t$. Let $\hat{D}_{t,j} = \sum_{k=0}^{l} D_{t+k,j}$. Then, the Bellman equation of the dynamic program is given as

$$f(\hat{x}_t, \tilde{y}_t) = \min_{y_t, z_t} \left\{ \sum_{j=1}^{J} \left(hE[\hat{x}_{t,j} + z_{t,j} - \hat{D}_{t,j}]^+ + pE[\hat{D}_{t,j} - \hat{x}_{t,j} - z_{t,j}]^+ \right) + E[f(\hat{x}_{t+1}, \tilde{y}_{t+1})] \right\} \tag{1}$$

$$s.t. \quad y_t \geq 0, \quad \sum_{j=1}^{J} z_{t,j} = y_{t-L}, \quad z_t \geq 0, \quad \hat{x}_{t+1} = \hat{x}_t + z_t - D_t.$$

Clark and Scarf (1960) have shown that the optimal policy if exists is very complex. Because of the curse of dimensionality of dynamic programming, this problem is also not computationally tractable.

2.2 Lower Bound Problem and Heuristic Policy

Federgruen and Zipkin (1984) propose a heuristic inventory control policy. In dosing so, they first obtain a tractable lower bound of (1) by relaxing a set of constraints. In (1), the amount of inventory allocated to each retailer should be non-negative, i.e., $z_t \geq 0$. Relaxing these constraints dramatically simplifies the problem. The optimal allocation decision is to set z_t such that the inventory position is the same for every retailer, i.e., $\hat{x}_{t,j} + z_{t,j} = (\sum_{k=1}^{J} \hat{x}_{t,k} + y_{t-L})/J$ for every j. This property reduces (1) to the problem of a single location inventory procurement problem. The leadtime demand of the relaxed problem is normally distributed with the mean value of $\mu J(L+l+1)$ and the standard deviation of $\sigma\sqrt{JL+J^2(l+1)}$. The optimal procurement policy of this relaxed problem is the well-known base-stock policy with the base-stock level of $\mu J(L+l+1) + \Phi^{-1}\left(\frac{p}{p+h}\right)\sigma\sqrt{JL+J^2(l+1)}$, where $\Phi(\cdot)$ is the c.d.f. of the standard normal distribution. The stationary distribution of x_t is also known in a functional form. We refer the reader to Federgruen and Zipkin (1984) and Gallego, Özer, and Zipkin (2007) for further details of this lower bound problem.

Using the optimal control parameter of the lower bound problem, Federgruen and Zipkin (1984) develop the heuristic policy. Inventory procurement decisions are made following the base-stock policy of the lower bound problem. Inventory allocation decisions are made to minimize instantaneous inventory costs;

$$\begin{aligned} \min_{z_t} \quad & \sum_{j=1}^{J} \left(hE[\hat{x}_{t,j} + z_{t,j} - \hat{D}_{t,j}]^+ + pE[\hat{D}_{t,j} - \hat{x}_{t,j} - z_{t,j}]^+ \right) \\ s.t. \quad & \sum_{j=1}^{J} z_{t,j} = y_{t-L}, \quad z_t \geq 0. \end{aligned} \tag{2}$$

The optimal solution of (2) is to divide y_{t-L} such that $\hat{x}_{t,j} + z_{t,j}$ is the same for all j to the extent possible given the constraint $z_t \geq 0$. When the inventory position $\hat{x}_{t,j}$ at a certain retailer is much larger than those at other retailers, no item will be allocated to the retailer, i.e., $z_{t,j} = 0$. The optimal z_t of the heuristic policy coincides with that of the lower bound problem if the difference among the inventory positions at different retailers are not too large.

2.3 Control Variates

The performance of the proposed heuristic policy can be measured by the difference between the expected cost under this policy and the optimal expected cost. Although one can easily solve (2) given realized demands, exact computation of the expected cost under the heuristic policy requires to repeat the task under all possible realizations of demands, which is not computationally tractable. Instead, one can estimate such performance metrics via Monte Carlo simulation.

We define x_t^H as the net inventory levels under the heuristic policy. Suppose that we need to estimate $E[g(x_t^H)]$ for a certain function $g : \mathbb{R}^{\mathbb{J}} \to \mathbb{R}$. We denote N independent samples of x_t^H by $x_t^{H,1}$, $x_t^{H,2}$, …, $x_t^{H,N}$. Then,

$$\widehat{g(x_t^H)}_{CMC} = \frac{1}{N} \sum_{n=1}^{N} g(x_t^{H,n}) \tag{3}$$

is the Crude Monte Carlo (CMC) estimator of $E[g(x_t^H)]$.

Next we construct a control variate for this estimator. Suppose that under the same realized demands, one system operates using the heuristic policy, which we call (**H**), and another system operates using the optimal policy of the lower bound problem, which we call (**B**). Although (**B**) is not a feasible policy in practice, one can numerically run this policy. Let x_t^H and x_t^B denote the inventory levels under (**H**) and (**B**), and let $(x_t^{H,n}, x_t^{B,n})$ for $n \in \{1,2,\ldots,N\}$ be independent samples of (x_t^H, x_t^B), where each pair of $x_t^{H,n}$ and $x_t^{B,n}$ are drawn from common realized demands. Then the following is a control variate estimator of $E[g(x_t^H)]$:

$$\widehat{g(x_t^H)}_{CV} = \frac{1}{N}\sum_{n=1}^{N}\left[g(x_t^{H,n}) - g(x_t^{B,n}) + E[g(x_t^B)]\right]. \tag{4}$$

Note that (4) is an unbiased estimator of $E[g(x_t^H)]$ because $E[g(x_t^{B,n}) - E[g(x_t^B)]] = 0$. Because the distribution of x_t^B is known in a functional form, one can easily compute $E[g(x_t^B)]$ without simulation. The inventory procurement decisions under (**H**) and (**B**) are the same, and the allocation decisions under (**B**) can be computed via a few additions and a division. Thus, the computational efforts for (3) and (4) are almost the same.

The degree of variance reduction that can be achieved by the control variate depends on the correlation between $g(x_t^H)$ and $g(x_t^B)$. The fact that the procurement orders made under the two policies are the same implies that the total amount of items that are received and distributed at the warehouse in each period is identical under the two policies. Although the exact amount allocated to each retailer can be different in the two cases, under both policies the retailer who fulfilled a larger demand in the previous period gets a larger amount of units. This observation implies that the net inventory levels, which determine costs, are similar under the two policies.

The following theorem formalizes the relation between the decisions made under the two policies.

Theorem 1 The following properties hold:
1. $Prob(\sum_{j=1}^J x_{t,j}^H = \sum_{j=1}^J x_{t,j}^B) = 1$.
2. If $\hat{x}_{t-1}^H + z_{t-1}^H = \hat{x}_{t-1}^B + z_{t-1}^B$, then $Prob(z_t^H = z_t^B) = Prob(\sum_{j=1}^J D_{t-L,j} \geq \sum_{j=1}^J D_{t-1,j} - J\min_k D_{t-1,k})$.
3. Both $z_{t,j}^H$ and $z_{t,j}^B$ increase in $D_{t-1,j}$, and decrease in $D_{t-1,k}$ for every $k \neq j$.

Proof. Part 1 is trivial because the procurement decisions made under (**H**) and (**B**) are always the same and $\sum_{j=1}^J z_{t,j} = y_{t-L}$ holds under both policies. We prove part 2. Note that the allocation decisions under the two policies are to minimize the instantaneous inventory costs under the constraints that $\sum_{j=1}^J z_{t,j} = y_{t-L}$. The difference is that under (**H**), additional constraints that $z_{t,j} \geq 0$ for every j need to be satisfied. Hence, under the same states $z_{t,j}^B = z_{t,j}^H$ holds if and only if $z_{t,j}^B \geq 0$ is satisfied for every j. Note that we have $Jz_{t,j}^B = \sum_{k=1}^J(\hat{x}_{t,k}^B - \hat{x}_{t,j}^B) + y_{t-L}^B = \sum_{k=1}^J(\hat{x}_{t-1,k}^B + z_{t-1,k}^B - D_{t-1,k} - \hat{x}_{t-1,j}^B + z_{t-1,j}^B + D_{t-1,j}) + y_{t-L}^B$. Because $\hat{x}_{t-1,j}^B + z_{t-1,j}^B$ is the same for every j and $y_{t-L}^B = \sum_{j=k}^J D_{t-L,k}$, we have $Jz_{t,j}^B = \sum_{k=1}^J(D_{t-1,j} - D_{t-1,k}) + \sum_{k=1}^J D_{t-L,j} = JD_{t-1,j} - \sum_{k=1}^J D_{t-1,k} + \sum_{k=1}^J D_{t-L,j}$. Hence, $z_{t,j}^B \geq 0$ holds for every j if and only $J\min_j D_{t-1,j} - \sum_{k=1}^J D_{t-1,k} + \sum_{k=1}^J D_{t-L,j} \geq 0$, which concludes the proof of part 2. For part 3, first note that $Jz_{t,j}^B = \sum_{k=1}^J(\hat{x}_{t-1,k}^B + z_{t-1,k}^B - D_{t-1,k} - \hat{x}_{t-1,j}^B + z_{t-1,j}^B + D_{t-1,j}) + y_{t-L}^B$, which increases in $D_{t-1,j}$ and decreases in $D_{t-1,k}$ for $k \neq j$. Similarly, the allocation decisions under (**H**), i.e., z_t^H, are to divide y_{t-L} such that $\hat{x}_{t,j} + z_{t,j} = \hat{x}_{t-1,j} + z_{t-1,j} - D_{t,j} + z_{t,j}$ is the same for all j to the extent possible given the constraint $z_t \geq 0$. Hence, $z_{t,j}$ increases in $D_{t,j}$ and decreases in $D_{t,k}$ for $k \neq j$. □

Part 1 of the theorem shows that the total net inventory levels at all retailers are always the same under the two policies. Part 2 shows the probability that the allocation decisions under the two policies coincide given that the states in the previous period are the same. In the probability function, $D_{t-L,j}$ and $D_{t-1,j}$ are i.i.d. random variables. The right-hand-side of the inequality is the sum of the difference between the demand at each retailer and the lowest demand at period $t-1$, i.e., $D_{t-1,j} - \min_k D_{t-1,k}$, whose probability distribution does not depend on μ. The left-hand-side of the inequality is the total demand at period $t-L$, which stochastically increases in μ. Hence, as the coefficient of variation of demands decreases, the probability that the two policies make the same allocation decisions converges to 1. Part 3 shows that even when the allocation decisions under the two policies are not the same, the impact of realized demands on allocation decisions has the same direction under the two policies.

2.4 Numerical Experiments

To show the impact of the proposed control variate method, we apply it to the estimation of two performance metrics of the heuristic policy. The first performance metric is the optimality gap. Let $c(x_t) = \sum_{j=1}^{J} h \max\{0, x_t\} + p \max\{0, -x_t\}$, which is the total holding and shortage cost at period t. Because $E[c(x_t^B)]$ is a lower bound of the optimal expected cost,

$$\frac{E[c(x_t^H)] - E[c(x_t^B)]}{E[c(x_t^B)]} \tag{5}$$

is an upper bound of the true optimality gap. We simply call (5) the optimality gap.

Under various model parameter sets, we estimate $E[c(x_t^H)]$ via both the crude Monte Carlo estimator (3) and the control variate estimator (4), and compare the variances of the two estimators. Then, we compute the optimality gap using the control variate estimator and the exact value of $E[c(x_t^B)]$. The set of model parameters used for the experiments are as follows: $L = 5$, $l = 5$, $p = 10$, $h = 1$, $\mu = 10$, $\sigma = \{2, 3, \ldots, 10\}$, and $J = \{5, 6, \ldots, 10\}$. We modified the base parameters used by Federgruen and Zipkin (1984) because the optimality gap under those parameters are too small, and the variances are reduced by more than 99% by our control variate. More specifically, we increased the leadtimes L and l, the demand uncertainty σ, and the number of retailers J so that the decisions made under the two policies deviate often. We use one million independent samples to obtain the estimators.

Figure 1 shows the percentage of the variance reduced by the control variate along with the optimality gap in each experiment. In all experiments, the variance of the control variate estimator is at least 94% smaller than that of the crude Monte Carlo estimator. As the optimality gap becomes smaller, the variance of the control variate estimator converges zero. When the optimality gap is zero, the decisions under both policies always coincide, and thus the estimator is equal to $E[c(x_t^B)]$ regardless of the sample size.

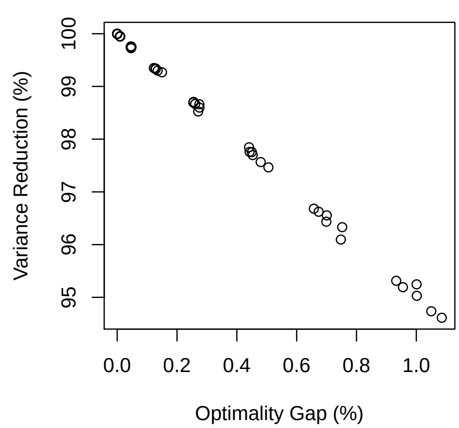

Figure 1: Impact of control variate on cost estimation.

Next we apply the control variate method to the second performance metric; type-2 service level. Type-2 service level, which is defined as the ratio of fulfilled demands to all demands, is given as

$$b(x_t) = 1 - \frac{\sum_{j=1}^{J} \max\{0, -x_t\}}{E[\sum_{j=1}^{J} D_{t-1}^{j}]}. \tag{6}$$

Using the same set of model parameters used in the previous experiment, we estimate $E[b(x_t^H)]$ via the crude Monte Carlo estimator and the control variate estimator, and compare their variances.

Figure 2 shows percentage of the variance reduced by the control variate and the service level in each experiment. As before, the control variate significantly reduces the variance of the estimator. Even when

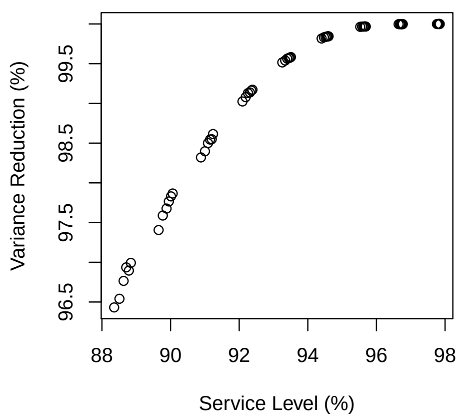

Figure 2: Impact of control variate on service level estimation.

the service level is below 90%, i.e., when only 90% of demands are filled on time, the control variate reduces more than 96% of the variance of the estimator. Hence, one can expect that our control variate performs extremely well under realistic model settings.

Considering the fact that the additional computational efforts required for the control variate is negligible, the proposed estimator can significantly reduce the time required to attain a given confidence level of the estimator. Although performance metrics similar to (5) and (6) are widely used in the inventory literature, the use of common samples and control variates for variance reduction has not been reported in this area. One potential reason for this oversight is the fact that the optimal expected cost under lower bound problems such as $E[c(x_t^B)]$ can be computed without simulation. Without the variance reduction benefit, one does not need to estimate $E[c(x_t^B)]$ separately.

3 JOINT PRICING AND INVENTORY CONTROL IN ASSEMBLE-TO-ORDER SYSTEMS

In this section, we consider the joint pricing and inventory control problem studied by Oh, Sourirajan, and Ettl (2014). For the sake of brevity, we provide a brief introduction of the problem and the heuristic policy. We then propose two control variates, and show their impacts via numerical experiments.

3.1 Formulation and Heuristic Policy

Consider an assemble-to-order system consisting of I components and J products. For one unit of product j, $B^{i,j}$ units of component i are required for each i. At the beginning of each period t, a decision maker places component replenishment orders y_t, and determines the prices of the products, p_t, which affect the random demands for the products $D_t(p_t)$. At the end of period t, the components ordered at the beginning of period t are received, and the decision maker determines how many units of products to produce, z_t. Unmet demands are backlogged, and incur unit shortage costs b. There are unit procurement costs c, unit assembly costs a, and unit holding costs h. The levels of component inventories at the beginning of period t are denoted by x_t, and the amounts of backlogged product demands are denoted by s_t.

This problem can be formulated as a dynamic program, which has two decision points in each period. At the end of period t, component allocation decisions are made;

$$\begin{aligned} g_t(u_t, D_t) \quad &= \quad \max_{z_t} \left\{ v_{t+1}(u_t - Bz_t, D_t - z_t) - a'z_t - h'(u_t - Bz_t) - b'(D_t - z_t) \right\} \\ &s.t. \quad Bz_t \le u_t, \quad 0 \le z_t \le D_t, \end{aligned} \tag{7}$$

and at the beginning of period t, component replenishment and pricing decisions are made;

$$v_t(x_t, s_t) \quad = \quad \max_{y_t \ge 0, p_t} \left\{ E[p_t{}'D_t(p_t)] - c'y_t + E[g_t(y_t + x_t, D_t(p_t) + s_t)] \right\}. \tag{8}$$

This dynamic program has $I + J$ state variables, and the optimal policy does not have a simple form. Oh, Sourirajan, and Ettl (2014) proposed a heuristic policy for this problem. They first obtained an upper bound of the optimal expected profit by relaxing the constraints $z_t \ge 0$ in (7) and $y_t \ge 0$ in (8). These relaxation decouples the dynamic program into T two-stage stochastic programs. The first-stage problem of the decoupled stochastic programs is a convex optimization problem for joint pricing and procurement decisions, and the second stage problem is a linear program for component allocation decisions. These problems can be solved via numerical methods.

The heuristic policy is also based on this relaxation idea. The inventory procurement policy is an independent base-stock policy, where the base-stock levels are obtained from the decoupled stochastic programs. If the inventory levels are below the base-stock levels, then the prices are set at the list prices, which are the optimal prices of the decoupled problem, and otherwise the prices are determined by a deterministic convex optimization problem. Finally, the inventory allocation decisions are made such that instantaneous inventory costs are minimized. We refer the reader to Oh, Sourirajan, and Ettl (2014) for further details about the heuristic policy.

3.2 Control Variates

We propose two control variates for the estimation of the expected profit under the heuristic policy. The first control variate is based on the upper bound problem, and is similar to (4). The second control variate also uses the upper bound problem, but further modifies it to reduce computational efforts.

We define q_T as the random vector that contains the history of all decision and state variables and realized demands from period 1 to period T. Then, the total expected profit is given as

$$\pi_T(q_T) = \sum_{t=1}^{T} p_t' D_t(p_t) - c'y_t - a'z_t - h'(y_t + x_t - Bz_t) - b'(D_t(p_t) - z_t). \tag{9}$$

We denote q_T under the heuristic policy by q_T^H, and N independent samples of q_T^H by $q_T^{H,1}$, $q_T^{H,2}$, ..., $q_T^{H,N}$. Then, the following is the crude Monte Carlo estimator of $E[\pi_T(q_T^H)]$:

$$\widehat{\pi_T(q_T^H)}_{CMC} = \frac{1}{N} \sum_{n=1}^{N} \pi_T(q_T^{H,n}). \tag{10}$$

As in the distribution system, the decisions made under the heuristic policy, which we call (**H**), and the optimal decisions of the upper bound problem, which we call (**B**), are similar. Under both policies, inventories are ordered up to the base-stock levels to the extent possible, and inventory allocation decisions are made to minimize instantaneous inventory costs. The same prices are offered if the inventory levels are below the base-stock levels. Hence, the total profit under (**B**) can be used as a control variate. We denote q_T under (**B**) that share the common demand realizations with q_T^H by q_T^B. Then the following is the first control variate estimator of $E[\pi_T(q_T^H)]$:

$$\widehat{\pi_T(q_T^H)}_{CV1} = \frac{1}{N}\sum_{n=1}^{N}\left[\pi_T(q_T^{H,n}) - \pi_T(q_T^{B,n}) + E[\pi_T(q_T^B)]\right]. \tag{11}$$

In the case of distribution systems, the decisions under the lower bound problem can be computed by a few additions and a division. Hence, the additional computational efforts for the control variate are absolutely negligible. In contrast, one needs to solve a linear program for each period to compute the component allocation decisions under (**B**). In the worst case, the computation time for (11) can be twice as much as that for (10) for a given sample size.

We propose another control variate, which requires negligible computational efforts. Note that the profit function (9) consists of revenue, procurement costs, assembly costs, holding costs, and shortage costs. Among the five, revenue and procurement costs are independent of component allocation z_t. Regardless of z_t, the prices are always set at fixed prices p_t^B, and ordering quantities are always equal to the realized component demands in the previous period under (**B**). Hence, the gross profit, i.e., revenue minus procurement costs, under (**B**) can be computed without any information on z_t. We define

$$\hat{\pi}_T^B = \sum_{t=1}^{T} p_t^{B\prime} D_t(p_t^B) - c' B D_t(p_t^B), \tag{12}$$

which indicates the total gross profit under (**B**). Then, the following is the second control variate estimator of $E[\pi_T(q_T^H)]$:

$$\widehat{\pi_T(q_T^H)}_{CV2} = \frac{1}{N}\sum_{n=1}^{N}\left[\pi_T(q_T^{H,n}) - \hat{\pi}_T^{B,n} + E[\hat{\pi}_T^B]\right]. \tag{13}$$

Note that the gross profit (12) does not include inventory holding and shortage costs. Hence, the variance of the estimator reduced by (13) can be limited when the holding and shortage costs are large compared to the gross profit. Although the variance reduction benefit of the second control variate can be smaller than that of the first control variate, the gross profit control variate requires negligible computational efforts.

3.3 Numerical Experiments

We conduct numerical experiments to assess the values of the control variates. We estimate $E[\pi_T(q_T^H)]$ using the three estimators (10), (11), and (13), and compared their variances. Oh, Sourirajan, and Ettl (2014) measured the optimality gap of the heuristic policy under 147 model parameter sets. We use the same setting for this study. We refer the reader to the paper for the details of these settings.

Figure 3 first shows the amount of the variance reduced by the first control variate (11) along with the optimality gap of the heuristic policy. In all 147 problem settings, the variance of the control variate estimator is at least 96% smaller than that of the crude Monte Carlo estimator. As in the case of distribution systems, the impact of this control variate gets stronger as the optimality gap decreases. When the optimality gap is small, there is a high correlation between the profit under the heuristic policy and the optimal profit of the upper bound problem.

Figure 4 shows the amount of the variance reduced by the second control variate (13). For this result, among the 147 model parameter sets we only use the 49 sets that are constructed by changing the holding

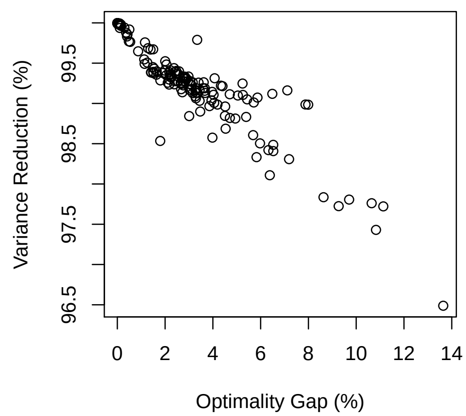

Figure 3: Impact of upper bound porift based control variate on profit estimation.

and shortage costs of the base set. Overall, the variance of the crude Monte Carlo estimator is reduced by 45% - 80% using the gross profit based control variate. The effectiveness of the control variate gets stronger as the ratio of holding costs to procurement costs gets smaller. In the experiment setting, shortage costs also increase proportionally as holding costs increase. When the holding and shortage costs are small, the difference between the net profit and the gross profit are small. Hence, the correlation between the profit under the heuristic policy and the gross profit of the upper bound problem is higher, resulting in a larger variance reduction.

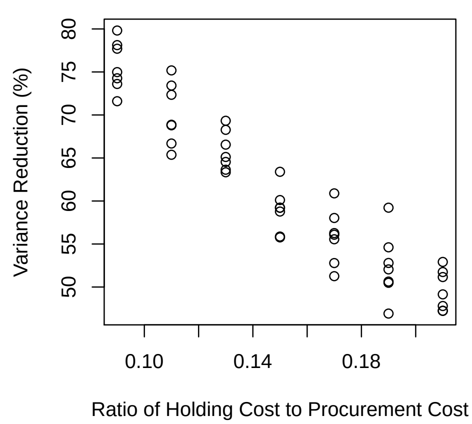

Figure 4: Impact of gross porift based control variate on profit estimation.

Figures 3 and 4 also show that the upper bound profit based control variate (11) greatly outperforms the gross profit based control variate i.e., (13). However, (13) is extremely easy to implement, whereas the implementation of (11) involves solving linear programs. The time for computing (11) for a given sample size can be as much as twice of the time for computing (13) for the same sample size.

4 CONCLUSION

In this paper, we have proposed control variate methods for the evaluation of heuristic policies in stochastic inventory control problems. If the optimality gap of a heuristic policy is small, i.e., if a heuristic policy is a good control policy, the decisions under the heuristic policy and those under the optimal policy must be small. We construct the control variates using this similarity. Although we applied the control variate methods to two inventory control problems, the method can also be applied to general Markov decision processes whose optimal solution is not known.

REFERENCES

Arrow, K. J., T. Harris, and J. Marchak. 1951. “Optimal inventory policy”. *Econometrica* 19:250–272.

Clark, A., and H. Scarf. 1960. “Optimal Policies for a Multi-Echelon Inventory Problem”. *Management Sci.* 6 (4): 475–490.

Doğru, M. K., M. I. Reiman, and Q. Wang. 2010. “A Stochastic Programming Based Inventory Policy for Assemble-to-Order Systems with Application to the W Model”. *Oper. Res.* 58 (4): 849–864.

Federgruen, A., and P. Zipkin. 1984. “Approximations of Dynamic, Multilocation Production and Inventory Problems”. *Management Sci.* 30 (1): 69–84.

Gallego, G., O. Özer, and P. Zipkin. 2007. “Bounds, Heuristics, and Approximations for Distribution Systems”. *Oper. Res.* 55 (3): 503–517.

Oh, S., K. Sourirajan, and M. Ettl. 2014. “Joint Pricing and Production Decisions in an Assemble-to-Order System”. *Manufacturing Service Oper. Management* 16 (4): 529–543.

Pang, Z., F. Y. Chen, and Y. Feng. 2012. “A note on the sturcture of joint inventory-pricing control with leadtimes”. *Oper. Res.* 60 (3): 581–587.

Veinott, A. 1965. “Optimal policy for a multi-product, dynamic, non-stationary inventory problem”. *Management Sci.* 12:206–222.

AUTHOR BIOGRAPHIES

SECHAN OH is a research scientist in the Industries and Solutions Department at IBM T.J. Watson Research Center. Prior to joining IBM, he did a Ph.D. in Management Science and Engineering at Stanford University. His research interests lie in modeling and analysis of business optimization problems with a current focus on pricing and supply chain management problems. His email address is seoh@us.ibm.com.