

Information Relaxation-Based Lower Bounds for The Stochastic Lot Sizing Problem with Advanced Demand Information

Awi Federgruen
Graduate School of Business
Columbia University
af7@columbia.edu

Daniel Guetta
Graduate School of Business
Columbia University
guetta@cantab.net

Garud Iyengar
Industrial Engineering & Operations Research
Columbia University
garud@ieor.columbia.edu

October 11, 2015

Abstract

We consider a periodic-review stochastic inventory control system consisting of a single retailer which faces random demands for a given item, and in which demand forecasts are dynamically updated. Replenishment orders are subject to a fixed and variable costs. We develop a general approach for finding lower bounds on the cost of such systems using an information relaxation. We test our approach in a model with advance demand information, and obtain good lower bounds over a range of problem parameters.

1 Introduction

Every inventory planning model which addresses demand risks builds on a representation of the anticipated future demand process. As such, the area of inventory planning is intimately connected with that of forecasting. For an inventory model to be usable, at least in a direct sense, it is therefore necessary to adopt a demand process with the same fundamental properties as the underlying forecast process.

Almost all forecast systems recognize that demands are typically correlated over time, often with demand volumes extending their impact onto future demands over a significant time lag. This gives rise to time series models such as AR and ARMA processes. Other systems employ econometric models in which demand volumes are explained by exogenous factors, such as economic indicators,

interest rates, commodity prices, exchange rates etc. . . , all of which evolve dynamically according to a separate dynamic process, for example a Markov process. In other settings, one deals with new products or new competitive dynamics requiring the need to learn about process parameters and adapt initial estimates in a Bayesian manner. In many systems, the forecasts themselves evolve dynamically. Heath and Jackson (1994) proposed a very general, and now widely used framework to model forecast evolutions with the so-called Martingale Method of Forecast Evolutions (MMFE). Finally, in many systems, part of future demands may be known, due to early advanced orders, while the remainder of the demand volume continues to be uncertain.

Almost invariably, optimal replenishment strategies within a given inventory model are obtained by representing the problem as a dynamic program or Markov Decision process. These methodologies are intrinsically challenged by the ubiquitous curse of dimensionality, a term already coined by their founding father, Richard Bellman. Any of the above forecasting features induces a multi-dimensional, often a high-dimensional state space, even in the simplest models dealing with a single item, sold and stored in a single location. It is for this reason, that , even sixty years after the initiation of the field of stochastic inventory theory, with the seminal papers by Arrow, Harris, and Marschak (1951) and Dvoretzky, Kiefer, and Wolfowitz (1953), and after thousands of publications, almost all inventory models make the simplistic assumption that *demands are independent across times*. A small number of papers have incorporated some of the above intertemporal interdependencies, but, almost invariably, to suggest heuristics that fail to be based on an underlying exact model.

The objective of this paper is to demonstrate how the recent methodology of information relaxations can be used to handle the above forecasting features. We illustrate this by considering a single item inventory model with advanced demand information. The methodology allows the user to develop accurate lower bounds for the optimal system wide costs. These lower bounds may be used, by themselves, as cost estimators in strategic studies. Alternatively, they may be used to benchmark heuristics and bound their optimality gaps.

We consider a stochastic inventory control model consisting of a single retailer which faces random demands for a given item. Replenishment are made from an outside supplier, and are associated with a given leadtime L , and subject to arbitrary state-independent constraints. Inventories are reviewed and orders are placed on a periodic basis. When a retailer runs out of stock, unmet demand is backlogged. Demand forecasts are dynamically updated – in other words, at the start of each period, information is revealed that updates the decision maker’s assumptions about demand

in this and future periods.

System-wide costs consist of three parts (i) orders placed with the external supplier incur both a fixed cost and a variable cost, assumed to be proportional to the order size (ii) inventories carried at the retailer incurs carrying costs assumed to be linear in the end-of-period inventory levels, or, more generally, given by convex functions of the latter (iii) the retailer incurs backlogging costs, given by a linear or convex cost function of the end-of-period backlog size. The objective is to minimize the expected discounted cost over a finite planning horizon.

This problem class includes many important examples. To name but a few, situations in which demands are correlated from period to period (for example, demands with an underlying Hidden Markov Model, or demands conforming a ARMA process), problems in which orders are placed a certain number of periods before they are delivered, and problems in which exogenous information is revealed that provides some insights into future demands. We will specifically be focusing on the advance demand information model introduced by Gallego and Özer (2001), in which demand in any given period t is revealed over the previous $N + 1$ periods $t - N, \dots, t$, so that our beliefs about the demand in period t is updated in each of these $N + 1$ periods. In a later paper, Özer and Wei (2004) consider the same model with the additional constraint that replenishment orders from the supplier must be bounded below a certain maximum capacity.

Problems with dynamic demand updates are pervasive in industry. In many industries such as the fashion industry, in which product turnover is high, the early performance of each product heavily informs the chance the product will perform well in the future. Even for long-standing products, it stands to reason that common exogenous market-related factors may affect sales in adjacent periods, inducing implicit correlations between these periods.

In spite of this pervasiveness, these problems are generally intractable for all but the simplest of model parameters. As a result, the literature has often stayed away from these problems, and has instead focused on problems with uncorrelated demands across time periods. A number of heuristics have been developed to make ordering decisions in these kinds of systems – for example, Levi and Shi (2013) and Shi et al. (2014) develop balancing algorithms for models with uncapacitated and capacitated ordering quantities respectively, and Truong (2014) performs a thorough analysis of a heuristic for problems with no fixed costs.

In evaluating the performance of these heuristics, however, almost all of these papers confine themselves to very small or otherwise simplified problem instances, for which an exact solution can be obtained by brute-force solution of the full DP formulation of the problem. They then compare

the performance of their heuristic with that of these exact solutions. Some more recent papers in the literature also bound the worst-case performance of their heuristics (in particular, Levi and Shi (2013) provide a worst-case performance guarantee of 3 in the uncapacitated case, and Shi et al. (2014) provide a worst-case performance guarantee of 4 in the capacitated case). However, whilst these bounds are impressive technically, the performance guarantees they provide leave much to be desired – indeed, in the instances considered by these papers, the heuristics usually perform significantly better. Unfortunately, as mentioned above, the actual performance of these heuristics can only be evaluated in instances where an exact solution is available.

To our knowledge, therefore, no general method exists for evaluating these heuristics on problem instances for which an exact solution is *not* available.

In this paper, we propose such a method. We use the *information relaxation* framework, first introduced by Brown, Smith, and Sun (2010) to obtain a lower bound on the optimal cost of these problems for any problem instance. We test our method on the capacitated and uncapacitated advance information models discussed in Gallego and Özer (2001) and Özer and Wei (2004). Whilst these papers only provide exact solutions for problems in which $N = L + 2$, we extend their analysis to also provide exact solutions for instances in which $N = L + 3$. For these instances, we show that our information relaxation lower bound performs excellently, and never diverges from the optimal solution by more than 10%.

Our lower bound can be compared to the performance of any heuristic, even in problem instances for which an exact DP solution is *not* available, and the resulting performance gap can be used to evaluate the heuristic in question. As such, our method provides a general method for the evaluation of any such heuristic, as a complement to any theoretical worst-case performance guarantees.

2 Literature Review

This paper considers models in which beliefs about future demand distributions are updated in each period. Much of the literature dealing with these kinds of models has relied on a dynamic programming framework. The approach has been effective in characterizing the structure of optimal policies for these problems. Indeed, for many of these models, it has been shown that state-dependent (s, S) policies are optimal. In an (s, S) policy, orders are placed whenever inventory falls below a level s to restore the inventory to a level S . In a *state-dependent* (s, S) policy, the threshold parameters s and S in each period depend on the initial state of the system in that

period. For example, Scarf (1960) and Veinott (1966) show that (s, S) policies are optimal in systems with independent demands, Iida and Zipkin (2006) and Sethi and Cheng (1997) show that (s, S) policies are optimal in situations where demand is exogenous and Markov modulated, and Gallego and Özer (2001) show that (s, S) policies are optimal under the advance demand information model described in the introduction. In many cases, however, the optimal policies are far more complicated. Shaoxiang and Lambrecht (1996) demonstrate that when simple capacity constraints are added on the replenishment quantities at the retailers the optimal policy admits an $X - Y$ band structure, with $X < Y$. If the inventory position is below X , order the full capacity. If it is above Y , order nothing. However, if the inventory position is located between the two thresholds, then the ordering policy is more complicated. The ordering policy can only be complicated by more complex demand structures; Özer and Wei (2004) consider models with such inventory thresholds in the advance demand information case, but restrict themselves to $X - Y$ band policies in which $X = Y$.

Unfortunately, even when an (s, S) policy can be shown to be optimal, these structural results are often difficult to use computationally. If demands are independent and the time horizon is finite, the resulting dynamic programs have a one-dimensional state space and are therefore tractable; they can be used to calculate optimal (s, S) policies. Federgruen and Zipkin (1984a) and Federgruen and Zheng (1992) provide efficient algorithms for calculating (s, S) policies in various infinite-horizon systems. For processes with Markov modulated demands, Song and Zipkin (1993) provide an algorithm to find optimal (s, S) policies using a modified value-iteration approach, but they impose strong assumptions on the structure of the underlying Markov process, and on the size of its state space. Gallego and Özer (2001) and Özer and Wei (2004) use backwards induction algorithms to solve the problem with advance demand information with uncapacitated and capacitated order sizes respectively. However, their approach involves the solution of a full dynamic program describing the problem, which does not scale to situations in which advance demand information is available many periods in advance.

Because of these computational difficulties, an extensive literature has developed around heuristics for this class of problems. We will provide a brief description of certain heuristics, but we refer the reader to Iida and Zipkin (2006), Lu, Song, and Regan (2006), Dong and Lee (2003), and Truong (2014) for more details. One notable class of heuristics has focused on myopic policies, which make ordering decisions to minimize costs in the period in which the order arrives without regard for future periods. *Look-ahead optimization* policies, make ordering decisions so as to min-

imize *all* future costs that will be incurred as a result of the order, but without regard to future orders (see, for example, Truong (2014)). Finally, a class of *balancing policies* has been developed which place an order to balance fixed costs of the order (if any) with expected future holding and backorder costs (see Levi and Shi (2013) and Shi et al. (2014) for an example of this algorithm in models with uncapacitated and capacitated order sizes respectively).

As mentioned in the introduction, however, most of these papers are only able to evaluate the performance of their heuristics against simple cases for which exact solutions are available. A few papers are able to provide worst-case guarantees for *any* problem instance, but these guarantees often promise only poor performance.

The *information relaxation* framework was first introduced in full generality by Brown, Smith, and Sun (2010), based on earlier work around option pricing by Haugh and Kogan (2004), Rogers (2002), and Andersen and Broadie (2004).

The information relaxation approach is based on the fact that it is often very easy to obtain an upper bound on the solution of a stochastic dynamic program¹ – standard Monte Carlo simulation techniques can be used to evaluate the cost of *any* feasible heuristic policy, and the resulting cost *must* be higher than the optimal cost.

Lower bounds, however, are more difficult to obtain. In some cases, relaxing certain constraints in the full dynamic program can lead to a collapse in the size of the state space. The solution of the resulting simplified program is necessarily a lower bound, because the optimal solution in the full dynamic program is necessarily feasible in the relaxed dynamic program. For an example of such an approach in a supply-chain setting, see Federgruen and Zipkin (1984b). See also Kunnumkal and Topaloglu (2008) and Federgruen, Guetta, and Iyengar (2015), who use *Lagrangian relaxation* rather than a full relaxation to obtain a lower bound.

In many dynamic programs, however, it is not possible to relax any of the explicit constraints in the dynamic program to obtain such a simplification. The information relaxation approach relaxes the implicit constraints that all policies be *non-anticipative*. Indeed, in any stochastic dynamic program, it is always assumed that in any given period, all decisions are made without any knowledge of the future. Relaxing this constraint often makes the dynamic program far easier to solve.

¹In all the dynamic programs we consider in this paper, our objective will be to minimize cost. Thus, we refer to ‘upper bounds’ obtained using a heuristic, and ‘lower bounds’ obtained using information relaxations. These results all carry to dynamic program in which our objective is to *maximize* some quantity.

Unfortunately, it is often the case that full information about the future allows the decision maker to make significantly better decisions – this results in a far lower optimal cost, and the resulting lower bounds are often very loose. To remedy this problem, Brown, Smith, and Sun (2010) suggest adding a *penalty* to the relaxed dynamic program, which penalizes the information relaxation. We consider these issues in greater detail in section 5.

Since its introduction in Brown, Smith, and Sun (2010), the information relaxation approach has received extensive interest in the literature. We describe a selection of papers relevant to our work, but refer the reader to Brown and Haugh (2015) for more details. The framework was first formalized in the paper above, inspired by a number of earlier papers that used similar methods (see, for example, Haugh and Kogan (2004), Andersen and Broadie (2004), Rogers (2002), and Glasserman (2003) who provides a nice overview of this work). Many of the applications of this method have been in the field of finance and option pricing, but some authors have considered applications in revenue and supply chain management. In their original paper, Brown, Smith, and Sun (2010) consider an adaptive inventory control problem introduced by Treharne and Sox (2002), in which the true demand distribution is not known, and may change in any time period – every time a new demand is observed, the decision maker’s belief about the demand distribution is updated to reflect this new observation. This model is a particular instance of the framework considered in this paper, but it assumes there are no fixed costs associated with orders from the supplier, and no capacity constraints. Brown and Smith (2014) consider an inventory management model with lost-sales (i.e., in which unfulfilled demand at the end of each period is lost rather than backlogged), again with no fixed ordering costs. Bernstein, Li, and Shang () study a joint inventory and pricing model. They develop a heuristic, and use an information-relaxation based method to evaluate the performance of their heuristic. Secomandi (2014) considers policies for managing commodity storage assets, using information relaxation bounds with penalties derived from a simplified model.

To our knowledge, the information relaxation framework has never been applied to obtain lower bounds in inventory management problems with fixed ordering costs, constrained order quantities and updates on the demand distributions. In addition, our information relaxation methodology is novel in that even when we assume future demands are known, we keep some uncertainty in our dynamic program. In particular, we account in period t for holding and backorder costs incurred in period $t + L$, and in calculating those costs L periods in advance, we do not assume full knowledge of future demand distributions. This allows us to retain some randomness even in the relaxed

problem, and results in significantly better lower bounds. This is somewhat analogous to the concept of conditional Monte Carlo simulation.

3 Model and Notation

The general description of our system is provided in the introduction. We consider a periodic review system with a finite planning horizon of $T < \infty$ periods. We discount future costs with a one-period discount factor $\beta \leq 1$. Our objective is to minimize expected discounted aggregate costs in our system. As in most standard inventory models, we assume that all stockouts at the retailers are fully backlogged. Our model allows for a non-stationary and correlated demand structure, and an arbitrary set of constraints on the ordering quantities.

The original description of the information relaxation in Brown, Smith, and Sun (2010) is couched in measure-theoretic terms. We refer the reader to that paper for a general exposition of the concept. In this paper, we restrict ourselves to a non-measure-theoretic description applied to the specific class of problems described in the introduction.

We first note some general conventions. We shall denote random variables using uppercase letters, and realizations of these random variables using lowercase letters. Vectors will be denoted using a bold font, and matrices using a blackboard font.

The demands D_1, \dots, D_T over our planning horizon are random. We denote realizations of these demands by d_1, \dots, d_T .

At the beginning of each period t , we observe an *information set* denoted \mathbf{F}_t – this contains all the information that is available at the start of period t ; namely, the full set of demand realizations d_1, \dots, d_{t-1} , and possibly some additional exogenous information. We denote realizations of this information set by \mathbf{f}_t . Once \mathbf{F}_t is observed, it induces a conditional joint distribution on each of the future demands (D_t, \dots, D_T), and on the future information sets $\mathbf{F}_{t+1}, \dots, \mathbf{F}_T$. We assume that for each $t = 1, \dots, T$ and each realization \mathbf{f}_t , the conditional expectation $\mathbb{E}[D_\tau | \mathbf{F}_t = \mathbf{f}_t]$ is well-defined and finite for every $\tau = t, \dots, T$.

Each new information set \mathbf{F}_t reveals new information, but also includes all the information in the sets $\mathbf{F}_1, \dots, \mathbf{F}_{t-1}$. In other words, knowledge of the value of \mathbf{F}_t for any given t implies knowledge of all \mathbf{F}_τ for $\tau < t$.

We shall assume that each information set \mathbf{F}_τ can be represented by a random vector in \mathbb{R}^I , where I therefore acts as the ‘size’ of our information sets. Because new information is revealed

over time, some information sets \mathbf{F}_t for $t < T$ may require fewer than I numbers to be described; we shall nevertheless also model these information sets as a random vector in \mathbb{R}^I , possibly with a number of entries set to equal zero with probability one.

Orders are placed from an external supplier, and arrive after a fixed lead time L . We denote the order placed in period t by W_t . In period t , each order incurs a fixed cost K_t and a variable procurement rate c_t per unit ordered. In each period t , orders are constrained to lie within a set \mathcal{W}_t .

We let x_t denote the total inventory position at the beginning of period t . This includes the total inventory on hand at the start of the period, as well as all orders currently in the pipeline, on their way to the retailer.

The state of our system, at the start of any period t , is then given by the $I + 1$ dimensional tuple (x_t, \mathbf{f}_t) .

At the end of each period, carrying and backlogging costs are assessed as a convex function of the end-of-period inventory and backlog size respectively. In particular, at the end of period t , a holding cost $h_t(\cdot)$ is charged as a function of the total inventory level and a backorder cost $p_t(\cdot)$ is charged as a function of the total backlogging size.

When we take expectations, we will write the variables over which the expectation is being taken in the *subscript* of the expectation symbol. For example, $\mathbb{E}_{\mathbb{F}, \mathcal{D}}[\cdot]$ will denote an expectation over all possible sample paths. Similarly, we will often take expectations over all sample paths on which \mathbf{F}_t is equal to some \mathbf{f}_t . We shall denote such an expectation by $\mathbb{E}_{\mathbb{F}, \mathcal{D} | \mathbf{F}_t = \mathbf{f}_t}[\cdot]$.

4 An Exact Dynamic Programming Formulation

In this section, we derive an exact dynamic programming formulation, which uses the $(I + 1)$ -tuple (x_t, \mathbf{f}_t) as the state space for the system at the start of any given period t .

Since any stockouts at a retailer are fully backlogged, a retailer's inventory level at the *end* of period $(t + L)$ equals its inventory position x_t at the beginning of period t , *plus* the order W_t placed in this period that will be delivered in period $t + L$, minus demand observed in the interval $[t, t + L]$. In particular, the inventory on hand at the end of period t is given by

$$x_t + W_t - \sum_{\tau=t}^{t+L} D_\tau$$

Thus, we can express expected carrying and backlogging costs at the end of period $t + L$ as a

function of x_t and W_t only; we call this function Q_t , where

$$Q_t(x_t + W_t | \mathbf{f}_t) = \beta^L \mathbb{E}_{\mathbf{D} | \mathbf{F}_t = \mathbf{f}_t} \left[h_t \left(x_t + W_t - \sum_{\tau=t}^{t+L} D_\tau \right) + p_t \left(x_t + W_t - \sum_{\tau=t}^{t+L} D_\tau \right) \right]$$

Let

$$V_t(x_t, \mathbf{f}_t) = \text{The expected minimal present value of costs incurred} \\ \text{in periods } t, t+1, \dots, T \text{ when starting in state } (x_t, \mathbf{f}_t)$$

It is easily verified that the value functions satisfy the following recursions

$$V_t(x_t, \mathbf{f}_t) = \min_{W_t \in \mathcal{W}_t} \{ K_t \mathbb{I}_{W_t > 0} + c_t W_t + Q_t(x_t + W_t | \mathbf{f}_t) \\ + \beta \mathbb{E}_{D_t, \mathbf{F}_{t+1} | \mathbf{F}_t = \mathbf{f}_t} [V_{t+1}(x_t + W_t - D_t, \mathbf{F}_{t+1})] \} \quad (1)$$

with $V_{T-L+1}(\cdot) = 0$.

Unfortunately, because \mathbf{f}_t is typically a high-dimensional vector, this dynamic program is almost always intractable. This has led to a number of heuristics, as discussed above.

5 The Information Relaxation

In this section, we consider a technique for obtaining a *lower bound* on the optimal cost of dynamic program (1).

To clarify our exposition, we shall develop some notation for the different types of value functions we will be using in this section:

- V_t will denote the full value functions, described in the previous section.
- v_t will denote *relaxed* value functions obtained by solving our problem on a single sample path, as described below.
- ν_t will denote *average* value functions, obtained by averaging the functions v_t over all sample paths.

Now, consider a given sample path, consisting of realizations $\mathbf{f} = \{\mathbf{f}_1, \dots, \mathbf{f}_T\}$ and $\mathbf{d} = \{d_1, \dots, d_T\}$. Along this sample path, we can define a path-specific value function $v_t^{\mathbf{f}, \mathbf{d}}(x_t)$ satisfying the following recursions

$$v_t^{\mathbf{f}, \mathbf{d}}(x_t) = \min_{W_t \in \mathcal{W}_t} \left\{ K_t \mathbb{I}_{W_t > 0} + c_t W_t + Q_t(x_t + W_t | \mathbf{f}_t) + \beta v_{t+1}^{\mathbf{f}, \mathbf{d}}(x_t + W_t - d_t) \right\} \quad (2)$$

With $v_{T-L+1}^{\mathbf{f},\mathbf{d}}(\cdot) = 0$. This dynamic program finds the decisions that minimize the objective in dynamic program (1) along a *specific* sample path. It is, however, important to note that this is *not* equivalent to minimizing the system costs along that sample path. Indeed, the holding and backlogging costs charged in this dynamic program through the functions $Q_t(\cdot)$ are *not* the actual costs that will be incurred along sample path \mathbf{f} , but *expected* costs that will be incurred in each period given the information available L periods earlier. This approach allows this pathwise DP to retain *some* of the randomness in our full system.

The benefits of this relaxation are twofold. First, dynamic program (2) is deterministic. Second, it has a one-dimensional state space. This makes the relaxation (2) far simpler to solve than the original DP in (1).

Of course, this relaxed DP only gives the optimal cost on one particular sample path. We can, however, find the average expected cost over *every* sample path, and this leads to the following approximate value functions

$$\nu_t(x_t, \mathbf{f}_t) = \mathbb{E}_{\mathbb{F}, \mathbf{D} | \mathbf{F}_t = \mathbf{f}_t} \left[v_t^{\mathbb{F}, \mathbf{D}}(x_t) \right] \quad (3)$$

We will formally show in Theorem 2 below that these functions are lower bounds on the true value functions $V_t(x_t, \mathbf{f}_t)$. Intuitively, these value functions are able to generate policies that make full use of all information about future uncertainty; they must, therefore, perform better than policies that cannot make use of this information.

We have therefore successfully generated a lower bound on the optimal value functions $V_t(x_t, \mathbf{f}_t)$. Unfortunately, policies that are able to make full use of all future information may perform considerably better than the optimal policies, resulting in a large gap between the exact solution and our lower bound.

To mitigate this problem and narrow this gap, we generate a set of penalties $r_t^{\mathbf{f},\mathbf{d}}(x_t, W_t, \mathbf{f}_t)$ in each period. The aim of these penalties is to penalize policies that use future information in making their decisions. Instead of solving problem (2) for each sample path, we then solve the following recursions

$$\tilde{v}_t^{\mathbf{f},\mathbf{d}}(x_t) = \min_{W_t \in \mathcal{W}_t} \left\{ K_t \mathbb{I}_{W_t > 0} + c_t W_t + Q_t(x_t + W_t | \mathbf{f}_t) + \beta \tilde{v}_{t+1}^{\mathbf{f},\mathbf{d}}(x_t + W_t - d_t) + \beta r_t^{\mathbf{f},\mathbf{d}}(x_t, W_t, \mathbf{f}_t) \right\} \quad (4)$$

with $\tilde{v}_{T-L+1}^{\mathbf{f},\mathbf{d}}(\cdot) = 0$. Even with the added penalty, our problem is still deterministic and still one-dimensional, and so the benefits mentioned above are preserved.

Various forms have been suggested in the literature for the penalty r_t . We will consider an approach based on *approximate value functions*. Suppose we are in possession of functions $\tilde{V}_t(x_t, \mathbf{f}_t)$ that approximate the full value functions $V_t(x_t, \mathbf{f}_t)$ from equation (1). We then define

$$r_t^{\mathbf{f}, \mathbf{d}}(x_t, W_t, \mathbf{f}_t) = \mathbb{E}_{D_t, \mathbf{F}_{t+1} | \mathbf{F}_t = \mathbf{f}_t} \left[\tilde{V}_{t+1}(x_t + W_t - D_t, \mathbf{F}_{t+1}) \right] - \tilde{V}_{t+1}(x_t + W_t - d_t, \mathbf{f}_{t+1}) \quad (5)$$

Intuitively, this penalty function captures the cost reduction resulting from information relaxation in period t under value function \tilde{V}_t . The first term gives the approximate cost-to-go assuming no future information is known, whereas the second value function gives the cost assuming knowledge of the value the uncertain random variables will take in the next period.

To formalize this intuition, we prove the following theorem, that shows that with the right choice of approximate value function \tilde{V} , the penalty above exactly captures this difference.

Theorem 1. *Suppose $\tilde{V}_t(\cdot) = V_t(\cdot)$. Then, for any \mathbf{f}, \mathbf{d} and x_t ,*

$$\tilde{v}_t^{\mathbf{f}, \mathbf{d}}(x_t) = V_t(x_t, \mathbf{f}_t)$$

In other words, the penalized, relaxed value function is equal to the full value function.

Proof. We prove this result by induction. Using our terminal conditions, the result is trivially true for $t = T - L + 1$. Now, suppose the result is true for period $t + 1$. Using this inductive hypothesis to replace \tilde{v}_{t+1} with V_{t+1} , equation (4) becomes

$$\begin{aligned} \tilde{v}_t^{\mathbf{f}, \mathbf{d}}(x_t) = \min_{W_t \in \mathcal{W}_t} \{ & K_t \mathbb{I}_{W_t > 0} + c_t W_t + Q_t(x_t + W_t | \mathbf{f}_t) \\ & + \beta V_{t+1}((x_t + W_t - d_t, \mathbf{f}_{t+1})) + r_t^{\mathbf{f}, \mathbf{d}}(x_t, W_t, \mathbf{f}_t) \} \end{aligned}$$

Using the definition of $r_t^{\mathbf{f}, \mathbf{d}}$ in equation (5), and canceling terms, we obtain

$$\begin{aligned} \tilde{v}_t^{\mathbf{f}, \mathbf{d}}(x_t) = \min_{W_t \in \mathcal{W}_t} \{ & K_t \mathbb{I}_{W_t > 0} + c_t W_t + Q_t(x_t + W_t | \mathbf{f}_t) \\ & + \beta \mathbb{E}_{D_t, \mathbf{F}_{t+1} | \mathbf{F}_t = \mathbf{f}_t} \left[\tilde{V}_{t+1}(x_t + W_t - D_t, \mathbf{F}_{t+1}) \right] \} \end{aligned}$$

Recalling that, from the statement of our Theorem, that $\tilde{V}_{t+1}(\cdot) = V_{t+1}(\cdot)$, the RHS of this equation is identical to that in the definition of $V_t(\cdot)$ in the full dynamic program (1). \square

Of course, we do not know the actual value function $V_{t+1}(\cdot)$, and so it is not possible to set $\tilde{V}_{t+1} = V_t$. This theorem, however, is useful in verifying our intuition above, and in proving that

there is at least one $\tilde{V}_t(\cdot)$ for which $\tilde{v}_t^{\mathbf{f}, \mathbf{d}}(x_t)$ is exactly equal to the true value function (Brown, Smith, and Sun (2010) call this *strong duality* in the context of general information relaxations).

Analogously to expression (3), we define a new value function that averages our penalized relaxed value function over sample paths

$$\tilde{v}_t(x_t, \mathbf{f}_t) = \mathbb{E}_{\mathbb{F}, \mathbf{D} | \mathbf{F}_t = \mathbf{f}_t} \left[\tilde{v}_t^{\mathbb{F}, \mathbf{D}}(x_t) \right] \quad (6)$$

We now show that whatever our choice of approximate value function $\tilde{V}_t(\cdot)$, the function $\tilde{v}_t(\cdot)$ defined in (6) provides a lower bound on the true value function.

Theorem 2. *For any t , x_t and \mathbf{f}_t*

$$\tilde{v}_t(x_t, \mathbf{f}_t) \leq V_t(x_t, \mathbf{f}_t)$$

Proof. Once again, we prove this by induction. The result is trivial for $t = T - L + 1$, using our terminal conditions. Now, suppose the result is true in period $t + 1$. Substituting (4) and (5) into (6), we obtain

$$\begin{aligned} \tilde{v}_t(x_t, \mathbf{f}_t) = \mathbb{E}_{\mathbb{F}, \mathbf{D} | \mathbf{F}_t = \mathbf{f}_t} \left[\min_{W_t \in \mathcal{W}_t} \left\{ K_t \mathbb{I}_{W_t > 0} + c_t W_t + Q_t(x_t + W_t | \mathbf{f}_t) + \beta \tilde{v}_{t+1}^{\mathbf{F}, \mathbf{D}}(x_t + W_t - D_t) \right. \right. \\ \left. \left. + \beta \mathbb{E}_{\hat{D}_t, \hat{\mathbf{F}}_{t+1} | \hat{\mathbf{F}}_t = \mathbf{f}_t} \left[\tilde{V}_{t+1}(x_t + W_t - \hat{D}_t, \hat{\mathbf{F}}_{t+1}) \right] - \beta \tilde{V}_{t+1}(x_t + W_t - D_t, \mathbf{F}_{t+1}) \right\} \right] \end{aligned}$$

Interchanging the outer expectation and the minimization can only increase the value of the function. Indeed, when the minimization is inside the expectation, a different W_t can be chosen for every point in the sample path. When the expectation is taken first, we are forced to pick *one* W_t for every point on the sample path, leading to a suboptimal result. Thus

$$\begin{aligned} \tilde{v}_t(x_t, \mathbf{f}_t) \leq \min_{W_t \in \mathcal{W}_t} \left\{ \mathbb{E}_{\mathbb{F}, \mathbf{D} | \mathbf{F}_t = \mathbf{f}_t} \left[K_t \mathbb{I}_{W_t > 0} + c_t W_t + Q_t(x_t + W_t | \mathbf{f}_t) + \beta \tilde{v}_{t+1}^{\mathbf{F}, \mathbf{D}}(x_t + W_t - D_t) \right. \right. \\ \left. \left. + \beta \mathbb{E}_{\hat{D}_t, \hat{\mathbf{F}}_{t+1} | \hat{\mathbf{F}}_t = \mathbf{f}_t} \left[\tilde{V}_{t+1}(x_t + W_t - \hat{D}_t, \hat{\mathbf{F}}_{t+1}) \right] - \beta \tilde{V}_{t+1}(x_t + W_t - D_t, \mathbf{F}_{t+1}) \right\} \right\} \end{aligned}$$

Consider, however, that the inner expectation and outer expectation above condition over precisely the same variables. We can therefore replace the variables \hat{V}_t and $\hat{\mathbf{F}}_{t+1}$ with V_t and \mathbf{F}_{t+1} . The last two terms then cancel, and we obtain

$$\begin{aligned} \tilde{v}_t(x_t, \mathbf{f}_t) &\leq \min_{W_t \in \mathcal{W}_t} \left\{ K_t \mathbb{I}_{W_t > 0} + c_t W_t + Q_t(x_t + W_t | \mathbf{f}_t) + \beta \mathbb{E}_{\mathbb{F}, \mathbf{D} | \mathbf{F}_t = \mathbf{f}_t} \left[\tilde{v}_{t+1}^{\mathbb{F}, \mathbf{D}}(x_t + W_t - D_t) \right] \right\} \\ &= \min_{W_t \in \mathcal{W}_t} \left\{ K_t \mathbb{I}_{W_t > 0} + c_t W_t + Q_t(x_t + W_t | \mathbf{f}_t) + \beta \mathbb{E}_{D_t, \mathbf{F}_{t+1} | \mathbf{F}_t = \mathbf{f}_t} \left[\tilde{v}_{t+1}(x_t + W_t - D_t, \mathbf{F}_{t+1}) \right] \right\} \end{aligned}$$

Using the inductive hypothesis

$$\begin{aligned} \tilde{v}_t(x_t, \mathbf{f}_t) \leq & \min_{W_t \in \mathcal{W}_t} \{K_t \mathbb{I}_{W_t > 0} + c_t W_t + Q_t(x_t + W_t | \mathbf{f}_t) \\ & + \beta \mathbb{E}_{D_t, \mathbf{F}_{t+1} | \mathbf{F}_t = \mathbf{f}_t} [V_{t+1}(x_t + W_t - D_t, \mathbf{F}_{t+1})]\} \end{aligned}$$

Comparing the RHS of this equation to the RHS of (1), we find that the RHS is precisely $V_t(x_t, \mathbf{f}_t)$, proving our theorem. \square

We now focus on the choice of approximate value function \tilde{V}_t . Theorem 2 shows that our penalized relaxed problem will provide a lower bound regardless of this choice. Our choice of approximate value function, however, determines how tight the lower bound will be. Instead of picking one specific form for the approximate value function, we pick a family of quadratic functions specified by three sets of parameters $\mathbf{a} = \{a_1, \dots, a_{T-L}\}$, $\mathbf{b} = \{\mathbf{b}_1, \dots, \mathbf{b}_{T-L}\}$, and $\mathcal{G} = \{\mathbb{G}_1, \dots, \mathbb{G}_{T-L}\}$. We will later optimize over these parameters to get the best possible lower bound.

$$\tilde{V}_t(x_t, \mathbf{f}_t) = \begin{pmatrix} x_t & (\mathbf{f}_t)^\top \end{pmatrix} \begin{pmatrix} a_{t-1} & \frac{1}{2}(\mathbf{b}_{t-1})^\top \\ \frac{1}{2}\mathbf{b}_{t-1} & \mathbb{G}_{t-1} \end{pmatrix} \begin{pmatrix} x_t \\ \mathbf{f}_t \end{pmatrix}$$

Using this form of approximate value function, simple algebraic manipulation leads to the following penalty

$$\begin{aligned} r_t^{\mathbf{f}, \mathbf{d}}(x_t, W_t, \mathbf{f}_t) &= \{\mathbf{b}_t \cdot [\mathbb{E}_{\mathbf{F}_{t+1} | \mathbf{F}_t = \mathbf{f}_t}(\mathbf{F}_{t+1}) - \mathbf{f}_{t+1}] - 2a_t [\mathbb{E}_{D_t | \mathbf{F}_t = \mathbf{f}_t}(D_t) - d_t]\} (x_t + W_t) \quad (7) \\ &+ a_t [\mathbb{E}_{D_t | \mathbf{F}_t = \mathbf{f}_t}(D_t^2) - d_t^2] - \mathbf{b}_t \cdot [\mathbb{E}_{D_t, \mathbf{F}_{t+1} | \mathbf{F}_t = \mathbf{f}_t}(D_t \mathbf{F}_{t+1}) - d_t \mathbf{f}_{t+1}] \quad (8) \\ &+ \left[\mathbb{E}_{\mathbf{F}_{t+1} | \mathbf{F}_t = \mathbf{f}_t} \left(\sum_{i=1}^I (\mathbf{F}_{t+1} \cdot \mathbb{G}_t^{\text{col } i}) \mathbf{F}_{t+1} \right) - \left(\sum_{i=1}^I (\mathbf{f}_{t+1} \cdot \mathbb{G}_t^{\text{col } i}) \mathbf{f}_{t+1} \right) \right] \quad (9) \end{aligned}$$

Note that of all the terms in the penalty, only those in (7) will contribute to tightening our lower bound. Indeed, the terms in (8) and (9) are independent of the state x_t and our decisions W_t and have mean zero. As such, when an average is taken over all sample paths in (6), these terms vanish.

In spite of this, these terms can be a useful part of our penalty. Indeed, because they are correlated with the relevant terms in (7), they can act as *control variates* and reduce the variance of the optimal costs in (6). For this reason, we keep the terms in (8). By ignoring the terms in (9), however, we obviate the need for the parameters \mathcal{G} which do not appear in the first terms in (7). This considerably simplifies our search over these parameters. We therefore henceforth set $\mathcal{G}_t = 0$ for all t .

For any combination of the remaining parameters $\mathbf{a} = \{a_1, \dots, a_{T-L}\}$, and $\mathbf{b} = \{\mathbf{b}_1, \dots, \mathbf{b}_{T-L}\}$, we use simulation to find a lower bound $\tilde{\nu}_t(x_t, \mathbf{f}_t)$. One small technical problem that can arise is that if the values of \mathbf{b} and \mathbf{a} are such that the coefficient of W_t is negative, the optimal solution will be to order an infinite amount of inventory and the resulting problem would be unbounded. The resulting lower bound would be $-\infty$, which is still a lower bound albeit not a very interesting one. To avoid this problem, we will introduce an artificial constraints that upper bounds the inventory that can be ordered in every period with a very large number (for example, equal to ten times the sum of the 99th percentile of demand in every period in our horizon). This does not in any way alter the original problem, but it does avoid this technical difficulty. When such ‘rogue’ values of \mathbf{b} and \mathbf{a} are encountered, the allocation will be set to its upper bounds. As we will shortly see, this will results in very large subgradients with respect to \mathbf{b} and \mathbf{a} that will ‘push us away’ from these extreme points.

All that remains to do is to *maximize* this lower bound with respect to the parameters \mathbf{a} and \mathbf{b} . In Theorem 3 below, we show that for any x_t and \mathbf{f}_t , $\tilde{\nu}_t(x_t, \mathbf{f}_t)$ is a concave function of \mathbf{a} and \mathbf{b} . This implies that any standard steepest ascent method is guaranteed to converge to the penalty parameters that lead to the largest lower bound.

We will need some additional notation. For any given sample path \mathbb{f}, \mathbf{d} , let

$$\begin{aligned} \tilde{W}_t^*(x_t, \mathbb{f}, \mathbf{d}) = \operatorname{argmin}_{W_t \in \mathcal{W}_t} \left\{ K_t \mathbb{I}_{W_t > 0} + c_t W_t + Q_t(x_t + W_t | \mathbf{f}_t) + \beta \tilde{v}_{t+1}^{\mathbb{f}, \mathbf{d}}(x_t + W_t - d_t) \right. \\ \left. + \beta r_t^{\mathbb{f}, \mathbf{d}}(x_t, W_t, \mathbf{f}_t) \right\} \end{aligned}$$

Furthermore, for that same sample path, define a sequence of variables $\{\chi_1(\mathbb{f}, \mathbf{d}), \dots, \chi_{T-L}(\mathbb{f}, \mathbf{d})\}$ by setting

$$\chi_1(\mathbb{f}, \mathbf{d}) = x_1 + \tilde{W}_1^*(x_1, \mathbb{f}, \mathbf{d})$$

and

$$\chi_t(\mathbb{f}, \mathbf{d}) = \chi_{t-1}(\mathbb{f}, \mathbf{d}) + \tilde{W}_t^*(\chi_{t-1}(\mathbb{f}, \mathbf{d}), \mathbb{f}, \mathbf{d})$$

Theorem 3. *For any given value of x_t and \mathbf{f}_t ,*

- (a) *The lower bound $\tilde{\nu}_t(x_t, \mathbf{f}_t)$ is a concave function of \mathbf{a} and \mathbf{b} .*
- (b) *$\tilde{\nu}_t(x_t, \mathbf{f}_t)$ has supergradients with respect to \mathbf{a} , given by*

$$[\nabla_{\mathbf{a}} \tilde{\nu}_t(x_t, \mathbf{f}_t)]_{\tau} = \mathbb{E}_{D_{\tau}, \chi_{\tau} | \mathbf{F}_t = \mathbf{f}_t} \left[\left(\mathbb{E}_{\hat{D}_{\tau} | \hat{\mathbf{F}}_t = \mathbf{f}_t} (\hat{D}_{\tau}^2) - D_{\tau}^2 \right) - 2 \left(\mathbb{E}_{\hat{D}_{\tau} | \hat{\mathbf{F}}_t = \mathbf{f}_t} (\hat{D}_{\tau}) - D_{\tau} \right) \chi_{\tau} \right]$$

and supergradients with respect to \mathbb{b} , given by

$$\begin{aligned} [\nabla_{\mathbb{b}} \tilde{v}_t(x_t, \mathbf{f}_t)]_{\tau} &= \mathbb{E}_{D_{\tau}, \chi_{\tau}, \mathbf{F}_{\tau+1} | \mathbf{F}_t = \mathbf{f}_t} \left[\left(\mathbb{E}_{\hat{\mathbf{F}}_{\tau+1} | \hat{\mathbf{F}}_t = \mathbf{f}_t} (\hat{\mathbf{F}}_{\tau+1}) - \mathbf{F}_{\tau+1} \right) \chi_{\tau} \right. \\ &\quad \left. + \mathbb{E}_{\hat{D}_{\tau}, \hat{\mathbf{F}}_{\tau+1} | \hat{\mathbf{F}}_t = \mathbf{f}_t} (\hat{D}_{\tau} \hat{\mathbf{F}}_{\tau+1}) - D_{\tau} \mathbf{F}_{\tau+1} \right] \end{aligned}$$

We will first need to prove the following Lemma.

Lemma 1. Consider a specific sample path $\mathbb{F} = \mathbf{f}$ and $\mathbf{D} = \mathbf{d}$.

- (a) $\tilde{v}_t^{\mathbf{f}, \mathbf{d}}(x_t)$ is concave in \mathbf{a} and \mathbb{b} .
- (b) The following gives the derivatives of $\tilde{v}_t^{\mathbf{f}, \mathbf{d}}(x_t)$ with respect to \mathbf{a} and \mathbb{b}

$$\begin{aligned} \left[\nabla_{\mathbf{a}} \tilde{v}_t^{\mathbf{f}, \mathbf{d}}(x_t) \right]_{\tau} &= (\mathbb{E}_{D_{\tau} | \mathbf{F}_t = \mathbf{f}_t} (D_{\tau}^2) - d_{\tau}^2) - 2 (\mathbb{E}_{D_{\tau} | \mathbf{F}_t = \mathbf{f}_t} (D_{\tau}) - d_{\tau}) \chi_{\tau}(\mathbf{f}, \mathbf{d}) \\ \left[\nabla_{\mathbb{b}} \tilde{v}_t^{\mathbf{f}, \mathbf{d}}(x_t) \right]_{\tau} &= (\mathbb{E}_{\mathbf{F}_{\tau+1} | \mathbf{F}_t = \mathbf{f}_t} (\mathbf{F}_{\tau+1}) - \mathbf{f}_{\tau+1}) \chi_{\tau}(\mathbf{f}, \mathbf{d}) + \mathbb{E}_{D_{\tau}, \mathbf{F}_{\tau+1} | \mathbf{F}_t = \mathbf{f}_t} (D_{\tau} \mathbf{F}_{\tau+1}) - d_{\tau} \mathbf{f}_{\tau+1} \end{aligned}$$

Proof. Combining (4) and the penalty given in (7) and (8), and recalling the definition of $\chi_t(\mathbf{f}, \mathbf{d})$, we can write

$$\begin{aligned} \tilde{v}_t^{\mathbf{f}, \mathbf{d}}(x_t) &= K_t \mathbb{I}_{\chi_t(\mathbf{f}, \mathbf{d}) - x_t > 0} + c_t (\chi_t(\mathbf{f}, \mathbf{d}) - x_t) + Q_t (\chi_t(\mathbf{f}, \mathbf{d}) | \mathbf{f}_t) + \beta \tilde{v}_{t+1}^{\mathbf{f}, \mathbf{d}} (\chi_t(\mathbf{f}, \mathbf{d}) - d_t) \\ &\quad + \{ \mathbf{b}_t \cdot [\mathbb{E}_{\mathbf{F}_{t+1} | \mathbf{F}_t = \mathbf{f}_t} (\mathbf{F}_{t+1}) - \mathbf{f}_{t+1}] - 2a_t [\mathbb{E}_{D_t | \mathbf{F}_t = \mathbf{f}_t} (D_t) - d_t] \} \chi_t(\mathbf{f}, \mathbf{d}) \\ &\quad + a_t [\mathbb{E}_{D_t | \mathbf{F}_t = \mathbf{f}_t} (D_t^2) - d_t^2] - \mathbf{b}_t \cdot [\mathbb{E}_{D_t, \mathbf{F}_{t+1} | \mathbf{F}_t = \mathbf{f}_t} (D_t \mathbf{F}_{t+1}) - d_t \mathbf{f}_{t+1}] \end{aligned}$$

To prove (a), assume $\tilde{v}_t^{\mathbf{f}, \mathbf{d}}(x_t)$ is convex for a given t . Then for $t - 1$, the function above is clearly linear in \mathbf{a} and \mathbb{b} . Thus, every term above is concave in these variables, as required.

To prove (b), recall the terminal condition $\tilde{v}_{T-L+1}^{\mathbf{f}, \mathbf{d}}(\cdot) = 0$. Ignoring terms that do not depend in \mathbf{a} and \mathbb{b} , and applying the definition of \tilde{v}_t recursively, we find that

$$\begin{aligned} \tilde{v}_t^{\mathbf{f}, \mathbf{d}}(x_t) &= \sum_{\tau=t}^{T-L} \{ \mathbf{b}_{\tau} \cdot [\mathbb{E}_{\mathbf{F}_{\tau+1} | \mathbf{F}_t = \mathbf{f}_t} (\mathbf{F}_{\tau+1}) - \mathbf{f}_{\tau+1}] - 2a_{\tau} [\mathbb{E}_{D_{\tau} | \mathbf{F}_t = \mathbf{f}_t} (D_{\tau}) - d_{\tau}] \} \chi_{\tau}(\mathbf{f}, \mathbf{d}) \\ &\quad + a_{\tau} [\mathbb{E}_{D_{\tau} | \mathbf{F}_t = \mathbf{f}_t} (D_{\tau}^2) - d_{\tau}^2] - \mathbf{b}_{\tau} \cdot [\mathbb{E}_{D_{\tau}, \mathbf{F}_{\tau+1} | \mathbf{F}_t = \mathbf{f}_t} (D_{\tau} \mathbf{F}_{\tau+1}) - d_{\tau} \mathbf{f}_{\tau+1}] \end{aligned}$$

The derivatives given above follow immediately. □

We are now ready to prove Theorem 3.

Proof of Theorem 3. Recall that by (6), $\tilde{v}_t(x_t, \mathbf{f}_t) = \mathbb{E}_{\mathbb{F}, \mathbf{D} | \mathbf{F}_t = \mathbf{f}_t} \left[\tilde{v}_t^{\mathbb{F}, \mathbf{D}}(x_t) \right]$. To prove Theorem 3, therefore, it suffices to show that we can interchange differentiation and expectation in this case. Our result then follows directly from Lemma 1.

To show this interchange is possible, it suffices to note that because of the artificial upper bound on order quantities above, the variables $\chi_t(\mathbf{f}, \mathbf{d})$ are always bounded. The derivatives of \tilde{v} are therefore also bounded, and so our interchange is justified. \square

6 Numerical Study

In this section, we carry out a numerical study to assess the gap between the lower bound discussed above and the optimal cost derived using an exact DP solution in the specific case of the uncapacitated and capacitated *advanced demand information* models, discussed by Gallego and Özer (2001) and Özer and Wei (2004).

In the advance demand information model each period's demand D_t is observed over a number of period $t-N, \dots, t$. In particular, we denote the part of period t 's demand observed in period τ by D_t^τ (so that $D_t = \sum_{\tau=t-N}^t D_t^\tau$). At the end of every period t , a vector $(D_t^t, \dots, D_{t+N}^t)$ is observed, containing the last part of period t 's demand, and parts of the demands in periods $t+1, \dots, t+N$. The costs and mechanics of the system are identical to those discussed in Section 3; in particular, we assume that deliveries from the depot face a leadtime L before they are delivered.

At the start of each period t , the state of the system consists of the current inventory position x_t together with all information previously observed pertaining to periods $t+L$ onwards (indeed, recall that in period t , we charge costs incurred in period $t+L$). In particular, as we begin period t , we shall have observed at least some part of the demand up to period $t+N-1$. We shall need the sum of all demands that have been observed in periods $t+L$ to $t+N-1$. This means that the size of our state space will need to be $N-L$. This problem is readily cast in the form described in Section 3.

We considered just over 810 instances of this problem, generated using all permutations of the following parameters:

- Leadtime $L \in \{0, 1, 2, 3, 4\}$
- Number of periods in the future for which we observe demand in period t : $N \in \{L+2, N+3\}$.
- Fixed ordering costs: $K \in \{0, 10, 50\}$.

- Backorder costs: $p \in \{1, 10, 50\}$.
- Maximum order size allowed from the external supplier: $C = \{3, 12, \infty\}$.
- In all cases, we assumed the demand in each period was Poisson distributed with a mean of 6. In the advance information framework, the demand in period t is revealed in periods $t, \dots, t + N$. To model this process, we assume that for each period t , each D_t^{t-N}, \dots, D_t^t is Poisson distributed with means $\lambda_0, \dots, \lambda_N$. We use three configurations for these means

$$\begin{aligned}
(\lambda_0, \dots, \lambda_N) &= \frac{6}{N}(1, \dots, 1) \\
(\lambda_0, \dots, \lambda_N) &= \frac{12}{(N+2)(N+1)}(N+1, N, \dots, 1) \\
(\lambda_0, \dots, \lambda_N) &= \frac{12}{(N+2)(N+1)}(1, 2, \dots, N+1)
\end{aligned}$$

The first configuration considers a situation in which demands are revealed uniformly across the $N + 1$ periods in question. The second configuration considers a situation in which more of the demand is revealed earlier, and the final configuration considers a situation in which more of the demand is revealed later.

In all cases, we used $h = 1$, $c = 2$, $T = 15$, and $\beta = 1$.

For each instance, we find the optimal cost using an exact dynamic programming formulation. It is worthwhile noting that, to our knowledge, our paper is the first to use an exact dynamic programming formulation to tackle the advance demand information problem with $N = L + 3$, resulting in a three-dimensional state-space; all previous papers have only considered cases in which $N = L + 2$, resulting in a two-dimensional state-space.

We then compute a lower bound using the information relaxation approach as follows. We begin with penalty parameters $\mathbf{a} = \mathbf{0}$ and $\mathbf{b} = 0$, and, using a specific seed to generate our random variables, we use simulations to estimate the value of the lower bound $\tilde{n}u_t$. Using a standard gradient descent algorithm, we modify the parameters \mathbf{a} and \mathbf{b} to obtain the highest possible lower bound for simulations generated using that particular seed. We then repeat this process for four more seeds, starting each search at $\mathbf{a} = \mathbf{0}$ and $\mathbf{b} = 0$. Using a sixth random seed, we use new simulations to evaluate the performance of the five pairs of parameters (\mathbf{a}, \mathbf{b}) obtained in our first five simulations and pick the best one. Finally, we evaluate the performance of our chosen parameters using a seventh seed.

To evaluate the performance of our lower bound, we calculate the percentage difference between our lower bound and the exact DP solution.

Figure 1 displays these results for 90 of the 810 instances in our numerical study. The remaining results are provided in an online appendix.

The results are generally excellent. In no instances is the gap between the upper and lower bound ever greater than 8%, and often the gaps observed are much smaller.

Looking at the first row in Figure 1, pertaining to $K = 50$ and $C = \infty$, it is clear that small backorder costs p tend to lead to better results (i.e., smaller gaps) than larger holding costs. It is also apparent that when $p = 10$ or $p = 50$, the addition of a penalty improves the performance of our lower bound, as evidenced by the fact the dotted lines there lie *outside* the confidence interval of our lower bound.

Looking at the second row in Figure 1, pertaining to $K = 10$ and $C = 12$, two trends are apparent. First, comparing these plots to those in the first row, it is apparent that lower fixed costs leads to better performance. Second, looking at each of the plot themselves, it becomes clear that larger values of L lead to considerably smaller gaps. The latter observation is readily explained. Recall that even while assuming full information, our lower bound nevertheless retains *some* randomness through the $Q(\dots)$ functions, which consider expected costs L periods from now. The larger L , the more randomness is retained in these functions, and the less impactful we might expect our relaxation to be.

Finally, considering the third line, we find that when capacity is severally constrained (in this case, with $c = 3$), results are uniformly excellent, even with large fixed ordering costs and large backorder costs p . This can again be understood in terms of our relaxation. When order quantities are so severely constrained, the optimal solution requires a maximum order (of three units) at all times, regardless of the state of the system. Having full knowledge of future information, therefore, does nothing to help us make better decisions, and does not significantly lower our optimal cost.

7 Conclusion

We have devised an approach to obtain lower bounds on the optimal costs of supply chain problems with dynamically updated demands. This class of models includes many important examples, from models with correlated demands to those with advance demand information.

Our lower bound approach is based on an information relaxation approach that solves the

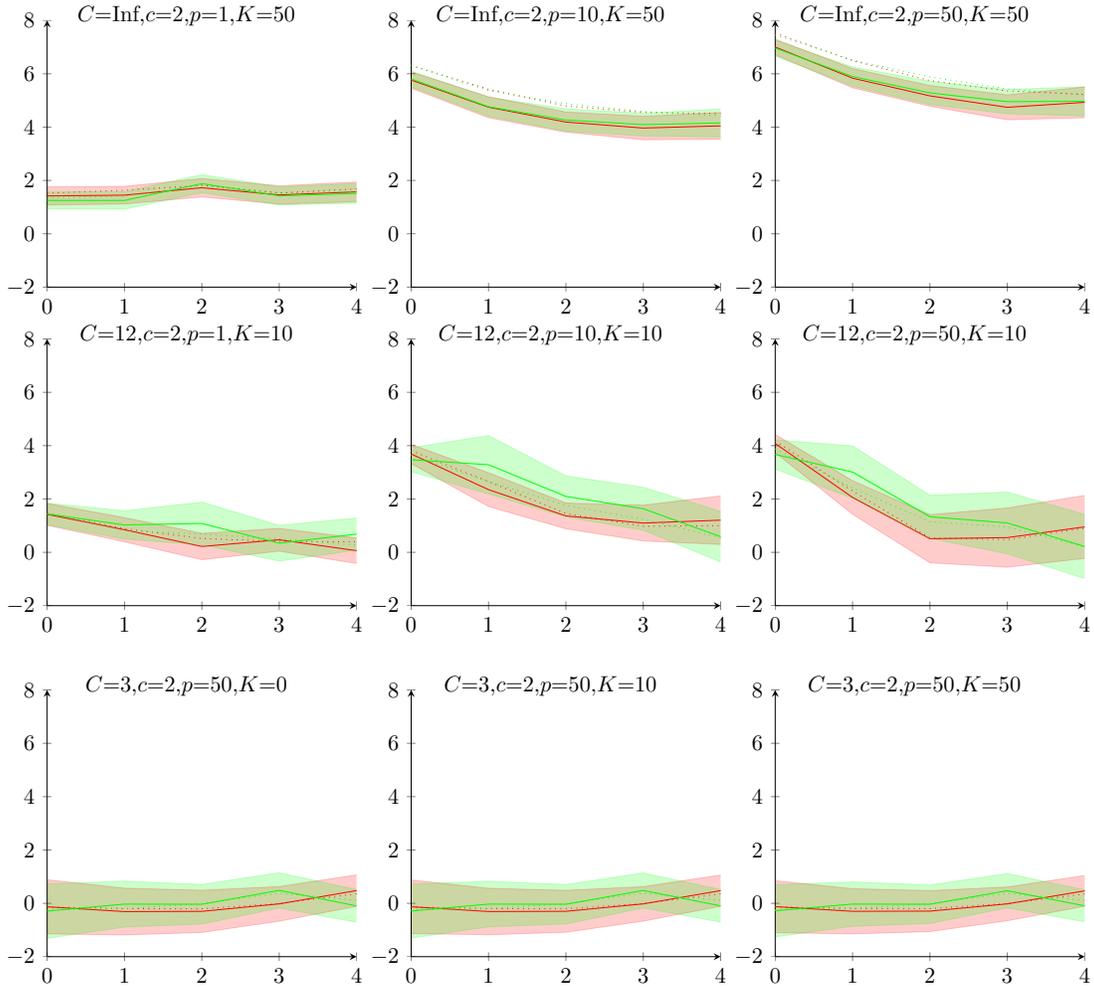


Figure 1: Results for 90 of the 810 instances in our numerical study. Each plot displays results for a particular combination of K , p , and C parameters. All results pertain to demand configurations in which the advance demand information is uniformly distributed over the $N + 1$ periods in question; the results are similar for other demand configurations. In each plot, the x -axis corresponds to different values of L , and the y -axis displays percentage differences between our lower bound and the exact DP solution. The red line corresponds to $N = L + 2$, and the green line corresponds to $N = L + 3$. The dotted line displays gaps for information relaxations with no penalty (i.e., with $\mathbf{a} = \mathbf{0}$ and $\mathbf{b} = 0$), and the solid lines represent the results with an optimal penalty. The shaded polygons display 95% confidence intervals around the solid lines.

problem for a number of sample paths, and averages the resulting cost over these sample paths. The resulting policies are able to make full use of future information, and therefore necessarily perform better than policies forced to rely on uncertain information. Even in the fully relaxed problems, we account in period t for *expected* holding and backorder costs incurred in period $t + L$. This expectation retains some uncertainty in our problem even in the fully relaxed problems.

We assessed the performance of these information relaxations by considering a set of instances for which an exact solution is available through the solution of an exact DP, and comparing our upper bound to the lower bound obtained using the information relaxation. We found that our lower bounds were generally excellent, with no instance yielding a gap larger than 8%.

In an attempt to narrow this gap further, we introduced a series of penalties based on an approach introduced by Brown, Smith, and Sun (2010). These penalties were successful in slightly reducing these already excellent gaps in a small number of instances.

As noted in other papers that use this upper/lower bound approach (see, for example, Federgruen, Guetta, and Iyengar (2015)), the ability to compute an accurate lower bound on the optimal cost for any set of parameters has fundamental benefits beyond the ability to test the accuracy of a specific heuristic policy. These bounds allow us to address various strategic questions. In the advance information case, for example, such a lower bound could allow us to evaluate the value of advance information – in particular, we could compute the lower bound for values values of N , and observe the reductions in cost – if any – that result from advance information about the demand.

Future work in this direction should first explore the use of different penalties to try and further improve the bounds above. For example, Brown and Smith (2011) suggest the use of linear approximations to non-linear penalties in the framework above. Desai, Farias, and Moallemi (2011) suggest what they call a ‘pathwise optimization’ approach that considers penalties consisting of linear combinations of a set of ‘basis penalties’, and finds the best information relaxation lower bound by finding the best such linear combination.

A second fruitful direction for future work is the use of information relaxation lower bounds to develop better heuristics for the problems at hand. This approach was used in the specific supply chain context considered Brown, Smith, and Sun (2010), but to our knowledge, no general method exists for obtaining such heuristics from lower bounds. Such methods could be powerful sources of new heuristics.

8 Acknowledgements

The authors thank Abdus Samad Khan, whose work in painstakingly running the numerical simulations described in this paper was invaluable in the writing of this paper.

References

- Andersen, L. and M. Broadie (2004). “Primal-dual simulation algorithm for pricing multidimensional American options”. In: *Management Science* 50.9, pp. 1222–1234.
- Arrow, K. J., T. Harris, and J. Marschak (1951). “Optimal inventory policy”. In: *Econometrica: Journal of the Econometric Society*, pp. 250–272.
- Bernstein, F., Y. Li, and K. Shang. “A Simple Heuristic for Joint Inventory and Pricing Models with Lead Time and Backorders”. In:
- Brown, D. B. and M. B. Haugh (2015). “Information Relaxation Bounds for Infinite Horizon Markov Decision Processes”. http://www.columbia.edu/~mh2078/infinite_horizon_current.pdf.
- Brown, D. B. and J. E. Smith (2011). “Dynamic portfolio optimization with transaction costs: Heuristics and dual bounds”. In: *Management Science* 57.10, pp. 1752–1770.
- (2014). “Information relaxations, duality, and convex stochastic dynamic programs”. In: *Operations Research* 62.6, pp. 1394–1415.
- Brown, D. B., J. E. Smith, and P. Sun (2010). “Information relaxations and duality in stochastic dynamic programs”. In: *Operations research* 58.4-part-1, pp. 785–801.
- Desai, V., V. Farias, and C. Moallemi (2011). “Bounds for Markov decision processes”. In: *Chapter in Reinforcement Learning and Approximate Dynamic Programming for Feedback Control (FL Lewis, D. Liu, eds.)*
- Dong, L. and H. L. Lee (2003). “Optimal policies and approximations for a serial multiechelon inventory system with time-correlated demand”. In: *Operations Research* 51.6, pp. 969–980.
- Dvoretzky, A., J. Kiefer, and J. Wolfowitz (1953). “On the optimal character of the (s, S) policy in inventory policy”. In: *Econometrica, October*.
- Federgruen, A., D. Guetta, and G. Iyengar (2015). “Two Echelon Distribution Systems with Random Demands and Storage Constraints”. Columbia University Working Paper.
- Federgruen, A. and Y-S. Zheng (1992). “An efficient algorithm for computing an optimal (r, Q) policy in continuous review stochastic inventory systems”. In: *Operations research* 40.4, pp. 808–813.

- Federgruen, A. and P. Zipkin (1984a). “An efficient algorithm for computing optimal (s, S) policies”. In: *Operations research* 32.6, pp. 1268–1285.
- (1984b). “Approximations of dynamic, multilocation production and inventory problems”. In: *Management Science* 30.1, pp. 69–84.
- Gallego, G. and Ö. Özer (2001). “Integrating replenishment decisions with advance demand information”. In: *Management science* 47.10, pp. 1344–1360.
- Glasserman, P. (2003). *Monte Carlo methods in financial engineering*. Vol. 53. Springer Science & Business Media.
- Haugh, M. B. and L. Kogan (2004). “Pricing American options: a duality approach”. In: *Operations Research* 52.2, pp. 258–270.
- Heath, David C and Peter L Jackson (1994). “Modeling the evolution of demand forecasts ITH application to safety stock analysis in production/distribution systems”. In: *IIE transactions* 26.3, pp. 17–30.
- Iida, T. and P. H. Zipkin (2006). “Approximate solutions of a dynamic forecast-inventory model”. In: *Manufacturing & Service Operations Management* 8.4, pp. 407–425.
- Kunnumkal, S. and H. Topaloglu (2008). “A duality-based relaxation and decomposition approach for inventory distribution systems”. In: *Naval Research Logistics (NRL)* 55.7, pp. 612–631. ISSN: 1520-6750. DOI: 10.1002/nav.20306. URL: <http://dx.doi.org/10.1002/nav.20306>.
- Levi, R. and C. Shi (2013). “Approximation algorithms for the stochastic lot-sizing problem with order lead times”. In: *Operations Research* 61.3, pp. 593–602.
- Lu, X., J-S. Song, and A. Regan (2006). “Inventory planning with forecast updates: Approximate solutions and cost error bounds”. In: *Operations research* 54.6, pp. 1079–1097.
- Özer, Ö. and W. Wei (2004). “Inventory control with limited capacity and advance demand information”. In: *Operations Research* 52.6, pp. 988–1000.
- Rogers, L. C. G. (2002). “Monte Carlo valuation of American options”. In: *Mathematical Finance* 12.3, pp. 271–286.
- Scarf, H. (1960). *The Optimality of (s, S) Policies in Dynamic Inventory Problems*. K. Arrow, S. Karlin, P. Suppes, eds., *Mathematical Models in the Social Sciences*.
- Secomandi, N. (2014). *Analysis and enhancement of practice-based policies for the real option management of commodity storage assets*. Tech. rep. Tepper Working Paper 2011-E11, Tepper School of Business, Carnegie Mellon University, Pittsburgh, PA, USA.

- Sethi, S. P. and F. Cheng (1997). “Optimality of (s, S) policies in inventory models with Markovian demand”. In: *Operations Research* 45.6, pp. 931–939.
- Shaoxiang, C. and M Lambrecht (1996). “XY band and modified (s, S) policy”. In: *Operations Research* 44.6, pp. 1013–1019.
- Shi, C. et al. (2014). “Approximation algorithms for capacitated stochastic inventory systems with setup costs”. In: *Naval Research Logistics (NRL)* 61.4, pp. 304–319.
- Song, J-S. and P. Zipkin (1993). “Inventory control in a fluctuating demand environment”. In: *Operations Research* 41.2, pp. 351–370.
- Treharne, J. T. and C. R. Sox (2002). “Adaptive inventory control for nonstationary demand and partial information”. In: *Management Science* 48.5, pp. 607–624.
- Truong, V-A. (2014). “Approximation algorithm for the stochastic multiperiod inventory problem via a look-ahead optimization approach”. In: *Mathematics of Operations Research* 39.4, pp. 1039–1056.
- Veinott, A. F. (1966). “On the Optimality of (s,S) Inventory Policies: New Conditions and a New Proof”. In: *SIAM Journal on Applied Mathematics* 14.5, pp. 1067–1083.

A Advanced Demand Information – Full Numerical Results

This appendix lists all the results from our numerical study in Figures 2-10. Each figure displays results for a particular value of K and a particular configuration of the advance information, comprising 90 of the 810 results in our study. In each plot, the x -axis corresponds to different values of L , and the y -axis displays percentage differences between our lower bound and the exact DP solution. The red line corresponds to $N = L + 2$, and the green line corresponds to $N = L + 3$. The dotted line displays gaps for information relaxations with no penalty (i.e., with $\mathbf{a} = \mathbf{0}$ and $\mathbf{b} = 0$, and the solid lines represent the results with an optimal penalty. The shaded polygons display 95% confidence intervals around the solid lines.

Recall three configurations of the advance information were used. In all cases, each period’s demand is Poisson distributed with a mean of 6. We assume that for each period t , each D_t^{t-N}, \dots, D_t^t

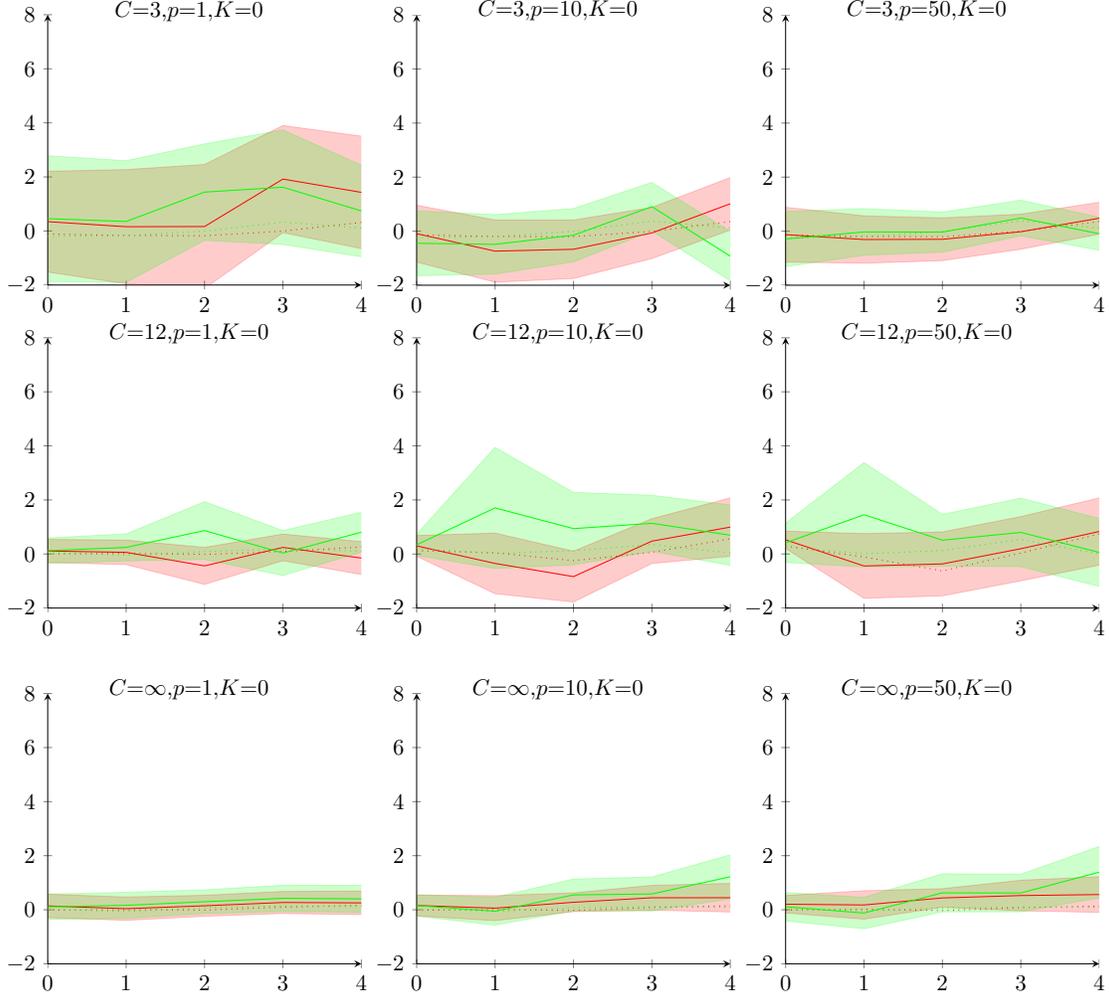


Figure 2: $K = 0$ and an equal split.

is Poisson distributed with means $\lambda_0, \dots, \lambda_N$. We use three configurations for these means

$$\begin{aligned}
 (\lambda_0, \dots, \lambda_N) &= \frac{6}{N}(1, \dots, 1) \\
 (\lambda_0, \dots, \lambda_N) &= \frac{12}{(N+2)(N+1)}(N+1, N, \dots, 1) \\
 (\lambda_0, \dots, \lambda_N) &= \frac{12}{(N+2)(N+1)}(1, 2, \dots, N+1)
 \end{aligned}$$

We call the first conguration ‘equal split’, the second ‘early bias’ and the last ‘late bias’.

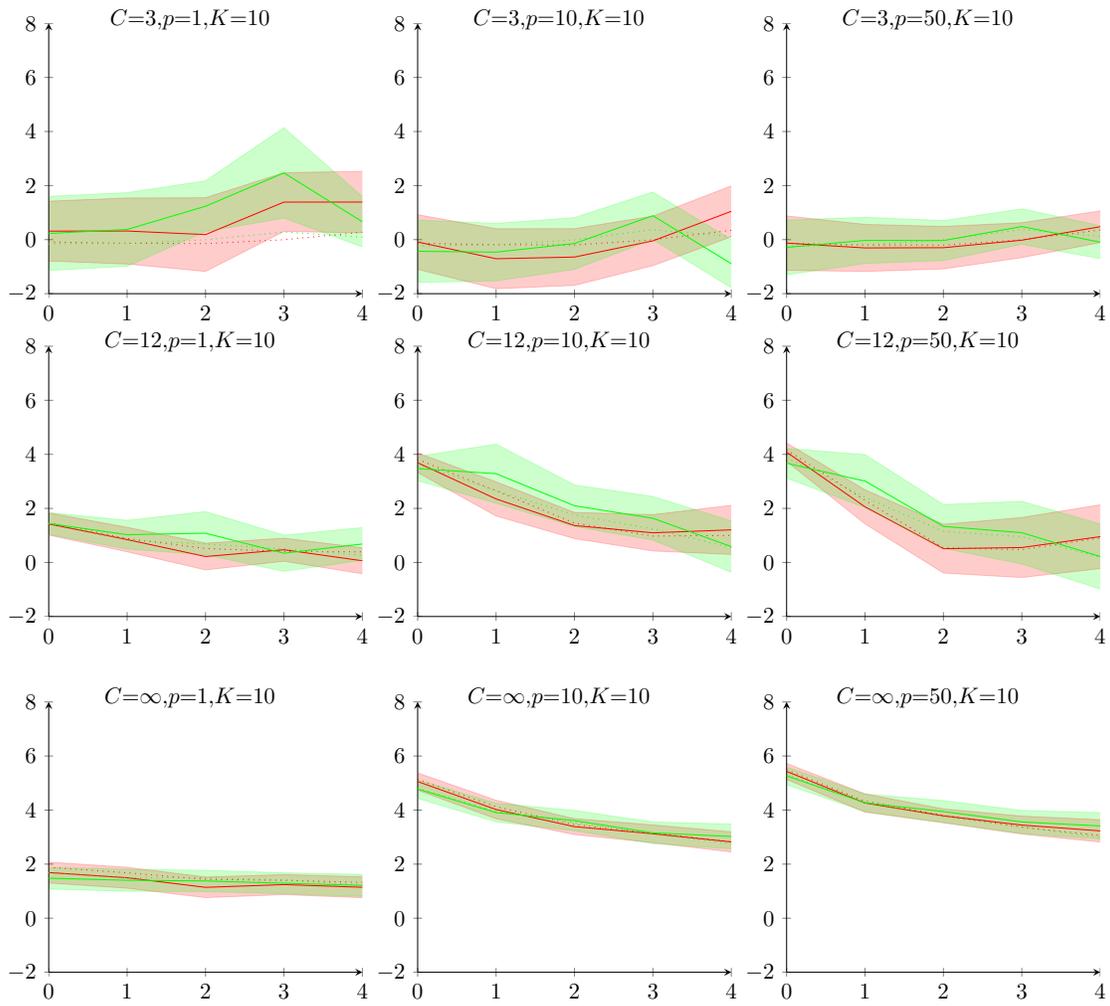


Figure 3: $K = 10$ and an equal split.

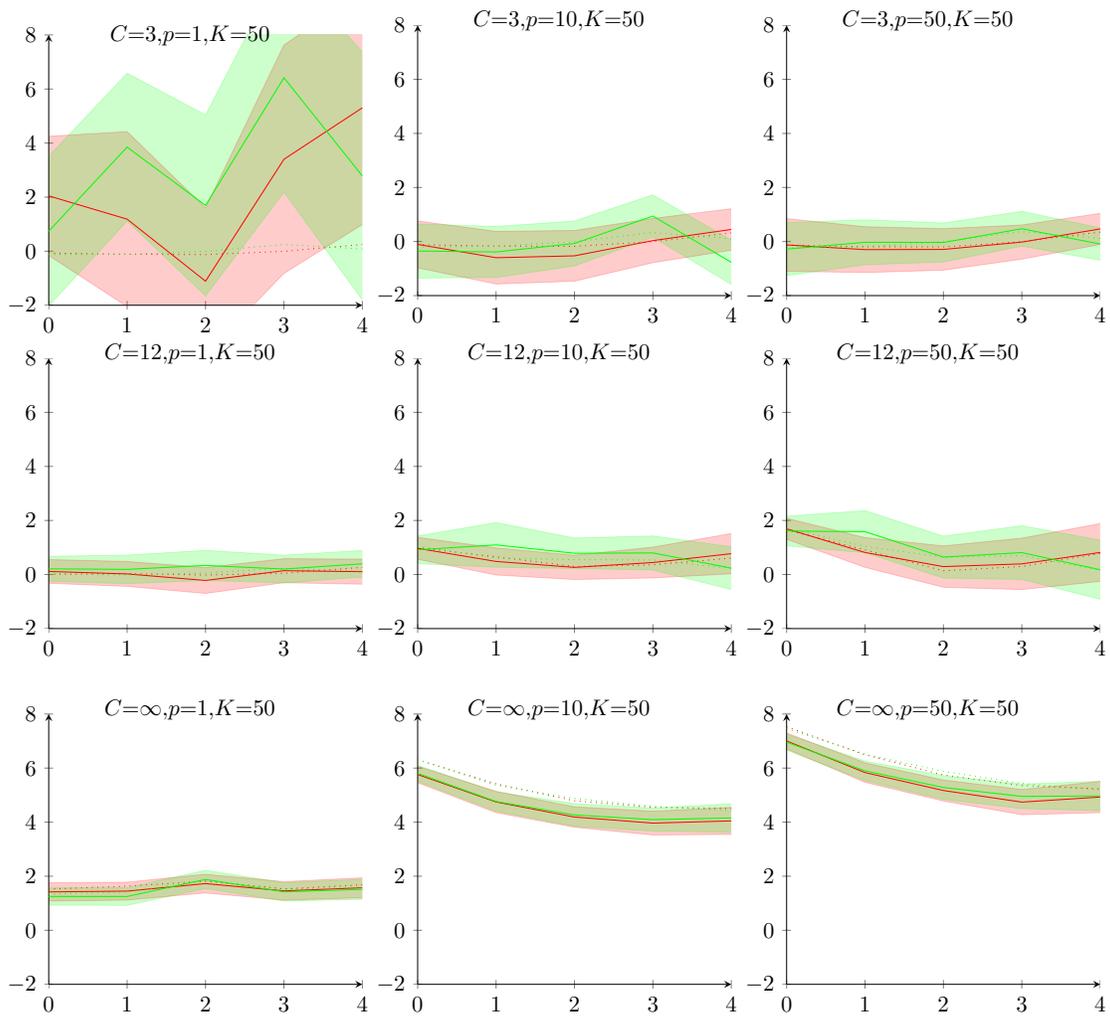


Figure 4: $K = 50$ and an equal split.

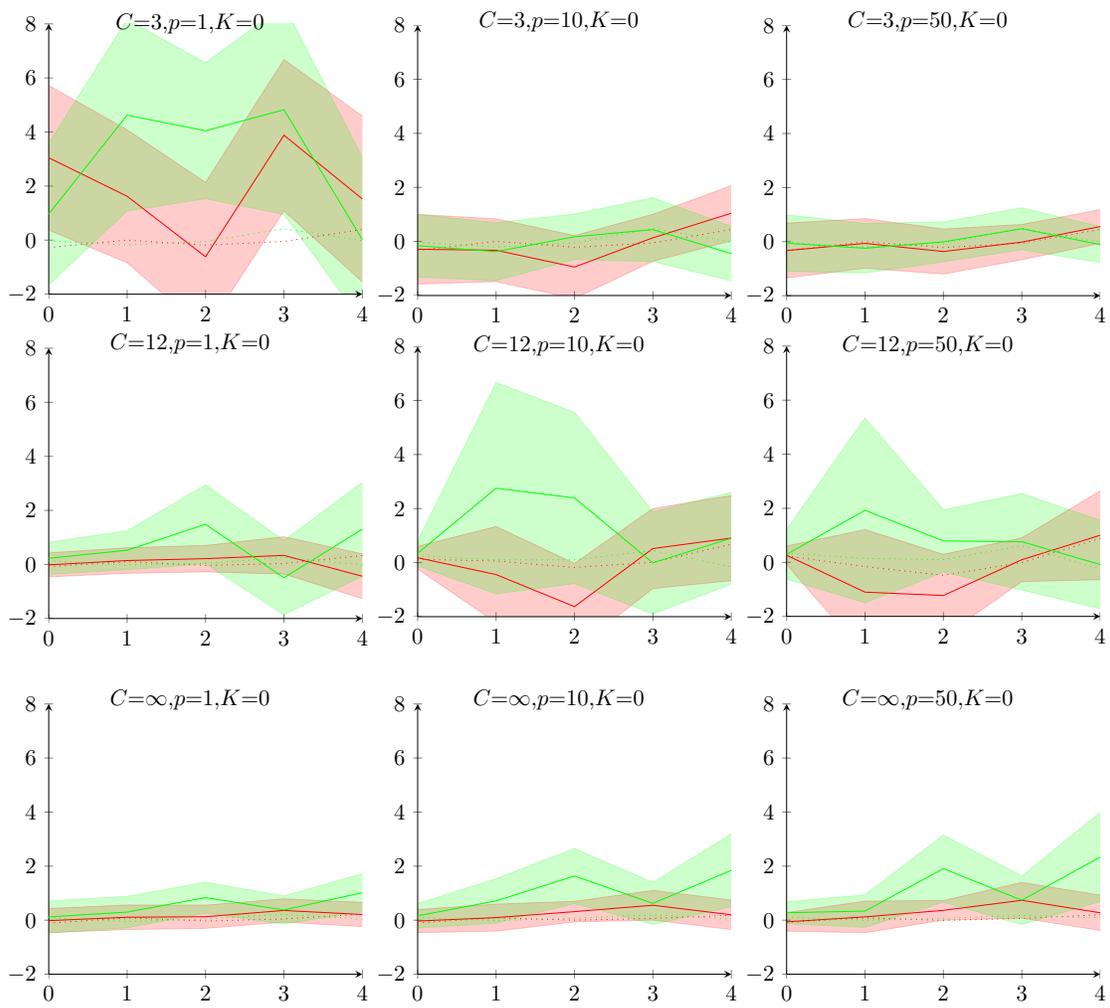


Figure 5: $K = 0$ and early bias.

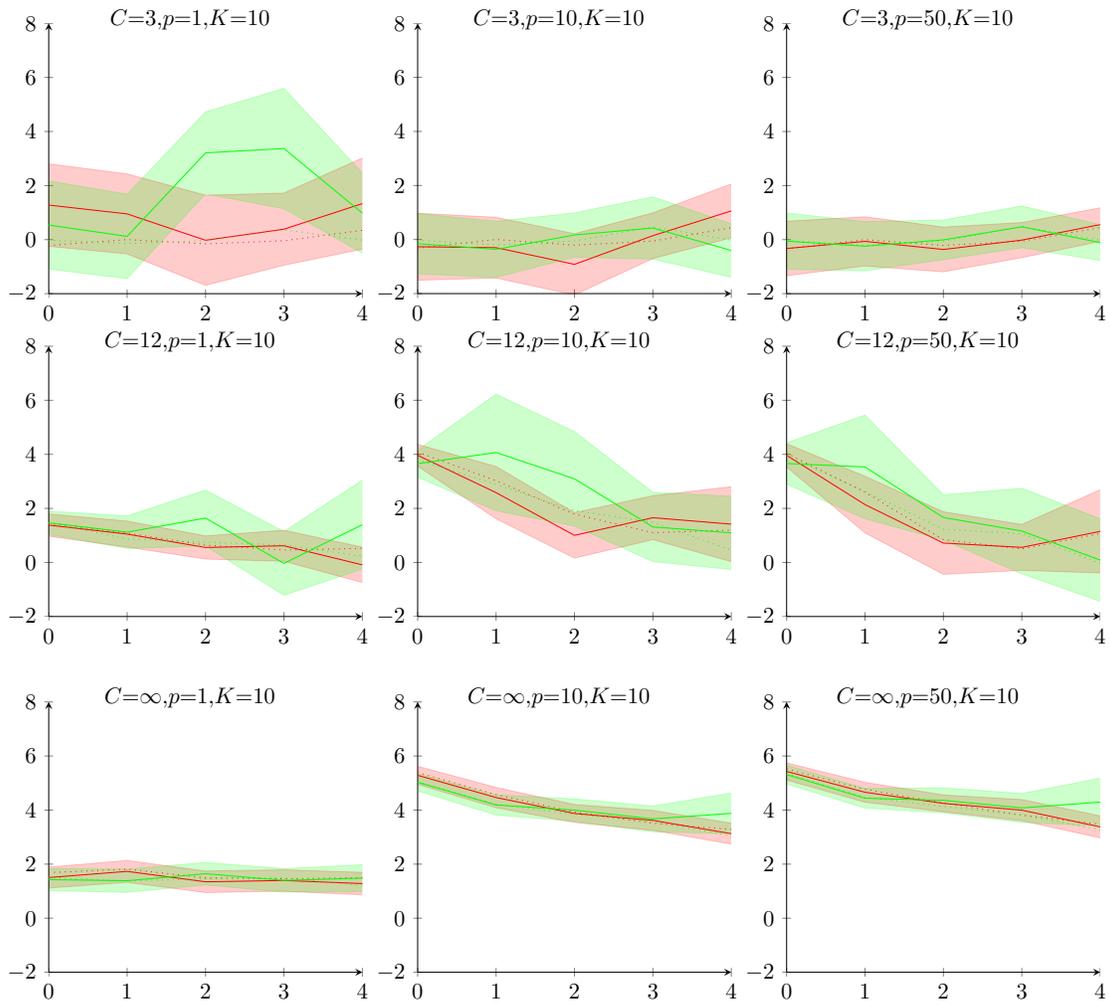


Figure 6: $K = 10$ and early bias.

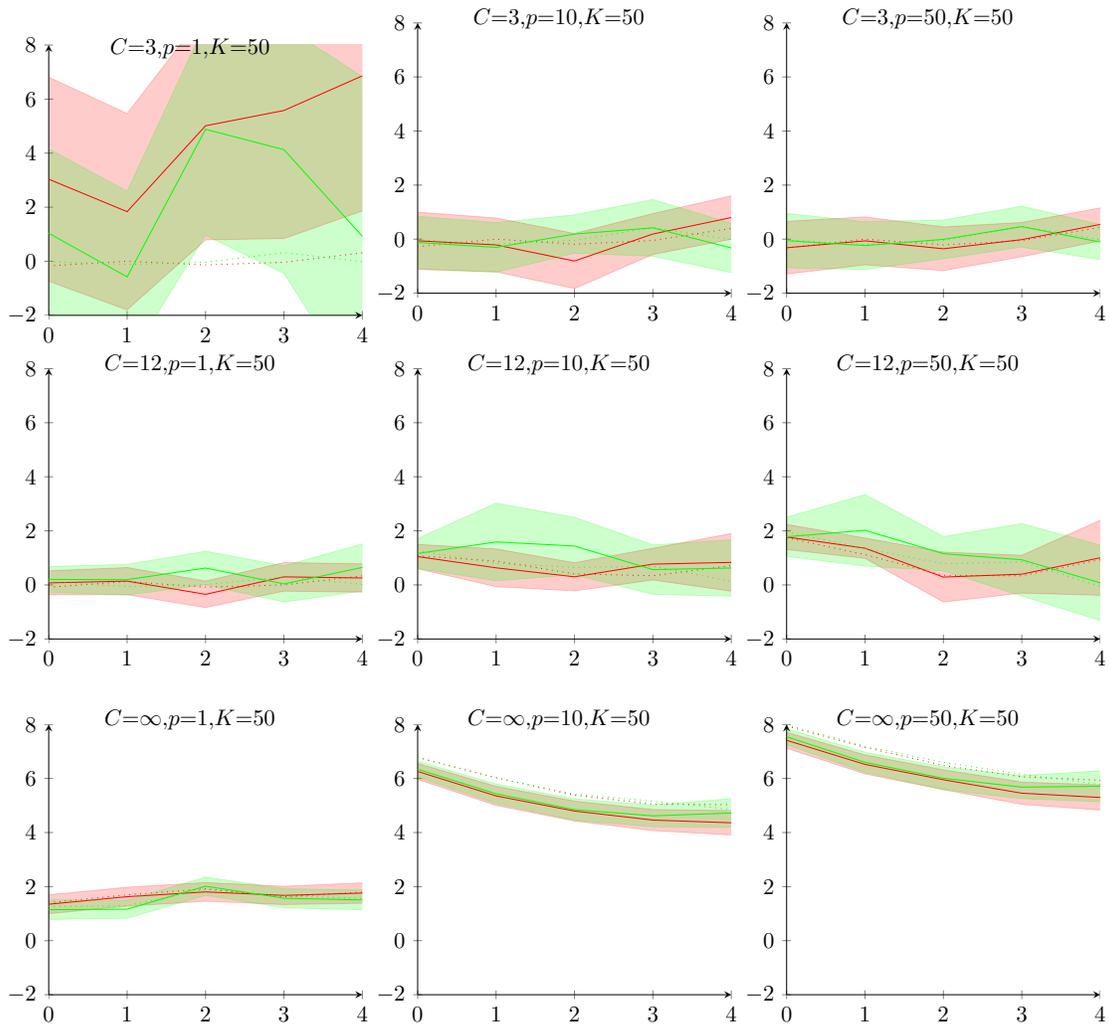


Figure 7: $K = 50$ and early bias.

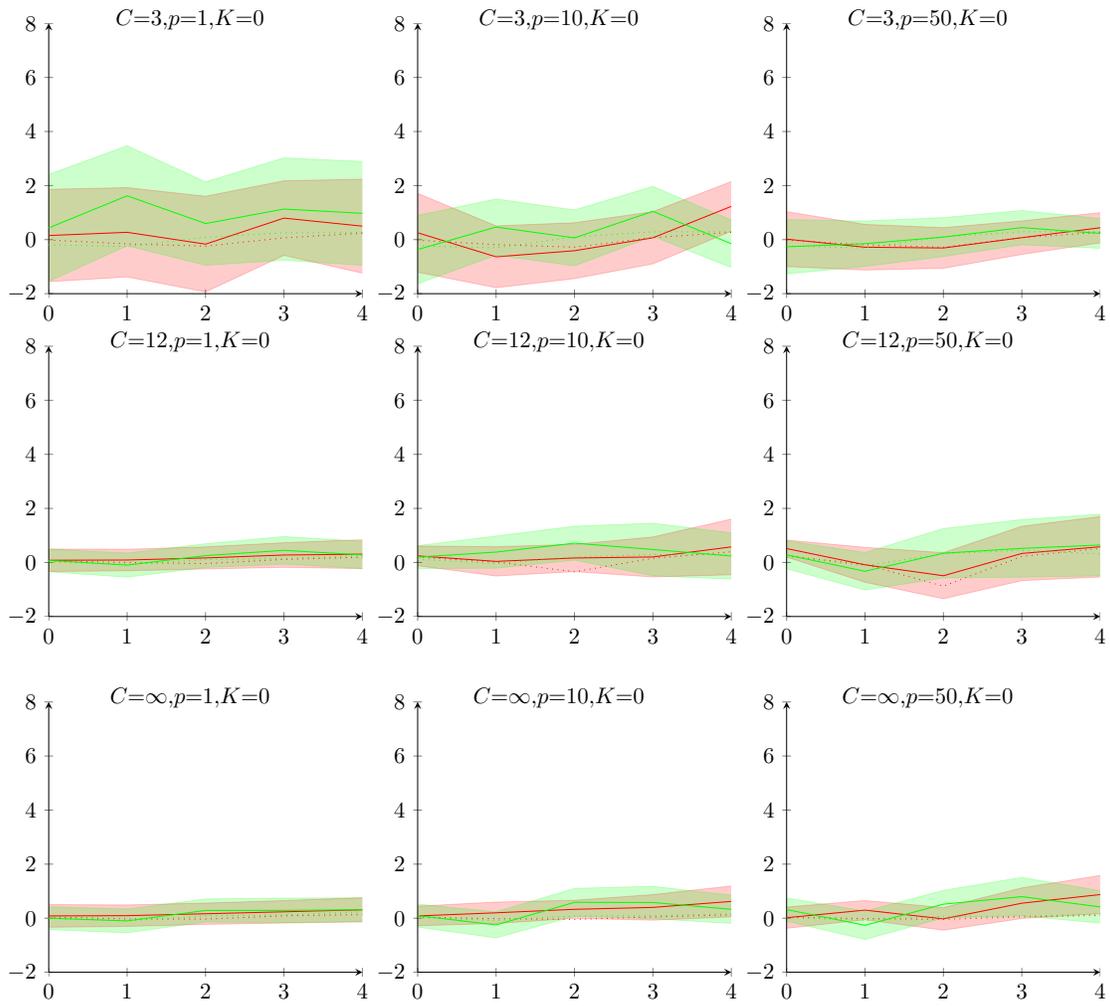


Figure 8: $K = 0$ and late bias.

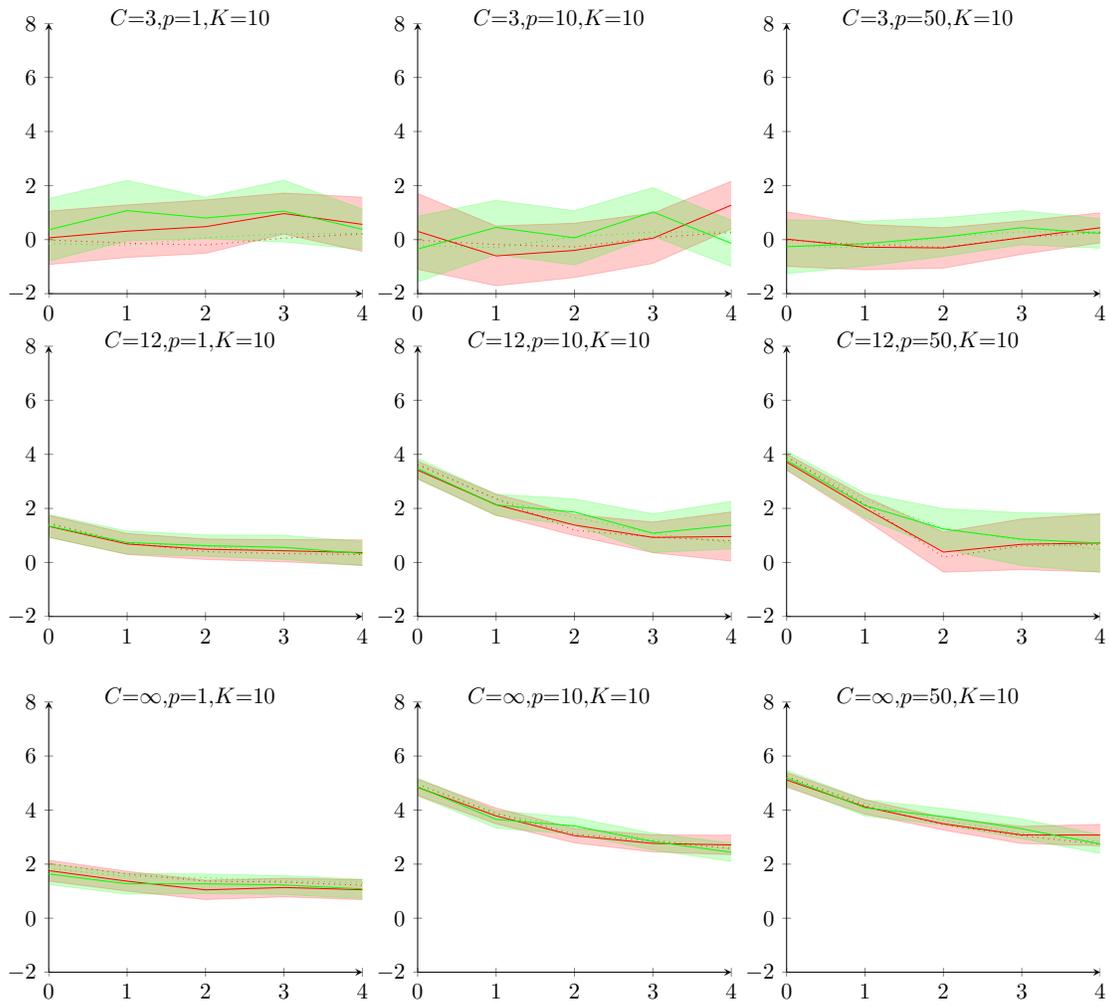


Figure 9: $K = 10$ and late bias.

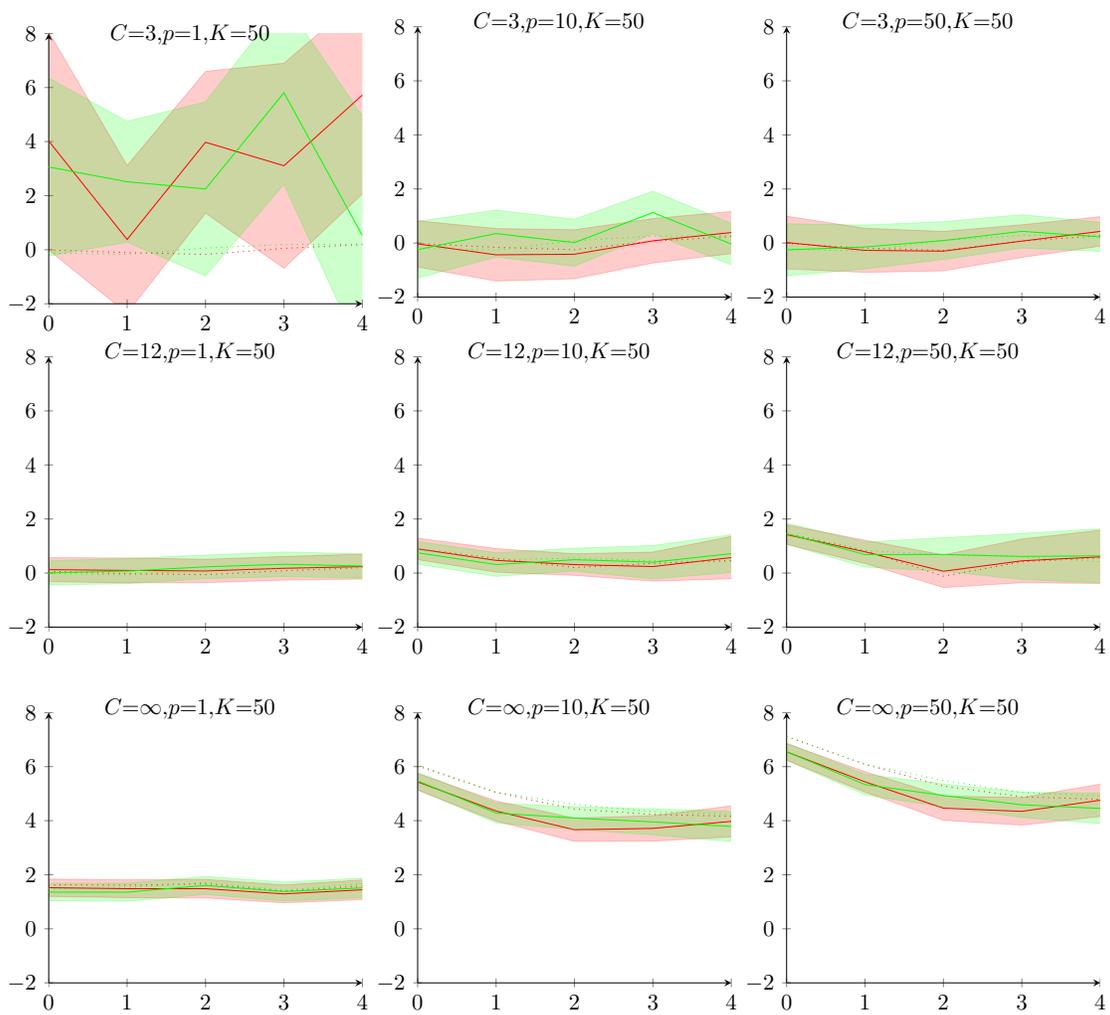


Figure 10: $K = 50$ and late bias.